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

August 27, 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.

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 diseasesb	INF	A00-B99	001–033, 034.1–1
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, 5
Liver diseases (exclude alcoholic liver disease)	LIV2	K71-K76	570, 571.4-571.9,
Renal diseases	CKD	N00-N08, N17-N19, N25-N27	580-589
Alzheimer's diseasec	AZD	F00, F01, F03, G30	290.0-290.2, 290.

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 and 20-year age groups, as well as 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 1.1), and that the uncertainty associated with such low numbers will be reflected in very wide confidence intervals for the younger ages.

```
cd /home/jimb0w/Documents/CM

import delimited "Consortium COD database v6.csv", clear
save uncleandbase, replace

set seed 3488717
use uncleandbase, clear
keep if substr(country,1,9)=="Australia"
keep if age_gp1!="" | age_gp4!=""
drop if cal < 2005
rename sex SEX
gen sex = 0 if SEX == "F"
replace sex = 1 if SEX == "M"
replace pys_nondm = pys_totpop-pys_dm
rename (alldeath_dm alldeath_nondm alldeath_totpop) (alldeath_d_dm alldeath_d_nondm alldeath_d_pop)

. ta age_gp1
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
       0-39 |         30       14.29       14.29
      40-49 |         30       14.29       28.57
      50-59 |         30       14.29       42.86
      60-69 |         30       14.29       57.14
      70-79 |         30       14.29       71.43
      80-89 |         30       14.29       85.71
        90+ |         30       14.29      100.00
------------+-----------------------------------
      Total |        210      100.00

. foreach i in alldeath can cvd chd cbd hfd res azd dmd inf flu ckd liv1 liv2 {
  2. di "`i´"
  3. ta age_gp1 if `i´_d_dm ==.
  4. gen max_`i´ = min(`i´_d_pop,5)
  5. quietly replace `i´_d_dm = runiformint(0,max_`i´) if `i´_d_dm ==.
  6. }
alldeath
no observations
can
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
       0-39 |          7      100.00      100.00
------------+-----------------------------------
      Total |          7      100.00
cvd
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
       0-39 |          6      100.00      100.00
------------+-----------------------------------
      Total |          6      100.00
chd
```

age_gp1	Freq.	Percent	Cum.
0-39	21	95.45	95.45
40-49	1	4.55	100.00
Total	22	100.00	

cbd

age_gp1	Freq.	Percent	Cum.
0-39	30	56.60	56.60
40-49	22	41.51	98.11
50-59	1	1.89	100.00
Total	53	100.00	

hfd

age_gp1	Freq.	Percent	Cum.
0-39	30	34.88	34.88
40-49	30	34.88	69.77
50-59	25	29.07	98.84
90+	1	1.16	100.00
Total	86	100.00	

res

age_gp1	Freq.	Percent	Cum.
0-39	30	47.62	47.62
40-49	30	47.62	95.24
50-59	3	4.76	100.00
Total	63	100.00	

azd

age_gp1	Freq.	Percent	Cum.
0-39	30	31.25	31.25
40-49	30	31.25	62.50
50-59	30	31.25	93.75
60-69	6	6.25	100.00
Total	96	100.00	

dmd
no observations
inf

age_gp1	Freq.	Percent	Cum.
0-39	30	58.82	58.82
40-49	17	33.33	92.16
50-59	1	1.96	94.12
90+	3	5.88	100.00
Total	51	100.00	

flu

age_gp1	Freq.	Percent	Cum.
0-39	30	35.29	35.29
40-49	30	35.29	70.59
50-59	22	25.88	96.47
60-69	3	3.53	100.00
Total	85	100.00	

ckd

age_gp1	Freq.	Percent	Cum.
0-39	30	44.12	44.12
40-49	30	44.12	88.24

```
      50-59 |          8       11.76      100.00
------------+-----------------------------------
      Total |         68      100.00
liv1
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
       0-39 |         30       46.88       46.88
      40-49 |          9       14.06       60.94
        90+ |         25       39.06      100.00
------------+-----------------------------------
      Total |         64      100.00
liv2
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
       0-39 |         30       36.14       36.14
      40-49 |         25       30.12       66.27
      50-59 |          3        3.61       69.88
        90+ |         25       30.12      100.00
------------+-----------------------------------
      Total |         83      100.00
. foreach i in alldeath can cvd chd cbd hfd res azd dmd inf flu ckd liv1 liv2 {
  2. di "`i´"
  3. count if `i´_d_dm > `i´_d_pop
  4. }
alldeath
  0
can
  0
cvd
  0
chd
  0
cbd
  0
hfd
  0
res
  0
azd
  0
dmd
  5
inf
  0
flu
  0
ckd
  0
liv1
  0
liv2
  0
. gen diff = dmd_d_dm-dmd_d_pop
. ta diff if diff >0
       diff |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |          3       60.00       60.00
          2 |          1       20.00       80.00
          5 |          1       20.00      100.00
------------+-----------------------------------
      Total |          5      100.00
. replace dmd_d_dm = dmd_d_pop if dmd_d_dm > dmd_d_pop
(5 real changes made)
```

We see that it is predominately younger age groups affected by missing data, which makes

sense. Also, there were some age groups in which the number of deaths due to diabetes among people with diabetes was greater than that recorded for the whole population. This likely has to do with differences with how we (Australian researchers) and the Australian Institute of Health and Welfare (who supplied the total population numbers) define residence in a state, or something similar. The differences were tiny, so I have just corrected the diabetes counts to not be more than the total population counts.

We should also check that the randomly generated death counts haven't produced nonsensical results. This could happen in three ways:

1. The number of deaths in each cause of death together is greater than for all causes
2. The number of deaths in CHD, CBD, and HFD together is greater than for CVD as a whole
3. The number of deaths in liver disease (excluding alcoholic liver disease) is greater than liver disease.

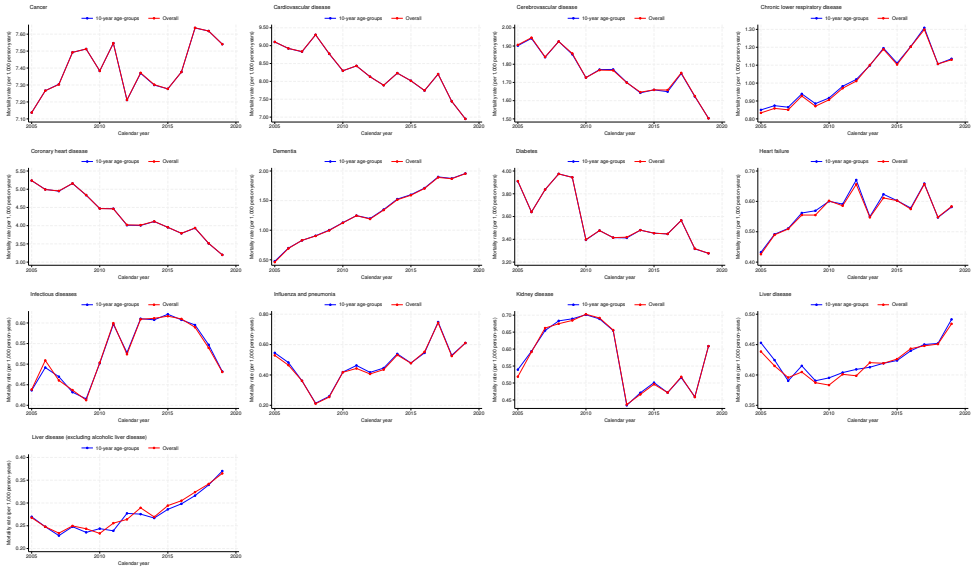
In the first case, we can just regenerate the random numbers until the error goes away; in the second, we can set the maximum number of deaths for the simulation of the other three causes of death as that for CVD (and because order matters, we will do this in the order of CVD, CBD, and HFD, based on their relative frequency in the overall population/ages where there is data); and for the third, we can set the maximum number of deaths for liver disease (excluding alcoholic liver disease) as that for overall liver diease.

```
. count if cvd_d_dm + can_d_dm + dmd_d_dm + inf_d_dm + flu_d_dm + res_d_dm + liv1_d_dm + ckd_d_dm +
> azd_d_dm > alldeath_d_dm
  0

. count if chd_d_dm + cbd_d_dm + hfd_d_dm > cvd_d_dm
  10

. ta age_gp1 if chd_d_dm + cbd_d_dm + hfd_d_dm > cvd_d_dm
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
       0-39 |         10      100.00      100.00
------------+-----------------------------------
      Total |         10      100.00

. replace max_chd = min(cvd_d_dm,5)
(5 real changes made)

. replace chd_d_dm = runiformint(0,max_chd) if chd_d_dm + cbd_d_dm + hfd_d_dm > cvd_d_dm
(9 real changes made)

. replace max_cbd = min(cvd_d_dm-chd_d_dm,5)
(12 real changes made)

. replace cbd_d_dm = runiformint(0,max_cbd) if chd_d_dm + cbd_d_dm + hfd_d_dm > cvd_d_dm
(5 real changes made)

. replace max_hfd = min(cvd_d_dm-chd_d_dm-cbd_d_dm,5)
(49 real changes made)

. replace hfd_d_dm = runiformint(0,max_hfd) if chd_d_dm + cbd_d_dm + hfd_d_dm > cvd_d_dm
(5 real changes made)

. count if chd_d_dm + cbd_d_dm + hfd_d_dm > cvd_d_dm
  0

. count if liv1_d_dm < liv2_d_dm
  27

. ta age_gp1 if liv1_d_dm < liv2_d_dm
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
       0-39 |         17       62.96       62.96
      40-49 |          2        7.41       70.37
```

```
        90+ |          8       29.63      100.00
------------+-----------------------------------
      Total |         27      100.00

. replace max_liv2 = min(liv1_d_dm,5)
(52 real changes made)

. replace liv2_d_dm = runiformint(0,max_liv2) if liv1_d_dm < liv2_d_dm
(27 real changes made)

. count if liv1_d_dm < liv2_d_dm
  0

. foreach i in alldeath can cvd chd cbd hfd res azd dmd inf flu ckd liv1 liv2 {
  2. quietly replace `i´_d_nondm = `i´_d_pop-`i´_d_dm
  3. }

. count if cvd_d_nondm + can_d_nondm + dmd_d_nondm + inf_d_nondm + flu_d_nondm + res_d_nondm + liv1_
> d_nondm + ckd_d_nondm + azd_d_nondm > alldeath_d_nondm
  0


*mkdir GPH
preserve
gen agegp = 1 if age_gp1!=""
replace agegp = 2 if age_gp4!=""
collapse (sum) pys_dm pys_nondm cvd_d_dm-azd_d_dm cvd_d_nondm-azd_d_nondm, by(calendar agegp)
foreach i in can cvd chd cbd hfd res azd dmd inf flu ckd liv1 liv2 {
if "`i´" == "can" {
local ii = "Cancer"
}
if "`i´" == "cvd" {
local ii = "Cardiovascular disease"
}
if "`i´" == "cbd" {
local ii = "Cerebrovascular disease"
}
if "`i´" == "res" {
local ii = "Chronic lower respiratory disease"
}
if "`i´" == "chd" {
local ii = "Coronary heart disease"
}
if "`i´" == "azd" {
local ii = "Dementia"
}
if "`i´" == "dmd" {
local ii = "Diabetes"
}
if "`i´" == "hfd" {
local ii = "Heart failure"
}
if "`i´" == "inf" {
local ii = "Infectious diseases"
}
if "`i´" == "flu" {
local ii = "Influenza and pneumonia"
}
if "`i´" == "ckd" {
local ii = "Kidney disease"
}
if "`i´" == "liv1" {
local ii = "Liver disease"
}
if "`i´" == "liv2" {
local ii = "Liver disease (excluding alcoholic liver disease)"
}
gen dm_`i´ = 1000*`i´_d_dm/pys_dm
twoway ///
(connected dm_`i´ cal if agegp == 1, col(blue)) ///
(connected dm_`i´ cal if agegp == 2, col(red)) ///
, graphregion(color(white)) ///
```

```
ytitle(Mortality rate (per 1,000 person-years)) ///
xtitle(Calendar year) ///
legend(order( ///
1 "10-year age-groups" ///
2 "Overall" ///
) cols(3) position(12) region(lcolor(none) color(none))) ///
ylabel(,angle(0) format(%9.2f)) ///
title("`ii´", placement(west) size(medium) col(black))
graph save GPH/dm_`i´_chk1, replace
}
restore
```

Figure 1.1: Crude mortality rate by age-grouping method, by cause of death. Australia. People with diabetes.




```
graph combine ///
GPH/dm_can_chk1.gph ///
GPH/dm_cvd_chk1.gph ///
GPH/dm_cbd_chk1.gph ///
GPH/dm_res_chk1.gph ///
GPH/dm_chd_chk1.gph ///
GPH/dm_azd_chk1.gph ///
GPH/dm_dmd_chk1.gph ///
GPH/dm_hfd_chk1.gph ///
GPH/dm_inf_chk1.gph ///
GPH/dm_flu_chk1.gph ///
GPH/dm_ckd_chk1.gph ///
GPH/dm_liv1_chk1.gph ///
GPH/dm_liv2_chk1.gph ///
, graphregion(color(white)) cols(4) altshrink
> tes.)
```

So, from Figure 1.1 we see that there doesn't appear to be any systematic issue introduced using random numbers. I will proceed using the most granular age groupings. I will assume the mid-point of the age interval for people with diabetes aged <40 is 35, for people without diabetes aged <40 is 20, and for both people with and without diabetes aged 90+ is 95.

```
keep if age_gp1!=""
```

```
replace country = substr(country,1,9)
gen age_dm = substr(age_gp1,1,2)
replace age_dm = "30" if age_dm == "0-"
destring age_dm, replace
replace age_dm = age_dm+5
gen age_nondm = substr(age_gp1,1,2)
replace age_nondm = "15" if age_nondm == "0-"
destring age_nondm, replace
replace age_nondm = age_nondm+5
keep country calendar sex alldeath_d_dm alldeath_d_nondm age_dm age_nondm pys_dm pys_nondm cvd_d_dm-
> azd_d_dm cvd_d_nondm-azd_d_nondm
save Australia, replace
```

1.2 Canada (Alberta)

For Canada (Alberta), we have the following variables (by age, sex, and calendar year): total population size, prevalence of diabetes, incidence of 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]. We can calculate person-years in people with diabetes, in a given calendar year, by adding the number of people with prevalent diabetes to half the number of people with incident diabetes and subtracting half the number of all-cause deaths [again, 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.

Canadian data restrictions prohibit the use of any cell count between 1 and 9 for people with diabetes and in the total population; thus, there are many blank values (see below). I will fill them in randomly, where the number can be any number from 1 to 9 with equal probability, unless the number of deaths in the total population for the age/sex group is <9 (after being randomly generated), in which case the upper bound will be the number of deaths in the total population.

Sense checks will be as for Australia, above.

```
use uncleandbase, clear
keep if substr(country,1,6)=="Canada"
rename sex SEX
gen sex = 0 if SEX == "F"
replace sex = 1 if SEX == "M"
replace pys_nondm = pys_totpop-pys_dm
rename (alldeath_dm alldeath_nondm alldeath_totpop) (alldeath_d_dm alldeath_d_nondm alldeath_d_pop)

. ta age_gp1
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
       0-39 |         32       14.29       14.29
      40-49 |         32       14.29       28.57
      50-59 |         32       14.29       42.86
      60-69 |         32       14.29       57.14
      70-79 |         32       14.29       71.43
      80-89 |         32       14.29       85.71
        90+ |         32       14.29      100.00
------------+-----------------------------------
      Total |        224      100.00

. foreach i in alldeath can cvd chd cbd hfd res azd dmd inf flu ckd liv1 liv2 {
  2. di "`i´"
  3. ta age_gp1 if `i´_d_pop ==.
  4. gen min_`i´ = max(`i´_d_dm,1) if `i´_d_dm!=.
  5. replace min_`i´ = 1 if `i´_d_dm==.
  6. replace `i´_d_dm=0 if `i´_d_pop==0
  7. quietly replace `i´_d_pop = runiformint(min_`i´,9) if `i´_d_pop==.
  8. ta age_gp1 if `i´_d_dm ==.
  9. gen max_`i´ = min(`i´_d_pop,9)
 10. quietly replace `i´_d_dm = runiformint(1,max_`i´) if `i´_d_dm ==.
 11. }
alldeath
no observations
(0 real changes made)
(0 real changes made)
no observations
can
no observations
(12 missing values generated)
```

```
(12 real changes made)
(0 real changes made)
```

age_gp1	Freq.	Percent	Cum.
0-39	6	50.00	50.00
40-49	6	50.00	100.00
Total	12	100.00	

```
cvd
no observations
(14 missing values generated)
(14 real changes made)
(0 real changes made)
```

age_gp1	Freq.	Percent	Cum.
0-39	7	50.00	50.00
40-49	7	50.00	100.00
Total	14	100.00	

```
chd
```

age_gp1	Freq.	Percent	Cum.
0-39	2	50.00	50.00
40-49	2	50.00	100.00
Total	4	100.00	

```
(24 missing values generated)
(24 real changes made)
(0 real changes made)
```

age_gp1	Freq.	Percent	Cum.
0-39	12	50.00	50.00
40-49	12	50.00	100.00
Total	24	100.00	

```
cbd
```

age_gp1	Freq.	Percent	Cum.
0-39	11	50.00	50.00
40-49	11	50.00	100.00
Total	22	100.00	

```
(62 missing values generated)
(62 real changes made)
(0 real changes made)
```

age_gp1	Freq.	Percent	Cum.
0-39	12	19.35	19.35
40-49	23	37.10	56.45
50-59	23	37.10	93.55
60-69	4	6.45	100.00
Total	62	100.00	

```
hfd
```

age_gp1	Freq.	Percent	Cum.
0-39	12	17.65	17.65
40-49	10	14.71	32.35
50-59	25	36.76	69.12
60-69	21	30.88	100.00
Total	68	100.00	

```
(61 missing values generated)
(61 real changes made)
(0 real changes made)
```

age_gp1	Freq.	Percent	Cum.

```
       0-39 |          1        1.64        1.64
      40-49 |          3        4.92        6.56
      50-59 |         23       37.70       44.26
      60-69 |         26       42.62       86.89
      70-79 |          3        4.92       91.80
        90+ |          5        8.20      100.00
------------+-----------------------------------
      Total |         61      100.00

res
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
       0-39 |         29       50.00       50.00
      40-49 |         29       50.00      100.00
------------+-----------------------------------
      Total |         58      100.00
(50 missing values generated)
(50 real changes made)
(0 real changes made)
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
       0-39 |          5       10.00       10.00
      40-49 |         18       36.00       46.00
      50-59 |         23       46.00       92.00
        90+ |          4        8.00      100.00
------------+-----------------------------------
      Total |         50      100.00

azd
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
       0-39 |          3       17.65       17.65
      40-49 |          3       17.65       35.29
      50-59 |          8       47.06       82.35
      60-69 |          3       17.65      100.00
------------+-----------------------------------
      Total |         17      100.00
(25 missing values generated)
(25 real changes made)
(0 real changes made)
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
      50-59 |         10       40.00       40.00
      60-69 |         12       48.00       88.00
      70-79 |          1        4.00       92.00
        90+ |          2        8.00      100.00
------------+-----------------------------------
      Total |         25      100.00

dmd
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
       0-39 |         13       48.15       48.15
      40-49 |         13       48.15       96.30
      50-59 |          1        3.70      100.00
------------+-----------------------------------
      Total |         27      100.00
(37 missing values generated)
(37 real changes made)
(0 real changes made)
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
       0-39 |         17       45.95       45.95
      40-49 |         18       48.65       94.59
      50-59 |          2        5.41      100.00
------------+-----------------------------------
      Total |         37      100.00

inf
```

age_gp1	Freq.	Percent	Cum.
0-39	14	45.16	45.16
40-49	14	45.16	90.32
50-59	3	9.68	100.00
Total	31	100.00	

(104 missing values generated)
(104 real changes made)
(0 real changes made)

age_gp1	Freq.	Percent	Cum.
0-39	13	12.50	12.50
40-49	26	25.00	37.50
50-59	28	26.92	64.42
60-69	11	10.58	75.00
70-79	4	3.85	78.85
80-89	3	2.88	81.73
90+	19	18.27	100.00
Total	104	100.00	

flu

age_gp1	Freq.	Percent	Cum.
0-39	24	45.28	45.28
40-49	24	45.28	90.57
50-59	5	9.43	100.00
Total	53	100.00	

(95 missing values generated)
(95 real changes made)
(0 real changes made)

age_gp1	Freq.	Percent	Cum.
0-39	14	14.74	14.74
40-49	22	23.16	37.89
50-59	28	29.47	67.37
60-69	22	23.16	90.53
70-79	7	7.37	97.89
90+	2	2.11	100.00
Total	95	100.00	

ckd

age_gp1	Freq.	Percent	Cum.
0-39	26	27.37	27.37
40-49	27	28.42	55.79
50-59	31	32.63	88.42
60-69	11	11.58	100.00
Total	95	100.00	

(96 missing values generated)
(96 real changes made)
(0 real changes made)

age_gp1	Freq.	Percent	Cum.
0-39	10	10.42	10.42
40-49	15	15.62	26.04
50-59	27	28.12	54.17
60-69	24	25.00	79.17
70-79	4	4.17	83.33
90+	16	16.67	100.00
Total	96	100.00	

liv1

age_gp1	Freq.	Percent	Cum.

```
       0-39 |         16       48.48       48.48
      80-89 |          1        3.03       51.52
        90+ |         16       48.48      100.00
------------+-----------------------------------
      Total |         33      100.00
(113 missing values generated)
(113 real changes made)
(0 real changes made)
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
       0-39 |         19       16.81       16.81
      40-49 |         28       24.78       41.59
      50-59 |         10        8.85       50.44
      60-69 |          3        2.65       53.10
      70-79 |          7        6.19       59.29
      80-89 |         24       21.24       80.53
        90+ |         22       19.47      100.00
------------+-----------------------------------
      Total |        113      100.00
liv2
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
       0-39 |         28       33.33       33.33
      40-49 |         26       30.95       64.29
      50-59 |          2        2.38       66.67
      60-69 |          1        1.19       67.86
      80-89 |          4        4.76       72.62
        90+ |         23       27.38      100.00
------------+-----------------------------------
      Total |         84      100.00
(126 missing values generated)
(126 real changes made)
(0 real changes made)
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
       0-39 |         11        8.73        8.73
      40-49 |         22       17.46       26.19
      50-59 |         26       20.63       46.83
      60-69 |          9        7.14       53.97
      70-79 |         11        8.73       62.70
      80-89 |         25       19.84       82.54
        90+ |         22       17.46      100.00
------------+-----------------------------------
      Total |        126      100.00
. foreach i in alldeath can cvd chd cbd hfd res azd dmd inf flu ckd liv1 liv2 {
  2. di "`i´"
  3. count if `i´_d_dm > `i´_d_pop
  4. }
alldeath
  0
can
  0
cvd
  0
chd
  0
cbd
  0
hfd
  0
res
  0
azd
  0
dmd
  0
```

```
inf
  0
flu
  0
ckd
  0
liv1
  0
liv2
  0
. count if cvd_d_pop + can_d_pop + dmd_d_pop + inf_d_pop + flu_d_pop + res_d_pop + liv1_d_pop + ckd_
> d_pop + azd_d_pop > alldeath_d_pop
  0
. count if cvd_d_dm + can_d_dm + dmd_d_dm + inf_d_dm + flu_d_dm + res_d_dm + liv1_d_dm + ckd_d_dm +
> azd_d_dm > alldeath_d_dm
  3
. quietly {
. count if chd_d_pop + cbd_d_pop + hfd_d_pop > cvd_d_pop
  0
. count if chd_d_dm + cbd_d_dm + hfd_d_dm > cvd_d_dm
  20
. ta age_gp1 if chd_d_dm + cbd_d_dm + hfd_d_dm > cvd_d_dm
```

age_gp1	Freq.	Percent	Cum.
0-39	9	45.00	45.00
40-49	10	50.00	95.00
50-59	1	5.00	100.00
Total	20	100.00	

```
. replace max_chd = min(cvd_d_dm,9)
(38 real changes made)
. replace chd_d_dm = runiformint(1,max_chd) if chd_d_dm + cbd_d_dm + hfd_d_dm > cvd_d_dm & inrange(c
> hd_d_dm,1,9)
(13 real changes made)
. replace max_cbd = min(cvd_d_dm-chd_d_dm,9)
(60 real changes made)
. replace cbd_d_dm = runiformint(0,max_cbd) if chd_d_dm + cbd_d_dm + hfd_d_dm > cvd_d_dm & inrange(c
> bd_d_dm,1,9)
(16 real changes made)
. replace max_hfd = min(cvd_d_dm-chd_d_dm-cbd_d_dm,9)
(102 real changes made)
. replace hfd_d_dm = runiformint(0,max_hfd) if chd_d_dm + cbd_d_dm + hfd_d_dm > cvd_d_dm & inrange(h
> fd_d_dm,1,9)
(1 real change made)
. count if chd_d_dm + cbd_d_dm + hfd_d_dm > cvd_d_dm
  0
. count if liv1_d_dm < liv2_d_dm
  32
. ta age_gp1 if liv1_d_dm < liv2_d_dm
```

age_gp1	Freq.	Percent	Cum.
0-39	5	15.62	15.62
40-49	8	25.00	40.62
60-69	2	6.25	46.88
80-89	8	25.00	71.88
90+	9	28.12	100.00
Total	32	100.00	

```
. replace max_liv2 = min(liv1_d_dm,9)
(124 real changes made)
. replace liv2_d_dm = runiformint(1,max_liv2) if liv1_d_dm < liv2_d_dm & inrange(liv2_d_dm,1,9)
(32 real changes made)
```

```
. count if liv1_d_dm < liv2_d_dm
  0
. foreach i in alldeath can cvd chd cbd hfd res azd dmd inf flu ckd liv1 liv2 {
  2. quietly replace `i´_d_nondm = `i´_d_pop-`i´_d_dm
  3. }
. count if cvd_d_nondm + can_d_nondm + dmd_d_nondm + inf_d_nondm + flu_d_nondm + res_d_nondm + liv1_
> d_nondm + ckd_d_nondm + azd_d_nondm > alldeath_d_nondm
  0


keep if age_gp1!=""
replace country = "Canada"
gen age_dm = substr(age_gp1,1,2)
replace age_dm = "30" if age_dm == "0-"
destring age_dm, replace
replace age_dm = age_dm+5
gen age_nondm = substr(age_gp1,1,2)
replace age_nondm = "15" if age_nondm == "0-"
destring age_nondm, replace
replace age_nondm = age_nondm+5
keep country calendar sex alldeath_d_dm alldeath_d_nondm age_dm age_nondm pys_dm pys_nondm cvd_d_dm-
> azd_d_dm cvd_d_nondm-azd_d_nondm
save Canada, replace
```

1.3 Denmark

For Denmark, we have the following variables (by age, sex, and calendar year): Person-years and deaths in people with and without diabetes. I.e., no further variables need to be derived. Denmark restricts counts between 1 and 3 for both people with and without diabetes. I will fill them in randomly, where the number can be any number from 1 to 3 with equal probability. I will assume the mid-point of the age interval for people aged <40 is 35 and for 90+ is 95.

```
use uncleandbase, clear
keep if substr(country,1,7)=="Denmark"
rename sex SEX
gen sex = 0 if SEX == "F"
replace sex = 1 if SEX == "M"
rename (alldeath_dm alldeath_nondm alldeath_totpop) (alldeath_d_dm alldeath_d_nondm alldeath_d_pop)
```

```
. ta age_gp1
```

age_gp1	Freq.	Percent	Cum.
0-39	36	14.29	14.29
40-49	36	14.29	28.57
50-59	36	14.29	42.86
60-69	36	14.29	57.14
70-79	36	14.29	71.43
80-89	36	14.29	85.71
90+	36	14.29	100.00
Total	252	100.00	

```
. foreach i in alldeath can cvd chd cbd hfd res azd dmd inf flu ckd liv1 liv2 {
  2. di "`i´"
  3. ta age_gp1 if `i´_d_nondm ==.
  4. quietly replace `i´_d_nondm = runiformint(1,3) if `i´_d_nondm==.
  5. ta age_gp1 if `i´_d_dm ==.
  6. quietly replace `i´_d_dm = runiformint(1,3) if `i´_d_dm ==.
  7. }
alldeath
no observations
no observations
can
no observations
```

age_gp1	Freq.	Percent	Cum.
0-39	23	100.00	100.00
Total	23	100.00	

```
cvd
no observations
```

age_gp1	Freq.	Percent	Cum.
0-39	24	96.00	96.00
40-49	1	4.00	100.00
Total	25	100.00	

```
chd
```

age_gp1	Freq.	Percent	Cum.
0-39	12	100.00	100.00
Total	12	100.00	

age_gp1	Freq.	Percent	Cum.
0-39	18	58.06	58.06
40-49	13	41.94	100.00

```
      Total |         31      100.00
cbd
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
       0-39 |          3      100.00      100.00
------------+-----------------------------------
      Total |          3      100.00
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
       0-39 |         11       29.73       29.73
      40-49 |         26       70.27      100.00
------------+-----------------------------------
      Total |         37      100.00
hfd
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
       0-39 |         14       31.82       31.82
      40-49 |         22       50.00       81.82
      50-59 |          8       18.18      100.00
------------+-----------------------------------
      Total |         44      100.00
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
       0-39 |          6       13.95       13.95
      40-49 |         15       34.88       48.84
      50-59 |         21       48.84       97.67
      60-69 |          1        2.33      100.00
------------+-----------------------------------
      Total |         43      100.00
res
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
       0-39 |         25       92.59       92.59
      40-49 |          2        7.41      100.00
------------+-----------------------------------
      Total |         27      100.00
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
       0-39 |          5       14.29       14.29
      40-49 |         28       80.00       94.29
      50-59 |          2        5.71      100.00
------------+-----------------------------------
      Total |         35      100.00
azd
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
       0-39 |          3       10.00       10.00
      40-49 |         13       43.33       53.33
      50-59 |         14       46.67      100.00
------------+-----------------------------------
      Total |         30      100.00
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
      40-49 |          1        3.85        3.85
      50-59 |          8       30.77       34.62
      60-69 |         16       61.54       96.15
        90+ |          1        3.85      100.00
------------+-----------------------------------
      Total |         26      100.00
dmd
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
       0-39 |          8       12.12       12.12
      40-49 |         21       31.82       43.94
      50-59 |         17       25.76       69.70
```

```
      60-69 |         13       19.70       89.39
      70-79 |          1        1.52       90.91
        90+ |          6        9.09      100.00
------------+-----------------------------------
      Total |         66      100.00
```

```
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
       0-39 |         11       91.67       91.67
      40-49 |          1        8.33      100.00
------------+-----------------------------------
      Total |         12      100.00
```

inf

```
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
       0-39 |         16       88.89       88.89
      40-49 |          2       11.11      100.00
------------+-----------------------------------
      Total |         18      100.00
```

```
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
       0-39 |         10       19.23       19.23
      40-49 |         20       38.46       57.69
      50-59 |         18       34.62       92.31
      60-69 |          2        3.85       96.15
        90+ |          2        3.85      100.00
------------+-----------------------------------
      Total |         52      100.00
```

flu

```
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
       0-39 |         22       84.62       84.62
      40-49 |          4       15.38      100.00
------------+-----------------------------------
      Total |         26      100.00
```

```
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
       0-39 |          8       15.09       15.09
      40-49 |         18       33.96       49.06
      50-59 |         23       43.40       92.45
      60-69 |          4        7.55      100.00
------------+-----------------------------------
      Total |         53      100.00
```

ckd

```
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
       0-39 |         20       37.74       37.74
      40-49 |         25       47.17       84.91
      50-59 |          8       15.09      100.00
------------+-----------------------------------
      Total |         53      100.00
```

```
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
       0-39 |          3        4.92        4.92
      40-49 |         11       18.03       22.95
      50-59 |         24       39.34       62.30
      60-69 |         11       18.03       80.33
      70-79 |          2        3.28       83.61
        90+ |         10       16.39      100.00
------------+-----------------------------------
      Total |         61      100.00
```

liv1

```
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
       0-39 |          7       29.17       29.17
        90+ |         17       70.83      100.00
------------+-----------------------------------
```

```
      Total |          24      100.00
```

age_gp1	Freq.	Percent	Cum.
0-39	15	28.30	28.30
40-49	12	22.64	50.94
80-89	11	20.75	71.70
90+	15	28.30	100.00
Total	53	100.00	

```
liv2
```

age_gp1	Freq.	Percent	Cum.
0-39	20	33.90	33.90
40-49	19	32.20	66.10
50-59	3	5.08	71.19
90+	17	28.81	100.00
Total	59	100.00	

age_gp1	Freq.	Percent	Cum.
0-39	4	4.82	4.82
40-49	12	14.46	19.28
50-59	23	27.71	46.99
60-69	6	7.23	54.22
70-79	7	8.43	62.65
80-89	18	21.69	84.34
90+	13	15.66	100.00
Total	83	100.00	

```
. count if cvd_d_nondm + can_d_nondm + dmd_d_nondm + inf_d_nondm + flu_d_nondm + res_d_nondm + liv1_
> d_nondm + ckd_d_nondm + azd_d_nondm > alldeath_d_nondm
  0

. count if cvd_d_dm + can_d_dm + dmd_d_dm + inf_d_dm + flu_d_dm + res_d_dm + liv1_d_dm + ckd_d_dm +
> azd_d_dm > alldeath_d_dm
  1

. count if chd_d_nondm + cbd_d_nondm + hfd_d_nondm > cvd_d_nondm
  0

. count if chd_d_dm + cbd_d_dm + hfd_d_dm > cvd_d_dm
  6

. ta age_gp1 if chd_d_dm + cbd_d_dm + hfd_d_dm > cvd_d_dm
```

age_gp1	Freq.	Percent	Cum.
0-39	5	83.33	83.33
40-49	1	16.67	100.00
Total	6	100.00	

```
. gen max_chd = min(cvd_d_dm,3)

. replace chd_d_dm = runiformint(1,max_chd) if chd_d_dm + cbd_d_dm + hfd_d_dm > cvd_d_dm & inrange(c
> hd_d_dm,1,3)
(4 real changes made)

. gen max_cbd = min(cvd_d_dm-chd_d_dm,3)

. replace cbd_d_dm = runiformint(0,max_cbd) if chd_d_dm + cbd_d_dm + hfd_d_dm > cvd_d_dm & inrange(c
> bd_d_dm,1,3)
(2 real changes made)

. gen max_hfd = min(cvd_d_dm-chd_d_dm-cbd_d_dm,3)

. replace hfd_d_dm = runiformint(0,max_hfd) if chd_d_dm + cbd_d_dm + hfd_d_dm > cvd_d_dm & inrange(h
> fd_d_dm,1,3)
(1 real change made)

. count if chd_d_dm + cbd_d_dm + hfd_d_dm > cvd_d_dm
  0

. count if liv1_d_nondm < liv2_d_nondm
  8
```

```
. ta age_gp1 if liv1_d_nondm < liv2_d_nondm
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
       0-39 |          1       12.50       12.50
        90+ |          7       87.50      100.00
------------+-----------------------------------
      Total |          8      100.00

. gen max_liv2 = min(liv1_d_nondm,3)

. replace liv2_d_nondm = runiformint(1,max_liv2) if liv1_d_nondm < liv2_d_nondm & inrange(liv2_d_non
> dm,1,3)
(8 real changes made)

. count if liv1_d_nondm < liv2_d_nondm
  0

. count if liv1_d_dm < liv2_d_dm
  9

. ta age_gp1 if liv1_d_dm < liv2_d_dm
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
      80-89 |          4       44.44       44.44
        90+ |          5       55.56      100.00
------------+-----------------------------------
      Total |          9      100.00

. replace max_liv2 = min(liv1_d_dm,3)
(73 real changes made)

. replace liv2_d_dm = runiformint(1,max_liv2) if liv1_d_dm < liv2_d_dm & inrange(liv2_d_dm,1,3)
(9 real changes made)

. count if liv1_d_dm < liv2_d_dm
  0
```

```
gen age_dm = substr(age_gp1,1,2)
replace age_dm = "30" if age_dm == "0-"
destring age_dm, replace
replace age_dm = age_dm+5
gen age_nondm = substr(age_gp1,1,2)
replace age_nondm = "15" if age_nondm == "0-"
destring age_nondm, replace
replace age_nondm = age_nondm+5
replace country = "Denmark"
keep country calendar sex alldeath_d_dm alldeath_d_nondm age_dm age_nondm pys_dm pys_nondm cvd_d_dm-
> azd_d_dm cvd_d_nondm-azd_d_nondm
save Denmark, replace
```

1.4 Finland

For Finland, we have the following variables (by age, sex, and calendar year): Person-years and deaths in people with and without diabetes. I.e., no further variables need to be derived. Finland restricts counts between 1 and 5 for both people with and without diabetes. I will fill them in randomly, where the number can be any number from 1 to 5 with equal probability. I will assume the mid-point of the age interval for people aged <40 is 35 and for 90+ is 95.

```
use uncleandbase, clear
keep if substr(country,1,7)=="Finland"
rename sex SEX
gen sex = 0 if SEX == "F"
replace sex = 1 if SEX == "M"
rename (alldeath_dm alldeath_nondm alldeath_totpop) (alldeath_d_dm alldeath_d_nondm alldeath_d_pop)

. ta age_gp1
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
       0-39 |         36       14.29       14.29
      40-49 |         36       14.29       28.57
      50-59 |         36       14.29       42.86
      60-69 |         36       14.29       57.14
      70-79 |         36       14.29       71.43
      80-89 |         36       14.29       85.71
        90+ |         36       14.29      100.00
------------+-----------------------------------
      Total |        252      100.00

. foreach i in alldeath can cvd chd cbd hfd res azd dmd inf flu ckd liv1 liv2 {
  2. di "`i´"
  3. ta age_gp1 if `i´_d_nondm ==.
  4. quietly replace `i´_d_nondm = runiformint(1,5) if `i´_d_nondm==.
  5. ta age_gp1 if `i´_d_dm ==.
  6. quietly replace `i´_d_dm = runiformint(1,5) if `i´_d_dm ==.
  7. }
alldeath
no observations
no observations
can
no observations
no observations
cvd
no observations
no observations
chd
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
       0-39 |          8       88.89       88.89
      40-49 |          1       11.11      100.00
------------+-----------------------------------
      Total |          9      100.00

    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
       0-39 |          7       87.50       87.50
      40-49 |          1       12.50      100.00
------------+-----------------------------------
      Total |          8      100.00

cbd
no observations
no observations
hfd
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
       0-39 |          5       12.50       12.50
      40-49 |          5       12.50       25.00
```

```
      50-59 |         15       37.50       62.50
      60-69 |         12       30.00       92.50
      70-79 |          1        2.50       95.00
      80-89 |          1        2.50       97.50
        90+ |          1        2.50      100.00
------------+-----------------------------------
      Total |         40      100.00
```

age_gp1	Freq.	Percent	Cum.
40-49	1	3.12	3.12
50-59	14	43.75	46.88
60-69	14	43.75	90.62
70-79	1	3.12	93.75
80-89	1	3.12	96.88
90+	1	3.12	100.00
Total	32	100.00	

res

age_gp1	Freq.	Percent	Cum.
0-39	24	57.14	57.14
40-49	18	42.86	100.00
Total	42	100.00	

age_gp1	Freq.	Percent	Cum.
0-39	3	33.33	33.33
40-49	6	66.67	100.00
Total	9	100.00	

azd

age_gp1	Freq.	Percent	Cum.
40-49	8	47.06	47.06
50-59	9	52.94	100.00
Total	17	100.00	

age_gp1	Freq.	Percent	Cum.
50-59	4	100.00	100.00
Total	4	100.00	

dmd

no observations

age_gp1	Freq.	Percent	Cum.
0-39	1	50.00	50.00
40-49	1	50.00	100.00
Total	2	100.00	

inf

age_gp1	Freq.	Percent	Cum.
0-39	6	30.00	30.00
40-49	12	60.00	90.00
50-59	2	10.00	100.00
Total	20	100.00	

age_gp1	Freq.	Percent	Cum.
0-39	3	25.00	25.00
40-49	7	58.33	83.33
50-59	2	16.67	100.00
Total	12	100.00	

flu

age_gp1	Freq.	Percent	Cum.
0-39	10	38.46	38.46
40-49	10	38.46	76.92
50-59	5	19.23	96.15
60-69	1	3.85	100.00
Total	26	100.00	

age_gp1	Freq.	Percent	Cum.
0-39	1	11.11	11.11
40-49	3	33.33	44.44
50-59	4	44.44	88.89
60-69	1	11.11	100.00
Total	9	100.00	

ckd

age_gp1	Freq.	Percent	Cum.
0-39	18	24.00	24.00
40-49	15	20.00	44.00
50-59	22	29.33	73.33
60-69	19	25.33	98.67
90+	1	1.33	100.00
Total	75	100.00	

age_gp1	Freq.	Percent	Cum.
0-39	3	8.11	8.11
40-49	4	10.81	18.92
50-59	12	32.43	51.35
60-69	16	43.24	94.59
90+	2	5.41	100.00
Total	37	100.00	

liv1

age_gp1	Freq.	Percent	Cum.
0-39	3	12.50	12.50
90+	21	87.50	100.00
Total	24	100.00	

age_gp1	Freq.	Percent	Cum.
0-39	1	5.88	5.88
90+	16	94.12	100.00
Total	17	100.00	

liv2

age_gp1	Freq.	Percent	Cum.
0-39	20	27.03	27.03
40-49	27	36.49	63.51
50-59	7	9.46	72.97
60-69	1	1.35	74.32
80-89	1	1.35	75.68
90+	18	24.32	100.00
Total	74	100.00	

age_gp1	Freq.	Percent	Cum.
0-39	2	5.26	5.26
40-49	12	31.58	36.84
50-59	6	15.79	52.63
60-69	1	2.63	55.26

```
      80-89 |          1        2.63       57.89
        90+ |         16       42.11      100.00
------------+-----------------------------------
      Total |         38      100.00
```

```
. count if cvd_d_nondm + can_d_nondm + dmd_d_nondm + inf_d_nondm + flu_d_nondm + res_d_nondm + liv1_
> d_nondm + ckd_d_nondm + azd_d_nondm > alldeath_d_nondm
  0

. count if cvd_d_dm + can_d_dm + dmd_d_dm + inf_d_dm + flu_d_dm + res_d_dm + liv1_d_dm + ckd_d_dm +
> azd_d_dm > alldeath_d_dm
  0

.
. count if chd_d_nondm + cbd_d_nondm + hfd_d_nondm > cvd_d_nondm
  6

. ta age_gp1 if chd_d_nondm + cbd_d_nondm + hfd_d_nondm > cvd_d_nondm
```

```
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
       0-39 |          2       33.33       33.33
      40-49 |          1       16.67       50.00
        90+ |          3       50.00      100.00
------------+-----------------------------------
      Total |          6      100.00
```

```
. gen max_chd = min(cvd_d_nondm,5)

. replace chd_d_nondm = runiformint(1,max_chd) if chd_d_nondm + cbd_d_nondm + hfd_d_nondm > cvd_d_no
> ndm & inrange(chd_d_nondm,1,5)
(1 real change made)

. gen max_cbd = min(cvd_d_nondm-chd_d_nondm,5)

. replace cbd_d_nondm = runiformint(0,max_cbd) if chd_d_nondm + cbd_d_nondm + hfd_d_nondm > cvd_d_no
> ndm & inrange(cbd_d_nondm,1,5)
(0 real changes made)

. gen max_hfd = min(cvd_d_nondm-chd_d_nondm-cbd_d_nondm,5)

. replace hfd_d_nondm = runiformint(0,max_hfd) if chd_d_nondm + cbd_d_nondm + hfd_d_nondm > cvd_d_no
> ndm & inrange(hfd_d_nondm,1,5)
(2 real changes made)

. count if chd_d_nondm + cbd_d_nondm + hfd_d_nondm > cvd_d_nondm
  3
```

This isn't an error with random number generation; the data itself is faulty.

```
. count if liv1_d_nondm < liv2_d_nondm
  6

. ta age_gp1 if liv1_d_nondm < liv2_d_nondm
```

```
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
        90+ |          6      100.00      100.00
------------+-----------------------------------
      Total |          6      100.00
```

```
. gen max_liv2 = min(liv1_d_nondm,5)

. replace liv2_d_nondm = runiformint(1,max_liv2) if liv1_d_nondm < liv2_d_nondm & inrange(liv2_d_non
> dm,1,5)
(6 real changes made)

. count if liv1_d_nondm < liv2_d_nondm
  0

. count if liv1_d_dm < liv2_d_dm
  7

. ta age_gp1 if liv1_d_dm < liv2_d_dm
```

```
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
       0-39 |          1       14.29       14.29
      40-49 |          1       14.29       28.57
        90+ |          5       71.43      100.00
------------+-----------------------------------
```

```
      Total |          7      100.00
. replace max_liv2 = min(liv1_d_dm,5)
(73 real changes made)
. replace liv2_d_dm = runiformint(1,max_liv2) if liv1_d_dm < liv2_d_dm & inrange(liv2_d_dm,1,5)
(7 real changes made)
. count if liv1_d_dm < liv2_d_dm
  0


gen age_dm = substr(age_gp1,1,2)
replace age_dm = "30" if age_dm == "0-"
destring age_dm, replace
replace age_dm = age_dm+5
gen age_nondm = substr(age_gp1,1,2)
replace age_nondm = "15" if age_nondm == "0-"
destring age_nondm, replace
replace age_nondm = age_nondm+5
keep country calendar sex alldeath_d_dm alldeath_d_nondm age_dm age_nondm pys_dm pys_nondm cvd_d_dm-
> azd_d_dm cvd_d_nondm-azd_d_nondm
save Finland, replace
```

1.5 France

For France, we have the following variables (by age, sex, and calendar year): person-years and deaths in people with and without diabetes. I.e., no further variables need to be derived. France has excluded counts between 1 and 4, which I will fill in randomly. I will assume the mid-point of the age interval for people aged <40 is 35 and for 90+ is 95.

```
use uncleandbase, clear
keep if substr(country,1,8)=="France_1"
rename sex SEX
gen sex = 0 if SEX == "F"
replace sex = 1 if SEX == "M"
rename (alldeath_dm alldeath_nondm alldeath_totpop) (alldeath_d_dm alldeath_d_nondm alldeath_d_pop)
recode dmd_d_nondm .=0
```

```
. ta age_gp1
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
      0-39  |         16       14.29       14.29
      40-49 |         16       14.29       28.57
      50-59 |         16       14.29       42.86
      60-69 |         16       14.29       57.14
      70-79 |         16       14.29       71.43
      80-89 |         16       14.29       85.71
        90+ |         16       14.29      100.00
------------+-----------------------------------
      Total |        112      100.00
. foreach i in alldeath can cvd chd cbd hfd res azd dmd inf flu ckd liv1 liv2 {
  2. di "`i´"
  3. ta age_gp1 if `i´_d_nondm ==.
  4. quietly replace `i´_d_nondm = runiformint(1,4) if `i´_d_nondm==.
  5. ta age_gp1 if `i´_d_dm ==.
  6. quietly replace `i´_d_dm = runiformint(1,4) if `i´_d_dm ==.
  7. }
alldeath
no observations
no observations
can
no observations
no observations
cvd
no observations
no observations
chd
no observations
no observations
cbd
no observations
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
      0-39  |          1       50.00       50.00
      40-49 |          1       50.00      100.00
------------+-----------------------------------
      Total |          2      100.00
hfd
no observations
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
      0-39  |          3       50.00       50.00
      40-49 |          3       50.00      100.00
------------+-----------------------------------
      Total |          6      100.00
res
no observations
```

```
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
       0-39 |          6       50.00       50.00
      40-49 |          6       50.00      100.00
------------+-----------------------------------
      Total |         12      100.00
azd
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
       0-39 |          6       50.00       50.00
      40-49 |          6       50.00      100.00
------------+-----------------------------------
      Total |         12      100.00
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
      40-49 |          1       50.00       50.00
      50-59 |          1       50.00      100.00
------------+-----------------------------------
      Total |          2      100.00
dmd
no observations
no observations
inf
no observations
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
       0-39 |          2       50.00       50.00
      40-49 |          2       50.00      100.00
------------+-----------------------------------
      Total |          4      100.00
flu
no observations
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
       0-39 |          4       50.00       50.00
      40-49 |          4       50.00      100.00
------------+-----------------------------------
      Total |          8      100.00
ckd
no observations
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
       0-39 |          8       50.00       50.00
      40-49 |          8       50.00      100.00
------------+-----------------------------------
      Total |         16      100.00
liv1
no observations
no observations
liv2
no observations
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
       0-39 |          5       50.00       50.00
      40-49 |          5       50.00      100.00
------------+-----------------------------------
      Total |         10      100.00

. count if cvd_d_nondm + can_d_nondm + dmd_d_nondm + inf_d_nondm + flu_d_nondm + res_d_nondm + liv1_
> d_nondm + ckd_d_nondm + azd_d_nondm > alldeath_d_nondm
  0

. count if cvd_d_dm + can_d_dm + dmd_d_dm + inf_d_dm + flu_d_dm + res_d_dm + liv1_d_dm + ckd_d_dm +
> azd_d_dm > alldeath_d_dm
  0

. count if chd_d_nondm + cbd_d_nondm + hfd_d_nondm > cvd_d_nondm
  0
```

```
. count if chd_d_dm + cbd_d_dm + hfd_d_dm > cvd_d_dm
  0
. count if liv1_d_nondm < liv2_d_nondm
  0
. count if liv1_d_dm < liv2_d_dm
  1
. ta age_gp1 if liv1_d_dm < liv2_d_dm
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
       0-39 |          1      100.00      100.00
------------+-----------------------------------
      Total |          1      100.00
. gen max_liv2 = min(liv1_d_dm,4)
. replace liv2_d_dm = runiformint(1,max_liv2) if liv1_d_dm < liv2_d_dm & inrange(liv2_d_dm,1,4)
(1 real change made)
. count if liv1_d_dm < liv2_d_dm
  0


replace country = substr(country,1,6)
gen age_dm = substr(age_gp1,1,2)
replace age_dm = "30" if age_dm == "0-"
destring age_dm, replace
replace age_dm = age_dm+5
gen age_nondm = substr(age_gp1,1,2)
replace age_nondm = "15" if age_nondm == "0-"
destring age_nondm, replace
replace age_nondm = age_nondm+5
keep country calendar sex alldeath_d_dm alldeath_d_nondm age_dm age_nondm pys_dm pys_nondm cvd_d_dm-
> azd_d_dm cvd_d_nondm-azd_d_nondm
save France, replace
```

1.6 Lithuania

For Lithuania, we have the following variables (by age, sex, and calendar year): total population size, prevalence of diabetes, incidence of diabetes, deaths in people with diabetes, and deaths in people without diabetes. 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. We can calculate person-years in people with diabetes, in a given calendar year, by adding the number of people with prevalent diabetes to half the number of people with incident diabetes and subtracting half the number of all-cause deaths. From there, person-years in people without diabetes is just person-years in the total population minus person-years in people with diabetes. [All of this has been performed before I got the dataset–JM.]

```
use uncleandbase, clear
keep if country == "Lithuania"
rename sex SEX
gen sex = 0 if SEX == "F"
replace sex = 1 if SEX == "M"
gen age_dm = substr(age_gp1,1,2)
replace age_dm = "30" if age_dm == "0-"
destring age_dm, replace
replace age_dm = age_dm+5
gen age_nondm = substr(age_gp1,1,2)
replace age_nondm = "15" if age_nondm == "0-"
destring age_nondm, replace
replace age_nondm = age_nondm+5
recode dmd_d_nondm .=0
rename (alldeath_dm alldeath_nondm alldeath_totpop) (alldeath_d_dm alldeath_d_nondm alldeath_d_pop)
count if cvd_d_nondm + can_d_nondm + dmd_d_nondm + inf_d_nondm + flu_d_nondm + res_d_nondm + liv1_d_
> nondm + ckd_d_nondm + azd_d_nondm > alldeath_d_nondm
count if cvd_d_dm + can_d_dm + dmd_d_dm + inf_d_dm + flu_d_dm + res_d_dm + liv1_d_dm + ckd_d_dm + az
> d_d_dm > alldeath_d_dm
count if chd_d_nondm + cbd_d_nondm + hfd_d_nondm > cvd_d_nondm
count if chd_d_dm + cbd_d_dm + hfd_d_dm > cvd_d_dm
count if liv1_d_nondm < liv2_d_nondm
count if liv1_d_dm < liv2_d_dm
count if liv1_d_dm < liv2_d_dm
keep country calendar sex alldeath_d_dm alldeath_d_nondm age_dm age_nondm pys_dm pys_nondm cvd_d_dm-
> azd_d_dm cvd_d_nondm-azd_d_nondm
save Lithuania, replace
```

1.7 Netherlands

For Netherlands, we have the following variables (by age, sex, and calendar year): Person-years and deaths in people with and without diabetes. I.e., no further variables need to be derived. Netherlands restricts counts between 0 and 9 for both people with and without diabetes. I will fill them in randomly, where the number can be any number from 0 to 9 with equal probability. I will assume the mid-point of the age interval for people aged <40 is 35 and for 90+ is 95.

Additionally, because the definition of diabetes in the Netherlands relies on 2 years of follow-up from 2006, we are dropping data from 2007.

```
. use uncleandbase, clear
. keep if country == "Netherlands"
(2,148 observations deleted)
. rename sex SEX
. gen sex = 0 if SEX == "F"
(98 missing values generated)
. replace sex = 1 if SEX == "M"
(98 real changes made)
. rename (alldeath_dm alldeath_nondm alldeath_totpop) (alldeath_d_dm alldeath_d_nondm alldeath_d_pop
> )

. ta age_gp1
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
        0-39 |         28       14.29       14.29
      40-49 |         28       14.29       28.57
      50-59 |         28       14.29       42.86
      60-69 |         28       14.29       57.14
      70-79 |         28       14.29       71.43
      80-89 |         28       14.29       85.71
        90+ |         28       14.29      100.00
------------+-----------------------------------
      Total |        196      100.00
. foreach i in alldeath can cvd chd cbd hfd res azd dmd inf flu ckd liv1 liv2 {
  2. di "`i´"
  3. ta age_gp1 if `i´_d_nondm ==.
  4. quietly replace `i´_d_nondm = runiformint(0,9) if `i´_d_nondm==.
  5. ta age_gp1 if `i´_d_dm ==.
  6. quietly replace `i´_d_dm = runiformint(0,9) if `i´_d_dm ==.
  7. }
alldeath
no observations
no observations
can
no observations
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
        0-39 |         25      100.00      100.00
------------+-----------------------------------
      Total |         25      100.00
cvd
no observations
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
        0-39 |         28      100.00      100.00
------------+-----------------------------------
      Total |         28      100.00
chd
no observations
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
```

0-39	28	60.87	60.87
40-49	18	39.13	100.00
Total	46	100.00	

cbd

age_gp1	Freq.	Percent	Cum.
0-39	2	100.00	100.00
Total	2	100.00	

age_gp1	Freq.	Percent	Cum.
0-39	28	49.12	49.12
40-49	28	49.12	98.25
50-59	1	1.75	100.00
Total	57	100.00	

hfd

age_gp1	Freq.	Percent	Cum.
0-39	26	65.00	65.00
40-49	14	35.00	100.00
Total	40	100.00	

age_gp1	Freq.	Percent	Cum.
0-39	28	41.18	41.18
40-49	28	41.18	82.35
50-59	12	17.65	100.00
Total	68	100.00	

res

age_gp1	Freq.	Percent	Cum.
0-39	28	87.50	87.50
40-49	4	12.50	100.00
Total	32	100.00	

age_gp1	Freq.	Percent	Cum.
0-39	28	47.46	47.46
40-49	28	47.46	94.92
50-59	3	5.08	100.00
Total	59	100.00	

azd

age_gp1	Freq.	Percent	Cum.
0-39	28	43.75	43.75
40-49	28	43.75	87.50
50-59	8	12.50	100.00
Total	64	100.00	

age_gp1	Freq.	Percent	Cum.
0-39	28	29.17	29.17
40-49	28	29.17	58.33
50-59	28	29.17	87.50
60-69	12	12.50	100.00
Total	96	100.00	

dmd

age_gp1	Freq.	Percent	Cum.
0-39	28	37.33	37.33

```
      40-49 |         27       36.00       73.33
      50-59 |         15       20.00       93.33
      60-69 |          3        4.00       97.33
        90+ |          2        2.67      100.00
------------+-----------------------------------
      Total |         75      100.00
```

```
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
        0-39 |         22       95.65       95.65
      40-49 |          1        4.35      100.00
------------+-----------------------------------
      Total |         23      100.00
```

```
inf
no observations
```

```
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
        0-39 |         28       43.75       43.75
      40-49 |         28       43.75       87.50
      50-59 |          8       12.50      100.00
------------+-----------------------------------
      Total |         64      100.00
```

```
flu
```

```
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
        0-39 |         11       57.89       57.89
      40-49 |          8       42.11      100.00
------------+-----------------------------------
      Total |         19      100.00
```

```
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
        0-39 |         28       34.57       34.57
      40-49 |         28       34.57       69.14
      50-59 |         25       30.86      100.00
------------+-----------------------------------
      Total |         81      100.00
```

```
ckd
```

```
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
        0-39 |         28       43.75       43.75
      40-49 |         27       42.19       85.94
      50-59 |          9       14.06      100.00
------------+-----------------------------------
      Total |         64      100.00
```

```
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
        0-39 |         28       32.56       32.56
      40-49 |         28       32.56       65.12
      50-59 |         27       31.40       96.51
      60-69 |          3        3.49      100.00
------------+-----------------------------------
      Total |         86      100.00
```

```
liv1
```

```
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
        0-39 |         13       65.00       65.00
        90+ |          7       35.00      100.00
------------+-----------------------------------
      Total |         20      100.00
```

```
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
        0-39 |         28       31.11       31.11
      40-49 |         27       30.00       61.11
      50-59 |          6        6.67       67.78
      80-89 |          1        1.11       68.89
        90+ |         28       31.11      100.00
------------+-----------------------------------
```

```
      Total |          90      100.00
liv2
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
       0-39 |          28       62.22       62.22
      40-49 |          10       22.22       84.44
        90+ |           7       15.56      100.00
------------+-----------------------------------
      Total |          45      100.00
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
       0-39 |          28       26.67       26.67
      40-49 |          28       26.67       53.33
      50-59 |          17       16.19       69.52
      60-69 |           2        1.90       71.43
      80-89 |           2        1.90       73.33
        90+ |          28       26.67      100.00
------------+-----------------------------------
      Total |         105      100.00

. count if cvd_d_nondm + can_d_nondm + dmd_d_nondm + inf_d_nondm + flu_d_nondm + res_d_nondm + liv1_
> d_nondm + ckd_d_nondm + azd_d_nondm > alldeath_d_nondm
  0

. count if cvd_d_dm + can_d_dm + dmd_d_dm + inf_d_dm + flu_d_dm + res_d_dm + liv1_d_dm + ckd_d_dm +
> azd_d_dm > alldeath_d_dm
  25

. quietly {
Done

. count if chd_d_pop + cbd_d_pop + hfd_d_pop > cvd_d_pop
  0

. count if chd_d_dm + cbd_d_dm + hfd_d_dm > cvd_d_dm
  28

. ta age_gp1 if chd_d_dm + cbd_d_dm + hfd_d_dm > cvd_d_dm
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
       0-39 |          26       92.86       92.86
      40-49 |           2        7.14      100.00
------------+-----------------------------------
      Total |          28      100.00

. gen max_chd = min(cvd_d_dm,9)

. replace chd_d_dm = runiformint(0,max_chd) if chd_d_dm + cbd_d_dm + hfd_d_dm > cvd_d_dm & inrange(c
> hd_d_dm,0,9)
(28 real changes made)

. gen max_cbd = min(cvd_d_dm-chd_d_dm,9)

. replace cbd_d_dm = runiformint(0,max_cbd) if chd_d_dm + cbd_d_dm + hfd_d_dm > cvd_d_dm & inrange(c
> bd_d_dm,0,9)
(26 real changes made)

. gen max_hfd = min(cvd_d_dm-chd_d_dm-cbd_d_dm,9)

. replace hfd_d_dm = runiformint(0,max_hfd) if chd_d_dm + cbd_d_dm + hfd_d_dm > cvd_d_dm & inrange(h
> fd_d_dm,0,9)
(20 real changes made)

. count if chd_d_dm + cbd_d_dm + hfd_d_dm > cvd_d_dm
  0

. count if liv1_d_nondm < liv2_d_nondm
  10

. ta age_gp1 if liv1_d_nondm < liv2_d_nondm
    age_gp1 |      Freq.     Percent        Cum.
------------+-----------------------------------
       0-39 |           8       80.00       80.00
        90+ |           2       20.00      100.00
------------+-----------------------------------
      Total |          10      100.00

. gen max_liv2 = min(liv1_d_nondm,9)
```

```
. replace liv2_d_nondm = runiformint(0,max_liv2) if liv1_d_nondm < liv2_d_nondm & inrange(liv2_d_non
> dm,0,9)
(10 real changes made)

. count if liv1_d_nondm < liv2_d_nondm
  0

. count if liv1_d_dm < liv2_d_dm
  43

. ta age_gp1 if liv1_d_dm < liv2_d_dm
```

age_gp1	Freq.	Percent	Cum.
0-39	11	25.58	25.58
40-49	12	27.91	53.49
50-59	5	11.63	65.12
80-89	1	2.33	67.44
90+	14	32.56	100.00
Total	43	100.00	

```
. replace max_liv2 = min(liv1_d_dm,9)
(83 real changes made)

. replace liv2_d_dm = runiformint(0,max_liv2) if liv1_d_dm < liv2_d_dm & inrange(liv2_d_dm,0,9)
(43 real changes made)

. count if liv1_d_dm < liv2_d_dm
  0


gen age_dm = substr(age_gp1,1,2)
replace age_dm = "30" if age_dm == "0-"
destring age_dm, replace
replace age_dm = age_dm+5
gen age_nondm = substr(age_gp1,1,2)
replace age_nondm = "15" if age_nondm == "0-"
destring age_nondm, replace
replace age_nondm = age_nondm+5
keep country calendar sex alldeath_d_dm alldeath_d_nondm age_dm age_nondm pys_dm pys_nondm cvd_d_dm-
> azd_d_dm cvd_d_nondm-azd_d_nondm
save Netherlands, replace
```

1.8 Scotland

For Scotland, 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]. I then calculate person-years in people without diabetes by substracting the person-years in people with diabetes from person-years in the total population; similarly for deaths in people without diabetes. There were a few age groups in whom the number of deaths from diabetes in people with diabetes was slightly greater than the total population deaths; I will simply make these zero in people without diabetes.

Also note we have received two different age groupings for Scotland for total population deaths – from 2006-2015: 0-39, 40-49, ..., 80+; from 2016-2020: 0-39, 40-49, ..., 90+. For the 80+ age grouping I will assume the mean age is 87.5 years.

```
. use uncleandbase, clear
. keep if country == "Scotland"
(2,104 observations deleted)
. rename sex SEX
. gen sex = 0 if SEX == "F"
(120 missing values generated)
. replace sex = 1 if SEX == "M"
(120 real changes made)
. rename (alldeath_dm alldeath_nondm alldeath_totpop) (alldeath_d_dm alldeath_d_nondm alldeath_d_pop
> )
. foreach i in alldeath can cvd chd cbd hfd res azd dmd inf flu ckd liv1 liv2 {
  2. quietly replace `i´_d_nondm = `i´_d_pop-`i´_d_dm
  3. di "`i´"
  4. ta `i´_d_nondm if `i´_d_nondm <0
  5. }
alldeath
no observations
can
no observations
cvd
no observations
chd
no observations
cbd
no observations
hfd
no observations
res
no observations
azd
no observations
dmd
```

dmd_d_nondm	Freq.	Percent	Cum.
-4	1	7.69	7.69
-2	2	15.38	23.08
-1	10	76.92	100.00
Total	13	100.00	

```
inf
no observations
flu
no observations
```

```
ckd
no observations
liv1
no observations
liv2
no observations
. replace dmd_d_nondm = 0 if dmd_d_nondm <0
(13 real changes made)
. replace pys_nondm = pys_totpop-pys_dm
(240 real changes made)
. gen age_dm = substr(age_gp1,1,2)
(30 missing values generated)
. replace age_dm = "30" if age_dm == "0-"
(30 real changes made)
. destring age_dm, replace
age_dm: all characters numeric; replaced as byte
(30 missing values generated)
. replace age_dm = age_dm+5
(210 real changes made)
. gen age_nondm = substr(age_gp1,1,2)
(30 missing values generated)
. replace age_nondm = substr(age_gp2,1,2) if cal <= 2015
(60 real changes made)
. replace age_nondm = "15" if age_nondm == "0-"
(30 real changes made)
. destring age_nondm, replace
age_nondm: all characters numeric; replaced as byte
(50 missing values generated)
. replace age_nondm = age_nondm+5
(190 real changes made)
. replace age_nondm = age_nondm+2.5 if age_nondm == 85 & cal <= 2015
variable age_nondm was byte now float
(20 real changes made)
. replace pys_dm =. if age_dm==.
(30 real changes made, 30 to missing)
. replace pys_nondm =. if age_nondm==.
(50 real changes made, 50 to missing)
. foreach i in alldeath can cvd chd cbd hfd res azd dmd inf flu ckd liv1 liv2 {
  2. replace `i´_d_dm = . if age_dm==.
  3. replace `i´_d_nondm = . if age_nondm==.
  4. }
(30 real changes made, 30 to missing)
(10 real changes made, 10 to missing)
(30 real changes made, 30 to missing)
(10 real changes made, 10 to missing)
(30 real changes made, 30 to missing)
(10 real changes made, 10 to missing)
(30 real changes made, 30 to missing)
(10 real changes made, 10 to missing)
(30 real changes made, 30 to missing)
(10 real changes made, 10 to missing)
(30 real changes made, 30 to missing)
(10 real changes made, 10 to missing)
(30 real changes made, 30 to missing)
(10 real changes made, 10 to missing)
(30 real changes made, 30 to missing)
(10 real changes made, 10 to missing)
(30 real changes made, 30 to missing)
(10 real changes made, 10 to missing)
(30 real changes made, 30 to missing)
(10 real changes made, 10 to missing)
(30 real changes made, 30 to missing)
(10 real changes made, 10 to missing)
```

```
(30 real changes made, 30 to missing)
(10 real changes made, 10 to missing)
(30 real changes made, 30 to missing)
(10 real changes made, 10 to missing)
(30 real changes made, 30 to missing)
(10 real changes made, 10 to missing)
. drop if age_dm==. & age_nondm==.
(10 observations deleted)
. count if cvd_d_nondm + can_d_nondm + dmd_d_nondm + inf_d_nondm + flu_d_nondm + res_d_nondm + liv1_
> d_nondm + ckd_d_nondm + azd_d_nondm > alldeath_d_nondm
  0
. count if cvd_d_dm + can_d_dm + dmd_d_dm + inf_d_dm + flu_d_dm + res_d_dm + liv1_d_dm + ckd_d_dm +
> azd_d_dm > alldeath_d_dm
  0
. count if chd_d_nondm + cbd_d_nondm + hfd_d_nondm > cvd_d_nondm
  0
. count if chd_d_dm + cbd_d_dm + hfd_d_dm > cvd_d_dm
  0
. count if liv1_d_nondm < liv2_d_nondm
  0
. count if liv1_d_dm < liv2_d_dm
  0


keep country calendar sex alldeath_d_dm alldeath_d_nondm age_dm age_nondm pys_dm pys_nondm cvd_d_dm-
> azd_d_dm cvd_d_nondm-azd_d_nondm
save Scotland, replace
```

1.9 South Korea

For South Korea, we have the following variables (by age, sex, and calendar year): Person-years and deaths in people with and without diabetes. I.e., no further variables need to be derived. Note that from 2007-2010, there is no data for people aged 90 and above, although this shouldn't have a huge impact on any results. I will assume the mid-point of the age interval for people aged <40 is 35 and for 90+ is 95.

```
use uncleandbase, clear
keep if country=="South Korea"
rename sex SEX
gen sex = 0 if SEX == "F"
replace sex = 1 if SEX == "M"
drop if age_gp4=="all ages"
drop if age_gp1 == "90+" & cal <= 2010
rename (alldeath_dm alldeath_nondm alldeath_totpop) (alldeath_d_dm alldeath_d_nondm alldeath_d_pop)
gen age_dm = substr(age_gp1,1,2)
replace age_dm = "30" if age_dm == "0-"
destring age_dm, replace
replace age_dm = age_dm+5
gen age_nondm = substr(age_gp1,1,2)
replace age_nondm = "15" if age_nondm == "0-"
destring age_nondm, replace
replace age_nondm = age_nondm+5
replace country = "SKorea"


. count if cvd_d_nondm + can_d_nondm + dmd_d_nondm + inf_d_nondm + flu_d_nondm + res_d_nondm + liv1_
> d_nondm + ckd_d_nondm + azd_d_nondm > alldeath_d_nondm
  0
. count if cvd_d_nondm + can_d_nondm + dmd_d_nondm + inf_d_nondm + flu_d_nondm + res_d_nondm + liv1_
> d_nondm + ckd_d_nondm + azd_d_nondm > alldeath_d_nondm
  0
. count if cvd_d_dm + can_d_dm + dmd_d_dm + inf_d_dm + flu_d_dm + res_d_dm + liv1_d_dm + ckd_d_dm +
> azd_d_dm > alldeath_d_dm
  0
. count if chd_d_nondm + cbd_d_nondm + hfd_d_nondm > cvd_d_nondm
  0
. count if chd_d_dm + cbd_d_dm + hfd_d_dm > cvd_d_dm
  0
. count if liv1_d_nondm < liv2_d_nondm
  0
. count if liv1_d_dm < liv2_d_dm
  0


keep country calendar sex alldeath_d_dm alldeath_d_nondm age_dm age_nondm pys_dm pys_nondm cvd_d_dm-
> azd_d_dm cvd_d_nondm-azd_d_nondm
save SKorea, replace
```

1.10 Summary

Table 1.2 shows a summary of the data included in this analysis.

```
*mkdir CSV
clear
foreach c in Australia Canada Denmark Finland France Lithuania Netherlands Scotland SKorea {
append using `c´
}
gen other_d_dm = alldeath_d_dm - ///
(cvd_d_dm + can_d_dm + dmd_d_dm + inf_d_dm + flu_d_dm + res_d_dm + liv1_d_dm + ckd_d_dm + azd_d_dm)
gen other_d_nondm = alldeath_d_nondm - ///
(cvd_d_nondm + can_d_nondm + dmd_d_nondm + inf_d_nondm + flu_d_nondm + res_d_nondm + liv1_d_nondm +
> ckd_d_nondm + azd_d_nondm)
save cleandbase, replace
use cleandbase, clear
foreach i in chd cbd hfd liv2 {
drop `i´_d_dm `i´_d_nondm
}
rename (liv1_d_dm liv1_d_nondm) (liv_d_dm liv_d_nondm)
save CMdata, replace
use cleandbase, clear
foreach i in alldeath can res azd dmd inf flu ckd liv1 liv2 {
drop `i´_d_dm `i´_d_nondm
}
save CMdataCVD, replace
use cleandbase, clear
foreach i in alldeath can cvd chd cbd hfd res azd dmd inf flu ckd {
drop `i´_d_dm `i´_d_nondm
}
save CMdataLIV, replace
use cleandbase, clear
foreach i in alldeath can cvd chd cbd hfd res dmd inf flu ckd liv1 liv2 {
drop `i´_d_dm `i´_d_nondm
}
save CMdataDEM, replace
foreach c in Australia Canada Denmark Finland France Lithuania Netherlands Scotland SKorea {
use `c´, clear
foreach i in chd cbd hfd liv2 {
drop `i´_d_dm `i´_d_nondm
}
rename (liv1_d_dm liv1_d_nondm) (liv_d_dm liv_d_nondm)
save `c´, replace
}
erase uncleandbase.dta
use CMdata, clear
bysort country (cal) : egen lb = min(cal)
bysort country (cal) : egen ub = max(cal)
tostring lb ub, replace
gen rang = lb+ "-" + ub
recode dmd_d_nondm .=0
collapse (sum) pys_dm pys_nondm cvd_d_dm-azd_d_dm cvd_d_nondm-azd_d_nondm other_d_dm other_d_nondm,
> by(country sex rang)
expand 2
bysort country sex : gen DM = _n-1
tostring sex pys_dm-DM, replace force format(%15.0fc)
gen pys = pys_dm if DM == "1"
replace pys = pys_nondm if DM == "0"
foreach i in can cvd res azd dmd inf flu ckd liv other {
gen `i´ = `i´_d_dm if DM == "1"
replace `i´ = `i´_d_nondm if DM == "0"
}
keep country-rang DM-other
order country rang DM sex
sort country rang DM sex
gen njm = _n
bysort country DM (njm) : replace DM ="" if _n!=1
bysort country (njm) : replace rang ="" if _n!=1
```

```
bysort country (njm) : replace country ="" if _n!=1
sort njm
replace DM = "No diabetes" if DM == "0"
replace DM = "Diabetes" if DM == "1"
replace sex = "Female" if sex == "0"
replace sex = "Male" if sex == "1"
drop njm
replace country = "South Korea" if country == "SKorea"
replace country = "Canada (Alberta)" if country == "Canada"
export delimited using CSV/T1.csv, delimiter(":") novarnames replace
```

Table 1.2: Summary of data included in the analysis.

Country	Period	Diabetes status	Sex	Person-years of follow-up	Death counts by cause of death									
					DEM	CAN	CVD	RES	DMD	INF	FLU	CKD	LIV	OTH
Australia	2005-2019	No diabetes	Female	131,621,405	190,397	239,116	33,800	70,553	5,305	10,965	15,879	11,931	4,680	139,032
			Male	128,841,354	236,115	207,582	36,838	34,544	4,803	11,194	11,719	10,180	10,231	162,782
		Diabetes	Female	5,208,083	31,240	40,204	5,188	8,626	18,118	2,774	2,609	2,981	1,635	20,385
			Male	5,889,560	51,030	50,476	6,561	6,731	21,098	3,210	2,781	3,318	3,110	26,028
Canada (Alberta)	2005-2020	No diabetes	Female	31,856,052	35,337	36,490	6,083	13,328	265	2,025	2,714	1,408	1,567	26,490
			Male	31,998,428	36,954	35,527	6,222	6,656	328	1,930	2,387	1,318	2,543	37,392
		Diabetes	Female	1,749,550	9,919	15,371	2,039	3,742	3,404	982	1,156	966	830	7,194
			Male	2,163,971	14,256	19,685	2,593	2,829	4,104	1,035	1,235	961	1,207	9,697
Denmark	2002-2019	No diabetes	Female	47,999,917	117,552	116,692	28,836	32,935	582	7,126	15,934	2,874	4,546	92,358
			Male	46,993,523	120,657	107,745	24,312	15,526	738	6,027	13,055	3,635	8,277	89,655
		Diabetes	Female	1,844,344	16,049	19,319	4,278	3,596	10,165	1,387	2,338	604	855	11,621
			Male	2,071,414	23,095	24,111	4,366	2,463	12,698	1,486	2,545	757	1,965	13,289
Finland	2000-2017	No diabetes	Female	46,015,832	74,713	103,001	5,687	49,654	0	2,416	6,452	866	4,280	67,154
			Male	44,057,257	81,409	102,432	11,814	23,448	0	2,260	6,018	942	10,650	93,103
		Diabetes	Female	2,911,525	19,558	54,506	1,400	13,944	4,486	840	1,928	371	1,351	12,002
			Male	3,023,860	24,133	51,384	2,836	6,970	4,715	726	1,668	357	3,266	14,478
France	2013-2020	No diabetes	Female	248,386,875	389,029	398,245	25,997	141,470	0	30,199	41,802	16,641	12,203	659,163
			Male	218,181,642	475,727	313,608	34,453	59,456	0	25,634	33,547	15,551	26,454	611,892
		Diabetes	Female	10,932,947	62,306	75,088	3,997	12,402	29,723	6,630	6,918	4,419	3,682	98,552
			Male	12,594,448	116,141	90,925	7,357	8,858	30,152	7,638	8,355	4,997	8,797	124,834
Lithuania	2014-2021	No diabetes	Female	11,564,058	24,203	85,770	1,107	2,179	0	2,041	1,499	368	2,061	19,037
			Male	10,078,198	31,081	63,343	2,897	883	0	2,557	2,530	332	3,775	30,722
		Diabetes	Female	653,558	4,511	18,406	262	297	2,199	498	308	51	378	3,751
			Male	446,916	4,407	11,431	385	100	1,609	363	277	30	523	3,282
Netherlands	2008-2021	No diabetes	Female	114,105,112	233,939	219,478	35,068	99,985	2,523	15,329	25,230	9,610	3,976	192,401
			Male	111,687,855	268,791	196,694	38,149	43,994	1,935	13,425	21,401	8,253	6,508	184,124
		Diabetes	Female	5,592,483	45,882	61,768	8,165	20,187	18,661	5,316	6,917	3,778	1,716	42,431
			Male	6,123,018	61,673	58,232	9,751	10,615	16,987	5,033	6,213	3,184	2,469	40,862
South Korea	2007-2019	No diabetes	Female	6,216,486	4,737	4,833	453	1,005	88	486	1,026	206	211	6,866
			Male	6,208,026	7,626	4,278	766	443	76	592	953	207	772	8,117
		Diabetes	Female	596,993	2,780	3,362	260	455	1,304	402	685	390	182	3,044
			Male	618,623	4,600	2,922	449	208	1,262	407	739	416	607	3,400
Scotland	2006-2020	No diabetes	Female	39,339,945	99,092	101,075	22,732	42,294	255	5,196	15,042	3,576	5,238	71,833
			Male	36,563,096	100,119	92,998	18,389	18,341	305	4,039	10,569	2,410	8,403	74,277
		Diabetes	Female	1,842,509	15,967	21,465	3,290	5,962	5,850	1,257	2,714	660	1,088	11,288
			Male	2,292,250	21,224	26,267	3,086	4,161	6,204	1,121	2,571	617	1,763	12,295

Abbreviations: DEM – Dementia; CAN – Cancer; CVD – Cardiovascular disease; RES – Chronic lower respiratory disease; DMD – Diabetes; INF – Infectious diseases; FLU – Influenza and pneumonia; CKD – Kidney disease; LIV – Liver disease; OTH – All other causes.

2 Crude rates

```
foreach c in Australia Canada Denmark Finland France Lithuania Netherlands Scotland SKorea {
use `c´, clear
if "`c´" == "Canada" {
local co = "Canada (Alberta)"
}
else if "`c´" == "SKorea" {
local co = "South Korea"
}
else {
local co = "`c´"
}
collapse (sum) pys_dm pys_nondm cvd_d_dm-azd_d_dm cvd_d_nondm-azd_d_nondm, by(calendar sex)
foreach i in can cvd res azd dmd inf flu ckd liv {
if "`i´" == "can" {
local ii = "Cancer"
}
if "`i´" == "cvd" {
local ii = "Cardiovascular disease"
}
if "`i´" == "res" {
local ii = "Chronic lower respiratory disease"
}
if "`i´" == "azd" {
local ii = "Dementia"
}
if "`i´" == "dmd" {
local ii = "Diabetes"
}
if "`i´" == "inf" {
local ii = "Infectious diseases"
}
if "`i´" == "flu" {
local ii = "Influenza and pneumonia"
}
if "`i´" == "ckd" {
local ii = "Kidney disease"
}
if "`i´" == "liv" {
local ii = "Liver disease"
}
foreach iii in dm nondm {
if "`iii´" == "dm" {
local dd = "with"
}
if "`iii´" == "nondm" {
local dd = "without"
}
gen `iii´_`i´ = 1000*`i´_d_`iii´/pys_`iii´
twoway ///
(connected `iii´_`i´ cal if sex == 0, col(red)) ///
(connected `iii´_`i´ cal if sex == 1, col(blue)) ///
, graphregion(color(white)) ///
ytitle(Mortality rate (per 1,000 person-years), margin(a+2)) ///
xtitle(Calendar year) ///
legend(order( ///
1 "Females" ///
2 "Males" ///
) cols(1) position(3) region(lcolor(none) color(none))) ///
ylabel(,angle(0) format(%9.2f)) ///
title("People `dd´ diabetes", placement(west) size(medium) col(black))
graph save GPH/cr_`i´_`iii´_`c´, replace
}
graph combine ///
GPH/cr_`i´_dm_`c´.gph ///
GPH/cr_`i´_nondm_`c´.gph ///
```

```
, graphregion(color(white)) cols(2) altshrink xsize(10)
}
}
```

Figure 2.1: Crude mortality rate by cause of death, sex, and diabetes status. Cancer. Australia.

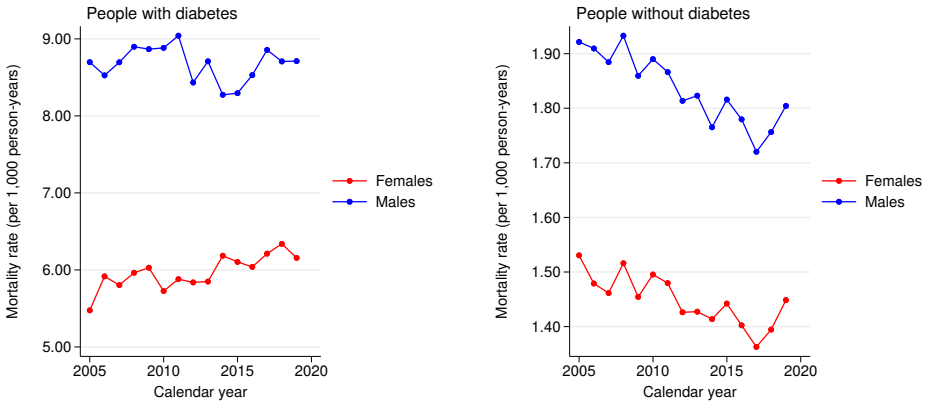


Figure 2.2: Crude mortality rate by cause of death, sex, and diabetes status. Cardiovascular disease. Australia.

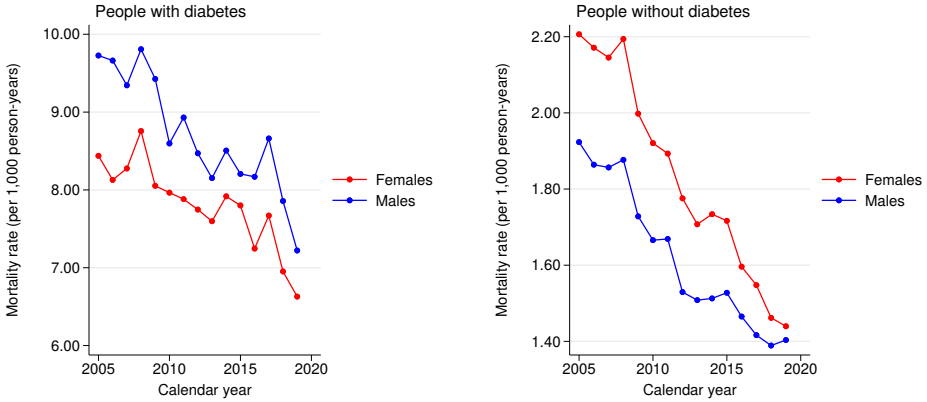


Figure 2.3: Crude mortality rate by cause of death, sex, and diabetes status. Chronic lower respiratory disease. Australia.

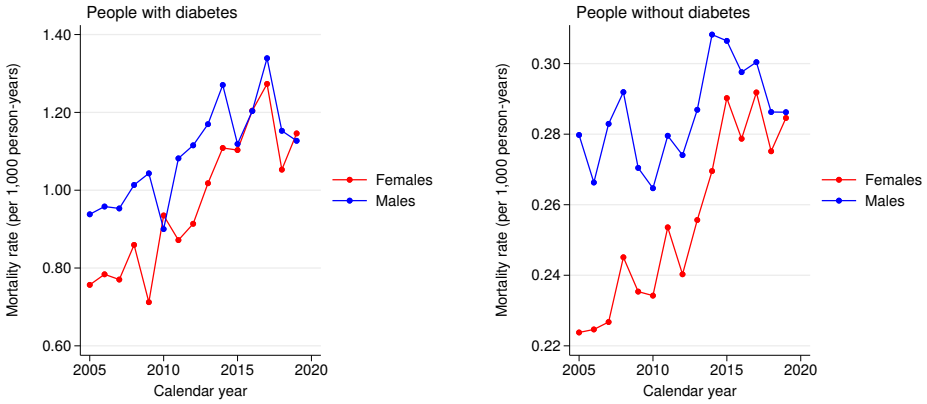


Figure 2.4: Crude mortality rate by cause of death, sex, and diabetes status. Dementia. Australia.

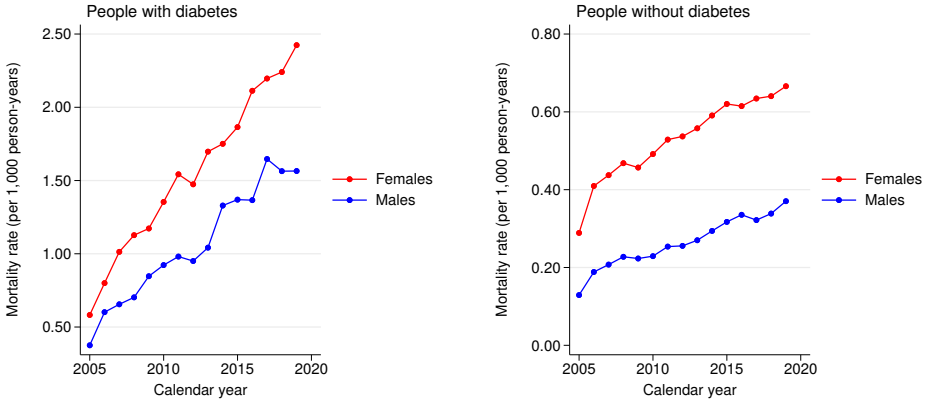


Figure 2.5: Crude mortality rate by cause of death, sex, and diabetes status. Diabetes. Australia.

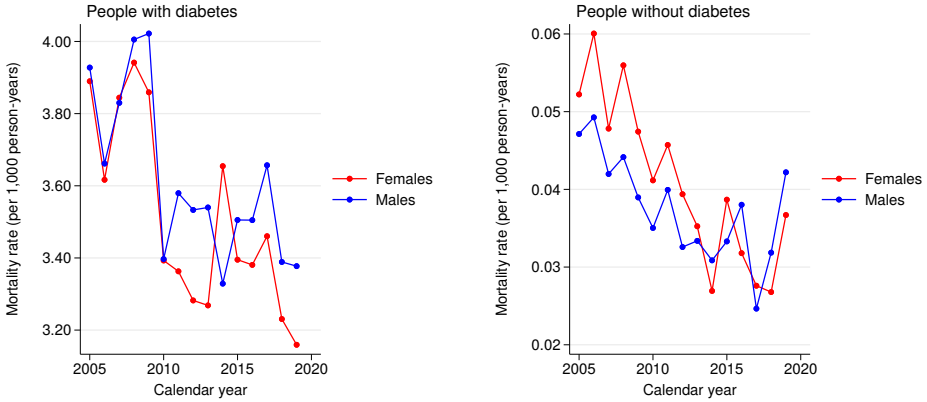


Figure 2.6: Crude mortality rate by cause of death, sex, and diabetes status. Infectious diseases. Australia.

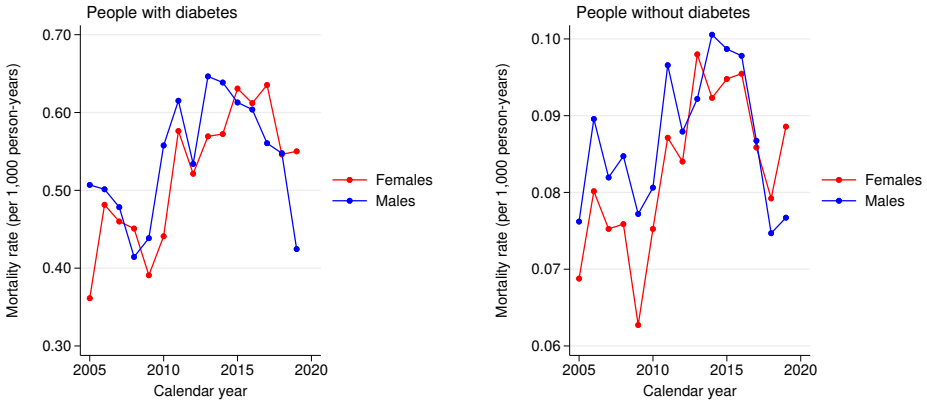


Figure 2.7: Crude mortality rate by cause of death, sex, and diabetes status. Influenza and pneumonia. Australia.

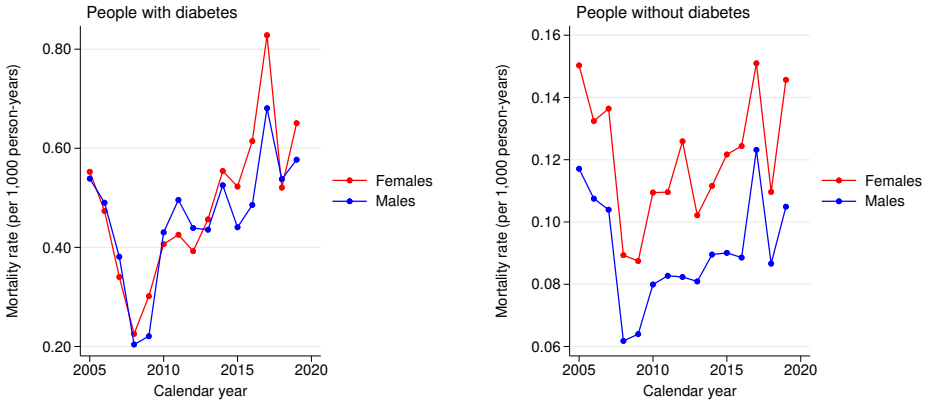


Figure 2.8: Crude mortality rate by cause of death, sex, and diabetes status. Kidney disease. Australia.

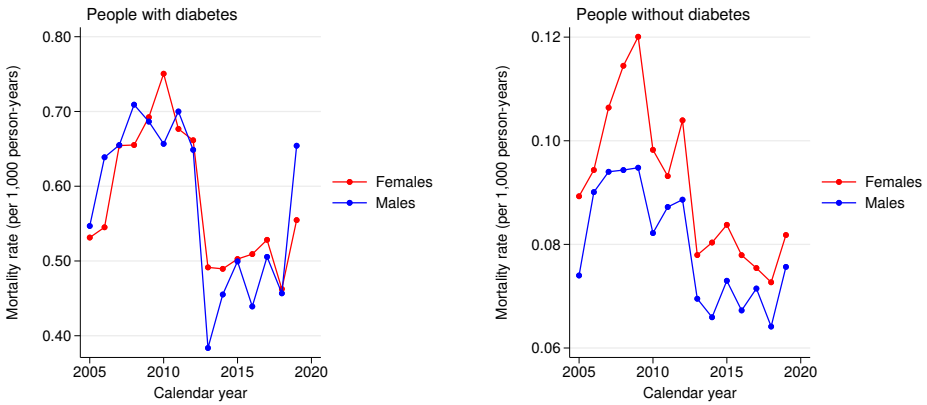


Figure 2.9: Crude mortality rate by cause of death, sex, and diabetes status. Liver disease. Australia.

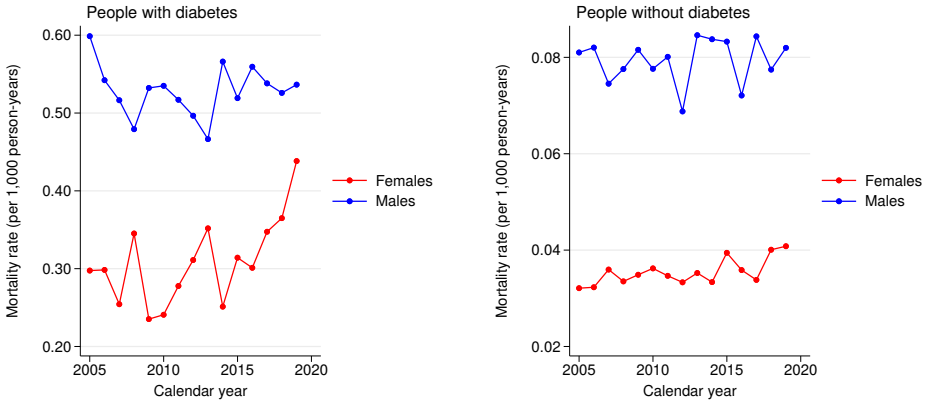


Figure 2.10: Crude mortality rate by cause of death, sex, and diabetes status. Cancer. Canada (Alberta).

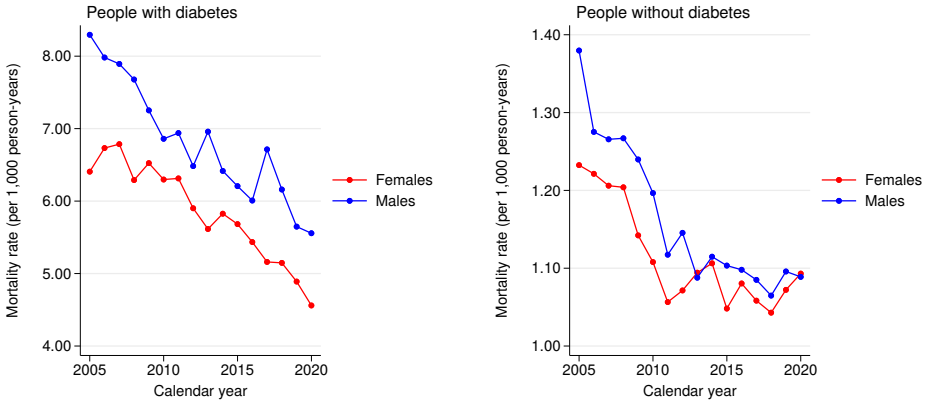


Figure 2.11: Crude mortality rate by cause of death, sex, and diabetes status. Cardiovascular disease. Canada (Alberta).

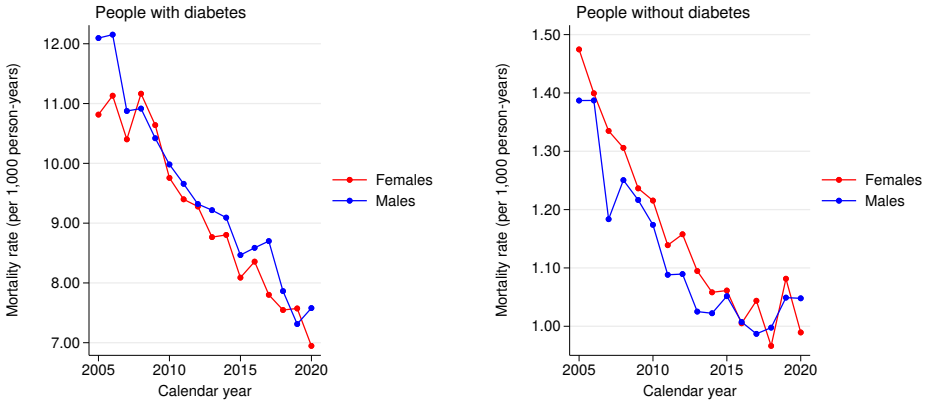


Figure 2.12: Crude mortality rate by cause of death, sex, and diabetes status. Chronic lower respiratory disease. Canada (Alberta).

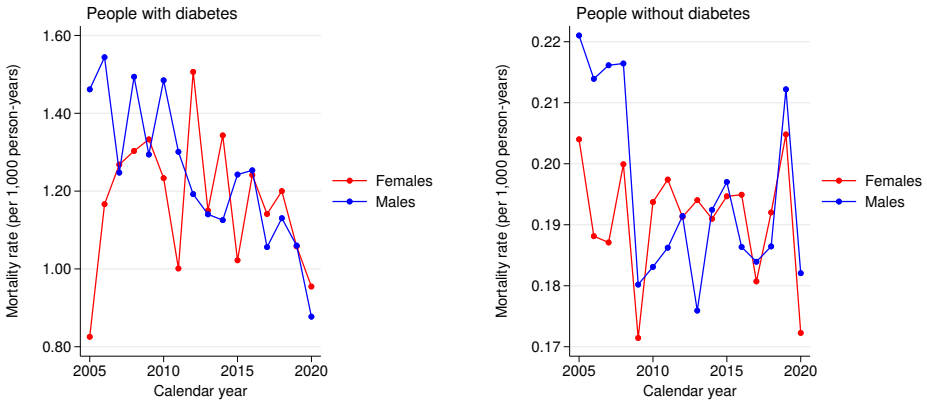


Figure 2.13: Crude mortality rate by cause of death, sex, and diabetes status. Dementia. Canada (Alberta).

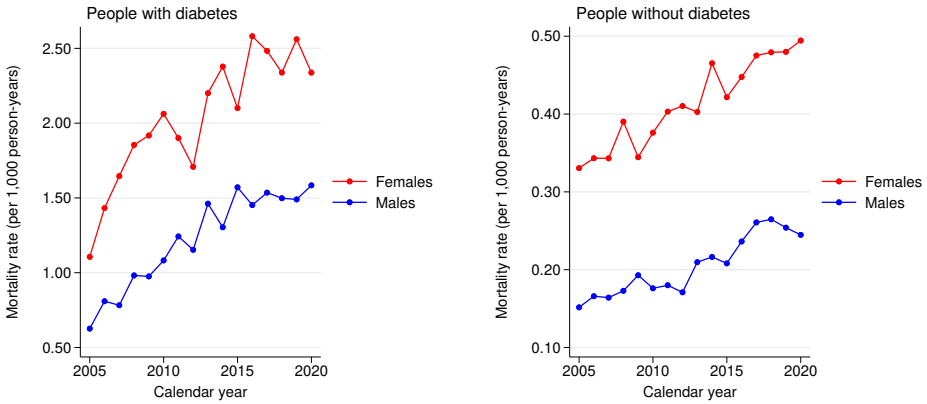


Figure 2.14: Crude mortality rate by cause of death, sex, and diabetes status. Diabetes. Canada (Alberta).

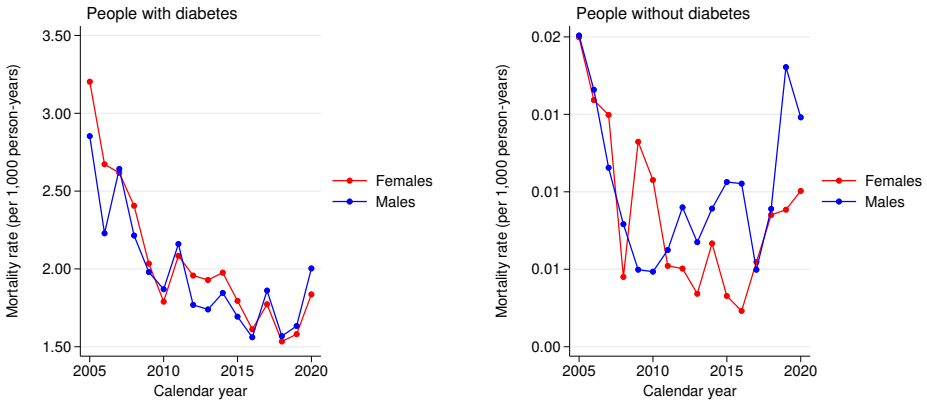


Figure 2.15: Crude mortality rate by cause of death, sex, and diabetes status. Infectious diseases. Canada (Alberta).

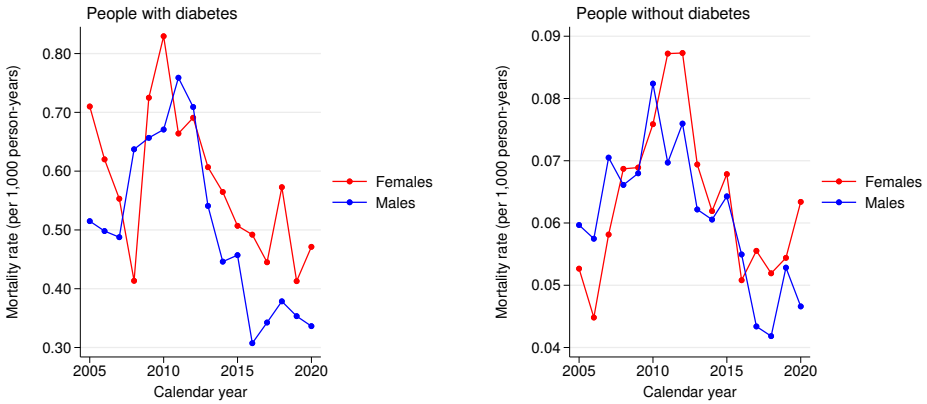


Figure 2.16: Crude mortality rate by cause of death, sex, and diabetes status. Influenza and pneumonia. Canada (Alberta).

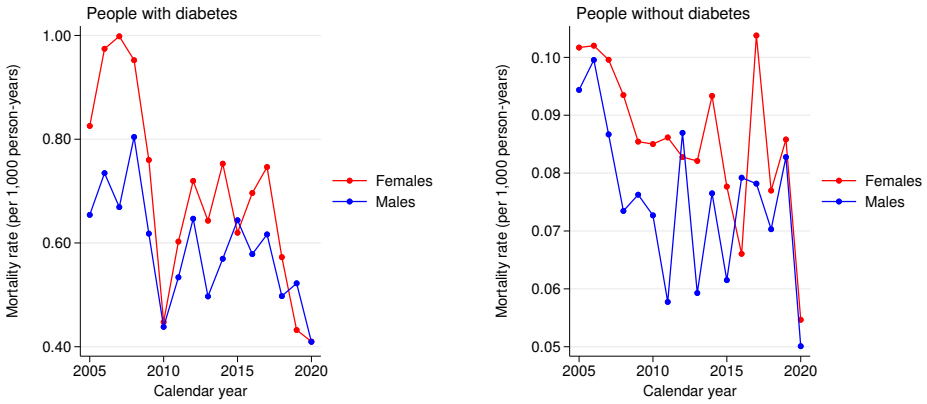


Figure 2.17: Crude mortality rate by cause of death, sex, and diabetes status. Kidney disease. Canada (Alberta).

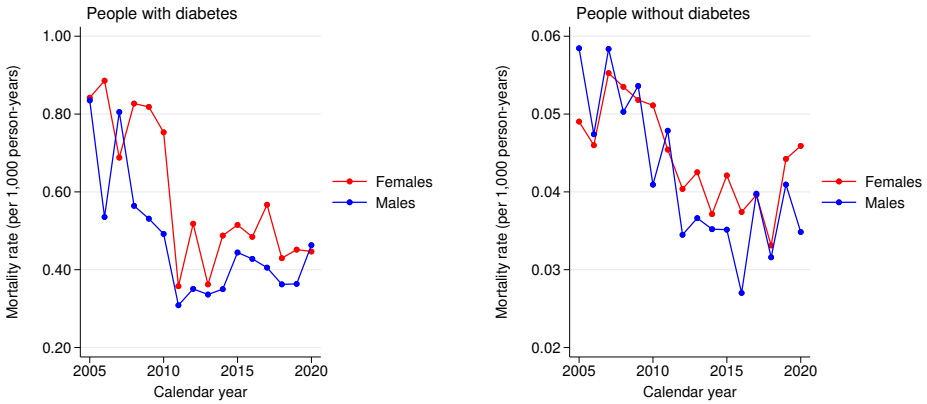


Figure 2.18: Crude mortality rate by cause of death, sex, and diabetes status. Liver disease. Canada (Alberta).

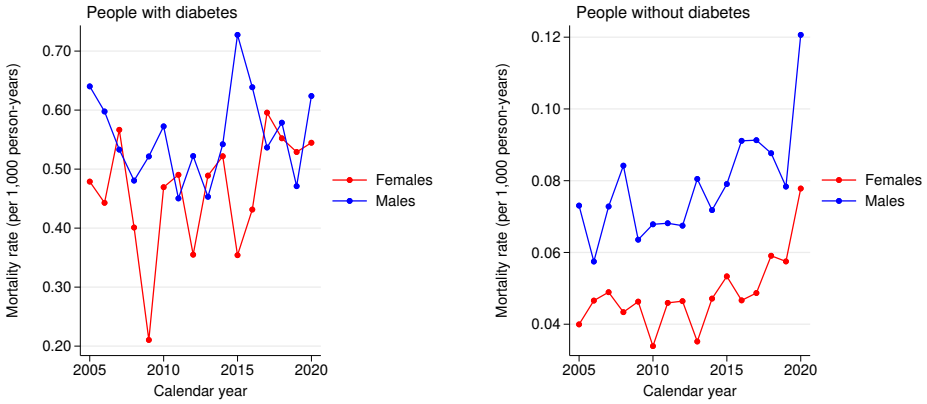


Figure 2.19: Crude mortality rate by cause of death, sex, and diabetes status. Cancer. Denmark.

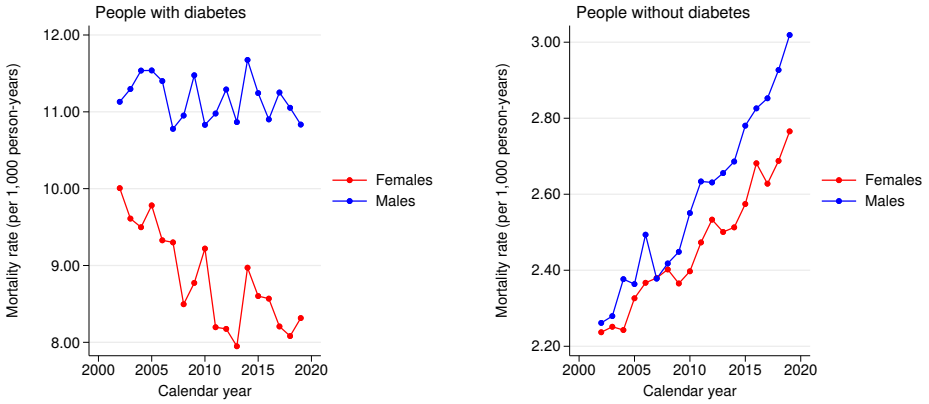


Figure 2.20: Crude mortality rate by cause of death, sex, and diabetes status. Cardiovascular disease. Denmark.

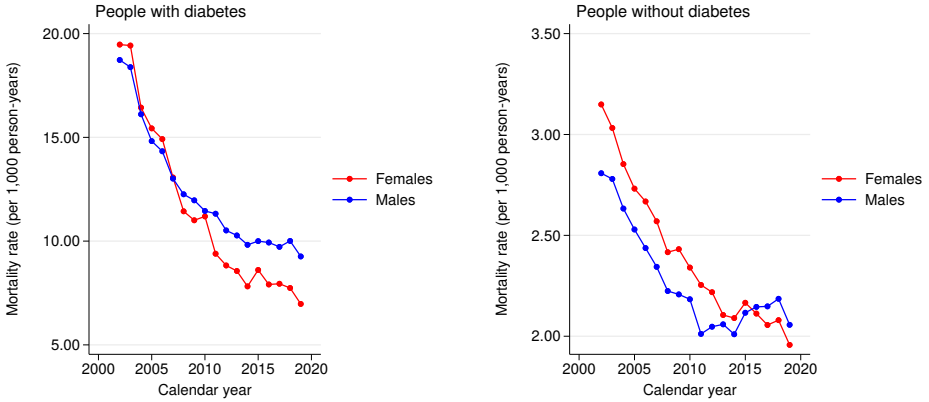


Figure 2.21: Crude mortality rate by cause of death, sex, and diabetes status. Chronic lower respiratory disease. Denmark.

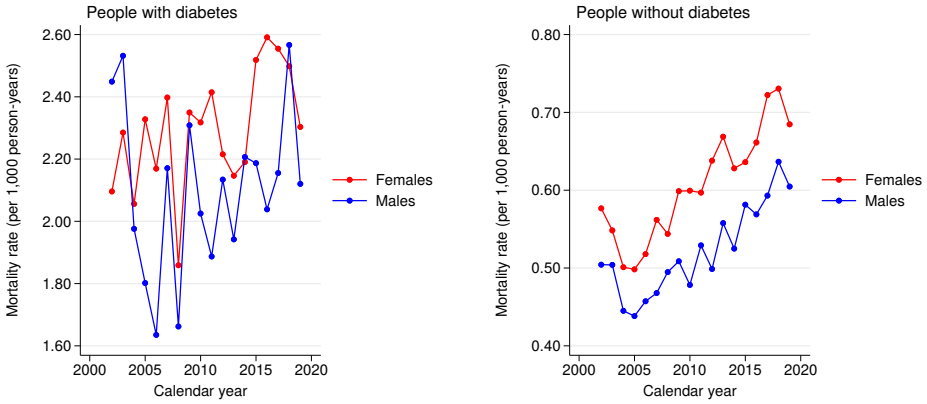


Figure 2.22: Crude mortality rate by cause of death, sex, and diabetes status. Dementia. Denmark.

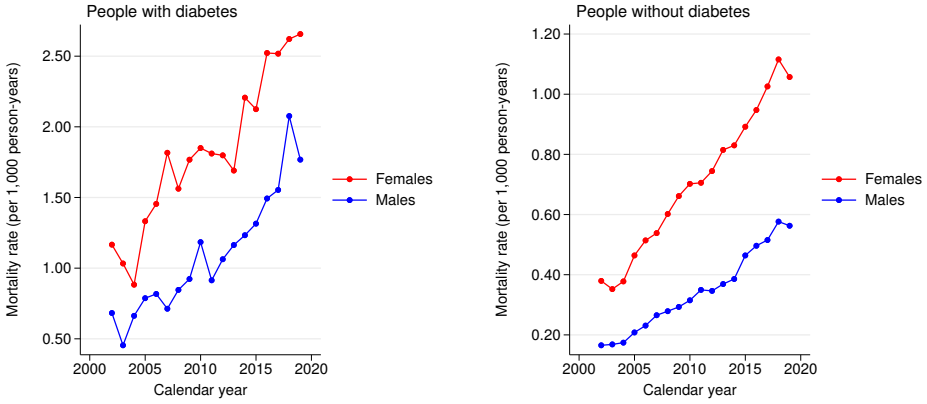


Figure 2.23: Crude mortality rate by cause of death, sex, and diabetes status. Diabetes. Denmark.

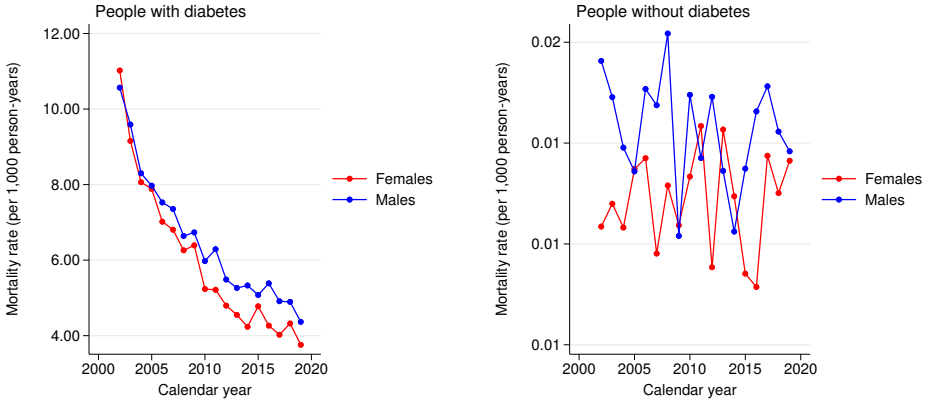


Figure 2.24: Crude mortality rate by cause of death, sex, and diabetes status. Infectious diseases. Denmark.

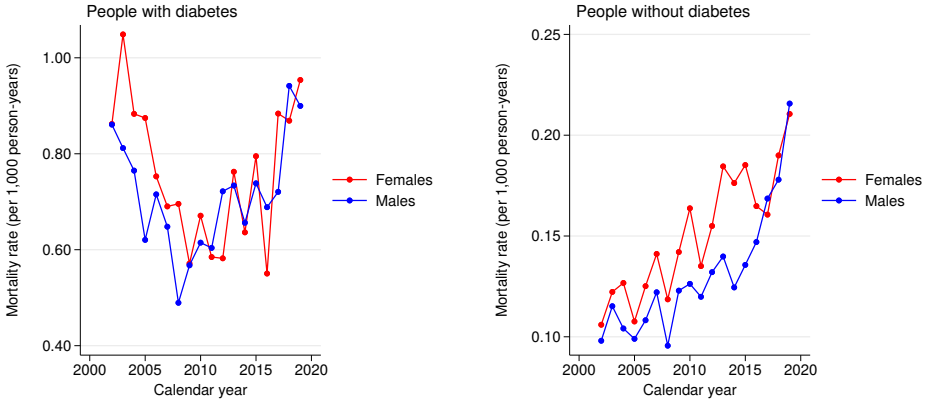


Figure 2.25: Crude mortality rate by cause of death, sex, and diabetes status. Influenza and pneumonia. Denmark.

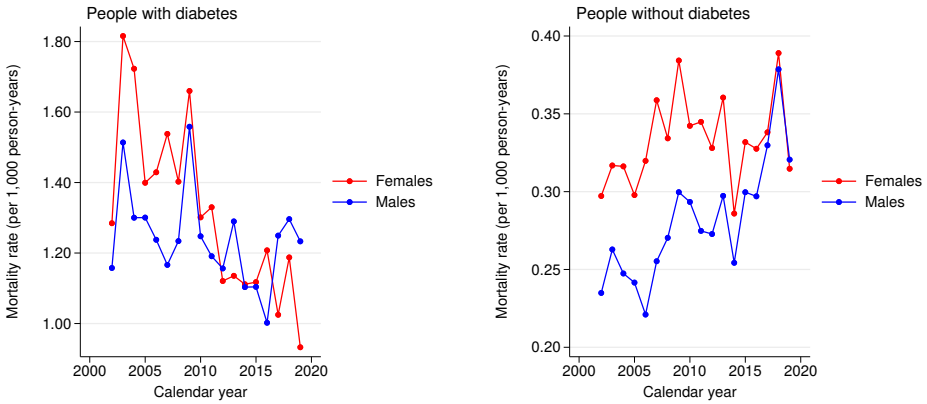


Figure 2.26: Crude mortality rate by cause of death, sex, and diabetes status. Kidney disease. Denmark.

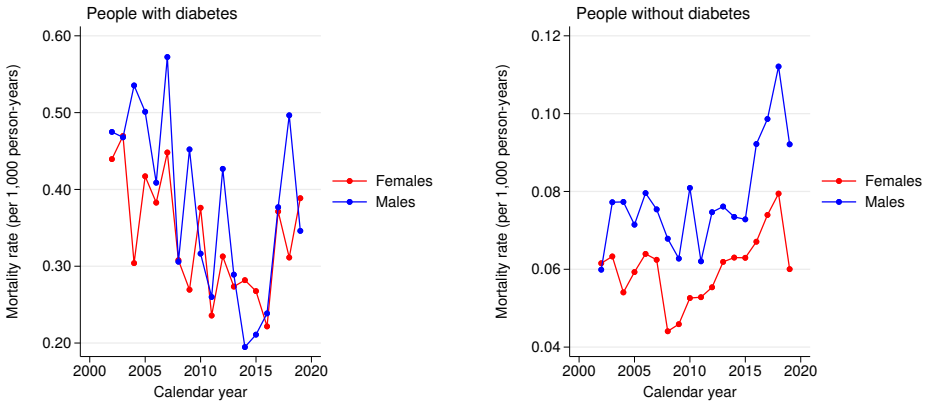


Figure 2.27: Crude mortality rate by cause of death, sex, and diabetes status. Liver disease. Denmark.

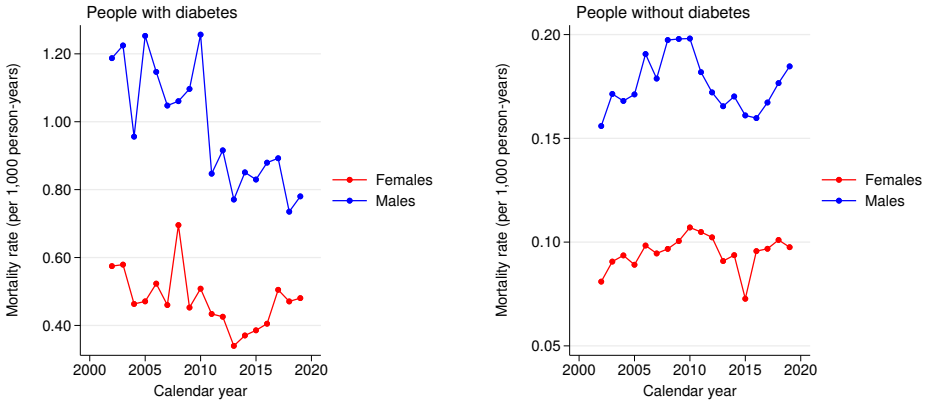


Figure 2.28: Crude mortality rate by cause of death, sex, and diabetes status. Cancer. Finland.

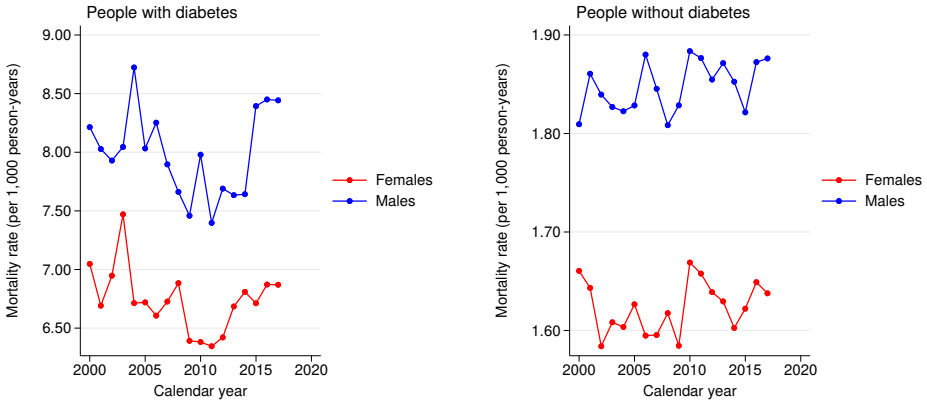


Figure 2.29: Crude mortality rate by cause of death, sex, and diabetes status. Cardiovascular disease. Finland.

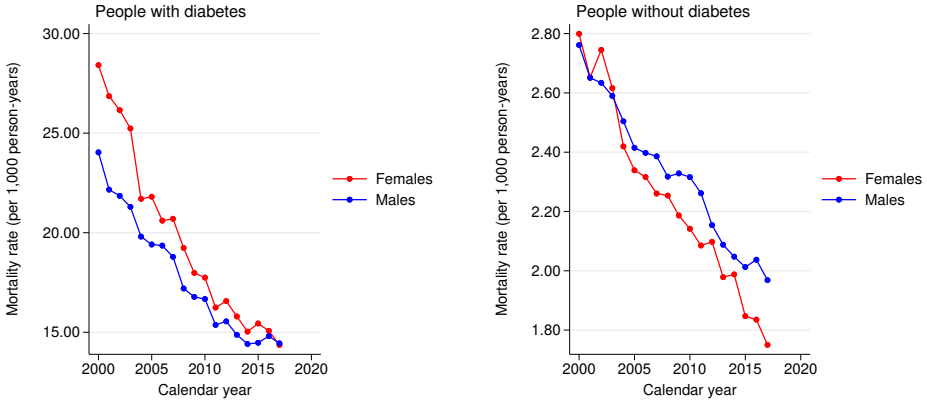


Figure 2.30: Crude mortality rate by cause of death, sex, and diabetes status. Chronic lower respiratory disease. Finland.

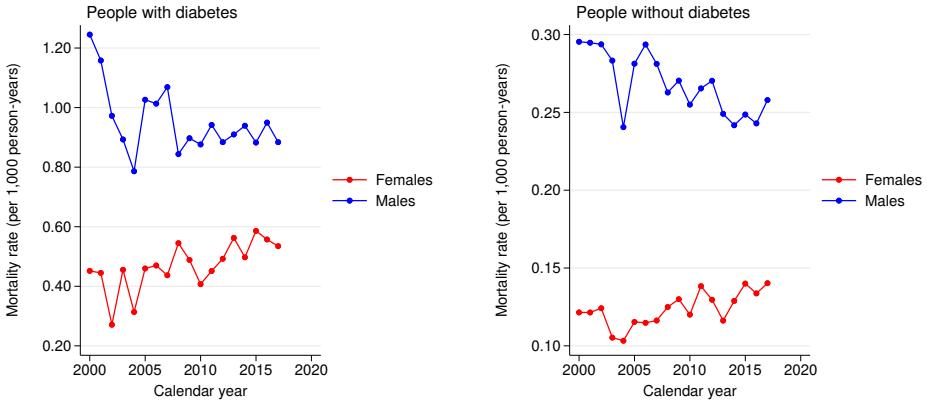


Figure 2.31: Crude mortality rate by cause of death, sex, and diabetes status. Dementia. Finland.

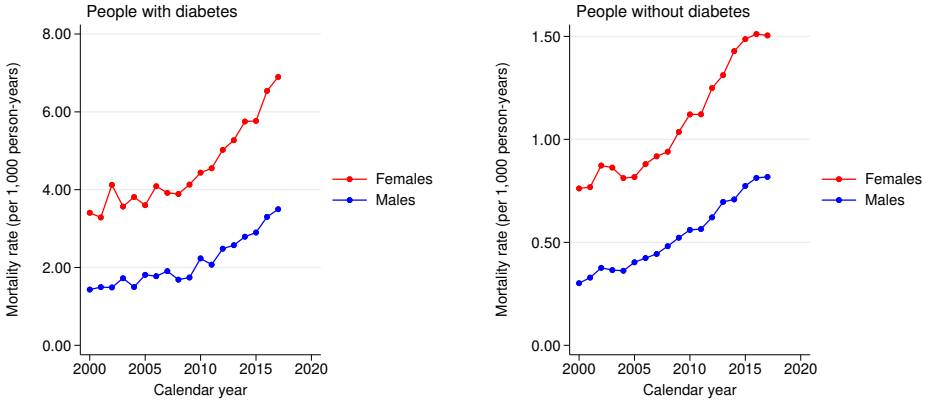


Figure 2.32: Crude mortality rate by cause of death, sex, and diabetes status. Diabetes. Finland.

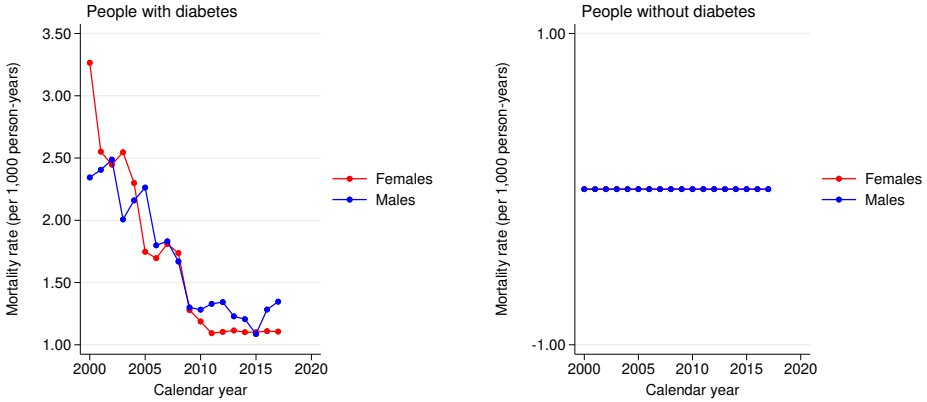


Figure 2.33: Crude mortality rate by cause of death, sex, and diabetes status. Infectious diseases. Finland.

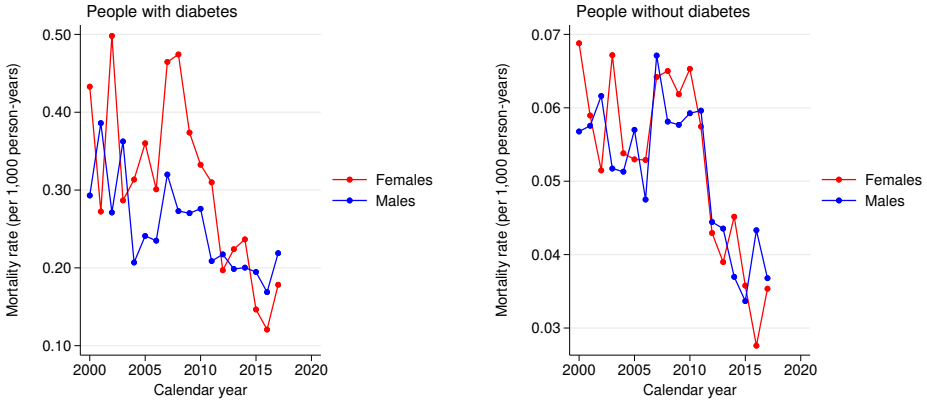


Figure 2.34: Crude mortality rate by cause of death, sex, and diabetes status. Influenza and pneumonia. Finland.

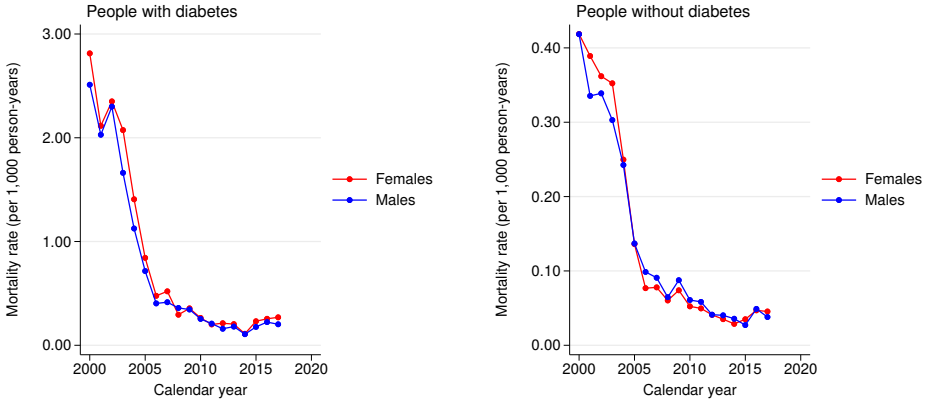


Figure 2.35: Crude mortality rate by cause of death, sex, and diabetes status. Kidney disease. Finland.

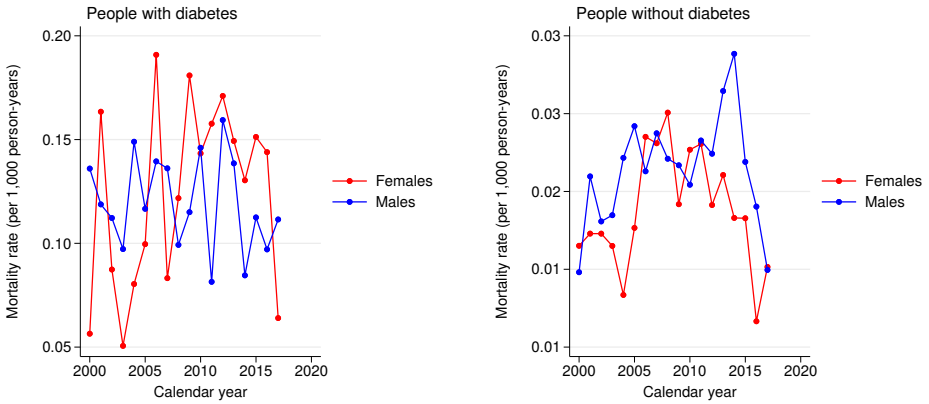


Figure 2.36: Crude mortality rate by cause of death, sex, and diabetes status. Liver disease. Finland.

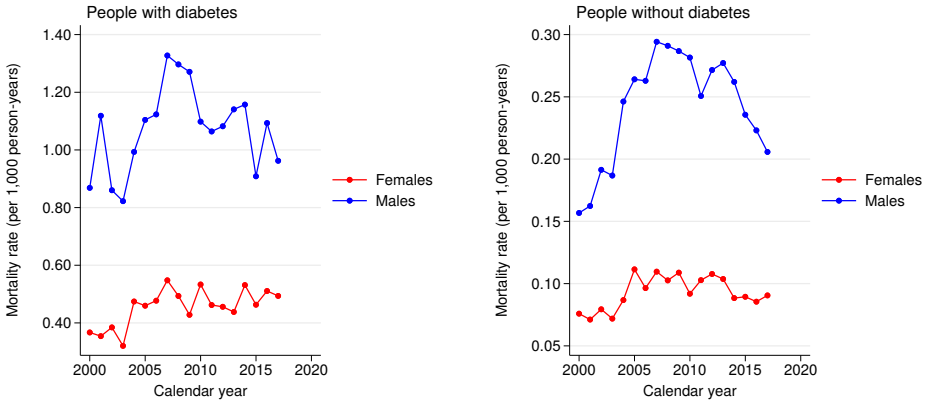


Figure 2.37: Crude mortality rate by cause of death, sex, and diabetes status. Cancer. France.

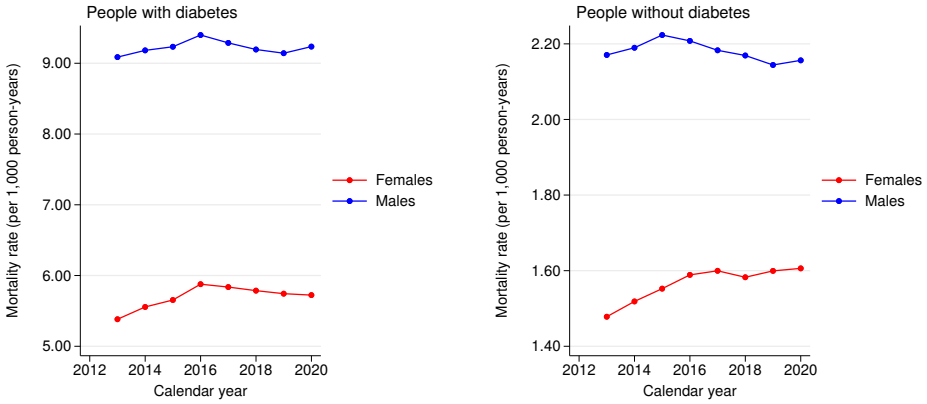


Figure 2.38: Crude mortality rate by cause of death, sex, and diabetes status. Cardiovascular disease. France.

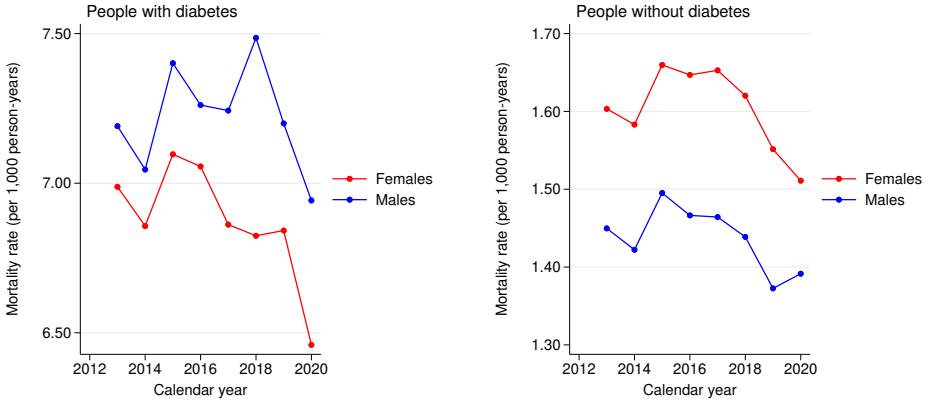


Figure 2.39: Crude mortality rate by cause of death, sex, and diabetes status. Chronic lower respiratory disease. France.

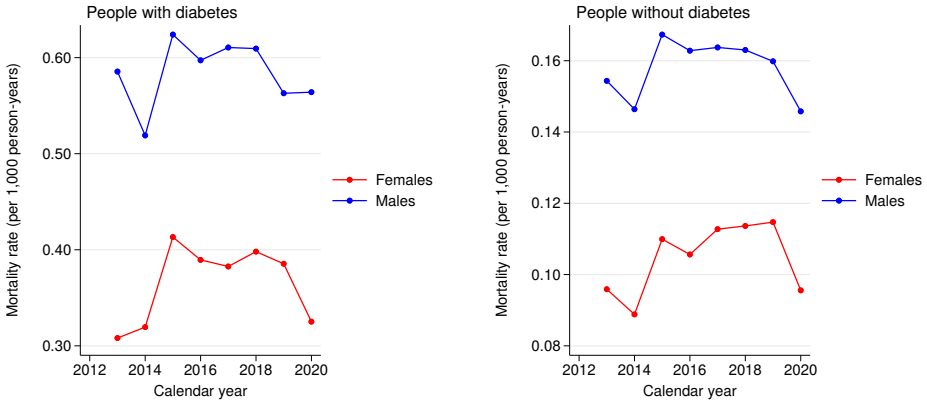


Figure 2.40: Crude mortality rate by cause of death, sex, and diabetes status. Dementia. France.

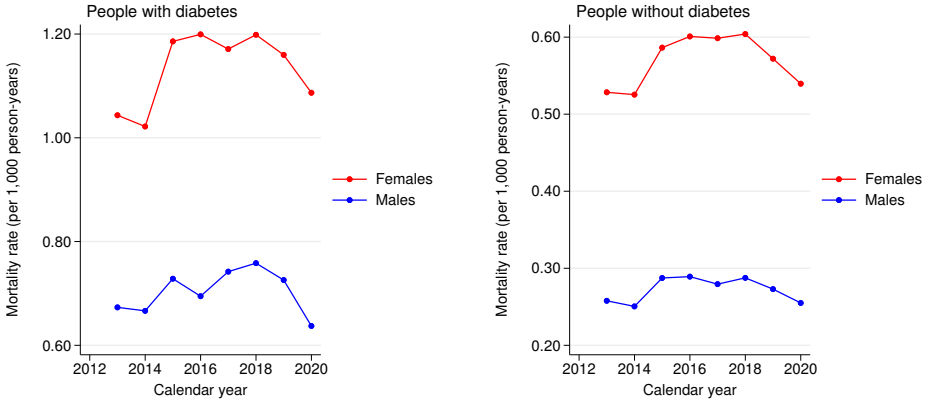


Figure 2.41: Crude mortality rate by cause of death, sex, and diabetes status. Diabetes. France.

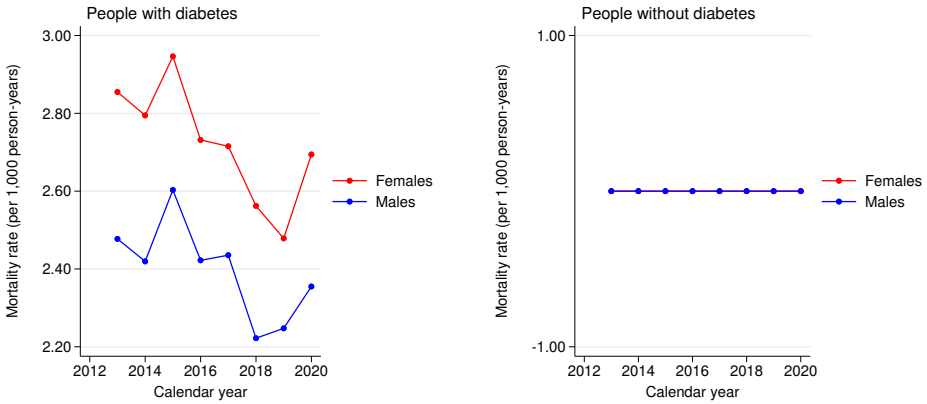


Figure 2.42: Crude mortality rate by cause of death, sex, and diabetes status. Infectious diseases. France.

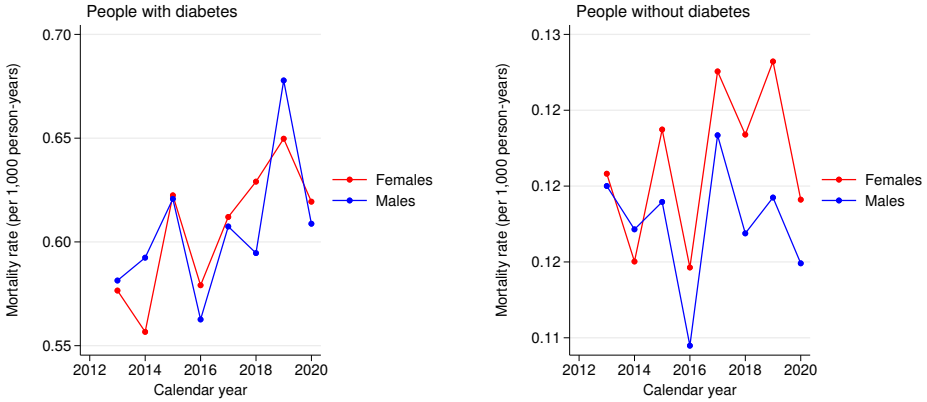


Figure 2.43: Crude mortality rate by cause of death, sex, and diabetes status. Influenza and pneumonia. France.

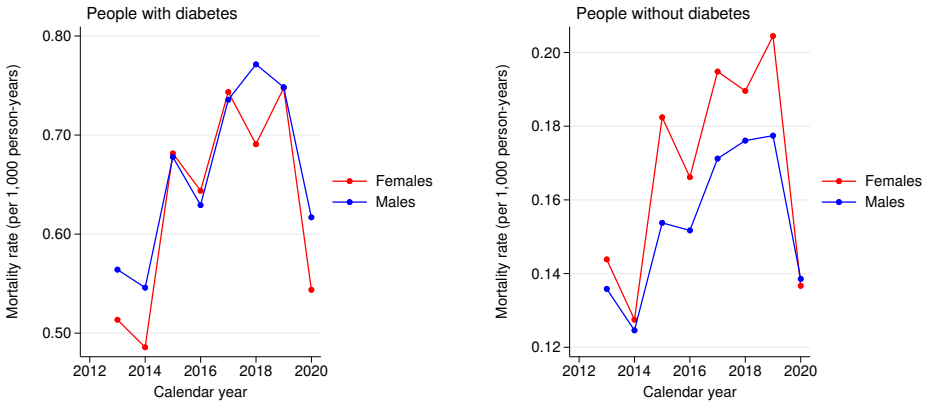


Figure 2.44: Crude mortality rate by cause of death, sex, and diabetes status. Kidney disease. France.

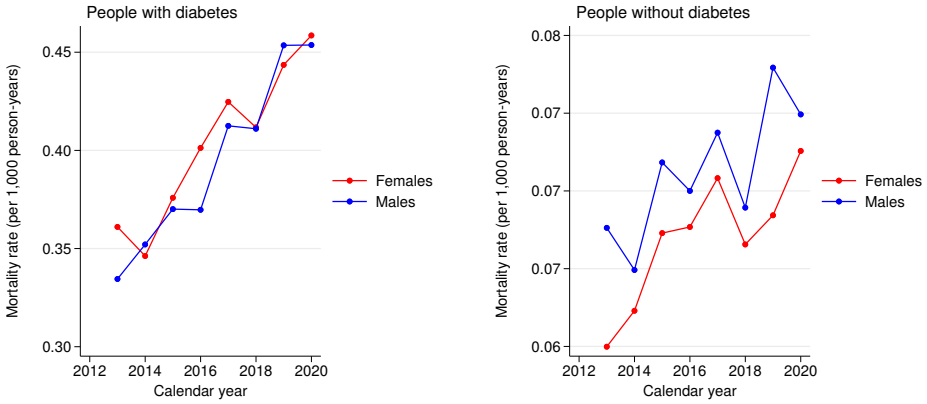


Figure 2.45: Crude mortality rate by cause of death, sex, and diabetes status. Liver disease. France.

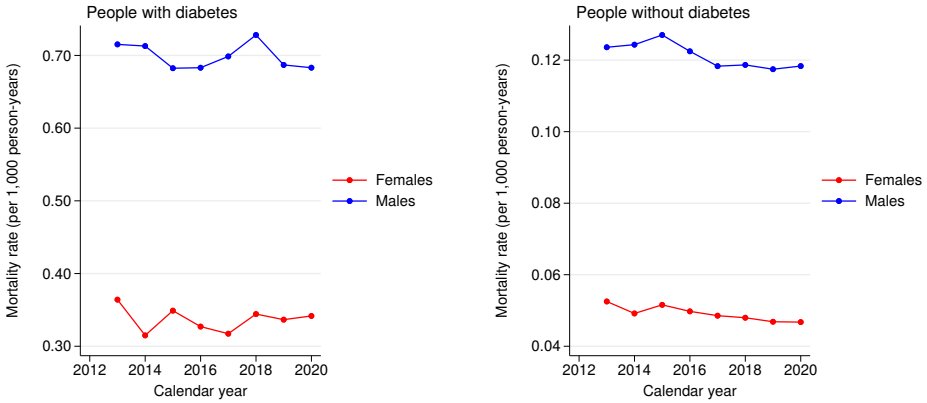


Figure 2.46: Crude mortality rate by cause of death, sex, and diabetes status. Cancer. Lithuania.

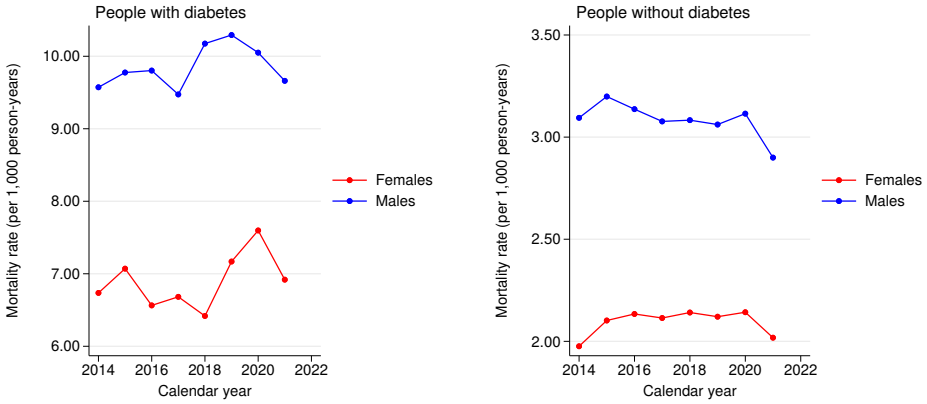


Figure 2.47: Crude mortality rate by cause of death, sex, and diabetes status. Cardiovascular disease. Lithuania.

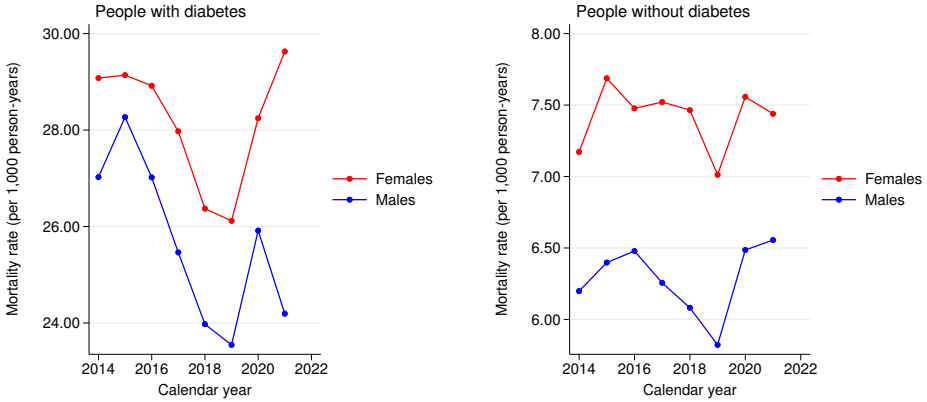


Figure 2.48: Crude mortality rate by cause of death, sex, and diabetes status. Chronic lower respiratory disease. Lithuania.

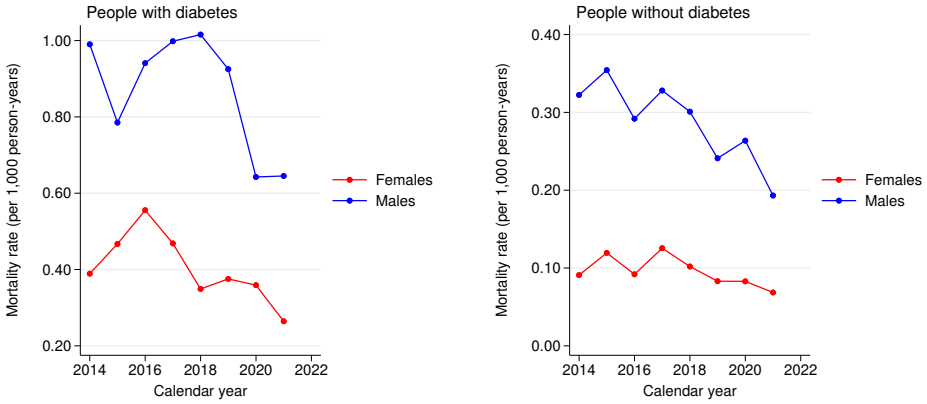


Figure 2.49: Crude mortality rate by cause of death, sex, and diabetes status. Dementia. Lithuania.

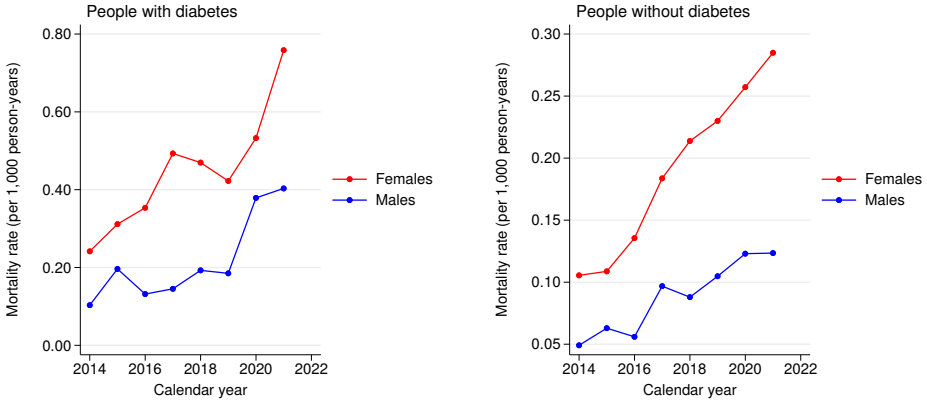


Figure 2.50: Crude mortality rate by cause of death, sex, and diabetes status. Diabetes. Lithuania.

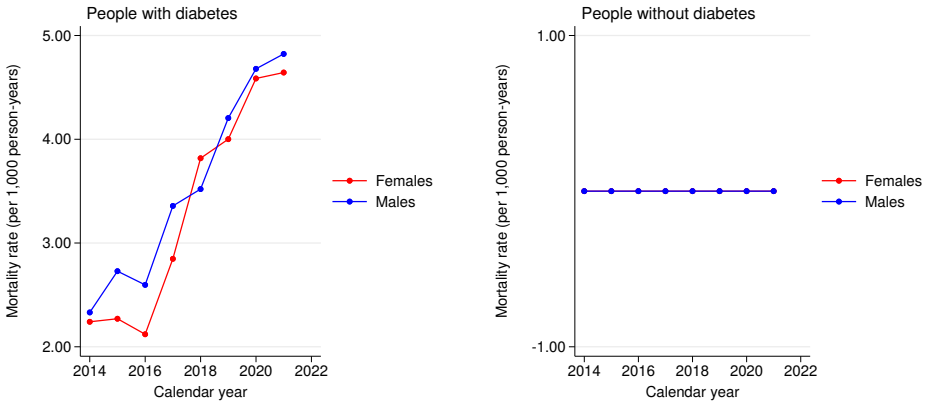


Figure 2.51: Crude mortality rate by cause of death, sex, and diabetes status. Infectious diseases. Lithuania.

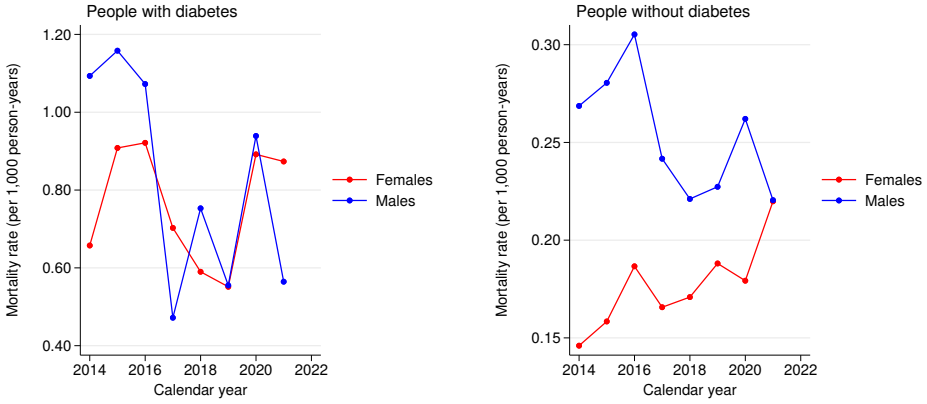


Figure 2.52: Crude mortality rate by cause of death, sex, and diabetes status. Influenza and pneumonia. Lithuania.

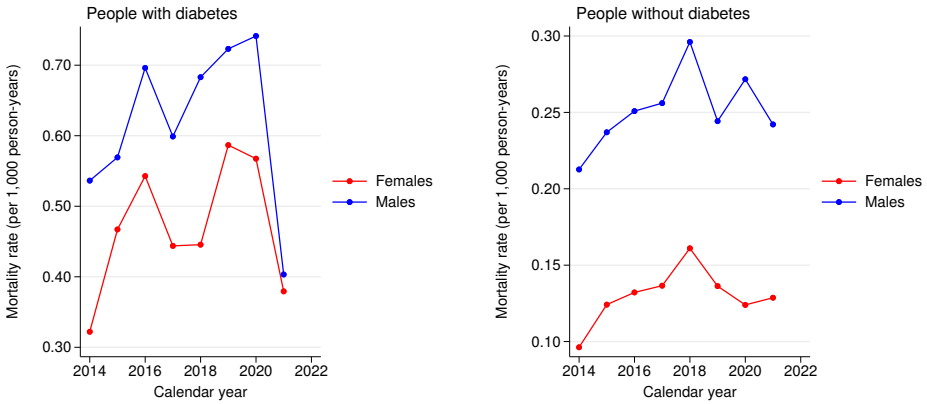


Figure 2.53: Crude mortality rate by cause of death, sex, and diabetes status. Kidney disease. Lithuania.

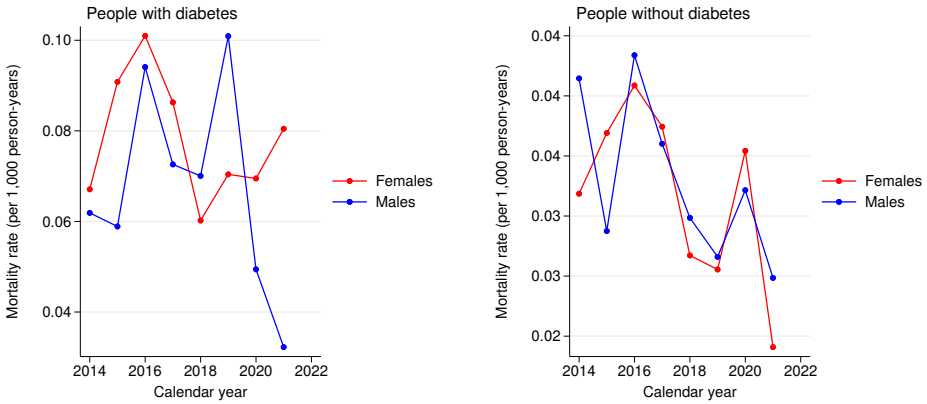


Figure 2.54: Crude mortality rate by cause of death, sex, and diabetes status. Liver disease. Lithuania.

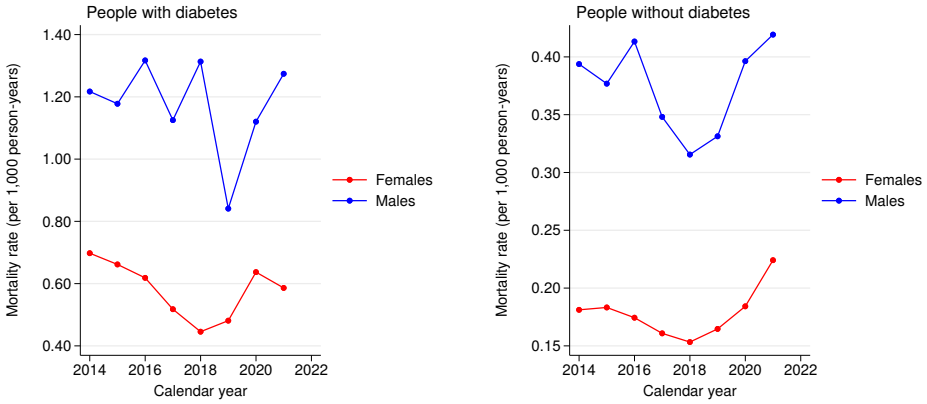


Figure 2.55: Crude mortality rate by cause of death, sex, and diabetes status. Cancer. Netherlands.

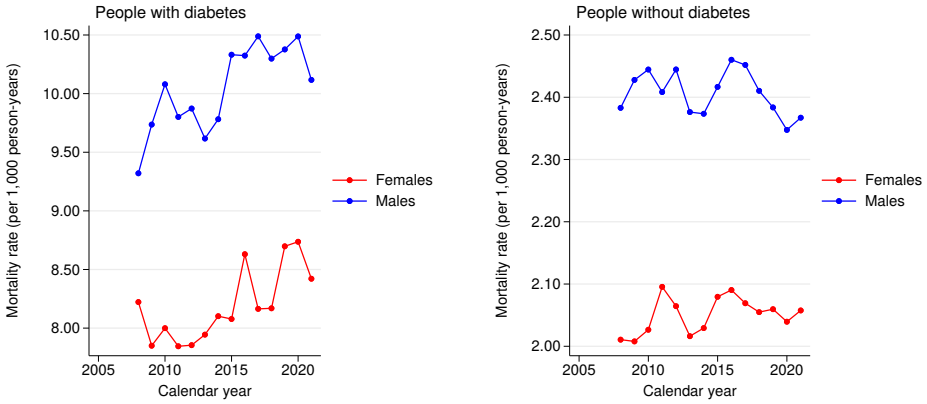


Figure 2.56: Crude mortality rate by cause of death, sex, and diabetes status. Cardiovascular disease. Netherlands.

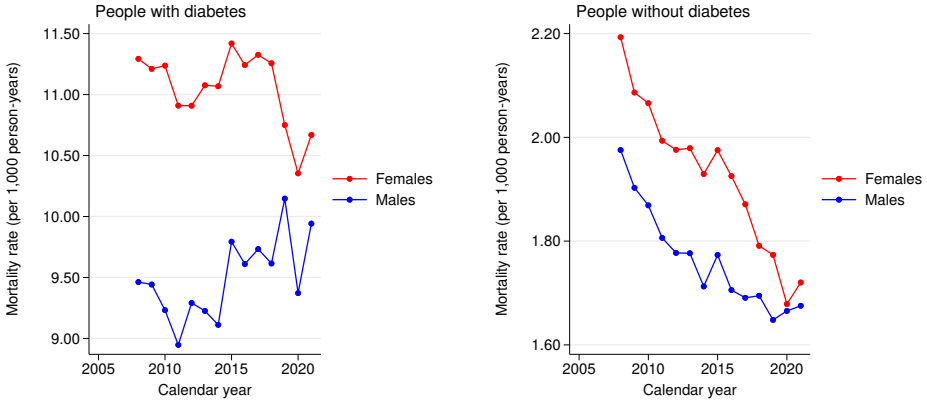


Figure 2.57: Crude mortality rate by cause of death, sex, and diabetes status. Chronic lower respiratory disease. Netherlands.

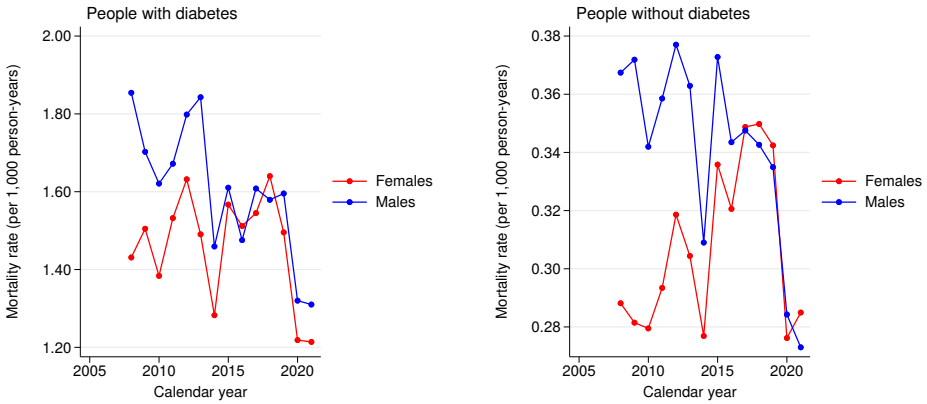


Figure 2.58: Crude mortality rate by cause of death, sex, and diabetes status. Dementia. Netherlands.

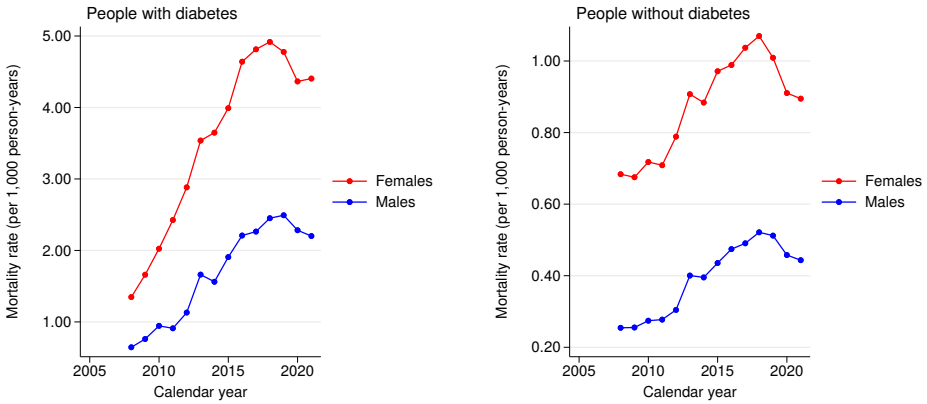


Figure 2.59: Crude mortality rate by cause of death, sex, and diabetes status. Diabetes. Netherlands.

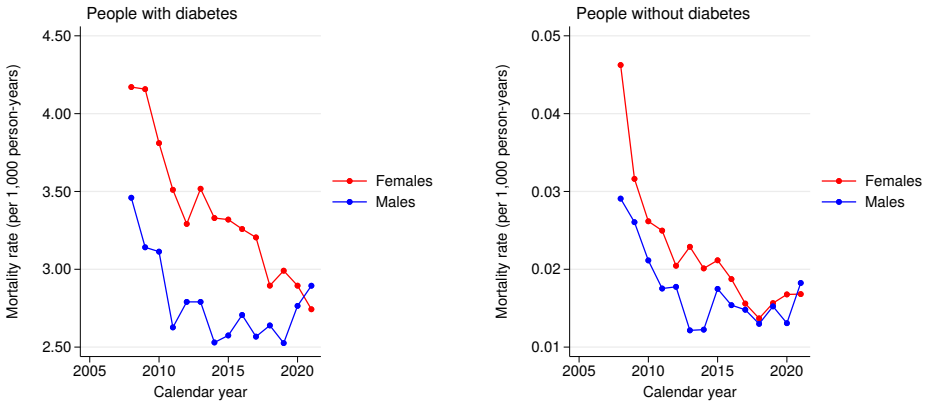


Figure 2.60: Crude mortality rate by cause of death, sex, and diabetes status. Infectious diseases. Netherlands.

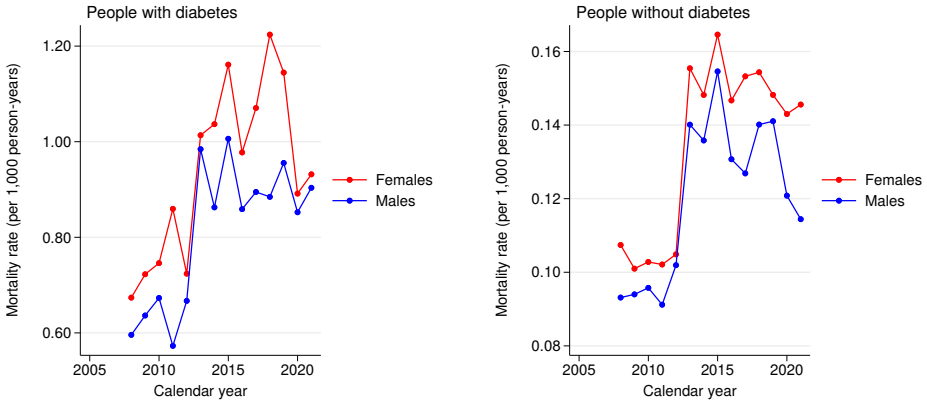


Figure 2.61: Crude mortality rate by cause of death, sex, and diabetes status. Influenza and pneumonia. Netherlands.

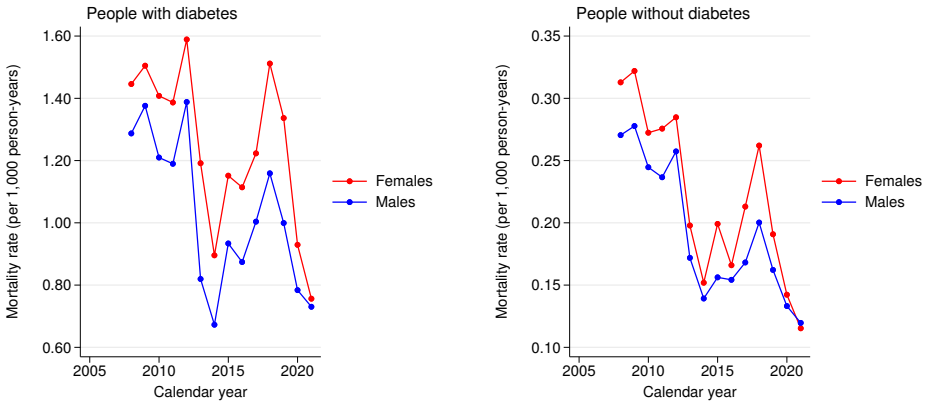


Figure 2.62: Crude mortality rate by cause of death, sex, and diabetes status. Kidney disease. Netherlands.

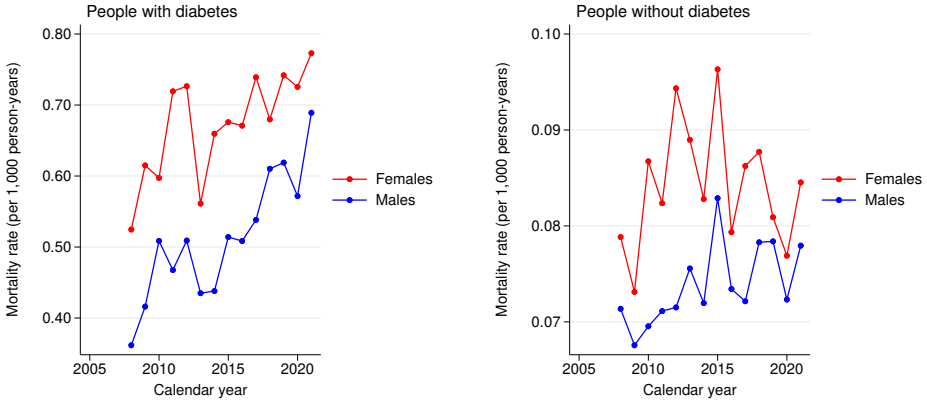


Figure 2.63: Crude mortality rate by cause of death, sex, and diabetes status. Liver disease. Netherlands.

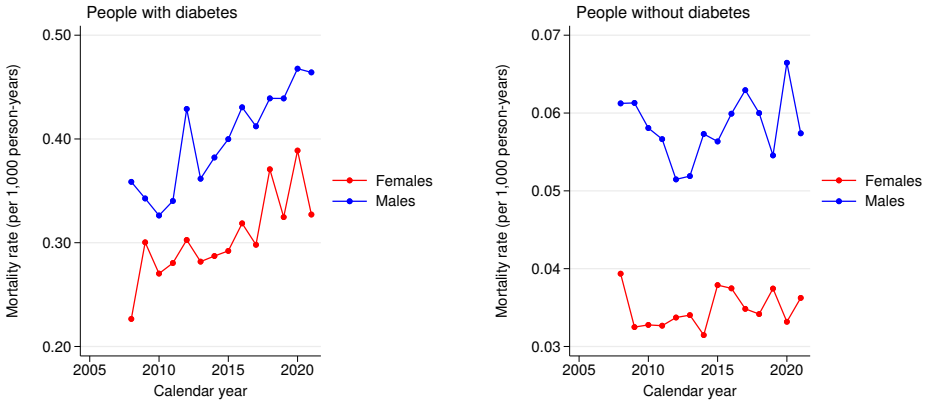


Figure 2.64: Crude mortality rate by cause of death, sex, and diabetes status. Cancer. Scotland.

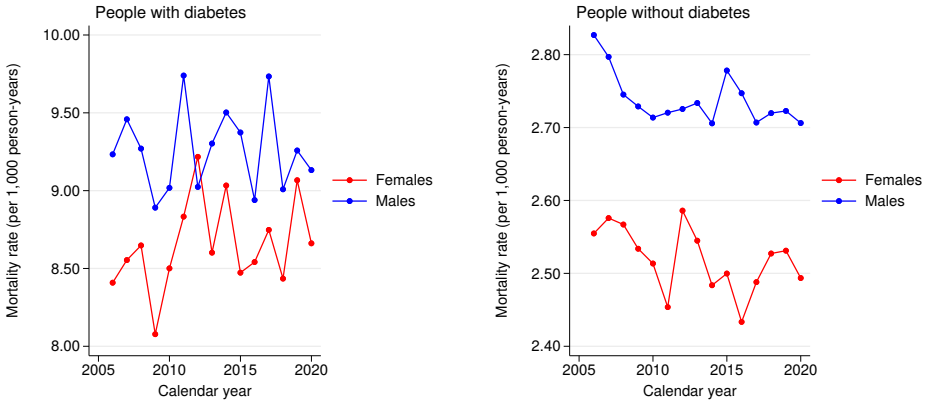


Figure 2.65: Crude mortality rate by cause of death, sex, and diabetes status. Cardiovascular disease. Scotland.

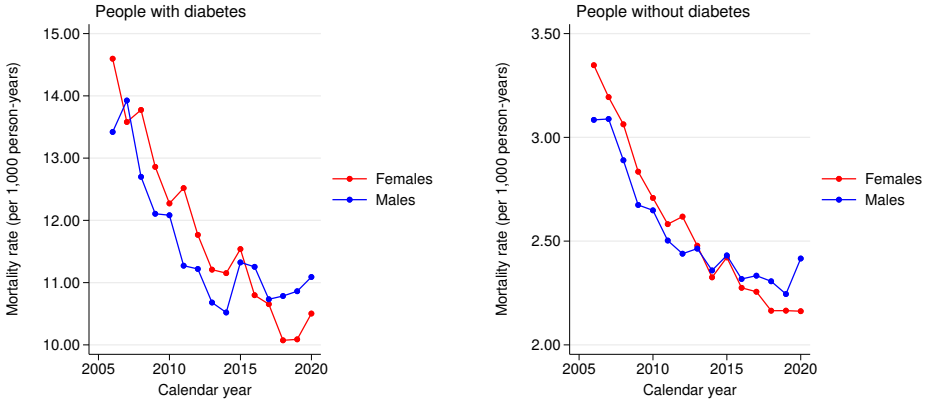


Figure 2.66: Crude mortality rate by cause of death, sex, and diabetes status. Chronic lower respiratory disease. Scotland.

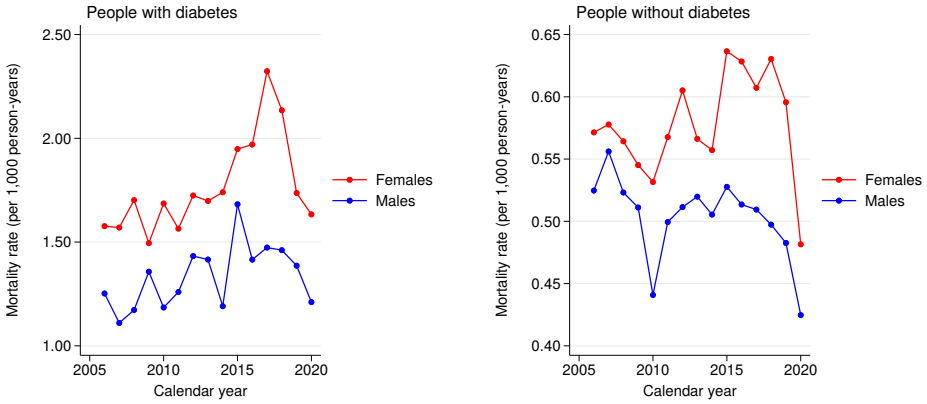


Figure 2.67: Crude mortality rate by cause of death, sex, and diabetes status. Dementia. Scotland.

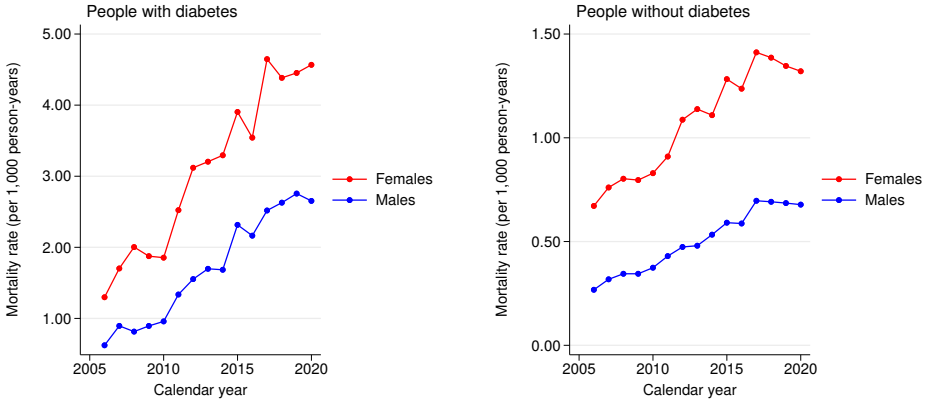


Figure 2.68: Crude mortality rate by cause of death, sex, and diabetes status. Diabetes. Scotland.

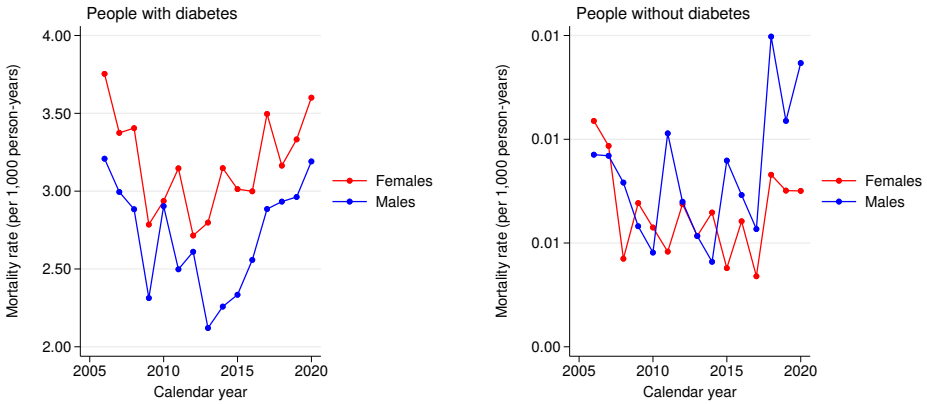


Figure 2.69: Crude mortality rate by cause of death, sex, and diabetes status. Infectious diseases. Scotland.

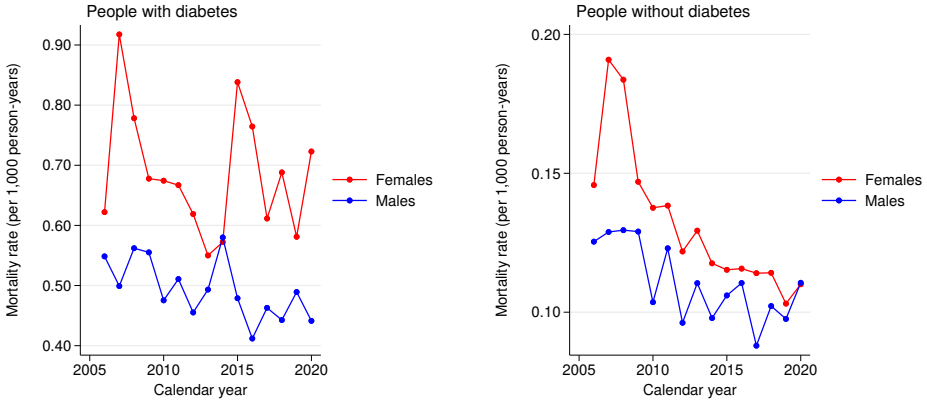


Figure 2.70: Crude mortality rate by cause of death, sex, and diabetes status. Influenza and pneumonia. Scotland.

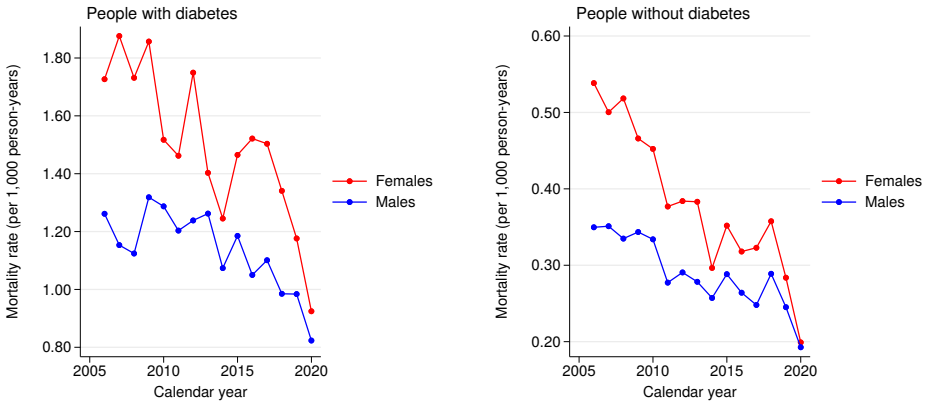


Figure 2.71: Crude mortality rate by cause of death, sex, and diabetes status. Kidney disease. Scotland.

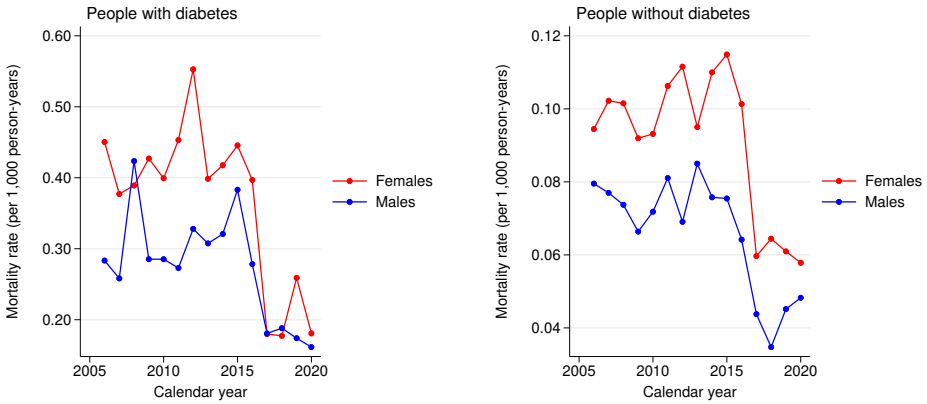


Figure 2.72: Crude mortality rate by cause of death, sex, and diabetes status. Liver disease. Scotland.

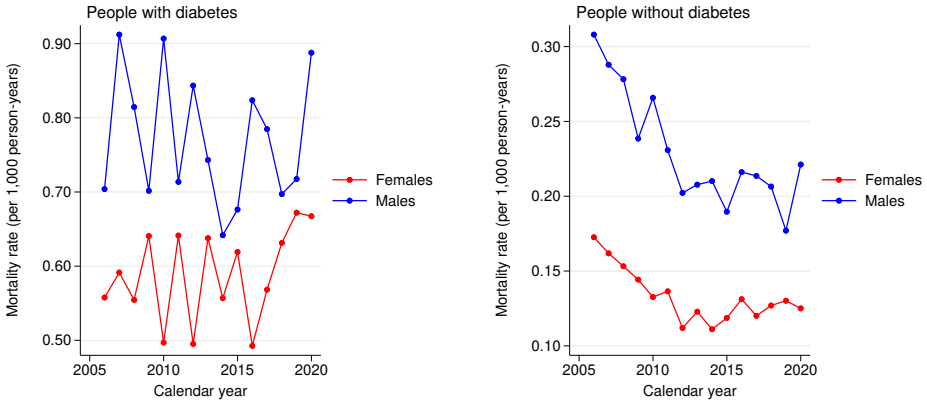


Figure 2.73: Crude mortality rate by cause of death, sex, and diabetes status. Cancer. South Korea.

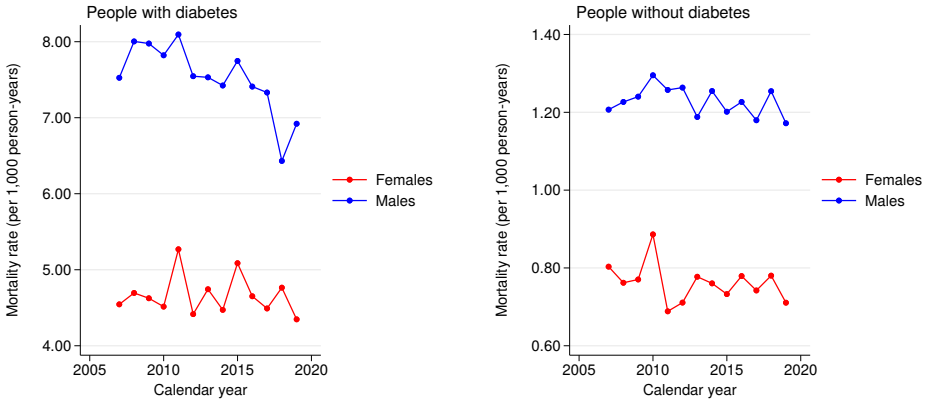


Figure 2.74: Crude mortality rate by cause of death, sex, and diabetes status. Cardiovascular disease. South Korea.

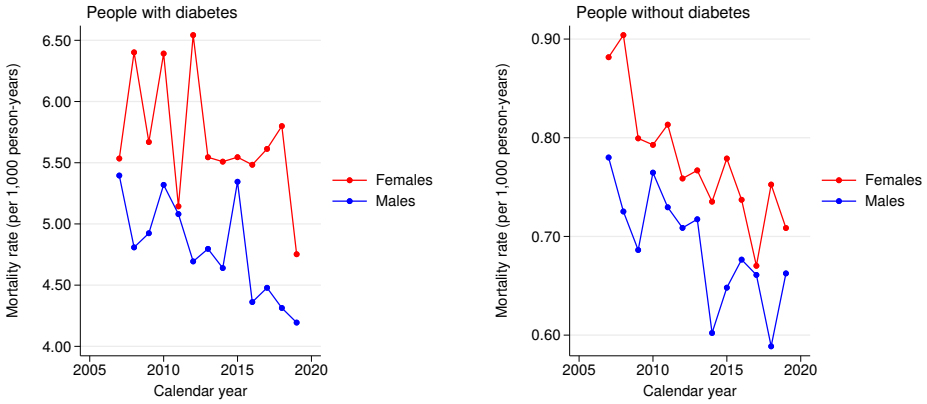


Figure 2.75: Crude mortality rate by cause of death, sex, and diabetes status. Chronic lower respiratory disease. South Korea.

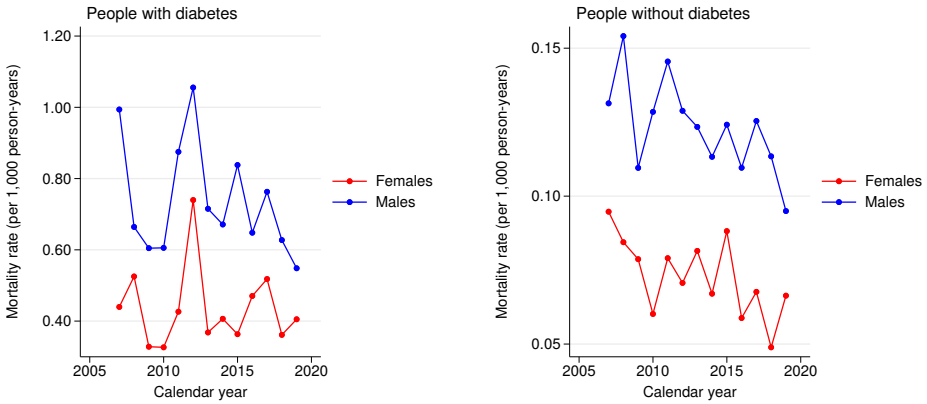


Figure 2.76: Crude mortality rate by cause of death, sex, and diabetes status. Dementia. South Korea.

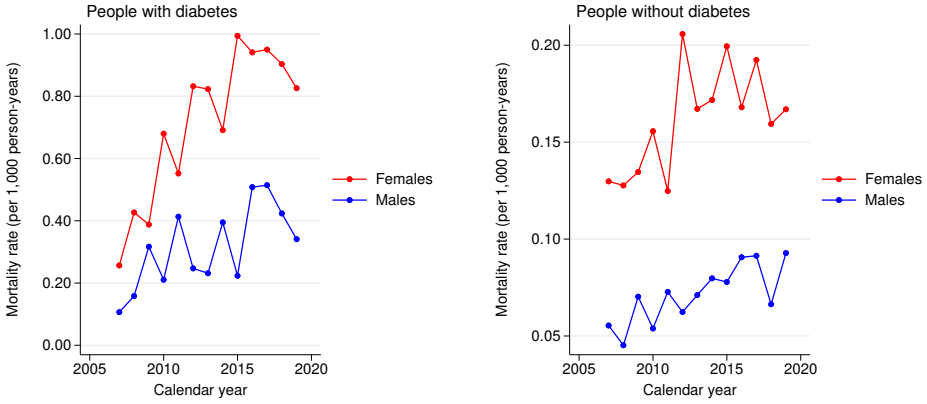


Figure 2.77: Crude mortality rate by cause of death, sex, and diabetes status. Diabetes. South Korea.

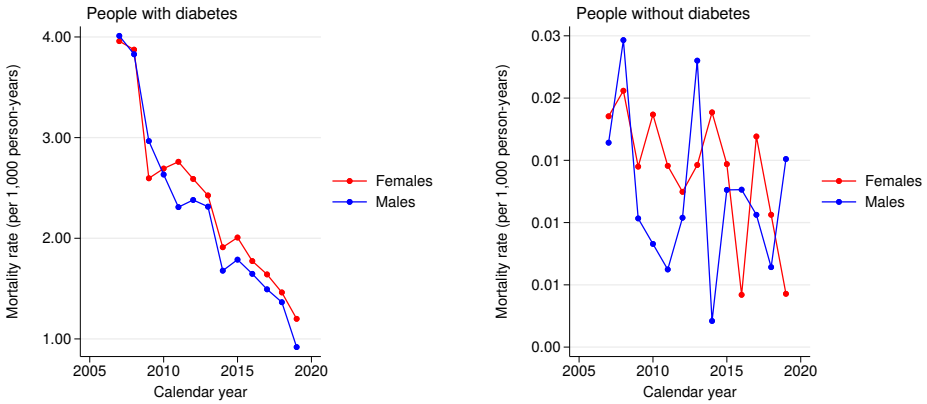


Figure 2.78: Crude mortality rate by cause of death, sex, and diabetes status. Infectious diseases. South Korea.

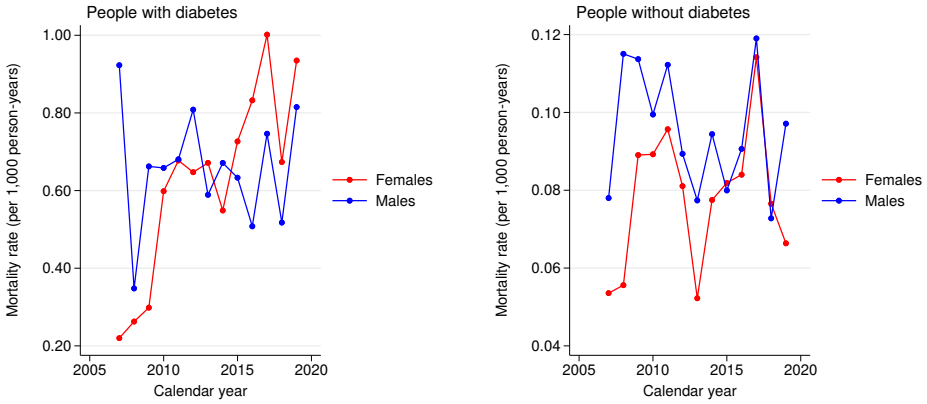


Figure 2.79: Crude mortality rate by cause of death, sex, and diabetes status. Influenza and pneumonia. South Korea.

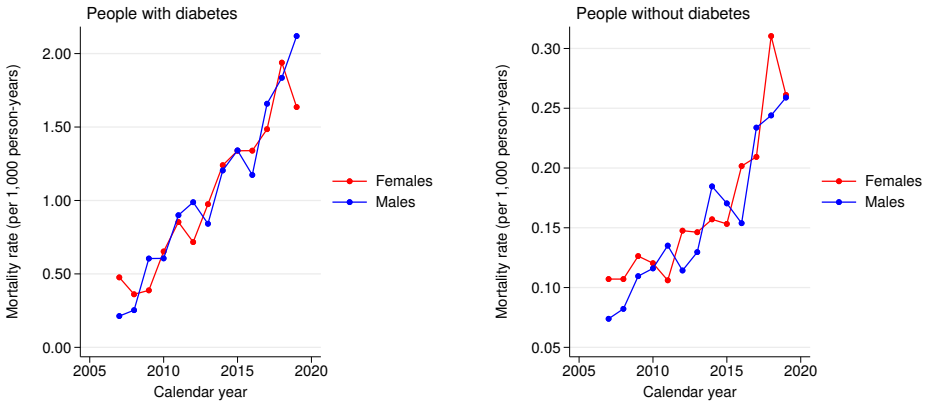


Figure 2.80: Crude mortality rate by cause of death, sex, and diabetes status. Kidney disease. South Korea.

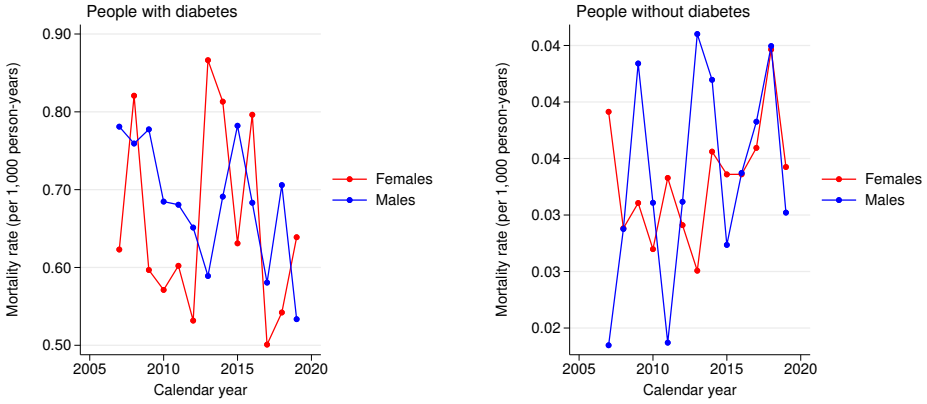
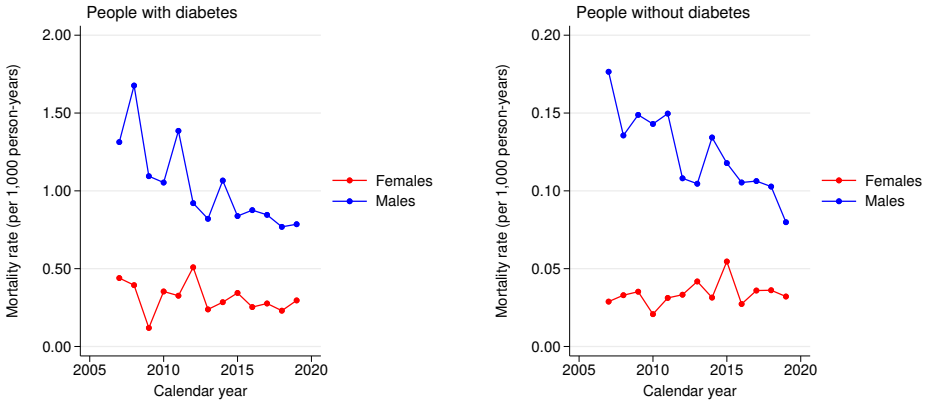


Figure 2.81: Crude mortality rate by cause of death, sex, and diabetes status. Liver disease. South Korea.



A few potential coding changes to note:

- Figure 2.8, Australia, kidney disease in 2013 – there is a big drop, suggesting a change in how kidney disease deaths were coded. Indeed, there were changes in coding kidney disease deaths by the Australian Burea of Statistics with the implementation of new software, which were as follows:
  - "N17-N19 Renal failure: There has been an increase in the number of conditions that have a causal relationship with renal failure. As a result, fewer deaths have been assigned to the code block N17-N19 as an underlying cause of death. Of note, E11 Non-insulin-dependent diabetes mellitus now combines with renal failure to form the code E11.2 Non-insulin- dependent diabetes mellitus with renal complications."
  - "N18 Chronic kidney disease: The title of code N18 has changed from Chronic renal failure to Chronic kidney disease. With the title update a coding change has occurred. Previously the term 'Chronic kidney disease' was coded to N03 Chronic nephritic syndrome. It is now coded to N18 Chronic kidney disease. Consequently, deaths assigned to N03 as an underlying cause have decreased."
- Figure 2.34, Finland, influenza and pneumonia from 2000-2005 – the continuous drop suggests a gradual change in how these deaths were coded.
- Figure 2.55 and Figure 2.56 and some others , Netherlands, Cancer and CVD in diabetes in 2007 – the big jump from 2007-2008 across multiple causes of death suggests something specific to the definition of diabetes might have changed this year. Indeed, looking at diabetes deaths in people without diabetes, the rate dramatically drops over this period.
- Figure 2.71, Scotland, kidney disease in 2017 – the big drop in 2017 for people with and without diabetes suggests a coding change.

We will still analyse this data, but we should note that the sudden changes could be coding changes that will influences estimates of the rate of change.

3 Cause-specific mortality rates

3.1 Methods

The methods are largely derived from Magliano et al. [**?**]. To generate age- and period-specific rates, which will be used to generate age-standardised rates, we will model mortality rates using age-period-cohort models [**?**]. Each model will be a Poisson model, parameterised using spline effects of age, period, and cohort (period − age), with log of person-years as the offset. Age is defined as above (i.e., the midpoint of the interval in most cases) and models are fit separately for each cause of death and country in people with and without diabetes and by sex. Because this will be $9 \times 9 \times 2 \times 2 = 324$ models, we won't check model fit for each model. Instead, to check model fit we will select a few at random and check the modelled and crude rates as well as the Pearson residuals. These models will be used to estimate mortality rates for single year ages and calendar years. These modelled rates will be used to generate age-standardised rates in people with and without diabetes by period, using direct standardisation (using the total diabetes population formed by pooling the consortium data).

```
*mkdir MD
foreach i in Australia Canada Denmark Finland France Lithuania Netherlands Scotland SKorea {
foreach ii in can cvd res azd dmd inf flu ckd liv {
foreach iii in dm nondm {
if "`ii´" == "dmd" & "`iii´" == "nondm" {
}
else {
foreach iiii in 0 1 {
use `i´, clear
keep if sex == `iiii´
replace calendar = calendar-2009.5
gen coh = calendar-age_`iii´
centile(age_`iii´), centile(5 35 65 95)
local A1 = r(c_1)
local A2 = r(c_2)
local A3 = r(c_3)
local A4 = r(c_4)
mkspline agesp = age_`iii´, cubic knots(`A1´ `A2´ `A3´ `A4´)
su(calendar), detail
local rang = r(max)-r(min)
if `rang´ < 10 {
centile calendar, centile(25 75)
local CK1 = r(c_1)
local CK2 = r(c_2)
mkspline timesp = calendar, cubic knots(`CK1´ `CK2´)
}
else if inrange(`rang´,10,14.9) {
centile calendar, centile(10 50 90)
local CK1 = r(c_1)
local CK2 = r(c_2)
local CK3 = r(c_3)
mkspline timesp = calendar, cubic knots(`CK1´ `CK2´ `CK3´)
}
else {
centile calendar, centile(5 35 65 95)
local CK1 = r(c_1)
local CK2 = r(c_2)
local CK3 = r(c_3)
local CK4 = r(c_4)
mkspline timesp = calendar, cubic knots(`CK1´ `CK2´ `CK3´ `CK4´)
}
centile(coh), centile(5 35 65 95)
local CO1 = r(c_1)
local CO2 = r(c_2)
```

```
local CO3 = r(c_3)
local CO4 = r(c_4)
mkspline cohsp = coh, cubic knots(`CO1´ `CO2´ `CO3´ `CO4´)
poisson `ii´_d_`iii´ agesp* timesp* cohsp*, exposure(pys_`iii´)
predict pred
gen OC = "`ii´"
gen DM = "`iii´"
save MD/RC_pred_`i´_`ii´_`iii´_`iiii´, replace
keep calendar
bysort cal : keep if _n == 1
expand 10
bysort cal : replace cal = cal+((_n-6)/10)
expand 700
bysort cal : gen age_`iii´ = (_n/10)+29.9
gen pys_`iii´ = 1
gen coh = calendar-age
mkspline agesp = age, cubic knots(`A1´ `A2´ `A3´ `A4´)
if `rang´ < 9.99 {
mkspline timesp = calendar, cubic knots(`CK1´ `CK2´)
}
else if inrange(`rang´,10,14.99) {
mkspline timesp = calendar, cubic knots(`CK1´ `CK2´ `CK3´)
}
else {
mkspline timesp = calendar, cubic knots(`CK1´ `CK2´ `CK3´ `CK4´)
}
mkspline cohsp = coh, cubic knots(`CO1´ `CO2´ `CO3´ `CO4´)
predict _Rate, ir
predict errr, stdp
replace _Rate = _Rate*1000
gen lb = exp(ln(_Rate)-1.96*errr)
gen ub = exp(ln(_Rate)+1.96*errr)
gen country = "`i´"
gen OC = "`ii´"
gen DM = "`iii´"
gen sex = `iiii´
replace cal = cal+2010
tostring age_`iii´, replace force format(%9.1f)
destring age_`iii´, replace
save MD/R_`i´_`ii´_`iii´_`iiii´, replace
}
}
}
}
}
set seed 1312
clear
gen A =.
foreach i in Australia Canada Denmark Finland France Lithuania Netherlands Scotland SKorea {
foreach ii in can cvd res azd dmd inf flu ckd liv {
foreach iii in dm nondm {
if "`ii´" == "dmd" & "`iii´" == "nondm" {
}
else {
foreach iiii in 0 1 {
local B = runiform()
append using MD/RC_pred_`i´_`ii´_`iii´_`iiii´
recode A .=`B´
keep if A > 0.985
}
}
}
}
}
br
save RCc, replace
set seed 1312
clear
```

```
gen A =.
foreach i in Australia Canada Denmark Finland France Lithuania Netherlands Scotland SKorea {
foreach ii in can cvd res azd dmd inf flu ckd liv {
foreach iii in dm nondm {
if "`ii´" == "dmd" & "`iii´" == "nondm" {
}
else {
foreach iiii in 0 1 {
local B = runiform()
append using MD/R_`i´_`ii´_`iii´_`iiii´
recode A .=`B´
keep if A > 0.985
}
}
}
}
}
save Rc, replace
```

```
. use Rc, clear
. bysort A : keep if _n == 1
(930,990 observations deleted)
. list country OC DM sex
```

	country	OC	DM	sex
1.	Finland	cvd	dm	0
2.	France	azd	nondm	1
3.	Australia	ckd	nondm	1
4.	Denmark	inf	dm	0
5.	France	liv	nondm	0
6.	Netherlands	azd	dm	1
7.	Lithuania	ckd	nondm	1
8.	Denmark	azd	dm	1
9.	Lithuania	cvd	dm	1
10.	Denmark	cvd	dm	1

```
forval i = 1/10 {
use Rc, clear
bysort A : keep if _n == 1
local c=country[`i´]
local o=OC[`i´]
local d=DM[`i´]
local s=sex[`i´]
if "`o´" == "can" {
local oo = "Cancer"
}
if "`o´" == "cvd" {
local oo = "Cardiovascular disease"
}
if "`o´" == "res" {
local oo = "Chronic lower respiratory disease"
}
if "`o´" == "azd" {
local oo = "Dementia"
}
if "`o´" == "dmd" {
local oo = "Diabetes"
}
if "`o´" == "inf" {
local oo = "Infectious diseases"
}
if "`o´" == "flu" {
local oo = "Influenza and pneumonia"
```

```
}
if "`o´" == "ckd" {
local oo = "Kidney disease"
}
if "`o´" == "liv" {
local oo = "Liver disease"
}
if "`d´" == "dm" {
local dd = "with"
}
if "`d´" == "nondm" {
local dd = "without"
}
if `s´ == 0 {
local ss = "Females"
}
if `s´ == 1 {
local ss = "Males"
}
use Rc, clear
keep if country == "`c´" & OC == "`o´" & sex == `s´ & DM == "`d´"
drop pys_nondm pys_dm
merge 1:1 age_`d´ sex cal using `c´
drop if _merge == 2
gen rate = 1000*`o´_d_`d´/pys_`d´
egen calmen = mean(calendar)
replace calmen = round(calmen,1)
local cmu = calmen[1]
twoway ///
(rarea ub lb age_`d´ if cale == `cmu´, color(black%30) fintensity(inten80) lwidth(none)) ///
(line _Rate age_`d´ if cale == `cmu´, color(black)) ///
(scatter rate age_`d´ if cale == `cmu´ & rate !=0, col(black)) ///
, graphregion(color(white)) ylabel(, angle(0) glpattern(solid) glcolor(gs10%20)) ///
xlabel(30(10)100, nogrid) ytitle("Mortality rate (per 1000 person-years)") ///
xtitle(Age) yscale(nolog) legend(order( ///
2 "Modelled" ///
3 "Crude" ///
) ring(0) cols(1) position(11) region(lcolor(none) col(none))) ///
title("`c´, `oo´, `ss´ `dd´ diabetes", col(black) placement(west) size(medium))
graph save GPH/Rc_`c´_`o´_`d´_`s´_age, replace
twoway ///
(rarea ub lb cale if age_`d´ == 45, color(gs0%30) fintensity(inten80) lwidth(none)) ///
(line _Rate cale if age_`d´ == 45, color(gs0)) ///
(scatter rate cale if age_`d´ == 45 & rate !=0, col(gs0)) ///
(rarea ub lb cale if age_`d´ == 65, color(gs5%30) fintensity(inten80) lwidth(none)) ///
(line _Rate cale if age_`d´ == 65, color(gs5)) ///
(scatter rate cale if age_`d´ == 65 & rate !=0, col(gs5)) ///
(rarea ub lb cale if age_`d´ == 85, color(gs10%30) fintensity(inten80) lwidth(none)) ///
(line _Rate cale if age_`d´ == 85, color(gs10)) ///
(scatter rate cale if age_`d´ == 85 & rate !=0, col(gs10)) ///
, graphregion(color(white)) ylabel(, angle(0) glpattern(solid) glcolor(gs10%20)) ///
ytitle("Mortality rate (per 1000 person-years)") ///
xtitle(Year) yscale(log) legend(order( ///
2 "Modelled" ///
2 "45" 5 "65" 8 "85" ///
3 "Crude" ///
3 "40-49" 6 "60-69" 9 "80-89" ///
) ring(0) cols(4) position(11) region(lcolor(none) col(none))) ///
title("`c´, `oo´, `ss´ `dd´ diabetes", col(black) placement(west) size(medium))
graph save GPH/Rc_`c´_`o´_`d´_`s´_period, replace
use RCc, clear
replace coh= coh+2010
keep if country == "`c´" & OC == "`o´" & sex == `s´ & DM == "`d´"
gen res = (`o´_d_`d´-pred)/sqrt(pred)
twoway ///
(scatter res age_`d´, col(black)) ///
, legend(off) ///
graphregion(color(white)) ///
```

```
ylabel(, format(%9.0f) grid angle(0) glpattern(solid) glcolor(gs10%20)) ///
ytitle("Pearson residuals", margin(a+2)) ///
xtitle("Age (years)") ///
title("`c´, `oo´, `ss´ `dd´ diabetes", col(black) placement(west) size(medium))
graph save GPH/RCc_`c´_`o´_`d´_`s´_age, replace
twoway ///
(scatter res cale, col(black)) ///
, legend(off) ///
graphregion(color(white)) ///
ylabel(, format(%9.0f) grid angle(0) glpattern(solid) glcolor(gs10%20)) ///
ytitle("Pearson residuals", margin(a+2)) ///
xtitle("Period") ///
title("`c´, `oo´, `ss´ `dd´ diabetes", col(black) placement(west) size(medium))
graph save GPH/RCc_`c´_`o´_`d´_`s´_period, replace
twoway ///
(scatter res coh, col(black)) ///
, legend(off) ///
graphregion(color(white)) ///
ylabel(, format(%9.0f) grid angle(0) glpattern(solid) glcolor(gs10%20)) ///
ytitle("Pearson residuals", margin(a+2)) ///
xtitle("Cohort") ///
title("`c´, `oo´, `ss´ `dd´ diabetes", col(black) placement(west) size(medium))
graph save GPH/RCc_`c´_`o´_`d´_`s´_cohort, replace
}


use Rc, clear
bysort A : keep if _n == 1
forval i = 1/10 {
local c`i´=country[`i´]
local o`i´=OC[`i´]
local d`i´=DM[`i´]
local s`i´=sex[`i´]
}
graph combine ///
GPH/Rc_`c1´_`o1´_`d1´_`s1´_age.gph ///
GPH/Rc_`c2´_`o2´_`d2´_`s2´_age.gph ///
GPH/Rc_`c3´_`o3´_`d3´_`s3´_age.gph ///
GPH/Rc_`c4´_`o4´_`d4´_`s4´_age.gph ///
GPH/Rc_`c5´_`o5´_`d5´_`s5´_age.gph ///
GPH/Rc_`c6´_`o6´_`d6´_`s6´_age.gph ///
GPH/Rc_`c7´_`o7´_`d7´_`s7´_age.gph ///
GPH/Rc_`c8´_`o8´_`d8´_`s8´_age.gph ///
GPH/Rc_`c9´_`o9´_`d9´_`s9´_age.gph ///
GPH/Rc_`c10´_`o10´_`d10´_`s10´_age.gph ///
, graphregion(color(white)) cols(2) altshrink xsize(3)
graph combine ///
GPH/Rc_`c1´_`o1´_`d1´_`s1´_period.gph ///
GPH/Rc_`c2´_`o2´_`d2´_`s2´_period.gph ///
GPH/Rc_`c3´_`o3´_`d3´_`s3´_period.gph ///
GPH/Rc_`c4´_`o4´_`d4´_`s4´_period.gph ///
GPH/Rc_`c5´_`o5´_`d5´_`s5´_period.gph ///
GPH/Rc_`c6´_`o6´_`d6´_`s6´_period.gph ///
GPH/Rc_`c7´_`o7´_`d7´_`s7´_period.gph ///
GPH/Rc_`c8´_`o8´_`d8´_`s8´_period.gph ///
GPH/Rc_`c9´_`o9´_`d9´_`s9´_period.gph ///
GPH/Rc_`c10´_`o10´_`d10´_`s10´_period.gph ///
, graphregion(color(white)) cols(2) altshrink xsize(3)
graph combine ///
GPH/RCc_`c1´_`o1´_`d1´_`s1´_age.gph ///
GPH/RCc_`c2´_`o2´_`d2´_`s2´_age.gph ///
GPH/RCc_`c3´_`o3´_`d3´_`s3´_age.gph ///
GPH/RCc_`c4´_`o4´_`d4´_`s4´_age.gph ///
GPH/RCc_`c5´_`o5´_`d5´_`s5´_age.gph ///
GPH/RCc_`c6´_`o6´_`d6´_`s6´_age.gph ///
GPH/RCc_`c7´_`o7´_`d7´_`s7´_age.gph ///
GPH/RCc_`c8´_`o8´_`d8´_`s8´_age.gph ///
GPH/RCc_`c9´_`o9´_`d9´_`s9´_age.gph ///
GPH/RCc_`c10´_`o10´_`d10´_`s10´_age.gph ///
```

Figure 3.1: Modelled and crude mortality rates by age for 10 randomly selected country/cause of death/diabetes status/sex combinations.

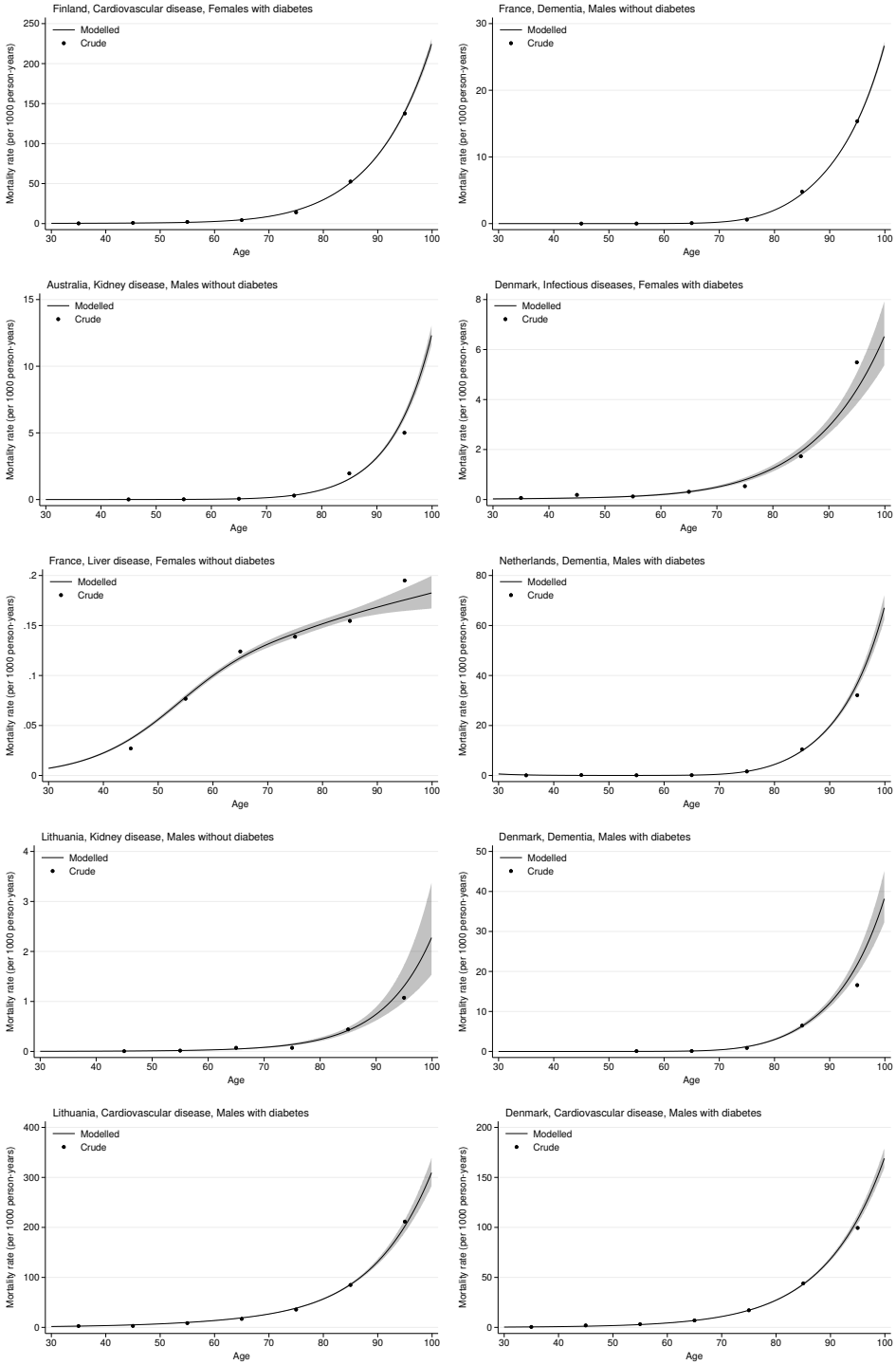


Figure 3.2: Modelled and crude mortality rates by year for 10 randomly selected country/cause of death/diabetes status/sex combinations.

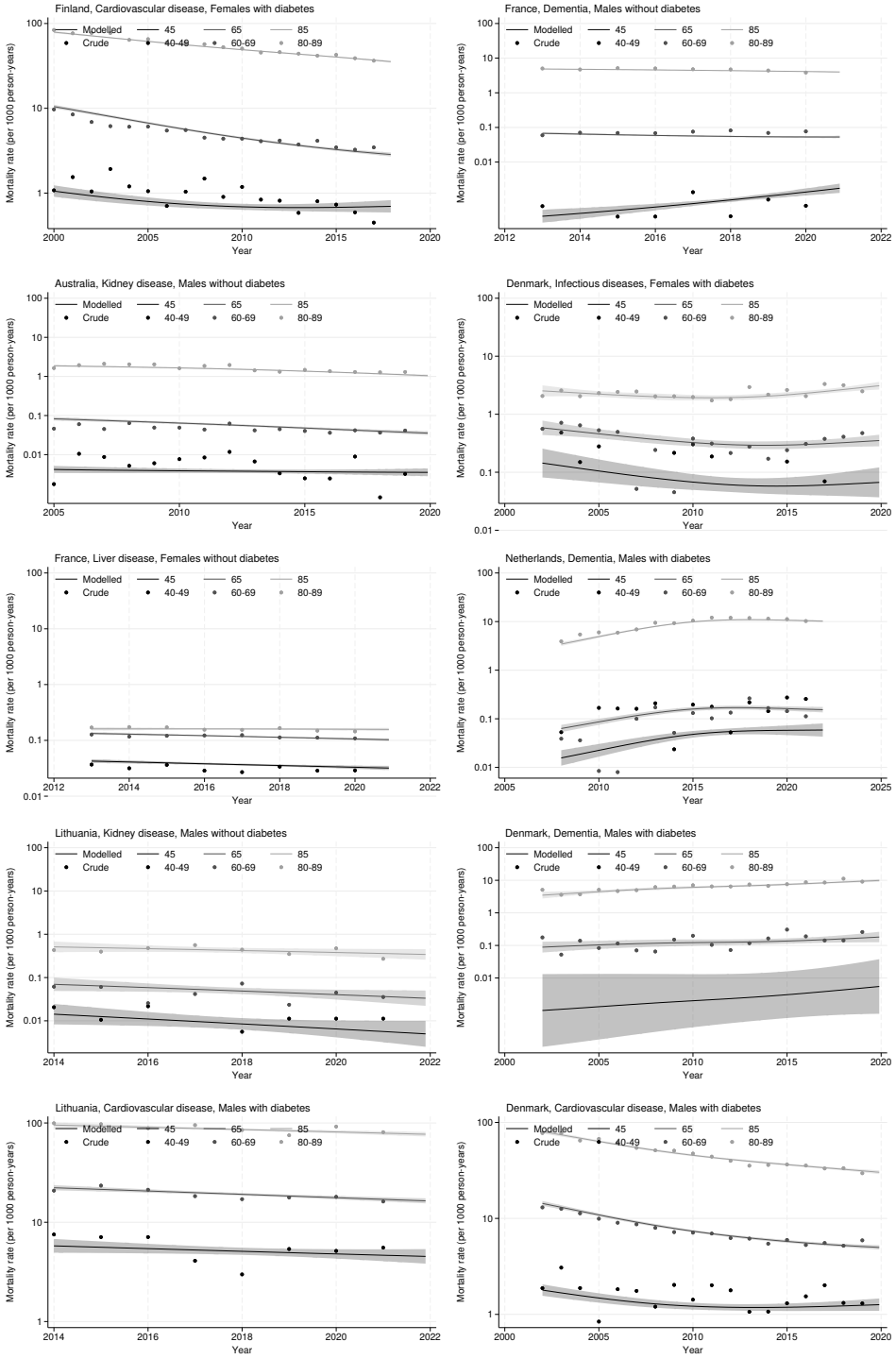


Figure 3.3: Pearson residuals by age for 10 randomly selected country/cause of death/diabetes status/sex combinations.

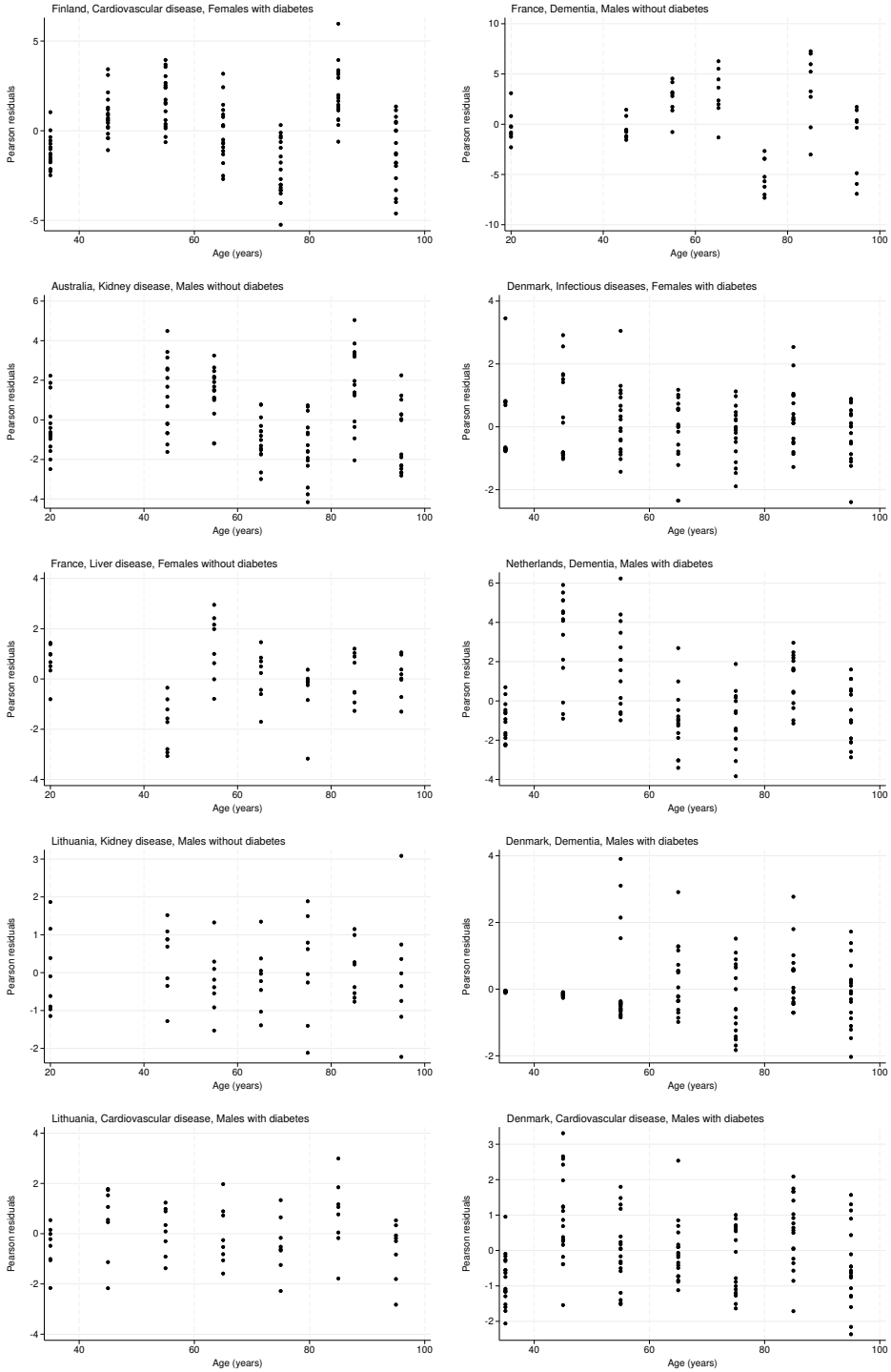
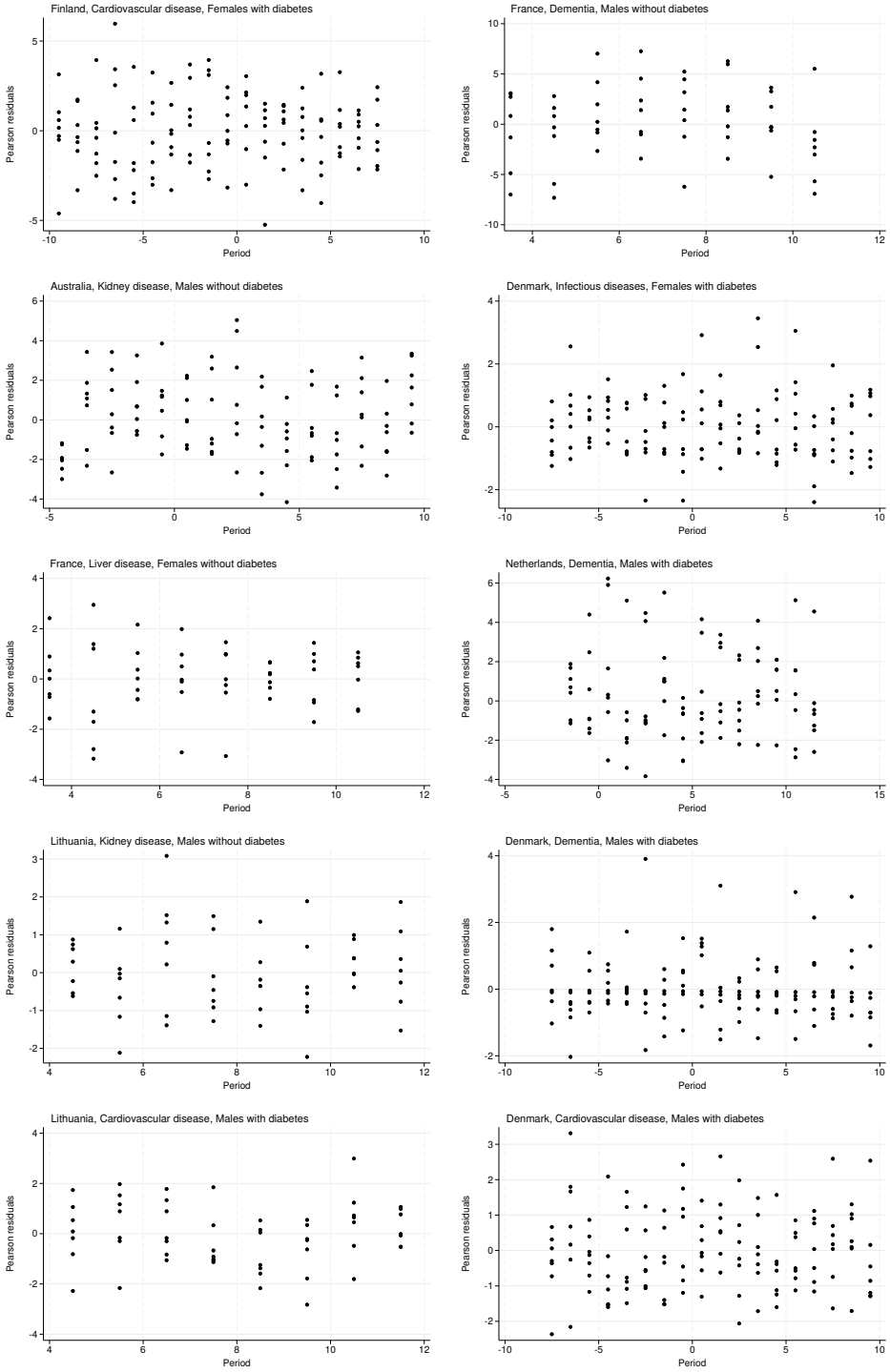


Figure 3.4: Pearson residuals by period for 10 randomly selected country/cause of death/diabetes status/sex combinations.



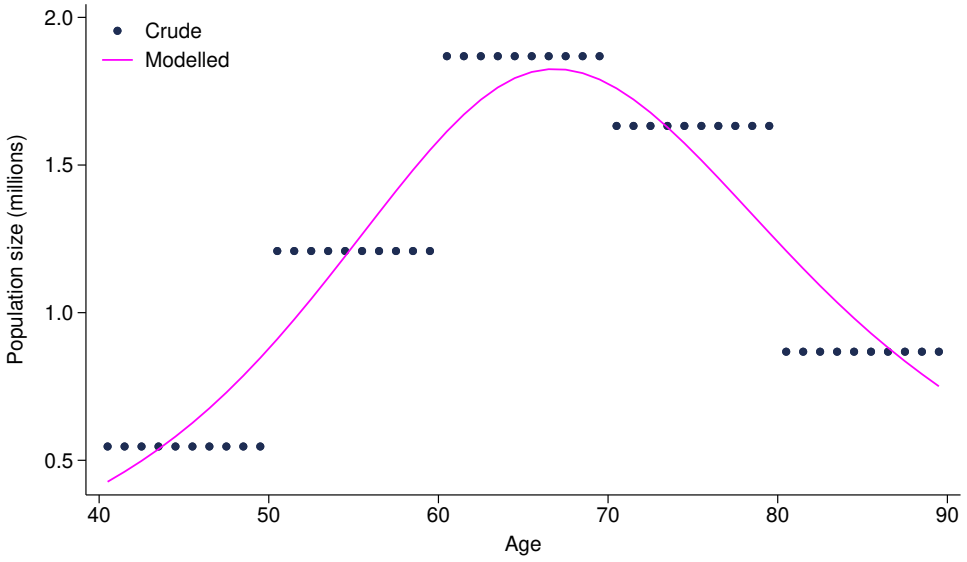
```
, graphregion(color(white)) cols(2) altshrink xsize(3)
graph combine ///
GPH/RCc_`c1´_`o1´_`d1´_`s1´_period.gph ///
GPH/RCc_`c2´_`o2´_`d2´_`s2´_period.gph ///
GPH/RCc_`c3´_`o3´_`d3´_`s3´_period.gph ///
GPH/RCc_`c4´_`o4´_`d4´_`s4´_period.gph ///
GPH/RCc_`c5´_`o5´_`d5´_`s5´_period.gph ///
GPH/RCc_`c6´_`o6´_`d6´_`s6´_period.gph ///
GPH/RCc_`c7´_`o7´_`d7´_`s7´_period.gph ///
GPH/RCc_`c8´_`o8´_`d8´_`s8´_period.gph ///
GPH/RCc_`c9´_`o9´_`d9´_`s9´_period.gph ///
GPH/RCc_`c10´_`o10´_`d10´_`s10´_period.gph ///
, graphregion(color(white)) cols(2) altshrink xsize(3)
```

We see that the models fit the data reasonably well (Figures 3.1- 3.4).

3.2 Age- and sex-standardised rates

We are going to calculate age-standardised mortality rates among people aged 40-89 years. We will first generate cause-specific mortality rates for people aged 40-89. Then, we will use direct standardisation to generate the age-standardised rates, using a reference population constructed by pooling the person-years among people with diabetes from all datasets. There will be two reference populations: first, one stratified by sex so that we can age and sex-standardise the overall results; second, one overall population to age-standardise the sex-stratified results to.

Figure 3.5: Pooled standard population




```
foreach i in Australia Canada Denmark Finland France Lithuania Netherlands Scotland SKorea {
use `i´, clear
collapse (sum) pys_dm, by(age_dm)
save `i´_pysdm, replace
}
clear
foreach i in Australia Canada Denmark Finland France Lithuania Netherlands Scotland SKorea {
append using `i´_pysdm
}
collapse (sum) pys_dm, by(age_dm)
drop if age_dm > 90 | age_dm == 35
expand 10
replace pys_dm=pys_dm/10
bysort age : replace age = age+_n-5.5
mkspline agesp = age, cubic knots(45 65 85)
glm pys_dm agesp*, family(gamma) link(log)
predict A
preserve
replace pys_dm = pys_dm/1000000
replace A = A/1000000
twoway ///
(scatter pys_dm age_dm, col(dknavy)) ///
(line A age_dm, col(magenta)) ///
, legend(symxsize(0.13cm) position(11) ring(0) region(lcolor(white) color(none)) ///
order(1 "Crude" ///
2 "Modelled") ///
```

Figure 3.6: Pooled standard population proportion

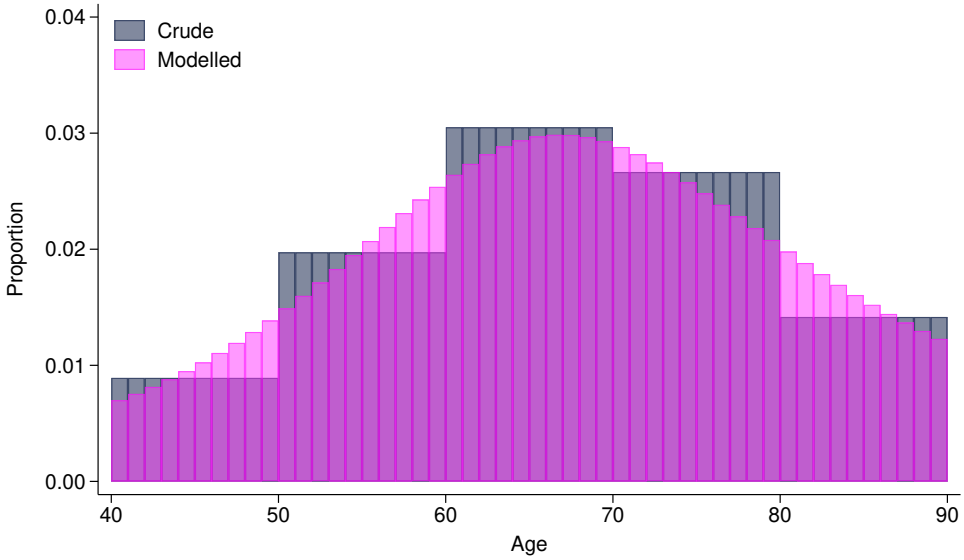
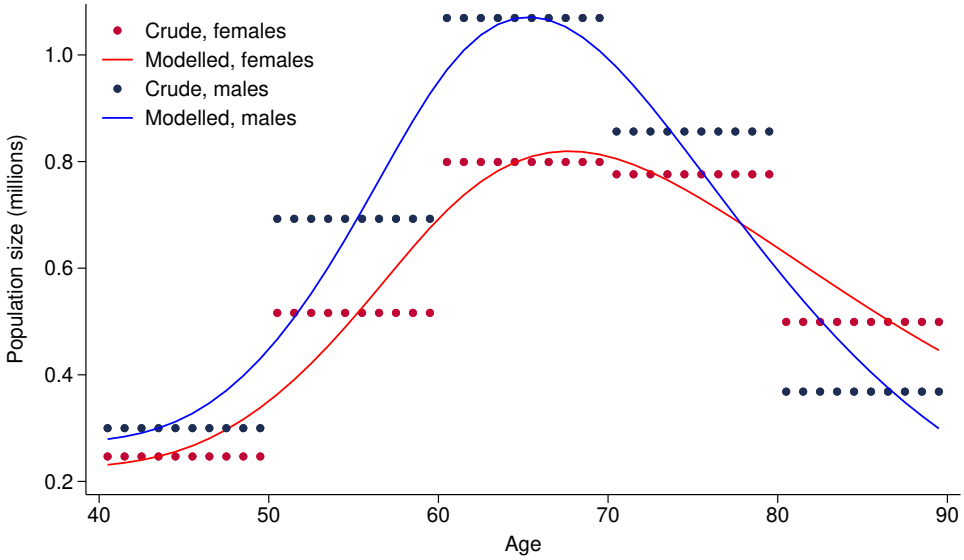


Figure 3.7: Pooled standard population by sex




```
cols(1)) ///
graphregion(color(white)) ///
ylabel(, format(%9.1f) angle(0) nogrid) ///
ytitle("Population size (millions)") xtitle("Age") xlabel(, nogrid)
restore
su(pys_dm)
gen age_dm_prop = pys_dm/r(sum)
su(A)
gen B = A/r(sum)
twoway ///
```

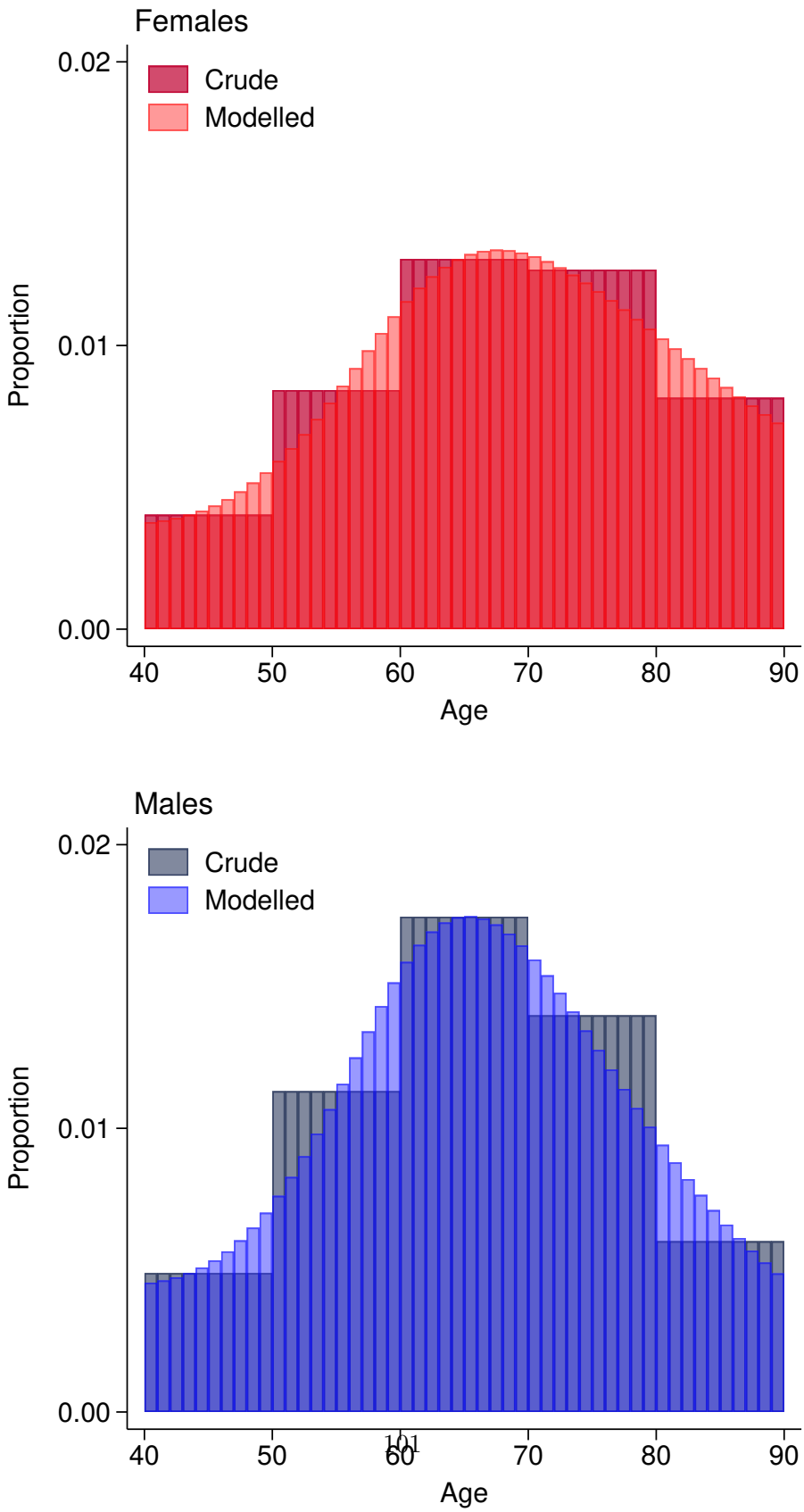
```
(bar age_dm_prop age_dm, color(dknavy%70)) ///
(bar B age_dm, color(magenta%50)) ///
, legend(symxsize(0.13cm) position(11) ring(0) region(lcolor(white) color(none)) ///
order(1 "Crude" ///
2 "Modelled") ///
cols(1)) ///
ylabel(0(0.01)0.04, angle(0) format(%9.2f) nogrid) ///
graphregion(color(white)) ///
ytitle("Proportion") xtitle("Age") xlabel(, nogrid)
keep age_dm B
replace age_dm = age-0.5
rename age_dm age
save refpop, replace
foreach i in Australia Canada Denmark Finland France Lithuania Netherlands Scotland SKorea {
use `i´, clear
collapse (sum) pys_dm, by(sex age_dm)
save `i´_pysdm_s, replace
}
clear
foreach i in Australia Canada Denmark Finland France Lithuania Netherlands Scotland SKorea {
append using `i´_pysdm_s
}
collapse (sum) pys_dm, by(sex age_dm)
drop if age_dm > 90 | age_dm == 35
expand 10
replace pys_dm=pys_dm/10
bysort sex age : replace age = age+_n-5.5
mkspline agesp = age, cubic knots(35 45 60 75 90)
glm pys_dm agesp* if sex == 0, family(gamma) link(log)
predict A0 if sex == 0
glm pys_dm agesp* if sex == 1, family(gamma) link(log)
predict A1 if sex == 1
preserve
replace pys_dm = pys_dm/1000000
replace A0 = A0/1000000
replace A1 = A1/1000000
twoway ///
(scatter pys_dm age_dm if sex == 0, col(cranberry)) ///
(line A0 age_dm, col(red)) ///
(scatter pys_dm age_dm if sex == 1, col(dknavy)) ///
(line A1 age_dm, col(blue)) ///
, legend(symxsize(0.13cm) position(11) ring(0) region(lcolor(white) color(none)) ///
order(1 "Crude, females" ///
2 "Modelled, females" ///
3 "Crude, males" ///
4 "Modelled, males") ///
cols(1)) ///
graphregion(color(white)) ///
ylabel(, format(%9.1f) angle(0) nogrid) xlabel(, nogrid) ///
ytitle("Population size (millions)") xtitle("Age")
restore
su(pys_dm)
gen age_dm_prop = pys_dm/r(sum)
gen A = A0
replace A = A1 if A ==.
su(A)
gen B = A/r(sum)
twoway ///
(bar age_dm_prop age_dm if sex == 0, color(cranberry%90)) ///
(bar B age_dm if sex == 0, color(red%50)) ///
, legend(symxsize(0.13cm) position(11) ring(0) region(lcolor(white) color(none)) ///
order(1 "Crude" ///
2 "Modelled") ///
cols(1)) ///
ylabel(0(0.01)0.02, angle(0) format(%9.2f) nogrid) xlabel(, nogrid) ///
graphregion(color(white)) ///
ytitle("Proportion") xtitle("Age") ///
title("Females", col(black) placement(west) size(medium))
```

```
graph save stdprop_0, replace
twoway ///
(bar age_dm_prop age_dm if sex == 1, color(dknavy%70)) ///
(bar B age_dm if sex == 1, color(blue%50)) ///
, legend(symxsize(0.13cm) position(11) ring(0) region(lcolor(white) color(none)) ///
order(1 "Crude" ///
2 "Modelled") ///
cols(1)) ///
ylabel(0(0.01)0.02, angle(0) format(%9.2f) nogrid) xlabel(, nogrid) ///
graphregion(color(white)) ///
ytitle("Proportion") xtitle("Age") ///
title("Males", col(black) placement(west) size(medium))
graph save stdprop_1, replace
graph combine ///
stdprop_0.gph stdprop_1.gph ///
, graphregion(color(white)) altshrink cols(1) xsize(2.5)
keep sex age_dm B
replace age_dm = age-0.5
rename age_dm age
save refpops, replace

quietly {
foreach i in Australia Canada Denmark Finland France Lithuania Netherlands Scotland SKorea {
foreach ii in can cvd res azd dmd inf flu ckd liv {
foreach iii in dm nondm {
if "`ii´" == "dmd" & "`iii´" == "nondm" {
}
else {
foreach iiii in 0 1 {
use `i´, clear
keep if sex == `iiii´
replace calendar = calendar-2009.5
gen coh = calendar-age_`iii´
centile(age_`iii´), centile(5 35 65 95)
local A1 = r(c_1)
local A2 = r(c_2)
local A3 = r(c_3)
local A4 = r(c_4)
mkspline agesp = age_`iii´, cubic knots(`A1´ `A2´ `A3´ `A4´)
su(calendar), detail
local rang = r(max)-r(min)
if `rang´ < 10 {
centile calendar, centile(25 75)
local CK1 = r(c_1)
local CK2 = r(c_2)
mkspline timesp = calendar, cubic knots(`CK1´ `CK2´)
}
else if inrange(`rang´,10,14.9) {
centile calendar, centile(10 50 90)
local CK1 = r(c_1)
local CK2 = r(c_2)
local CK3 = r(c_3)
mkspline timesp = calendar, cubic knots(`CK1´ `CK2´ `CK3´)
}
else {
centile calendar, centile(5 35 65 95)
local CK1 = r(c_1)
local CK2 = r(c_2)
local CK3 = r(c_3)
local CK4 = r(c_4)
mkspline timesp = calendar, cubic knots(`CK1´ `CK2´ `CK3´ `CK4´)
}
centile(coh), centile(5 35 65 95)
local CO1 = r(c_1)
local CO2 = r(c_2)
local CO3 = r(c_3)
local CO4 = r(c_4)
mkspline cohsp = coh, cubic knots(`CO1´ `CO2´ `CO3´ `CO4´)
```

```
poisson `ii´_d_`iii´ agesp* timesp* cohsp*, exposure(pys_`iii´)
keep sex calendar pys_`iii´ age_`iii´
if "`i´" == "Scotland" & "`iii´" == "nondm" {
keep if inrange(age_`iii´,40,89)
expand 10 if age_`iii´!=87.5
expand 20 if age_`iii´==87.5
replace pys = pys/10 if age_`iii´!=87.5
replace pys = pys/20 if age_`iii´==87.5
bysort cal age : replace age = age+_n-6 if age_`iii´!=87.5
bysort cal age : replace age = age+_n-8.5 if age_`iii´==87.5
drop if age_`iii´ >= 90
}
else {
keep if inrange(age_`iii´,40,89)
expand 10
replace pys = pys/10
bysort cal age : replace age = age+_n-6
}
gen coh = calendar-age
mkspline agesp = age, cubic knots(`A1´ `A2´ `A3´ `A4´)
if `rang´ < 9.99 {
mkspline timesp = calendar, cubic knots(`CK1´ `CK2´)
}
else if inrange(`rang´,10,14.99) {
mkspline timesp = calendar, cubic knots(`CK1´ `CK2´ `CK3´)
}
else {
mkspline timesp = calendar, cubic knots(`CK1´ `CK2´ `CK3´ `CK4´)
}
mkspline cohsp = coh, cubic knots(`CO1´ `CO2´ `CO3´ `CO4´)
predict _Rate, ir
save MD/STDi_`i´_`ii´_`iii´_`iiii´, replace
rename age_`iii´ age
merge m:1 age using refpop
drop _merge
gen double expdeath = _Rate*B
bysort cal : egen double expdeath1 = sum(expdeath)
gen stdrate = 1000*expdeath1
gen SEC1 = ((B^2)*(_Rate*(1-_Rate)))/pys_`iii´
bysort cal : egen double SEC2 = sum(SEC1)
gen double SE = sqrt(SEC2)
gen lb = 1000*(expdeath1-1.96*SE)
gen ub = 1000*(expdeath1+1.96*SE)
bysort cal (age) : keep if _n == 1
count if lb < 0
if r(N) != 0 {
noisily di "`i´" " " "`ii´" " " "`iii´" " " "`iiii´"
}
keep cal stdrate lb ub sex
gen country = "`i´"
gen OC = "`ii´"
gen DM = "`iii´"
replace cal = cal+2009.5
save MD/STD_`i´_`ii´_`iii´_`iiii´, replace
}
clear
append using MD/STDi_`i´_`ii´_`iii´_0 MD/STDi_`i´_`ii´_`iii´_1
rename age_`iii´ age
merge m:1 sex age using refpops
drop _merge
gen double expdeath = _Rate*B
bysort cal : egen double expdeath1 = sum(expdeath)
gen stdrate = 1000*expdeath1
gen SEC1 = ((B^2)*(_Rate*(1-_Rate)))/pys_`iii´
bysort cal : egen double SEC2 = sum(SEC1)
gen double SE = sqrt(SEC2)
gen lb = 1000*(expdeath1-1.96*SE)
gen ub = 1000*(expdeath1+1.96*SE)
```

```
bysort cal (age) : keep if _n == 1
count if lb < 0
if r(N) != 0 {
noisily di "`i´" " " "`ii´" " " "`iii´"
replace lb = 0.001 if lb < 0
}
keep cal stdrate lb ub
gen country = "`i´"
gen OC = "`ii´"
gen DM = "`iii´"
replace cal = cal+2009.5
save MD/STD_`i´_`ii´_`iii´, replace
}
}
}
}
}

. quietly {
Lithuania ckd dm 1
SKorea azd dm 1
```

Figure 3.8: Pooled standard population proportion by sex



The standardisation confidence interval crosses 0 for CKD deaths among males from Lithuania, reflecting extremely low numbers. I simply won't plot these. Similarly for Dementia deaths among males from South Korea.

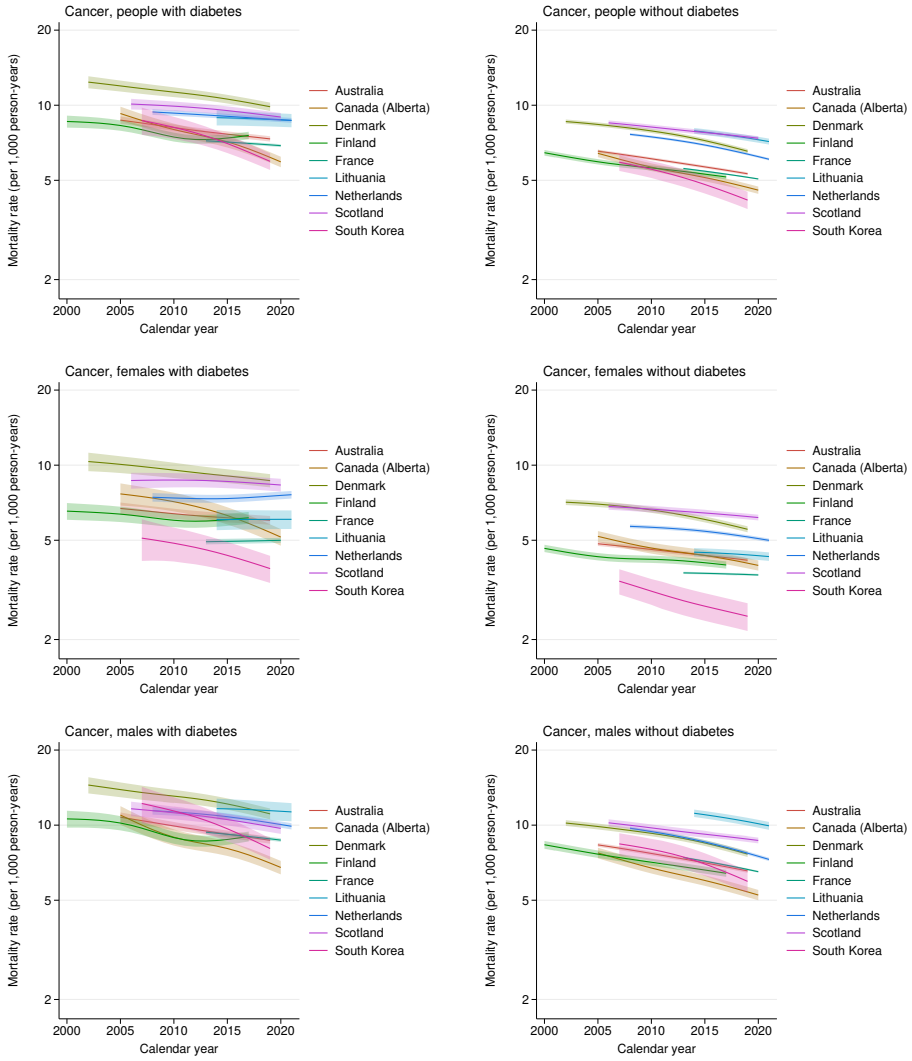
```
clear
set obs 1
gen country = "Lithuania"
save MD/STD_Lithuania_ckd_dm_1, replace
clear
set obs 1
gen country = "SKorea"
save MD/STD_SKorea_azd_dm_1, replace

*ssc install palettes
*ssc install colrspace
foreach ii in can cvd res azd dmd inf flu ckd liv {
foreach iii in dm nondm {
if "`ii´" == "dmd" & "`iii´" == "nondm" {
}
else {
if "`ii´" == "can" {
local oo = "Cancer"
local ylab = "2 5 10 20"
local yform = "%9.0f"
local yrange = "1.8 20"
}
if "`ii´" == "cvd" {
local oo = "Cardiovascular disease"
local ylab = "2 5 10 20 50"
local yform = "%9.0f"
local yrange = "1.8 50"
}
if "`ii´" == "res" {
local oo = "Chronic lower respiratory disease"
local ylab = "0.1 0.2 0.5 1 2"
local yform = "%9.1f"
local yrange = "0.05 4"
}
if "`ii´" == "azd" {
local oo = "Dementia"
local ylab = "0.01 0.02 0.05 0.1 0.2 0.5 1 2 5"
local yform = "%9.2f"
local yrange = "0.005 5"
}
if "`ii´" == "dmd" {
local oo = "Diabetes"
local ylab = "1 2 5 10"
local yform = "%9.0f"
local yrange = "0.5 10"
}
if "`ii´" == "inf" {
local oo = "Infectious diseases"
local ylab = "0.1 0.2 0.5 1 2"
local yform = "%9.1f"
local yrange = "0.05 2"
}
if "`ii´" == "flu" {
local oo = "Influenza and pneumonia"
local ylab = "0.1 0.2 0.5 1 2"
local yform = "%9.1f"
local yrange = "0.05 3"
}
if "`ii´" == "ckd" {
local oo = "Kidney disease"
local ylab = "0.01 0.02 0.05 0.1 0.2 0.5 1 2"
local yform = "%9.2f"
local yrange = "0.005 2"
```

```
}
if "`ii´" == "liv" {
local oo = "Liver disease"
local ylab = "0.02 0.05 0.1 0.2 0.5 1 2"
local yform = "%9.2f"
local yrange = "0.02 2.1"
}
if "`iii´" == "dm" {
local w = "with"
}
if "`iii´" == "nondm" {
local w = "without"
}
clear
foreach i in Australia Canada Denmark Finland France Lithuania Netherlands Scotland SKorea {
append using MD/STD_`i´_`ii´_`iii´
}
replace country = "Canada (Alberta)" if country == "Canada"
replace country = "South Korea" if country == "SKorea"
preserve
bysort country : keep if _n == 1
forval i = 1/9 {
local C`i´ = country[`i´]
}
restore
colorpalette hue, n(9) luminance(50) nograph
twoway ///
(rarea ub lb calendar if country == "`C1´", color("`r(p1)´%30") fintensity(inten80) lwidth(none)) //
> /
(line stdrate calendar if country == "`C1´", color("`r(p1)´") lpattern(solid)) ///
(rarea ub lb calendar if country == "`C2´", color("`r(p2)´%30") fintensity(inten80) lwidth(none)) //
> /
(line stdrate calendar if country == "`C2´", color("`r(p2)´") lpattern(solid)) ///
(rarea ub lb calendar if country == "`C3´", color("`r(p3)´%30") fintensity(inten80) lwidth(none)) //
> /
(line stdrate calendar if country == "`C3´", color("`r(p3)´") lpattern(solid)) ///
(rarea ub lb calendar if country == "`C4´", color("`r(p4)´%30") fintensity(inten80) lwidth(none)) //
> /
(line stdrate calendar if country == "`C4´", color("`r(p4)´") lpattern(solid)) ///
(rarea ub lb calendar if country == "`C5´", color("`r(p5)´%30") fintensity(inten80) lwidth(none)) //
> /
(line stdrate calendar if country == "`C5´", color("`r(p5)´") lpattern(solid)) ///
(rarea ub lb calendar if country == "`C6´", color("`r(p6)´%30") fintensity(inten80) lwidth(none)) //
> /
(line stdrate calendar if country == "`C6´", color("`r(p6)´") lpattern(solid)) ///
(rarea ub lb calendar if country == "`C7´", color("`r(p7)´%30") fintensity(inten80) lwidth(none)) //
> /
(line stdrate calendar if country == "`C7´", color("`r(p7)´") lpattern(solid)) ///
(rarea ub lb calendar if country == "`C8´", color("`r(p8)´%30") fintensity(inten80) lwidth(none)) //
> /
(line stdrate calendar if country == "`C8´", color("`r(p8)´") lpattern(solid)) ///
(rarea ub lb calendar if country == "`C9´", color("`r(p9)´%30") fintensity(inten80) lwidth(none)) //
> /
(line stdrate calendar if country == "`C9´", color("`r(p9)´") lpattern(solid)) ///
, legend(symxsize(0.13cm) position(3) region(lcolor(white) color(none)) ///
order(2 "`C1´" ///
4 "`C2´" ///
6 "`C3´" ///
8 "`C4´" ///
10 "`C5´" ///
12 "`C6´" ///
14 "`C7´" ///
16 "`C8´" ///
18 "`C9´") ///
cols(1)) ///
graphregion(color(white)) ///
ylabel(`ylab´, format(`yform´) grid glpattern(solid) glcolor(gs10%20) angle(0)) ///
yscale(log range(`yrange´)) ///
```

```
xscale(range(2000 2020)) ///
xlabel(2000(5)2020, nogrid) ///
ytitle("Mortality rate (per 1,000 person-years)", margin(a+2)) ///
xtitle("Calendar year") ///
title("`oo´, people `w´ diabetes", placement(west) color(black) size(medium))
graph save GPH/STD_GPH_`ii´_`iii´, replace
forval iiii = 0/1 {
if `iiii´ == 0 {
local s = "females"
}
if `iiii´ == 1 {
local s = "males"
}
clear
foreach i in Australia Canada Denmark Finland France Lithuania Netherlands Scotland SKorea {
append using MD/STD_`i´_`ii´_`iii´_`iiii´
}
replace country = "Canada (Alberta)" if country == "Canada"
replace country = "South Korea" if country == "SKorea"
preserve
bysort country : keep if _n == 1
forval i = 1/9 {
local C`i´ = country[`i´]
}
restore
colorpalette hue, n(9) luminance(50) nograph
twoway ///
(rarea ub lb calendar if country == "`C1´", color("`r(p1)´%30") fintensity(inten80) lwidth(none)) //
> /
(line stdrate calendar if country == "`C1´", color("`r(p1)´") lpattern(solid)) ///
(rarea ub lb calendar if country == "`C2´", color("`r(p2)´%30") fintensity(inten80) lwidth(none)) //
> /
(line stdrate calendar if country == "`C2´", color("`r(p2)´") lpattern(solid)) ///
(rarea ub lb calendar if country == "`C3´", color("`r(p3)´%30") fintensity(inten80) lwidth(none)) //
> /
(line stdrate calendar if country == "`C3´", color("`r(p3)´") lpattern(solid)) ///
(rarea ub lb calendar if country == "`C4´", color("`r(p4)´%30") fintensity(inten80) lwidth(none)) //
> /
(line stdrate calendar if country == "`C4´", color("`r(p4)´") lpattern(solid)) ///
(rarea ub lb calendar if country == "`C5´", color("`r(p5)´%30") fintensity(inten80) lwidth(none)) //
> /
(line stdrate calendar if country == "`C5´", color("`r(p5)´") lpattern(solid)) ///
(rarea ub lb calendar if country == "`C6´", color("`r(p6)´%30") fintensity(inten80) lwidth(none)) //
> /
(line stdrate calendar if country == "`C6´", color("`r(p6)´") lpattern(solid)) ///
(rarea ub lb calendar if country == "`C7´", color("`r(p7)´%30") fintensity(inten80) lwidth(none)) //
> /
(line stdrate calendar if country == "`C7´", color("`r(p7)´") lpattern(solid)) ///
(rarea ub lb calendar if country == "`C8´", color("`r(p8)´%30") fintensity(inten80) lwidth(none)) //
> /
(line stdrate calendar if country == "`C8´", color("`r(p8)´") lpattern(solid)) ///
(rarea ub lb calendar if country == "`C9´", color("`r(p9)´%30") fintensity(inten80) lwidth(none)) //
> /
(line stdrate calendar if country == "`C9´", color("`r(p9)´") lpattern(solid)) ///
, legend(symxsize(0.13cm) position(3) region(lcolor(white) color(none)) ///
order(2 "`C1´" ///
4 "`C2´" ///
6 "`C3´" ///
8 "`C4´" ///
10 "`C5´" ///
12 "`C6´" ///
14 "`C7´" ///
16 "`C8´" ///
18 "`C9´") ///
cols(1)) ///
graphregion(color(white)) ///
ylabel(`ylab´, format(`yform´) grid  glpattern(solid) glcolor(gs10%20) angle(0)) ///
yscale(log range(`yrange´)) ///
```

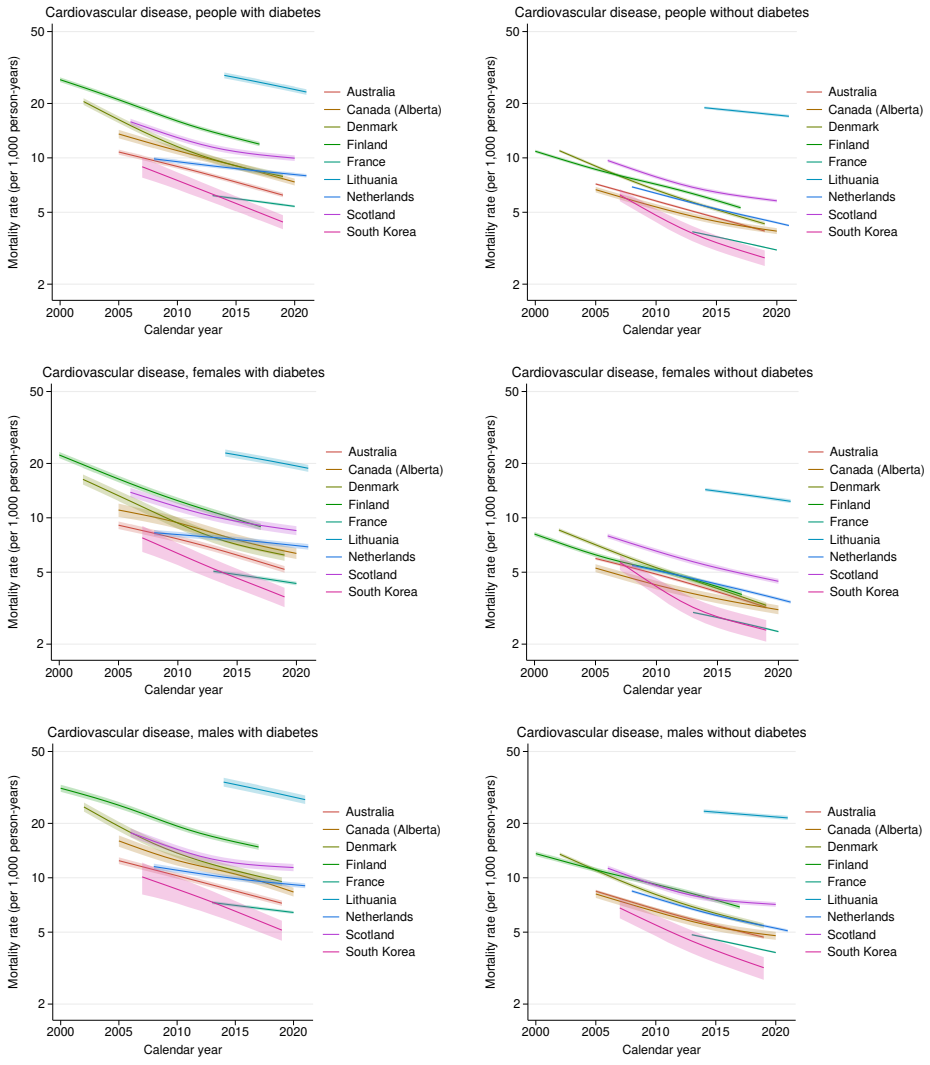
```
xscale(range(2000 2020)) ///
xlabel(2000(5)2020, nogrid) ///
ytitle("Mortality rate (per 1,000 person-years)", margin(a+2)) ///
xtitle("Calendar year") ///
title("`oo´, `s´ `w´ diabetes", placement(west) color(black) size(medium))
graph save GPH/STD_GPH_`ii´_`iii´_`iiii´, replace
}
}
}
}
```

Figure 3.9: Age-standardised mortality rate by cause of death, people aged 40-89. Cancer.



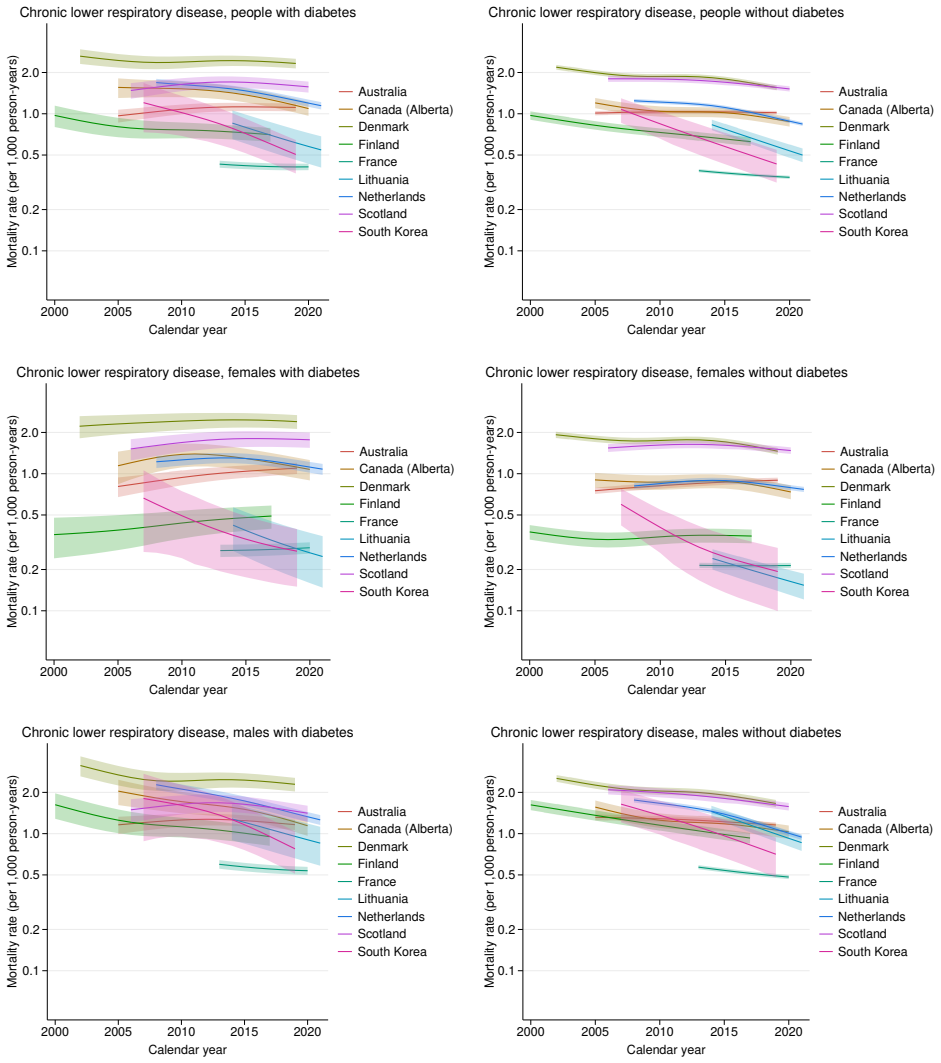

```
foreach ii in can cvd res azd dmd inf flu ckd liv {
if "`ii´" == "can" {
local oo = "Cancer"
}
if "`ii´" == "cvd" {
```

Figure 3.10: Age-standardised mortality rate by cause of death, people aged 40-89. Cardiovascular disease.



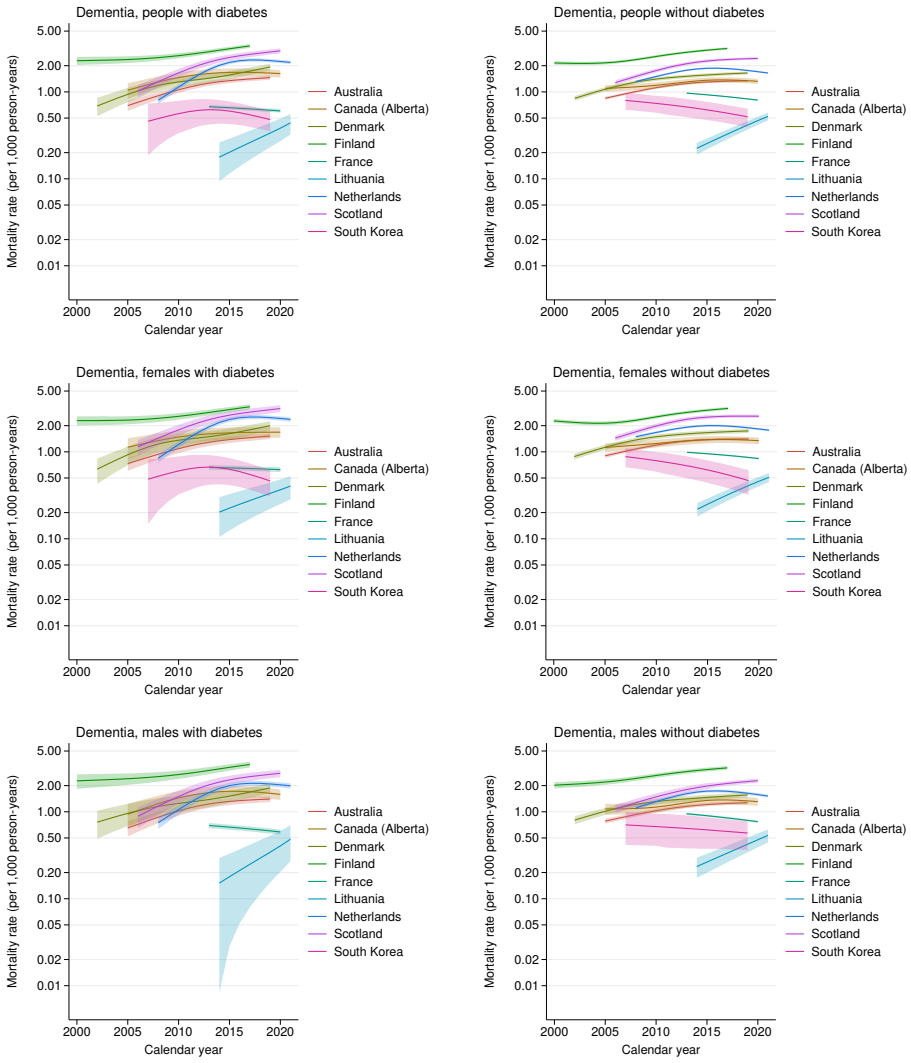

```
local oo = "Cardiovascular disease"
}
if "`ii´" == "res" {
local oo = "Chronic lower respiratory disease"
}
if "`ii´" == "azd" {
local oo = "Dementia"
}
if "`ii´" == "dmd" {
local oo = "Diabetes"
}
if "`ii´" == "inf" {
local oo = "Infectious diseases"
}
if "`ii´" == "flu" {
```

Figure 3.11: Age-standardised mortality rate by cause of death, people aged 40-89. Chronic lower respiratory disease.




```
local oo = "Influenza and pneumonia"
}
if "`ii´" == "ckd" {
local oo = "Kidney disease"
}
if "`ii´" == "liv" {
local oo = "Liver disease"
}
if "`ii´" == "dmd" {
graph combine ///
GPH/STD_GPH_`ii´_dm.gph ///
GPH/STD_GPH_`ii´_dm_0.gph ///
GPH/STD_GPH_`ii´_dm_1.gph ///
, graphregion(color(white)) cols(1) altshrink xsize(2)
}
```

Figure 3.12: Age-standardised mortality rate by cause of death, people aged 40-89. Dementia.




```
else {
graph combine ///
GPH/STD_GPH_`ii´_dm.gph ///
GPH/STD_GPH_`ii´_nondm.gph ///
GPH/STD_GPH_`ii´_dm_0.gph ///
GPH/STD_GPH_`ii´_nondm_0.gph ///
GPH/STD_GPH_`ii´_dm_1.gph ///
GPH/STD_GPH_`ii´_nondm_1.gph ///
, graphregion(color(white)) cols(2) altshrink xsize(4)
}
}
```

Figure 3.13: Age-standardised mortality rate by cause of death, people aged 40-89. Diabetes.

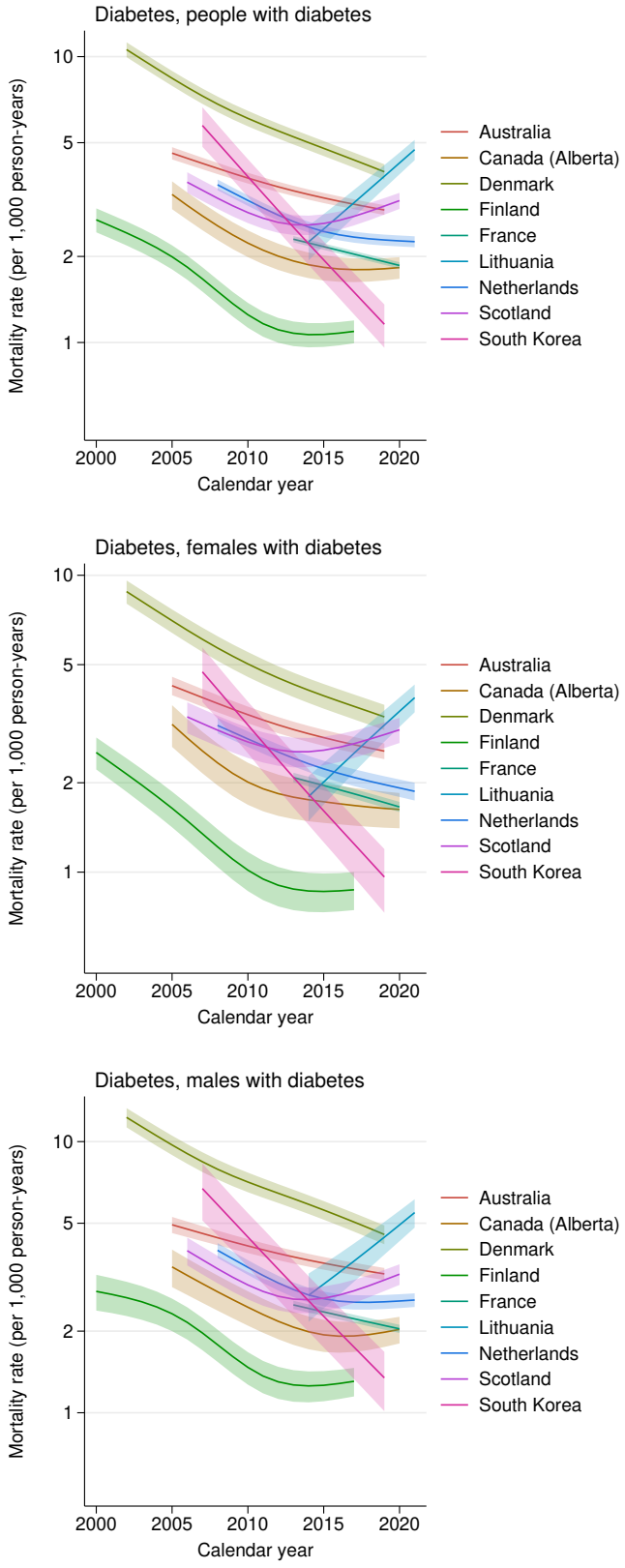


Figure 3.14: Age-standardised mortality rate by cause of death, people aged 40-89. Infectious diseases.

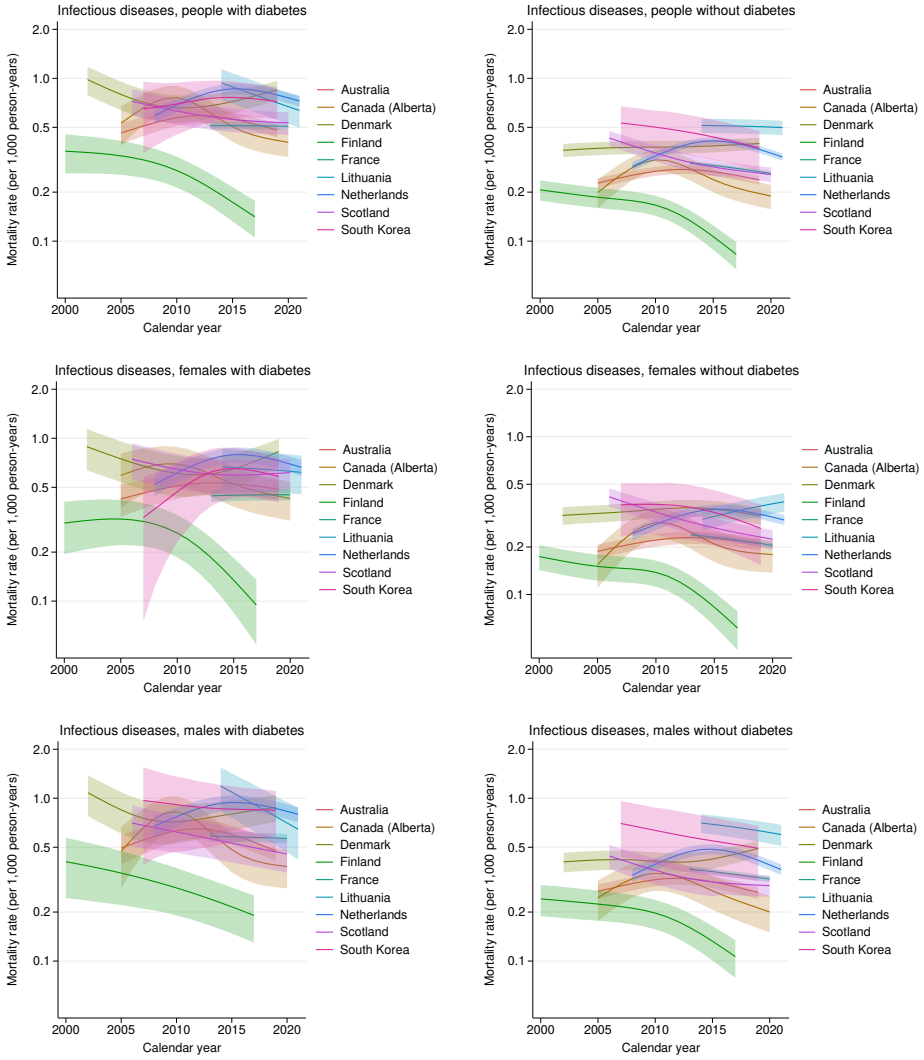


Figure 3.15: Age-standardised mortality rate by cause of death, people aged 40-89. Influenza and pneumonia.

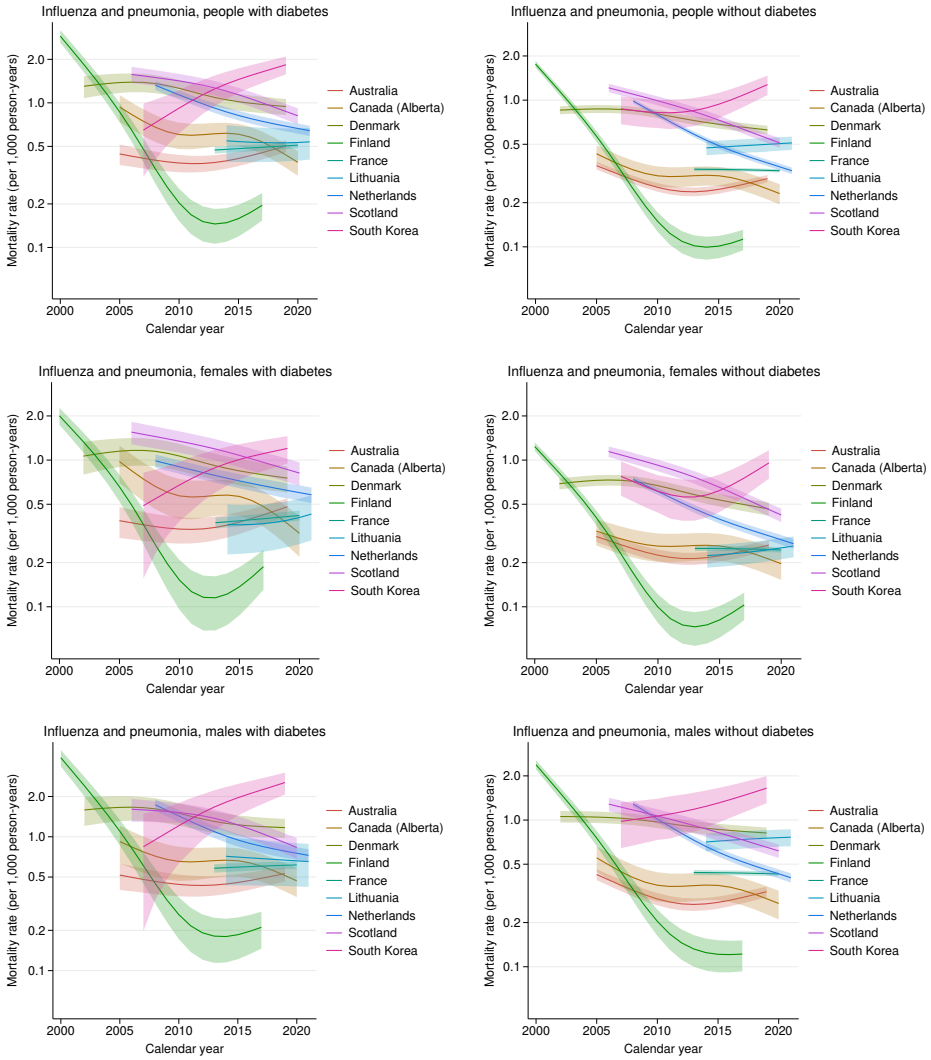


Figure 3.16: Age-standardised mortality rate by cause of death, people aged 40-89. Kidney disease.

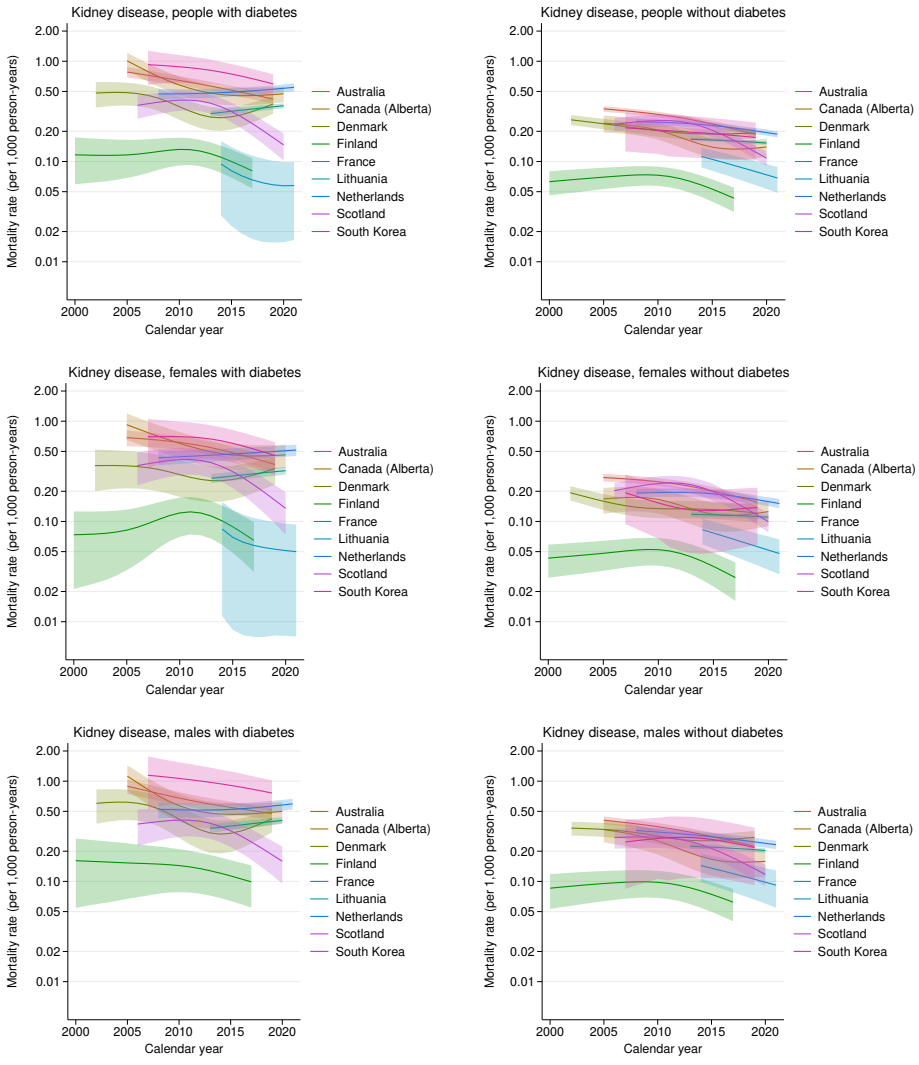
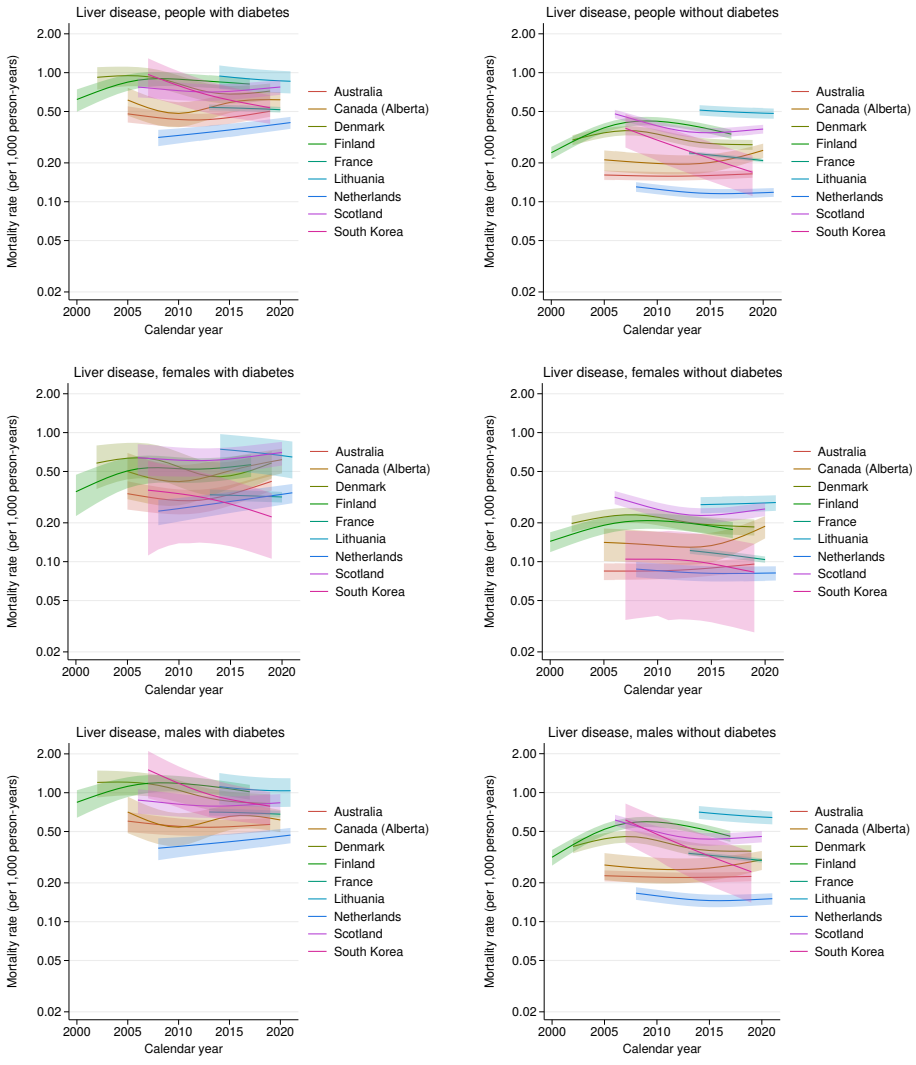


Figure 3.17: Age-standardised mortality rate by cause of death, people aged 40-89. Liver disease.



References