Reference Adjusted Survival Measures as an Alternative to Net Survival in Population-Based Cancer Studies

Rebecka V Johansson[1], Therese M-L Andersson[1], Paul C Lambert[1,2]

[1] Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
[2] Cancer Registry of Norway, Norwegian Institute of Public Health, Oslo, Norway

Outline

- Master’s thesis
  - Background
  - Research questions
  - Simulation study
    - Results
- Generating relative survival data
  - popmortsim
  - sessions

Background

What is the goal?

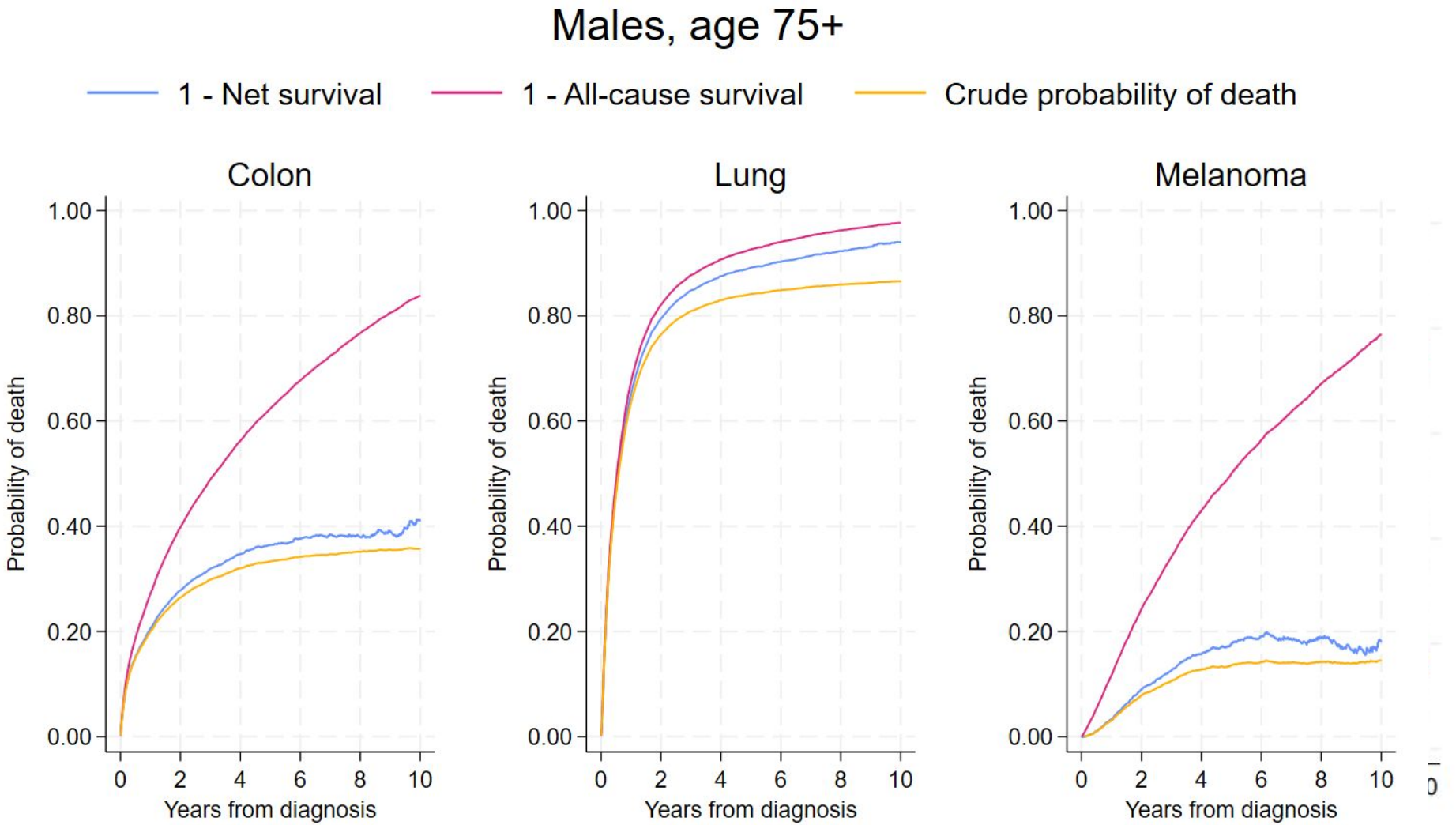
Be able to compare populations in terms of cancer mortality, e.g. males/females, countries

Most common measure:

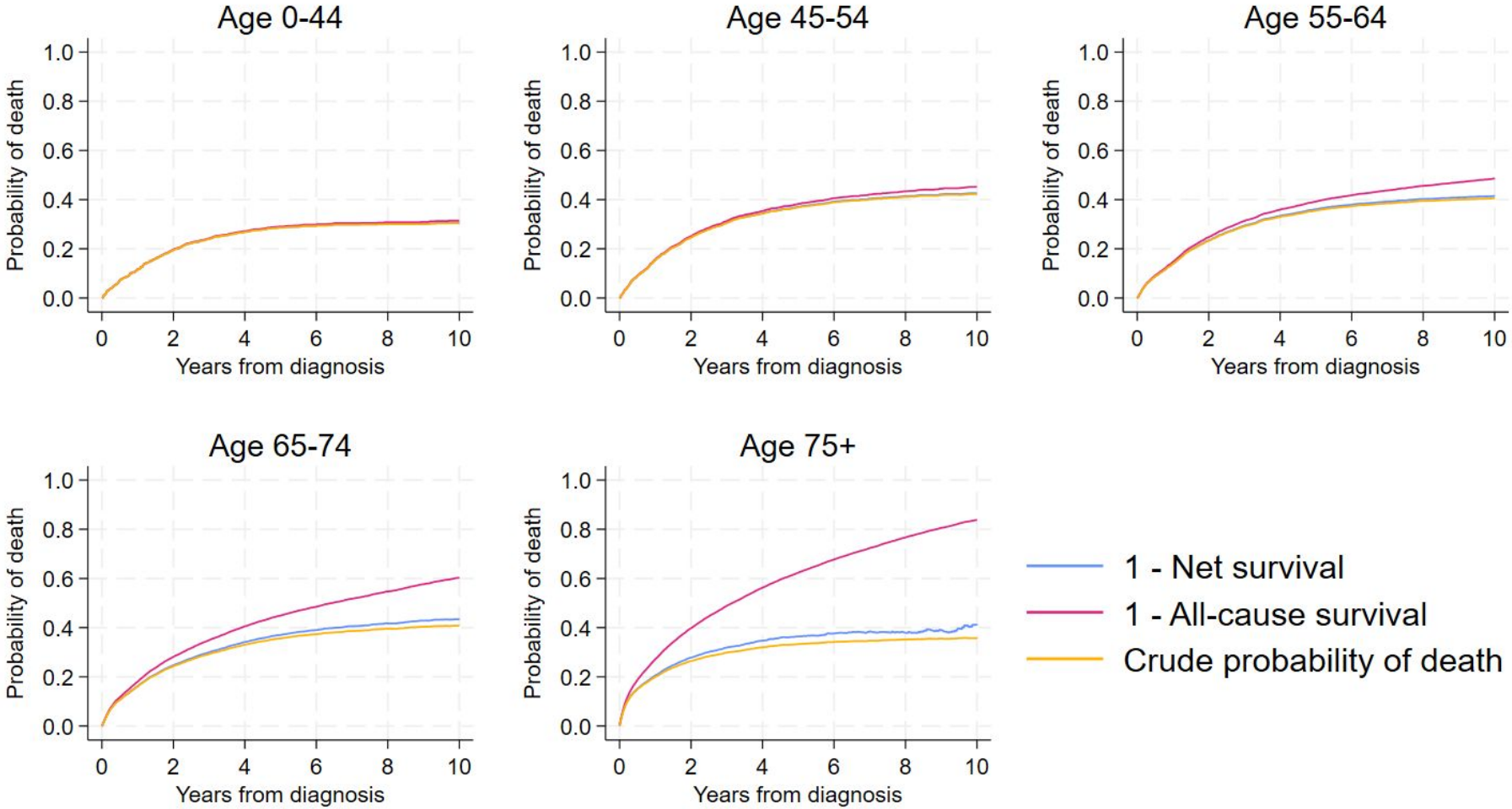
Net survival - Probability of surviving the cancer under study if it is impossible to die from anything else

“Real world” measures

- All cause survival - Probability of surviving any cause
- Crude probability of death due to cancer - Probability of dying of the cancer under study in the presence of competing risks


Males, age 75+
1 - Net survival
1 - All-cause survival
Crude probability of death
Colon
Lung
Melanoma
Probability of death
1.00
0.80
0.60
0.40
0.20
0.00
0
2
4
6
8
10
Years from diagnosis

Colon, males


Age 0-44
Age 45-54
Age 55-64
Age 65-74
Age 75+
Probability of death
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10
Years from diagnosis
1 - Net survival
1 - All-cause survival
Crude probability of death

Estimating net survival (relative survival framework)

- Mortality rate for individual i: $h_i(t \mid \mathbf{x}_i) = h_i^*(t \mid \mathbf{x}_i) + \lambda_i(t \mid \mathbf{x}_i)$
- Survival for individual i: $S_i(t \mid \mathbf{x}_i) = S_i^*(t \mid \mathbf{x}_i) R_i(t \mid \mathbf{x}_i)$
- Relative survival for individual i:

$$R_i(t \mid \mathbf{x}_i) = \frac{S_i(t \mid \mathbf{x}_i)}{S_i^*(t \mid \mathbf{x}_i)} = \exp\Big(- \int_0^t \lambda_i(u \mid \mathbf{x}_i) du \Big)$$

The Pohar Perme estimator [1]

- Estimates marginal (average) net survival
- People still alive are upweighted $w_i^*(t) = \frac{1}{S_i^*(t)}$
- Cumulative excess hazard: $\hat{H}_{PP}(t) = \int_0^t \frac{\sum_{i=1}^n w_i^*(u) Y_i(u) \{dN_i(u) - dH_i^*(u)\}}{\sum_{i=1}^n w_i^*(u) Y_i(u)}$

In Stata - stpp

```
stpp Rpp using ${POPMORT}, /// population mortality file
    datebirth(birthdate) /// date at birth
    datediag(diagdate) /// date at diagnosis
    pmother(sex) /// other variables in population mortality file
    by(sex) /// seperate estimates by sex
    list(1 5 10) /// display 1,5 and 10 year survival
    allcause(AC) /// calculate all-cause survival
    crudeprob(CP_c CP_o) /// calculate crude probabilities of death
    deathprob /// calculate death probabilities instead
    indweights(wts) //  individual weights for age-standardization
```

Population mortality (popmort) file

	_year	_age	sex	rate	prob
1	1950	0	1	.023527	.9767476
2	1950	1	1	.002013	.997989
3	1950	2	1	.001607	.9983943
4	1950	3	1	.001324	.9986769
5	1950	4	1	.000973	.9990275
6	1950	5	1	.000955	.9990454
7	1950	6	1	.000808	.9991923
8	1950	7	1	.000652	.9993482
9	1950	8	1	.000727	.9992732
10	1950	9	1	.000528	.9994721

Reference adjustment [2]

Assign reference (common) expected rates for other-cause mortality → fair comparisons

$$S_i^R(t \mid \mathbf{x}_i) = S_i^{**}(t \mid \mathbf{x}_i) R_i(t \mid \mathbf{x}_i) = \frac{S_i^{**}(t \mid \mathbf{x}_i) S_i(t \mid \mathbf{x}_i)}{S_i^{*}(t \mid \mathbf{x}_i)}$$

Weights: $w_i(t) = \left\{ \frac{S_i^{**}(t)}{S_i^{*}(t)} \right\}$

Cumulative hazard: $\hat{H}_R(t) = \int_0^t \frac{\sum_{i=1}^n w_i(u) Y_i(u) \{ dN_i(u) - dH_i^{*}(u) + dH_i^{**}(u) \}}{\sum_{i=1}^n w_i(u) Y_i(u)}$

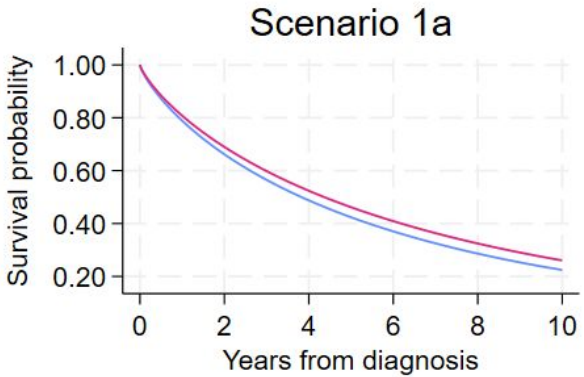
In Stata - stpp

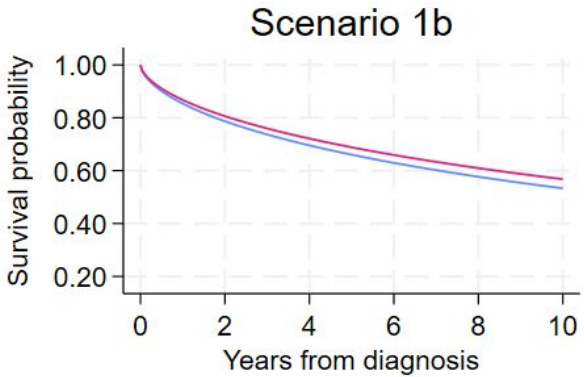
```
stpp Rpp using ${POPMORT}, /// population mortality file
    datebirth(birthdate) /// date at birth
    datediag(diagdate) /// date at diagnosis
    pmother(sex) /// other variables in population mortality file
    by(sex) /// seperate estimates by sex
    list(1 5 10) /// display 1,5 and 10 year survival
    allcause(AC) /// calculate all-cause survival
    crudeprob(CP_c CP_o) /// calculate crude probabilities of death
    deathprob /// calculate death probabilities instead
    indweights(wts) /// individual weights for age-standardization
    using2(popmort_refadj.dta) // population mortality file for the reference rates
```

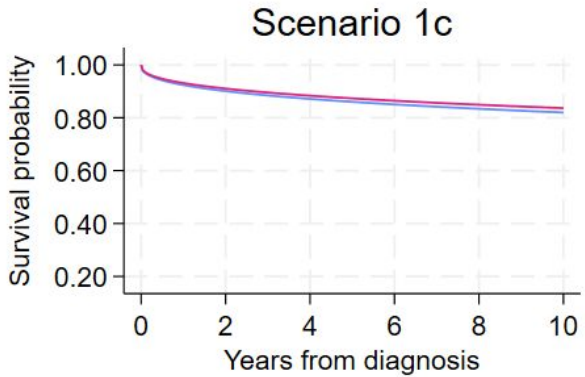
Research questions

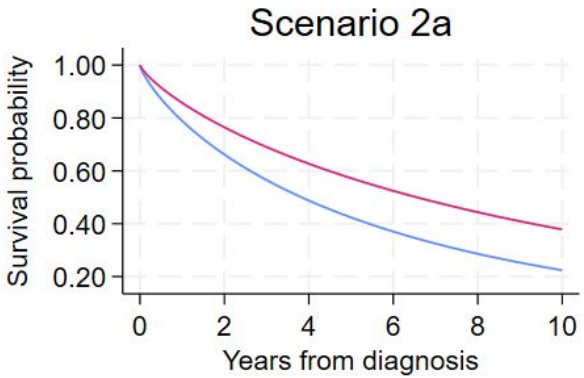
- Does reference adjustment have **higher power** than net survival to detect differences between groups, particularly for long-term survival?
- How **robust** are net survival and reference adjustment to errors in the population mortality file?

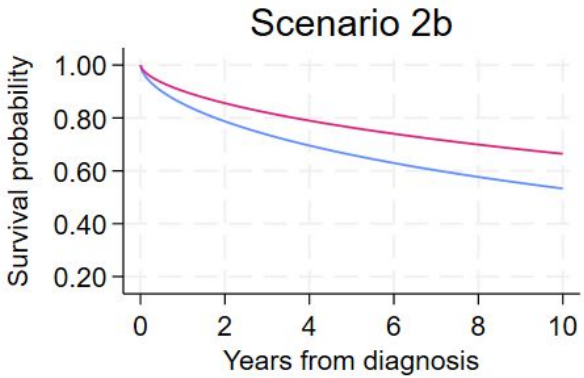
Simulation Study

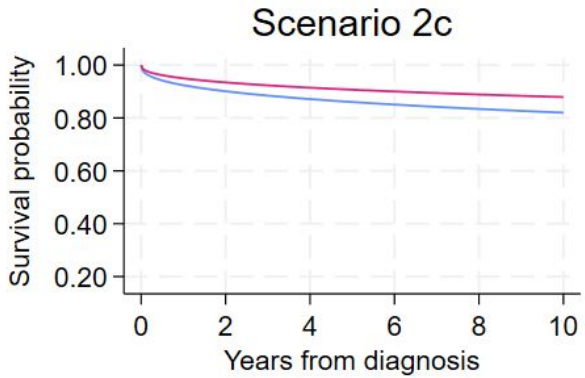

Scenario 1a
Survival probability
1.00
0.80
0.60
0.40
0.20
0
2
4
6
8
10
Years from diagnosis


Scenario 1b
Survival probability
1.00
0.80
0.60
0.40
0.20
0
2
4
6
8
10
Years from diagnosis


Scenario 1c
Survival probability
1.00
0.80
0.60
0.40
0.20
0
2
4
6
8
10
Years from diagnosis

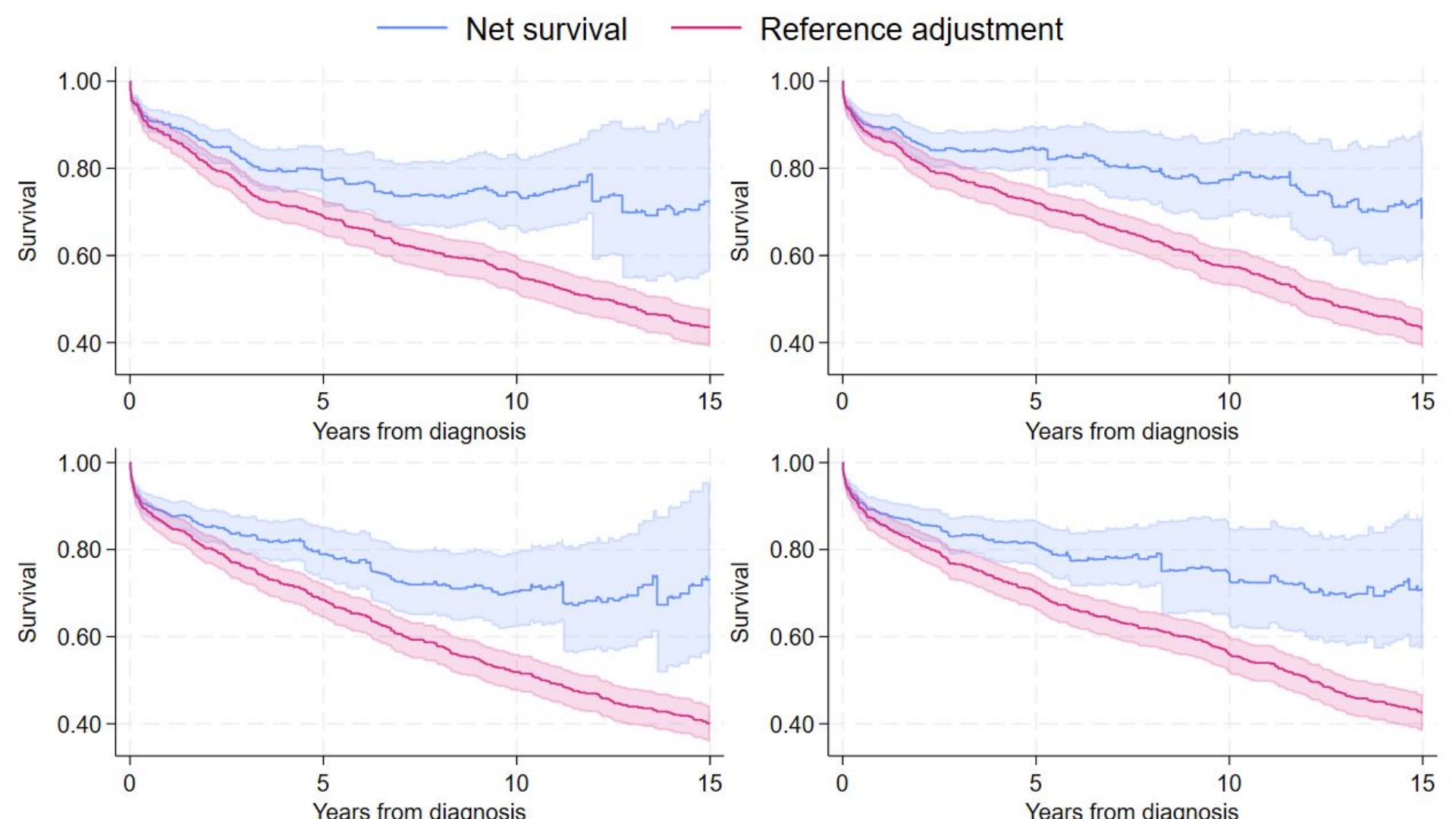

Scenario 2a
Survival probability
1.00
0.80
0.60
0.40
0.20
0
2
4
6
8
10
Years from diagnosis


Scenario 2b
Survival probability
1.00
0.80
0.60
0.40
0.20
0
2
4
6
8
10
Years from diagnosis


Scenario 2c
Survival probability
1.00
0.80
0.60
0.40
0.20
0
2
4
6
8
10
Years from diagnosis

Research questions

- **Does reference adjustment have higher power than net survival to detect differences between groups, particularly for long-term survival?**
- How robust are net survival and reference adjustment to errors in the population mortality file?

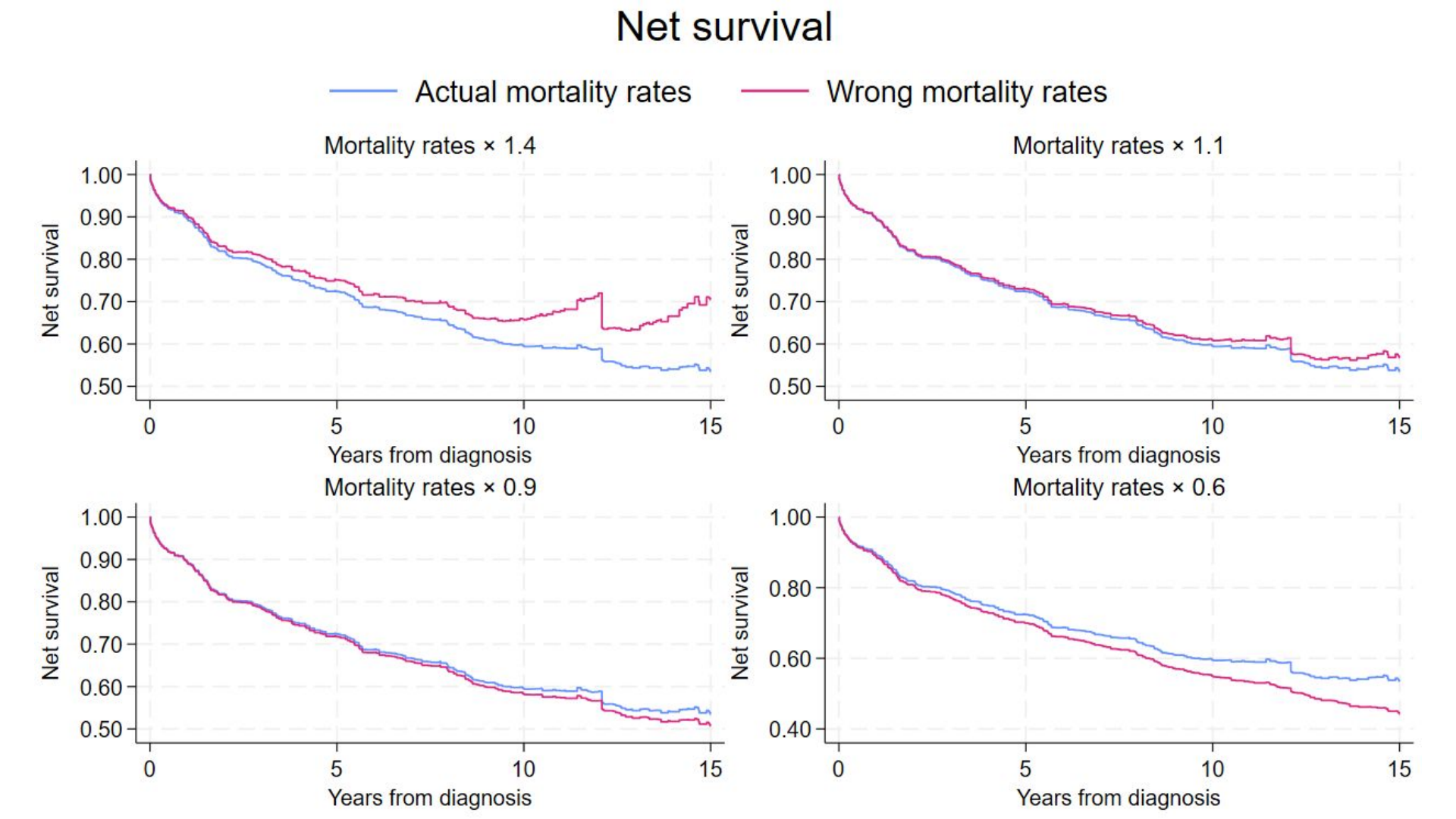

Net survival
Reference adjustment
Survival
1.00
0.80
0.60
0.40
0
5
10
15
Years from diagnosis

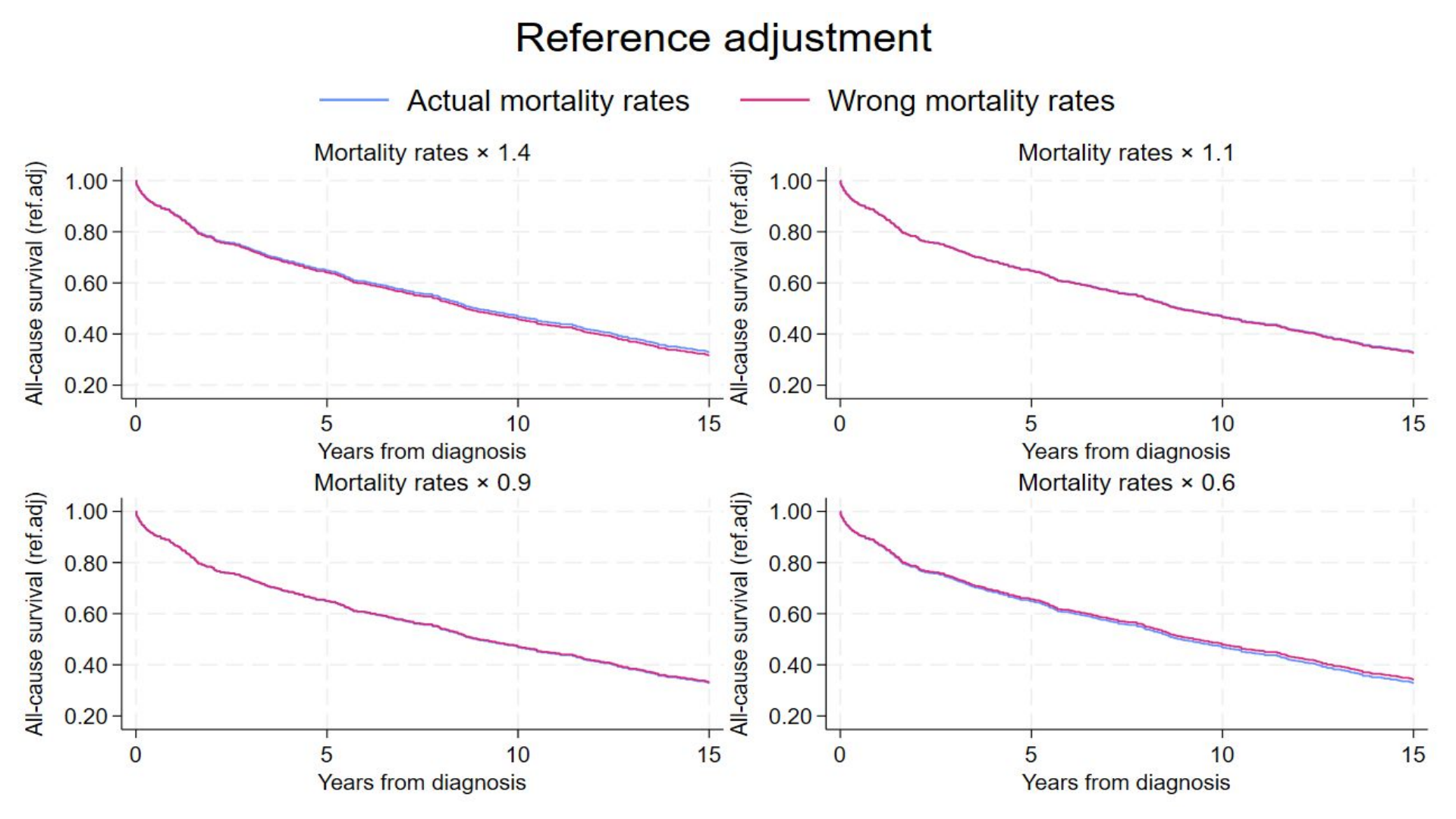
Result: Power advantage

		Net			RA		
Scenario	Time	Bias	Power	Cov.	Bias	Power	Cov.
2a	1	0.0009	0.8463	0.9495	0.0009	0.8432	0.9474
	5	0.0019	0.9921	0.9611	0.0008	0.9942	0.9563
	10	0.0022	0.9726	0.9674	0.0003	0.9932	0.9553
	15	0.0007	0.8663	0.9626	-0.0003	0.9726	0.9542
2b	1	0.0005	0.6758	0.9511	0.0005	0.6737	0.9484
	5	0.0024	0.8789	0.9595	0.0010	0.9095	0.9547
	10	0.0022	0.7063	0.9668	-0.0006	0.9211	0.9653
	15	0.0019	0.4389	0.9616	0.0001	0.8811	0.9689
2c	1	0.0003	0.3768	0.9495	0.0006	0.3821	0.9495
	5	0.0012	0.3532	0.9537	0.0001	0.4153	0.9658
	10	0.0014	0.1579	0.9658	-0.0015	0.3237	0.9758
	15	-0.0028	0.0737	0.9547	-0.0008	0.2468	0.9742

Research questions

- Does reference adjustment have higher power than net survival to detect differences between groups, particularly for long-term survival?
- **How robust are net survival and reference adjustment to errors in the population mortality file?**


Net survival
Actual mortality rates
Wrong mortality rates
Mortality rates × 1.4
Mortality rates × 1.1
Mortality rates × 0.9
Mortality rates × 0.6
Net survival
Years from diagnosis
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0
5
10
15


Reference adjustment
Actual mortality rates
Wrong mortality rates
Mortality rates × 1.4
Mortality rates × 1.1
Mortality rates × 0.9
Mortality rates × 0.6
All-cause survival (ref.adj)
1.00
0.80
0.60
0.40
0.20
0
5
10
15
Years from diagnosis

Result: Robustness to wrong mortality rates

Scenario	Measure	Sex	Mortality rates ×1.4	×1.1	×0.9	×0.6
2c				Time: 10		
	Net	Male	0.2044 (0.672)	0.0329 (0.877)	-0.0370 (0.884)	-0.1104 (0.066)
		Female	0.1703 (0.598)	0.0304 (0.865)	-0.0305 (0.896)	-0.0969 (0.079)
	RA	Male	-0.0002 (0.939)	-0.0002 (0.939)	-0.0002 (0.939)	-0.0002 (0.939)
		Female	-0.0161 (0.832)	-0.0041 (0.928)	0.0043 (0.929)	0.0174 (0.860)
				Time: 15		
	Net	Male	0.5232 (0.748)	0.0679 (0.905)	-0.0645 (0.855)	-0.1785 (0.021)
		Female	0.4209 (0.708)	0.0606 (0.884)	-0.0574 (0.862)	-0.1652 (0.008)
	RA	Male	-0.0006 (0.939)	-0.0006 (0.939)	-0.0006 (0.939)	-0.0006 (0.939)
		Female	-0.0186 (0.783)	-0.0044 (0.919)	0.0055 (0.923)	0.0212 (0.811)

Bias (Coverage)

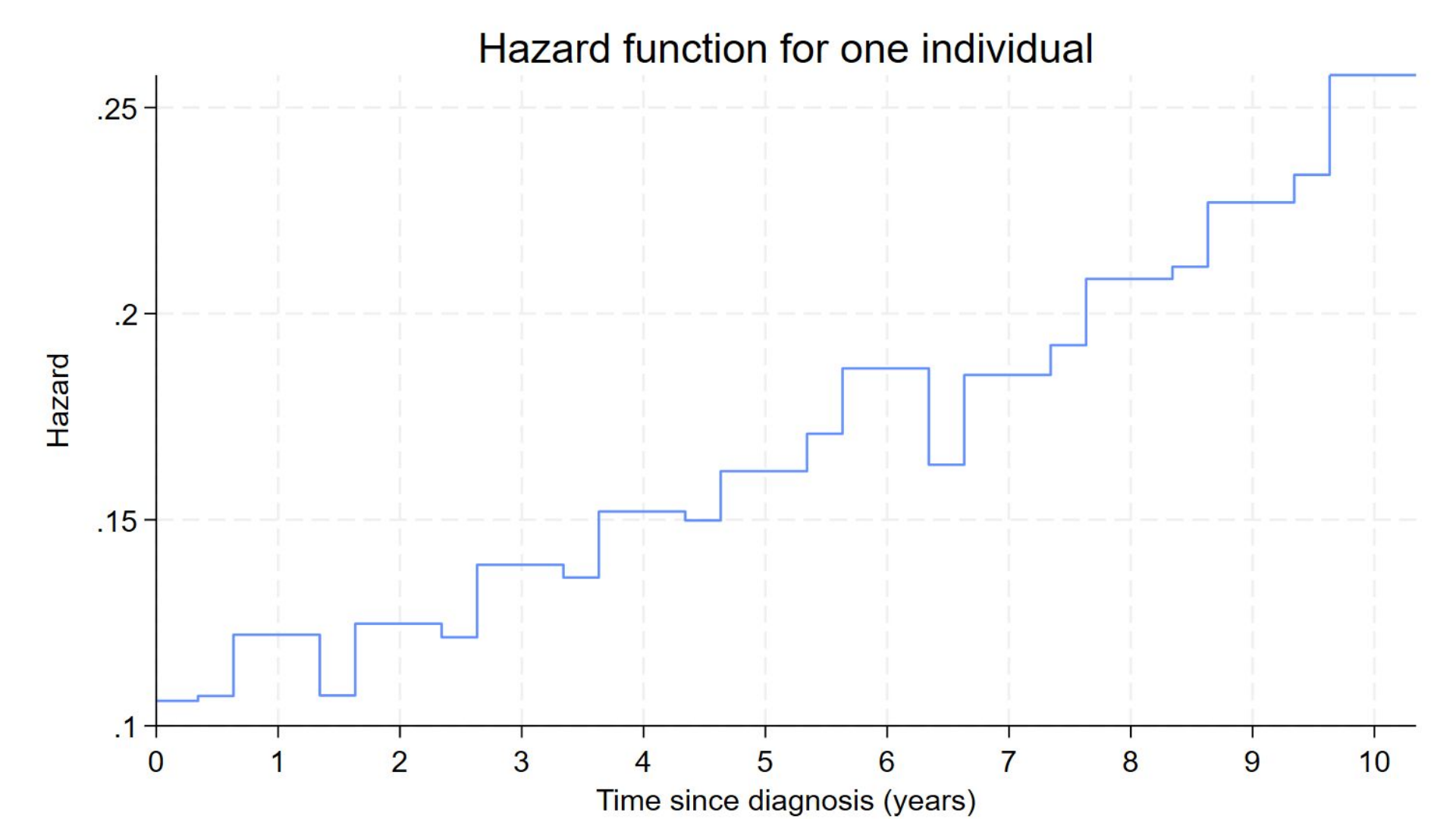
Generating relative survival data

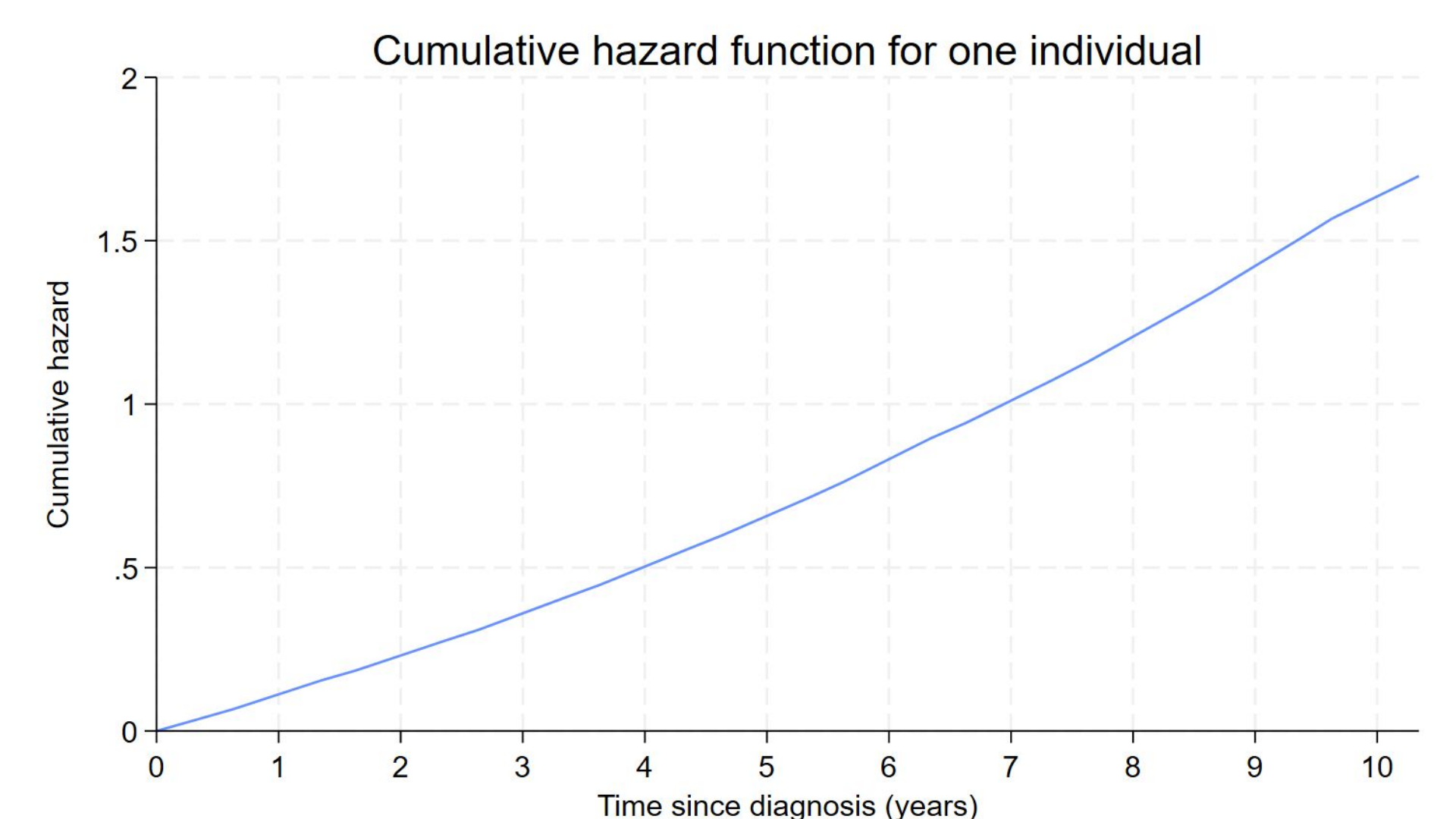
Idea

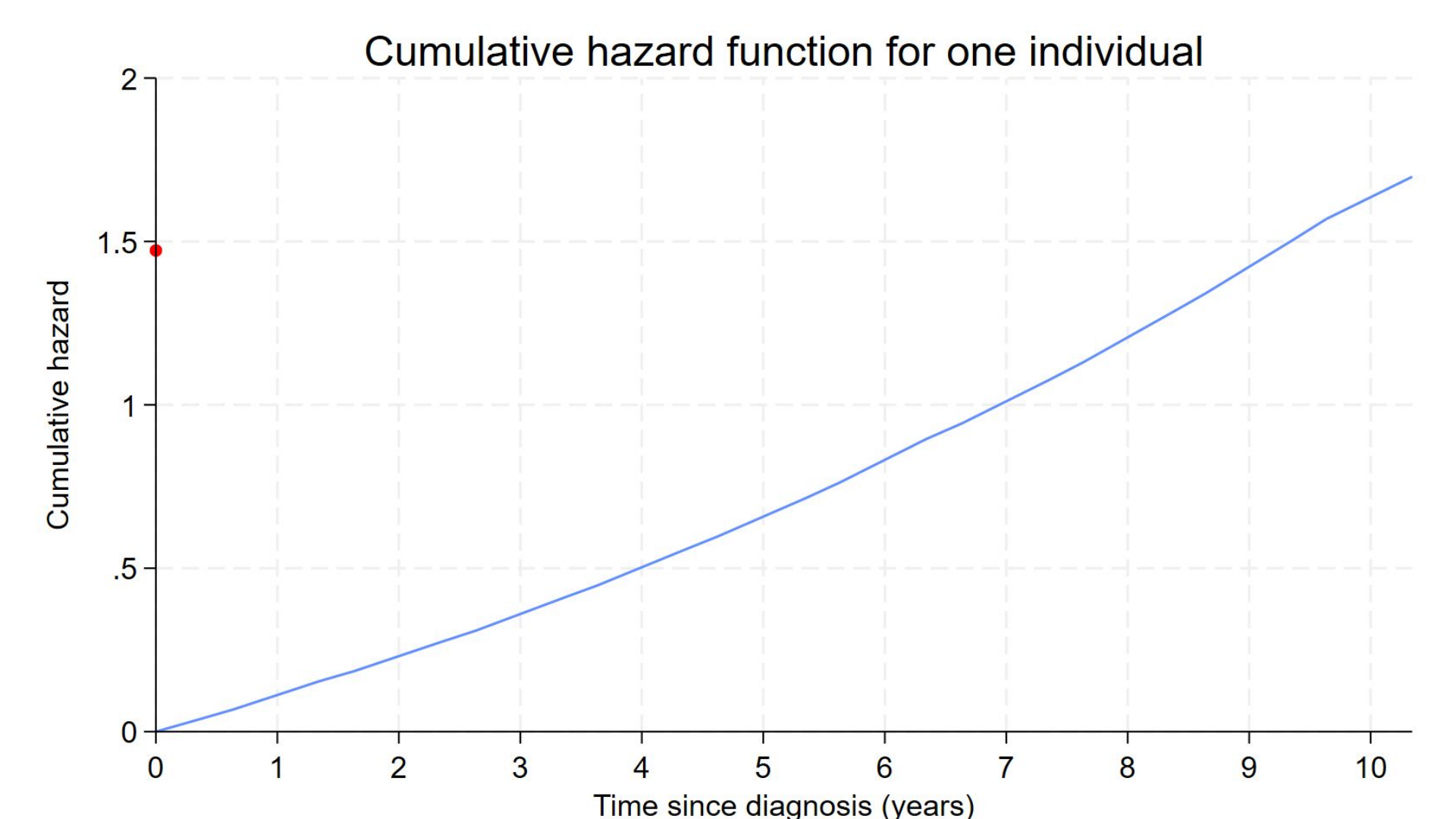
- Want a survival time due to cancer and a survival time due to other causes
  - Survival time due to cancer → easy to simulate
  - Survival time due to other causes
    - Need popmort file
    - Previous method
      - Required looping & many merges
      - Made unrealistic assumptions

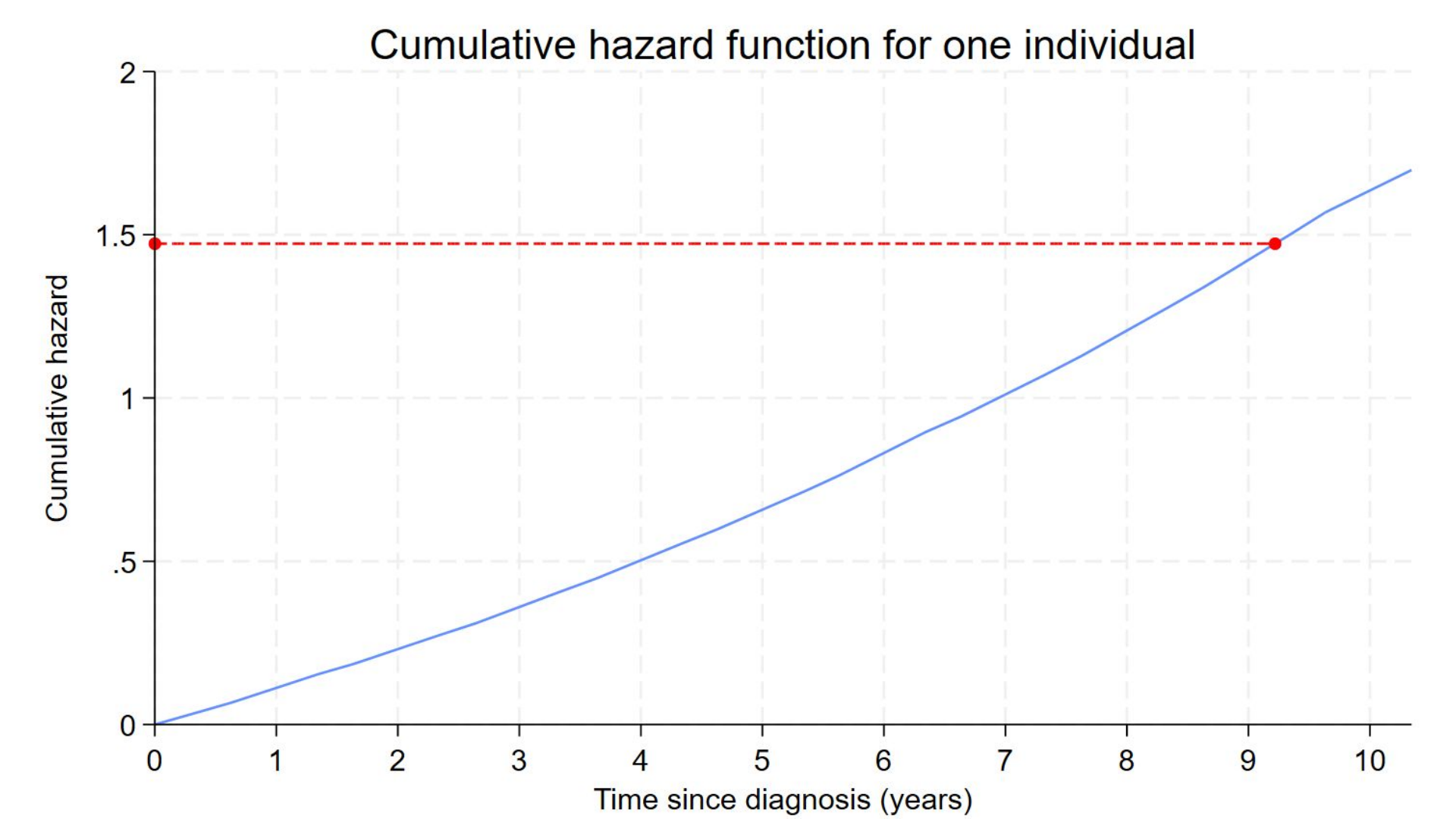
popmortsim

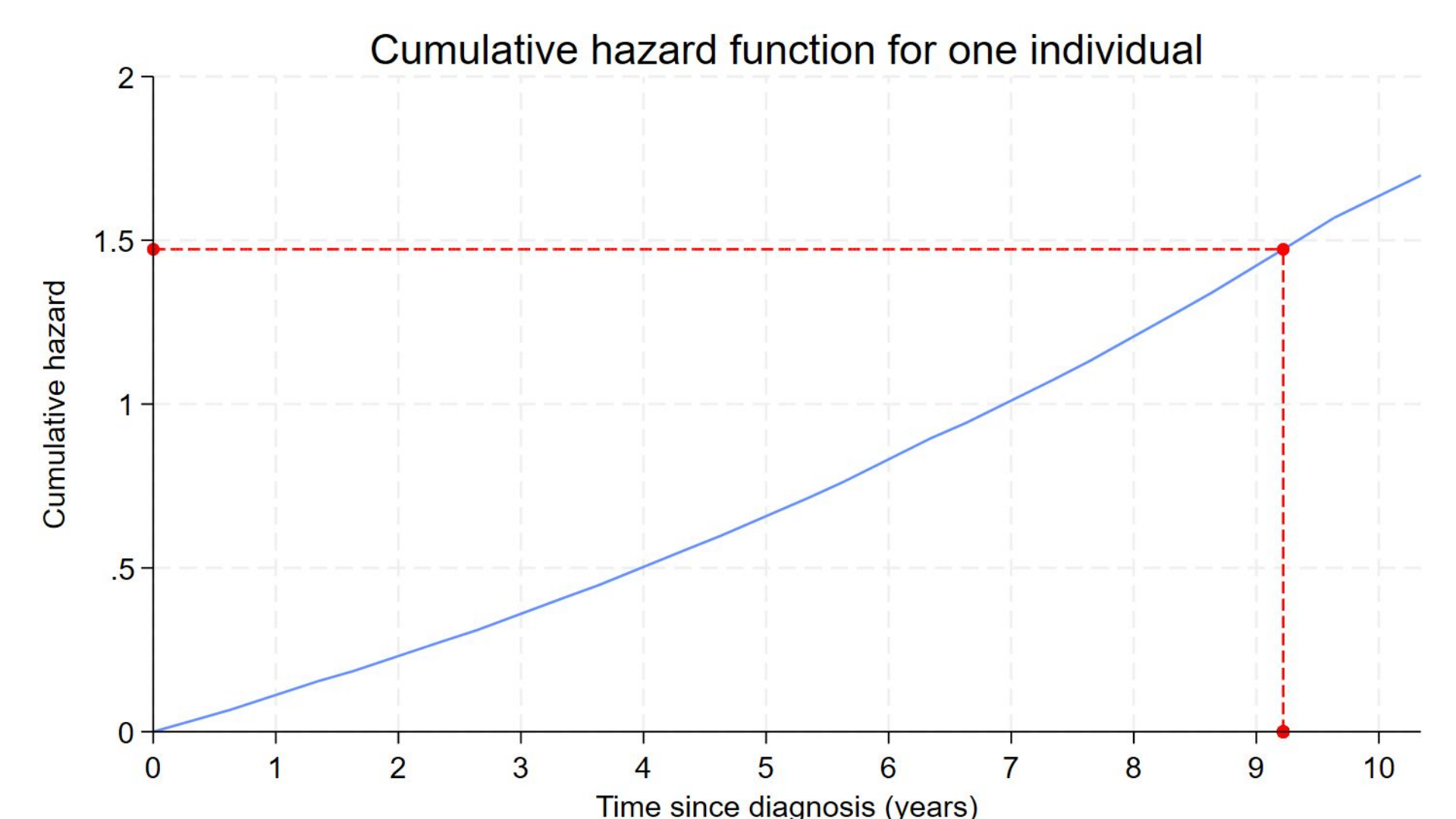
- More realistic
- Idea
  - For x years of follow-up → 2x+1 changes in hazard
  - Can expand data
    - 1 row = 1 interval with constant hazard
  - One merge
  - Generate random threshold
  - Use cumulative hazard to find a survival time


Hazard function for one individual
.25
.2
.15
.1
Hazard
0
1
2
3
4
5
6
7
8
9
10
Time since diagnosis (years)


Cumulative hazard function for one individual
Cumulative hazard
0
.5
1
1.5
2
0
1
2
3
4
5
6
7
8
9
10
Time since diagnosis (years)


Cumulative hazard function for one individual
Cumulative hazard
0
.5
1
1.5
2
0
1
2
3
4
5
6
7
8
9
10
Time since diagnosis (years)


Cumulative hazard function for one individual
Cumulative hazard
2
1.5
1
.5
0
0
1
2
3
4
5
6
7
8
9
10
Time since diagnosis (years)


Cumulative hazard function for one individual
Cumulative hazard
2
1.5
1
.5
0
0
1
2
3
4
5
6
7
8
9
10
Time since diagnosis (years)

Input → Output

```
popmortsim time death using popmort2022, ///
    agediag(agediag) datediag(datediag) pmother(sex)
```

	agediag	datediag	sex
1	81.36668	29aug1987	1

	agediag	datediag	sex	time	death
1	81.36668	29aug1987	1	9.219528	1

sessions

- Idea
  - Each simulation takes time to run but they are independent
  - Can send each scenario to a new Stata session
  - Want to only start a session is there are sufficient resources
  - Substantially reduces time

Summary and Conclusions

- Reference adjustment has **higher power** to detect differences in the non-parametric case.
- Reference adjustment is more **robust** to errors in the population mortality file. Net survival becomes biased.
- In Stata we use:
  - stpp to estimate net survival and reference adjusted all-cause survival non-parametrically
  - popmortsim to make simulation of relative survival data more realistic
  - sessions to send each scenario to a different Stata session for increased speed

References

[1] Perme MP et al. On Estimation in Relative Survival. En. Biometrics 2012 Mar; 68:113–20. doi: 10.1111/j.1541-0420.2011.01640.x.

[2] Lambert PC et al. Reference-adjusted and standardized all-cause and crude probabilities as an alternative to net survival in population-based cancer studies. en. International Journal of Epidemiology 2020 Oct; 49:1614–23. doi: 10.1093/ije/dyaa112.

Thank you for listening!