Social inequality in mortality among adults and elderly in northern Sweden 1851–2013[1]

Göran Broström and Sören Edvinsson

April 27, 2017

[1]Seminar April 27, CEDAR, Umeå

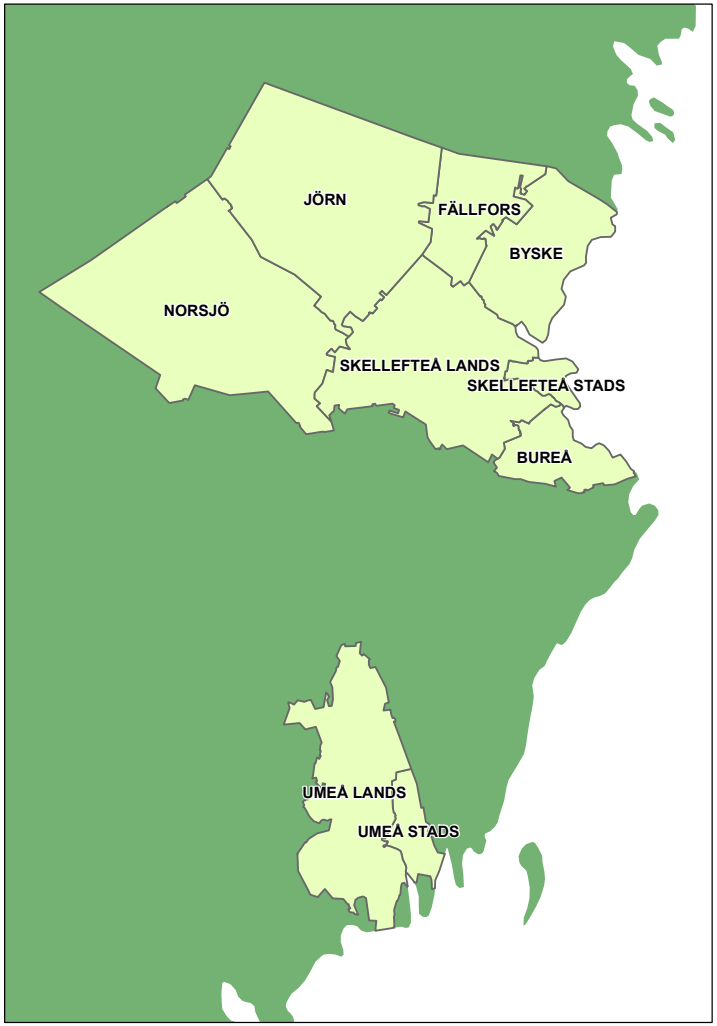
Questions

1. Has inequality in mortality between social classes increased in the adult and elderly population?
2. Are there any gender differences in the effect of social position?
3. Is social position equally important among the retired population as for those in working age?
4. Is there a difference between classes in mortality from cardiovascular diseases and cancers respectively?
5. Are there large differences in survival depending on level of education and income categories?

Outline

1. Data management
2. Modeling considerations
3. Analysis
4. Results
5. Alternative Results

Here!



Skellefteå:

- Small town, heavy industry in the nineteenth century.

Umeå:

- Not so small town, older. Center for administration, education, and military.

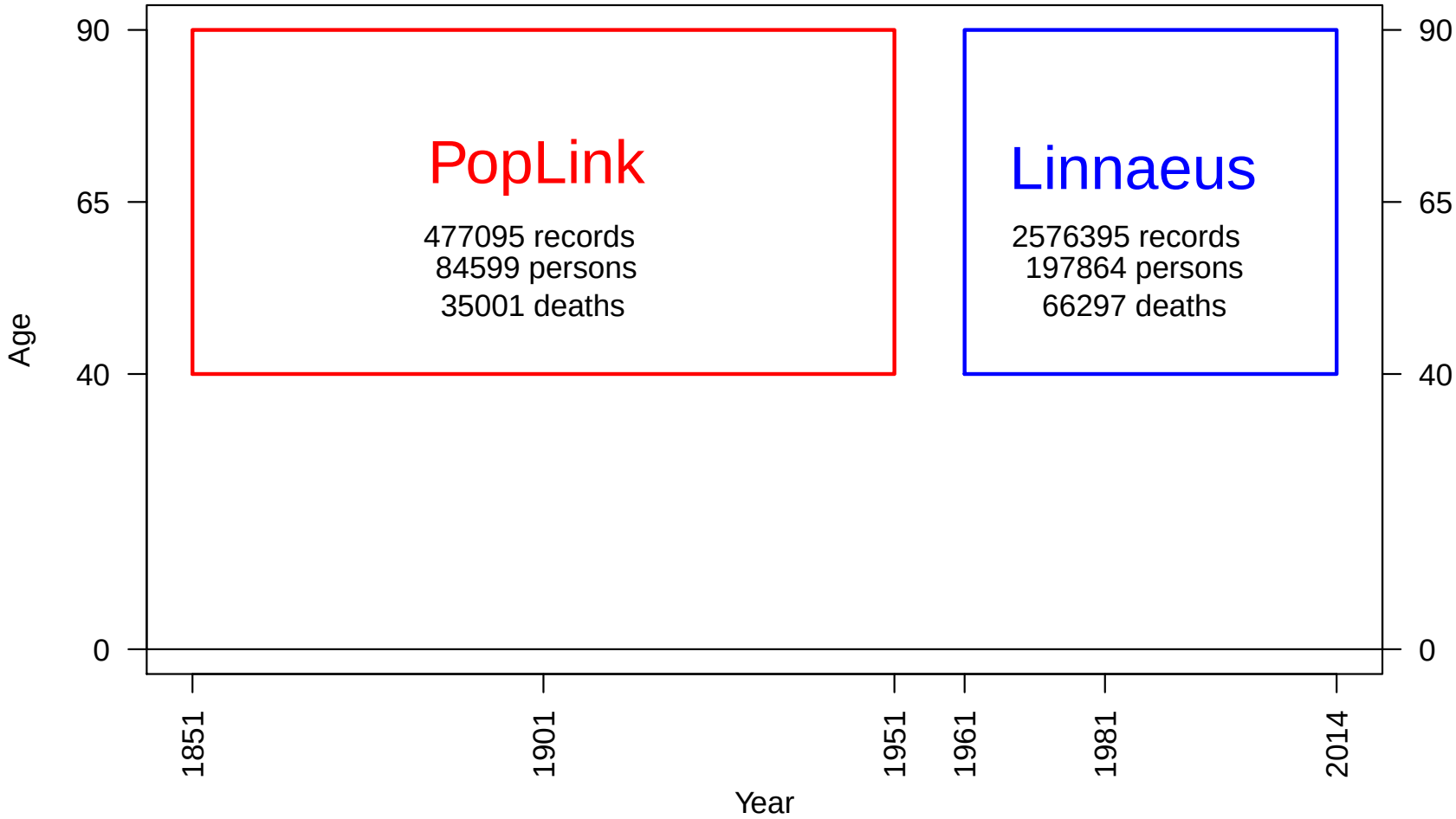
Rural areas:

- Very similar: Small farms dominating.

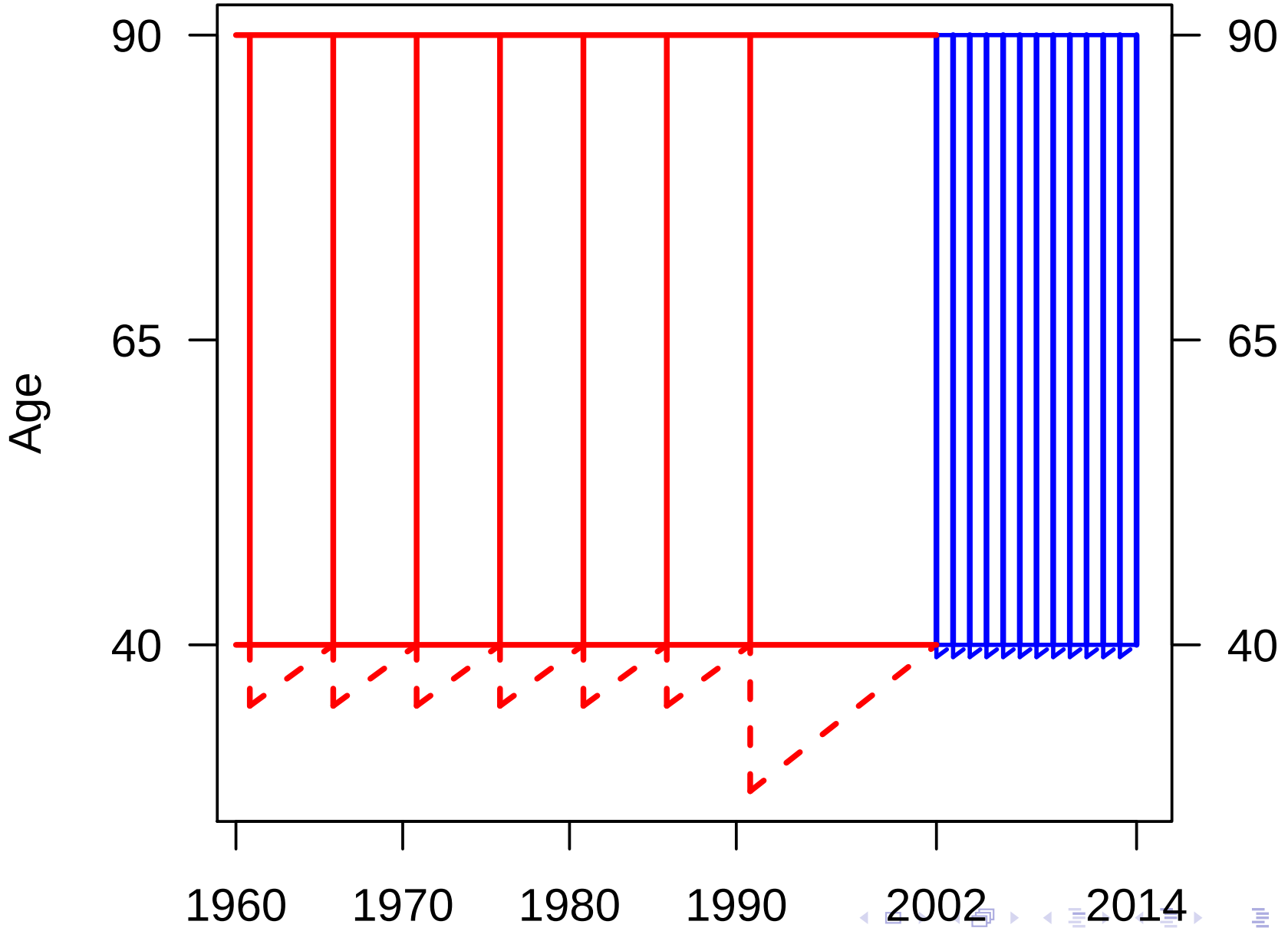
Data sources

- Swedish church book records (1851–1950).
  - digitized by the Demographic Data Base at Umeå University,
    - POPLINK
  - `http://www.cedar.umu.se/`
- The Linnaeus data base (1960–2013)
  - Census data 1960, 1965, . . . , 1990
  - Population registers (LISA, 1986–2013)
    - Income
    - Education
  - Death information from National Board of Health and Welfare
    - Death dates from 1 January 1961 to 31 December 2013.
    - Causes of death (ULORSAK, "main cause of death")

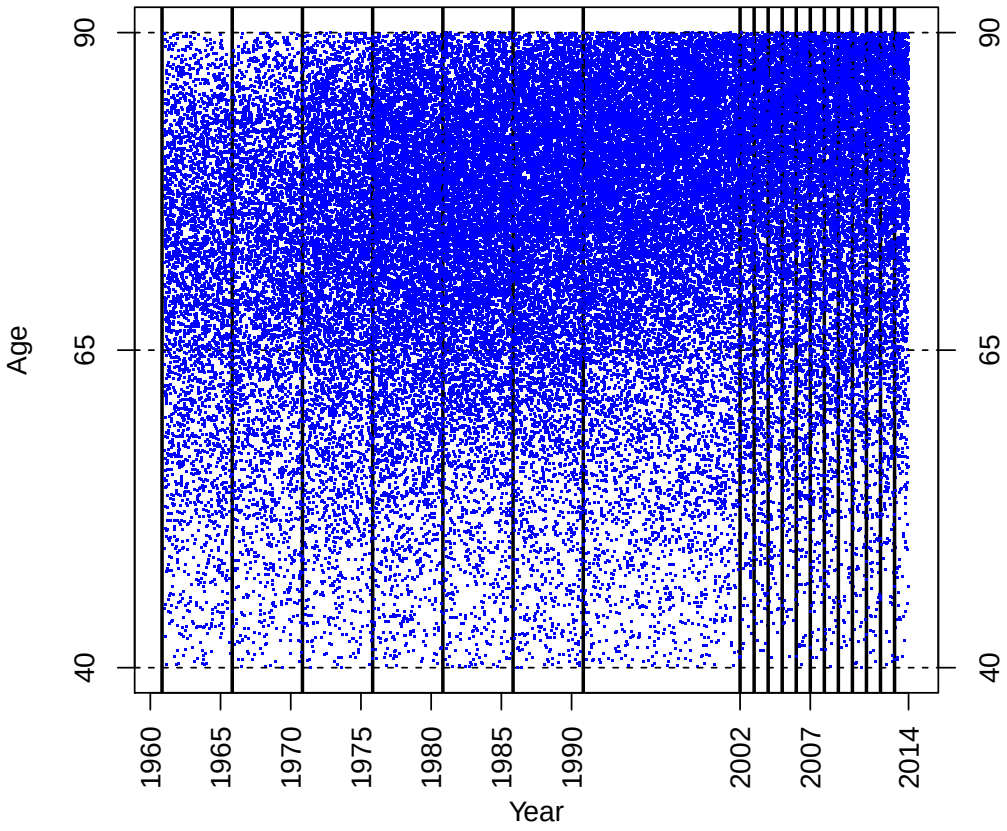
The sampling frame



The FOB and Lisa data



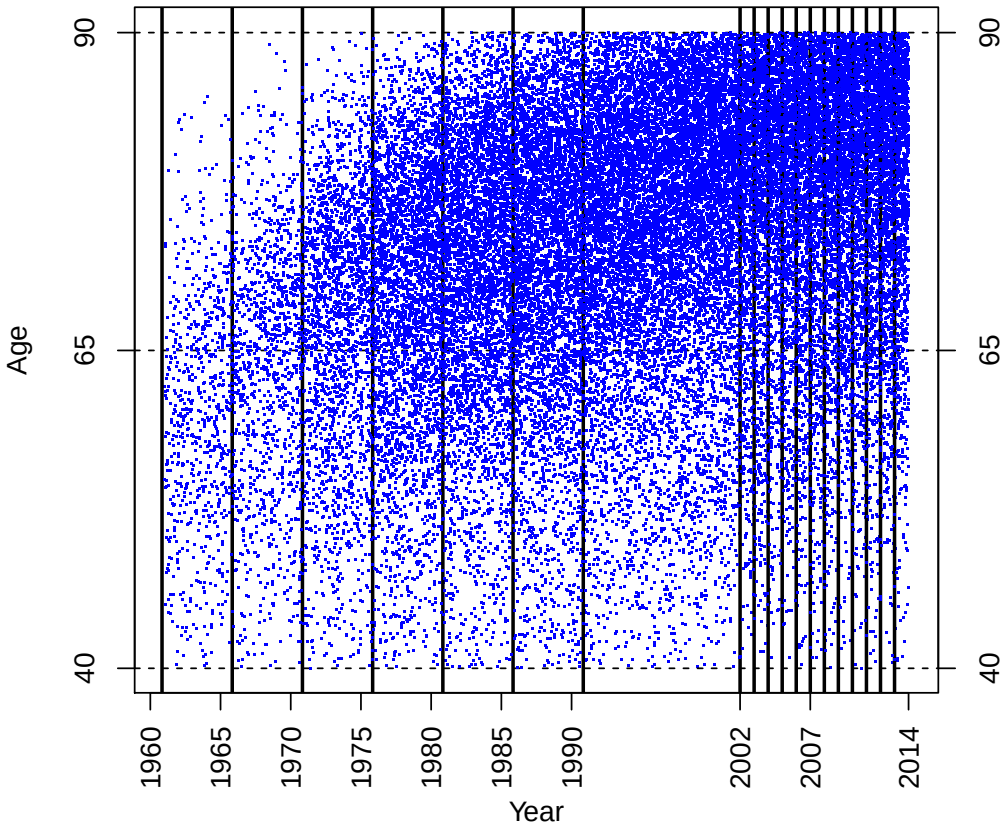
The Linnaeus data with deaths



	Number	Deaths
Women	97988	30275
Men	99876	36022
Total	197864	66297

Note: Persons with missing HISCLASS (occupation) are included.

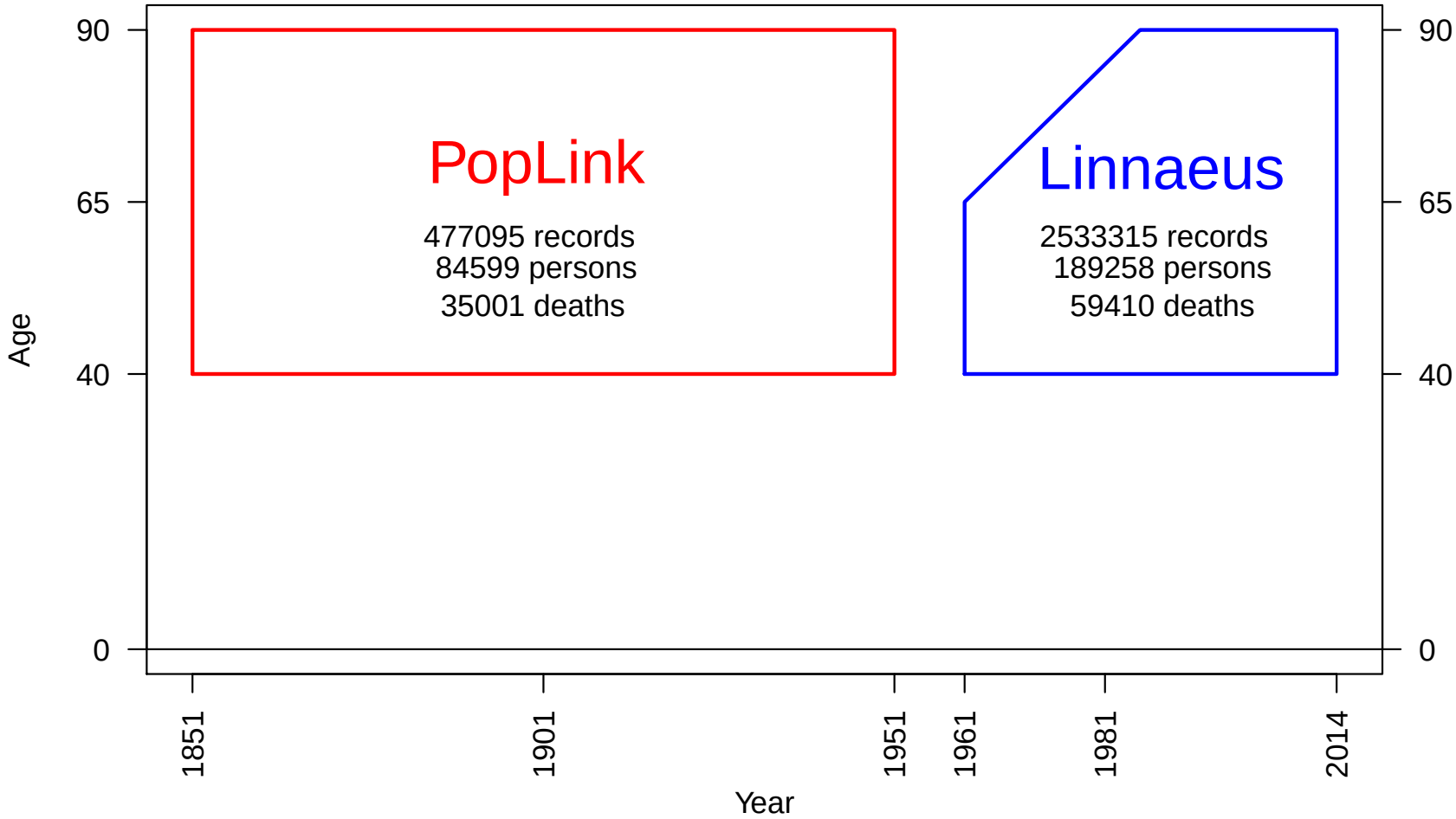
The Linnaeus data with deaths



	Number	Deaths
Women	70249	13888
Men	86427	28991
Total	156676	42879

Note: Persons with missing HISCLASS (occupation) are excluded.

The sampling frame, II



Variables

sex	
birth date	By quarter (15 Feb, May, Aug, Nov)
death date	By day (Socialstyrelsen)
cause of death	from Socialstyrelsen
parish	Urban/Rural
HISCLASS	elite, middle, worker.
income	1990–2005, grouped by quartiles.
education	1990–2005, seven categories.
period	1851–1875, . . . , 2008–2013 (10 periods).
cohort	1761–1771, . . . , 1971–1974 (22 birth cohorts).

Hisclass

1. `elite`, HISCLASS 1 and 2. Higher managers and professionals.
2. `middle`, HISCLASS 3, 4, 5, 6, and 8. Lower managers, lower white collar, and farmers.
3. `worker`, HISCLASS 7, 9, 10, 11 and 12. Workers of different skills including farm workers.

Preliminary Cox regression

```
library(eha)
system.time(fit <- coxph(Surv(enter, exit, event) ~
                              hisclass * period + civst + urban + sex,
                         data = vb0))

   user  system elapsed
151.360   0.444 151.806

system.time(dr <- drop1(fit, test = "Chisq"))

   user  system elapsed
455.236   0.944 456.188
```

It takes time (more than 10 minutes)!

ANOVA

	Df	AIC	LRT	Pr(>Chi)
<none>		2227937.83		
civst	2	2228508.16	574.34	0.0000
urban	1	2227962.96	27.13	0.0000
sex	1	2231526.62	3590.79	0.0000
hisclass:period	27	2228575.58	691.75	0.0000

And with so much data, everything tends to be statistically significant.

Test of Proportional Hazards

```
    user   system  elapsed
1063.928    0.916 1064.028
```

Another 19 minutes!

Proportional hazards, results

```
#system.time(cz <- cox.zph(fit))
cz$table[NROW(cz$table), ]

    rho   chisq         p
     NA 2064.07    0.0000
```

- As usual, highly significant result (PH assumption violated).
- Should we worry?
  - Yes
  - No

Sufficient statistics

Aggregate:

```
vb0Tab <- aggregate(vb0[, c("event", "exposure")],
                    vb0[, c("sex", "age", "period", "cohort",
                            "urban", "civst", "hisclass")],
                    FUN = sum)
vb0Tab$age <- factor(vb0Tab$age)
nr <- NROW(vb0Tab)
```

- Gives a table assuming piecewise constant hazards on five-year intervals: 40–45, 45–50, . . . , 85–90 ("age").
- Number of cells: $2 \times 10 \times 10 \times 20 \times 2 \times 3 \times 4 = 96000$
- Number of non-empty cells: 10785.

`http://capa.ddb.umu.se/ds/aggregate.html`

PH regression with the table

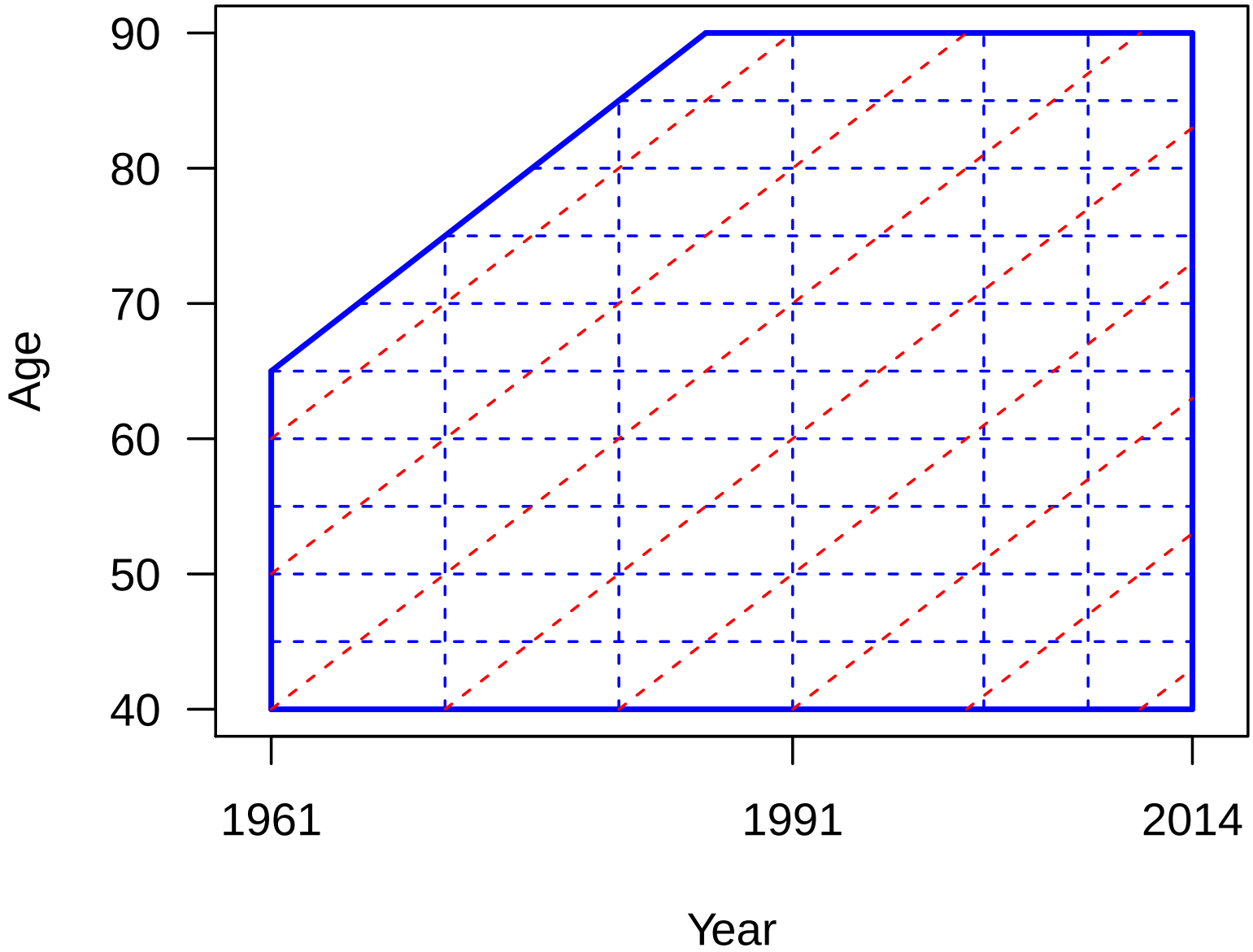
- Piecewise constant hazard model generates the same likelihood as a Poisson model.
  - Does not imply that the number of deaths follow a Poisson distribution!

```
system.time(fitp <- glm(event ~ offset(log(exposure)) + age +
                  hisclass * period + civst + urban + sex,
              data = vb0Tab, family = poisson))

   user  system elapsed
  0.884   2.216   0.570

system.time(dr <- drop1(fitp, test = "Chisq"))

   user  system elapsed
  2.476   6.440   1.164
```

ANOVA with the table

	Df	Deviance	AIC	LRT	Pr(>Chi)
<none>		17539.79	43607.22		
age	9	116442.66	142492.08	98902.87	0.0000
civst	2	18254.46	44317.88	714.67	0.0000
urban	1	17553.35	43618.78	13.56	0.0002
sex	1	21008.66	47074.08	3468.87	0.0000
hisclass:period	27	18167.89	44181.31	628.09	0.0000

Age–Period–Cohort data



Counting "cells"

- There are 115 cells in one figure
- There are $2 \times 2 \times 3 \times 4 = 48$ such figures.
- In total: $115 \times 48 = 5520$ combinations.
- Number of non-empty combinations: 4072.

	sex	age	period	cohort	urban	civst	hisclass	event	exposure
158	female	70	1981-1990	1901	FALSE	unmarried	elite	0	5.3750000
159	male	75	1981-1990	1901	FALSE	unmarried	elite	0	6.5000000
160	female	75	1981-1990	1901	FALSE	unmarried	elite	2	11.6873370
161	male	80	1981-1990	1901	FALSE	unmarried	elite	2	1.8648946
162	female	80	1981-1990	1901	FALSE	unmarried	elite	0	6.3761636
163	female	85	1981-1990	1901	FALSE	unmarried	elite	0	0.3755818

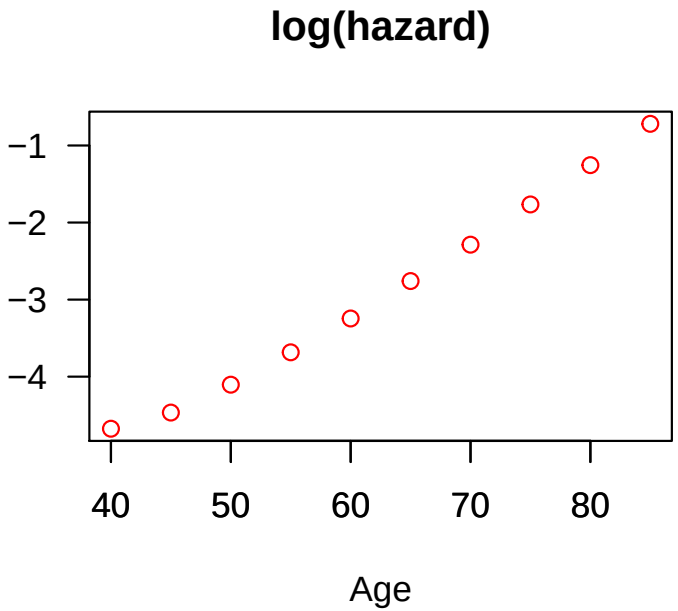
An anonymous data set!

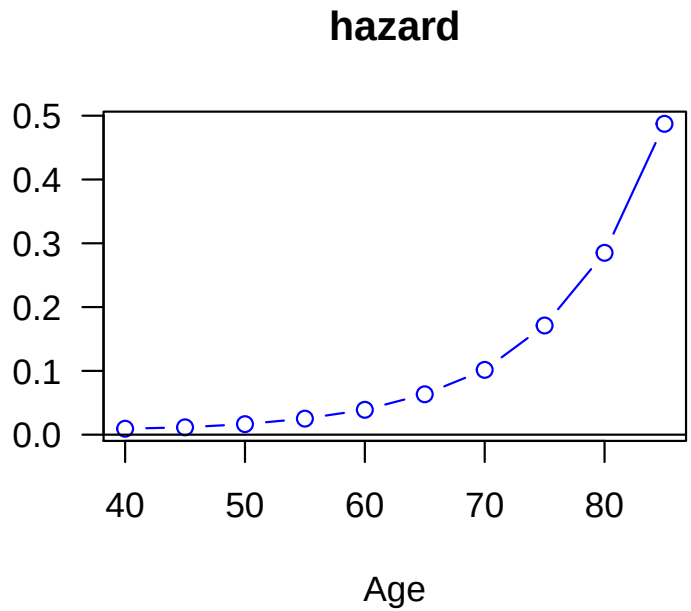
The 'baseline hazard function'

	Estimate	Std. Error
(Intercept)	-4.68	0.13
age45	0.21	0.02
age50	0.57	0.02
age55	0.99	0.02
age60	1.43	0.02
age65	1.92	0.02
age70	2.39	0.02
age75	2.91	0.02
age80	3.42	0.02
age85	3.96	0.02

- (Intercept) is log hazard in 40–45. Reference interval.
- age45 is the difference in log hazard between 45–50 and 40–45,
- age50 is the difference in log hazard between 50–55 and 40–45,
- Etcetera ...

Baseline hazards



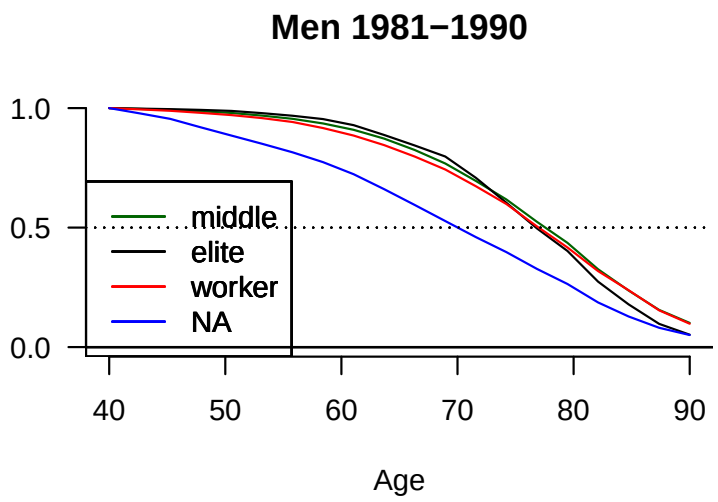


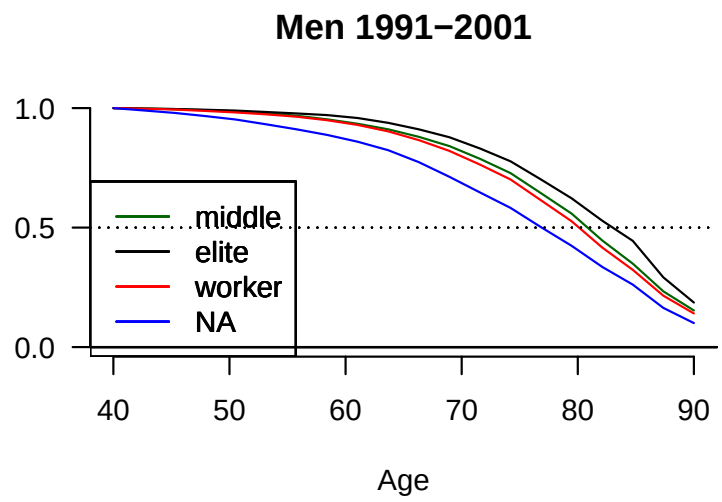
Why not try:

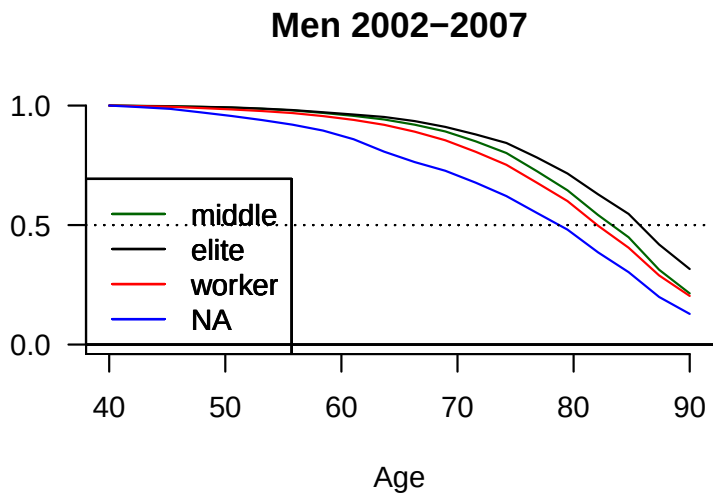
$$\log\{h(t)\} = \alpha + \beta t, \quad 40 < t \leq 90,$$

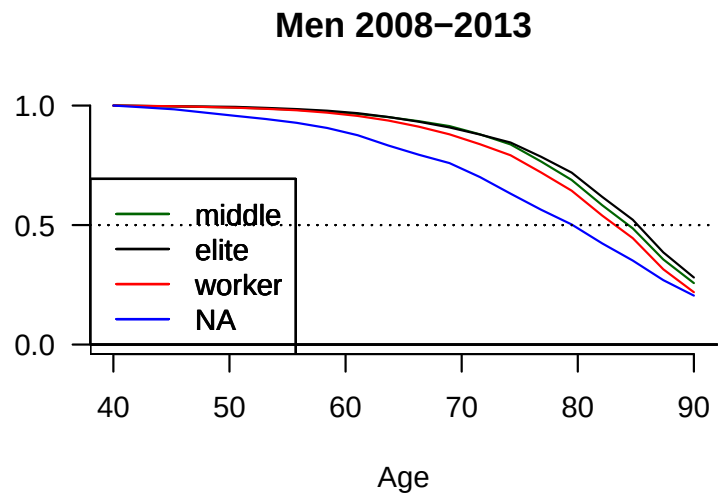
a Gomperz distribution?

Survival analysis, men 1961-2013

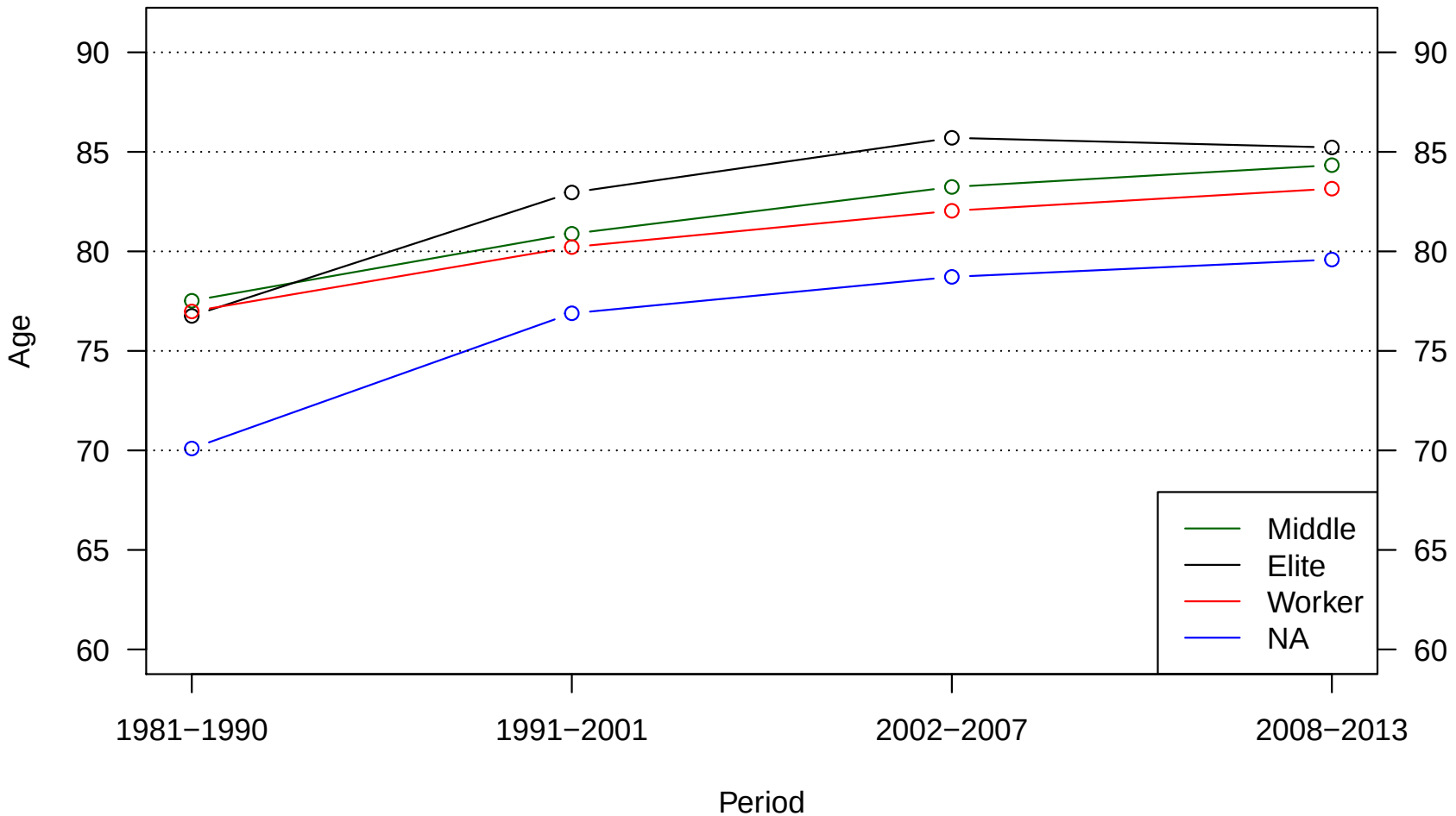




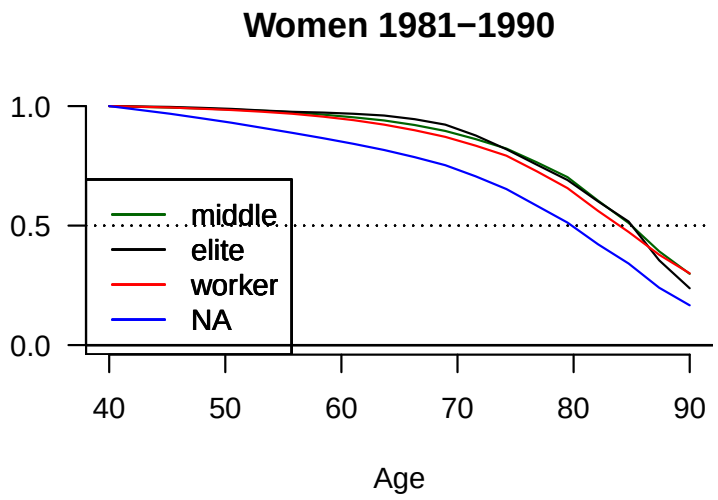


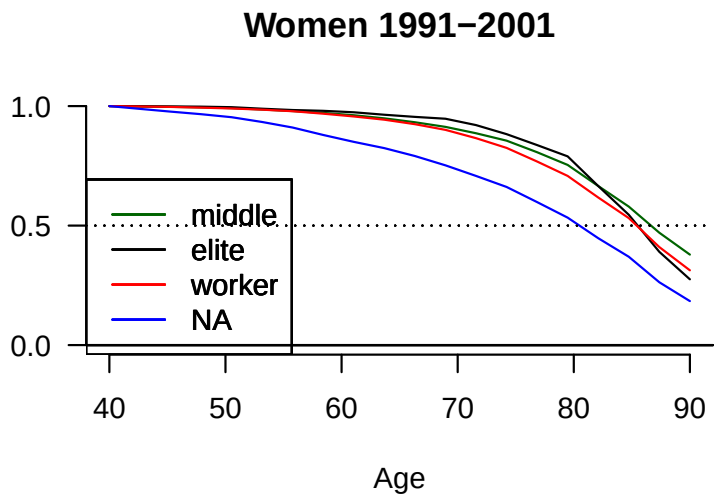


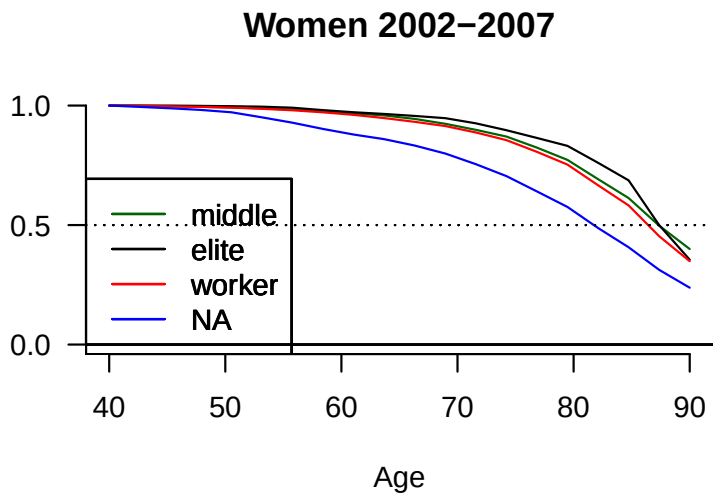
Median life after 40, men 1961-2013

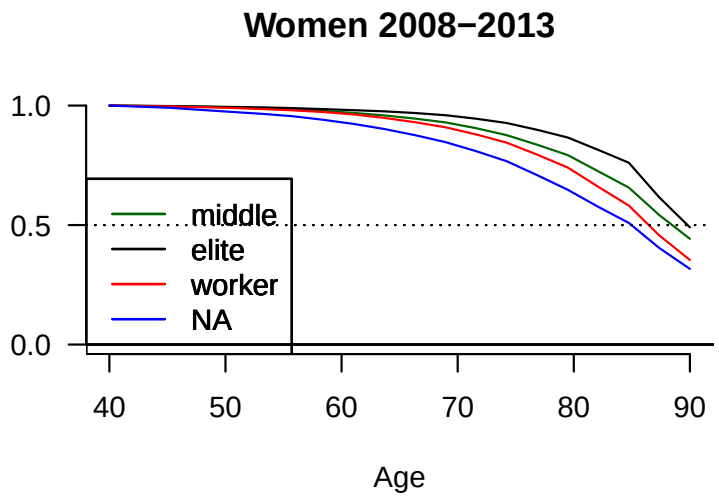


Survival analysis, women 1961-2013

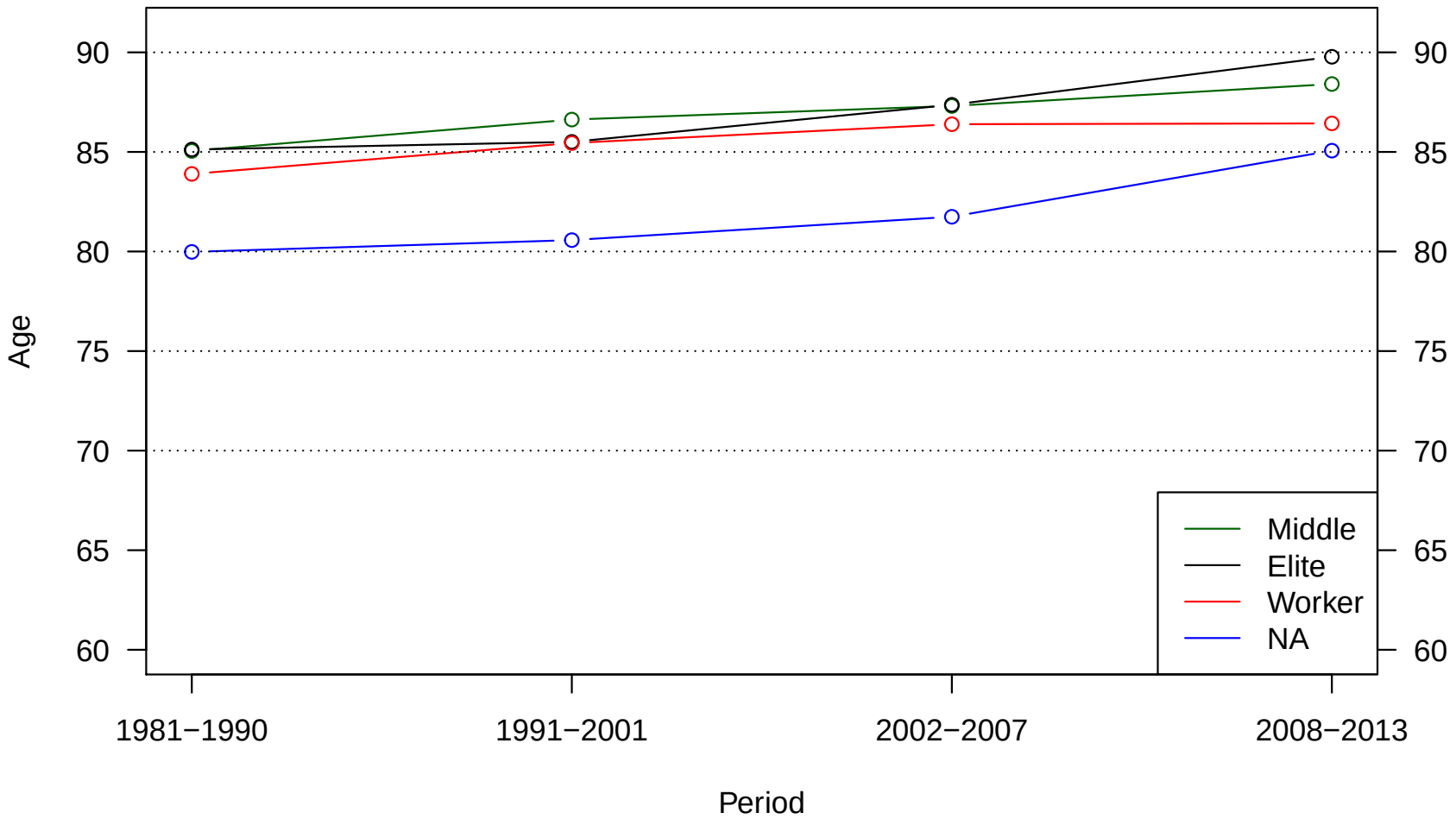




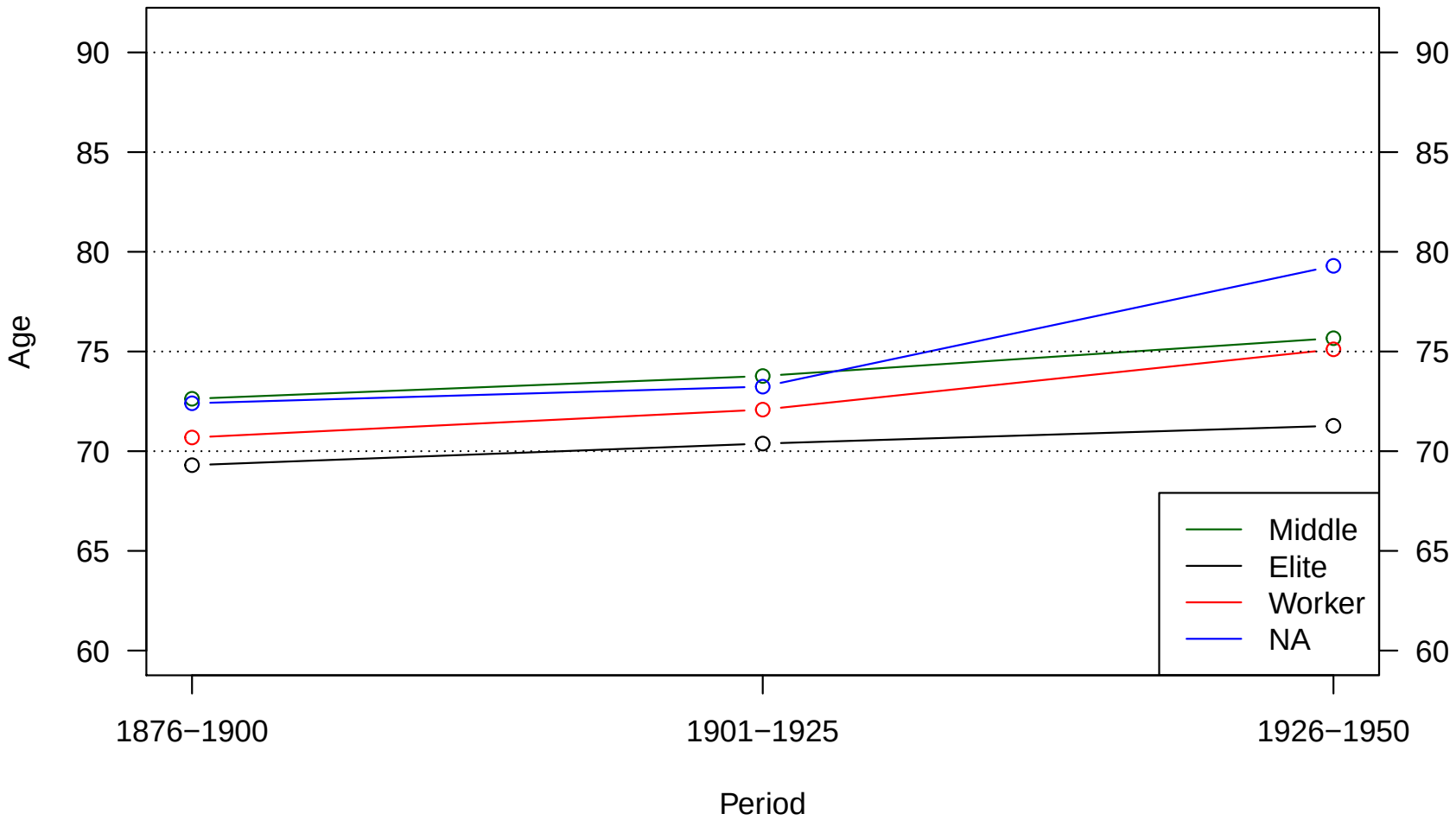




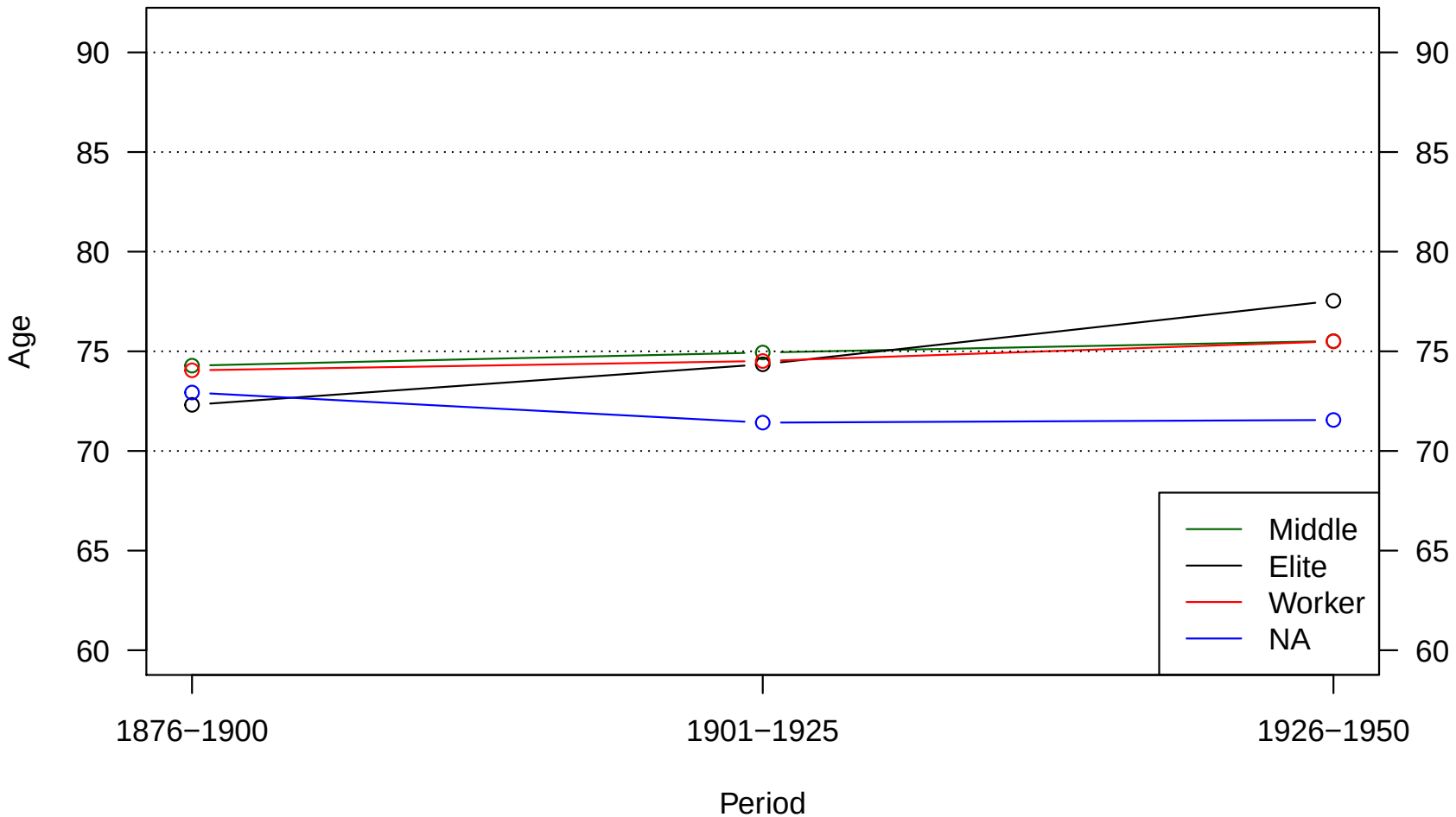
Median life after 40, women 1961-2013



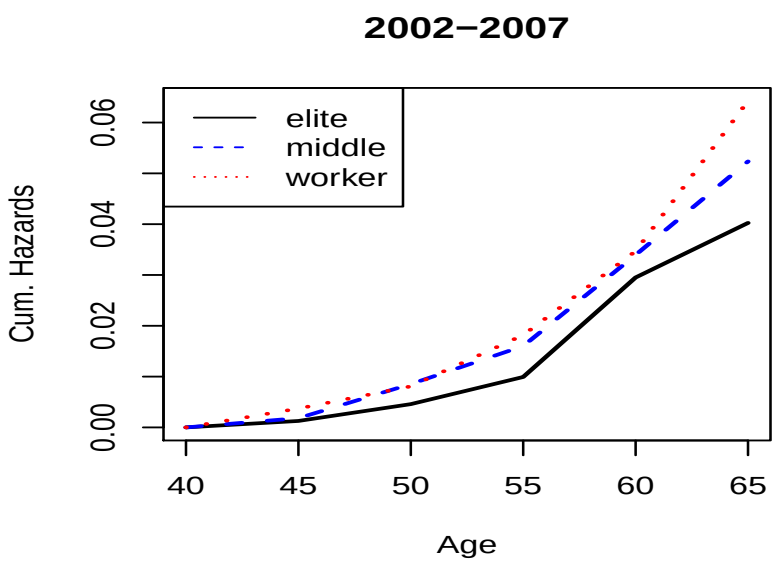
Median life after 40, men 1851-1950

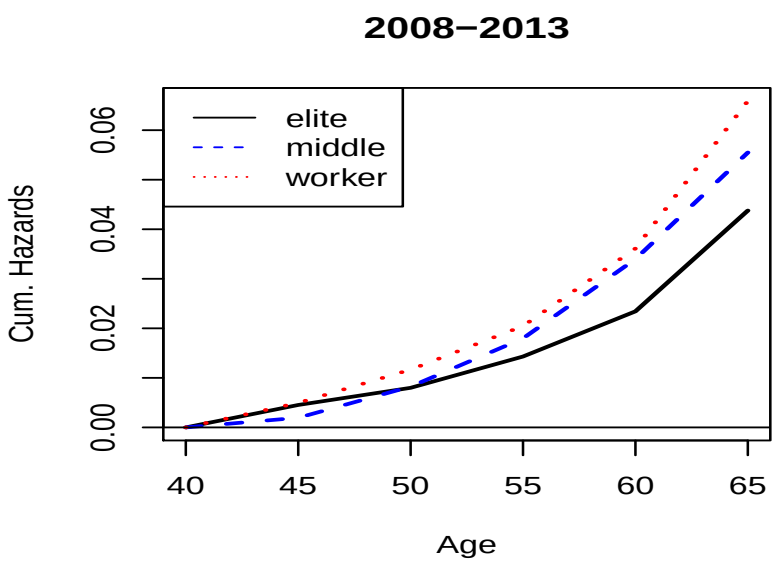


Median life after 40, women 1851-1950

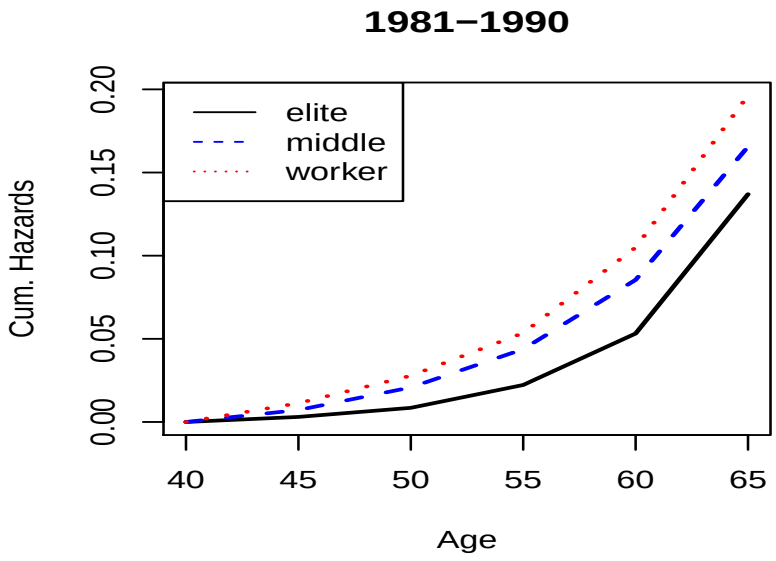


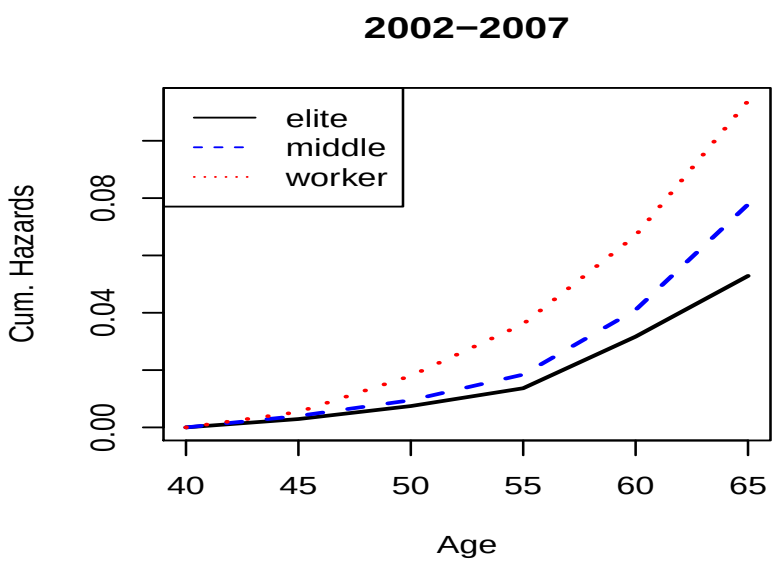
Proportional hazards for women?

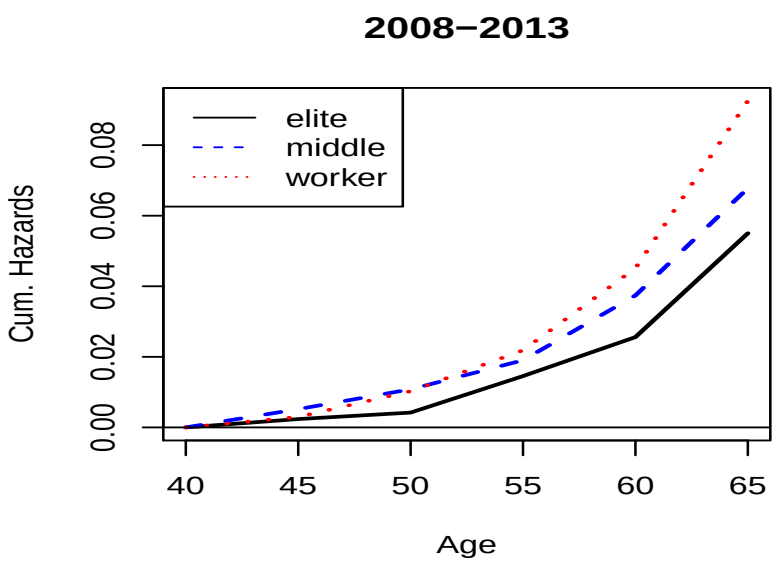




Proportional hazards for men?







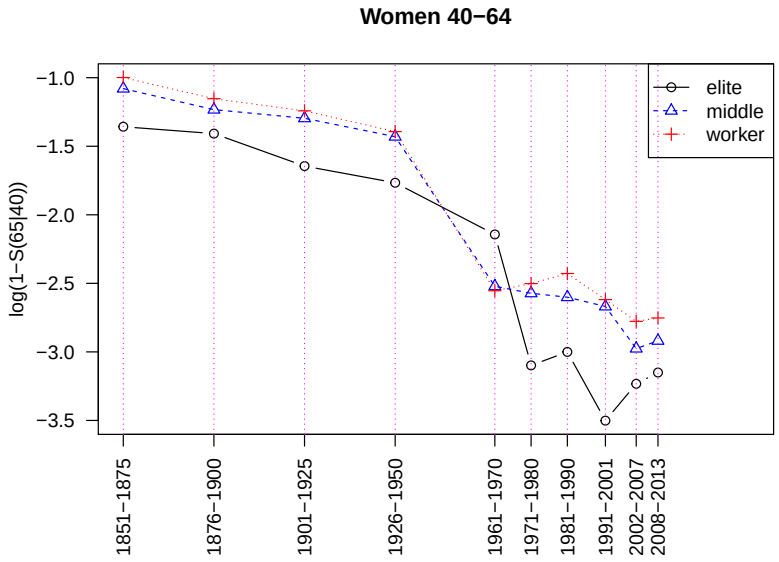
No proportional hazards (wrt HISCLASS)!

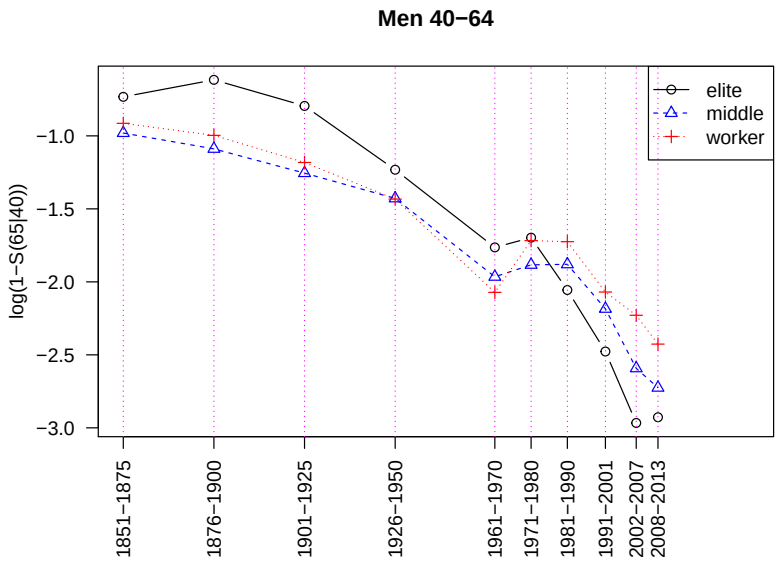
- Fit each HISCLASS separately, for each time period.
- Use the cumulative hazard at 65 (or 90) for comparison.

So, we fit separate PH models for each combination of

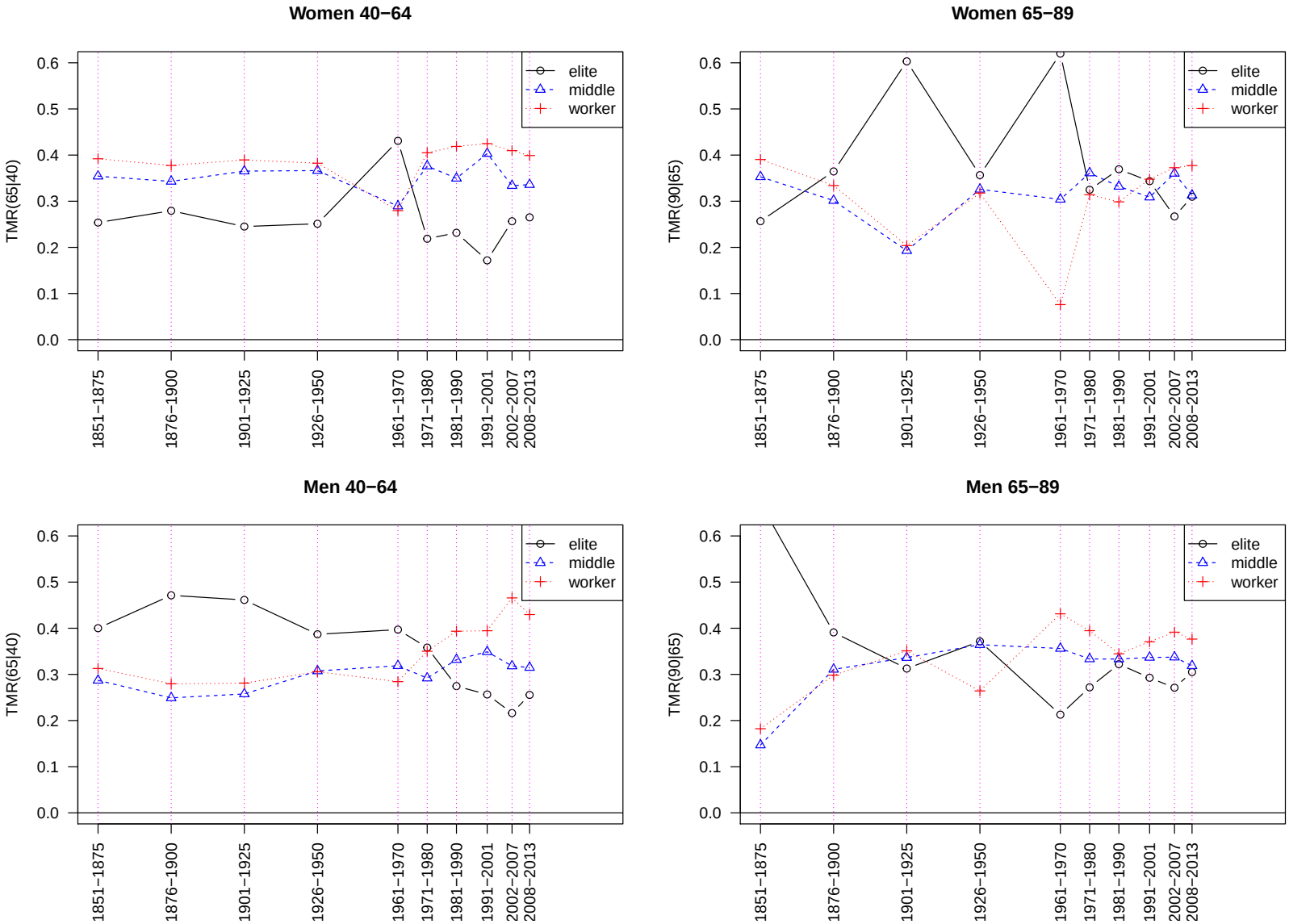
- `sex`
- `period`: 1851–75, 1876–1900, 1901–10, etc.
- `hisclass`
- `age group`: 40–64, 65–89.
- `cause of death`

Women and men, age 40–64

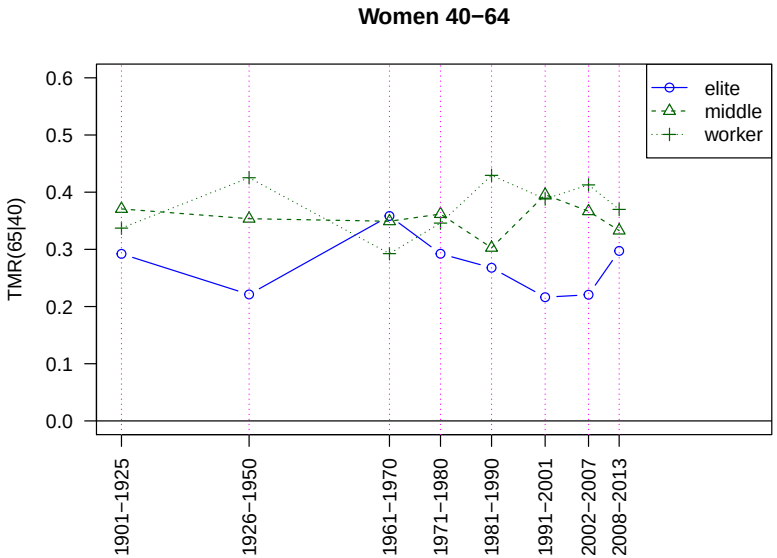


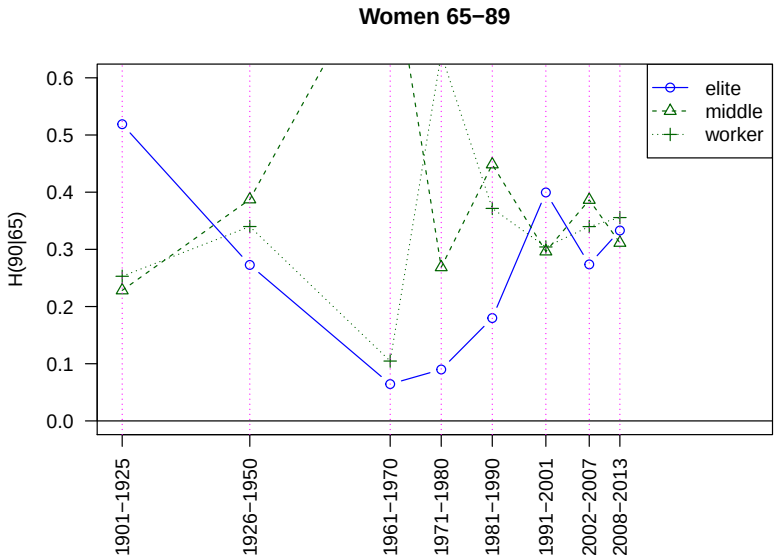


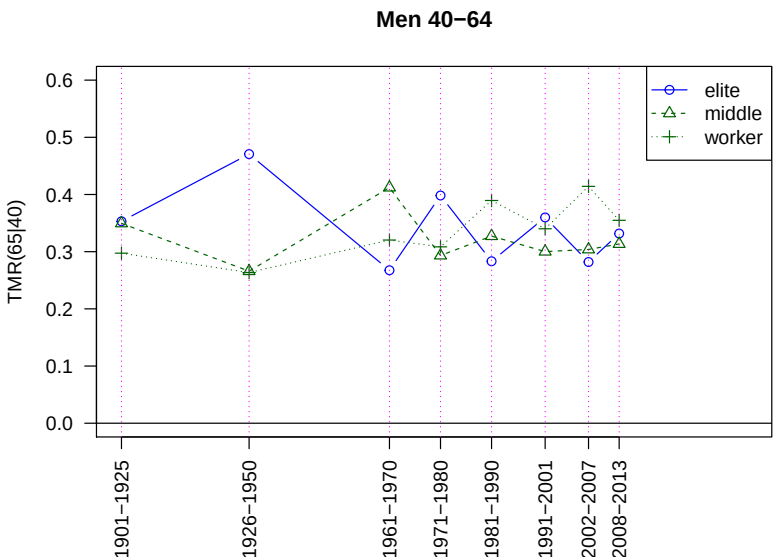
Relative differences, ages 40–64 and 65–89

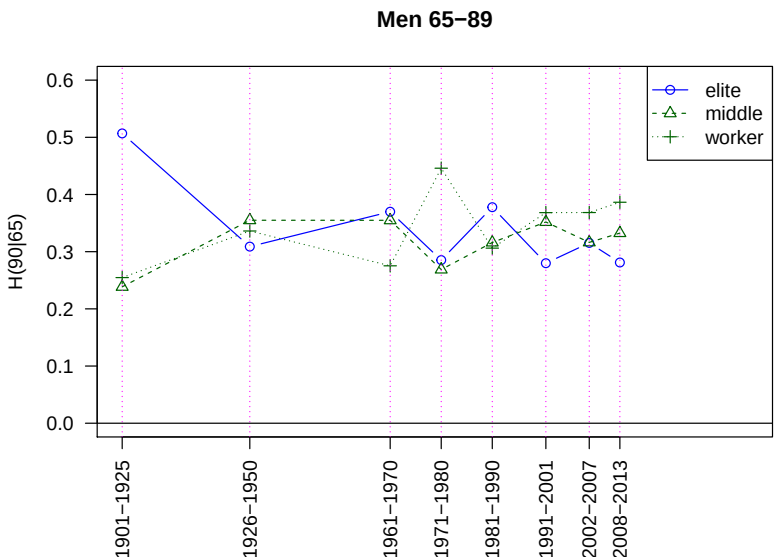


Cancer mortality, ages 40-64 and 65-89

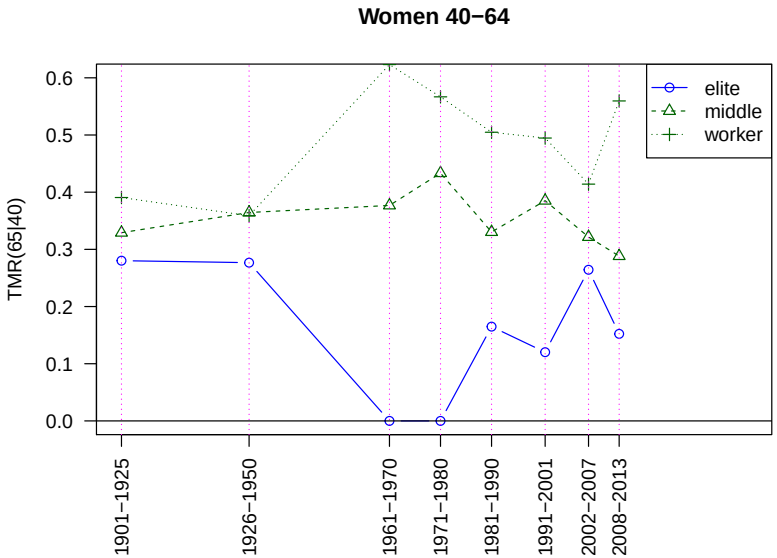


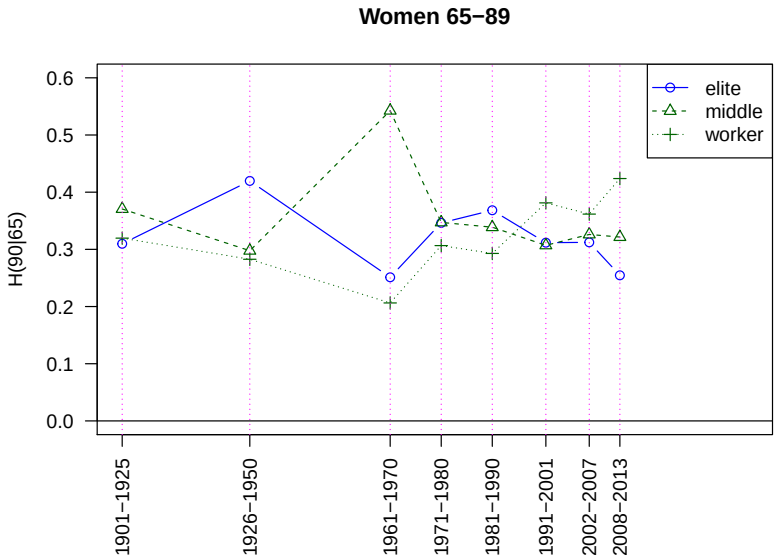


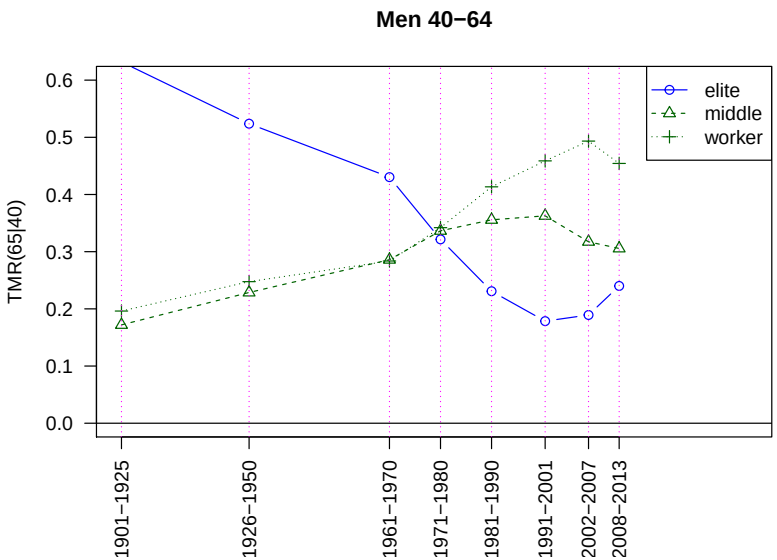


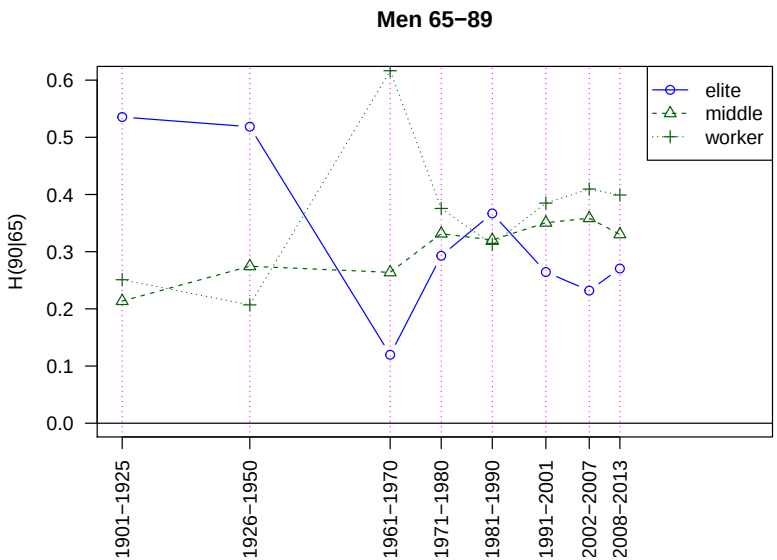


Cardiovascular mortality

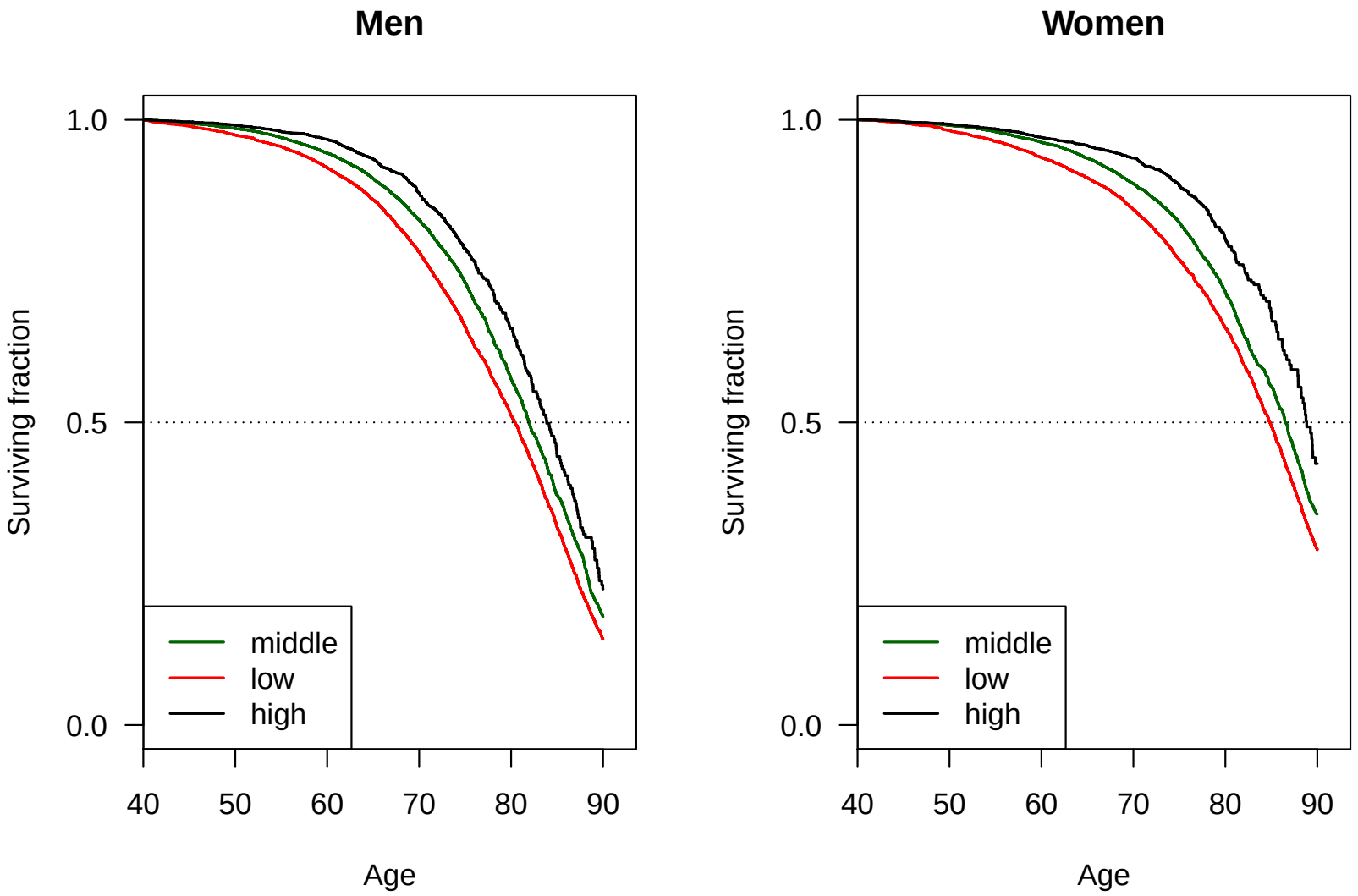




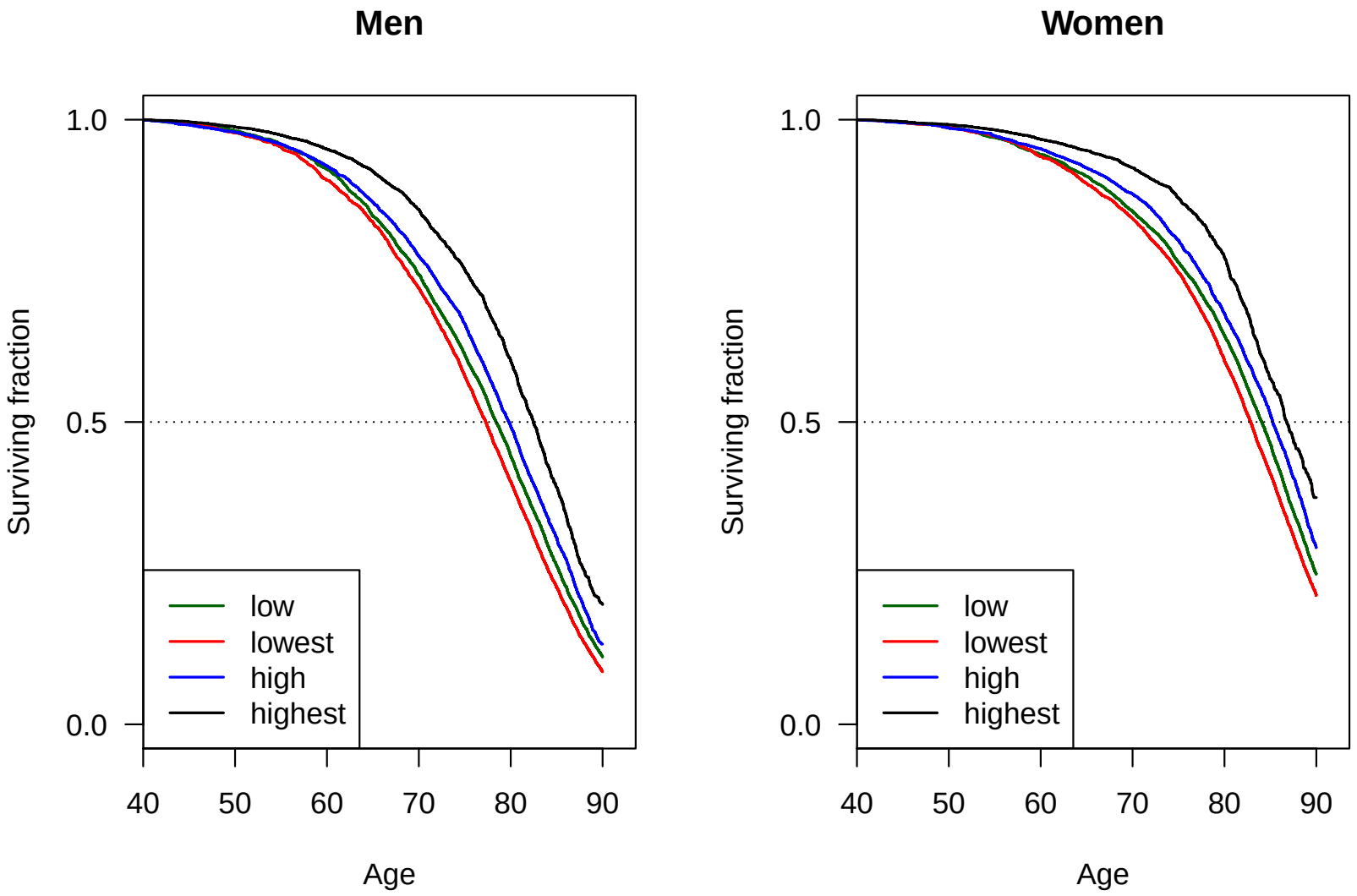




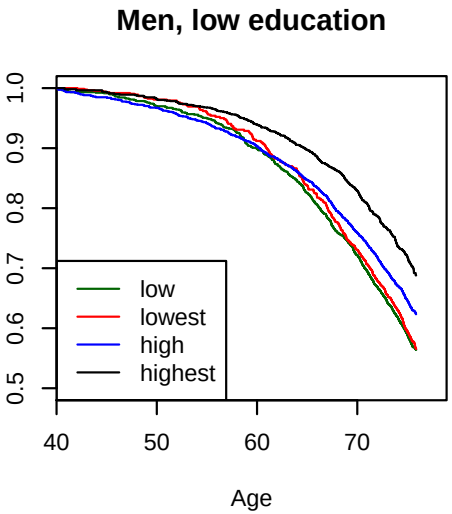
Education 1990–2013

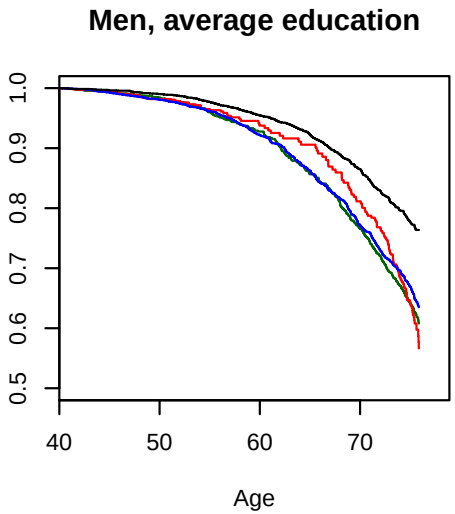


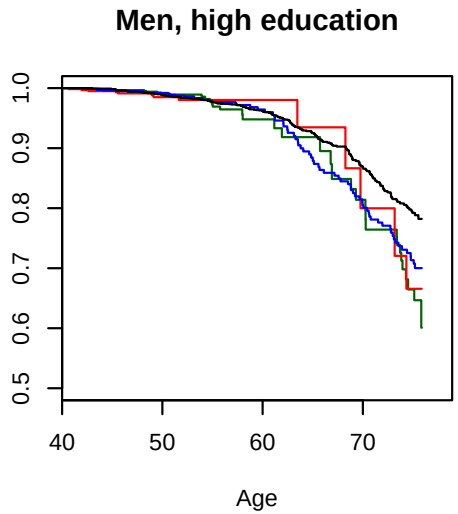
Income 1986–2005

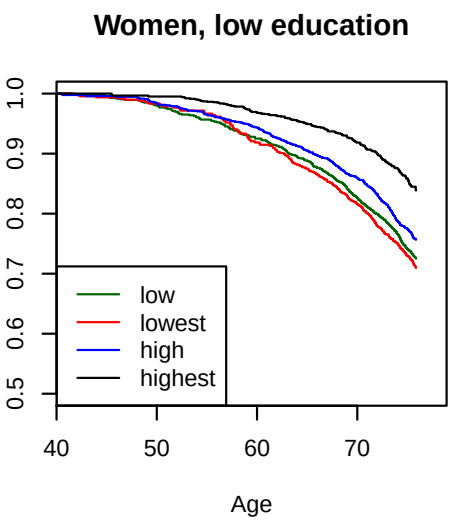


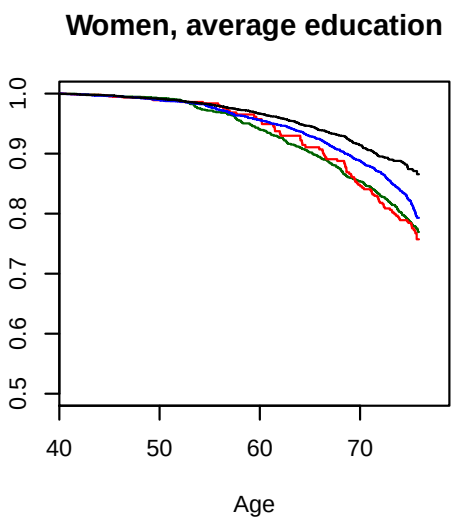
Income by education, 1990–2005, ages 40–75!

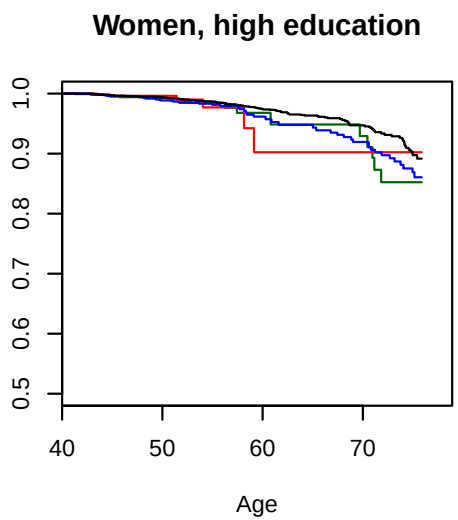












Answers(?)

1. Has inequality in mortality between social classes increased?
   - Not really
2. Are there any gender differences in the effect of social position?
   - Yes
3. Is social position equally important among the retired population as for those in working age?
   - Not really
4. Is there a difference between classes in mortality from cardiovascular diseases and cancers respectively?
   - Yes
5. Are there large differences in survival depending on level of education and income categories?
   - Yes, definitely.

Conclusion

- Improved survival characterised all HISCLASSes.
- There are gender differences in the social pattern
- No difference in the social pattern in the two studied age groups.
- The upper class(es) gradually take over as the most healthy group.
- Education and income are the most important variables during late time period.