



**Actuarial  
Research Centre**

Institute and Faculty  
of Actuaries

# Beyond Proportional Hazards

**Technical Workshop**

Wednesday 30 October 2019, Staple Inn Hall

**ARC Research Programme - Use of Big Health and Actuarial Data for understanding Longevity and Morbidity Risks**

**Programme's webpage - <http://bit.ly/arctechworkshop>**

**Get in touch - [arc@actuaries.org.uk](mailto:arc@actuaries.org.uk)**



**Actuarial  
Research Centre**

Institute and Faculty  
of Actuaries

# Cox regression, the proportional hazards assumption and time-varying covariates

Dr Ilyas Bakbergenuly (UEA)

The ‘**Use of Big Health and Actuarial Data for understanding Longevity and Morbidity Risks**’ research programme is being funded by the Actuarial Research Centre.

30/10/2019

[www.actuaries.org.uk/arc](http://www.actuaries.org.uk/arc)



## Actuarial Research Centre

Institute and Faculty  
of Actuaries

### Disclaimer

The views expressed in this presentation are those of invited contributors and not necessarily those of the Institute and Faculty of Actuaries. The Institute and Faculty of Actuaries does not endorse any of the views stated, nor any claims or representations made in this presentation and accept no responsibility or liability to any person for loss or damage suffered as a consequence of their placing reliance upon any view, claim or representation made in this presentation. The information and expressions of opinion contained in this presentation are not intended to be a comprehensive study, nor to provide actuarial advice or advice of any nature and should not be treated as a substitute for specific advice concerning individual situations. On no account may any part of this presentation be reproduced without the written permission of the Institute and Faculty of Actuaries .

# Objectives

- What is a hazard ratio and Cox proportional hazards model.
- Describe methods to check the assumption of proportional hazards in the Cox model
- Describe methods how to deal with non-proportional hazards in the Cox model

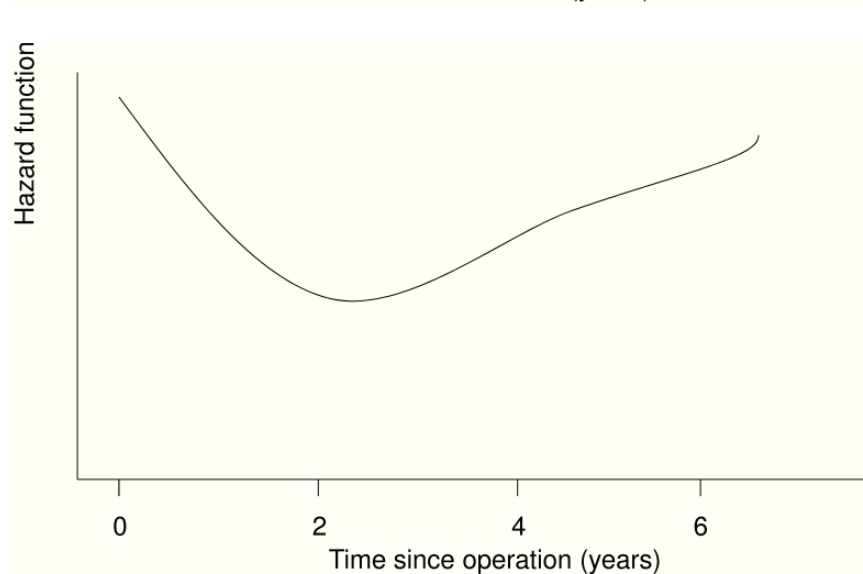
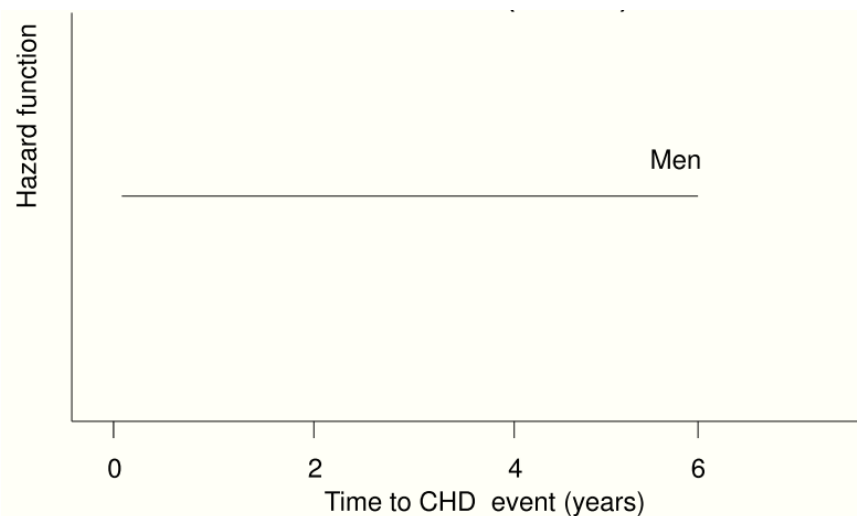


Actuarial  
Research Centre

Institute and Faculty  
of Actuaries

# Hazard aka “force of mortality” and “mortality intensity”

- Hazard is an instantaneous failure rate at time  $t$ 
  - Probability that an individual will experience the event at time  $t$  given that the event has not yet occurred.



# Cox proportional hazards regression

- The type of regression model typically used in survival analysis in medicine is the Cox proportional hazards regression model.
- The Cox model estimates the hazard  $\mu_i(t)$  for subject  $i$  for time  $t$  by multiplying the baseline hazard function  $\mu_0(t)$  by the subject's risk score  $r_i$  as

$$\mu_i(t, \beta, Z_i) = \mu_0(t) r_i(\beta, Z_i) = \mu_0(t) e^{\beta Z_i}$$

- The risk factors  $Z$  have a log-linear contribution to the force of mortality which does not depend on time  $t$ .



# Hazard ratio (HR)

- Taking a ratio of the hazard functions for two subjects  $i$  and  $j$  who differ in one risk factor  $z$  (with the values  $z_0$  and  $z_1$ , respectively) but not in the other risk factors,

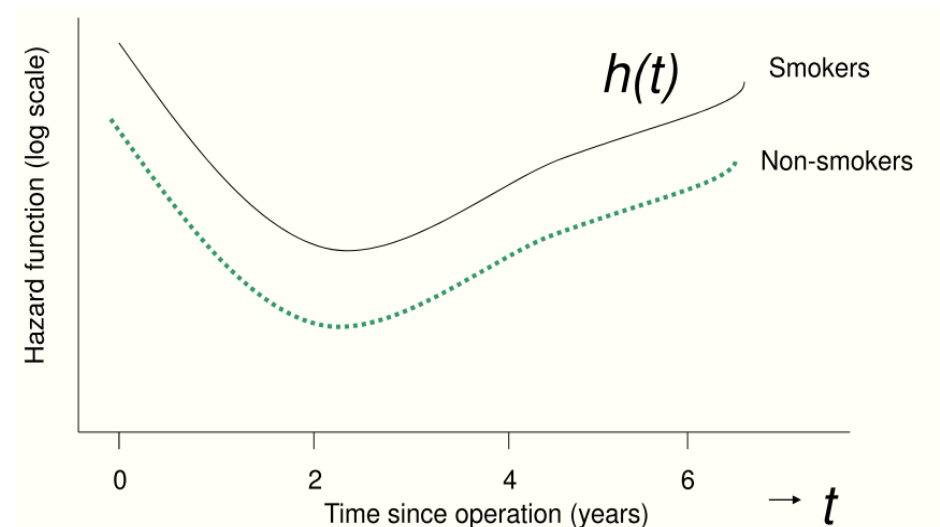
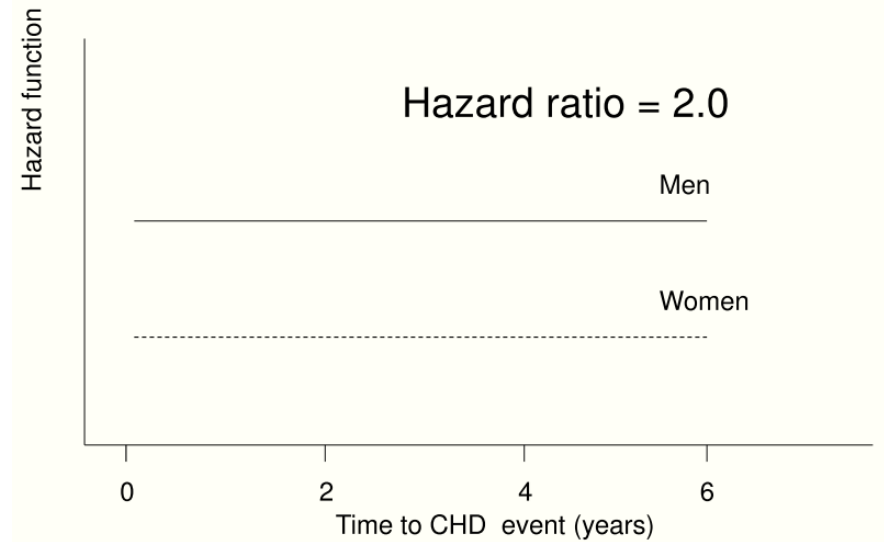
$$\text{HR}(t, \beta, Z) = \frac{\mu_i(t, \beta, Z_i)}{\mu_j(t, \beta, Z_j)} = \frac{\mu_0(t) e^{\beta Z_i}}{\mu_0(t) e^{\beta Z_j}} = \frac{e^{\beta z z_1}}{e^{\beta z z_0}} = e^{\beta z (z_0 - z_1)}.$$

- This means that the baseline hazard  $\mu_0(t)$  does not have to be specified and the hazard ratio  $e^{\beta z (z_0 - z_1)}$  is constant with respect to time  $t$ .
- Because of this, the Cox model does not make any assumptions about the shape of the baseline hazard.
- $e^{\beta z (z_0 - z_1)}$  is an adjusted HR, i.e. **all other risks are already accounted for by the model.**



# Hazard ratio

- Comparison of two hazard functions
- Cox model assumes constant hazard ratio over time





# Proportional hazards assumption

- Graphical methods:
  - Comparison of Kaplan-Meier estimates by group
  - Plot (minus the log cumulative baseline hazard) for each group against (log survival time)
- Formal tests:
  - Grambsch and Therneau's test based on Schoenfeld residuals
  - Include interaction between covariate and a function of time
    - Log(time) often used but could be any function of time



# Example: Cox model for death from Parkinson's disease

- Data: parkison disease
  - Sample of 520 patients
  - Study period of 17 years
- Outcome: time to death (266 events)
- Exposure: new vs standard treatment
- Covariates:

Sex (baseline male / female)

Age (baseline 25-59 / 60-69 / 70-92)

	exp(coef)	exp(-coef)	lower .95	upper .95
treat	1.216	0.8221	0.9549	1.55

Concordance= 0.527 (se = 0.016 )

	exp(coef)	exp(-coef)	lower .95	upper .95
treat	1.216	0.8224	0.9545	1.549
sex	1.031	0.9701	0.8099	1.312

Concordance= 0.522 (se = 0.018 )

	exp(coef)	exp(-coef)	lower .95	upper .95
treat	1.1615	0.8610	0.9101	1.4822
sex	0.7412	1.3491	0.5774	0.9516
agegrp2	3.4363	0.2910	2.3853	4.9504
agegrp3	7.7408	0.1292	5.3363	11.2286

Concordance= 0.706 (se = 0.019 )

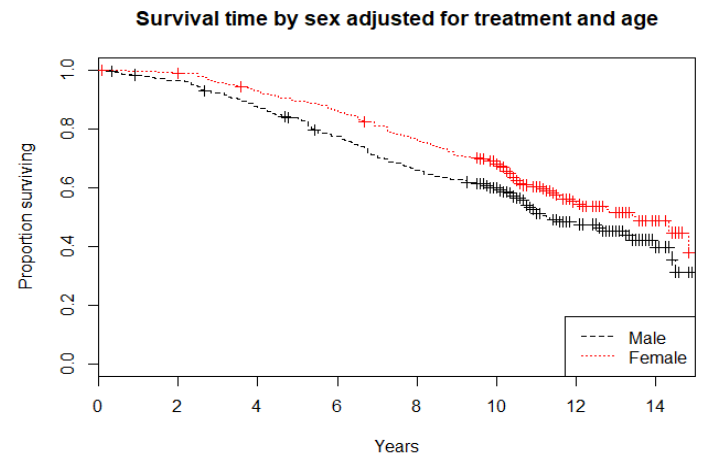
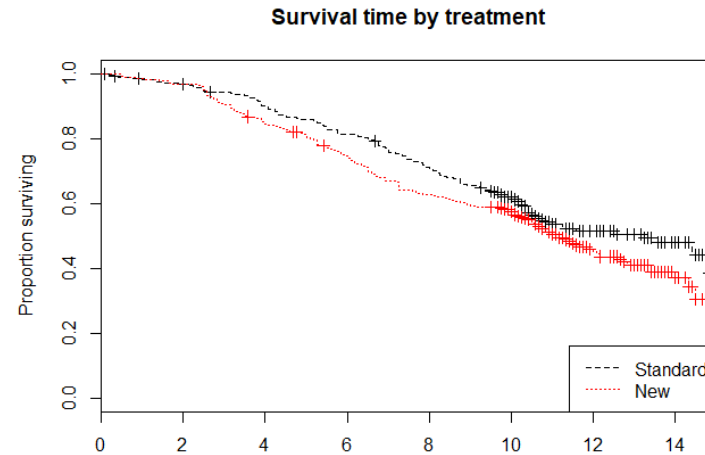


Actuarial  
Research Centre

Institute and Faculty  
of Actuaries

# Kaplan-Meier plots by levels of a factor

- Estimated survival function
  - Does not adjust for other covariates!
  - Crossing of hazard lines indicates non-proportional hazards
  - Otherwise, can be difficult to judge



Actuarial  
Research Centre

Institute and Faculty  
of Actuaries

# Complementary log-log plot of $S(t;Z)$

- From the hazard function of the PH model, we obtain the survivor function

$$S(t; Z) = \exp\{ -M_0(t) e^{\beta Z} \}$$

where  $M_0(t)$  is the cumulative hazard corresponding to  $\mu_0(t)$ .

- Hence  $\ln\{ -\ln S(t; Z) \} = \ln(M_0(t)) + \beta Z$ .
- Hence any two such functions,  $S(t; z_1)$  and  $S(t; z_2)$  for different values of the covariate vector  $z$ , will be parallel.
- Plot  $\ln\{ -\ln S(t; Z) \}$  vs  $t$  or a function of  $t$ .

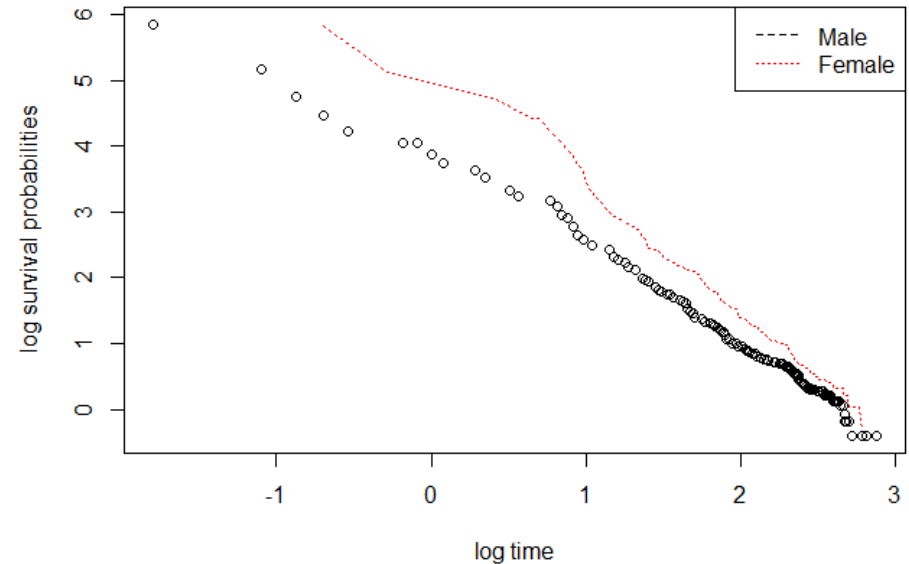


Actuarial  
Research Centre

Institute and Faculty  
of Actuaries

# Complementary log-log plot for Parkinson's data

- Can be unadjusted or adjusted (here adjusted for treatment and age group)
- Proportional hazards assumption violated if curves are not parallel to each other
- Plot vs  $\log(t)$  shows straight lines for Weibull distribution.



This only works if there are few covariates and few distinct values, only then  $S(t;Z)$  is reliably estimated for each  $Z$  value.



Actuarial  
Research Centre

Institute and Faculty  
of Actuaries

# Residuals

- Residual is the difference between an observed value and a predicted value.
- Due to censoring, this is not straightforward in survival analysis
- Therefore, there are many types of residuals
- Here we are going to concentrate on **Cox-Snell residuals** and **Schoenfeld residuals**



Actuarial  
Research Centre

Institute and Faculty  
of Actuaries

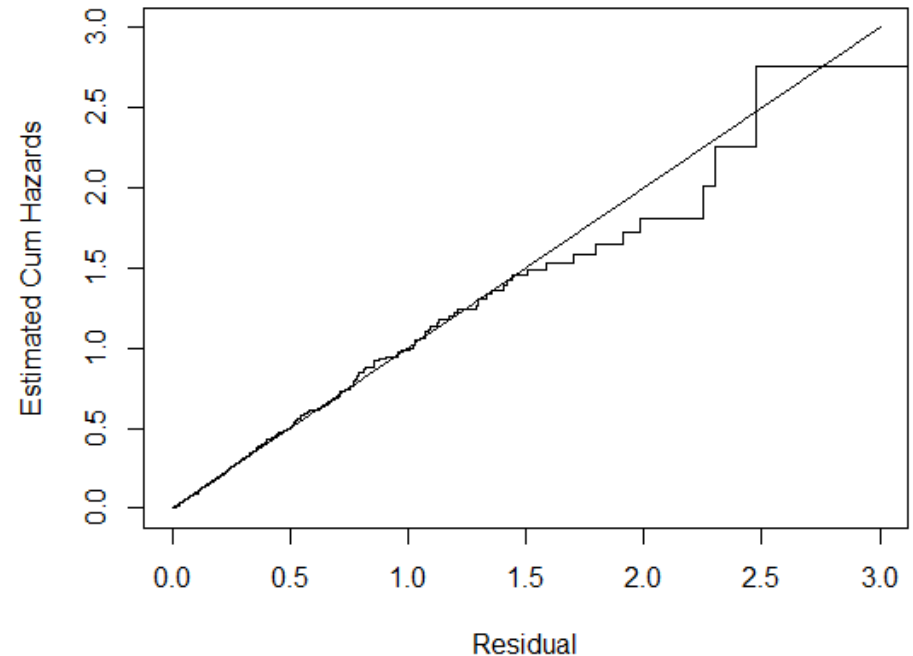
# Cox-Snell residuals

- In order to assess an overall goodness of fit of a model, we use Cox-Snell residuals
- Cox-Snell residuals are  $-\log(\hat{S}(t; Z))$ , i.e. estimated cumulative hazards at the time of death or censoring
- If the model is correct, Cox-Snell residuals should have exponential distribution  $\exp(1)$



# Cox-Snell residuals

- Overall goodness-of-fit
  - The first survival model for Parkinson's data with treatment, sex, and age group. Graph indicates good fit.
- Plot of Cox-Snell residuals is just a QQ-plot for exponential distribution



Actuarial  
Research Centre

Institute and Faculty  
of Actuaries



# Schoenfeld residuals

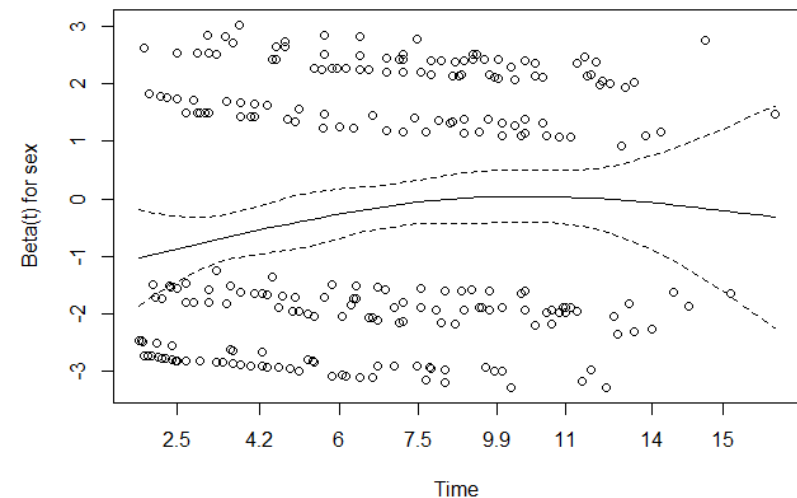
- Schoenfeld residuals are the differences between the covariate value  $Z_i$  of subject  $i$  who experienced an event at time  $t_i$  and the weighted average of all covariate values across all subjects at risk at  $t_i$
- Schoenfeld residuals are used for testing the proportionality of hazards assumption using Grambsch and Therneau's test



# Grambsch and Therneau test

- Testing correlation between Schoenfeld residuals and survival time
- Significant correlation indicates non-proportional hazards

	rho	chisq	p
treat	-0.06626	1.156192	0.2823
sex	0.13732	5.459043	0.0195
agegrp2	0.00114	0.000344	0.9852
agegrp3	-0.07785	1.658215	0.1978
GLOBAL	NA	8.123383	0.0872



Actuarial  
Research Centre

Institute and Faculty  
of Actuaries

# Cox model with time-varying coefficients

$$\mu(t, \beta, Z) = \mu_0(t) e^{\beta(t) Z}$$

Write the time-varying coefficients as

$$\beta_j(t) = \beta_j + \theta_j g_j(t), \quad j=1, \dots, p$$

where  $g_j(t)$  is known. A standard choice is  $g_j(t) = \log(t)$ .

Test  $H_0: \theta=0$  (as a vector and for each component.).



# Testing interaction of covariate with time

- Significant correlation indicates non-proportional hazards
- NB: very sensitive

n= 520, number of events= 266

	coef	exp(coef)	se(coef)	z	Pr(> z )	
treat	4.078e+00	5.905e+01	5.989e-01	6.810	9.76e-12	***
sex	3.109e+00	2.239e+01	4.942e-01	6.290	3.17e-10	***
agegrp2	1.814e+01	7.566e+07	2.184e+00	8.307	< 2e-16	***
agegrp3	1.903e+01	1.843e+08	2.149e+00	8.857	< 2e-16	***
treat:log(time)	-2.054e+00	1.283e-01	2.900e-01	-7.081	1.43e-12	***
sex:log(time)	-1.590e+00	2.040e-01	2.493e-01	-6.377	1.81e-10	***
agegrp2:log(time)	-7.222e+00	7.300e-04	8.989e-01	-8.035	8.88e-16	***
agegrp3:log(time)	-7.597e+00	5.019e-04	8.938e-01	-8.500	< 2e-16	***

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1



Actuarial  
Research Centre

Institute and Faculty  
of Actuaries

# What if the proportional hazards assumption is not met?

- Stratify the analysis on violating variable:  $\mu_s(t, \beta, Z') = \mu_{0s}(t)e^{\beta Z'}$  for  $Z'$  being all covariates but that one.
  - Fit one model: allow baseline hazards to vary by group but assume covariate effects are the same across strata. **Only if the variable is of no direct interest.** (There should be no significant interactions between covariates and stratum variable)
  - Fit separate models: allow both baseline hazards and hazard ratios to vary by group



# What if the proportional hazards assumption is not met?

- Include time-dependent effect
  - Split follow-up time such that the hazards are proportional within these time bands
  - Continuous (could be any function of time)



**Actuarial  
Research Centre**

Institute and Faculty  
of Actuaries

# Stratified analysis

- Check for interactions
- Fit one stratified Cox model  
(n=520, events=266)
- Fit separate models
  - Male (n=283, events=141)
  - Female (n=237, events=125)
- Easy procedure but comes at the cost of no estimate for the effect of the violated variable associated with the outcome

	exp(coef)	exp(-coef)	lower .95	upper .95
treat	1.0395	0.9620	0.4877	2.216
sex	0.7895	1.2666	0.2964	2.103
agegrp2	2.1979	0.4550	0.7261	6.653
agegrp3	23.7526	0.0421	7.8517	71.855
treat:sex	1.0938	0.9142	0.6695	1.787
sex:agegrp2	1.3487	0.7415	0.6348	2.865
sex:agegrp3	0.4752	2.1045	0.2243	1.007

	exp(coef)	exp(-coef)	lower .95	upper .95
treat	1.162	0.8605	0.9099	1.484
agegrp2	3.392	0.2948	2.3549	4.887
agegrp3	7.389	0.1353	5.0865	10.735

	exp(coef)	exp(-coef)	lower .95	upper .95
treat	1.126	0.88817	0.8067	1.571
agegrp2	2.918	0.34272	1.8230	4.670
agegrp3	10.540	0.09488	6.5101	17.063

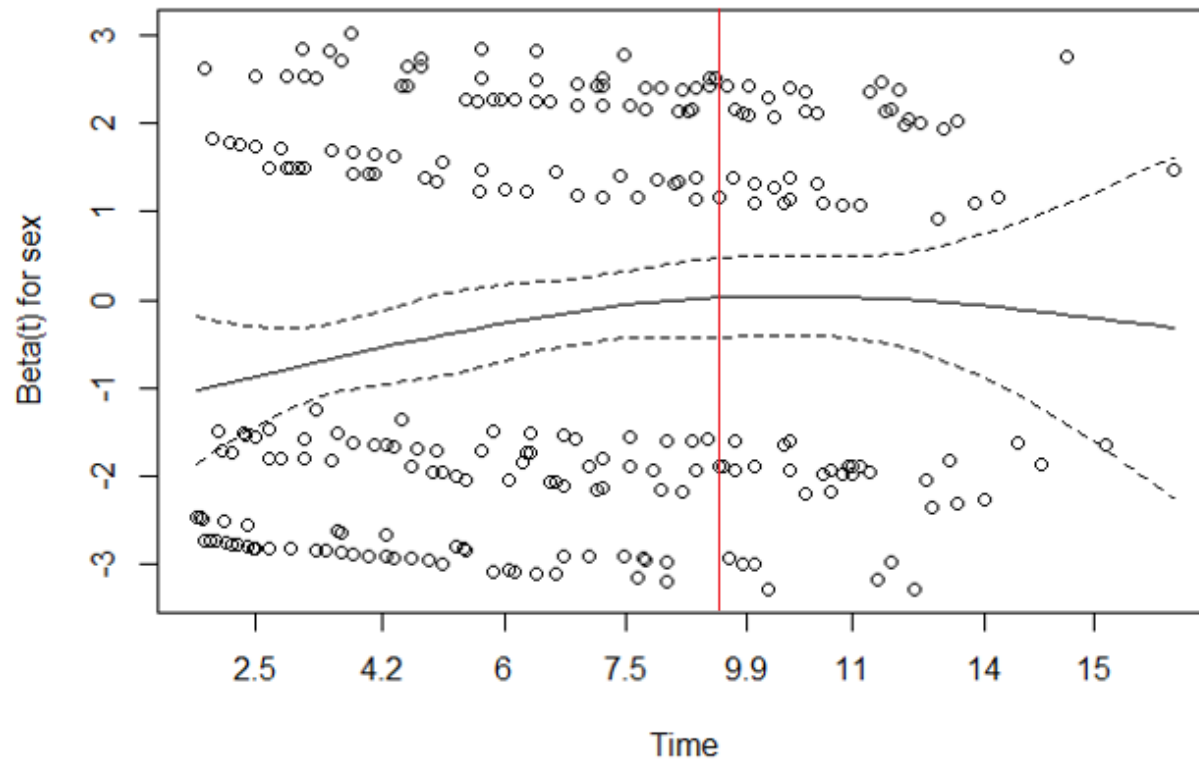
	exp(coef)	exp(-coef)	lower .95	upper .95
treat	1.258	0.7950	0.8737	1.811
agegrp2	4.005	0.2497	2.2151	7.242
agegrp3	5.335	0.1875	2.9578	9.622



Research Centre

Institute and Faculty  
of Actuaries

# Schoenfeld Residuals plot of effect of sex over time



Actuarial  
Research Centre

Institute and Faculty  
of Actuaries



# Step-wise time-dependent hazards

- Split follow-up time in intervals in which the proportional hazards assumption is no longer violated
- Create time dependent effect
  - Here: 0 = male's hazard (baseline),  
1=female's hazard 0-9 years,  
2=female's hazard 9+ years
- Fit model with time dependent effect
- More time consuming procedure due to creating the most effective time intervals

	exp(coef)	exp(-coef)	lower .95	upper .95
treat	1.1636	0.8594	0.9117	1.4852
t_sex1	0.6887	1.4520	0.5162	0.9188
t_sex2	0.9118	1.0968	0.5678	1.4641
agegrp2	3.4215	0.2923	2.3750	4.9291
agegrp3	7.5985	0.1316	5.2343	11.0305

	rho	chisq	p
treat	-0.06453	1.101428	0.2940
t_sex1	0.09955	2.762015	0.0965
t_sex2	0.02402	0.158846	0.6902
agegrp2	0.00158	0.000657	0.9795
agegrp3	-0.07938	1.708571	0.1912
GLOBAL	NA	6.446771	0.2651



Actuarial  
Research Centre

Institute and Faculty  
of Actuaries

# References

- Therneau, T.M. and Grambsch, P.M., 2013. *Modeling survival data: extending the Cox model*. Springer Science & Business Media.
- Martinussen T, Scheike TH. *Dynamic Regression Models for Survival Data*. Springer: New York, 2006.





**Actuarial  
Research Centre**

Institute and Faculty  
of Actuaries

# The Actuarial Research Centre (ARC)

A gateway to global actuarial research

The Actuarial Research Centre (ARC) is the Institute and Faculty of Actuaries' (IFoA) network of actuarial researchers around the world.

The ARC seeks to deliver cutting-edge research programmes that address some of the significant, global challenges in actuarial science, through a partnership of the actuarial profession, the academic community and practitioners.

The **'Use of Big Health and Actuarial Data for understanding Longevity and Morbidity Risks'** research programme is being funded by the ARC.

# Questions

# Comments

The views expressed in this presentation are those of the presenter.



**Actuarial  
Research Centre**

Institute and Faculty  
of Actuaries



**Actuarial  
Research Centre**

Institute and Faculty  
of Actuaries

**UEA**  
University of East Anglia

# Linking survival modelling results to life expectancy differentials

Elena Kulinskaya  
School of Computing Sciences,  
University of East Anglia

The **'Use of Big Health and Actuarial Data for understanding Longevity and Morbidity Risks'** research programme is being funded by the Actuarial Research Centre.

30/10/2019

[www.actuaries.org.uk/arc](http://www.actuaries.org.uk/arc)



## Actuarial Research Centre

Institute and Faculty  
of Actuaries

### Disclaimer

The views expressed in this presentation are those of invited contributors and not necessarily those of the Institute and Faculty of Actuaries. The Institute and Faculty of Actuaries does not endorse any of the views stated, nor any claims or representations made in this presentation and accept no responsibility or liability to any person for loss or damage suffered as a consequence of their placing reliance upon any view, claim or representation made in this presentation. The information and expressions of opinion contained in this presentation are not intended to be a comprehensive study, nor to provide actuarial advice or advice of any nature and should not be treated as a substitute for specific advice concerning individual situations. On no account may any part of this presentation be reproduced without the written permission of the Institute and Faculty of Actuaries .

# Quantifying Longevity Changes

- Medical and social advances are the major drivers in the longevity increase. But how to quantify this relationship?
- In medicine, Randomized Control Trials (RCTs) are considered to be the gold standard.
- RCTs estimate the hazard or force of mortality in a (selective) sample of people and summarised over the observed (limited) time period.
- New health interventions are usually based on these estimated hazards obtained from clinical trials. A lengthy lead time would be needed to observe their effect on population longevity.



# Our approach, 1

- Our research uses The Health Improvement Network (THIN) primary care data to develop statistical models of longevity.
- The advantage of using individual-level medical data is that it is possible to model both the uptake of medical treatment and the effect of that treatment on longevity conditional on the individual sociodemographic and health factors instead of the aggregated profile.
- We carefully design each observational study and match cases to controls. Survival models, usually the Cox regression, are fitted to such individual level data.
- The conclusions are generalisable to the general population.





# Hazard ratio

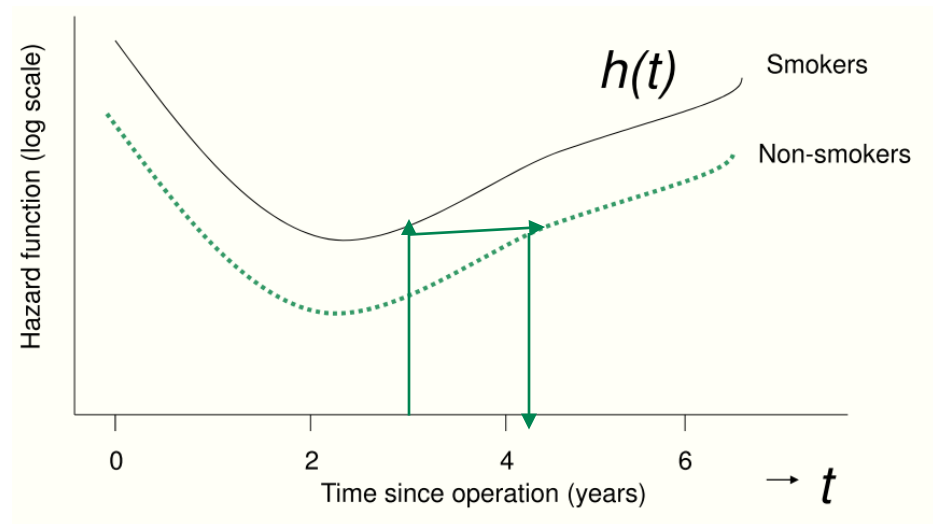
- The type of regression model typically used in survival analysis in medicine is the Cox proportional hazards regression model.
- The Cox model estimates the hazard  $\mu_i(x)$  for subject  $i$  at time  $x$  as  $\mu_i(x, \beta, Z_i) = \mu_0(x) r_i(\beta, Z_i) = \mu_0(x) e^{\beta Z_i}$
- Taking a ratio of the hazard functions for two subjects  $i$  and  $j$  who differ in one risk factor  $z$  and not in the other risk factors,

$$\mu(x, \beta, Z) = \frac{\mu_i(x, \beta, Z_i)}{\mu_j(x, \beta, Z_j)} = \frac{\mu_0(x) e^{\beta Z_1}}{\mu_0(x) e^{\beta Z_0}} = \frac{e^{\beta z z_1}}{e^{\beta z z_0}} = e^{\beta z (z_0 - z_1)}$$



# From a hazard to effective age

- For simplicity, consider a binary risk factor with the reference value  $y = 0$  and (at risk)  $y = 1$ .
- On the log scale, the log-hazards are  $\lambda_1(t) = \lambda_0(t) + \beta$ . This means that the log-hazard lines differ only by an increment
- For a monotone-increasing hazard, find the (unique) time increment  $\Delta(t)$  such that  $\lambda_1(t) = \lambda_0(t + \Delta(t))$



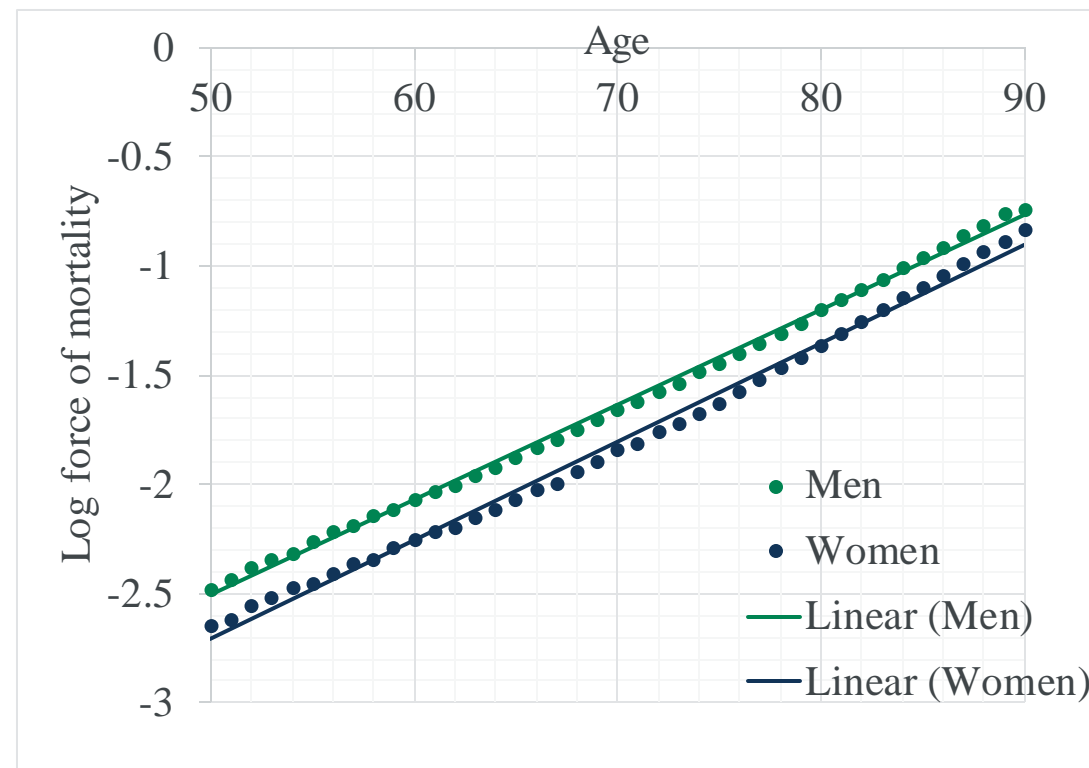
Value of  $t + \Delta(t)$  is, by definition, the **effective age** of the person with risk  $y = 1$  at chronological age  $t$ .



# Our approach, 2: for an individual

- For an individual, the hazard ratios obtained from the survival models are translated into effective age changes.
- Effective age at  $y=1$  is the average chronological age with the same hazard as when  $y=0$ .
- Effective ages are often used by insurers as a way of applying the correct rating to an underwritten life.

Log force of mortality for UK population based on 2010 period life table (Office for National Statistics 2017).



# What does HR mean for an individual

- Using Gompertz law  $\lambda_0(t)=a+bt$ , the increase in annual hazard of mortality associated with ageing one year is approximately constant between ages 50 and 90.

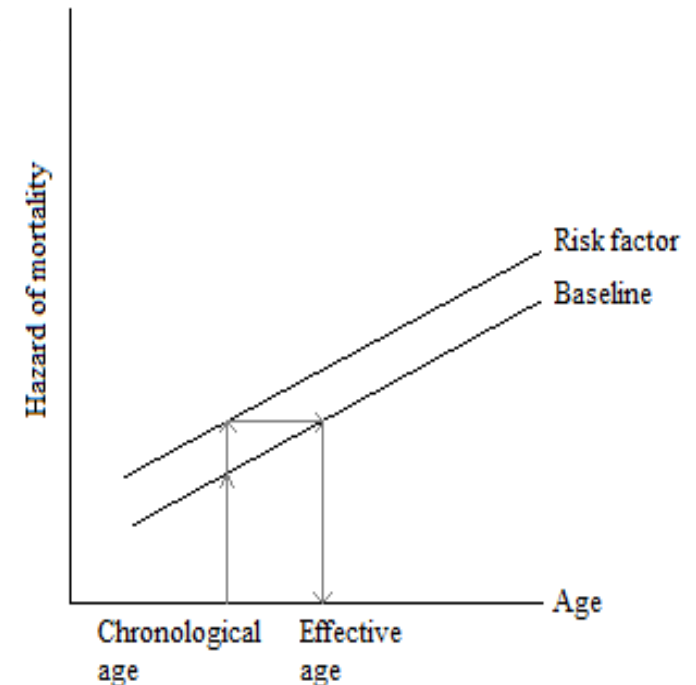
For  $y=1$ ,  $\lambda_1(t) = a+bt+ \beta=a+b(t+\Delta) \rightarrow \Delta=\beta/b$

- For England and Wales in 2010-2012, the increase in the hazard between those ages was approximately 1.1 per year.
- A HR can be translated to the numbers of years gained in effective age as

$$\Delta = \log(\text{HR}) / \log(1.1) \approx 10 * \log(\text{HR}).$$

[Brenner, 1993; Spiegelhalter, 2016]

- For LE at age  $t$ ,  $e_1(t) = e_0(t + \Delta)$ .



# Our approach, 3: Period life expectancy

- Consider a population consisting of  $J$  risk groups of prevalence  $q_j$ ,  $j=1, \dots, J$ ;  $\sum q_j = 1$ .
- Treatment of interest ( $i=0, 1$ ) is prevalent in the population from age  $T$  but its effects vary across risk groups.
- Prevalence of the treatment of interest in group  $j$  at age  $T$  is  $p_{j,1}$
- Then the overall survival function  $S(T)$  at age  $x=T$  is the weighted mean of the survival functions in the individual risk groups

$$S(T) = \left[ \sum_j q_j p_{j,1} S_{1j}(T) + \sum_j q_j (1 - p_{j,1}) S_{0j}(T) \right] / \sum_j q_j.$$



# The Cox model

- Assume that the hazards are proportional, so that the hazards

$\mu_{ij} = \mu_0(x)\mu_{ij}(Y)$ , where  $\mu_0(x)$  is the baseline hazard at age  $x$

and  $\log(\mu_{ij}(Y)) = a_{ij} = a_0(T) + \alpha_i + \beta_j + \gamma_{ij} + \beta^T Y$ ,

where  $a_0(T)$  is the baseline value which may depend on intervention time  $T$ ,  $\alpha_i$ ,  $\beta_j$  and  $\gamma_{ij}$  are the main effects and interaction of risk group  $j$  and treatment  $I$ , and the other covariates  $Y$  have no interactions with the treatment or the risk of interest.



# Survival function under Gompertz-Cox model

- The log-hazards in a risk group (i,j) are just  $a_{ij} + bx$ , i.e. the straight lines with the same slopes but differing intercepts.

- The survival functions are  $S_{ij}(x) = \exp(-e^{a_{ij} + bx} b^{-1})$ .  
Substituting the  $a_{ij}$ , the survival functions at age  $x > T$  are

$$S_{ij}(x|Y) = \exp(-e^{a_0(T) + \alpha_i + \beta_j + \gamma_{ij} + \beta^T Y} b^{-1} (1 - e^{bx})).$$

- Assuming that the prevalences do not depend on  $Y$ ,  $Y$  can be integrated out to obtain

$$S(x|T) = \sum_j q_j p_{j,1} \exp(-e^{a_0(T) + \alpha_1 + \beta_j + \gamma_{1j}} b^{-1} (1 - e^{bx})) + \sum_j q_j (1 - p_{j,1}) \exp(-e^{a_0(T) + \beta_j} b^{-1} (1 - e^{bx})).$$



# Finding component survival functions

$$S(x|T) = \sum_j q_j p_{j,1} \exp(-e^{a_0(T) + \alpha_1 + \beta_j + \gamma_{1j}} b^{-1} (1 - e^{bx})) + \sum_j q_j (1 - p_{j,1}) \exp(-e^{a_0(T) + \beta_j} b^{-1} (1 - e^{bx})).$$

- This is a non-linear equation with one unknown,  $a_0$ . The left-hand side is given by the period life-table, and the slope  $b$  should be determined for a particular population of interest. As  $S(x)$  is a decreasing function of  $a_0$ , it has a unique solution.
- After solving for  $a_0(T)$ , we can find component survival functions  $S_{ij}(x)$  for any set of prevalences  $\{q_j\}$  and  $\{p_{j,1}\}$ .





# Estimating changes in life expectancy

- For a Gompertz distribution  $G(a,b)$ , the LE is

$$e_{G(a,b)}(z) = \frac{b^{-1} \exp(b^{-1}e^a) E_1(b^{-1}e^{a+bx})}{\exp(-e^a b^{-1}(e^{bx} - 1))}.$$

So we can find component LEs  $e_{ij}(z)$  for each component distribution  $G(a_{ij},b)$ . Then, the life expectancy at age  $z$  is

$$e(z) = \frac{\int_z^\infty S(x)dx}{S(z)} = \frac{\sum_k w_k S_k(z) \int_z^\infty S_k(x)dx / S_k(z)}{\sum w_k S_k(z)} = \frac{\sum_k w_k S_k(z) e_k(z)}{\sum w_k S_k(z)}.$$

- Taking all  $p_{j,1} = 0$ , we obtain a hypothetical life expectancy  $e_0(z)$  if there were no intervention of interest, and, for all  $p_{j,1} = 1$ , a hypothetical life expectancy  $e_1(z)$  with full uptake of the intervention.



# Case study: survival benefits of statins

- We used the data for QRISK2 groups 10–19% and  $\geq 20\%$  at ages 70 and 75 (yob 1920-1940, observed 1987-2011) from Gitsels et al. [2016]
- We fitted the same Cox models after adding the QRISK2 group to the predictors. The final models adjusted for sex, birth cohort, socioeconomic status, diabetes, hyper-cholesterolaemia, blood pressure regulating drugs, body mass index, and smoking status. The models included a random effect on general practice.
- Interactions between statins, QRISK2 groups and the other risk factors were tested, but none was significant.
- We also used the adjusted HRs for all-cause mortality of heart attack survivors, from Gitsels et al. [2017], as a substitute for HRs for CVD sufferers. The HRs are given in Table 1.

# Table 1. Population characteristics, statins study by Gitsels et al. 2016

Cohort	Cardiac risk	Women % (Statins %)	Men % (Statins %)
Age 70	QRISK2 10-19%	80 (9.5)	17 (5.4)
N=247,149, FU=7 years	QRISK2 >20%	20 (28.2)	83 (17.4)
Age 75	QRISK2 10-19%	15 (4.6)	0 (0.0)
N=194,085, FU=6 years	QRISK2 >20%	85 (19.6)	100 (19.1)

\* FU= average follow-up



**Actuarial  
Research Centre**

Institute and Faculty  
of Actuaries

# Table 2. Hazard ratios of statins and of cardiac risk groups

Cardiac risk	Age	HR statins (vs no statins)	Changes in effective age (men) <sup>1</sup>	Changes in effective age (women) <sup>1</sup>
No heart attack	70	0.84(0.80, 0.88)	-1.69 (-2.16,-1.24)	-1.57 (-2.01, -1.15)
	75	0.82 (0.79, 0.86)	-1.92 (-3.18,-1.46)	-1.79 (-2.97, -1.36)
Heart attack	70	0.74 (0.70, 0.78)	-2.91 (-3.45, -2.40)	-2.72 (-3.22, -2.24)
	75	0.77 (0.74, 0.81)	-2.53 (-2.91, -2.04)	-2.36 (-2.72, -1.90)
Cardiac risk	Age	HR Cardiac Risk	Changes in effective age (men)	Changes in effective age (women)
QRISK2 10-19%	70	0.80 (0.77, 0.83)	-2.16 (-2.53, -1.80)	-2.01 (-2.36, -1.68)
	75	0.87 (0.80, 0.94)	-1.35 (-2.16, -0.60)	-1.26 (-2.01, -0.56)
QRISK2>20%	70	1	0	0
	75	1	0	0
Heart attack	70	1.50 (1.42,1.59)	3.92 (3.39, 4.48)	3.66 (3.17, 4.19)
	75	1.45 (1.38,1.53)	3.59 (3.11, 4.11)	3.35 (2.91, 3.84)

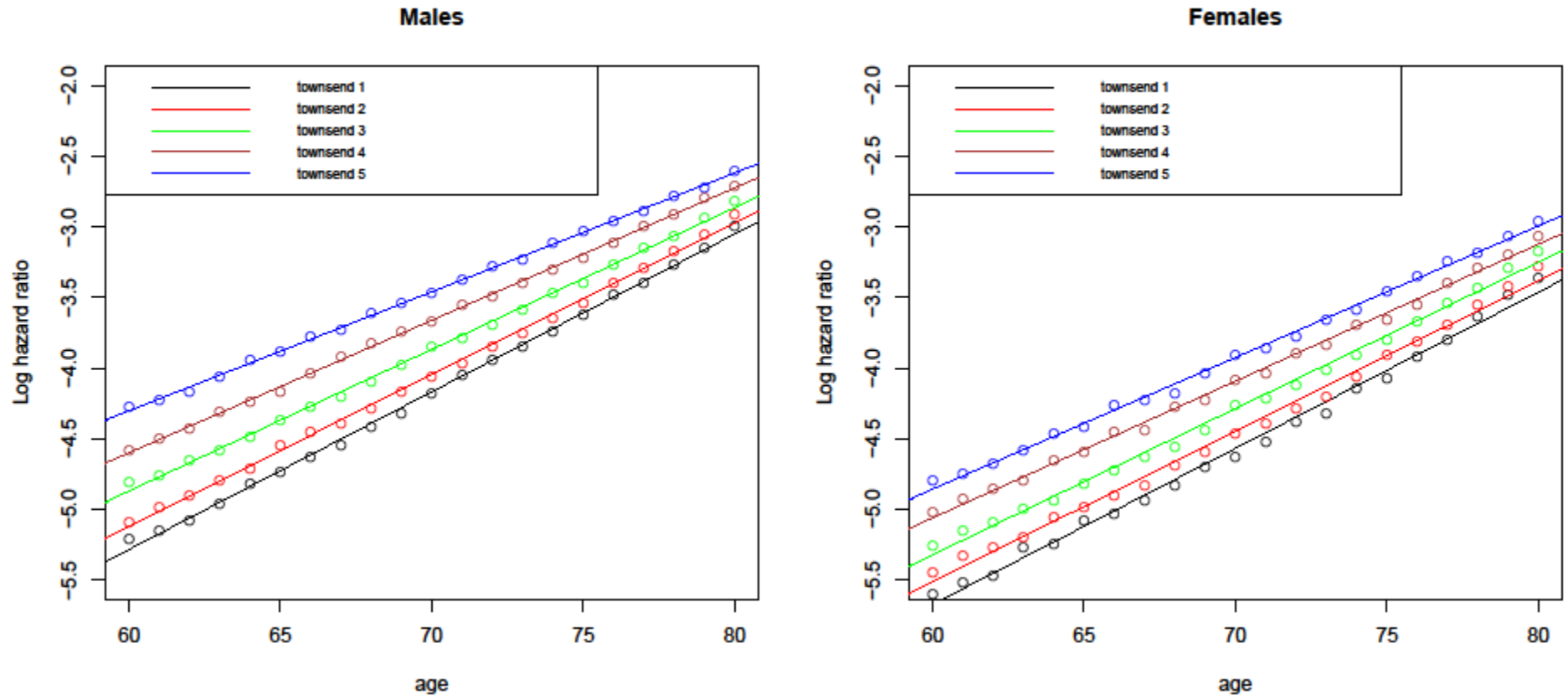
<sup>1</sup> based on Gompertz distributions with  $b=0.1034$  for men and  $0.1108$  for women.

# Info on the prevalences of the risk groups and the treatment

- Prevalence of risk groups (q's on slide 9): QRISK2 score increases with age and by age 70, there were practically no patients with a QRISK2 score of  $< 10\%$  and by age 75, there were no male patients with a QRISK2 score of  $< 20\%$ .
- Prevalence of treatment(p's on slide 9): At the end of study period in 2010, statins were prescribed in 20% of patients with a QRISK2 score of  $< 20\%$ , in 45% of patients with a QRISK2 score of  $\geq 20\%$ , and in 90% of patients with CVD.
- Given cardiac risk group, statins were prescribed more in women, in younger patients, and in patients from less deprived areas.



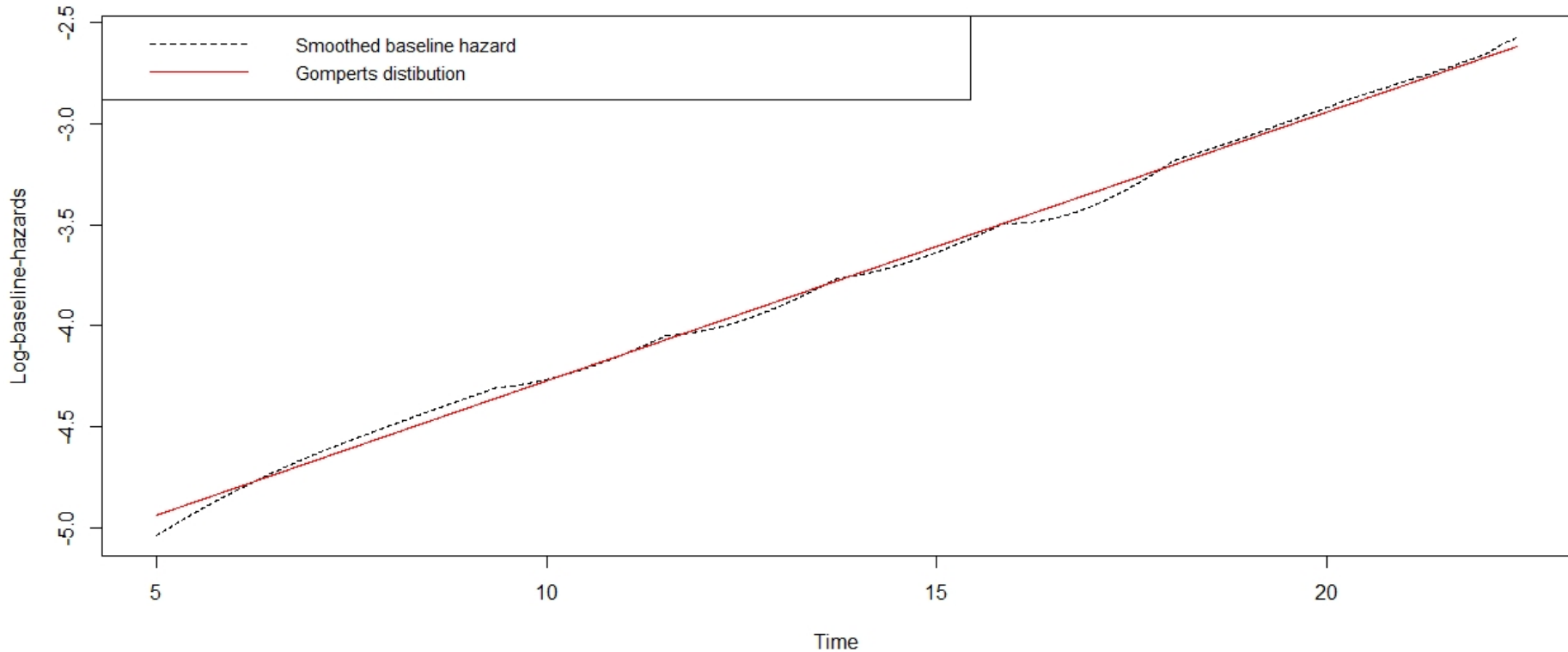
# Fitting Gompertz distribution to period life tables



Log-hazards between the ages 70- 90 from the ONS period life table centered at 2010 (circles) and fitted regression lines by Townsend score quintiles and sex.



# Baseline hazard in the statins survival model



The baseline hazard is well approximated by the Gompertz hazard.



Actuarial  
Research Centre  
Institute and Faculty  
of Actuaries

# Calculating component life expectancies

- Since the mortality rates, the cardiac risk distribution and the statin prescription rates differ by gender and by socio-economic status, we analysed the life tables separately for each Townsend score quintile- by-gender combination.
- For each life table, we substitute the  $S(x)$  at age  $x=70$  or  $75$  (obtained from the fitted Gompertz distribution  $G(a,b)$ ) into the left-hand side of the equation on top of slide 13, and solve for the value of  $a_0(T)$ .
- These values were used to calculate period life tables for component cardiac risk by statin prescription subpopulations for each  $(i,j)$  combination.



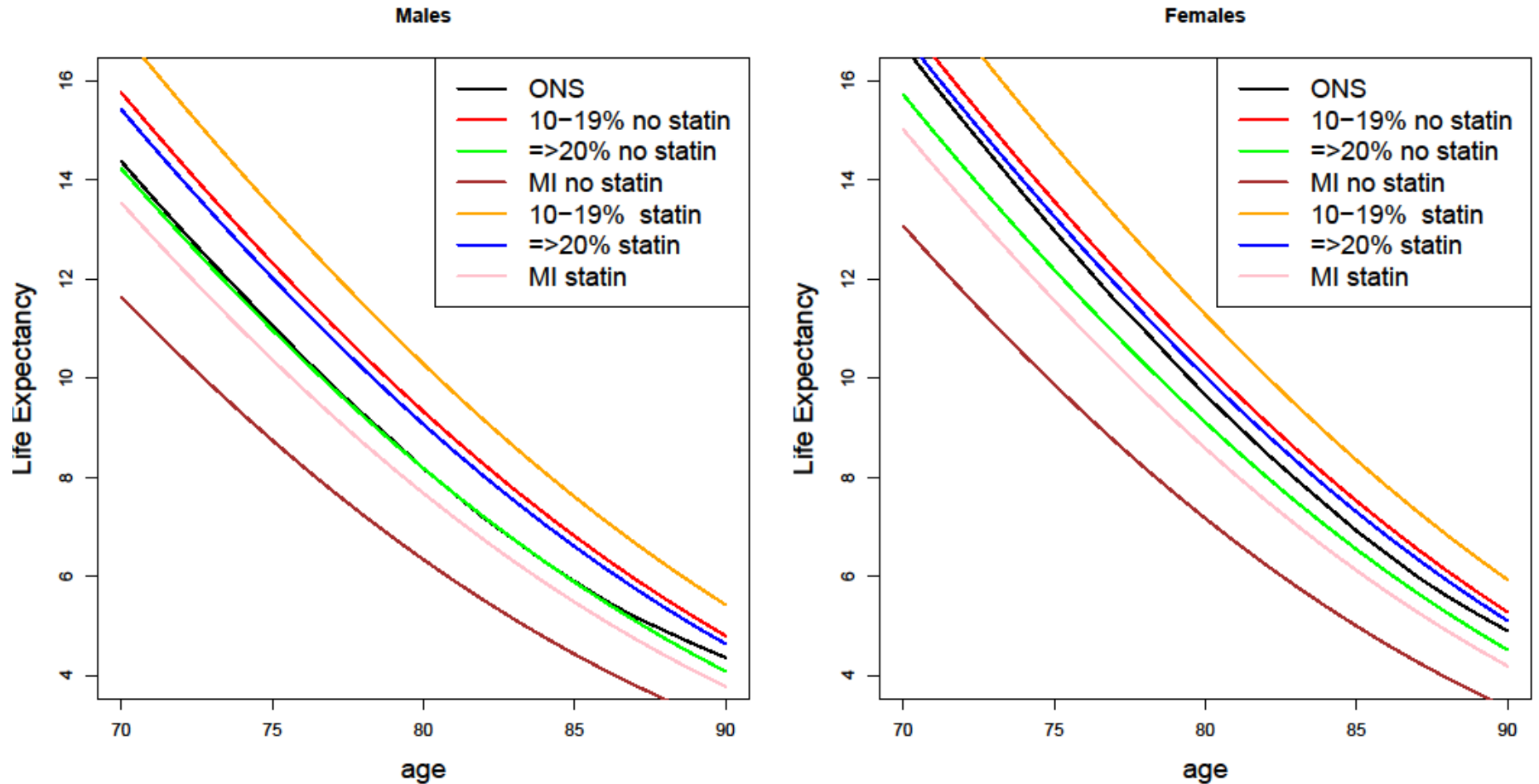


# Results, 1

- Increase in individual LE due to statins depends on cardiac risk, and is highest for heart attack survivors (1.41–2.02 years), and is comparable in the two QRISK2 groups (1.14–1.35 years across ages 70 and 75). The effect of statins increases with deprivation.
- We also calculated the period LE and its increase due to statins in each cardiac risk group for the total England and Wales population by averaging the LE across all TS quintiles, and plotted the results.



# Life expectancy by cardiac risk group with and without statins for ages 70 – 90 based on the ONS period life table centered at 2010



# Results, 2

- We also calculated national life expectancy with and without statins, by averaging the LE across cardiac risk groups, taking  $p = 0$  (for no statins) and  $p = 1$  (for statins).
- The national life expectancy for women aged 70 or 75 would be increased by up to 0.91 or 0.79 years, respectively, if all eligible women under the current guideline of primary and secondary prevention of CVD were prescribed statins.
- Similarly, the national life expectancy for men aged 70 or 75 would be increased by up to 0.79 or 0.63 years. The most improvement would come from the areas of medium deprivation.



# An app for general public

Information about you

Age (60-85)

Sex  Male  Female

Postcode  ?

Smoking status

Weight (kg)

Height (m)

**Health profile**

Diabetes

Chronic kidney disease

Cardiovascular disease


Systolic blood pressure  ?

On blood pressure treatment?

Hypertension

High cholesterol level

**Calculate**



Your body mass index is

Your risk of having a heart attack or stroke within the next 10 years is  %

Your life expectancy is  years without statin

years with statin



Actuarial  
Research Centre

Institute and Faculty  
of Actuaries

# References:

1. Brenner H, Gefeller O, Greenland S. (1993) Risk and rate advancement periods as measures of exposure impact on the occurrence of chronic diseases. *Epidemiol Camb Mass.* 4(3):229–36.
2. Gitsels L.A., Kulinskaya E., Steel N. (2016) Survival Benefits of Statins for Primary Prevention: A Cohort Study. *PLoS ONE* 11(11): e0166847. doi:10.1371/journal.pone.0166847
3. Gitsels LA, Kulinskaya E, Steel N Survival prospects after acute myocardial infarction in the UK: a matched cohort study 1987–2011 *BMJ Open* 2017;7:e013570. doi:10.1136/bmjopen-2016-013570.
4. Kulinskaya E., Gitsels, LA. and Bakbergenuly, I., 2019. Calculation of changes in individual and period life expectancy based on proportional hazards model of an intervention. *Insurance Mathematics and Economics*, under review.
5. Spiegelhalter (2016) How old are you, really? Communicating chronic risk through ‘effective age’ of your body and organs. *BMC Medical Informatics and Decision Making*, 16:104



Questions

Comments



**Actuarial  
Research Centre**

Institute and Faculty  
of Actuaries



**Actuarial  
Research Centre**

Institute and Faculty  
of Actuaries

**UEA**  
University of East Anglia

# Modelling non-proportional hazards: time-dependent coefficients, parametric “double Cox” regression and Landmark analysis

Elena Kulinskaya  
School of Computing Sciences,  
University of East Anglia

The ‘**Use of Big Health and Actuarial Data for understanding Longevity and Morbidity Risks**’ research programme is being funded by the Actuarial Research Centre.

30/10/2019

[www.actuaries.org.uk/arc](http://www.actuaries.org.uk/arc)



## Actuarial Research Centre

Institute and Faculty  
of Actuaries

### Disclaimer

The views expressed in this presentation are those of invited contributors and not necessarily those of the Institute and Faculty of Actuaries. The Institute and Faculty of Actuaries does not endorse any of the views stated, nor any claims or representations made in this presentation and accept no responsibility or liability to any person for loss or damage suffered as a consequence of their placing reliance upon any view, claim or representation made in this presentation. The information and expressions of opinion contained in this presentation are not intended to be a comprehensive study, nor to provide actuarial advice or advice of any nature and should not be treated as a substitute for specific advice concerning individual situations. On no account may any part of this presentation be reproduced without the written permission of the Institute and Faculty of Actuaries .



# What if the proportional hazards assumption is not met?

- For a Cox model  $\mu(t|\beta, Z) = \mu_0(t)\exp(Z^T\beta)$  we discussed two ways to cope with non-proportionality:
- Stratify the analysis on violating variable:  $\mu_s(t|\beta, Z) = \mu_{0s}(t)e^{Z^T\beta}$ 
  - baseline hazards vary by strata  $s$ ;
  - Here we add an option of modelling shape of baseline hazards
- Include time-varying effects:  $\mu(t, |\beta, Z) = \mu_0(t)e^{Z^T\beta(t)}$ 
  - Coefficients  $\beta(t)$  are continuous functions of time
  - Use landmark analysis



# Parametric “Double-Cox” regression

Components:

- A baseline hazard function (which **changes over time**).
- The risk factors  $Z$  have a log-linear contribution to the force of mortality which does not depend on time  $t$ .

The Cox parametric regression model

$$\mu(t|Z) = \mu_0(t|Z) \exp(Z^T \beta)$$

Baseline hazard function

$\beta$  is a vector of unknown parameters for scale and  $Z$  is a vector of covariates

Weibull or Gompertz baseline hazard function with scale  $\lambda$  and shape  $k$ . Shape  $k$  is modelled as  $k=k(Z)$ .

Additional regression model to allow varying shape depending on covariates

$$\mu_0(t|Z) = \frac{k(Z)}{\lambda} \left(\frac{t}{\lambda}\right)^{k(Z)-1}$$

$$\mu_0(t|Z) = \lambda \exp(k(Z)t)$$

$$k(Z) = k_0 e^{Z^T \beta_k}$$



Actuarial Research Centre

Institute and Faculty of Actuaries

# Cox model with shared frailty

**Proportional hazards model with frailty:**

$$\mu(t|U, Z) = \mu_0(t)Ue^{Z^T\beta},$$

For mathematical convenience, it is frequently assumed that frailty  $U$  is gamma-distributed with mean 1 and unknown variance  $\sigma^2$ :

$$U \sim \text{Gamma}(\sigma^{-2}, \sigma^{-2}).$$

The frailty variance  $\sigma^2$  characterizes heterogeneity in the population.

**Shared frailty assumption:**

All patients from the same **unit** /clients from the same company are in the same cluster  $j$ ,  $j=1, \dots, J$  and share the same frailty  $U_j$ .



# “Double-Cox” model with shared frailty

- Standard shared frailty Cox model :  $\mu(t|U, Z) = \mu_0(t)Ue^{Z^T\beta}$ ;
- Baseline hazard  $\mu_0(t)=\mu_0(t; \lambda, k)$ ;
- Cox-like parameterization for the shape of the baseline hazard function:  
 $k(Z)=k_0e^{Z^T\beta_k}$ ;
- Frailty  $U \sim \text{Gamma}$  with mean 1 and variance  $\sigma^2$ .
- If needed, competing risks can be introduced through correlated shared frailty components.

Find MLE of the vector of unknown parameters  $\theta=(\lambda, k_0, \sigma^2, \beta, \beta_k)$ .

This model was introduced in [1] for analysis of time to revision/  
time to death after hip replacement.

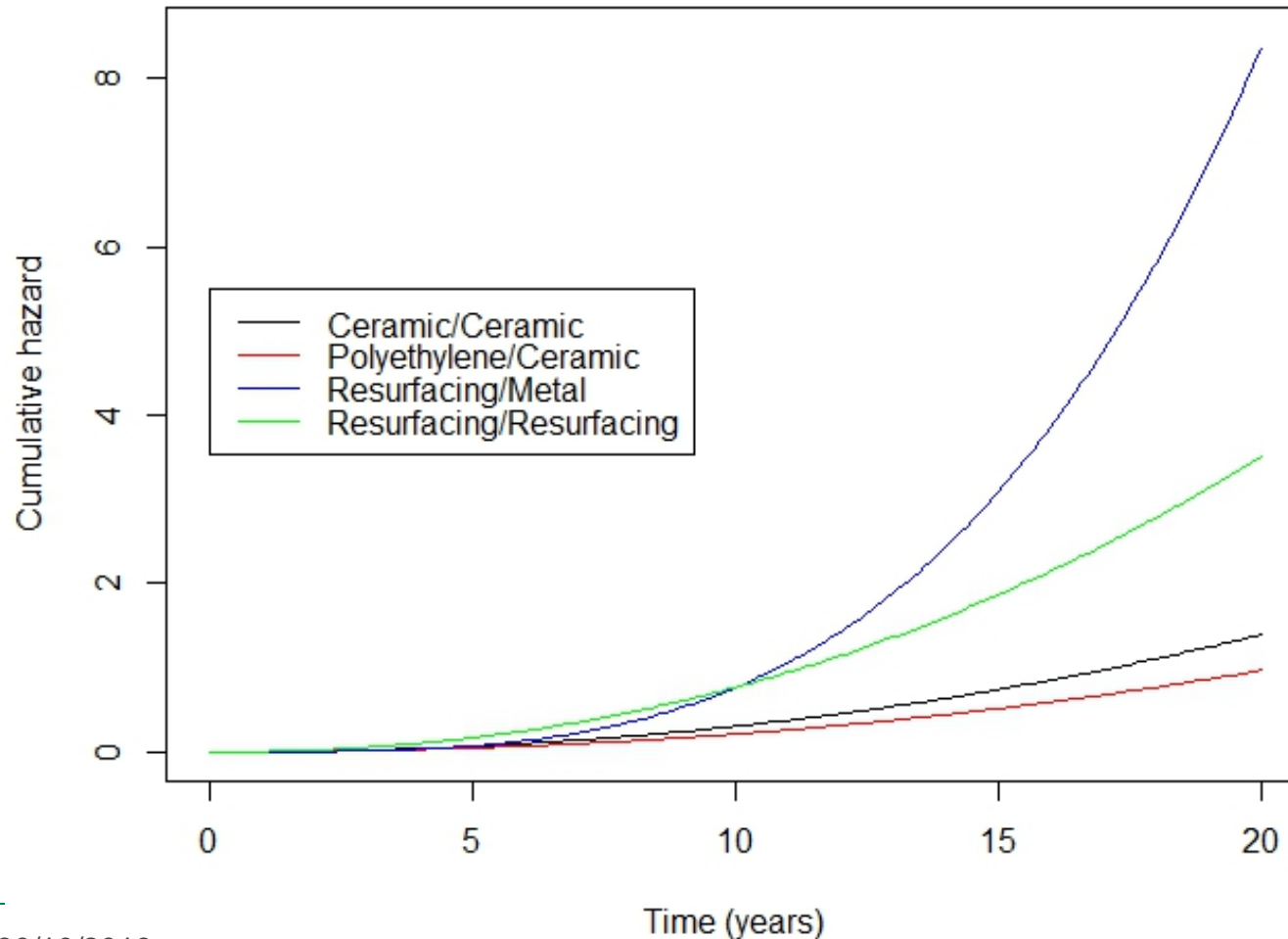


Actuarial  
Research Centre

Institute and Faculty  
of Actuaries

# Different shapes of cumulative hazards for revision surgery after hip replacement

Cumulative hazard function by type of bearing



Actuarial  
Research Centre

Institute and Faculty  
of Actuaries

# Extended Cox regression with time-varying covariates and regression effects

A model may include both constant and time-varying effects:

$$\mu(t, |\beta, Z) = \mu_0(t) e^{Z(t)^T \beta(t) + X(t)^T \gamma}$$

- Here  $Z(t)$  and  $X(t)$  are time-varying covariates (updated over time).
- $Z(t)$  are covariates with time-varying hazards  $\beta(t)$ , and  $X(t)$  covariates have constant hazards  $\gamma$ .
- See Ch. 6 in the book by Martinussen&Scheike [2] and their R package *timereg* for analysis of extended multiplicative hazards models.
- Their program *timecox* can test for and fit models with both constant and time-varying effects.



Actuarial  
Research Centre

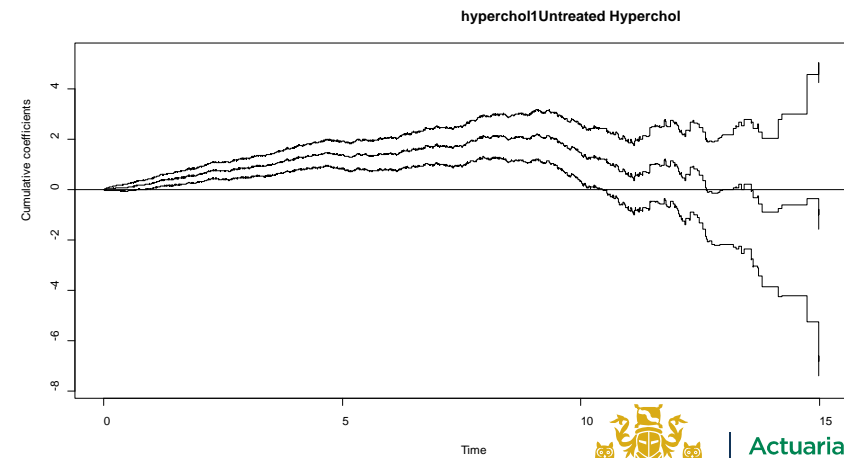
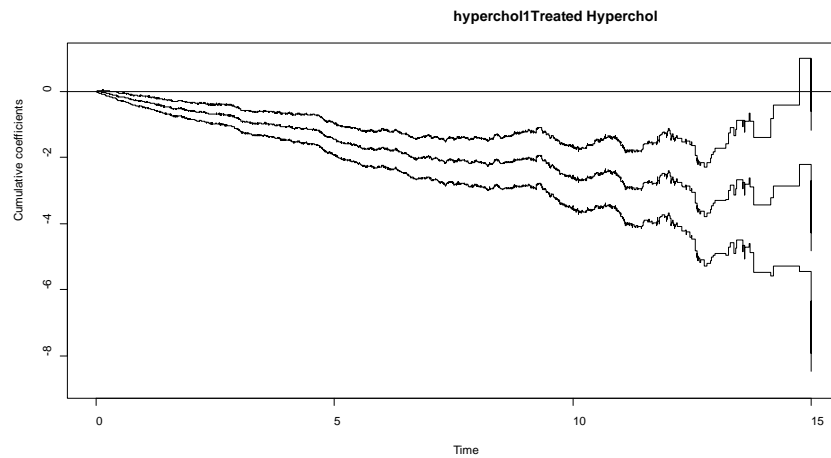
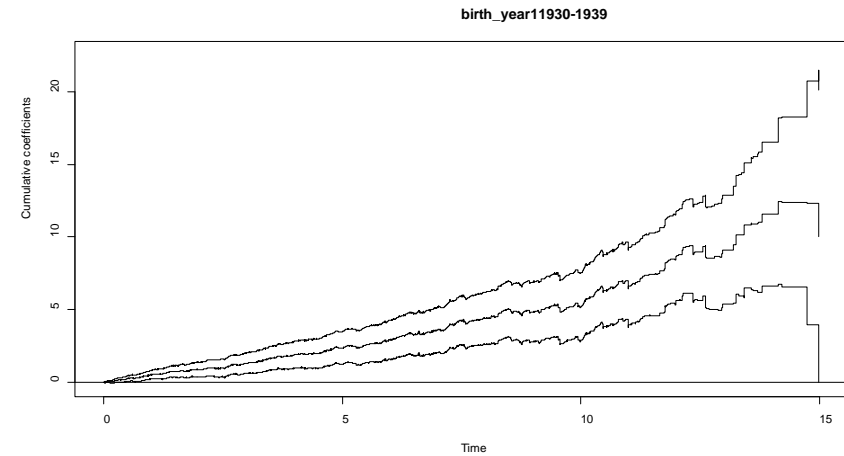
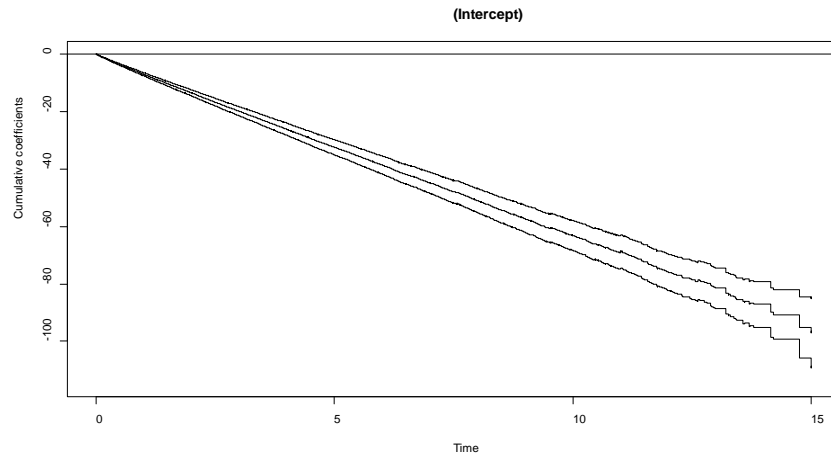
Institute and Faculty  
of Actuaries

# Inference in extended Cox model

- It is easier to estimate cumulative regression coefficients  $B(t) = \int_0^t \beta(s) ds$ , their estimates are  $n^{1/2}$ -consistent and asymptotically Normal.
- This allows to draw confidence bands for  $B(t)$  and to test hypotheses about them.
- A simple test of  $\beta_p(t) = \beta_p$  is based on maximum deviation of the cumulative coefficient  $B_p(t)$  from a straight line over an interval  $[0, T]$ .
- Similarly, cumulative residuals are used for various diagnostic purposes.



# Plots of cumulative coefficients for DM2 study





# Robustness of the Cox model

Consider once more the extended Cox model

$$\mu(t|\beta, Z) = \mu_0(t)e^{Z^T\beta(t)}.$$

The cumulative hazard  $M(t|Z) = -\ln(S(t|Z))$ . The ratio

$$\frac{M(t|Z)}{M_0(t)} = \frac{\int \mu_0(s) \exp(Z^T\beta(s)) ds}{\int \mu_0(s) ds} \approx \frac{\exp\left(\int \mu_0(s) (Z^T\beta(s)) ds\right)}{\int \mu_0(s) ds} = \exp(Z^T\bar{\beta}(t)),$$

where  $\bar{\beta}(t) = \frac{\int \mu_0(s)\beta(s) ds}{\int \mu_0(s) ds}$ , if the variance  $\frac{\int \mu_0(s) (Z^T(\beta(s) - \bar{\beta}(t)))^2 ds}{\int \mu_0(s) ds}$  is

small. This means that the Cox model gives approximately correct predictions of surviving up to time  $t$ .



Actuarial  
Research Centre

Institute and Faculty  
of Actuaries

# What is landmark analysis

In the landmarking approach, dynamic predictions for the conditional survival after  $t=t_{LM}$  is used on current information for all patients still alive just prior to  $t_{LM}$ . [Van Houwelingen, H. and Putter, H., 2011]

The sliding landmark model is the simple Cox model

$$h(t|x, t_{LM}, w) = h_0(t|t_{LM}, w) \exp(x^T \beta_{LM}), \quad s \leq t \leq s + w$$

for the data set obtained by truncation at  $s = t_{LM}$  and administrative censoring at  $t_{LM}+w$ .

$h_0(t|t_{LM}, w)$  is the baseline hazard or force of mortality.

This is a convenient way to obtain a dynamic prediction without fitting a complicated model with time-varying effects.



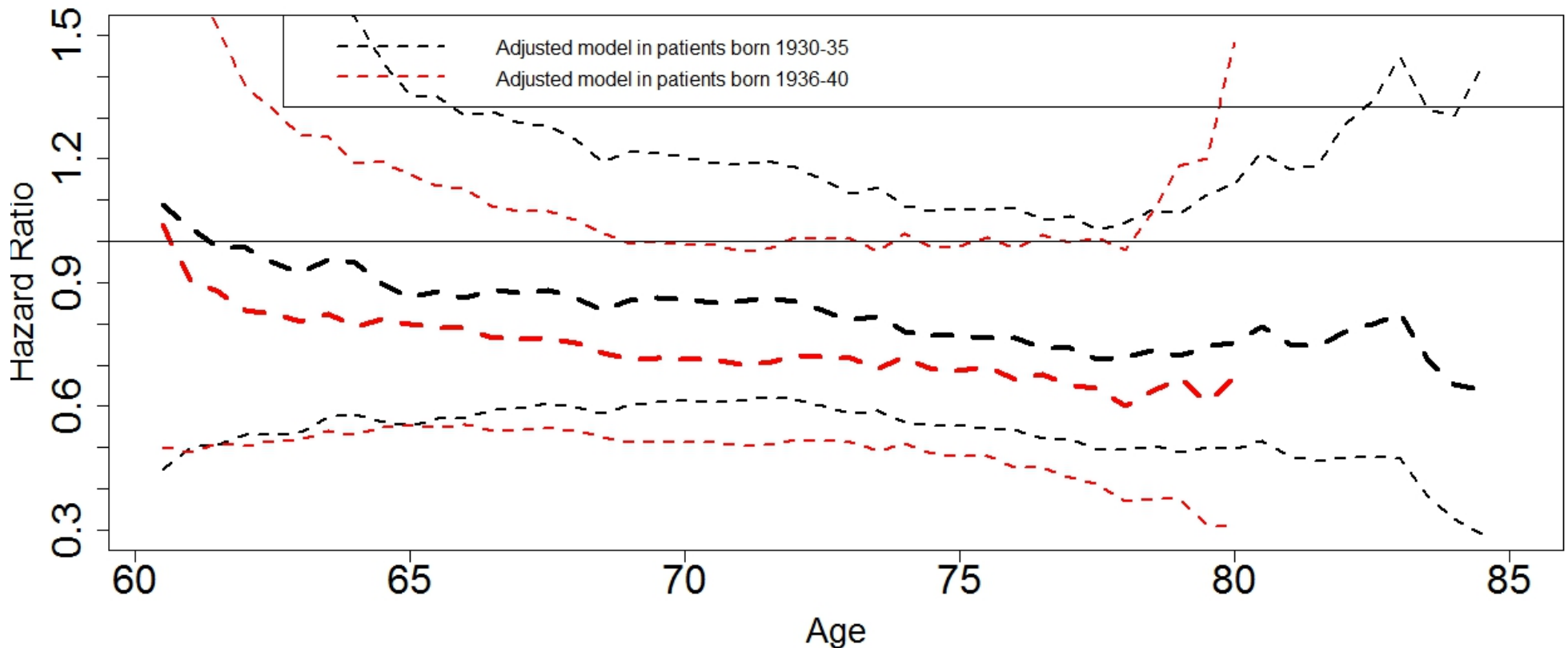
# Super-prediction data set

- Fix the prediction window  $w$ ; [say,  $w=5$  years]
- Select a set of prediction time points  $\{s_1, \dots, s_L\}$ ,  $20 \leq L \leq 100$ ; [say, every 6 months.]
- Create a prediction data set for each  $t_{LM}=s_l$  by truncation and administrative censoring;
- Stack all these data into a single “Super-prediction data set”. The subsets corresponding to a given prediction time  $t_{LM}=s_l$  are “strata”.
- The risk set  $R(t_i)$  for an event time  $t_i$  is present in all strata with  $s \leq t_i \leq s + w$ . Passing from stratum  $s$  to  $s+1$  corresponds to sliding the window over the time range.
- Individual life  $j$  contributes up to  $w/|s_{l+1} - s_l|$  times in each prediction window. [10 times when  $w=5$  and the time shift  $s_{l+1} - s_l$  is 6 months.]



# Sliding Cox model results (crude model)

## Hazard of all-cause mortality associated with statin prescription



# Integrated partial log-likelihood landmark model - ipl

The landmark super prediction model with window  $w$  and letting the regression coefficients  $\beta_{LM}$  depend on time  $t_{LM}$  is given by

$$h(t|x, t_{LM} = s, w) = h_0(t|s, w) \exp(x^T \beta_{LM}(s)), \quad s \leq t \leq s + w$$

where

$$\beta_{LM}(s) = \sum_{j=1}^m \gamma_j f_j(s).$$

- $f_j(s)$  are the basis functions,  $f_1(s)=1$ ,  $f_j(0)=0$  for  $j>1$ , and  $\gamma_j$  are the parameters, with  $\beta_{LM}(0) = \gamma_1$ .
- The parameters of this model are estimated by maximizing the integrated (over  $s$ ) partial log-likelihood introduced by van Houwelingen (2007).
- This approach is based on a stratified (on  $s$ ) analysis with smooth landmark dependent effect  $\beta_{LM}(s)$  and separate estimated baseline hazards for each stratum.



Actuarial  
Research Centre

Institute and Faculty  
of Actuaries

# Pseudo-partial log-likelihood landmark model - *ipl*\*

In the *ipl*\* model, the baseline hazard is modelled directly as

$$h_0(t|s, w) = h_0(t) \exp(\theta(s)),$$

where  $\theta(s) = \sum_{j=1}^m \eta_j g_j(s)$

for proper basis functions  $g_j(s)$  with  $g_j(s_1) = 0$ , resulting in

$$h(t|x, t_{LM} = s, w) = h_0(t) \exp(x^T \beta_{LM}(s) + \theta(s)), \quad s \leq t \leq s + w,$$

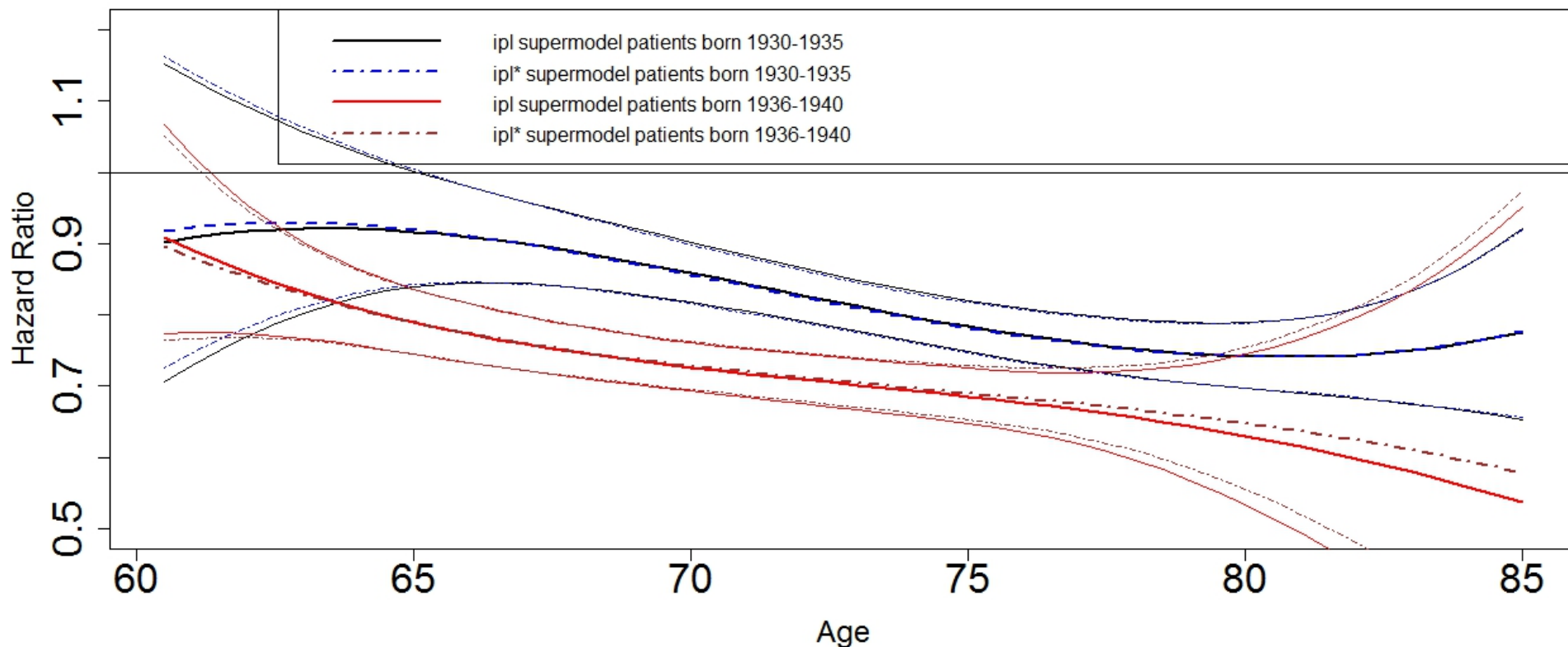
where  $\beta_{LM}(s)$  and  $\theta(s)$  are the  $m$ th degree polynomials in  $s$ .



Actuarial  
Research Centre

Institute and Faculty  
of Actuaries

# Adjusted hazard of all-cause mortality associated with current statin prescription



Actuarial  
Research Centre

Institute and Faculty  
of Actuaries

# Predicted probabilities of survival in a window

Predictions for all  $s \in [s_1, s_L]$  in the *ipl*\* model are obtained from estimated cumulative hazards

$$H(s + w|x, t_{LM} = s) = \exp(x^T \beta_{LM}(s) + \theta(s)) (H_0^*(s+w) - H_0^*(s-))$$

This is because in the *ipl*\* model

$$h(t|x, t_{LM} = s, w) = h_0(t) \exp(x^T \beta_{LM}(s) + \theta(s)), \quad s \leq t \leq s + w,$$

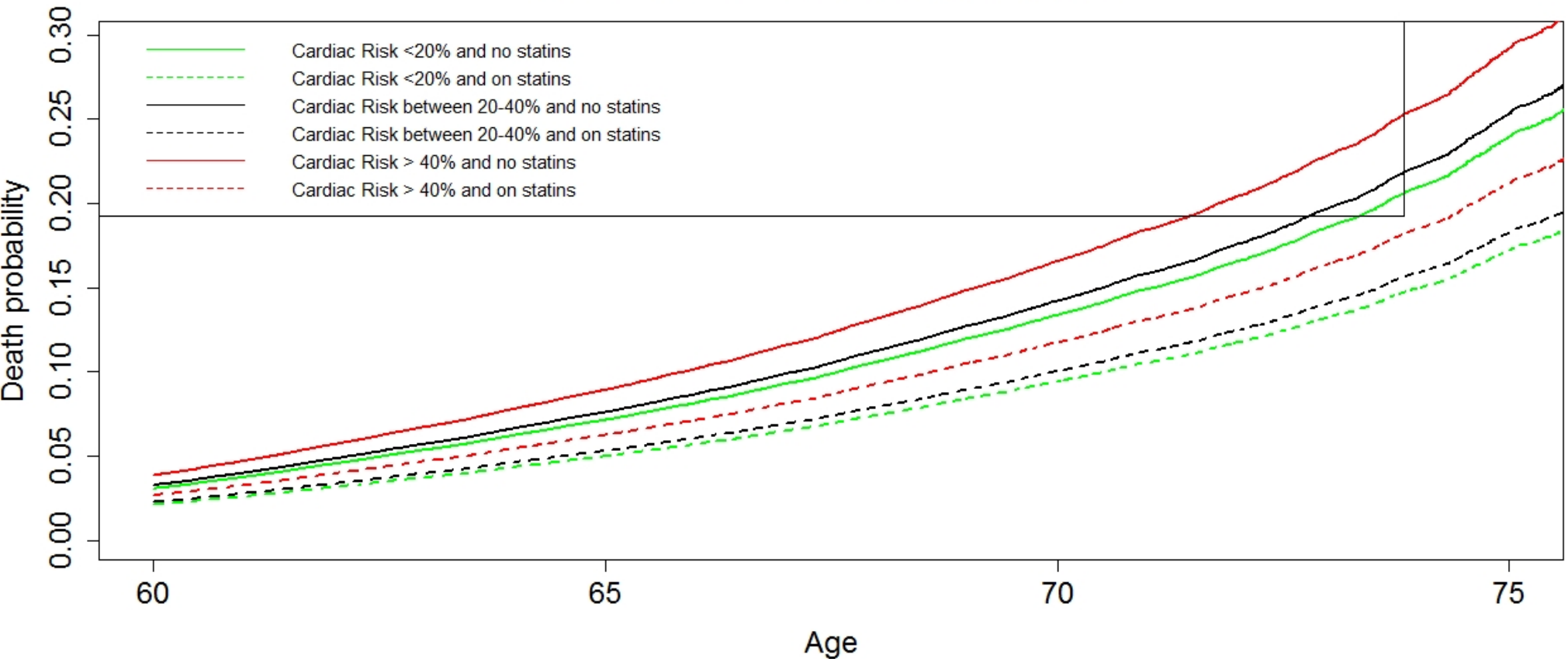
only the baseline hazard  $h_0(t)$  depends on  $t$ .



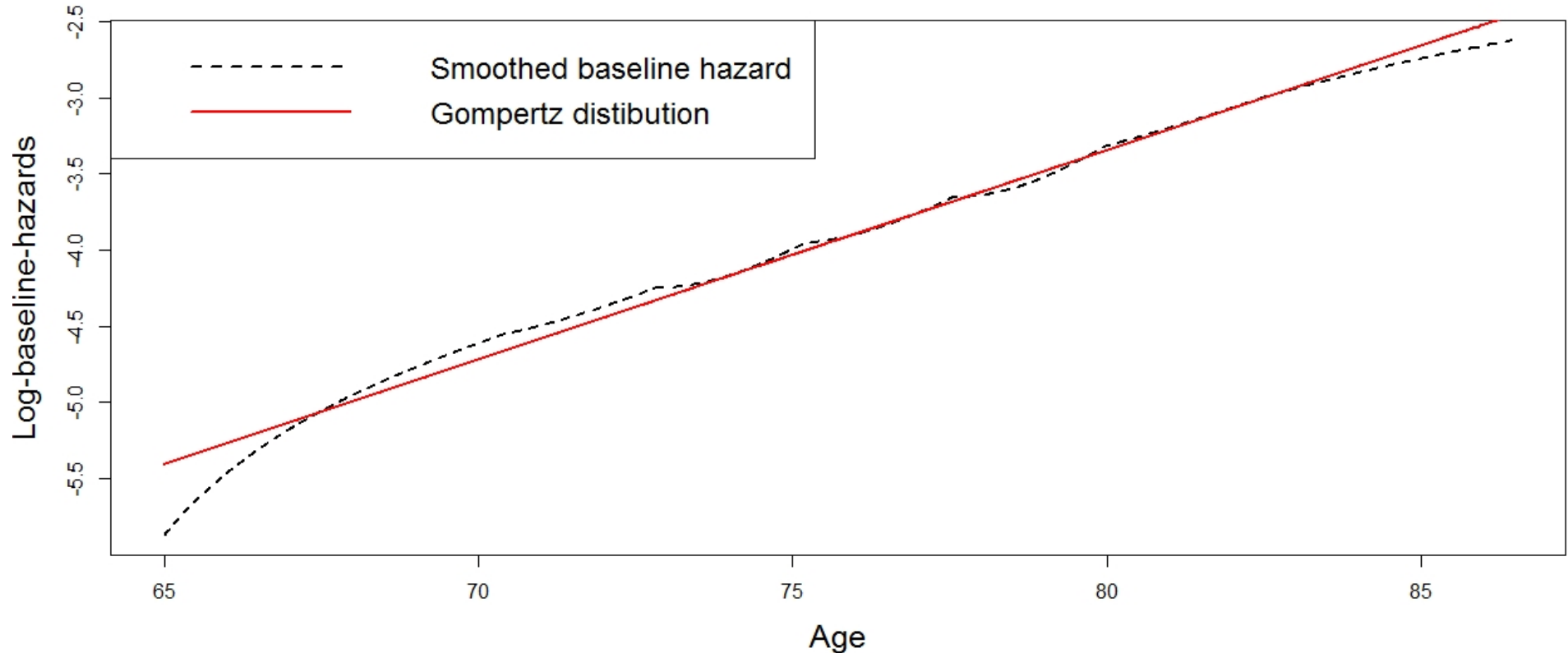


# Probabilities of death for 1936-1940 cohort

Dynamic prediction with 10 year window



# Baseline hazard in the statins landmark model



The baseline hazard is well approximated by the Gompertz hazard



Actuarial  
Research Centre  
Institute and Faculty  
of Actuaries

# The *ipl*\* landmark model in actuarial research

In the *ipl*\* model, the hazards are modelled as

$$h(t|x, t_{LM} = s, w) = h_0(t) \exp(x^T(s)\beta_{LM}(s) + \theta(s)), \quad s \leq t \leq s + w,$$

where  $\beta_{LM}(s)$  and  $\theta(s)$  are the  $k$ th and the  $(k-1)$ th degree polynomials of  $s = t - t_0$ .

The log-hazards are  $\lambda(t|x, t_0) = \lambda_0(t) + x^T(s)\beta_{LM}(s) + \theta(s)$ .

For Gompertz baseline hazard,  $\lambda_0(t) = a + bt$ .

Values of  $a$  and  $b$  can be estimated from the estimated baseline hazard or substituted for a particular population. Next, we can obtain cumulative hazards, survival and period life expectancy for various scenarios of changing risks  $x(s)$ .



# Discussion and conclusions

- The most general form of extended Cox regression with time-dependent effects is difficult to use. To make it relevant to actuarial research we also need to consider the shape of the baseline hazards.
- Parametric “double-Cox” model is a useful replacement for the stratified Cox model which also models shape of baseline hazards and can be easily used for actuarial purposes.
- Landmark analysis is a convenient way to model dynamically changing survival data. The  $ipl^*$  model conveniently lends itself to actuarial modelling.
- Extra development is required to use the results for population LE projections using methodology similar to that in Kulinskaya et al. (2019) .



# References:

1. Begun A., Kulinskaya E. and MacGregor A. , 2019. Risk-adjusted CUSUM control charts for shared frailty survival models with application to hip replacement outcomes: a study using the NJR dataset. *BMC Medical Research Methodology*, in print.
2. Martinussen T, Scheike TH. *Dynamic Regression Models for Survival Data*. Springer: New York, 2006.
3. Van Houwelingen, H.C., 2007. Dynamic prediction by landmarking in event history analysis. *Scandinavian Journal of Statistics*, 34(1), pp.70-85.
4. Van Houwelingen, H. and Putter, H., 2011. *Dynamic prediction in clinical survival analysis*. CRC Press.
5. Gitsels, L.A., Bakbergenuly, I., Steel, N. and Kulinskaya, E., 2019. Dynamic prediction of long-term survival effects of statin therapy: a landmarking analysis in the general population. *British Medical Journal* (under review)
6. Kulinskaya E., Gitsels, LA. and Bakbergenuly, I., 2019. Calculation of changes in individual and period life expectancy based on proportional hazards model of an intervention. *Insurance Mathematics and Economics*, under review.



Questions

Comments



**Actuarial  
Research Centre**

Institute and Faculty  
of Actuaries



**Actuarial  
Research Centre**

Institute and Faculty  
of Actuaries

**UEA**  
University of East Anglia

# Landmark analysis of survival benefits of statin prescription

Dr Ilyas Bakbergenuly (UEA)

School of Computing Sciences,  
University of East Anglia

The **'Use of Big Health and Actuarial Data for understanding Longevity and Morbidity Risks'** research programme is being funded by the Actuarial Research Centre.

30/10/2019

[www.actuaries.org.uk/arc](http://www.actuaries.org.uk/arc)



## Actuarial Research Centre

Institute and Faculty  
of Actuaries

### Disclaimer

The views expressed in this presentation are those of invited contributors and not necessarily those of the Institute and Faculty of Actuaries. The Institute and Faculty of Actuaries does not endorse any of the views stated, nor any claims or representations made in this presentation and accept no responsibility or liability to any person for loss or damage suffered as a consequence of their placing reliance upon any view, claim or representation made in this presentation. The information and expressions of opinion contained in this presentation are not intended to be a comprehensive study, nor to provide actuarial advice or advice of any nature and should not be treated as a substitute for specific advice concerning individual situations. On no account may any part of this presentation be reproduced without the written permission of the Institute and Faculty of Actuaries .



# Introduction

- We aim to demonstrate the use of landmark analysis in actuarial research using the statin survival benefits as a case study.
- Statins have been widely prescribed for cardiac prevention
- Clinical trials have demonstrated the survival benefits of statin prescription
- The threshold of cardiac risk at which to prescribe statins is still controversial, especially at older ages where everyone would be eligible solely due to their age.
- Little is known about the effect of long-term prescription in the general population, where sequential treatment decisions are made according to the latest clinical guidelines.



# The Health Improvement Network (THIN) data

- Anonymised electronic primary care medical records (Vision)
- Data collection began in 2003 using Read codes
- 11 million patients, 3.7 million active patients
- 562 general practices, covering 6.2% of the UK population
- Diagnoses, prescriptions, consultations, postcode deprivation



## Subset of THIN selected for our research:

- 110,243 patients who turned 60 between 1990 and 2000 and did not have a previous statin prescription or a cardiovascular disease diagnosis



Actuarial  
Research Centre  
Institute and Faculty  
of Actuaries

# Primary prevention of CVD

Primary prevention: no previous history of CVD

- Example: lipid-lowering therapy - statins

National Institute of Health and Clinical Excellence (NICE):

- Offer atorvastatin 20 mg for the primary prevention of CVD to people who have a 10% or greater 10-year risk of developing CVD.
- Estimate the level of risk using the QRISK2 assessment tool
- [www.nice.org.uk/guidance/cg181/](http://www.nice.org.uk/guidance/cg181/)
- [www.qrisk.org/2016/](http://www.qrisk.org/2016/)



Up to 17 million UK residents eligible for statins

About you

Age (25-84):

Sex:  Male  Female

Ethnicity:

UK postcode: leave blank if unknown

Postcode:

Clinical information

Smoking status:

Diabetes status:

Angina or heart attack in a 1st degree relative < 60?

Chronic kidney disease?

Atrial fibrillation?

On blood pressure treatment?

Rheumatoid arthritis?

Leave blank if unknown

Cholesterol/HDL ratio:

Systolic blood pressure (mmHg):

Body mass index

Height (cm):

Weight (kg):

Calculate risk over  years.

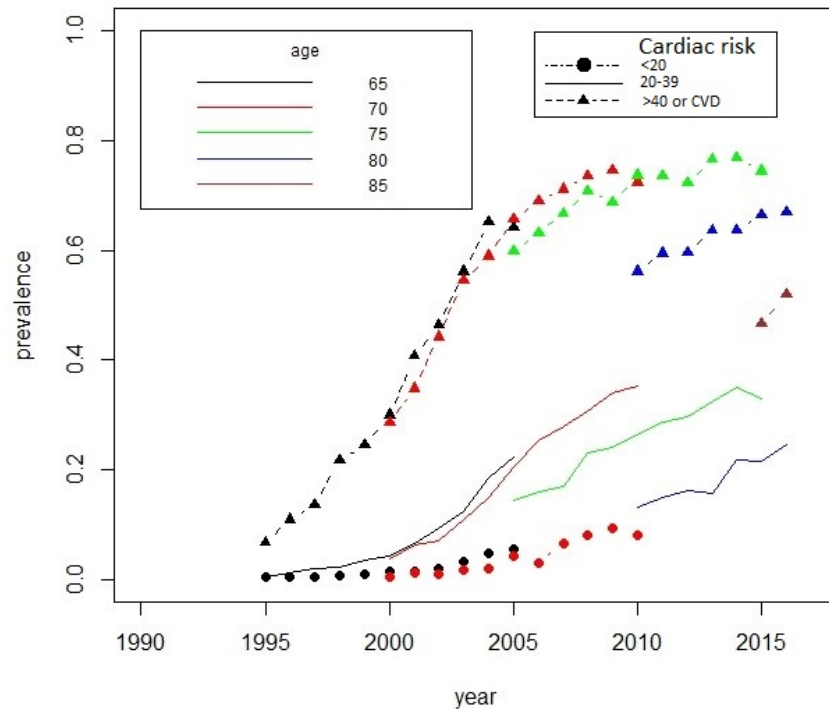


Actuarial  
Research Centre

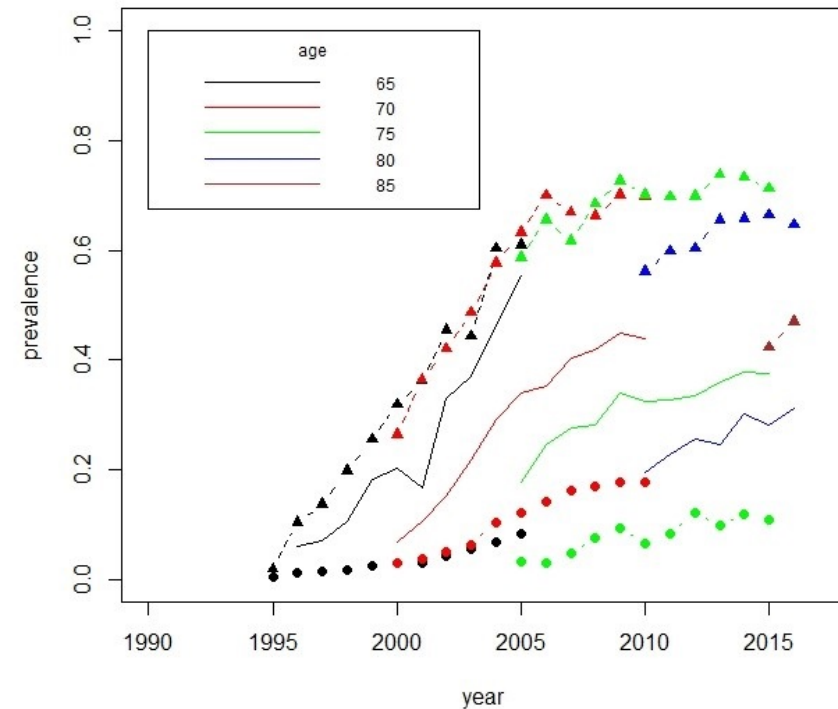
Institute and Faculty  
of Actuaries

# Prevalence of statin prescription

statin prescription in male patients



statin prescription in female patients



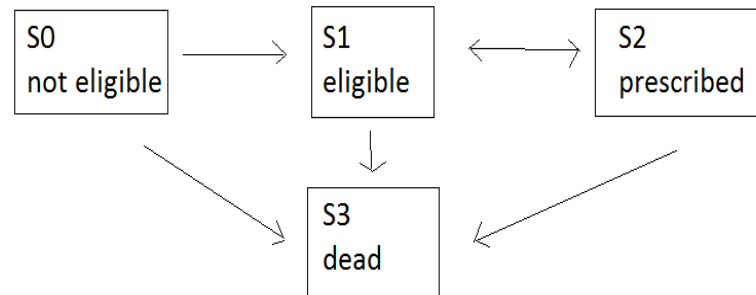
Prevalence of statin prescription differs by calendar year, age, sex and cardiac risk group



Actuarial  
Research Centre  
Institute and Faculty  
of Actuaries

# Statistical Analysis Options

- Objective: dynamically predict the survival benefits associated with statin therapy over the course of 25 years.
- The original plan was to develop a model with the following states: S0 not eligible for statins, S1 eligible for statins, S2 prescribed statins, and S3 death.



- Alternatively, develop a survival model with time-dependent predictors and parameters.
- Or use landmark analysis.



# Adherence to statin prescription

Number of arm switches	0	1	2	3	4	5+
% of patients	51.1	40.7	6.2	1.5	0.4	0.1

- 51.1% were never prescribed statins;
- 40.7% were prescribed at some age and stayed on statins;
- 6.2% dropped off statins permanently;
- 1.5% dropped off statins and then came back on to stay;
- 0.5% had 4 or more switches;
- the maximum was 9 switches for 1 person



# Data preparation and analysis

**Data:** Medical history was updated every half a year (landmark) until end of follow-up (death, deregistered or end of study).

**Imputation:** Due to missing data at early ages, multiple imputation was performed using joint modelling at age 60. The method of last observation carried forward was used for missingness in follow up.

**Analysis:** Landmark analyses were carried out by fitting Cox proportional hazards regression of all cause mortality associated with current statin prescription at each landmark from age 60 to 85 and adjusted for medical history.

We separately conducted three landmark analyses: with window widths 5, 10 and 30 years.



# The four stages of modelling process

- A Cox model was fitted on complete cases at baseline age to inform the imputation model. Both models included all medical history if prevalent.
- Cox models were fitted on the imputed datasets at ages 65, 70, 75, 80 and 85 to inform the final landmark model. These models included all medical history and tested for interactions between statin prescription, sex, year of birth and cardiac risk.
- The final, fully adjusted, Cox landmark models were fitted at 10 imputed datasets. The landmarking was smoothed with an integrated partial log-likelihood (ipl) and with Pseudo-partial log-likelihood (ipl\*).
- Ten landmark models pooled using Rubin's rules.





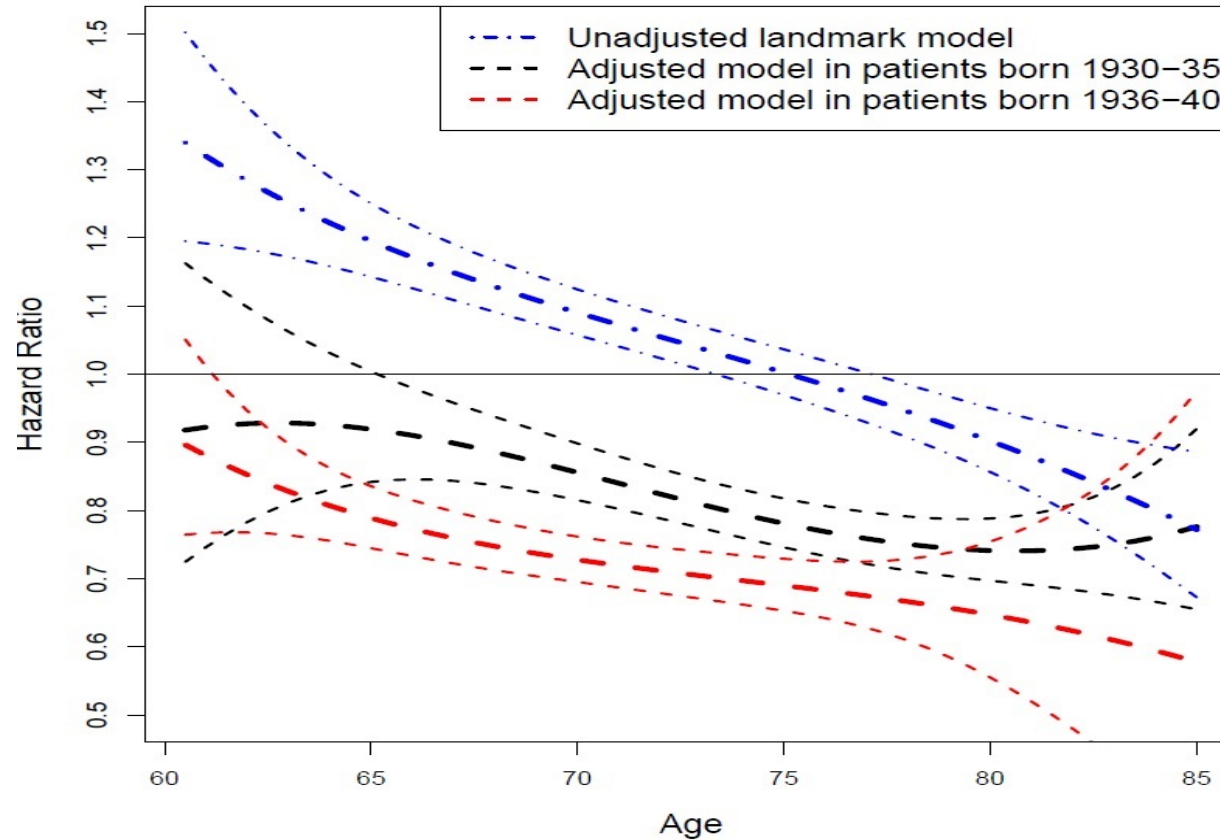
# The statistical model for survival benefits of statins was adjusted for:

- Cardiac risk at three levels: low ( $QRISK2 \leq 20\%$ ), medium ( $QRISK2$  of 20-39%) and high ( $QRISK2 \geq 40$  or CVD diagnosis)
- Sex, birth cohort, Townsend deprivation quintile, chronic kidney disease, diabetes, treated hypertension, hypercholesteremia, aspirin, BMI, alcohol consumer status, smoking status and general practice



# Hazard of all-cause mortality associated with statin prescription (30 years window)

Hazard of all-cause mortality associated with statin prescription



# Why statins are more beneficial in younger cohort: better drugs?

www.pulseoday.co.uk/gps-set-for-mass-drug-switch-to-atorvastatin-after-analysis-shows-price-could-fall-by-95/13482969/article

We use cookies to give you the best experience. If you do nothing we'll assume that it's ok. [More info](#) Agree

**PULSE** At the heart of general practice since 1960

OPINION 'After about 20 patients, there is not an iota of empathy left'

NEWS ↓ VIEWS ↓ CLINICAL ↓ PARTNERS ↓ SESSIONALS ↓ TRAINEE PULSE MAGAZINE EVENTS JOBS

## GPs set for mass drug switch to atorvastatin after analysis shows price could fall by 95%

22 February 2012

Share Print Rate Save Comment

SHARE ON FACEBOOK

Exclusive GPs are set to be enrolled in schemes to switch patients en masse to atorvastatin in the wake of an analysis for the Government's

MOST POPULAR MOST COMMENTED

### Reduction in low-density lipoprotein cholesterol

Statins are grouped by NICE into three different intensity categories according to the percentage reduction in LDL cholesterol

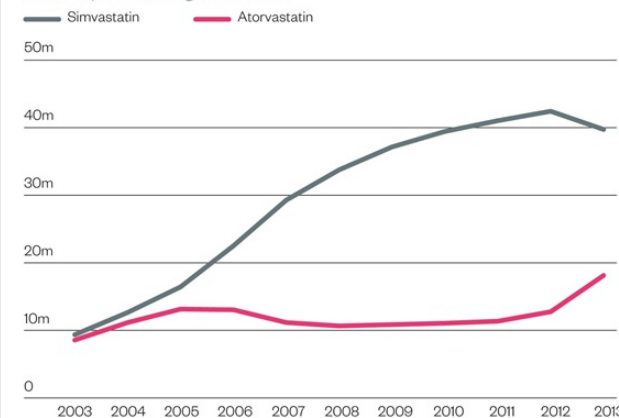
- 20% to 30%: low-intensity statin
- 31%-40%: medium-intensity statin
- Above 40%: high-intensity statin

	5mg	10mg	20mg	40mg	80mg
Fluvastatin	—	—	●	●	●
Pravastatin	—	●	●	●	—
Simvastatin	—	●	●	●	●
Atorvastatin	—	●	●	●	●
Rosuvastatin	●	●	●	●	—

Source: National Institute for Health and Care Excellence

### Atorvastatin vs simvastatin

Items dispensed in England (millions)



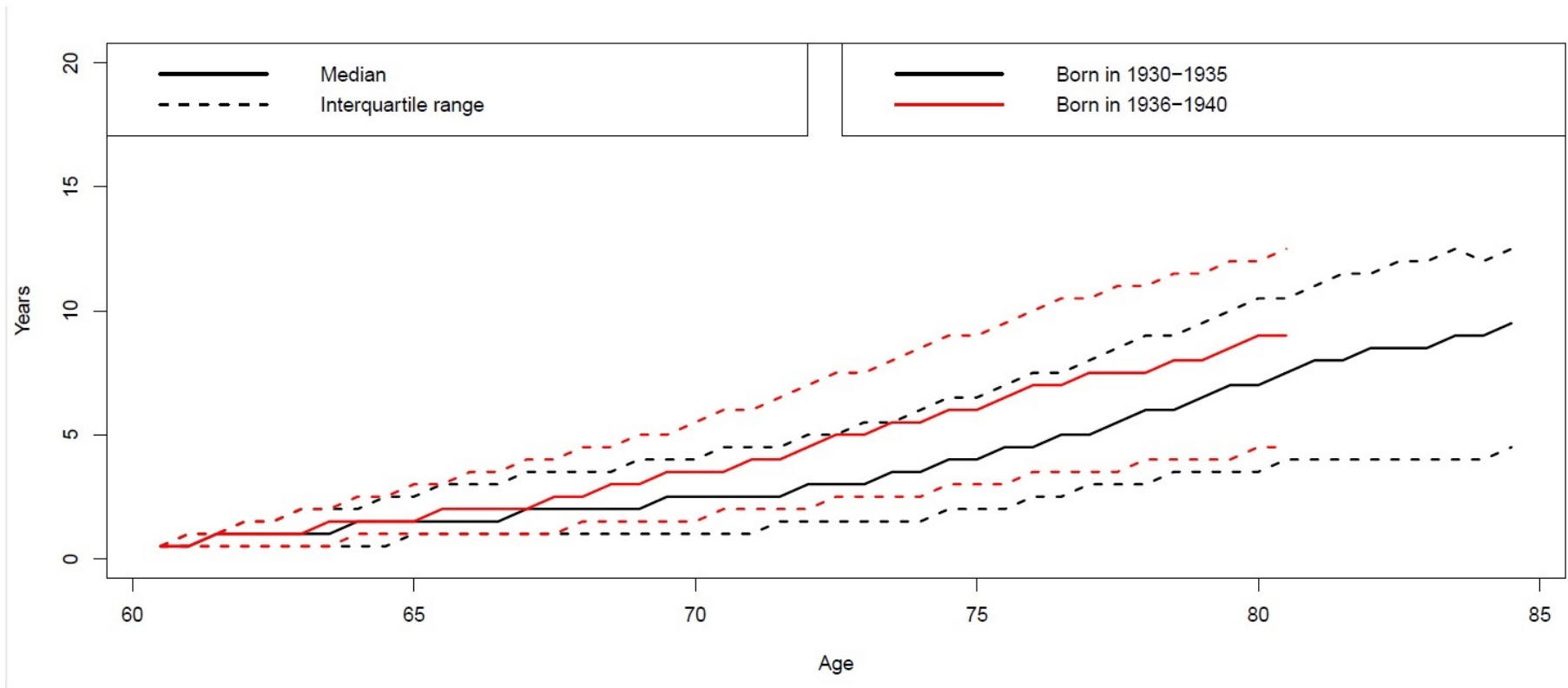
Cerivastatin was withdrawn from the world market in 2001 and the clinical guidelines changed from simvastatin to atorvastatin in 2014. But in 2014 our patients were 79-89 years old.



Actuarial Research Centre

Institute and Faculty of Actuaries

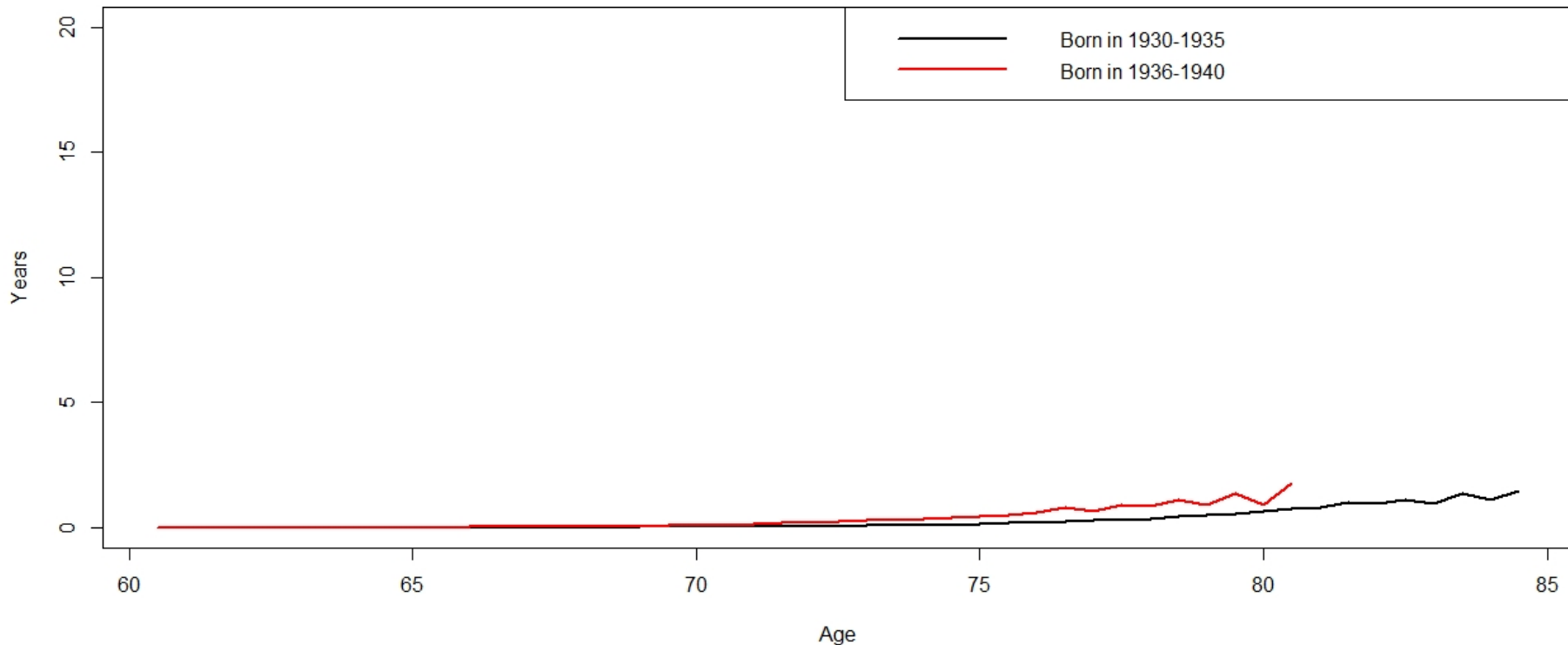
# Length of prior prescription for patients on statins at age s



Actuarial  
Research Centre

Institute and Faculty  
of Actuaries

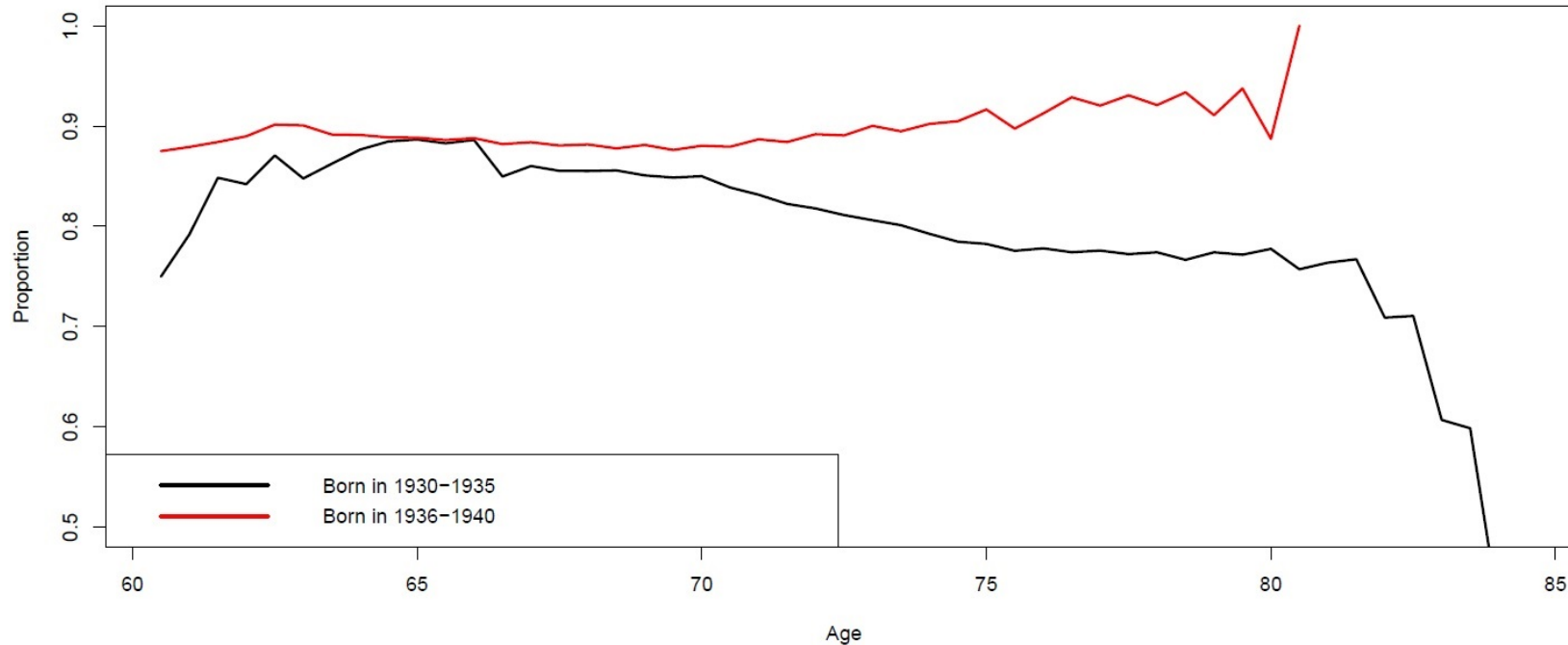
# Average length of prior prescription for patients not on statins at age s



Actuarial  
Research Centre

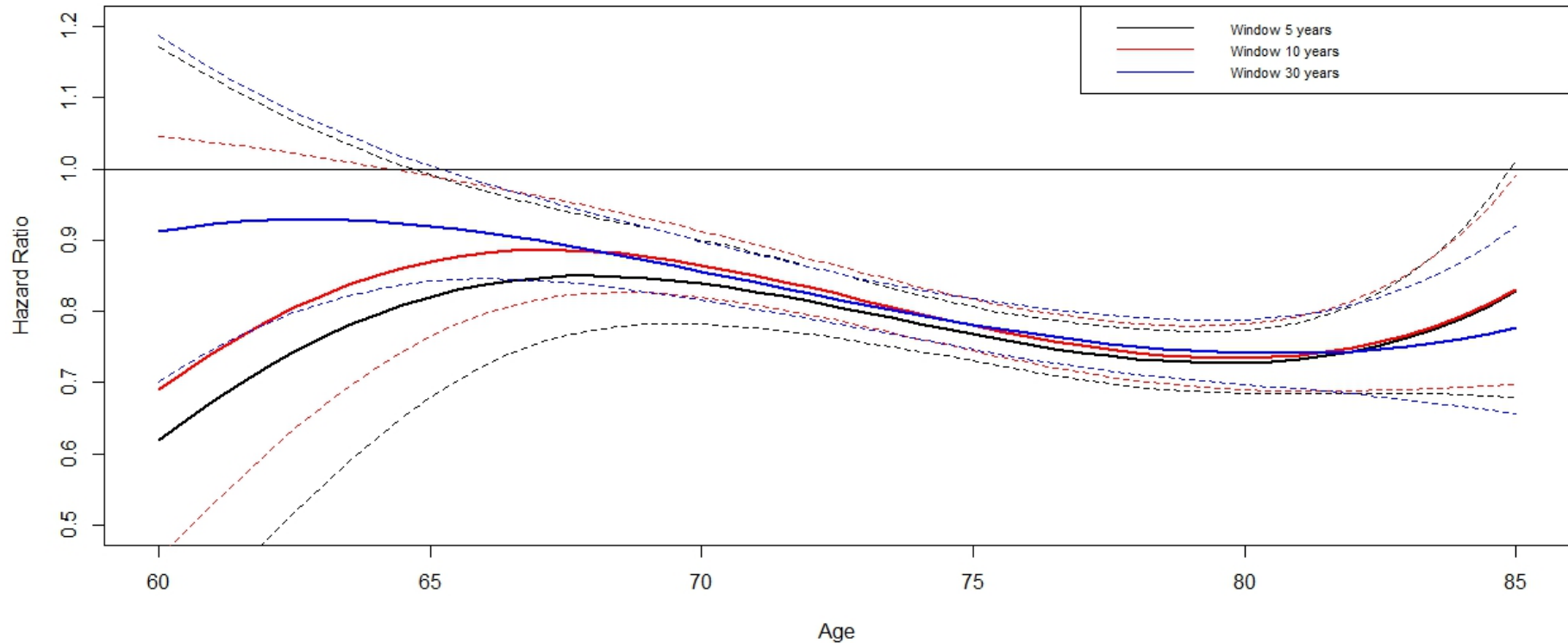
Institute and Faculty  
of Actuaries

# Proportion of patients on statins at age s with at least 75% adherence at follow up



# HRs of all-cause mortality estimated in 5, 10 and 30 years window

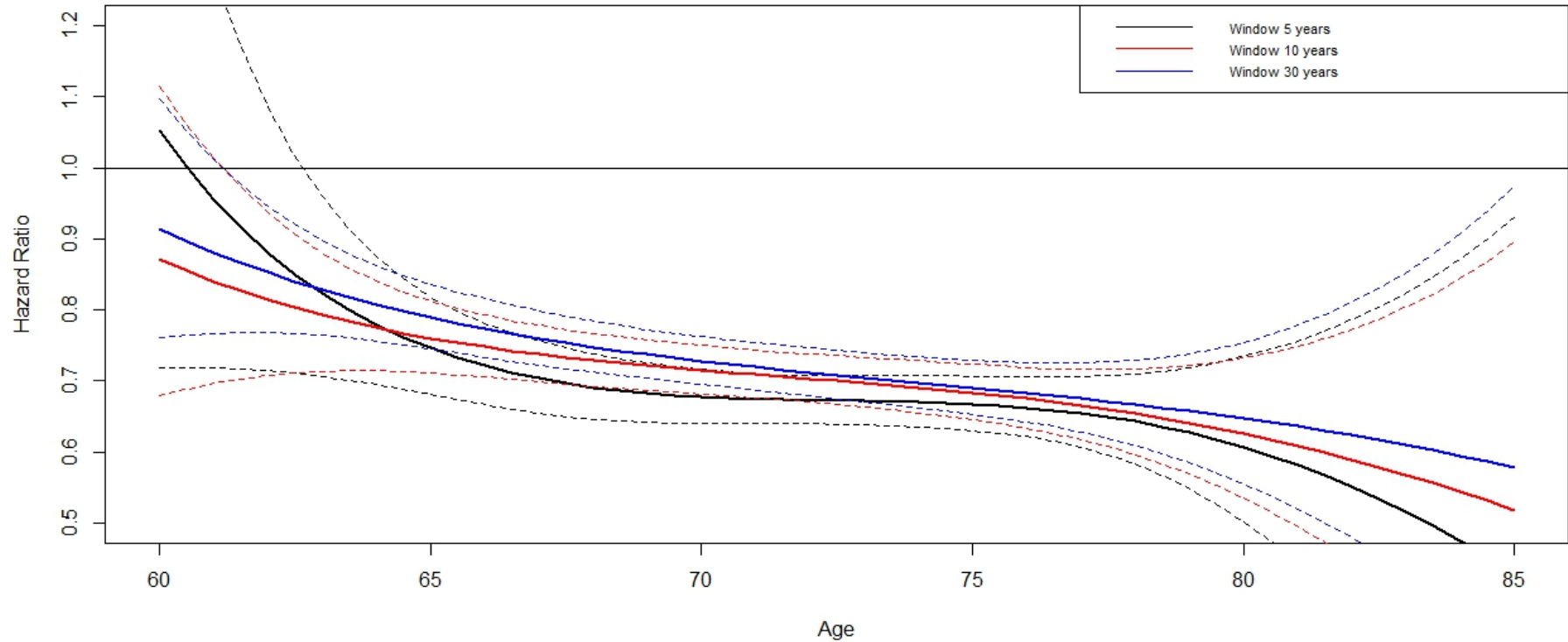
Born in 1930-1935



Actuarial  
Research Centre  
Institute and Faculty  
of Actuaries

# HRs of all-cause mortality estimated in 5, 10 and 30 years window

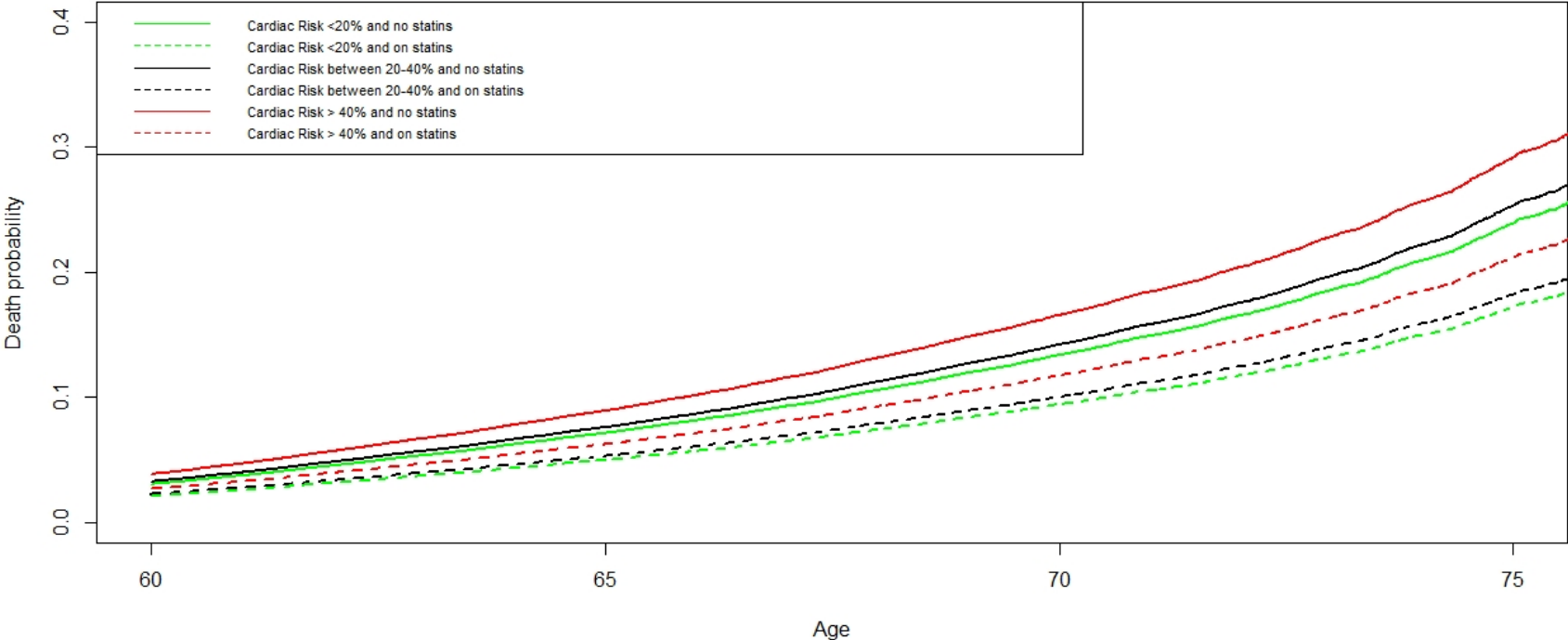
Born in 1936-1940





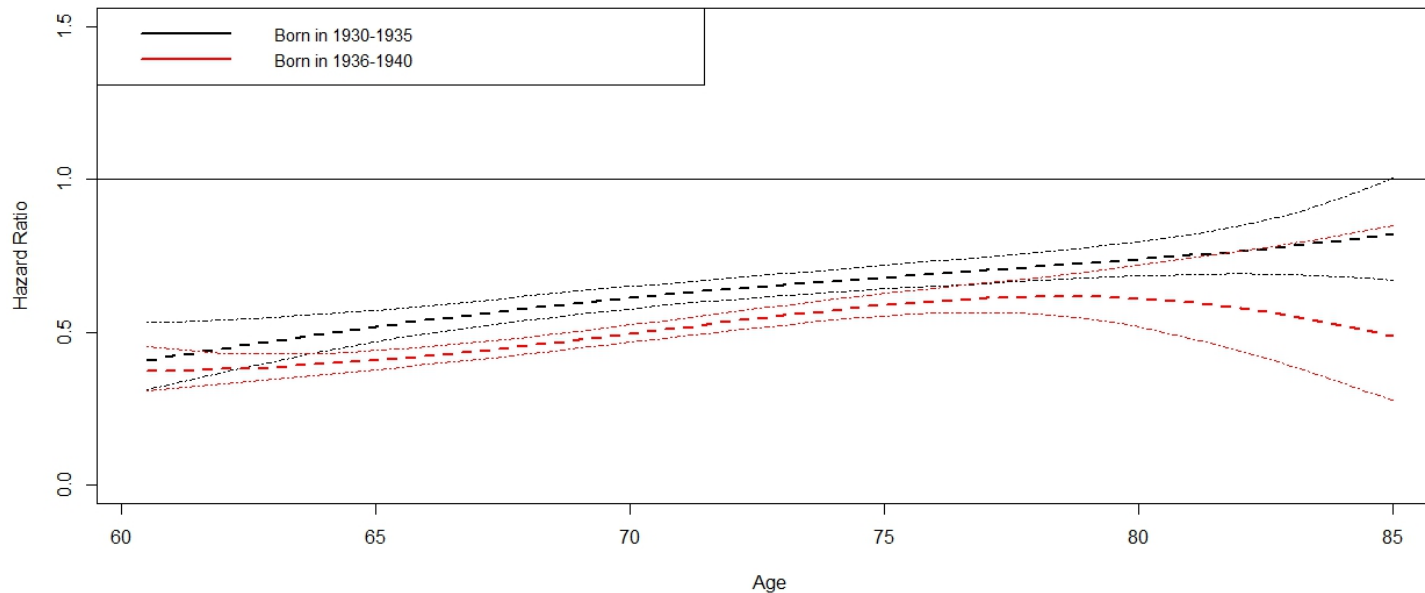
# Probabilities of death for 1936-1940 cohort

Dynamic prediction with 10 year window



# Do survival benefits really increase at older ages? Here controls never were on statins:

We also performed an analysis keeping only patients who never were prescribed statins in the control group. Younger patients do better!



Similar HRs (0.74 ,0.63 vs 0.74, 0.61 here) only from age 80!



Actuarial  
Research Centre  
Institute and Faculty  
of Actuaries

# Summary of results on statins

- The prevalence of statin prescription increased substantially by age with nearly half of the study population having had a prescription by age 75 and 57% by age 85 at the end of the study.
- The adherence to statin prescription was high, with 77% adhering more than 75% of the time and only 5% adhering less than 25% of the time
- In “current knowledge” landmark analysis, statin prescription was associated with increasing survival benefits at older ages and was significant at the earliest from age 62 onward. Benefits seemed to decrease with age in our sensitivity analysis based on the full knowledge of statin history.
- Statin prescription was more effective in patients born in later years due to the changing availability and recommended dosages of statin types resulting in more effective treatment but did not differ by sex or cardiac risk.
- Therefore, age alone can be used to decide on initiating and staying on statin therapy based on the predicted overall effect (which tallied up benefits and harms).



# Discussion and conclusions on statins

- After adjustment for cardiac risk and related medical history, it appears that statin therapy is especially beneficial at older ages and in people born at later years in a realistic “current knowledge” scenario. The benefits of statins in earlier ages may be underestimated as more people will get statin prescription later.
- This study adjusted for cardiac risk groups defined by the changing clinical guidelines on the eligibility of statin prescription. However we did not distinguish between recommended types and doses of statins. This might partly explain why statin prescription was associated with greater survival benefits in patients born in later years.
- We used statin prescription as a proxy for statin intake. Lower intake than prescription would result in more conservative findings and thus imply that statins could be even more beneficial.



# References:

1. Gitsels, L.A., Kulinskaya, E. and Steel, N., 2016. Survival benefits of statins for primary prevention: a cohort study. *PloS one*, 11(11), p.e0166847.
2. Gitsels, L.A., Bakbergenuly, I., Steel, N. and Kulinskaya, E., 2019. Dynamic prediction of long-term survival effects of statin therapy: a landmarking analysis in the general population. *British Medical Journal* (under review)
3. Van Houwelingen, H. and Putter, H., 2011. *Dynamic prediction in clinical survival analysis*. CRC Press.



Questions

Comments



**Actuarial  
Research Centre**

Institute and Faculty  
of Actuaries



**Actuarial  
Research Centre**

Institute and Faculty  
of Actuaries

# Stroke Mortality and Morbidity in the UK

Padma Chutoo (PhD candidate)  
University of East Anglia

The **'Use of Big Health and Actuarial Data for understanding Longevity and Morbidity Risks'** research programme is being funded by the Actuarial Research Centre.

30/10/2019

[www.actuaries.org.uk/arc](http://www.actuaries.org.uk/arc)



## Actuarial Research Centre

Institute and Faculty  
of Actuaries

### Disclaimer

The views expressed in this presentation are those of invited contributors and not necessarily those of the Institute and Faculty of Actuaries. The Institute and Faculty of Actuaries does not endorse any of the views stated, nor any claims or representations made in this presentation and accept no responsibility or liability to any person for loss or damage suffered as a consequence of their placing reliance upon any view, claim or representation made in this presentation. The information and expressions of opinion contained in this presentation are not intended to be a comprehensive study, nor to provide actuarial advice or advice of any nature and should not be treated as a substitute for specific advice concerning individual situations. On no account may any part of this presentation be reproduced without the written permission of the Institute and Faculty of Actuaries .



# Overview

- Stroke definition and statistics
- Study description
- Patient Numbers
- Kaplan Meier plots
- Checking the Cox's Proportional hazard Assumption
- Parametric regression fits
- Double Cox-Weibull model specification
- IS model and hazard plots
- Overview of Multiple Imputation
- Future works

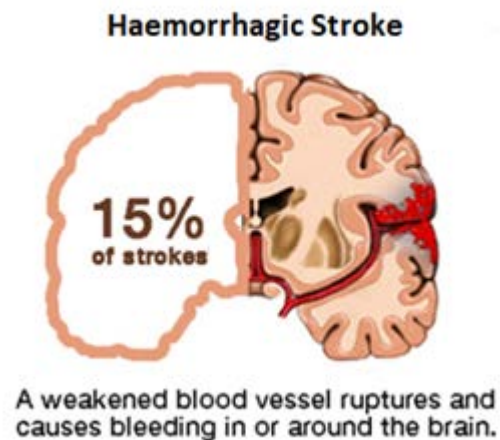
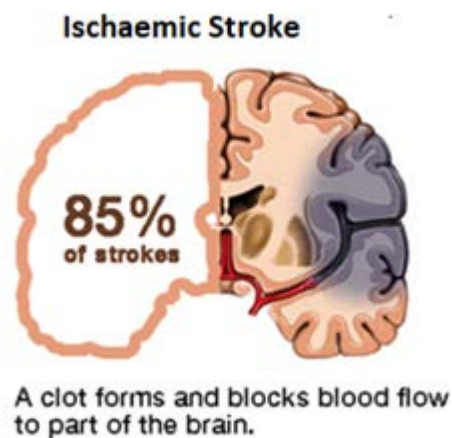


**Actuarial  
Research Centre**

Institute and Faculty  
of Actuaries

# What is Stroke?

- **Ischemic stroke** is caused by a blood clot that blocks or plugs a blood vessel in the brain.
- **Haemorrhagic stroke** is caused by a blood vessel that breaks and bleeds into the brain



- **Transient Ischemic Attacks** or TIAs, are “mini-strokes” whereby the symptoms from the clot appear temporarily. TIAs are warning signs that should be taken seriously.

Source : [medstarhealth.org](http://medstarhealth.org)



**Actuarial  
Research Centre**

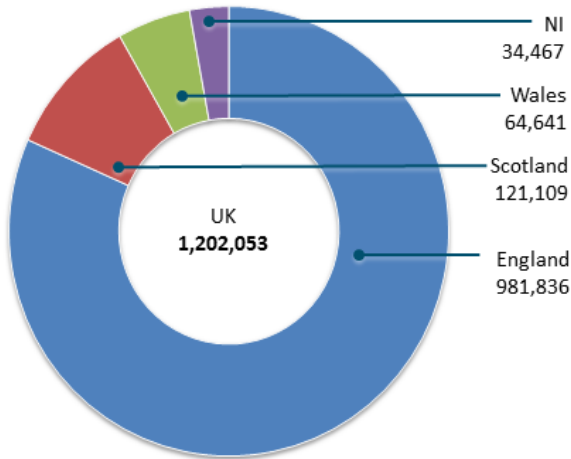
Institute and Faculty  
of Actuaries

# Stroke Statistics

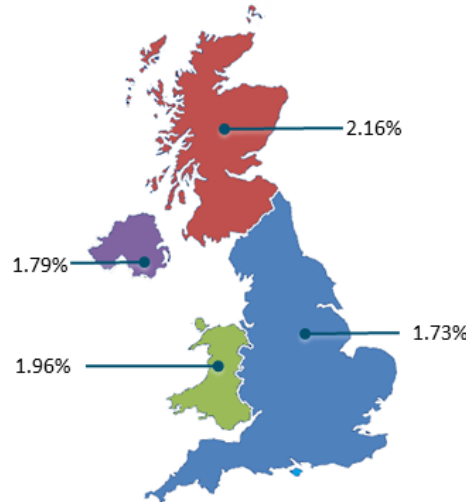
Stroke is the **fourth** single leading cause of death in the UK.



## STROKE SURVIVORS UK



Reference: Health and Social Care Information Centre. Quality and Outcomes Framework (QOF) – 2014-2015.



Stroke is the **biggest single cause of major disability** in the United Kingdom. Almost two-thirds of stroke survivors leave hospital with disability.

Stroke burden is projected to rise from around **38 million** Disability-Adjusted Life Years (DALYs) globally in 1990 to **61 million** DALYs in 2020.



The NHS and social care costs of stroke are around **£1.7 billion** a year in England



**Actuarial Research Centre**  
Institute and Faculty of Actuaries

# Stroke study: brief description

- Objective: impact of 1<sup>st</sup> **ischaemic stroke** and **transient ischaemic attack (TIA)** on longevity and morbidity risks.
- The study period is from 1986 up to 2017.
- Design: case/control 1:3
- Exclusion criteria: prior major cancers, dementia, chronic kidney disease stages 3+ and haemorrhagic stroke.
- The primary outcome is all-cause mortality. The secondary outcomes are further strokes, dementia (Alzheimer's and vascular dementia), heart failure, myocardial infarction, pulmonary arterial disease.



# Stroke study: brief description

## Variables of interest:

- **Drugs:** Antihypertensive drugs, Anticoagulant drugs, Lipid regulating drugs and antidiabetic drugs.
- **Medical conditions:** Asthma, Atrial Fibrillation, CKD, CHD, PAD, Hypothyroidism, COPD, Diabetes, Hypercholesterolemia, Hypertension, Depression.
- **Demographical and lifestyle conditions:** Blood-Pressure, Cholesterol, BMI, gender, date of birth, age at entry, smoking status, alcohol status and IMD Decile.



# Patient numbers

## IS dataset

### Full case dataset

( N = 25,711)

Cases = 8,983

Controls = 16,728

## TIA dataset

### Full case dataset

( N = 24,797)

Cases = 9,377

Controls = 15,420

### Multiple Imputation dataset

( N = 75,769)

Cases = 20,250

Controls = 55,519

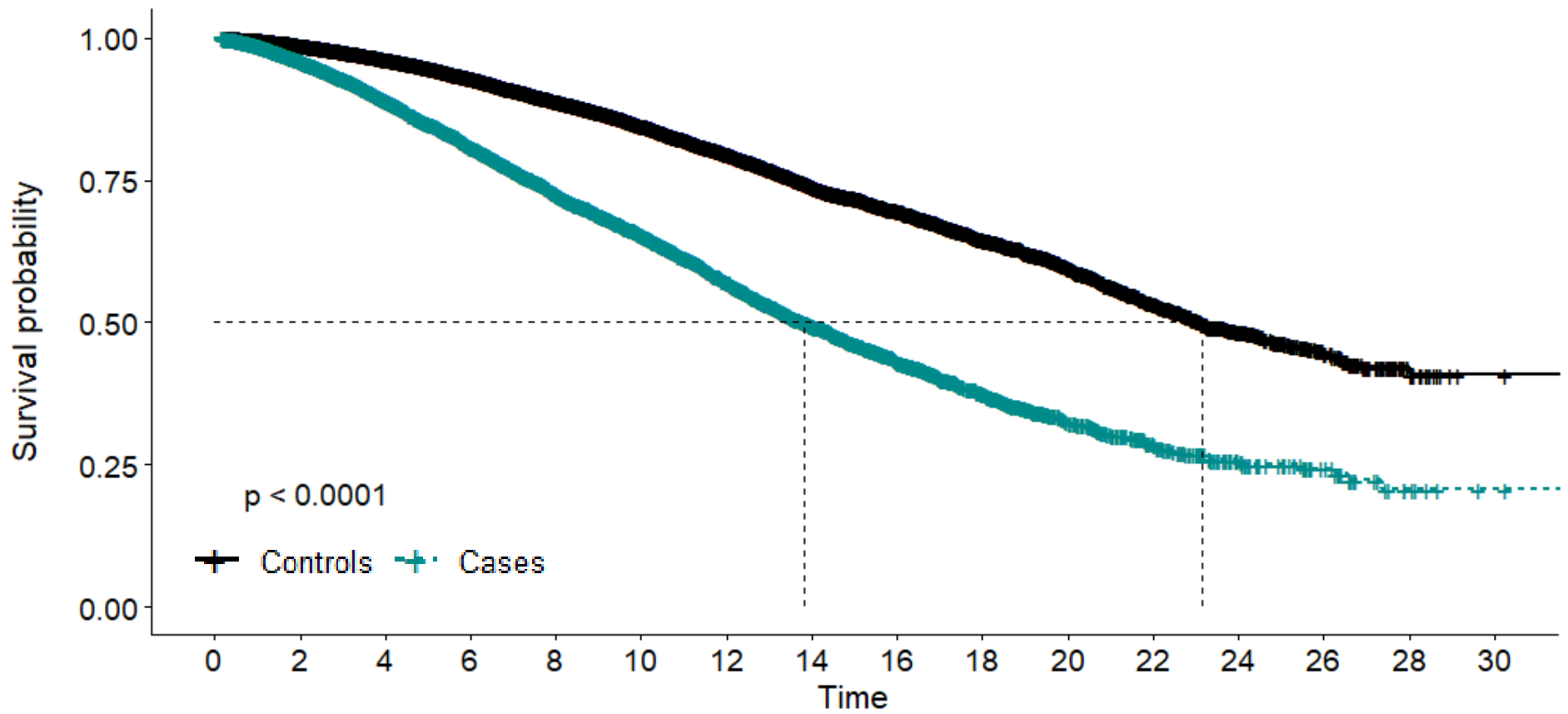
### Multiple Imputation dataset

( N = 74,037)

Cases = 20,633

Controls = 53,374

# Unadjusted Kaplan Meier plot IS cases and controls

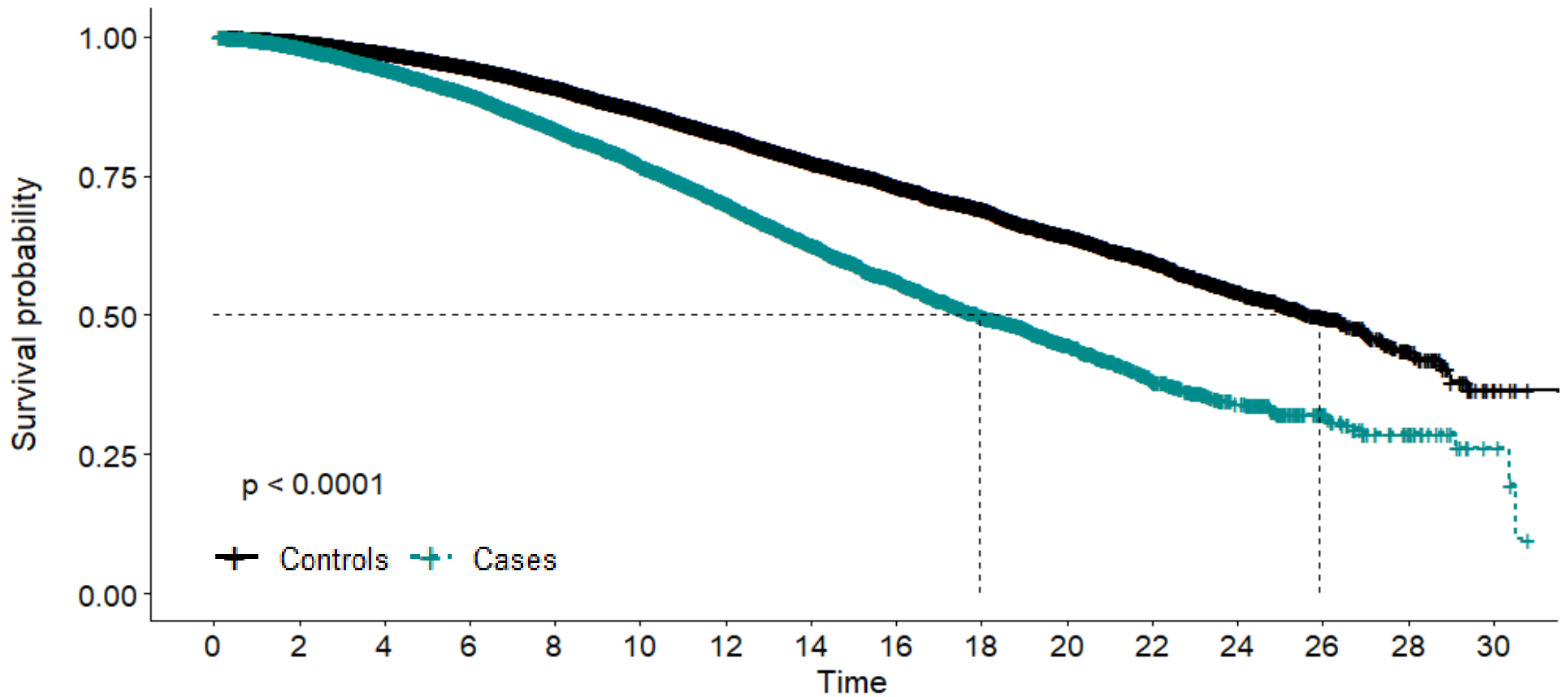


Number at risk

	0	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30
Controls	16728	15860	13956	11782	9535	7520	5712	4024	2672	1670	1034	607	317	133	31	4
Cases	8983	7997	6451	4963	3724	2748	1941	1255	786	456	235	131	63	29	8	2

Time

# Unadjusted Kaplan Meier plot TIA cases and controls



Number at risk

	0	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30
Controls	15420	14985	13796	12259	10591	8901	7143	5341	3807	2580	1640	1002	617	264	97	12
Cases	9377	8829	7812	6666	5478	4363	3307	2338	1515	929	565	295	159	60	25	5

Time



## Violations to Cox's Proportional hazard assumption ( $\alpha = 0.05$ )

	rho	chisq	p
birth_cohort1921 to 1930	-0.03057	4.6059	0.03186147087697
birth_cohort1931 to 1940	-0.04986	11.4182	0.00072727164999
birth_cohort1941 to 1960	-0.08358	32.0438	0.00000001507344
age_cat2	0.00991	0.4616	0.49688725441191
age_cat3	-0.01266	0.7385	0.39013103374282
age_cat4	-0.02933	3.9631	0.04650827842605
sexMale	0.00256	0.0294	0.86394990495609
groupscases	-0.03331	4.7344	0.02956567737229
IMD_Quintile2	0.01582	1.2685	0.26005731908003
IMD_Quintile3	0.00982	0.5247	0.46886479768015
IMD_Quintile4	0.02516	3.5033	0.06124845077876
IMD_Quintile5	0.02632	4.0789	0.04342093793658
BMI_catObese	0.03568	5.6952	0.01701099437364
BMI_catOverweight	0.01743	1.3773	0.24055963462973
BMI_catUnderweight	-0.01279	0.7491	0.38677422168589
⋮	⋮	⋮	⋮
AF_factorYes and Treated	-0.00210	0.0205	0.88620238544372
AF_factorYes and Untreated	0.01561	1.1125	0.29153432590237
hypertension_factorYes and Treated	-0.04421	8.7999	0.00301247436989
hypertension_factorYes and Untreated	-0.00589	0.1542	0.69459488132253
age_cat2:groupscases	-0.01262	0.6992	0.40306144688028
age_cat3:groupscases	-0.00481	0.1024	0.74901958328828
age_cat4:groupscases	0.00344	0.0526	0.81864457985083
sexMale:groupscases	0.00573	0.1442	0.70412354400852
groupscases:hypertension_factorYes and Treated	0.02637	3.1294	0.07689249209760
groupscases:hypertension_factorYes and Untreated	0.01609	1.1510	0.28334103952037
GLOBAL	NA	134.1296	0.00000000000228

### Comments

Covariates violating the Cox's PH assumption :  
Birth cohort, Age category, case/control, BMI, IMD, hypertension and antiplatelet.

The global test was highly significant providing evidence of non-proportionality.

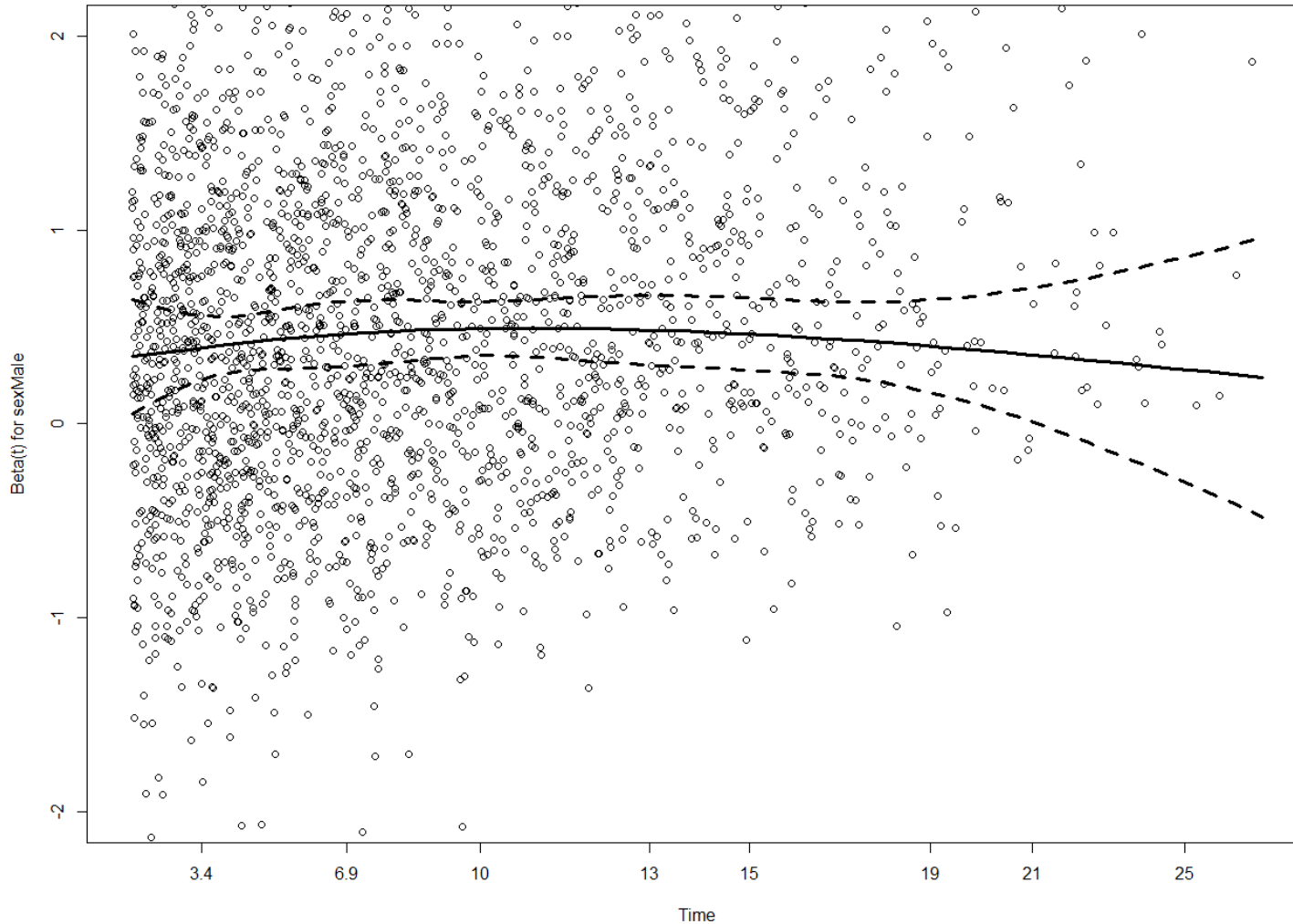


Actuarial  
Research Centre

Institute and Faculty  
of Actuaries

# Graphical diagnostics based on the scaled Schoenfeld residuals.

## A covariate which does not violate the Cox's PH assumption



### Comments:

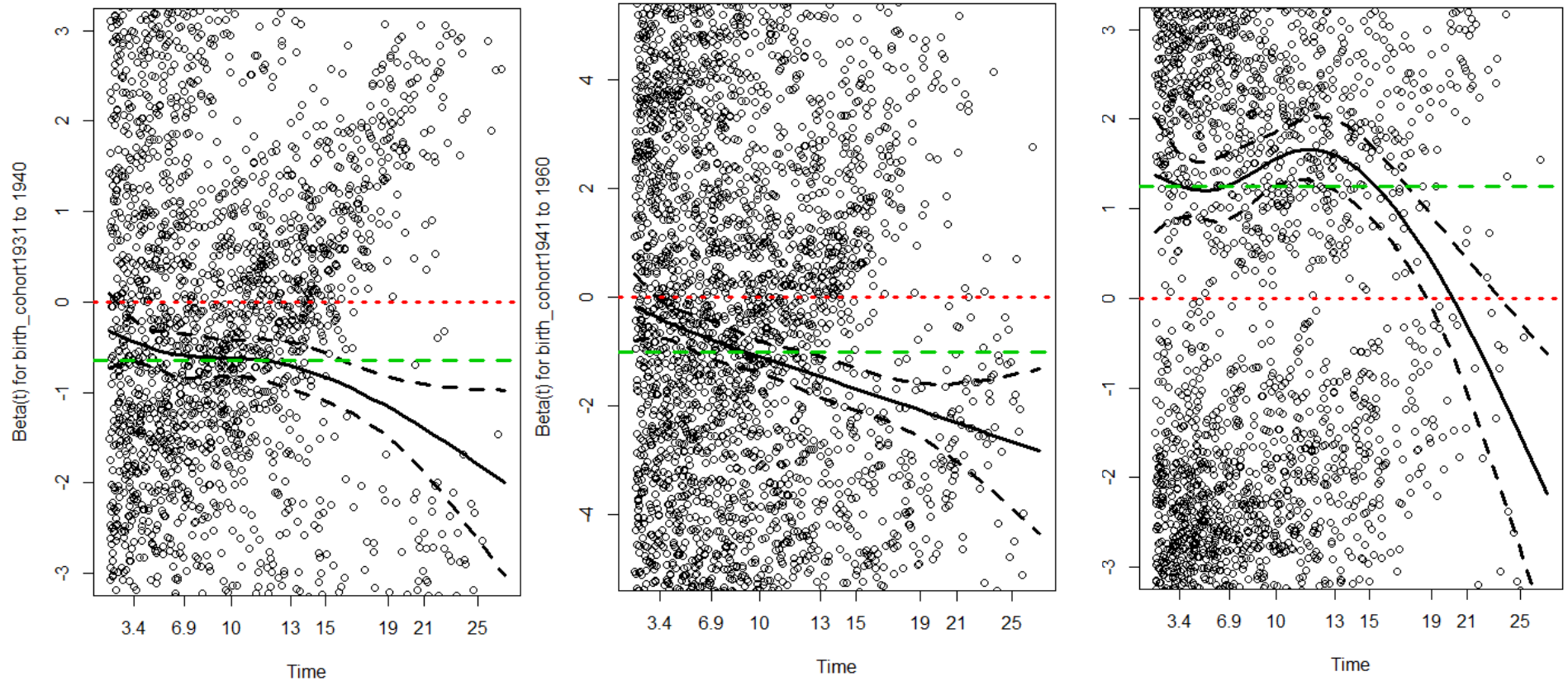
There is no distinct pattern of the residuals with time, so this covariate is not time-dependent.



Actuarial  
Research Centre

Institute and Faculty  
of Actuaries

# Graphical diagnostics based on the scaled Schoenfeld residuals.



- ..... Reference line for null effect
- - - Average hazard over time
- Time-varying hazard

## Comments:

A non-zero slope is evidence against proportionality.

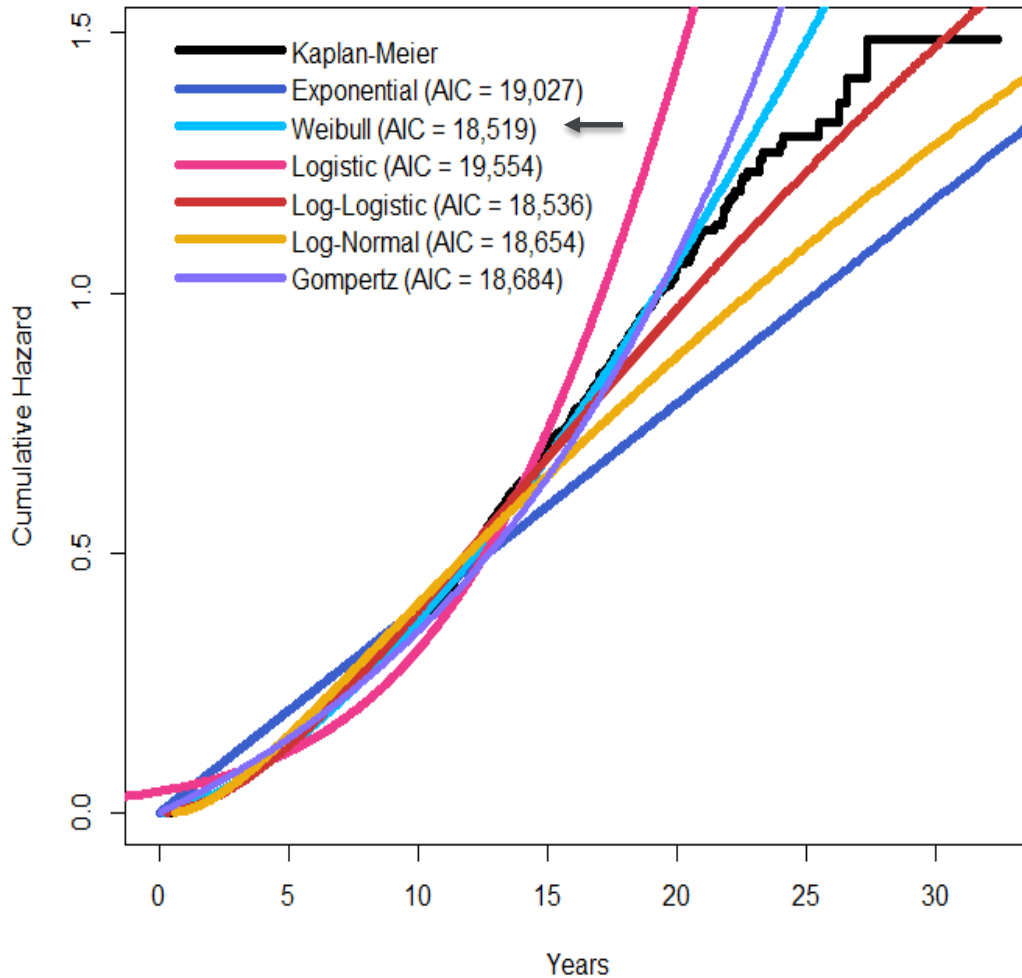


Actuarial  
Research Centre

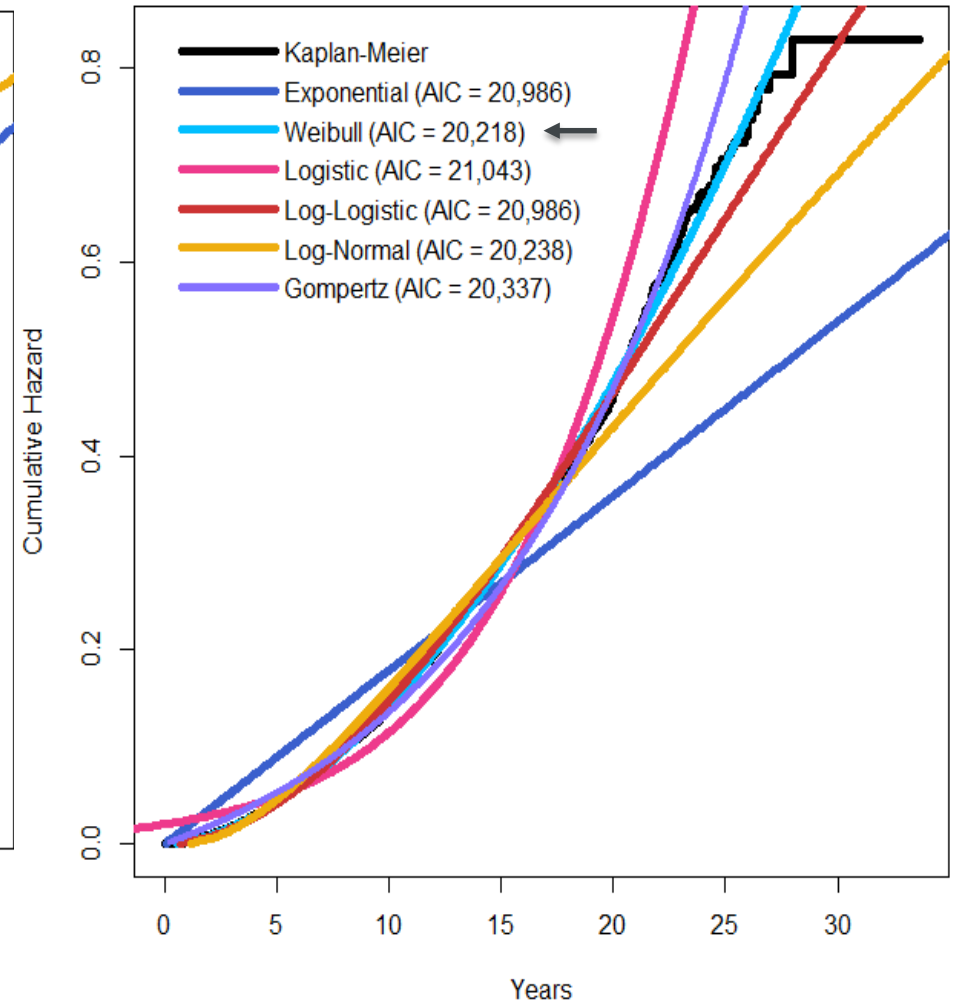
Institute and Faculty  
of Actuaries

# Distribution fitting : IS

Parametric Regression Fits  
IS cases



Parametric Regression Fits  
IS controls



# Parametric “Double-Cox” regression

Components:

- A baseline hazard function (which **changes over time**).
- The risk factors  $Z$  have a log-linear contribution to the force of mortality which does not depend on time  $t$ .

The Cox parametric regression model

$$\mu(t|Z) = \mu_0(t|Z) \exp(Z^T \beta)$$

Baseline hazard function

$\beta$  is a vector of unknown parameters for scale and  $Z$  is a vector of covariates

Weibull baseline hazard function with scale  $\lambda$  and shape  $k$ . Shape  $k$  is modelled as  $k=k(Z)$ .

$$\mu_0(t|Z) = \frac{k(Z)}{\lambda} \left(\frac{t}{\lambda}\right)^{k(Z)-1}$$

$$k(Z) = k_0 e^{Z^T \beta_k}$$

Additional regression model to allow varying shape depending on covariates



Actuarial  
Research Centre

Institute and Faculty  
of Actuaries

# IS model

The scale model includes the following main effects and interactions :

- Birth cohort
- IMD in Quintiles
- Body Mass index
- Antiplatelet therapy
- Chronic Pulmonary Disorder
- Chronic Kidney Disease ( stages 1-3)
- Heart Failure
- Myocardial Infarction
- Peripheral Arterial Disease
- Atrial fibrillation
- Diabetes
- Anticoagulant therapy
- Smoking
- Interaction of IS diagnosis with Antihypertensive treatment
- Interaction of IS diagnosis with sex
- Interaction of IS diagnosis with age

The shape model includes the following main effects :

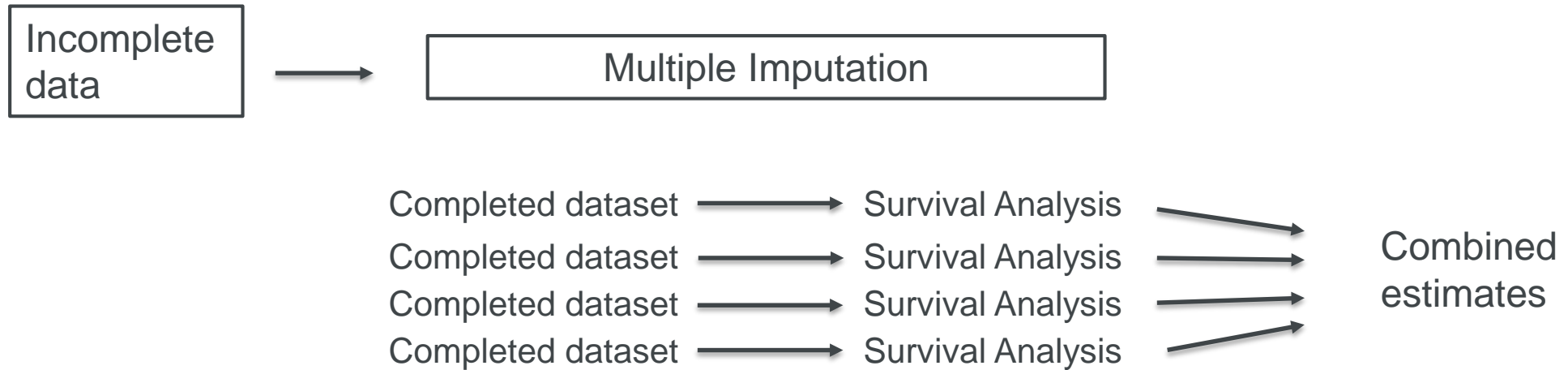
- Birth cohort
- Antiplatelet therapy



**Actuarial  
Research Centre**

Institute and Faculty  
of Actuaries

# Overview of Multiple Imputation



**Step 1 :**  
*Generate multiple sets of imputed values to produce multiple imputed datasets.*

**Step 2 :**  
*Perform survival analysis on each dataset.*

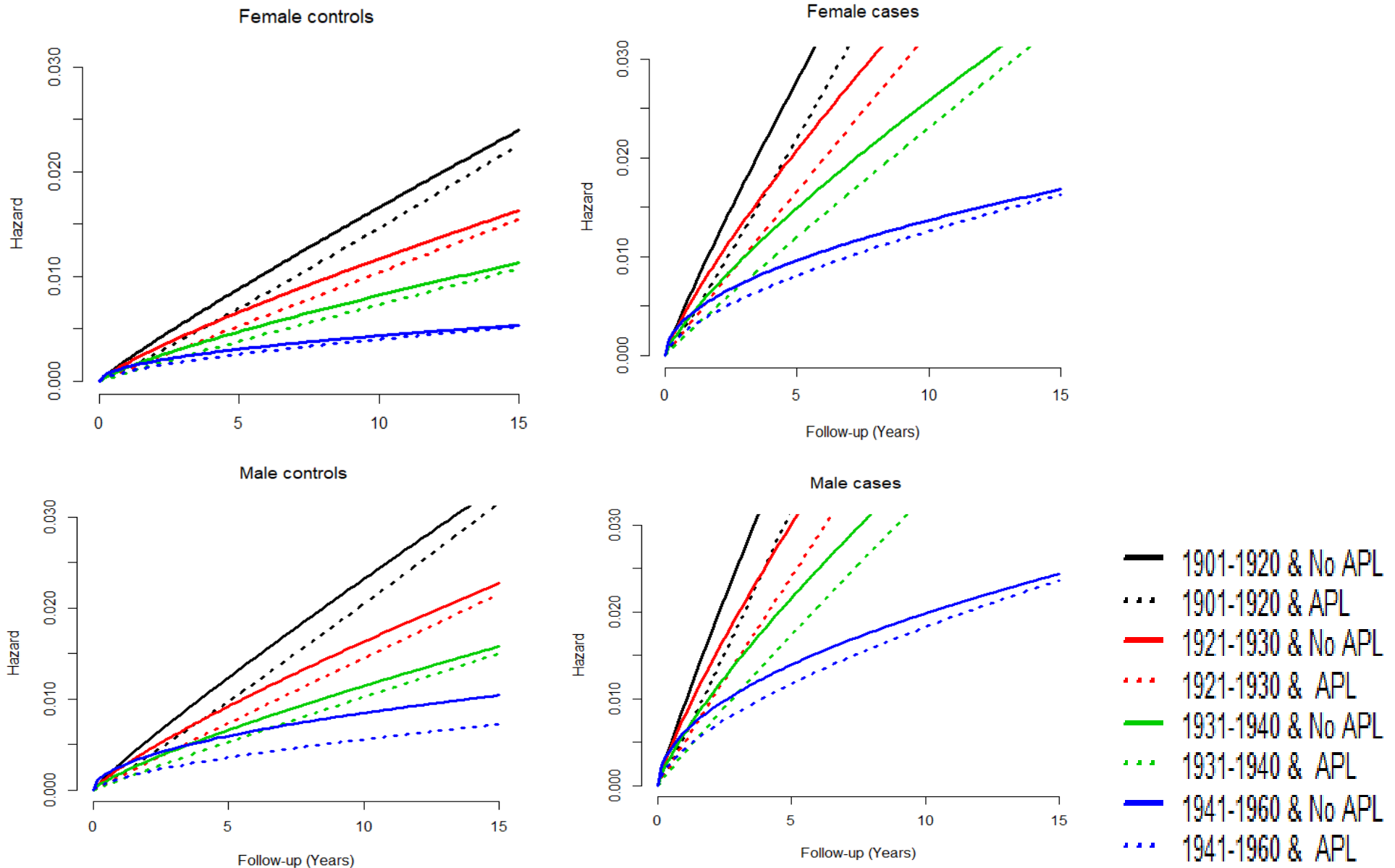
**Step 3 :**  
*Pool the results using Rubin's rules.*



**Actuarial  
Research Centre**

Institute and Faculty  
of Actuaries

# Hazard curves demonstrating the birth cohort effect : IS



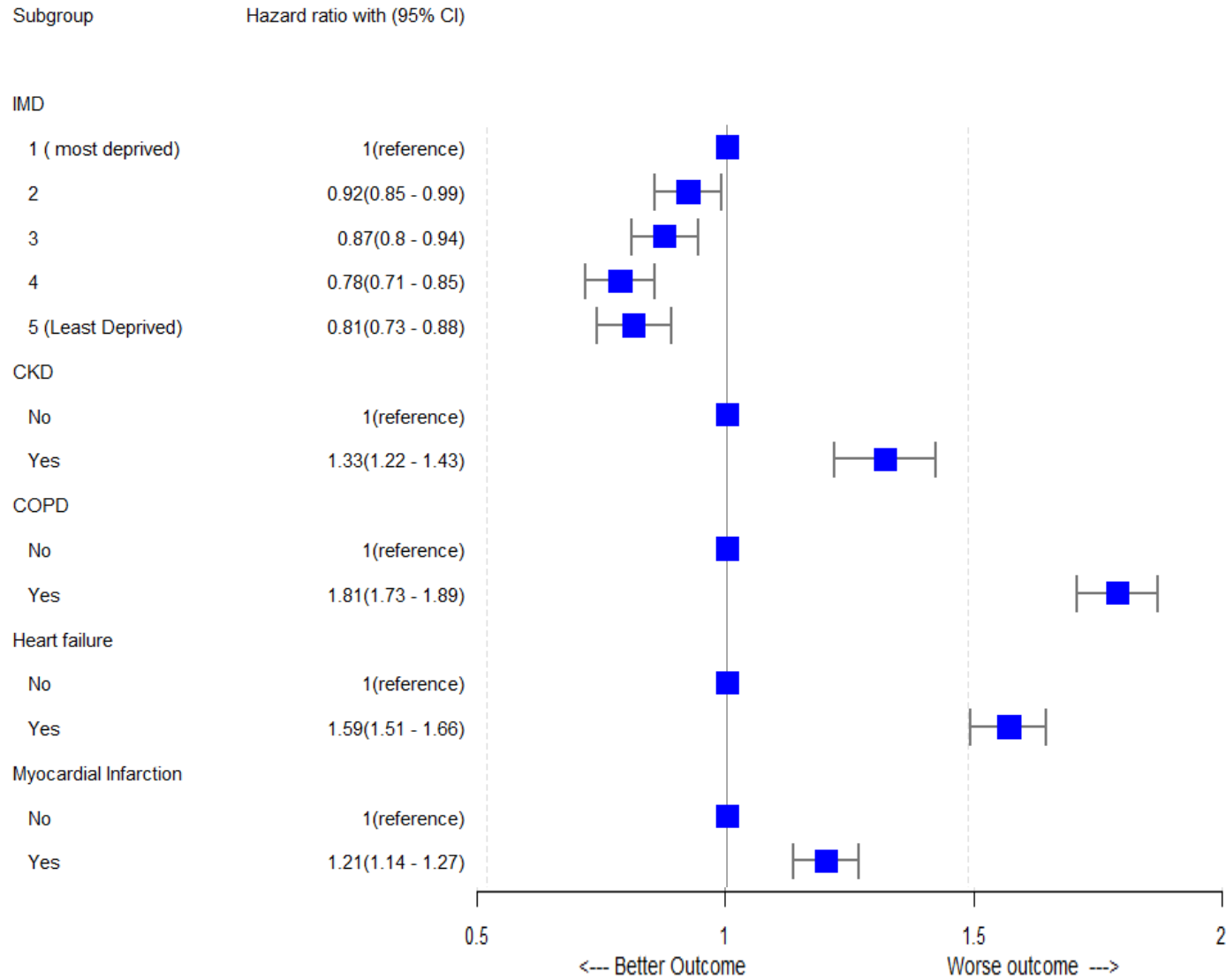
30/10/2019

Hazard curves for healthy cases and controls, aged 39-60 years with IMD Quintile = 1 across different birth cohorts and APL( antiplatelet intake).



# Forest plot : IS

## Adjusted Hazard Ratio

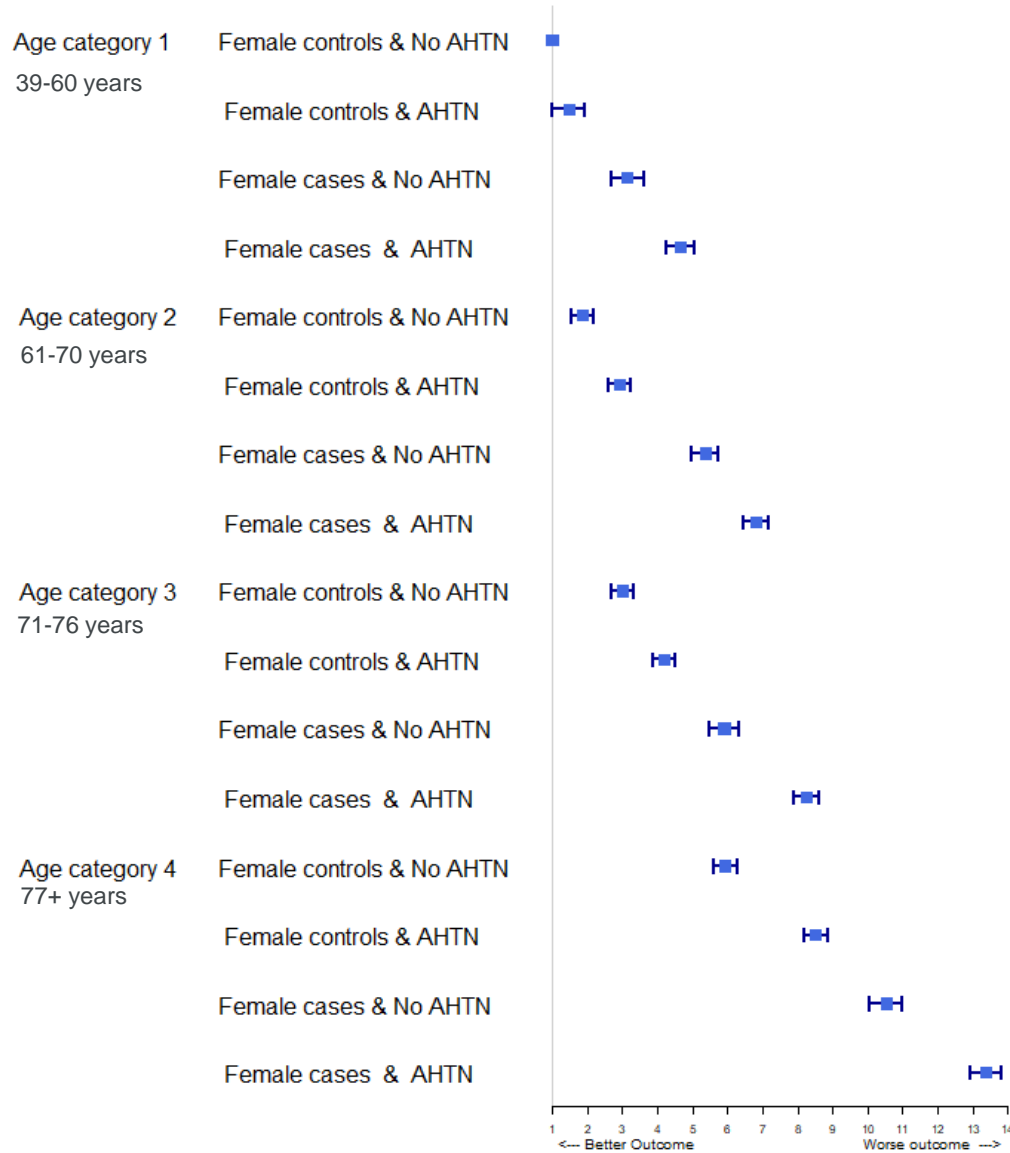


**Actuarial  
Research Centre**

Institute and Faculty  
of Actuaries

# Forest plot : IS model

## Adjusted Hazard ratios



**Actuarial  
Research Centre**

Institute and Faculty  
of Actuaries

# Future Works:

- Write up two papers: on TIA and on IS
- Translation of models into actuarial analysis



**Actuarial  
Research Centre**

Institute and Faculty  
of Actuaries



**Actuarial  
Research Centre**

Institute and Faculty  
of Actuaries

# The Actuarial Research Centre (ARC)

## A gateway to global actuarial research

The Actuarial Research Centre (ARC) is the Institute and Faculty of Actuaries' (IFoA) network of actuarial researchers around the world.

The ARC seeks to deliver cutting-edge research programmes that address some of the significant, global challenges in actuarial science, through a partnership of the actuarial profession, the academic community and practitioners.

The **'Use of Big Health and Actuarial Data for understanding Longevity and Morbidity Risks'** research programme is being funded by the ARC.

[www.actuaries.org.uk/arc](http://www.actuaries.org.uk/arc)



**Actuarial  
Research Centre**

Institute and Faculty  
of Actuaries

# Impact of Diabetes Mellitus II on Longevity and Morbidity Risks: Full Case Analysis

Njabulo Ncube (PhD Candidate)  
University of East Anglia (UK)



## Actuarial Research Centre

Institute and Faculty  
of Actuaries

### Disclaimer

The views expressed in this presentation are those of invited contributors and not necessarily those of the Institute and Faculty of Actuaries. The Institute and Faculty of Actuaries does not endorse any of the views stated, nor any claims or representations made in this presentation and accept no responsibility or liability to any person for loss or damage suffered as a consequence of their placing reliance upon any view, claim or representation made in this presentation. The information and expressions of opinion contained in this presentation are not intended to be a comprehensive study, nor to provide actuarial advice or advice of any nature and should not be treated as a substitute for specific advice concerning individual situations. On no account may any part of this presentation be reproduced without the written permission of the Institute and Faculty of Actuaries .

# Presentation Outline

## □ Introduction

- Purpose of the Study
- Why Diabetes Mellitus II?

## □ Study Design

- Selection Criteria
- Study Sample
- Statistical Models

## □ Results

## □ Further Data Modelling



# Introduction

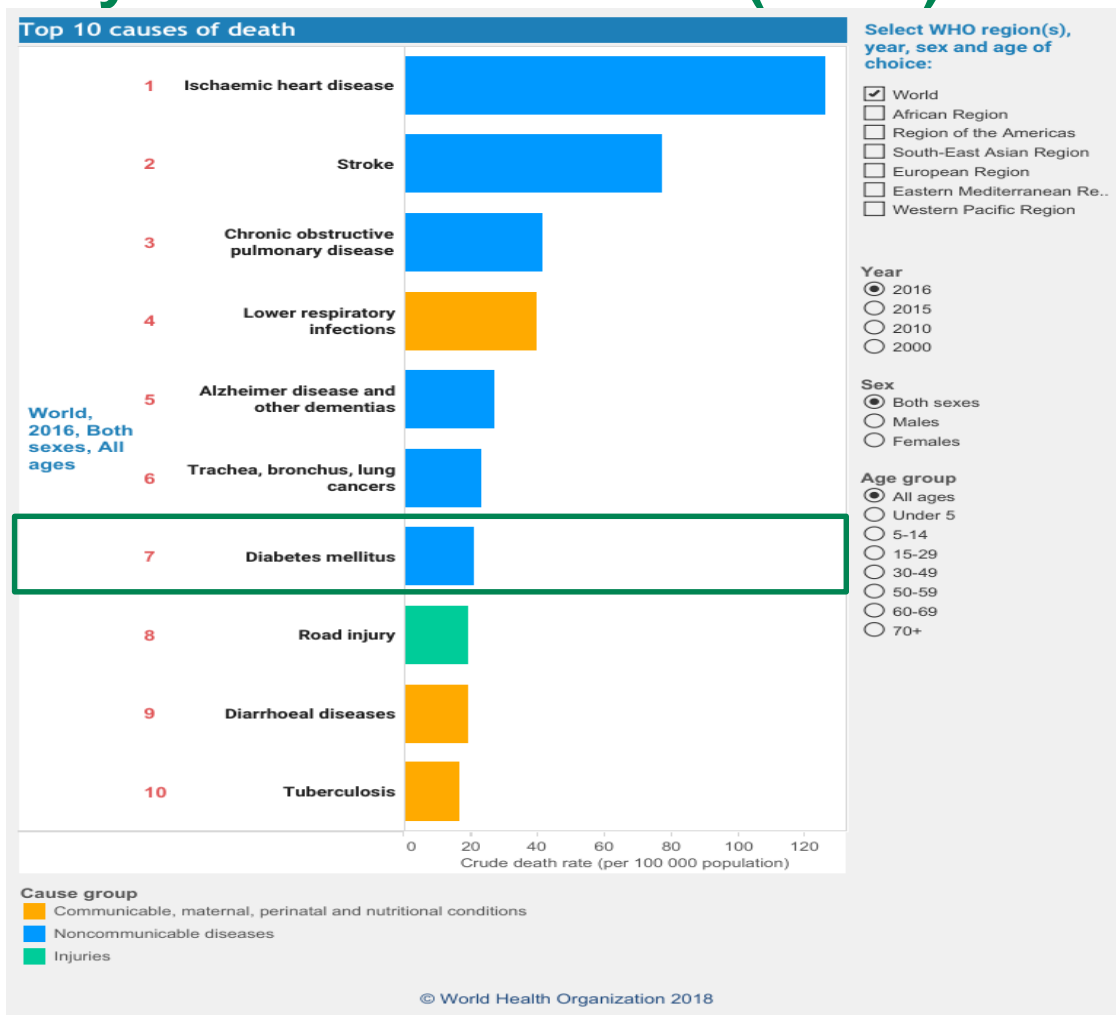
## Purpose of the Study

- ❑ To derive, analyse and model the impact of diabetes mellitus II (DM-II) on longevity and morbidity risks.
- ❑ Primary Outcome: all-cause mortality.
- ❑ Secondary Outcomes: amputation, cognitive impairment, Chronic Kidney Disease (CKD) Stages 3 to 5, heart failure (HF), myocardial infarction (MI), pulmonary vascular disease (PVD), stroke, cancer and cognitive impairment including dementia.





# Why Diabetes Mellitus II (DM-II)



Source: WHO (2018)

## DM-II: Rankings among the Top Ten

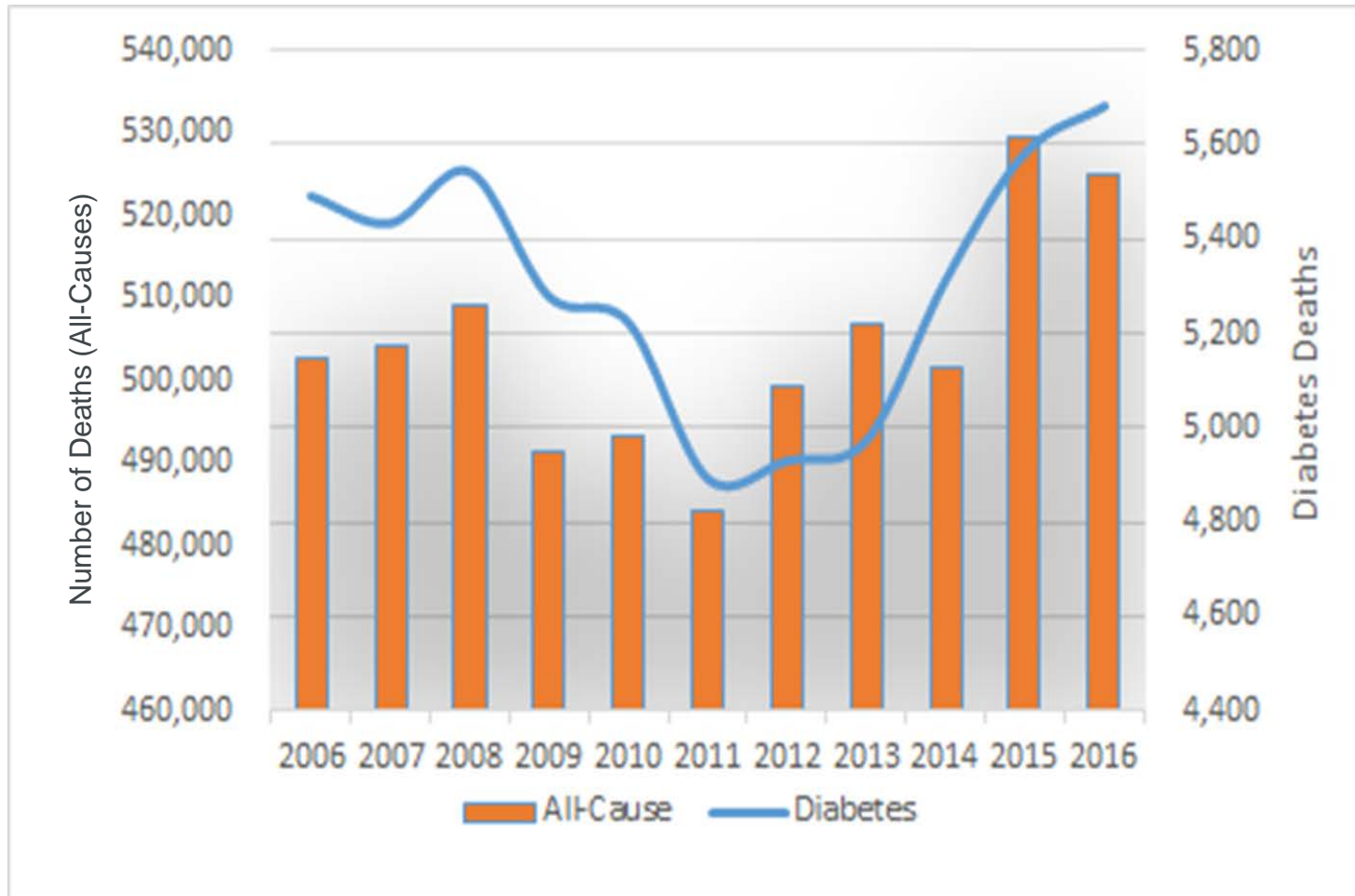
Year	All Ages	50-59	60-69	70+
2016	7	6	5	6
2015	7	6	5	6
2010	10	7	5	7
2000	15	9	6	7



**Actuarial Research Centre**

Institute and Faculty of Actuaries

# Why DM-II (cntd.)



Source: ONS (2017)



Actuarial  
Research Centre  
Institute and Faculty  
of Actuaries

# Study Design

## Selection Criteria

- ❑ UK THIN database.
- Patients diagnosed with DM-II (cases) from 1984 and, aged 40 years and above were matched (1:3) to non – diabetics (controls) by practice, age and sex.
- Excluded patients with severe medical conditions diagnosed (e.g. cancer) before entry date.
- The follow up period is from 1984 up to 2017.



# Study Design

## ☐ Variables of Interest – at entry

Demographic
1. Age Group
2. Birth Year
3. Gender
4. General Practice (Frailty)

Life Style and Socio-economic
1. Smoking Status
2. Townsend Deprivation Index
3. Body Mass Index (BMI)

Medical Conditions
1. Case-Control Indicator
2. Angina
3. Atrial Fibrillation (AF)
4. HF
5. Hypercholesterolemia
6. Hypertension
7. MI
8. PVD



Interactions e.g. Age Group and Gender, Case-Control and Smoking status



**Actuarial  
Research Centre**

Institute and Faculty  
of Actuaries

# Study Design

## Full Case Analysis - Selection Criteria

Included Patients with complete records on

- Smoking status,
- Alcohol consumption status,
- Townsend deprivation score,
- BMI,
- Blood Pressure (BP),
- Blood lipid ratio and
- High-density lipoproteins (HDL).



**Actuarial  
Research Centre**

Institute and Faculty  
of Actuaries

# Study Sample

## Total Study Sample

- 108 282 (57% Males) Cases.
- 253 800 (55% Males) Controls.

## Full Case Study Sample

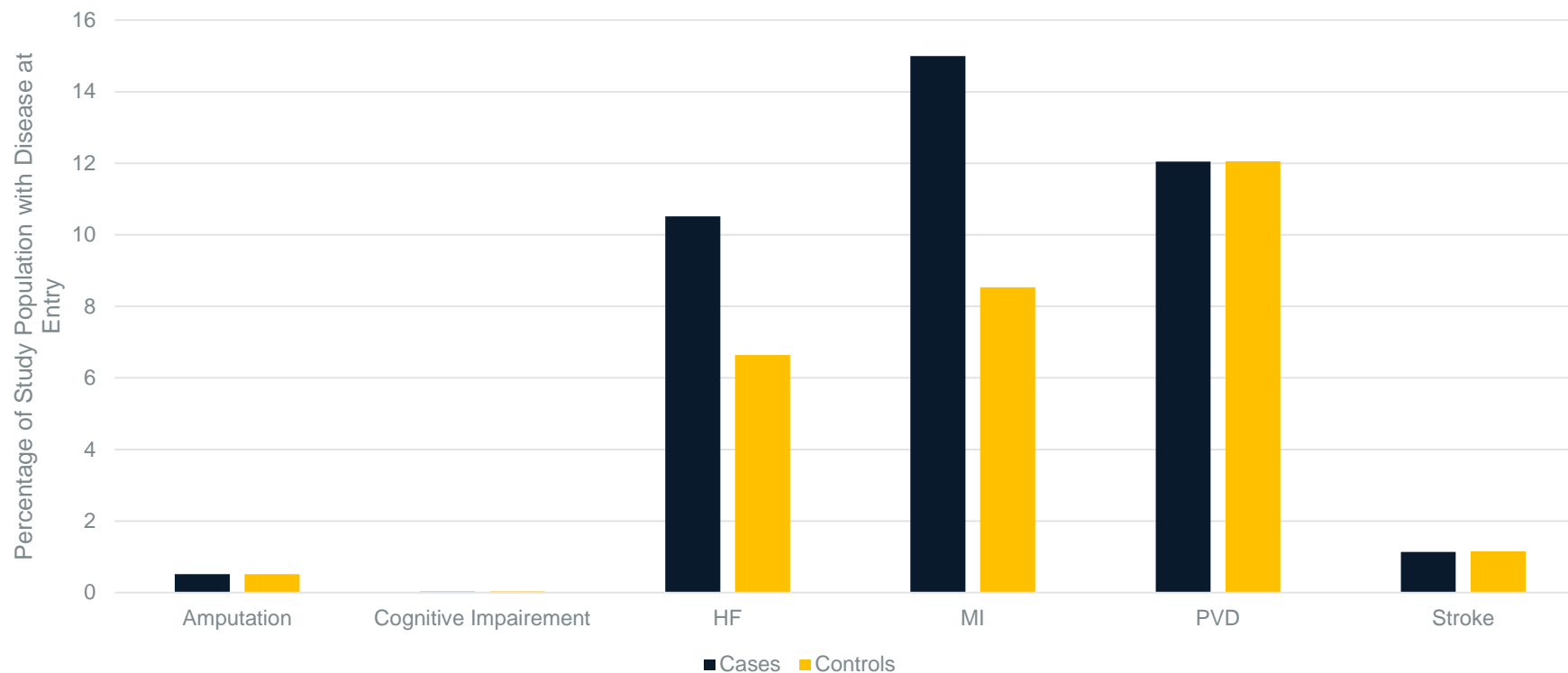
- 20 213 (57.7% Males) Cases.
- 28 693 (56.2% Males) Controls.

## Distribution of the Study Sample by Age Group, Sex and Case-Control Status



**Actuarial Research Centre**  
 Institute and Faculty of Actuaries

## Prevalence of Some Medical Conditions at Entry Date



<b>Cases</b>	560 (0.52%)	27 (0.02%)	11,388 (10.52%)	16,242 (15.00%)	13,048 (12.05%)	1,230 (1.14%)
<b>Controls</b>	1,317 (0.52%)	73 (0.03%)	16,857 (6.64%)	21,665 (8.54%)	30,595 (12.05%)	2,920 (1.15%)



**Actuarial  
Research Centre**

Institute and Faculty  
of Actuaries

# Statistical Models for All-Cause Mortality

- Cox Regression for DM – II

Backward elimination was used for variable selection ( $\alpha_{main} = 0.05$ ,  $\alpha_{interactions} = 0.01$ )



- Case-control indicator,
  - Age group,
  - Birth Year,
  - Gender,
  - Smoking status,
  - Townsend deprivation index,
  - HF,
  - Hypercholesterolemia,
  - Hypertension,
  - MI,
  - PVD,
  - BMI
- and interactions





# Assessing PH Assumption ( $\alpha = 0.05$ )

	rho	chisq	p
Case-Control [Cases]			
Age Group [50-59]	0.002855	0.042302	0.837045
Age Group [60+]	-0.01454	1.12216	0.289454
Birth Year [1930-1939]	0.048563	13.64601	0.000221
Birth Year [1940-1949]	0.027559	4.442733	0.03505
Gender [Male]	-0.00143	0.010588	0.918042
Smokes [Former]	-0.00845	0.3666	0.544863
Smokes [Smoker]	-0.02358	2.866207	0.090458
Townsend [Less Deprived]	-0.01738	1.583886	0.208202
Townsend [2]	-0.01299	0.872694	0.350211
Townsend [4]	-0.01032	0.554632	0.456431
Townsend [Most Deprived]	-0.00078	0.00316	0.955175
HF [Yes]	-0.00032	0.000549	0.981302
Hypercholesterolemia [Treated]	0.033125	5.858287	0.015504
Hypercholesterolemia [Untreated]	-0.03202	5.509689	0.018911
Hypertension [Treated]	0.036188	6.896064	0.008639
Hypertension [Untreated]	-0.01671	1.454762	0.227765
MI [Yes]	-0.01977	2.116238	0.145744
PVD [Yes]	-0.00438	0.099198	0.752794
BMI [Overweight]	-0.01086	0.605801	0.436373
BMI [Obese]	-0.00255	0.033734	0.854273
Case-Control [Cases]:Smokes [Former]	0.025331	3.357494	0.0669
Case-Control [Cases]:Smokes [Smoker]	0.02132	2.33379	0.126593
Case-Control [Cases]:Hypercholesterolemia [Treated]	-0.03117	5.062485	0.024449
Case-Control [Cases]:Hypercholesterolemia [Untreated]	-0.01483	1.147207	0.284135
Case-Control [Cases]:MI [Yes]	0.018921	1.870298	0.171441
Case-Control [Cases]:PVD [Yes]	0.005052	0.132072	0.716293
Case-Control [Cases]:BMI [Overweight]	-0.01035	0.555738	0.455983
Case-Control [Cases]:BMI [Obese]	-0.02668	3.697196	0.054504
Age Group [50-59]:Gender [Male]	0.004786	0.118789	0.730352
Age Group [60+]:Gender [Male]	0.013832	1.006281	0.315796
Birth Year [1930-1939]:Gender [Male]	-0.02832	4.510593	0.033686
Birth Year [1940-1949]:Gender [Male]	-0.01775	1.778759	0.182302
Smokes [Former]:BMI [Overweight]	-0.00518	0.138134	0.710143
Smokes [Smoker]:BMI [Overweight]	0.005923	0.18299	0.668816
Smokes [Former]:BMI [Obese]	-0.00536	0.149176	0.699324
Smokes [Smoker]:BMI [Obese]	-0.00344	0.060556	0.805619
Townsend [Less Deprived]:BMI [Overweight]	0.002167	0.024508	0.8756
Townsend [2]:BMI [Overweight]	0.00821	0.349348	0.554483
Townsend [4]:BMI [Overweight]	0.006623	0.228463	0.632666
Townsend [Most Deprived]:BMI [Overweight]	0.002618	0.035617	0.850308
Townsend [Less Deprived]:BMI [Obese]	0.017672	1.63206	0.201418
Townsend [2]:BMI [Obese]	0.01169	0.707889	0.400146
Townsend [4]:BMI [Obese]	0.010126	0.53434	0.464788
Townsend [Most Deprived]:BMI [Obese]	0.008117	0.341958	0.558701
GLOBAL	NA	103.2145	0.0000018

## Variables violating the PH Assumption

- Year of Birth
- Hypercholesterolemia
- Hypertension



Actuarial  
Research Centre

Institute and Faculty  
of Actuaries

# Validating PH Assumption results using timecox ( $\alpha = 0.05$ )

## Test for Time Invariant Effects

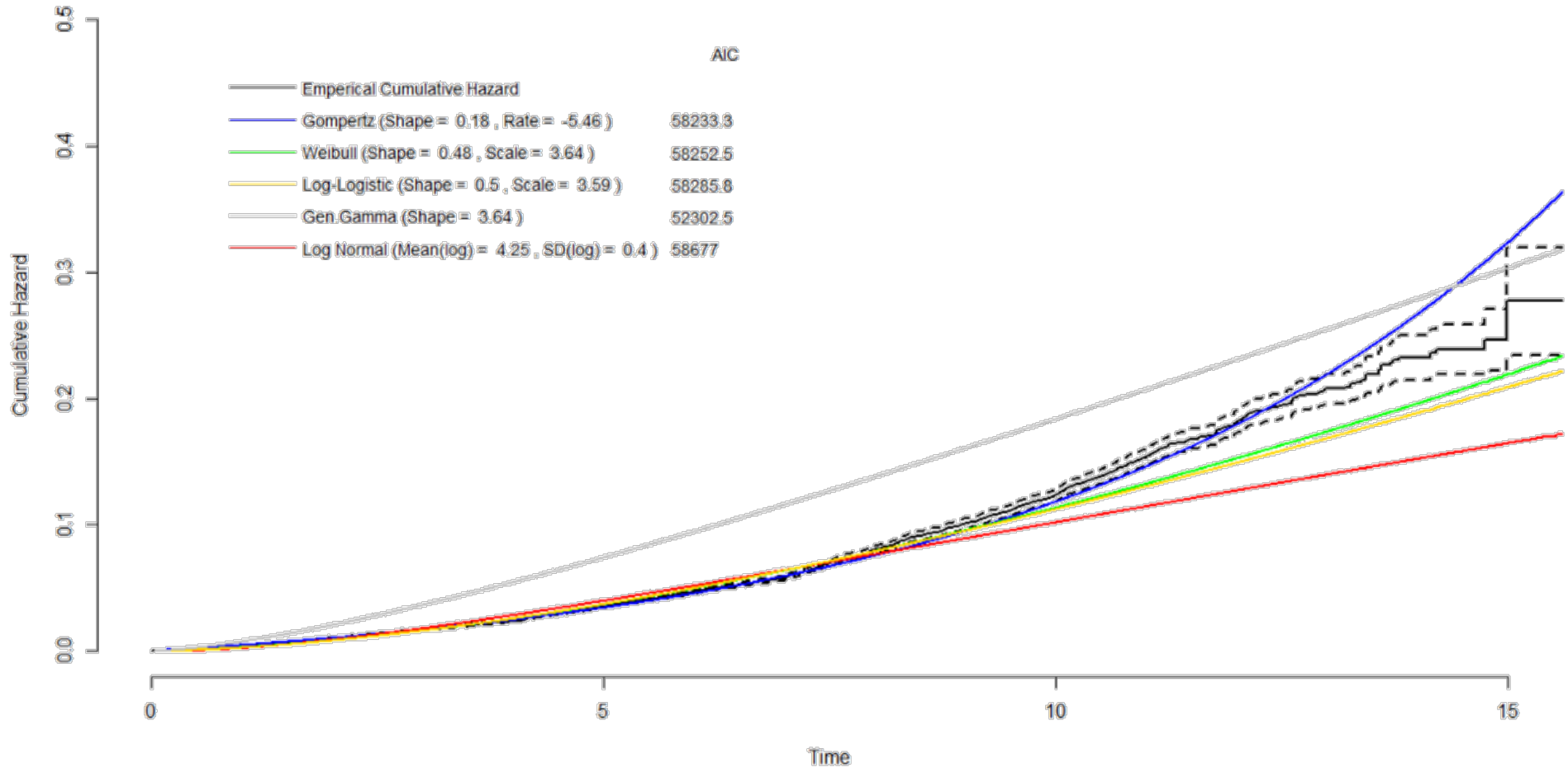
	Kolmogorov-Sminorv Test	p-value: $H_0: \beta(t) = \beta$
Intercept	2.72	0.207
Birth Year [1930-1939]	2.99	0.25
Birth Year [1940-1949]	2.53	0.217
Hypercholesterolemia [Treated]	2.52	0.735
Hypercholesterolemia [Untreated]	3.16	0.029
Hypertension [Treated]	4.9	0.159
Hypertension [Untreated]	2.31	0.558
Birth Year [1930-1939]:const(Gender) [Male]	5.2	0.127
Birth Year [1940-1949]:const(Gender) [Male]	4.5	0.324

**Only hypercholesterolemia has time variant effects**



Actuarial  
Research Centre  
Institute and Faculty  
of Actuaries

# Estimating the Baseline Function using flexsurvreg package



Actuarial  
Research Centre

Institute and Faculty  
of Actuaries

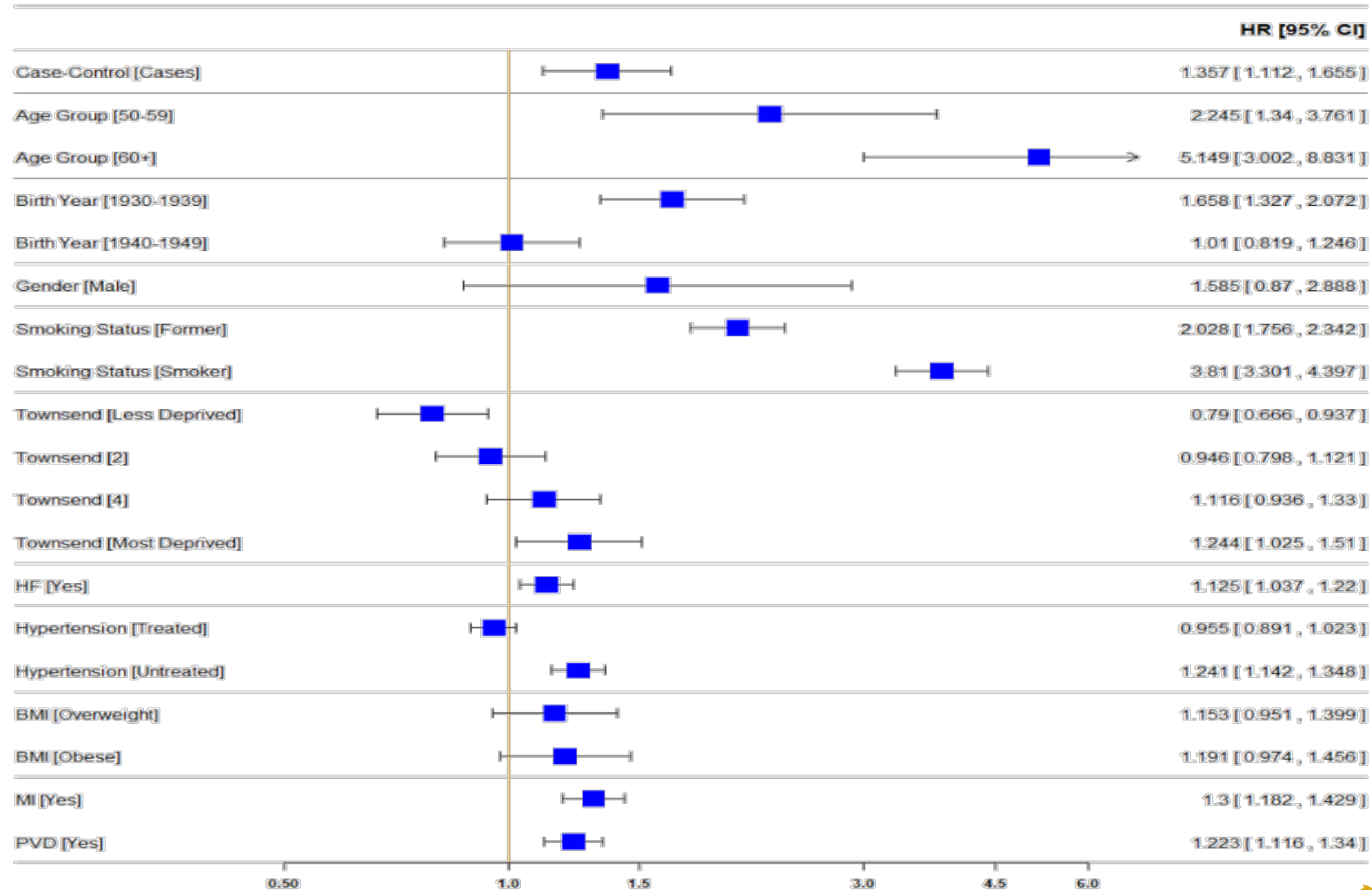
# Gompertz-Cox Regression

- Distribution
  - Gompertz distribution.
- Shape Model
  - Hypercholesterolemia.
- Scale Model
  - All covariates and interactions as in Cox Model.



# Adjusted Hazard Ratios

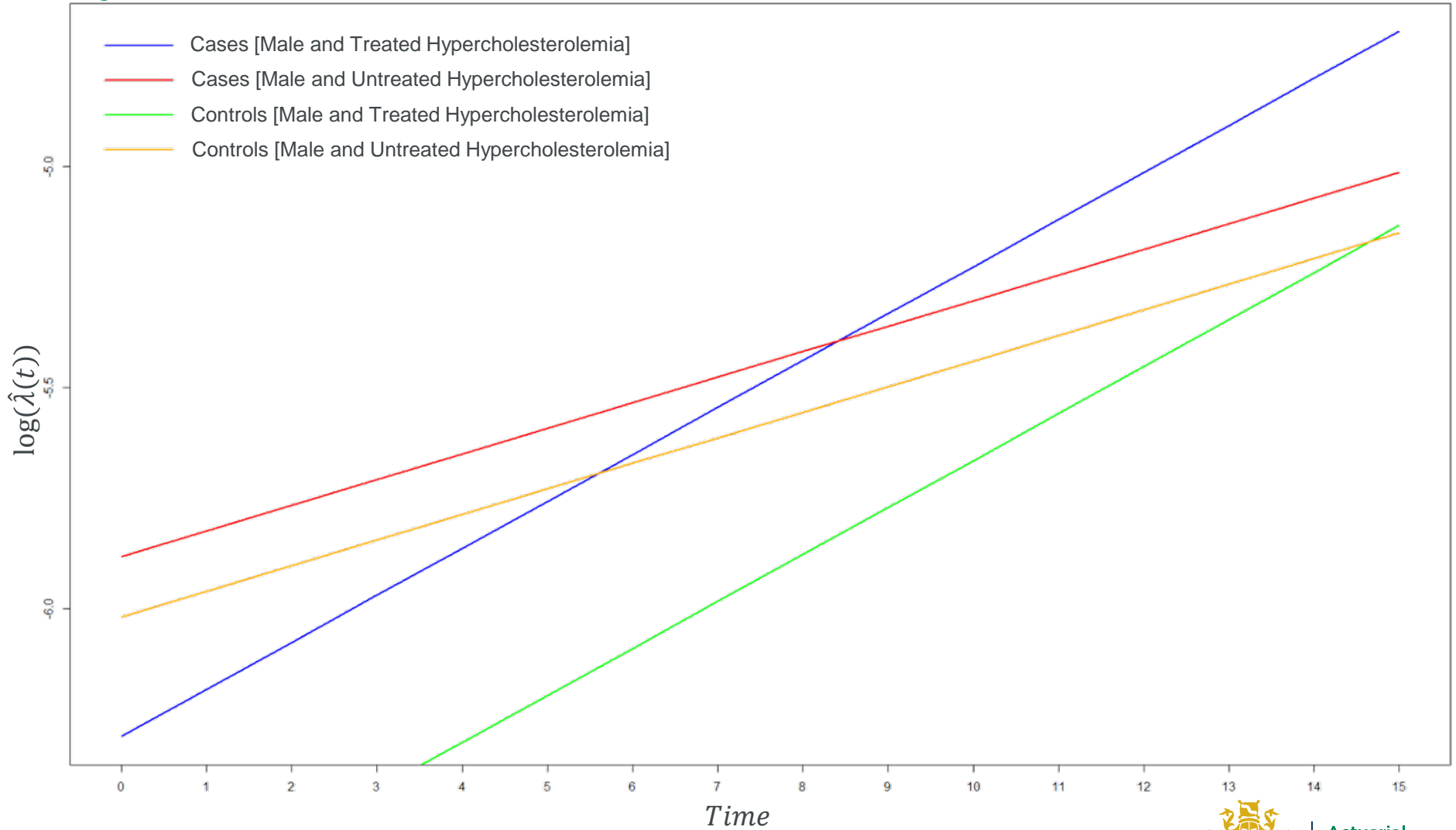
Adjusted Hazard Ratios for Scale Cox Model



Actuarial  
Research Centre

Institute and Faculty  
of Actuaries

# Adjusted Hazard Function



# Further Work

1. Imputed Data Model (Mortality)
2. Translation into Actuarial Models (Mortality)
3. Morbidity Models (Cancer, CKD Stages 3 – 5)
4. Translation into Actuarial Models
5. Publish at least 2 papers



# References

1. Begun A., Kulinskaya E. and MacGregor A. (2019). Risk-adjusted CUSUM control charts for shared frailty survival models with application to hip replacement outcomes: a study using the NJR dataset. *BMC Medical Research Methodology*, in print.
2. Martinussen T, Scheike T H. (2006), *Dynamic Regression Models for Survival Data*. Springer: New York.
3. WHO (2018). Top Ten Causes of Death.  
[https://www.who.int/gho/mortality\\_burden\\_disease/causes\\_death/top\\_10/en/](https://www.who.int/gho/mortality_burden_disease/causes_death/top_10/en/). Accessed: 23/10/2019
4. ONS (2017). Deaths in England and Wales.  
[www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/](http://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/). Accessed: 22/02/2018.







**Actuarial  
Research Centre**

Institute and Faculty  
of Actuaries

# The Actuarial Research Centre (ARC)

A gateway to global actuarial research

The Actuarial Research Centre (ARC) is the Institute and Faculty of Actuaries' (IFoA) network of actuarial researchers around the world.

The ARC seeks to deliver cutting-edge research programmes that address some of the significant, global challenges in actuarial science, through a partnership of the actuarial profession, the academic community and practitioners.



**Actuarial  
Research Centre**

Institute and Faculty  
of Actuaries



# Hormone Replacement Therapy and its effects on Morbidity and Longevity of Women

Nurunnahar Akter

School of Computing Sciences, University of East Anglia

The '**Use of Big Health and Actuarial Data for understanding Longevity and Morbidity Risks**' research programme is being funded by the Actuarial Research Centre.

October 30, 2019

[www.actuaries.org.uk/arc](http://www.actuaries.org.uk/arc)





## Actuarial Research Centre

Institute and Faculty  
of Actuaries

### Disclaimer

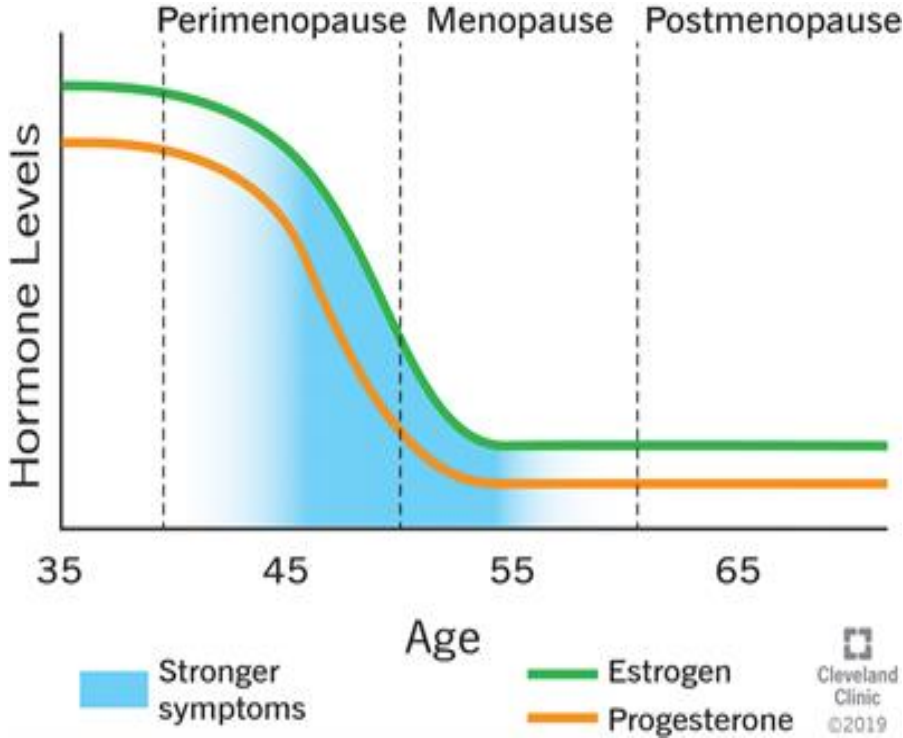
The views expressed in this presentation are those of invited contributors and not necessarily those of the Institute and Faculty of Actuaries. The Institute and Faculty of Actuaries does not endorse any of the views stated, nor any claims or representations made in this presentation and accept no responsibility or liability to any person for loss or damage suffered as a consequence of their placing reliance upon any view, claim or representation made in this presentation. The information and expressions of opinion contained in this presentation are not intended to be a comprehensive study, nor to provide actuarial advice or advice of any nature and should not be treated as a substitute for specific advice concerning individual situations. On no account may any part of this presentation be reproduced without the written permission of the Institute and Faculty of Actuaries .

# Outline

- Brief description of Hormone Replacement Therapy
- Study design and selection criteria
- Distribution of the study population
- Hazards of selected medical conditions at follow-up
- Complete case analysis
- Results



# Menopause and its Symptoms:



**MENOPAUSE Symptoms**

- HOT FLASHES CONTINUES FOR YEARS (Illustration of a woman with a fan)
- SLEEP PROBLEMS (Illustration of a woman sitting up in bed)
- MEMORY ISSUES (Illustration of a woman with a question mark over her head)
- EXTREME TIREDNESS (Illustration of a woman sitting on a pink sofa)
- WEIGHT GAIN (Illustration of a woman with a large belly)
- EXTREME SKIN DRYNESS (Illustration of a woman's face)
- BONE LOSS (Illustration of two vertebrae)
- MORE BAD HAIR DAYS (Illustration of a woman pulling at her hair)

THE YOGA ASSISTANT logo is in the bottom right corner of the infographic.



**Actuarial Research Centre**  
 Institute and Faculty of Actuaries

# Brief Description of Hormone Replacement Therapy (HRT)

## What is HRT?

- HRT is mainly used to relieve women from menopausal symptoms
- It has been used for more than sixty years
- HRT contains female sex hormones estrogen and/or progesterone
- First available in the United Kingdom in 1965

## Routes of Administration

- Oral tablets, transdermal patches, injections, topical gels, and ointments.



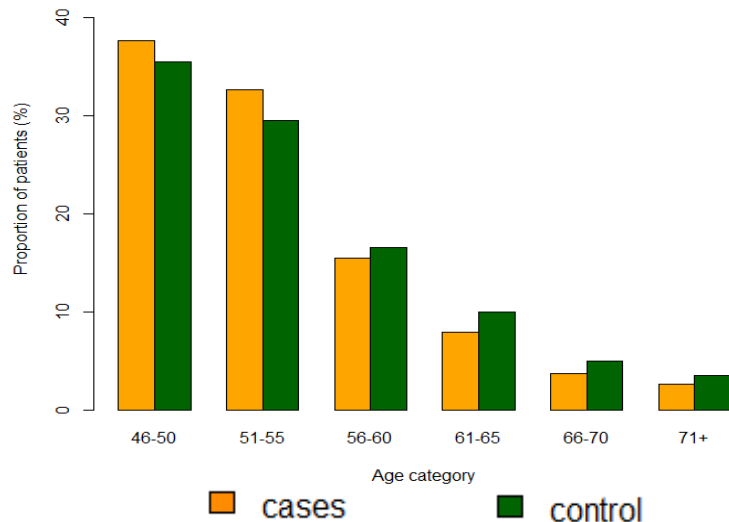
# Study design and patients selection criteria

- Cases are patients of age 46 years and above who received any kind of HRT.
- Controls are matched with cases by year of birth and general practice (GP).
- Patients with all kinds of cancer, acute myocardial infarction (AMI), serious heart failure, stroke (except TIA), chronic kidney disease (CKD) stage 3-5, dementia, oophorectomy before 45, premature ovarian insufficiency, premature menopause and surgical menopause are excluded.
- Primary outcome of interest is all-cause mortality. Secondary outcomes are osteoporosis, dementia, cardiovascular disease, type II diabetes, and hormonal cancers.
- **Follow up period between 1984 to 2017.**
- **Working data consists of 112,354 cases and 245,320 matched controls.**



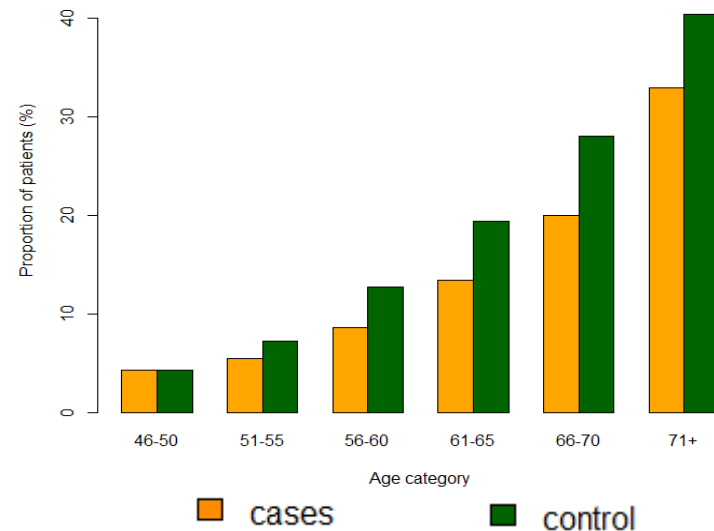
# Age distribution at first HRT prescription and death experience at follow-up

Proportion of Study Population by Age Category at Baseline



Age-group	46-50	51-55	56-60	61-65	66-70	71+
Cases	42269	36680	17362	8930	4185	2928
Controls	87125	72497	40688	24382	12160	8468

Death Experience of Study Population by Age Category



Age-group	46-50	51-55	56-60	61-65	66-70	71+
Cases	1809	1985	1498	1199	837	965
Controls	3747	5260	5166	4718	3414	3423

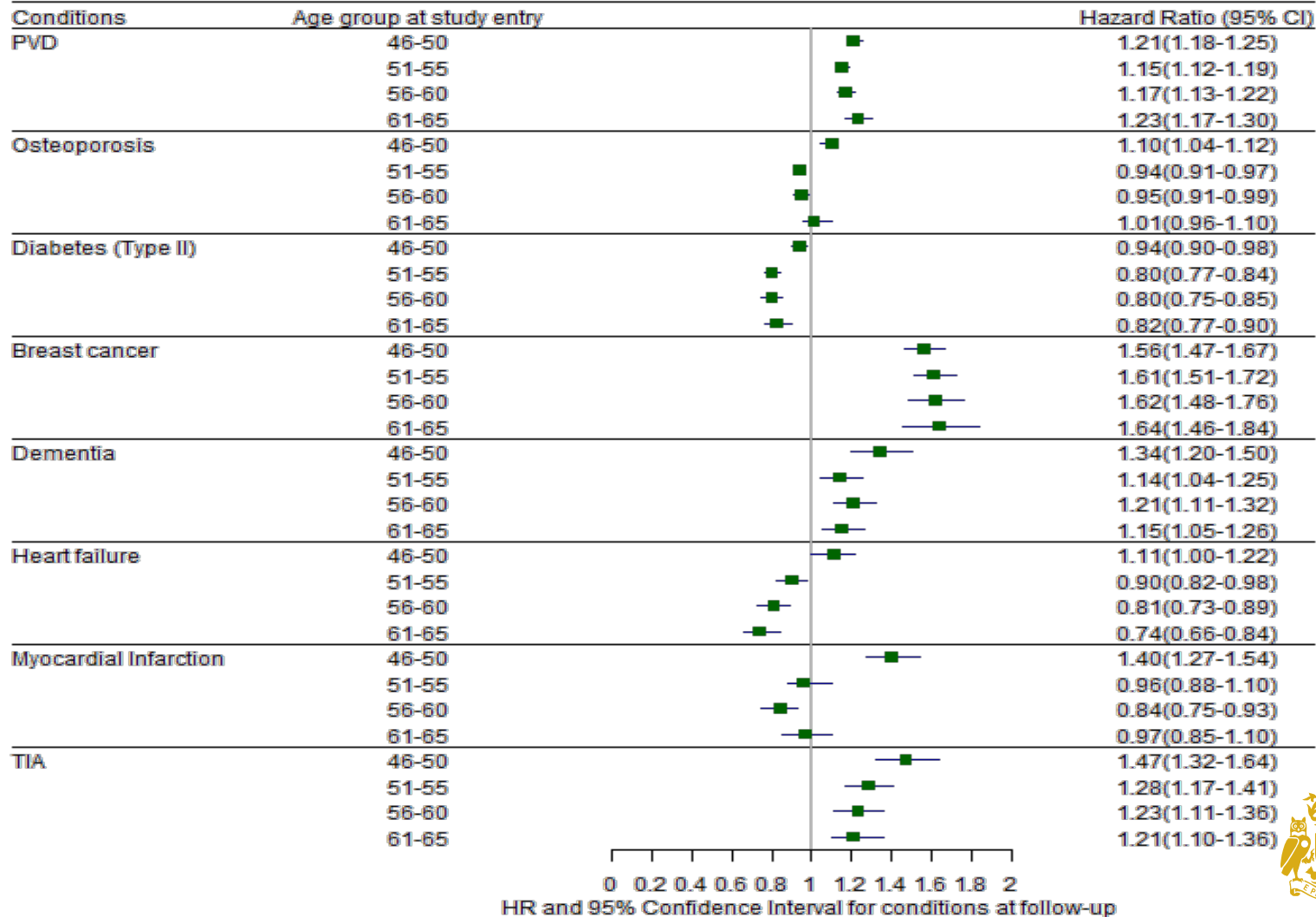
- Majority of women started HRT between 46-55 years of age
- There are more death in controls than cases in all age category



Actuarial  
Research Centre  
Institute and Faculty  
of Actuaries



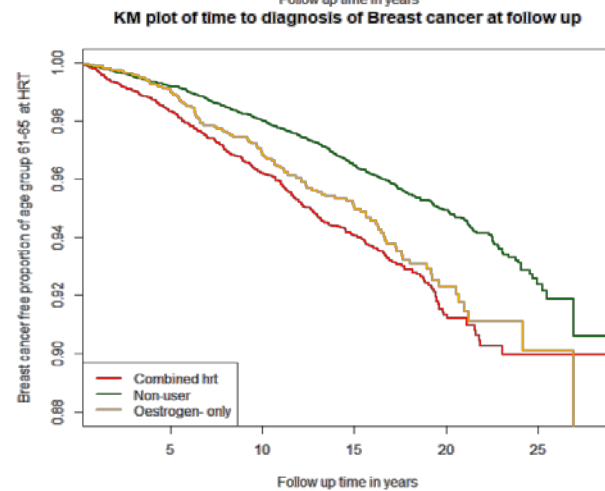
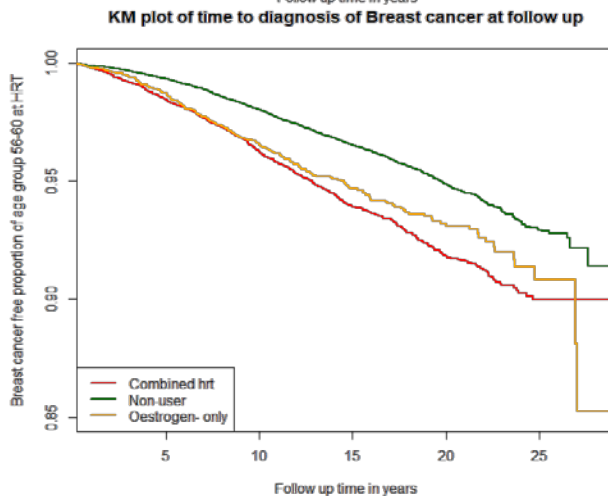
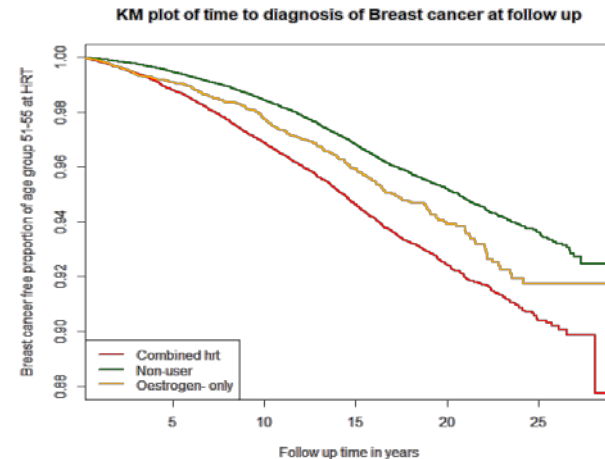
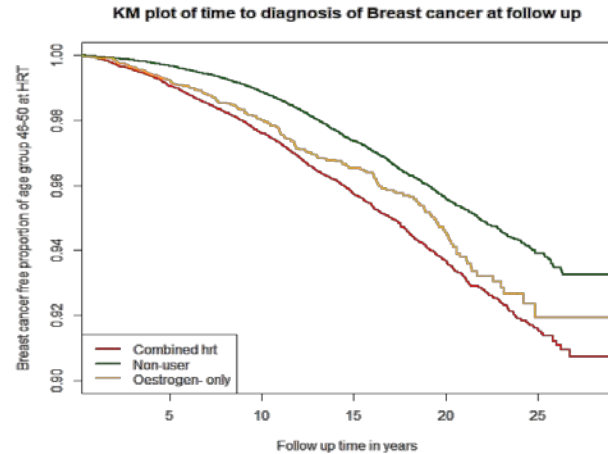
# Hazard ratios and 95% confidence intervals of the conditions developed at follow up



Actuarial  
Research Centre

Institute and Faculty  
of Actuaries

# Time to diagnosis of Breast cancer at follow up by age category at HRT and its type



- In all age category HRT users developed more breast cancer than non-user at follow up



Actuarial  
Research Centre

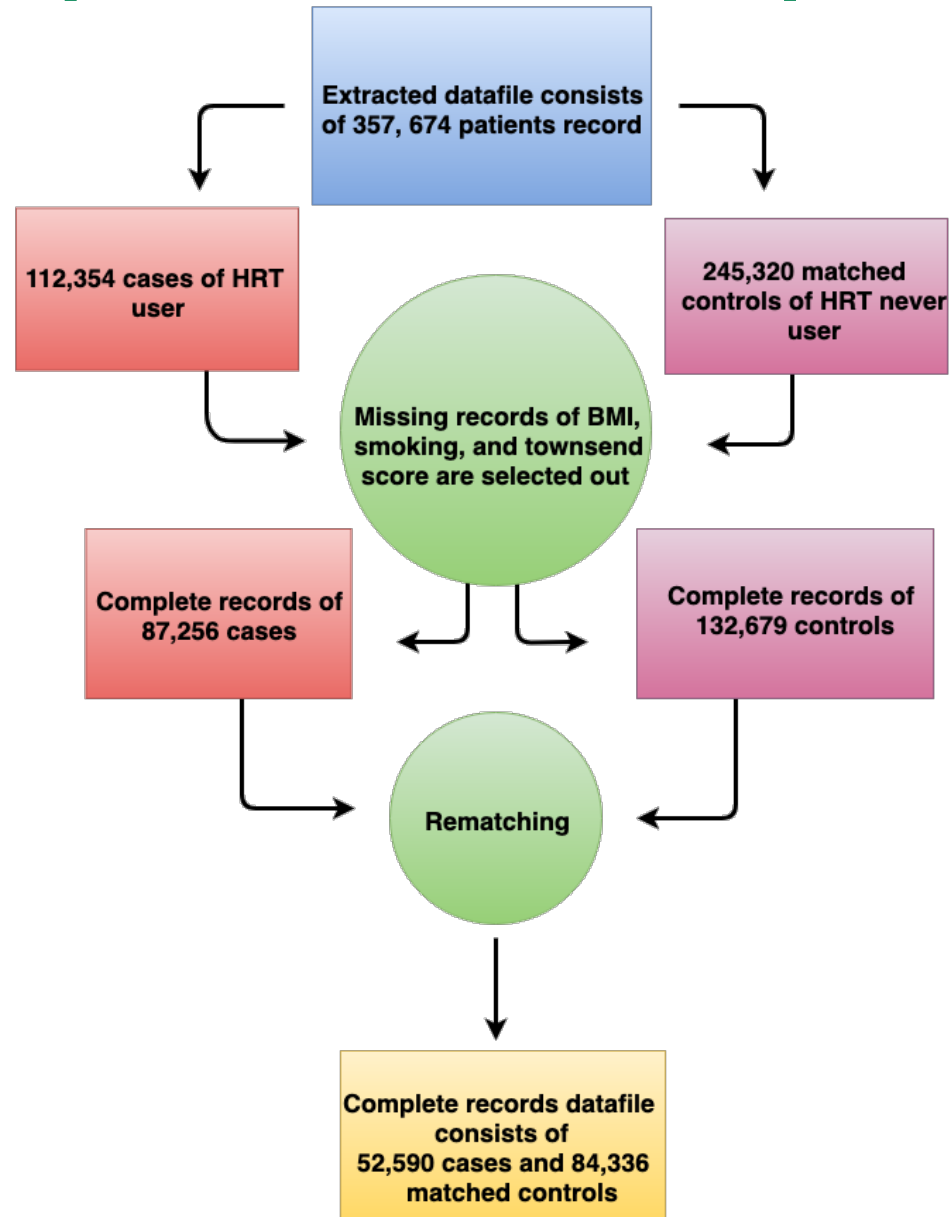
Institute and Faculty  
of Actuaries

# Survival model of all-cause mortality

- The following predictors were used in the survival modelling:
  - **Socio-economic status:** Townsend score
  - **Lifestyle:** Smoking status, body mass index (BMI)
  - **Health:** Type II diabetes, hypertension, hypercholesterolaemia, peripheral vascular disease (PVD)/peripheral arterial disease (PAD), coronary heart disease (CHD), oophorectomy/hysterectomy status, systolic and diastolic blood pressure
  - **Demography:** Age category at first HRT and birth cohort
  - **Medication:** HRT (estrogen-only, estrogen and progesterone), antihypertensive drugs
- Patients with complete information for all of the above covariates has been selected for full case analysis
- Final model also included interactions of smoking with BMI and type II diabetes



# Selection of patients with complete records:



Actuarial  
Research Centre

Institute and Faculty  
of Actuaries

# Grambsch and Therneau test

	rho	chisq	p
hrt_cat_1Combined	-0.002697	0.074411	0.785020029057
hrt_cat_10estrogen-only	-0.000269	0.000753	0.978111154356
Age.cat51-55	-0.031085	10.057256	0.001517490554
Age.cat56-60	-0.027581	7.989151	0.004705847114
Age.cat61-65	-0.040494	17.268775	0.000032447539
B.cohort1931-1940	-0.026251	7.131065	0.007575972010
B.cohort1941-1950	-0.046838	22.712986	0.000001880925
B.cohort1951-1960	-0.042161	18.070751	0.000021284617
hypertension_cat_treated	-0.014381	2.178700	0.139932572339
hypertension_cat_untreated	0.012651	1.655491	0.198213197224
bmi_cat0bese	0.003277	0.110294	0.739809895849
bmi_cat0Overweight	-0.004197	0.181653	0.669956369590
smoking_cat_current	-0.051673	27.301012	0.000000174120
smoking_cat_ex	-0.023847	5.853903	0.015542544963
townsend_2	0.001919	0.038061	0.845320438510
townsend_3	-0.004281	0.188900	0.663833458455
townsend_4	-0.003795	0.148532	0.699942607104
townsend_5	-0.007716	0.615187	0.432841333884
diabetes_type2	-0.006944	0.522423	0.469809744403
osteoporosis	0.004521	0.230037	0.631496621577
CHD	-0.004256	0.188971	0.663774173883
opho.hysboth_removed	0.008301	0.719502	0.396307206913
opho.hysopho_without_hys	0.001018	0.010713	0.917564502212
bmi_cat0bese:smoking_cat_current	0.025841	6.906091	0.008590260141
bmi_cat0Overweight:smoking_cat_current	0.014928	2.301700	0.129232523293
bmi_cat0bese:smoking_cat_ex	0.016734	2.898047	0.088686922137
bmi_cat0Overweight:smoking_cat_ex	0.023806	5.863101	0.015461528170
smoking_cat_current:diabetes_type2	0.005911	0.362035	0.547378311009
smoking_cat_ex:diabetes_type2	-0.016739	2.914064	0.087810369554
GLOBAL	NA	103.604595	0.00000000257

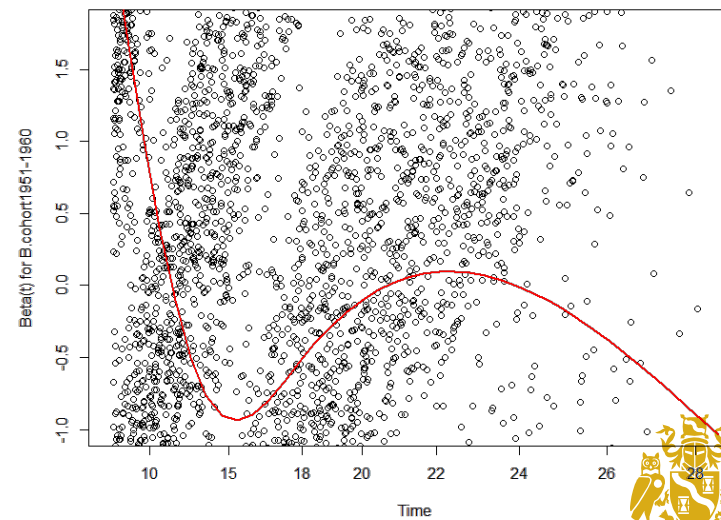
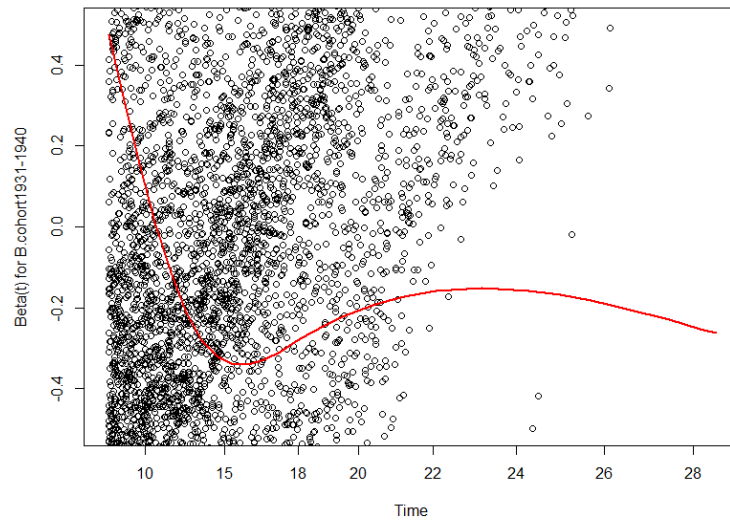
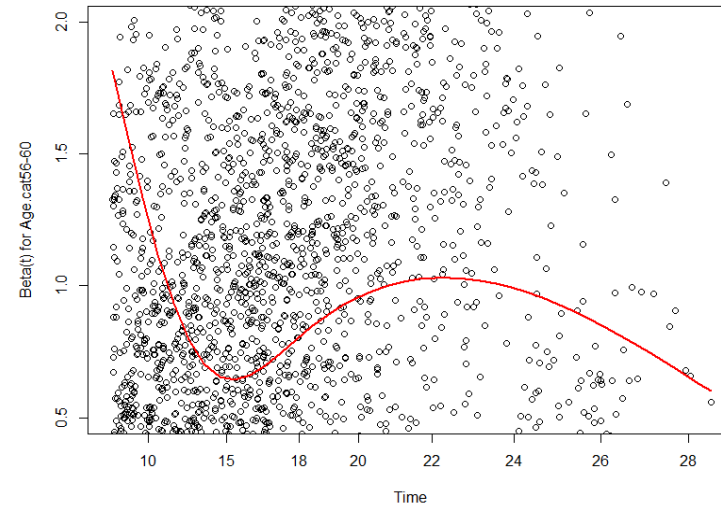
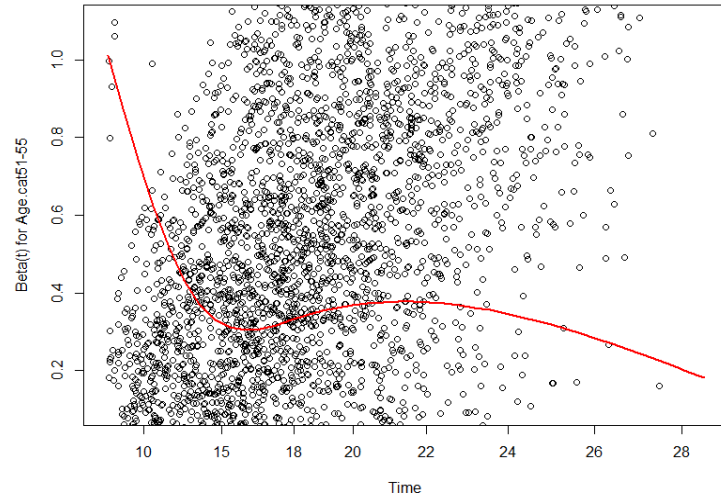
A significant p-value (<0.05) is an indication of violation of the proportional hazard assumption in the Cox PH model



Actuarial  
Research Centre

Institute and Faculty  
of Actuaries

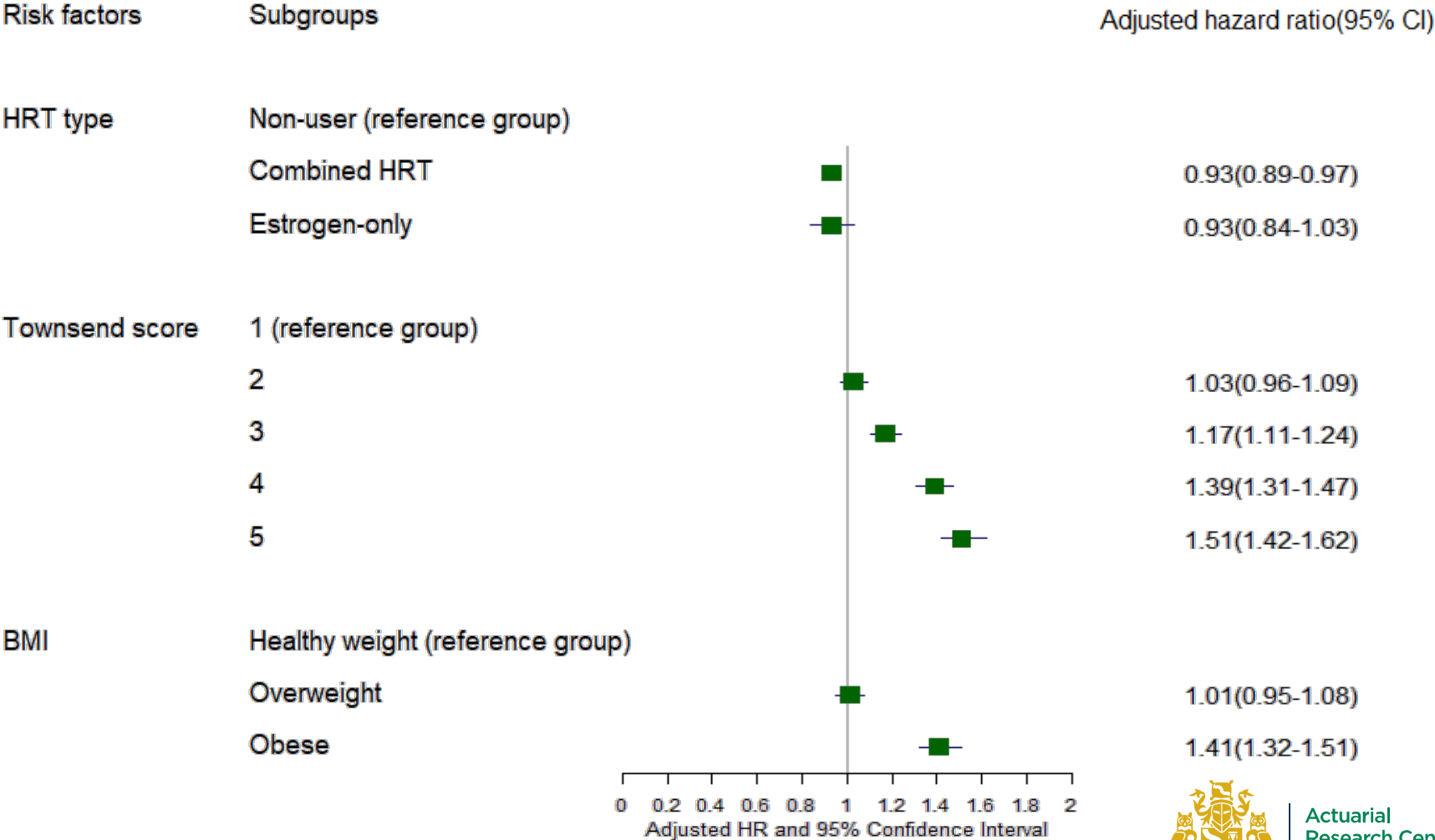
# Plots of residuals:



Actuarial  
Research Centre

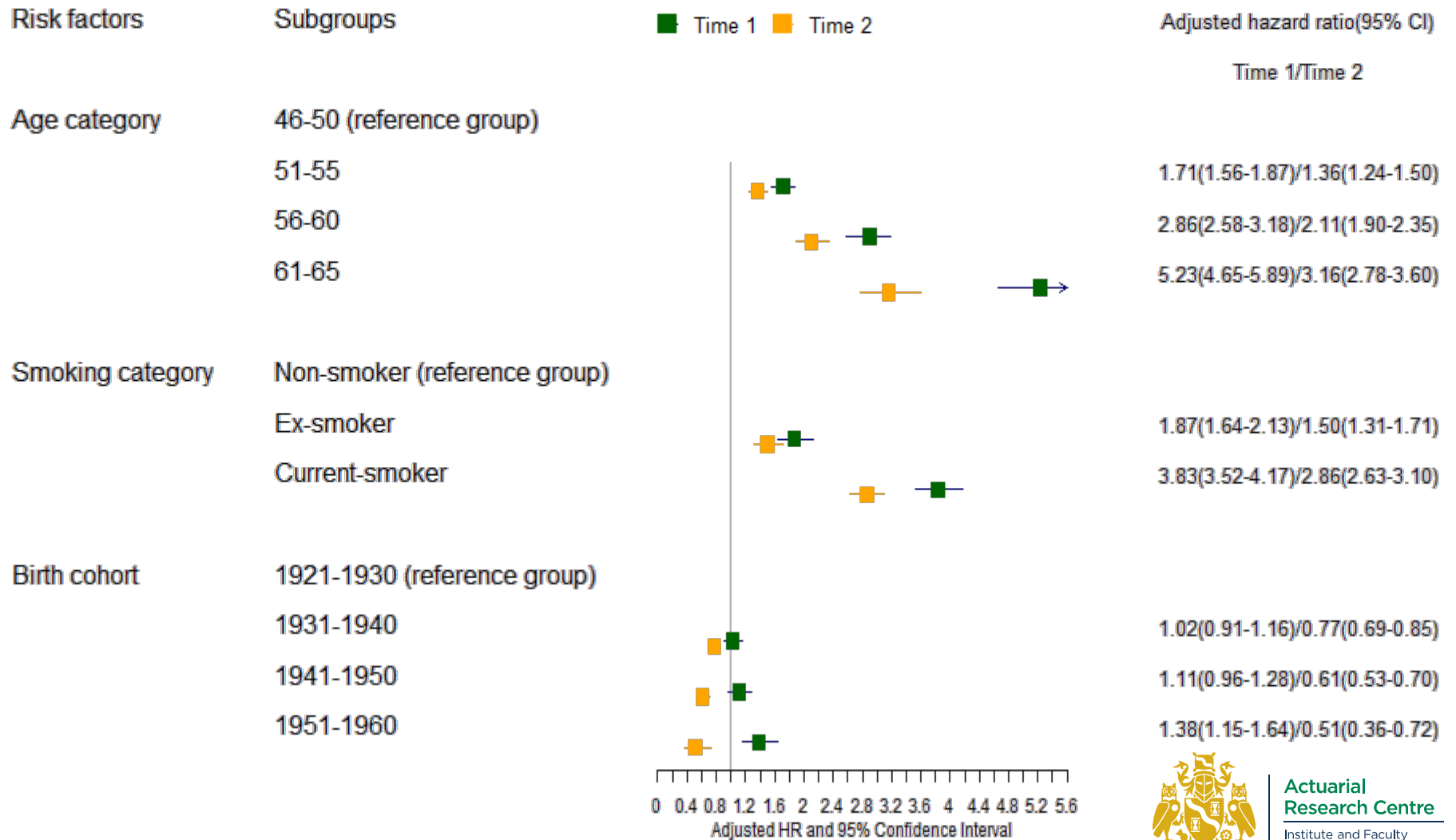
Institute and Faculty  
of Actuaries

# HRT model, Forest plot 1



Actuarial  
Research Centre  
Institute and Faculty  
of Actuaries

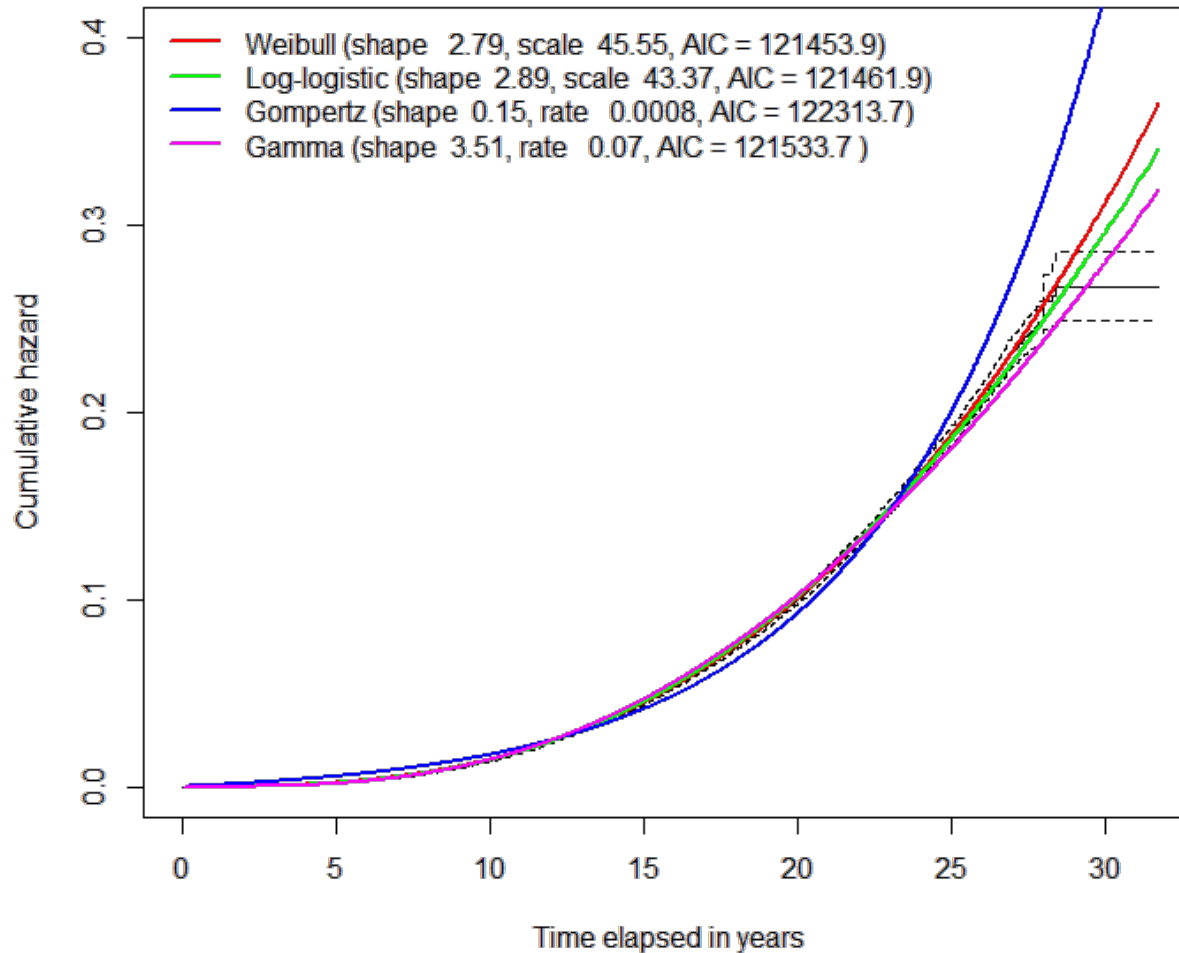
# HRT model, Forest plot 2



**Actuarial  
Research Centre**  
Institute and Faculty  
of Actuaries



# Baseline hazard function fitted with different parametric distributions:



Actuarial  
Research Centre

Institute and Faculty  
of Actuaries

# Future Work:

- Multiple imputation
- Models for imputed data.
- Translation of models into actuarial analysis
- Landmark analysis



**Actuarial  
Research Centre**

Institute and Faculty  
of Actuaries



## Actuarial Research Centre

Institute and Faculty  
of Actuaries

# The Actuarial Research Centre (ARC)

## A gateway to global actuarial research

The Actuarial Research Centre (ARC) is the Institute and Faculty of Actuaries' (IFoA) network of actuarial researchers around the world.

The ARC seeks to deliver cutting-edge research programmes that address some of the significant, global challenges in actuarial science, through a partnership of the actuarial profession, the academic community and practitioners.