Actuarial Research Centre

Institute and Faculty of Actuaries

Use of Primary Health Care Records Data in Actuarial Research

PI Elena Kulinskaya (UEA)
CoIs Nigel Wright (Aviva), Nicholas Steel (NMS)
Team: Ilyas Bakbergenuly, Padma Chutoo, Nurunnahar Akter, Njabulo Ncube

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

9/03/21

Contents

- Introduction
- Hazards, hazards ratios and the Cox Regression
- Landmark analysis
- Double-Cox model
- Modelling survival after Stroke and TIA - *Padma Chutoo*
- Does HRT increase Life expectancy? *Nurunnahar Akter*
- Modelling survival of people with T2M - *Njabulo Ncube*
- R software - *Ilyas Bakbergenuly*
- Summary and discussion

AVIVA

UEA University of East Anglia

Actuarial Research Centre
Institute and Faculty of Actuaries

Aims and data

Aims: Development of novel statistical and actuarial methods for modelling mortality and morbidity and evaluating longevity improvement based on Big Health and Actuarial Data.

Data: Our research uses The Health Improvement Network (THIN) primary care data to develop statistical models of longevity

- Our subset of THIN includes all patients born before 1960 and followed to 01.01.2017, this is 3.5 million patients
- 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 lifestyle and health factors instead of the aggregated profile

AVIVA

University of East Anglia

Actuarial Research Centre
Institute and Faculty of Actuaries

Target Conditions and interventions

AVIVA University of East Anglia Institute and Faculty of Actuaries

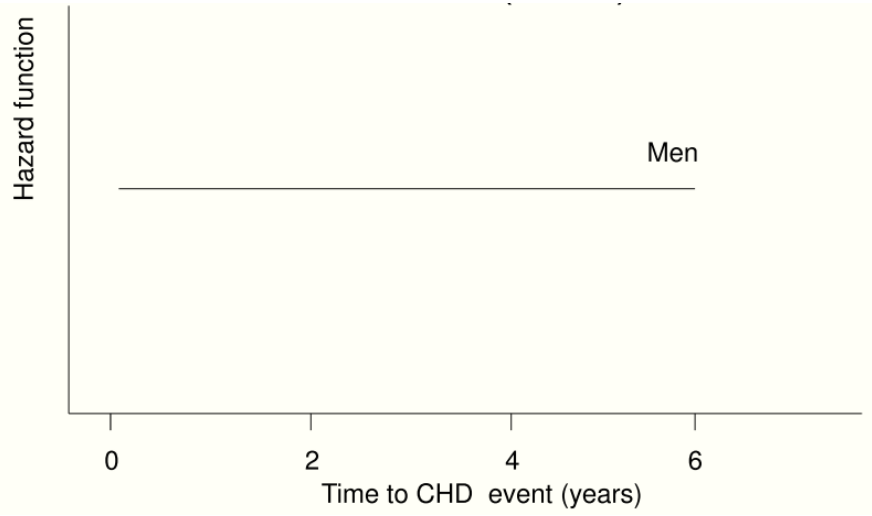
Contents

- Introduction
- **Hazards, hazards ratios and the Cox Regression**
- Landmark analysis
- Double-Cox model

AVIVA University of East Anglia Actuarial Research Centre Institute and Faculty of Actuaries

Hazard aka "force of mortality" and "mortality intensity"

- Hazard is an instantaneous failure rate at time (age) *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 or force of mortality $\mu_i(t)$ 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*.
- The hazard ratio for subjects i and j does not include the baseline hazard, and is constant over time (age).

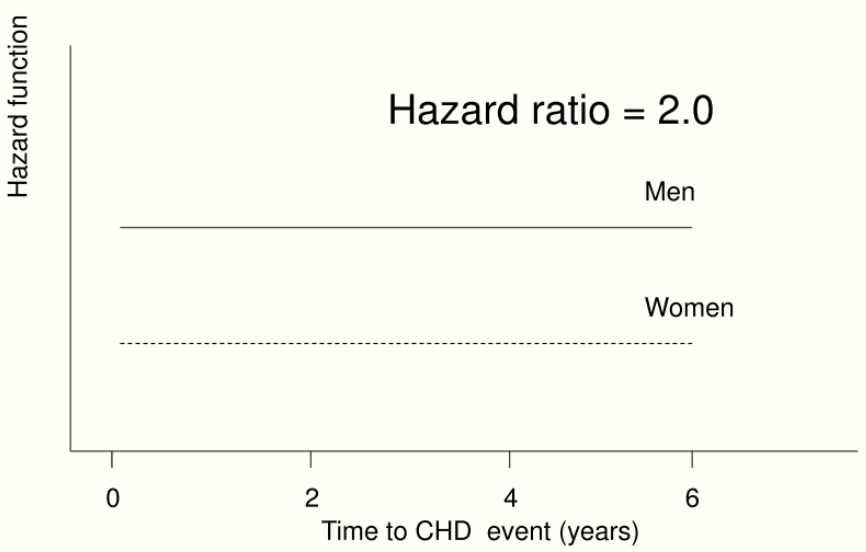
AVIVA

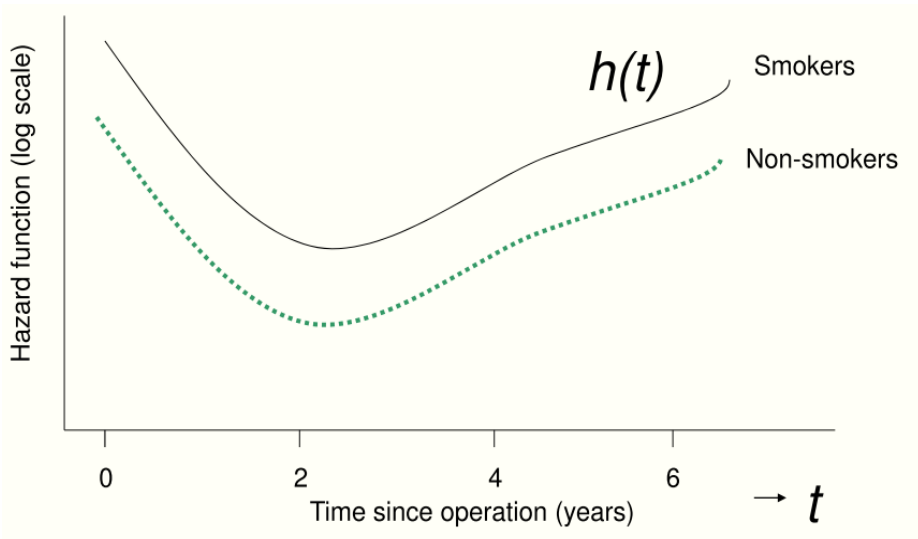
University of East Anglia

Actuarial Research Centre
Institute and Faculty of Actuaries

Hazard ratio (HR)

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





Gompertz baseline hazards

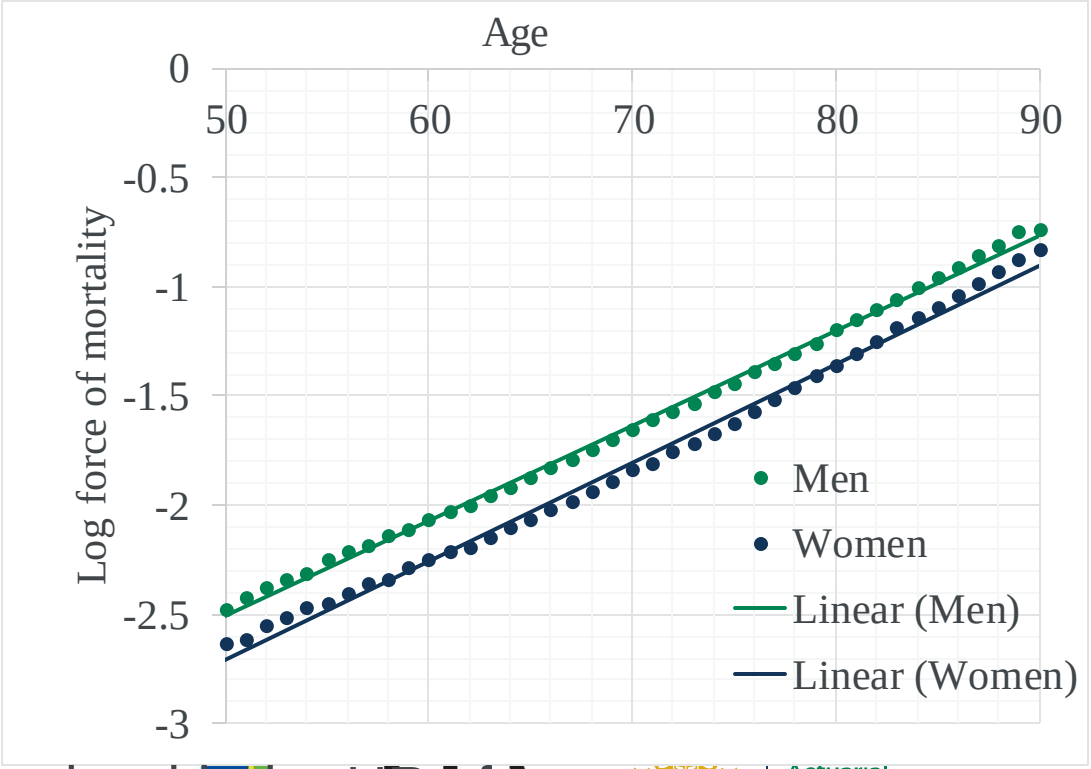
- It is well accepted that the Gompertz distribution provides a good description of human mortality between ages 50 and 95 (Brenner et al. 1993, Spiegelhalter 2016).

Using Gompertz law,

$$\ln \mu_0(t) = \lambda_0(t) = a + bt,$$

- For England and Wales in 2010-2012, the increase in the hazard between those ages was approximately 1.1 per year.
- We use Gompertz baseline hazards with the HRs from Cox regression. [3]

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



AVIVA University of East Anglia Actuarial Research Centre Institute and Faculty of Actuaries

What if the proportional hazards assumption is not met?

- For a Cox model $\mu(t|\beta, Z) = \mu_0(t)\exp(Z^{\mathrm{T}}\beta)$ we use two ways to cope with non-proportionality:
- Use landmark analysis which amalgamates a smoothed series of piecewise constant hazards
  - This is a comparatively new statistical method [Van Houwelingen, H. and Putter, H., 2011]
  - We published papers on linking landmark analysis to longevity [4] and on landmark analysis of statins [5] .
- Include additionally a model for shape of baseline hazards- "Double-Cox" model
  - This is a new statistical model developed by us to deal with these kinds of data [2]

AVIVA

University of East Anglia

Actuarial Research Centre
Institute and Faculty of Actuaries

Contents

- Introduction
- Hazards, hazards ratios and the Cox Regression
- **Landmark analysis**
- Double-Cox model

AVIVA

University of East Anglia

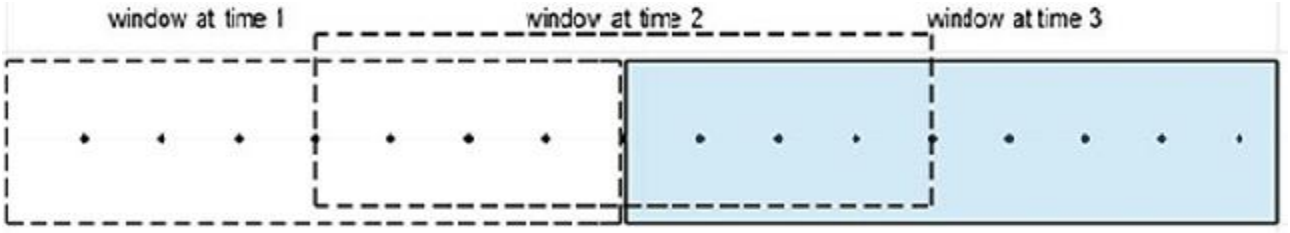
Actuarial Research Centre
Institute and Faculty of Actuaries

Landmark analysis

- Landmark Analysis accounts for time-dependent effects by fitting a series of Cox regression models within a sliding window $[t_{LM}, t_{LM}+w]$ for a series of points $s = t_{LM}$.

$$\mu\,(t|x, t_{LM}, w) = \mu_0(t|t_{LM}, w)\exp(Z^T\beta_{LM})\,, \qquad s \le t \le s + w$$

- Each consecutive data set *s* is obtained by truncation at $s = t_{LM}$ and administrative censoring at t_{LM}+w.



Dynamic prediction for the conditional survival after t=t_{LM} is based on current information for all patients still alive just prior to t_{LM}. [Van Houwelingen, H. and Putter, H., 2011]

AVIVA

University of East Anglia

Actuarial Research Centre
Institute and Faculty of Actuaries

Application: Landmark Analysis of Statins

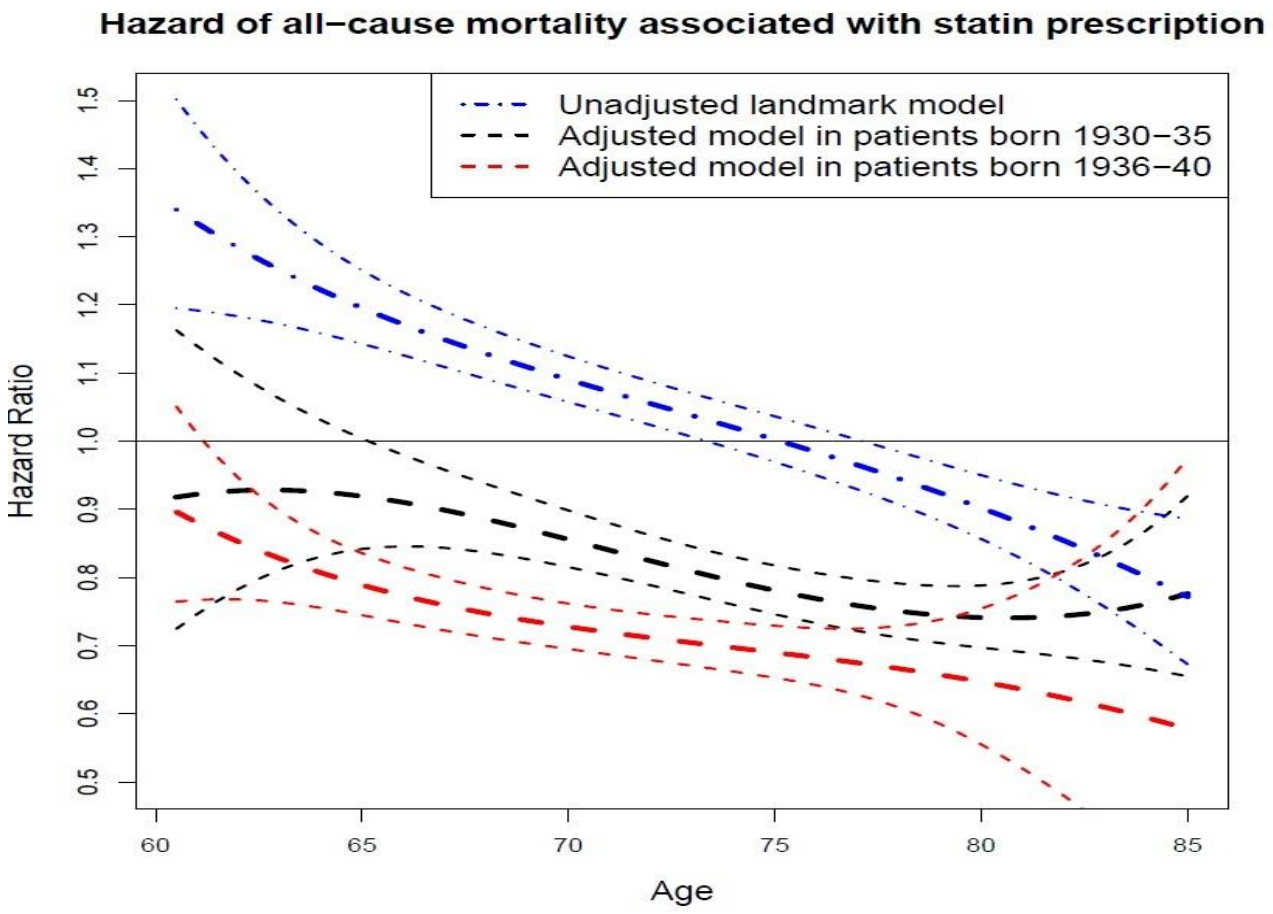
Data: 110,243 patients who turned 60 between 1990 and 2000 and did not have a previous statin prescription or a cardiovascular disease diagnosis.
Medical history was updated every half a year (landmark) until end of follow-up (death, deregistered or end of study).

Imputation: Due to missing data at early ages, multiple imputation was performed using joint modelling at age 60.

Analysis: Landmark analysis was 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 (51 landmark points) and adjusted for medical history [5]

AVIVA University of East Anglia Actuarial Research Centre Institute and Faculty of Actuaries

Hazard of all-cause mortality associated with statin prescription



The raw regression coefficients β_{LM} and the baseline hazards are smoothed over time as the *m*th degree polynomials in s. Their, coefficients are estimated using pseudo-partial log-likelihood.

We add Gompertz baseline hazards to obtain survival functions and the life expectancy (LE). [4]

AVIVA

University of East Anglia

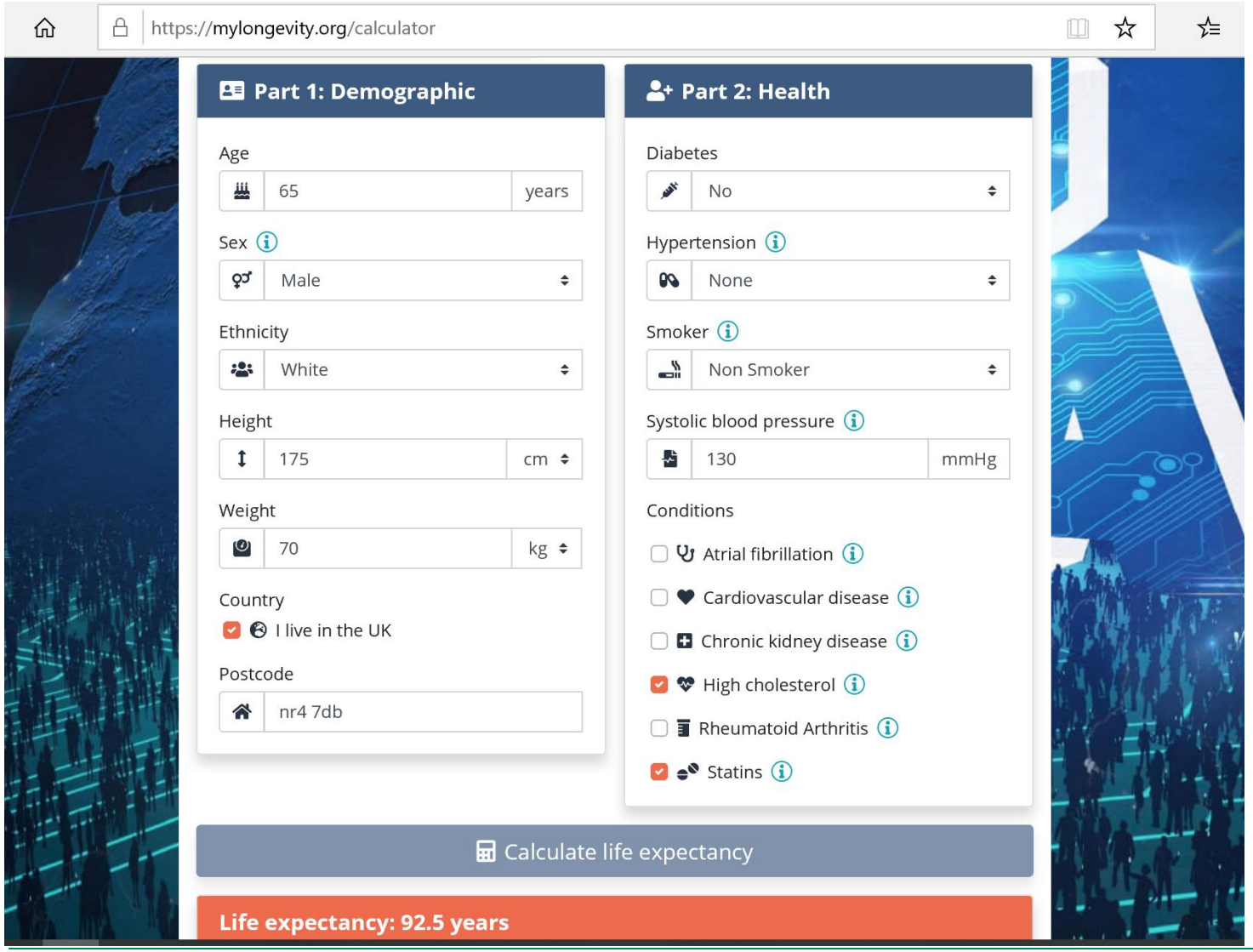
Actuarial Research Centre
Institute and Faculty of Actuaries

Calculating component life expectancies

- Since the mortality rates and the prevalences of the factors differ by gender and by socio-economic status (SES), we analysed the life tables separately for each SES quintile-by-gender combination.
- For each life table, we considered, all combinations of statin use (2 levels), smoking (3), hypertension (3), diabetes (2), hypercholesterolaemia (2), BMI category (3) and cardiac risk (3), **648 combinations** in total, and estimated their prevalences at 51 landmark points.
- We estimated survival functions and LE for each risk group at each landmark point. The overall survival function is the weighted mean of the survival functions in the individual risk groups. This was used for calibration of LEs [3, 4]

AVIVA University of East Anglia Actuarial Research Centre Institute and Faculty of Actuaries

mylongevity.org app



Our results are implemented in the LE calculator app

(web developer George Oastler).

Actuarial Research Centre
Institute and Faculty of Actuaries

Contents

- Introduction
- Hazards, hazards ratios and the Cox Regression
- Landmark analysis
- **Double-Cox model**
- Modelling survival after Stroke and TIA - *Padma Chutoo*

AVIVA

University of East Anglia

Actuarial Research Centre
Institute and Faculty of Actuaries

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

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

β is a vector of unknown parameters for scale and Z is a vector of covariates

Baseline hazard function

U ~ Gamma with mean 1 and variance σ^2 is a shared (across a cluster, e.g. GP practice) frailty. σ^2 is population heterogeneity

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

Additional regression model to allow varying shape depending on covariates

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

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

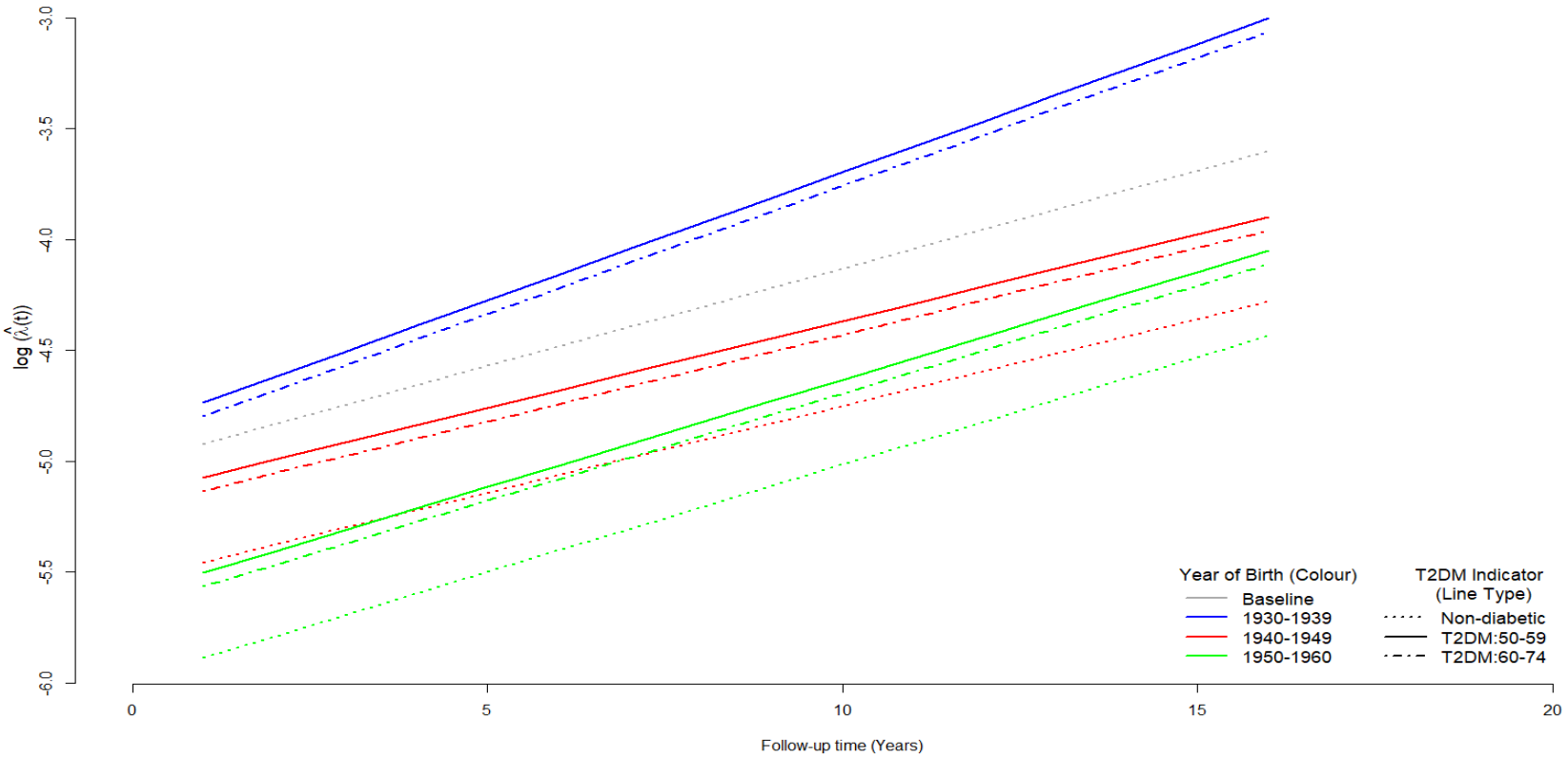
$$\mathrm{k(Z)} = k_0 e^{Z^{\mathrm{T}}\beta_k}$$

AVIVA

University of East Anglia

Actuarial Research Centre
Institute and Faculty of Actuaries

Shape effects of birth cohorts in T2DM



Log cumulative hazards of all-cause mortality by birth cohort and T2DM diagnosis for patients with treated hypertension

AVIVA

University of East Anglia

Actuarial Research Centre
Institute and Faculty of Actuaries

Target Conditions and interventions

AVIVA University of East Anglia Institute and Faculty of Actuaries

References

1. Gitsels L.A., Kulinskaya E., Steel N. (2016). PLoS ONE **11**(11): e0166847.
2. Begun A., Kulinskaya E. and MacGregor A. (2019) *BMC Med Res Methodol* **19,** 217 (2019).
3. Kulinskaya E. , Gitsels, LA. Bakbergenuly, I. and Wright N.R. (2020) *Insurance Mathematics and Economics*, **93**, 27-35
4. Kulinskaya, E. , Gitsels, L.A., Bakbergenuly, I. and Wright, N.R. (2020) *Insurance Mathematics and Economics* https://doi.org/10.1016/j.insmatheco.2020.11.001
5. Gitsels, LA. , Bakbergenuly, I., Steel N. and Kulinskaya E. (2021) *Family Medicine and Community Health Journal*, 0:e000780. doi:10.1136/fmch-2020-000780 14/01/2021

AVIVA

University of East Anglia

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.

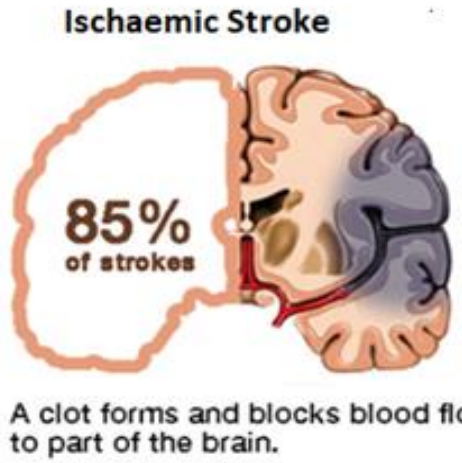
Outline of presentation

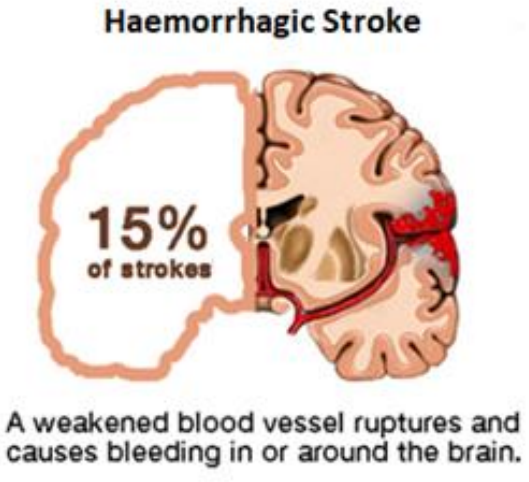
- Overview of Transient Ischaemic Attack (TIA) and Ischaemic stroke (IS).
- Survival Models' findings.
- Some practical examples : How TIA and ischaemic stroke affect life expectancy?

AVIVA University of East Anglia Actuarial Research Centre Institute and Faculty of Actuaries

What is Stroke?

- **Ischaemic 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 Ischaemic 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

AVIVA

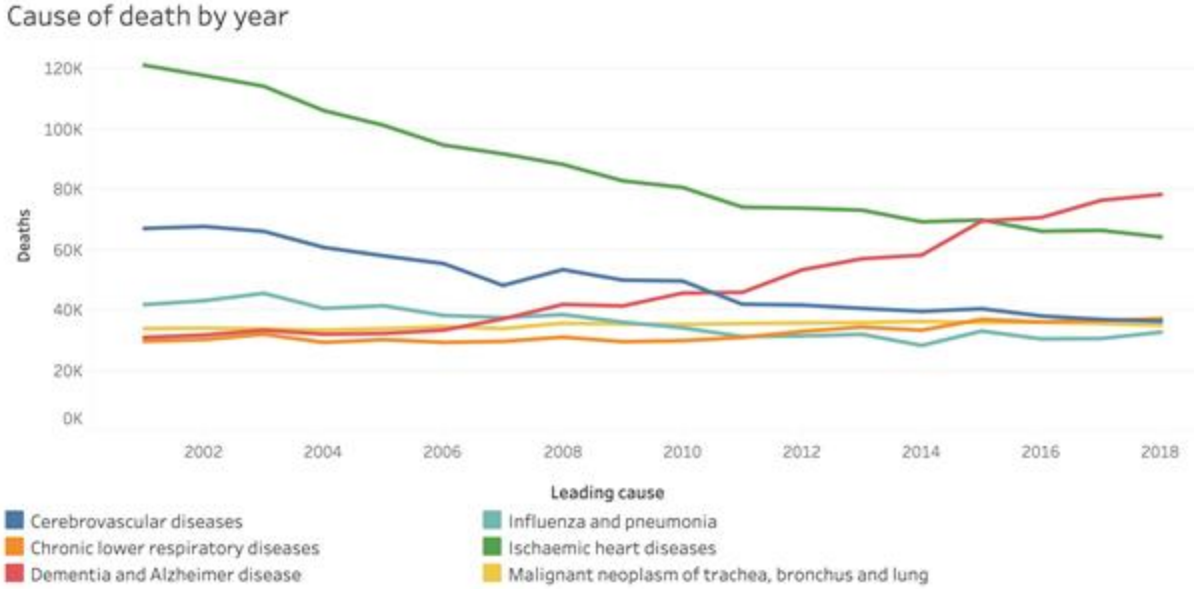
UEA University of East Anglia

Actuarial Research Centre
Institute and Faculty of Actuaries

Stroke : a serious medical condition

Leading causes of death

Stroke is a type of cerebrovascular disease, which is **one of the leading causes of death** in the UK. Stroke accounts for roughly 75% of deaths from cerebrovascular diseases [Stroke statistics, 2021].



Leading cause of disability

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

Stroke prevalence is projected to increase by 120% between 2015 and 2035 and the associated societal costs will almost treble [King et al.,2020].

AVIVA

UEA University of East Anglia

Actuarial Research Centre
Institute and Faculty of Actuaries

Stroke study: brief description

- Objective: impact of 1st **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.

AVIVA

University of East Anglia

Actuarial Research Centre
Institute and Faculty of Actuaries

Stroke study: brief description

Variables of interest:

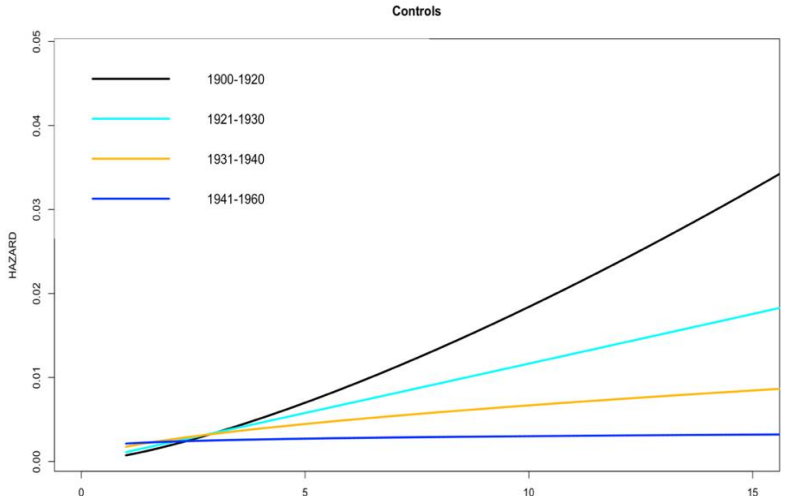
- **Drugs:** Antihypertensive drugs, Antiplatelet drugs, Anticoagulant drugs, Lipid regulating drugs and antidiabetic drugs.
- **Medical conditions:** Asthma, Atrial Fibrillation, CKD, CHD, PAD, Hypothyroidism, COPD, Diabetes, Hypercholesterolemia, Hypertension, Depression.
- **Demographical and lifestyle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

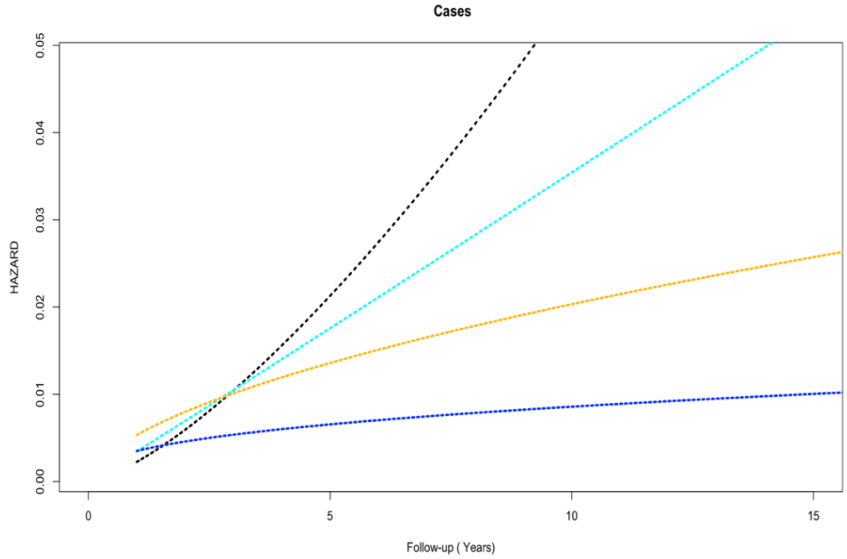
AVIVA

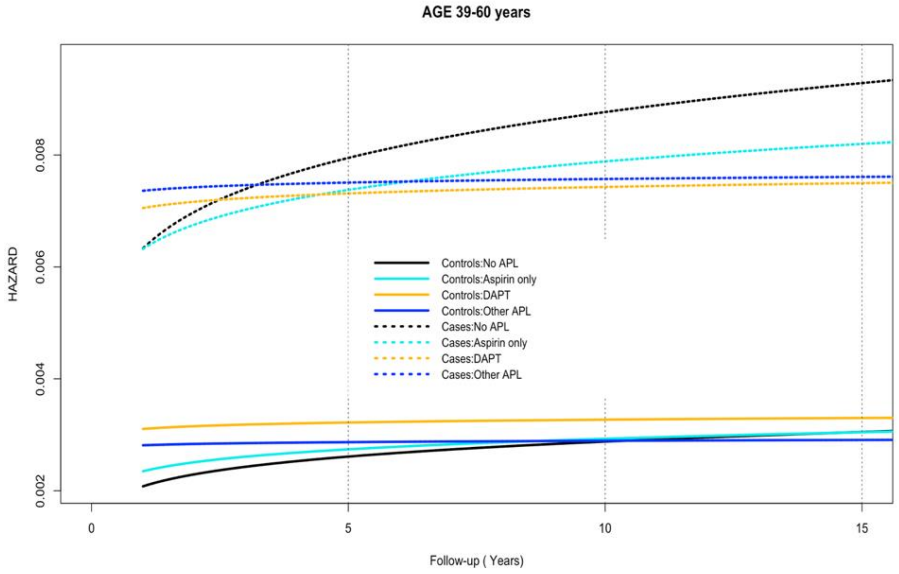
University of East Anglia

Actuarial Research Centre
Institute and Faculty of Actuaries

Findings: Survival after TIA







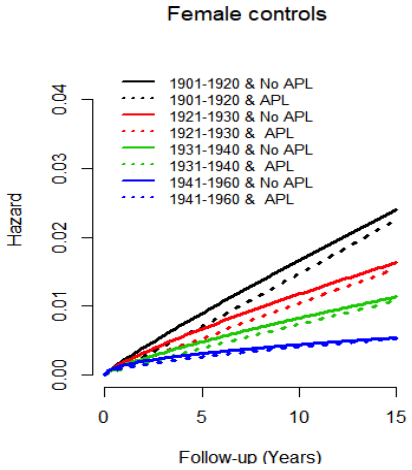
- High burden of cardio-vascular comorbidities is more prevalent among TIA patients than their matched controls.
- The overall risk of death remains high for 10 to 15 years after a TIA event.
- Aspirin provides long-term marginal survival benefits to TIA patients.

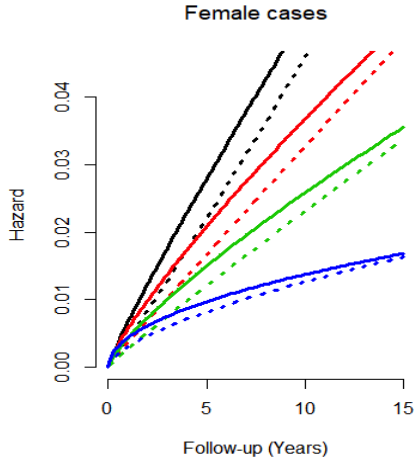
AVIVA

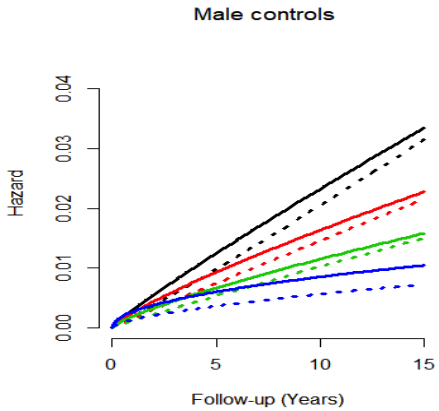
University of East Anglia

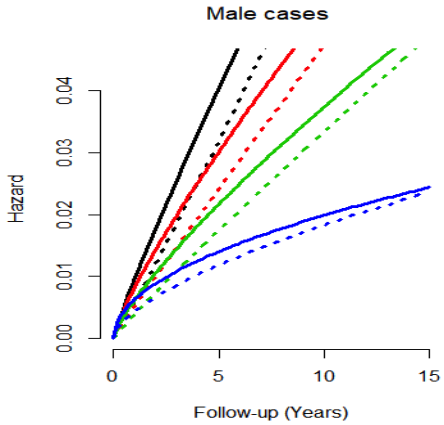
Actuarial Research Centre
Institute and Faculty of Actuaries

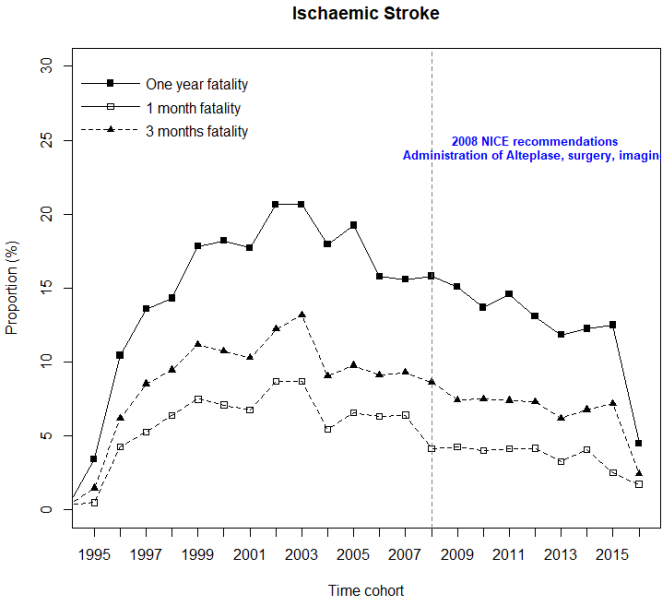
Findings : Survival after IS

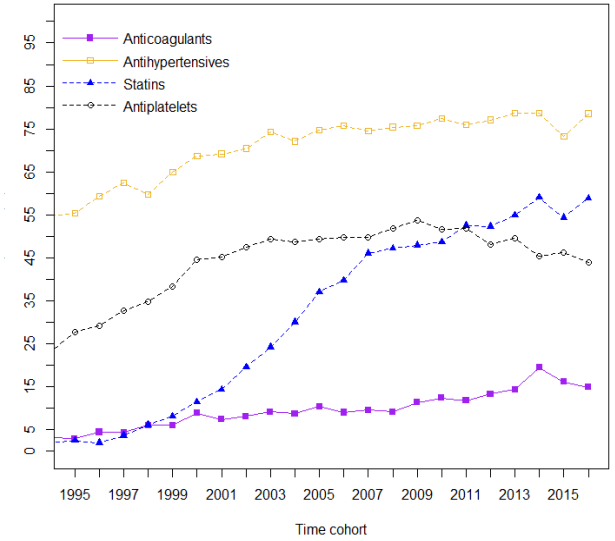












- Stroke survival outcome in England has improved over time.
- Hypertension is associated with poor survival outcomes.
- Pre-morbid use of preventive antiplatelet treatment declined after year 2010 in England. Yet, our finding shows that that antiplatelets provide a survival benefit to the cases.

AVIVA

University of East Anglia

Actuarial Research Centre
Institute and Faculty of Actuaries

Longevity models : Some examples

Medical condition	Gender	Age at diagnosis	Years after diagnosis	Number of years lost
IS	Male	50	2	14.1
IS	Female	50	2	16.0
TIA	Male	50	2	6.9
TIA	Female	50	2	8.2
IS	Male	60	2	9.9
IS	Female	60	2	11.2
TIA	Male	60	2	5
TIA	Female	60	2	5.9
IS	Male	70	2	6.9
IS	Female	70	2	7.8
TIA	Male	70	2	3.7
TIA	Female	70	2	4.1

References

- King, D., Wittenberg, R., Patel, A., Quayyum, Z., Berdunov, V. and Knapp, M., 2020. The future incidence, prevalence and costs of stroke in the UK. *Age and ageing*, *49*(2), pp.277-282.
- Stroke Association. 2021. Stroke statistics. [online] Available at: <https://www.stroke.org.uk/what-is-stroke/stroke-statistics> [Accessed 29 January 2021].
- Begun, A., Kulinskaya, E. and MacGregor, A.J., 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, 19(1), p.217.

AVIVA

University of East Anglia

Actuarial Research Centre
Institute and Faculty of Actuaries

Actuarial
Research Centre

Institute and Faculty
of Actuaries

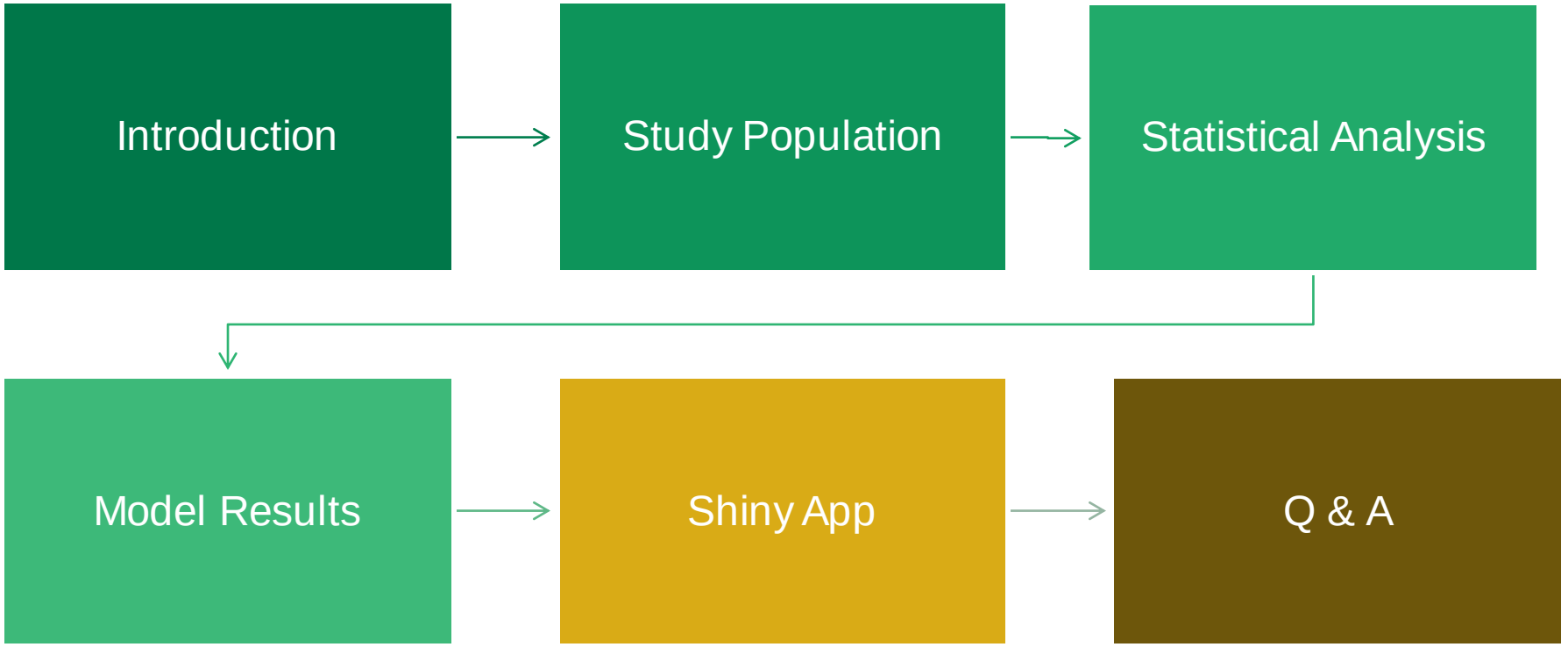
On the Survival of Diabetes Type II individuals in the UK

Njabulo Ncube

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

9 March 2021

Presentation Outline



AVIVA

University of East Anglia

Actuarial Research Centre
Institute and Faculty of Actuaries

Introduction

- *Objective*: estimating survival of diabetes type II (T2DM) individuals in the UK using Big Health Data.
- *Study design*: retrospective matched case-control study
- *Selection Criteria*:
  individuals born between 1930 and 1960 inclusive;
  diagnosed with T2DM between 2000 and 2016, inclusive;
  aged 50 to 74 years and with no serious medical conditions at diagnosis.
- *Matching*: 1:3 by general practice, gender and age

AVIVA University of East Anglia Actuarial Research Centre Institute and Faculty of Actuaries

Study Population

Total Size: 221 182

Number of Deaths: 29 618

Table 1: Percentage of study population by selected demographic variables

Description		2000 – 2004		2005 – 2009		2010 – 2016	
		T2DM	Controls	T2DM	Controls	T2DM	Controls
Birth Cohort:	1930 – 1939	28.1	71.9	30.2	60.8	34.7	65.3
	1940 – 1949	28.3	71.7	30.7	60.3	35.5	64.5
	1950 – 1960	28.9	71.1	30.9	69.1	36.5	63.5
Gender:	Male	58.7	56.8	59.7	56.5	60.3	57
Age at Entry:	50 – 59	42.1	41.9	42	41.1	40.2	39.9
	60 – 74	57.9	58.1	58	58.9	59.8	60.1

Table 2: Missingness

Variable	Number	%
Smoking Status	13 425	6.1
BMI	34 924	15.8
Townsend Deprivation Index	14 497	6.6

AVIVA

University of East Anglia

Actuarial Research Centre

Institute and Faculty of Actuaries

Statistical Analysis

- *Model*: Begun et. al.(2019) Gompertz-Cox Model with gamma frailty.
- *Variable Selection*: Backward elimination

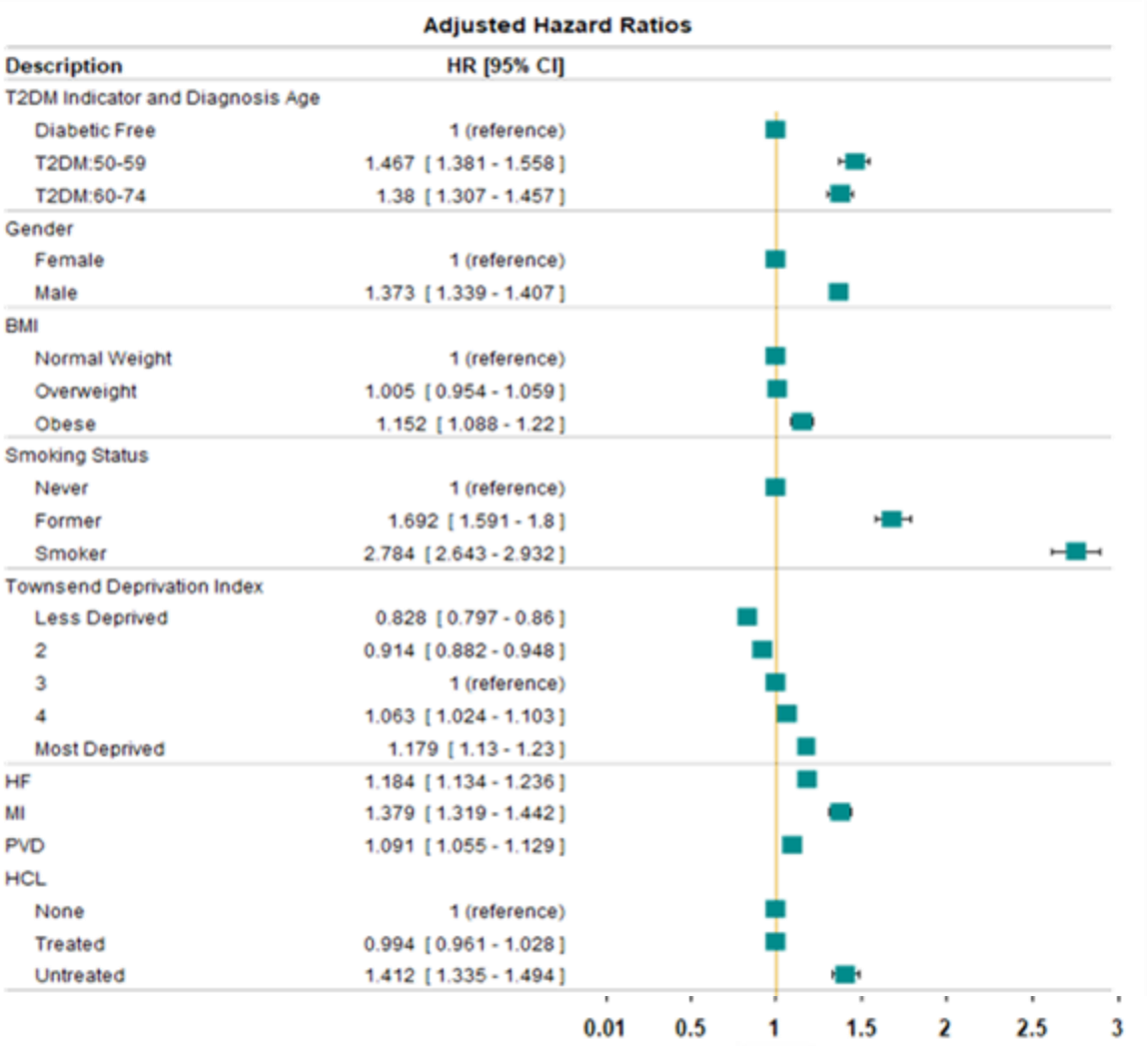
  - using complete cases

  $\alpha = 5\%$ for main effects, $\alpha = 1\%$ for interactions
- *Missingness*: multiple imputation – R jomo package
- *Discrimination*: Coefficient of Concordance (0.76), Likelihood and AIC
- Shape parameters: Birth cohort, Atrial Fibrillation, Hypertension

AVIVA University of East Anglia Actuarial Research Centre Institute and Faculty of Actuaries

Results

(a)



HCL: hypercholesterolemia
HF: heart failure
MI: myocardial infarction
PVD: pulmonary vascular disease

AVIVA

University of East Anglia

Actuarial Research Centre
Institute and Faculty of Actuaries

Results

(b)

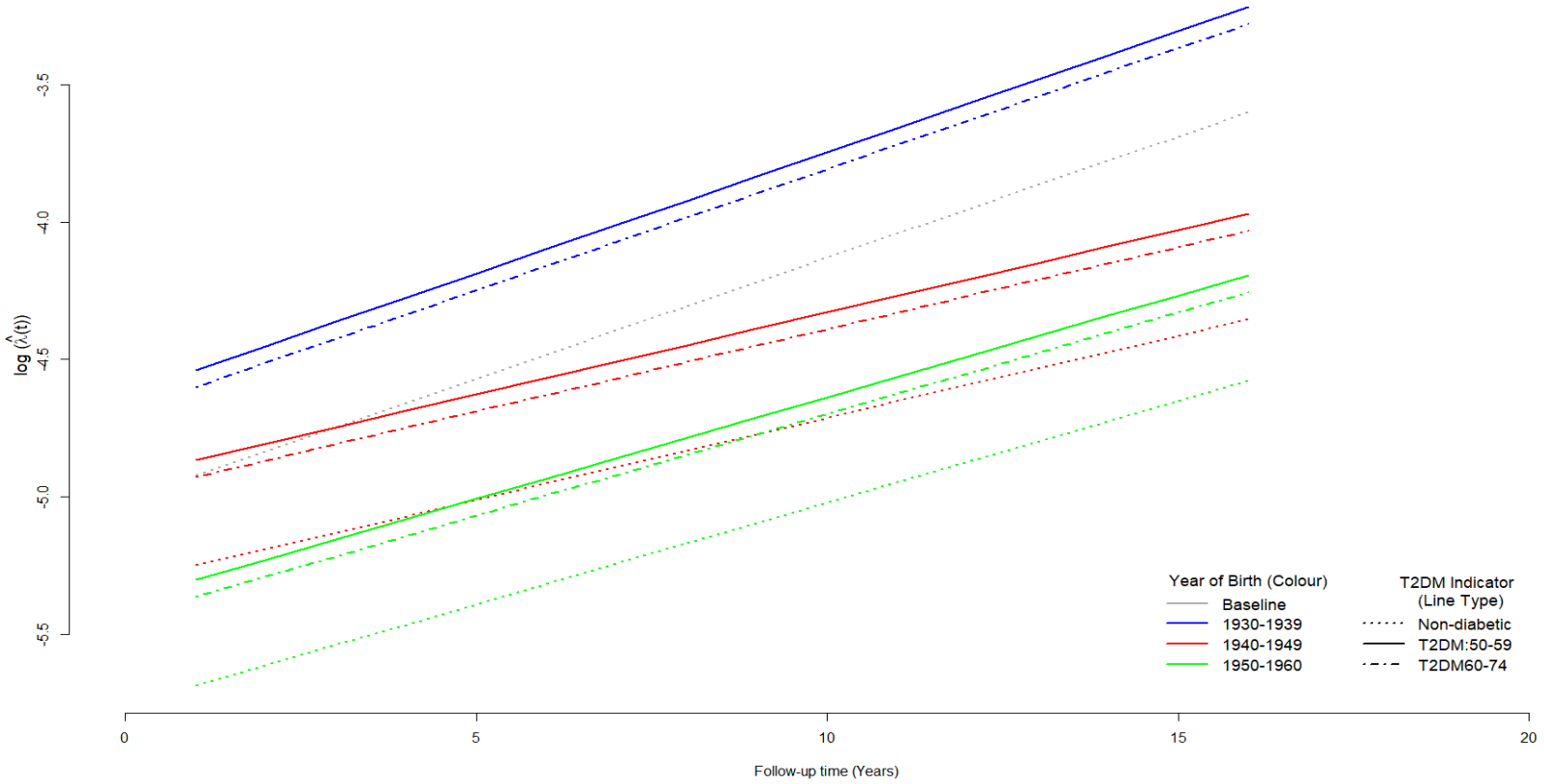


Fig. 2: Log scale mortality hazards by birth cohort and T2DM diagnosis

AVIVA

University of East Anglia

Actuarial Research Centre

Institute and Faculty of Actuaries

Shiny Application

- estimates the life expectancy of an individual with given conditions from the entry age (diagnosis age for T2DM) to $\omega = 100$ years of age.
- uses age as a time scale.
- plots the hazard, cumulative hazard or survival functions.
- has an option to plot the hazard or cumulative hazard functions at the log scale.

AVIVA University of East Anglia Actuarial Research Centre Institute and Faculty of Actuaries

Shiny Application

- Application has been published for the purpose of this presentation only. It won't be accessible until it has been fully developed.

Actuarial Translations

AVIVA

University of East Anglia

Actuarial Research Centre
Institute and Faculty of Actuaries

Actuarial Research Centre

Institute and Faculty of Actuaries

The Effect of HRT on the Survival of UK Women: A Retrospective Cohort Study 1984-2017

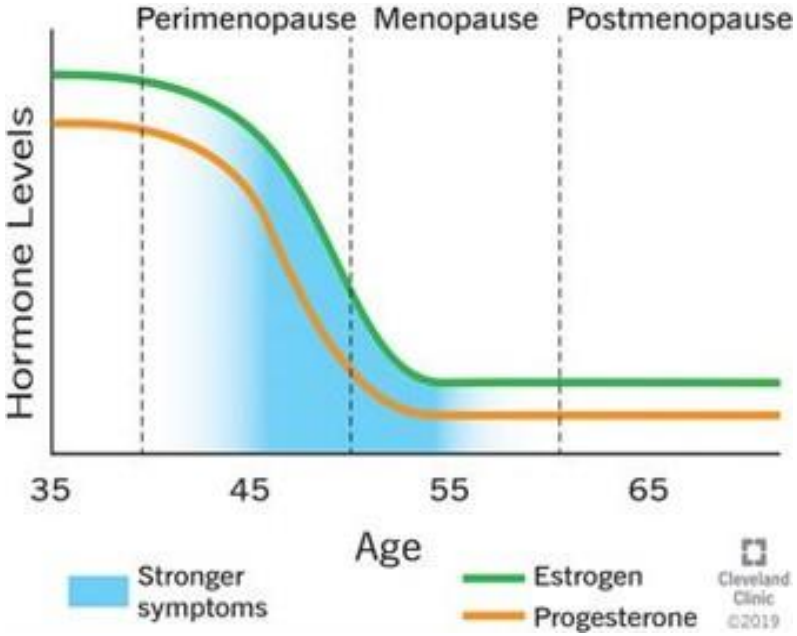
Nurunnahar Akter

Co-investigators: Ilyas Bakbergenuly, Nicholas Steel, Elena Kulinskaya
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.

Hormone Replacement Therapy (HRT)

- Around 80% women in the western countries suffer from menopausal symptoms
- HRT is mainly used to treat menopausal symptoms caused by deficiency of female sex hormones oestrogen and progesterone
- First available in the United Kingdom in 1965
- The routes of administration are oral tablets, transdermal patches, injections, topical gels and ointments.



(British Menopause Society, 2019)

AVIVA

University of East Anglia

Actuarial Research Centre
Institute and Faculty of Actuaries

Study design and model selection

- Patients prescribed any kind of oral and/or transdermal HRT at age 46 years or above are the exposures
- Outcome is death from any cause after starting HRT
- Any kind of cancer, acute myocardial infraction (AMI), serious heart failure, stroke (except TIA), chronic kidney disease (CKD) stage 3-5, dementia, oophorectomy before 45 years, premature ovarian insufficiency, premature menopause, and surgically induced menopause are excluded
- Analysis included **105,199** cases who born between 1921 to 1960, and started HRT between 46 to 65 years of age, and **224,643** matched controls
- Length of follow-up was up to 32 years between 1984-2017
- Covariates included in the final model are type of HRT, birth cohort, age at HRT, deprivation status, hypertension and its treatments, coronary heart disease, uterine/ovarian status, interaction of smoking with body mass index (BMI), and smoking with type II diabetes
- Incomplete records in BMI, smoking, deprivation status, systolic/diastolic blood pressure were dealt with by multilevel multiple imputation using the '**jomo**' package in R programming software.

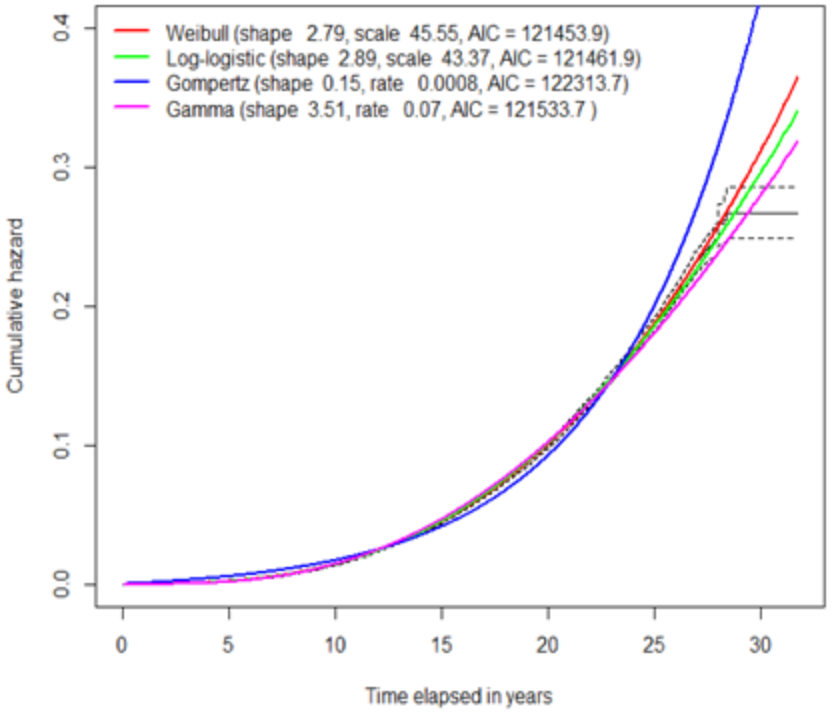
AVIVA

University of East Anglia

Actuarial Research Centre
Institute and Faculty of Actuaries

Weibull-Double-Cox survival model

- Proportional hazards (PH) assumptions were checked by Grambsch and Therneau test
- Birth cohort and age category at study entry violated the PH assumptions
- Baseline hazards of the study population were fitted with different parametric distributions
- Weibull distribution provided the best fit to baseline hazards
- Weibull-Double-Cox model was fitted to handle the non-proportional hazards.



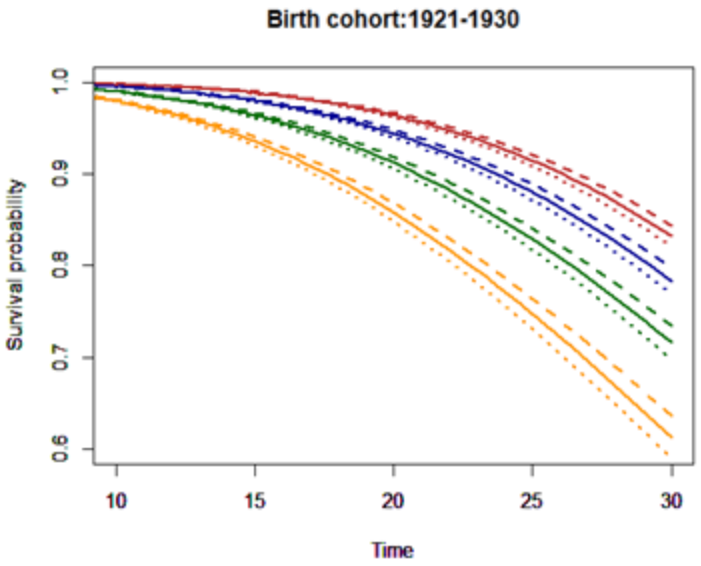
The value of Akaike's Information Criterion (AIC) is an indication of the goodness-of-fit. The lower the AIC the better the fit.

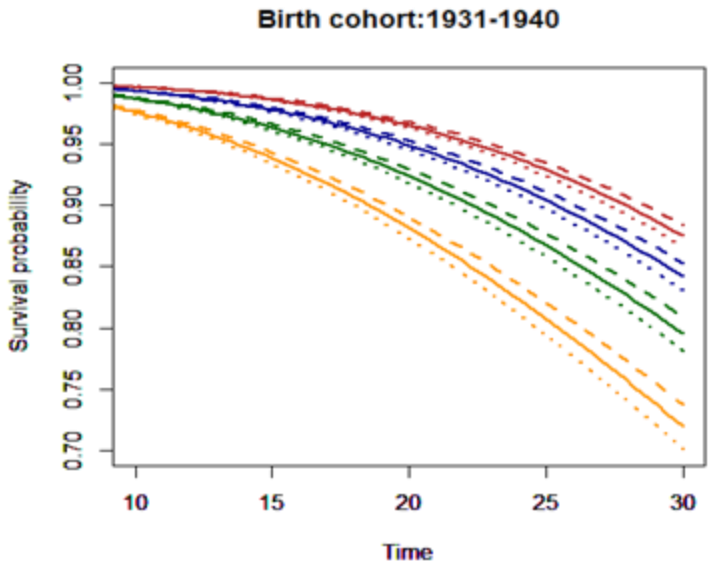
AVIVA

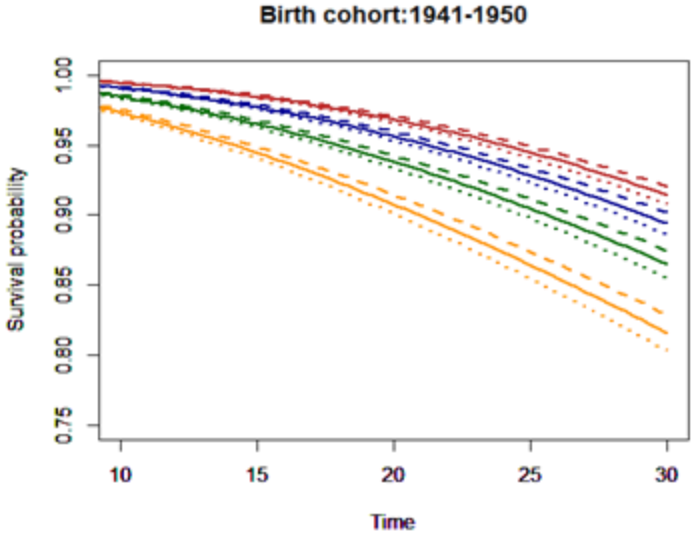
University of East Anglia

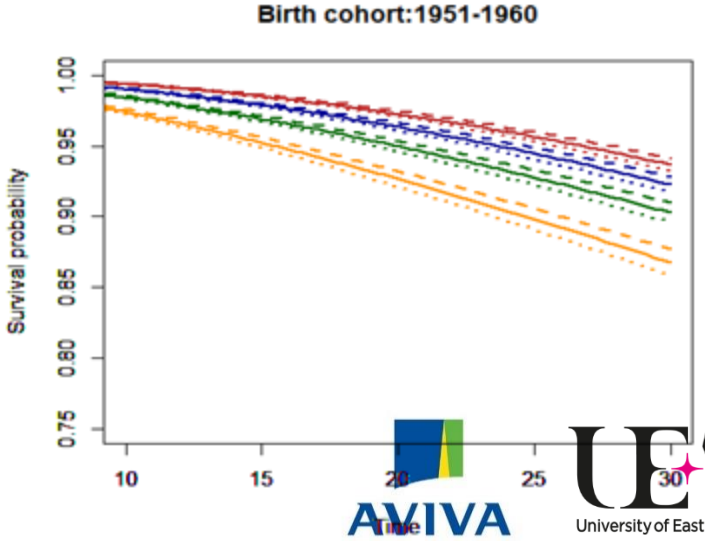
Actuarial Research Centre
Institute and Faculty of Actuaries

Survival from study entry by birth cohort









Red: Age 46-50
Blue: Age 51-55
Green: Age 56-60
Yellow: Age 61-65

Dashed lines: Combined HRT

Dotted lines: Oestrogen-only HRT

Solid lines: Non-users

Time: time from study entry

- Age groups are based on age at first HRT treatment
- Survival prospects improved in later birth cohorts

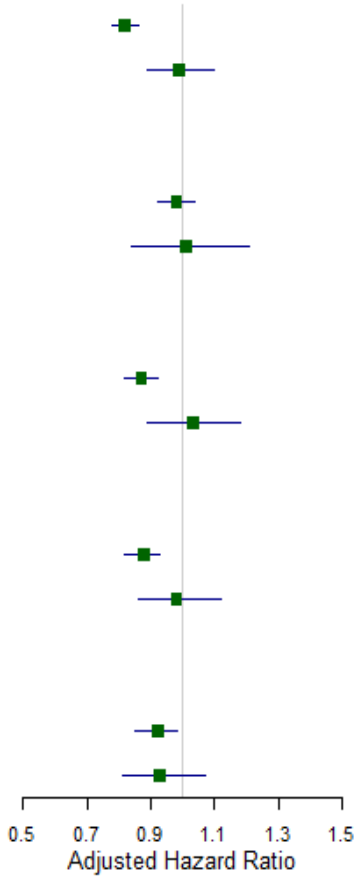
AVIVA

University of East Anglia

Actuarial Research Centre
Institute and Faculty of Actuaries

Age-specific HRT effects

Age cohort	HRT treatment	Adjusted HR (95% CI)
Overall	No (reference group)	
	Combined HRT	0.91 (0.88-0.94)
	Estrogen-only	0.99 (0.93-1.07)
46-50	No	
	Combined HRT	0.98 (0.92 - 1.04)
	Estrogen-only	1.01 (0.84 - 1.21)
51-55	No	
	Combined HRT	0.87 (0.82-0.92)
	Estrogen-only	1.03 (0.89-1.18)
56-60	No	
	Combined HRT	0.88 (0.82-0.93)
	Estrogen-only	0.98 (0.86-1.12)
61-65	No	
	Combined HRT	0.92 (0.85 - 0.98)
	Estrogen-only	0.93 (0.81 - 1.07)



- Adjusted effects of HRT on all-cause mortality were estimated on full data and age subgroups at first treatment separately
- No effect of estrogen-only HRT
- Overall, combined HRT reduces mortality by 9%
- Highest reduction (13%) was in women who started treatment at age 51 to 55 years
- No significant reduction in combined HRT users of age group 46 to 50 at first treatment.

AVIVA

University of East Anglia

Actuarial Research Centre
Institute and Faculty of Actuaries

Ongoing work

- Development of a survival model with age as a time-scale
- Calculation of life expectancy using the age model.

AVIVA

University of East Anglia

Actuarial Research Centre
Institute and Faculty of Actuaries

Actuarial Research Centre

Institute and Faculty of Actuaries

R Software

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.

09/03/2021

Objectives

- Introduce R package "mylongevity" for calculation of life expectancies
- Functions using standard Cox regression and results from landmark analysis
- Functions using double-Cox parametric survival model
- Example of using functions from the R package
- Summary

AVIVA University of East Anglia Actuarial Research Centre Institute and Faculty of Actuaries

Mylongevity

- R package mylongevity is available in GitHub repository "ilyasstatistics/mylongevity".
- The package can be installed by using an R command - devtools::install_github("ilyasstatistics/mylongevity")
- In order to install the package a software Rtools40 has to be installed in advance
- R package mylongevity consists of 5 main functions available to user

AVIVA

University of East Anglia

Actuarial Research Centre
Institute and Faculty of Actuaries

Function mylongevity_method1()

- mylongevity_method1 (data, indexes_of_variables, working_directory)
- This method matches the attributes of inputted data with results of life expectancies obtained from the landmark analysis as described in Kulinskaya et al. [4]. For a set of given attributes, this method produces a table with life expectancies at a given age similar to those provided on the website https://mylongevity.org/, see also Kulinskaya et al. [3].

AVIVA

University of East Anglia

Actuarial Research Centre
Institute and Faculty of Actuaries

Function mylongevity_method2()

- mylongevity_method2 (data, indexes_of_variables, list_of_variables, age, a, b, working_directory)
- This method calculates life expectancies using the log-hazard ratio estimates from the landmark analysis Kulinskaya et al. [4] and the weights of the risk groups are estimated from the given dataset. These weights correspond to frequencies of risk profiles in the dataset.
- a and b are the Gompertz shape and scale parameters supplied by the user.

AVIVA

University of East Anglia

Actuarial Research Centre
Institute and Faculty of Actuaries

Function mylongevity_method3()

- mylongevity_method3 (data, indexes_of_variables, list_of_variables, age, a, b, T_start_indicator, T_stop_indicator, status_indicator, working_directory)
- This method fits the Cox regression to the provided data, eliminates non-significant terms, calculates weights for each combination of risk factors in the final Cox regression model and computes life expectancies based on the methodology from Kulinskaya et al. [3].
- a and b are the Gompertz shape and scale parameters supplied by the user.

AVIVA

University of East Anglia

Actuarial Research Centre
Institute and Faculty of Actuaries

Data entry

Age: – years

Gender : values "M" for males and "F" for females

Townsend: deprivation index 1 (least deprived), 2, 3, 4, 5 (most deprived).

smokerCategory : 1 - no smoker, 2 - ex smoker, 3 – smoker

htn : 1 – no-hypertension, 2 - treated hypertension, 3 – untreated hypertension

diabetesCategory : 0 – no diabetes, 1 – diabetes

hcl : 0 – no hypercholesteromia, 1 – hypercholesteromia

bmiCategory: 1 – Healthy weight (BMI<25), 2 – Overweight (BMI≥25 and BMI<30), 3 – Obese (BMI≥30)

qRiskCategory: 2 - low risk (QRISK2<20), 0 - moderate risk (QRISK2≥20 and QRISK2<40), 1 - high risk (QRISK2≥40 or diagnosis of CVD)

Statins: 0 - no statin, 1 – yes statin

AVIVA

University of East Anglia

Actuarial Research Centre
Institute and Faculty of Actuaries

Data and results from mylongevity_method1()

```
> life_expectancy_table=mylongevity_method1(data=data_of_clients,indexes_of_variables=indexes_for_columns)
>
>
> life_expectancy_table
   age gender townsend smokerCategory htn diabetesCategory hcl bmiCategory qRiskCategory statins          b         a0        LE
1   75      M        2              1   1                 0   0           2              0       0 0.11250341 -12.285306 13.608318
2   83      M        3              2   3                 0   0           2              0       0 0.10568125 -10.777331  4.773028
3   66      M        5              2   1                 1   0           3              1       0 0.08613558  -8.800226 10.529908
4   83      M        1              2   1                 1   0           2              0       0 0.11764571 -12.081031  5.699709
5   79      M        5              2   3                 0   0           3              1       0 0.08613558  -8.978484  5.705869
6   85      F        1              3   2                 1   1           2              2       1 0.12311139 -12.839448  5.824969
7   71      F        1              3   2                 1   1           2              1       0 0.12311139 -11.984656  9.650149
8   70      F        5              3   2                 0   0           1              1       1 0.09637933 -10.102172 11.483401
9   78      F        3              1   1                 0   1           2              1       1 0.11162558 -12.492568 13.276157
10  72      F        2              2   1                 0   0           2              1       0 0.11735626 -12.185365 12.627901
>
```

AVIVA

University of East Anglia

Actuarial Research Centre
Institute and Faculty of Actuaries

Functions for double-Cox parametric survival model

- Similar functions double_cox_longevity1() and double_cox_longevity2() will be available for double-Cox parametric survival model
- These two methods fit double-Cox parametrical survival regression model and calculate life expectancies based on given parameters for Gompertz or Weibull baseline risk.

AVIVA

University of East Anglia

Actuarial Research Centre
Institute and Faculty of Actuaries

Functions for double-Cox parametric survival model

- double_cox_longevity1(data,dist,cluster,formula.shape, formula.scale,age_of_diagnosis,time_past_from_diagnosis, working_directory,name_for_age_factor=NULL)
- double_cox_longevity2(data,dist,cluster,formula.shape, formula.scale,age_of_diagnosis,time_past_from_diagnosis, working_directory,name_for_age_factor=NULL)
- dist is “Gomperz” or “Weibull” distribution
- cluster is the membership variable for shared frailty

AVIVA

University of East Anglia

Actuarial Research Centre
Institute and Faculty of Actuaries

Summary

- The functions in mylongevity package can be applied to dataset of interest in order to provide survival analysis and to calculate life expectancies based on our research
- The functions could be applied with our existing models and results, or to develop new survival models from scratch.
- The package will be publicly available in Github

AVIVA

University of East Anglia

Actuarial Research Centre
Institute and Faculty of Actuaries

Actuarial Research Centre

Institute and Faculty of Actuaries

Use of Primary Health Care Records Data in Actuarial Research

Summary and discussion

PI Elena Kulinskaya (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.

9/03/21

Observation studies vs Clinical trials

- The randomized controlled trials are essential to evaluate the effectiveness of medications and to obtain regulatory approval for their use in clinical practice.
- Yet, RCTs lack generalizability and long-term effects. This is due to stringent entry criteria and short follow-up.
- Observational studies fill this gap by assessing long-term effects of medications on infrequent outcomes, including mortality.
- The use of computerized health databases, has led to an explosion in the last decade in the number of published observational database studies of the impact of medications.

[S.Suissa, https://doi.org/10.1093/aje/kwm324]

AVIVA

University of East Anglia

Actuarial Research Centre
Institute and Faculty of Actuaries

Possible biases in observation studies

- Well-known biases in observation studies are the "immortality bias“ and the “bias by indication“.
- Immortal time in observational studies can bias the results in favour of the treatment group. It refers to a period of follow-up during which, by design, death or the study outcome cannot occur (i.e. waiting time for a treatment).
- The “bias by indication“ arises when GPs prescribe drugs to sicker people.
- Other (self)selection biases, when say better-off people get more treatments.
- Unavoidable unmeasured confounding – by e.g. family history, social factors, etc that are just not there in primary care data

AVIVA University of East Anglia Actuarial Research Centre Institute and Faculty of Actuaries

Target Conditions and interventions

AVIVA University of East Anglia Institute and Faculty of Actuaries

Design of an observational study is chosen to minimize possible biases

- We take a lot of care at the design stage.
- Diabetes/ Stroke/ HRT studies: case-control studies from the date of diagnosis or treatment. Controls matched by age/sex/GP practice.
- Selection biases are partly eliminated by matching, and mostly by careful statistical modelling which includes all available confounders and sophisticated imputation to fill missing values.

AVIVA

UEA University of East Anglia

Actuarial Research Centre
Institute and Faculty of Actuaries

Missing data and Multiple imputation

- Missing data are unavoidable in big health databases, potentially leading to bias and loss of precision.
- Informative missingness can create selection bias when using complete case analysis.
- Appropriate Multiple imputation (MI) methods are used to reduce these selection biases.
- We use multivariate MI methods to generate 10 imputed datasets, and then amalgamate their results using Rubin's rules.

AVIVA

University of East Anglia

Actuarial Research Centre
Institute and Faculty of Actuaries

Models for actuarial implementation

- We spend a lot of time and effort trying to obtain good models for actuarial implementation of our results.
- Two designs:
  - People are followed up from a fixed age A (our landmark analysis)
    - ❖ Time scale A+t directly corresponds to age.
  - Case-control studies from the date of diagnosis or treatment.
    - ❖ time scale is age or time from study entry; ages at entry differ.

AVIVA University of East Anglia Actuarial Research Centre Institute and Faculty of Actuaries

Time scale matters for model interpretability

- Our original models for chronic medical conditions used time from the study entry as the timescale and birth cohort and age-of-diagnosis cohort as extra predictors. This parametrisation is useful for medical applications.
- For actuarial implementation, we need the models with continuous age. We attempted to change the timescale to age but could not built interpretable models. This happens because age at study entry is the age of diagnosis and the survival time is not independent of it.
- We now reverted to models with continuous age of entry as a predictor. We use these models to obtain relative survival benefits/harms.

AVIVA

University of East Anglia

Actuarial Research Centre
Institute and Faculty of Actuaries

Summary

Aims: Development of novel statistical and actuarial methods for modelling mortality and morbidity and evaluating longevity improvement based on Big Health and Actuarial Data.

Target conditions and interventions: Stroke, Diabetes, MI, Hip replacement, Hypertension, statin and HRT use.

Novel methodology: Use of the Cox regression, Landmark analysis and Double-Cox modelling in actuarial calculations.

Implementation:

mylongevity app (214,237 users to end January 2021)

R software.

AVIVA

University of East Anglia

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.

www.actuaries.org.uk/arc

9 March 2021

Questions

Comments

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

AVIVA

University of East Anglia

Actuarial Research Centre
Institute and Faculty of Actuaries