The authors declare no competing interests.

# A lifetime cardiovascular model developed using clinical trials and calibrated in large population cohort

**Runguo Wu**[1], Claire Williams[2], Iryna Schlackow[2], Junwen Zhou[2], Jonathan Emberson[2], Christina Reith[2], Anthony Keech[4], John Robson[1], Kenneth Wilkinson[3], Jane Armitage[2], Rory Collins[2], Alastair Gray[2], John Simes[4], Colin Baigent[2], Borislava Mihaylova[1,2]

[1]Wolfson Institute of Population Health, Queen Mary University of London, London, UK
[2]Nuffield Department of Population Health, University of Oxford, Oxford, UK
[3]Oxfordshire, UK; [4]NHMRC Clinical Trials Centre, University of Sydney, Sydney, Australia

# Outline

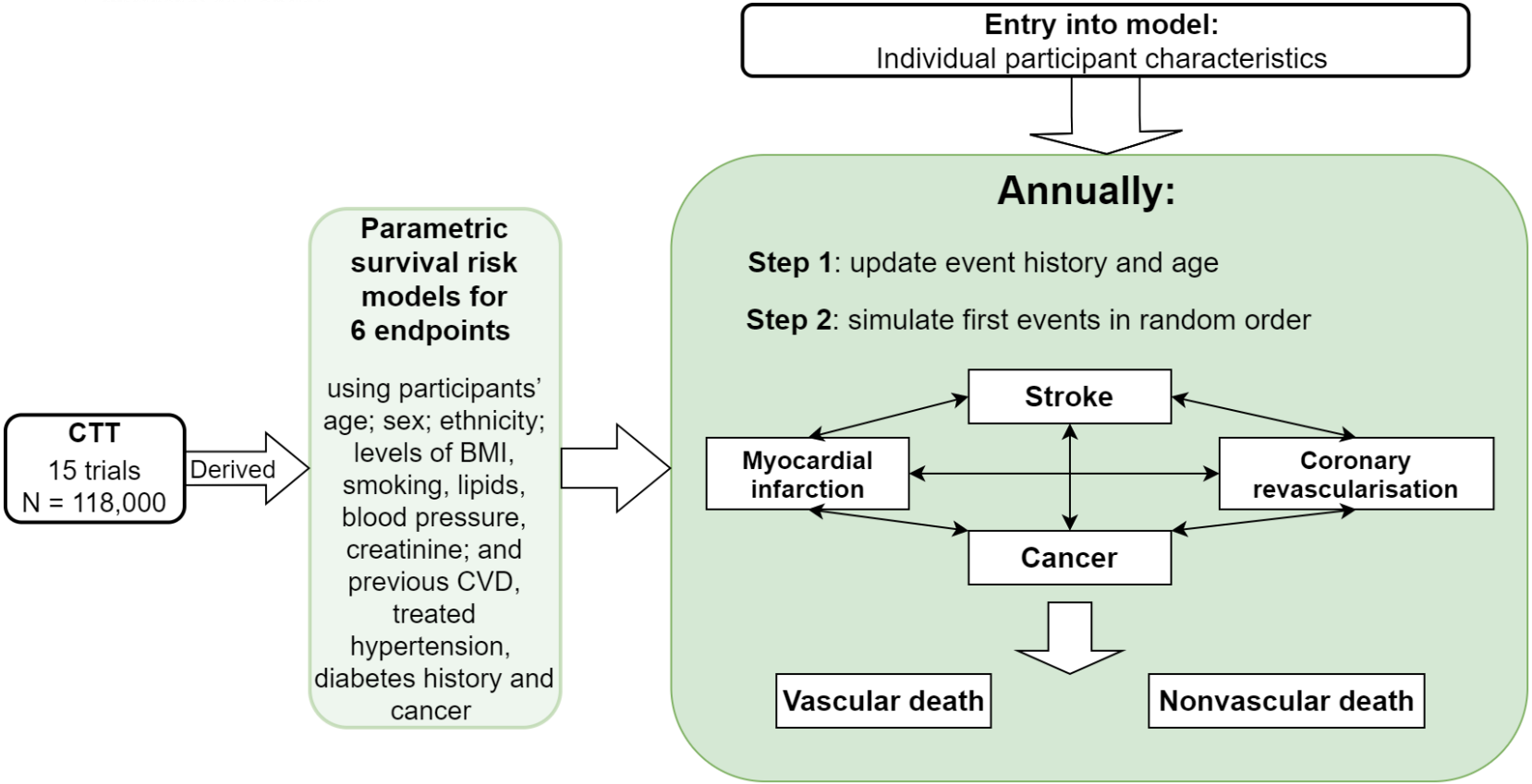
- Aim: an approach of developing a model in one database, and transferring it to another.
- A micro-simulation CVD model developed using individual participant data (IPD) from trials.
- Calibrated and further developed using a large contemporary middle-aged UK population cohort.

# Databases

CTT Collaboration
Cholesterol Treatment Trialists' Collaboration

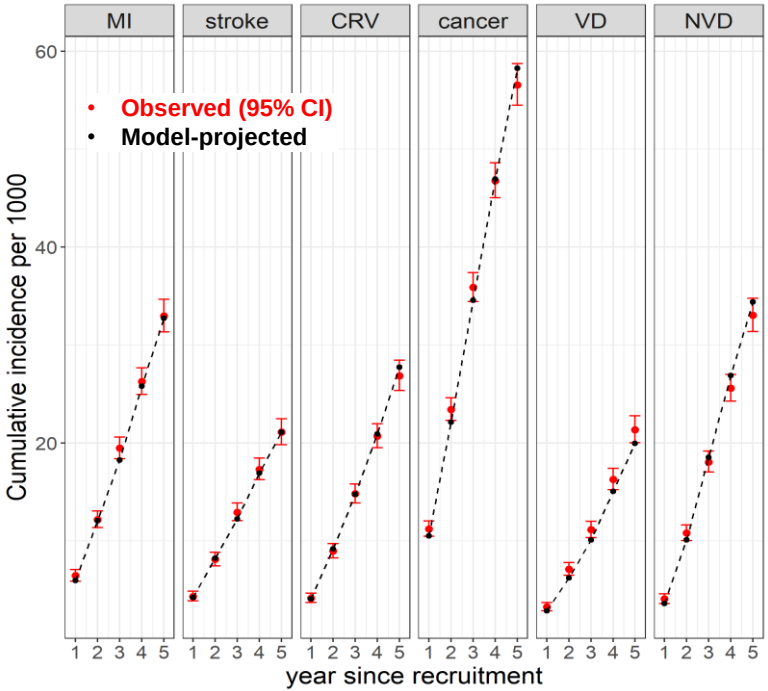
- 15 large trials across multiple countries (from 1980s to 2010)
- 110,000 individuals
- Detailed patient characteristics at entry and vascular, cancer events, mortality during follow-up
- About 5 years of follow-up

# CTT discrete event micro-simulation model

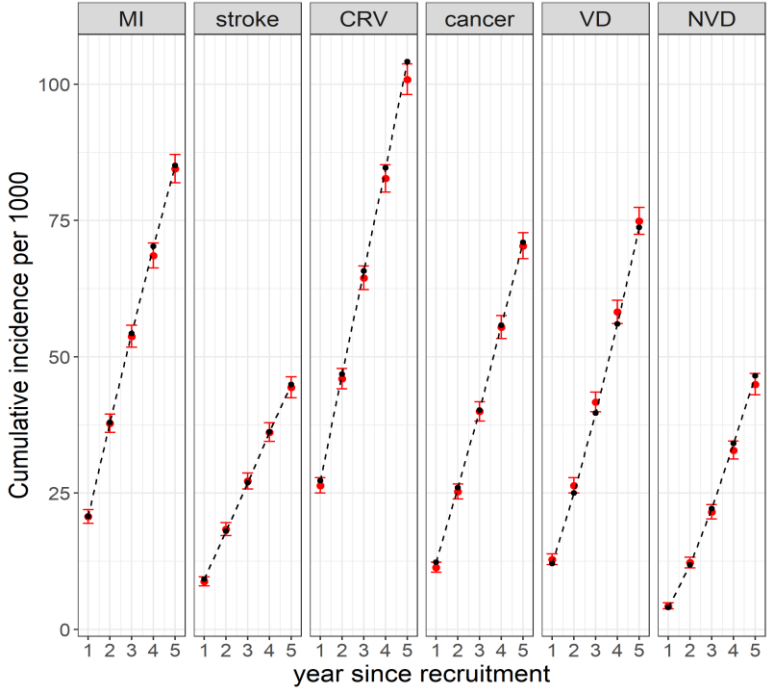


# Internal validation in CTT

**Without previous CVD history**




**With CVD history**




MI: myocardial infarction; CRV: coronary revascularisation; VD: vascular death; NVD: nonvascular death

# Databases

CTT Collaboration
Cholesterol Treatment Trialists' Collaboration

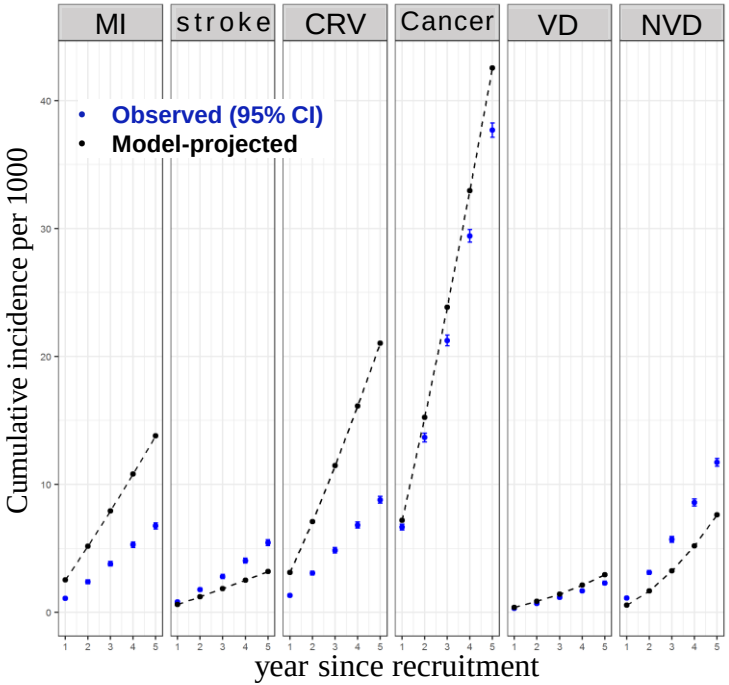
- 15 large trials across multiple countries (from 1980s to 2010)
- 110,000 individuals
- Detailed patient characteristics at entry and vascular, cancer events, mortality during follow-up
- About  5 years of follow-up



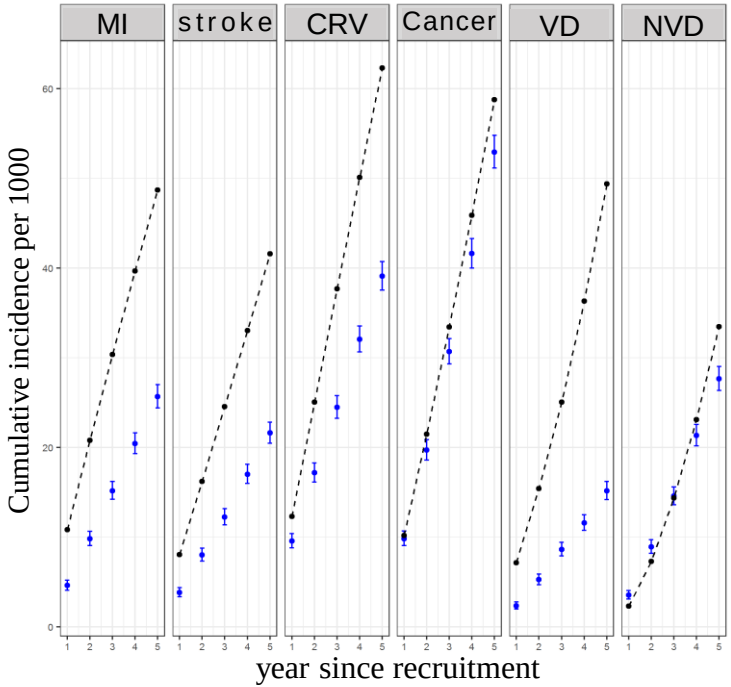

- A large cohort enrolled in 2006 - 2010
- 500,000 UK participants
- Very detailed individual characteristics at entry and linked electronic health records of primary and inpatient care, cancer registries and mortality records
- About 9 years of follow-up until 2020

# CTT model external validation in UKB data

**Without previous CVD history**




**With CVD history**




MI: myocardial infarction; CRV: coronary revascularisation; VD: vascular death; NVD: nonvascular death

# Poor performance due to different populations

- **Different countries**
  - CTT is a combination of multinational trial populations
  - UKB is a (healthier) UK cohort
- **Calendar time:** UKB data collection started 10-15 years later than CTT trials
- **Different distributions of patient sociodemographic and clinical risk factors**

| | Without previous CVD | | With previous CVD | |
|---|---|---|---|---|
| | **UKB** | **CTT** | **UKB** | **CTT** |
| | **N = 444,852** | **N = 68,018** | **N = 57,641** | **N = 49,878** |
| | **Mean (SD) / % count** | | | |
| Age at recruitment | 56 (8.1) | 62.26 (9.21) | 60.4 (7) | 62.66 (9.13) |
| Female | 56% | 35.4% | 41% | 21.6% |
| White | 95% | 72.3% | 95% | 78.0% |
| Black | 2% | 9.0% | 1% | 2.5% |
| Other ethnicity | 4% | 18.7% | 4% | 1.5% |
| Current smoker | 10% | 20.4% | 12% | 19.2 |
| Baseline diabetes | 4% | 22.2% | 13% | 15.9% |
| Hypertension treatment | 16% | 52.2% | 46% | 51.1% |
| Baseline cancer | 7% | 0.0% | 10% | 0.1% |
| CVD history | | | | |
| Other CHD only | | | 50% | 19.3% |
| MI only | | | 4% | 49.9% |
| PAD only | | | 12% | 6.4% |
| Stroke only | | | 9% | 7.8% |
| Two or more | | | 25% | 16.7% |
| LDL (mmol/L) | 3.6 (0.82) | 3.5 (0.89) | 3.1 (0.87) | 3.8 (0.85) |
| HDL (mmol/L) | 1.5 (0.37) | 1.3 (0.38) | 1.3 (0.36) | 1.1 (0.31) |
| Creatinine (umol/L) | 72 (15) | 91 (24) | 78 (32) | 98 (23) |
| Blood pressure (mmHg) | 138 (19)/82 (10) | 142 (20)/83 (11) | 139 (19)/81 (10) | 139 (22)/81 (11) |

# UKB contribution in model development

1. **Contemporary UK population** → Useful to calibrate model for use in UK
2. **More detailed information of patient characteristics, e.g. smoking and ethnicity** → Refine the existing patient characteristics
3. **Further patient characteristics e.g. diet, physical activity, mental health, deprivation** → Introduce new patient characteristics
4. **Further disease endpoints during follow-up (e.g. incident diabetes)** → Add further endpoints of interest

# Calibration of risk equations (steps 1-3)

- **Calibrate intercept and shape**

$$f(y_{new}) = a + b_{overall} * XB$$

- **Modify the definition of covariate and recalibrate its slope**

$$f(y_{new}) = a + b_{overall} * XB_{sub} + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n$$

- **Add covariates of interest only available in new data**

$$f(y_{new}) = a + b_{overall} * XB + b_1 cov_1 + b_2 cov_2 + \ldots + b_n cov_n$$

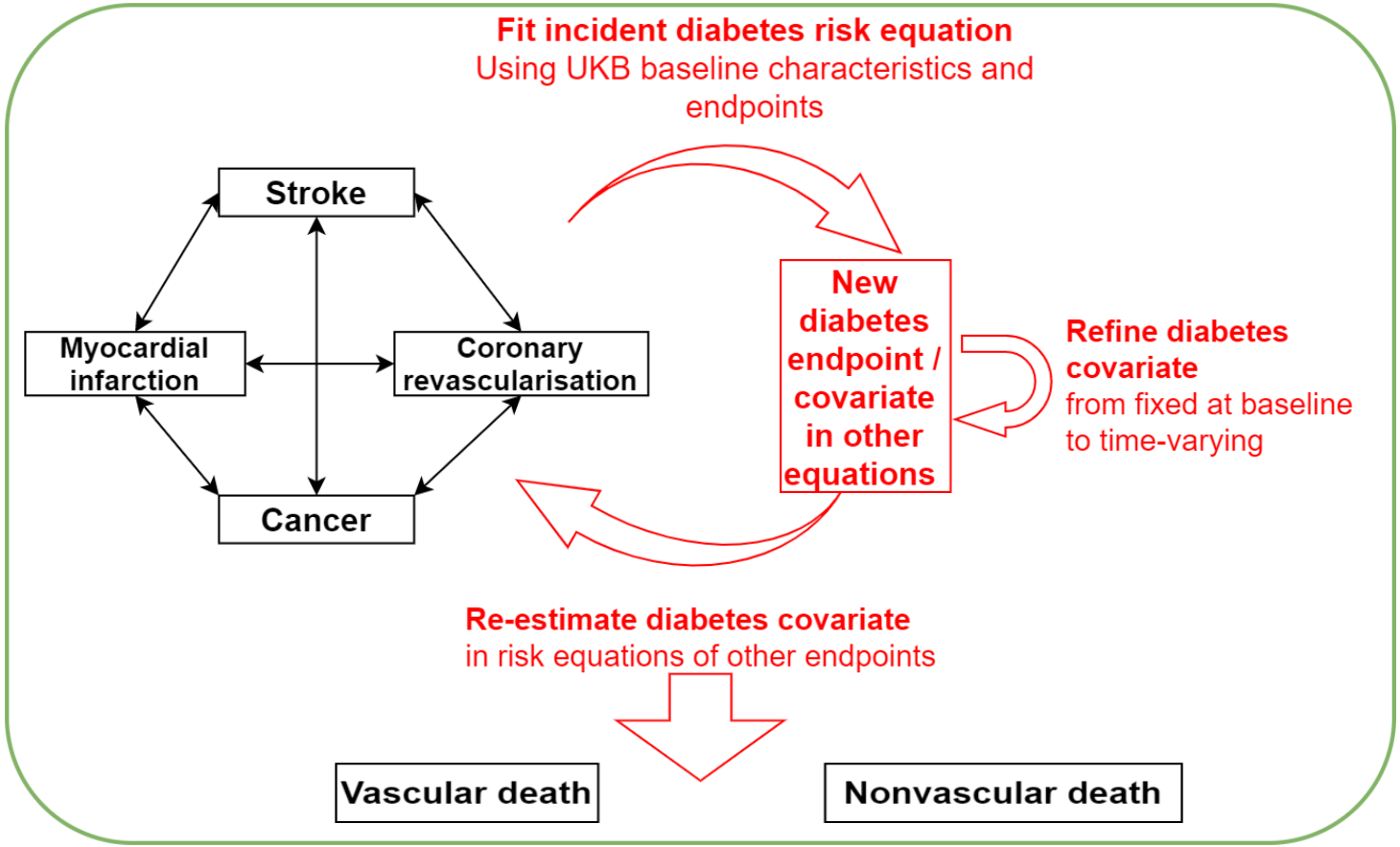
Smoking; ethnicity; Diabetes; Cancer; Age and sex (if needed)

Physical activity; diet quality; mental illness; deprivation; Type 1 diabetes

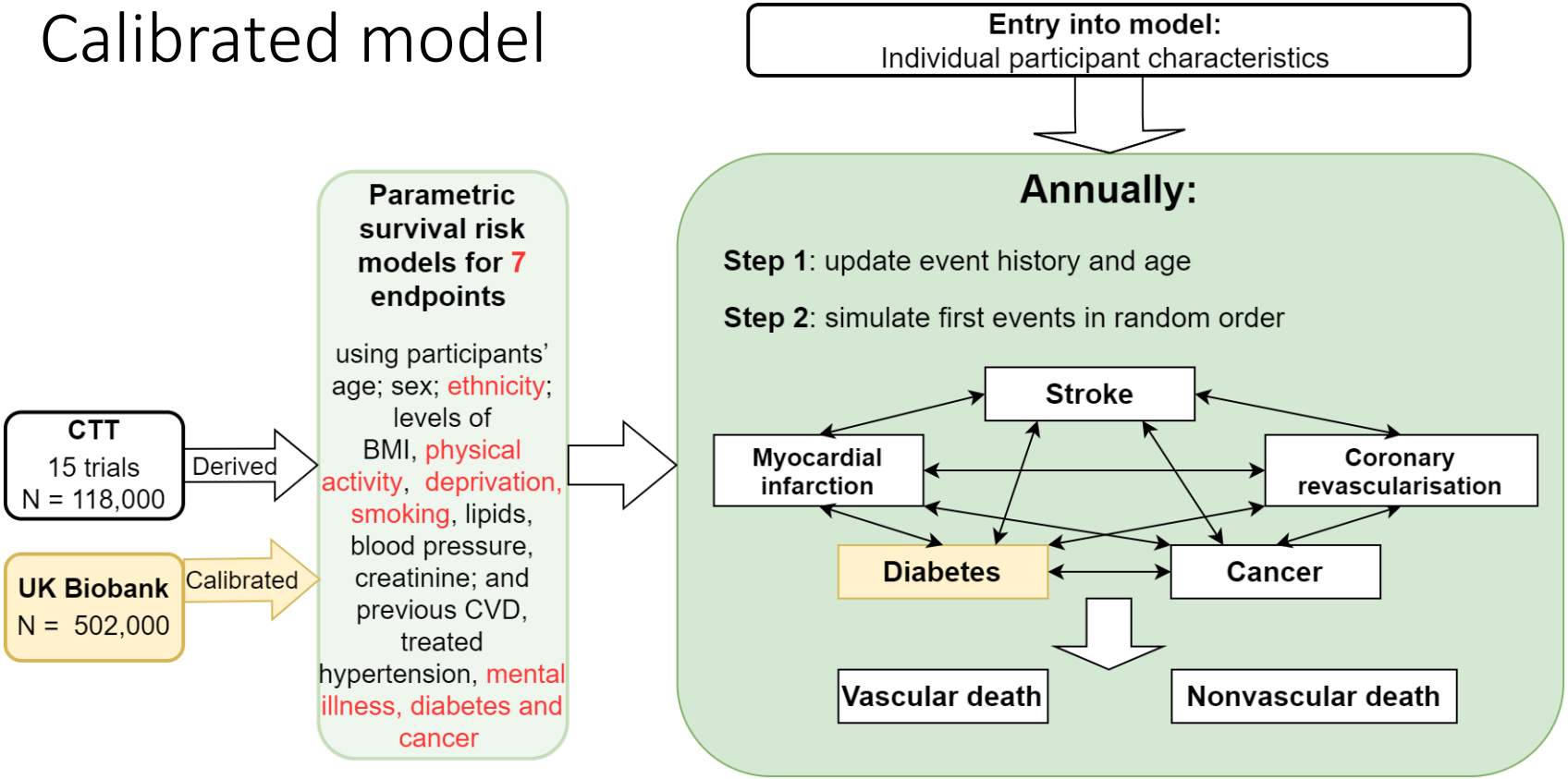
$f(y_{new})$ is a parametric survival function of time to event data; $a$ is the intercept; $XB$ is the linear predictor as the combination of CTT regression coefficients and the predictor values in the new data; $b_{overall}$ is the scaling factor to be estimated and to adjust CTT coefficients; $XB_{sub}$ is the linear predictor excluding covariates $x$; $cov$ are the new covariates.

NOTE: The approach relies on proportional hazard assumption holding for associations between risk factors and endpoints.

# Step 4: add diabetes endpoint and refine diabetes covariate

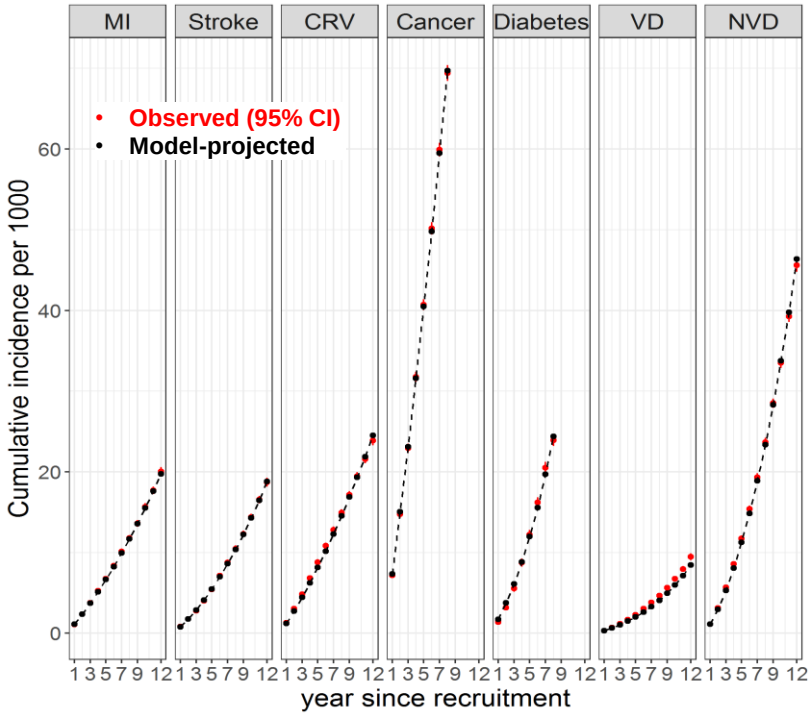


# Overview of the Calibrated model

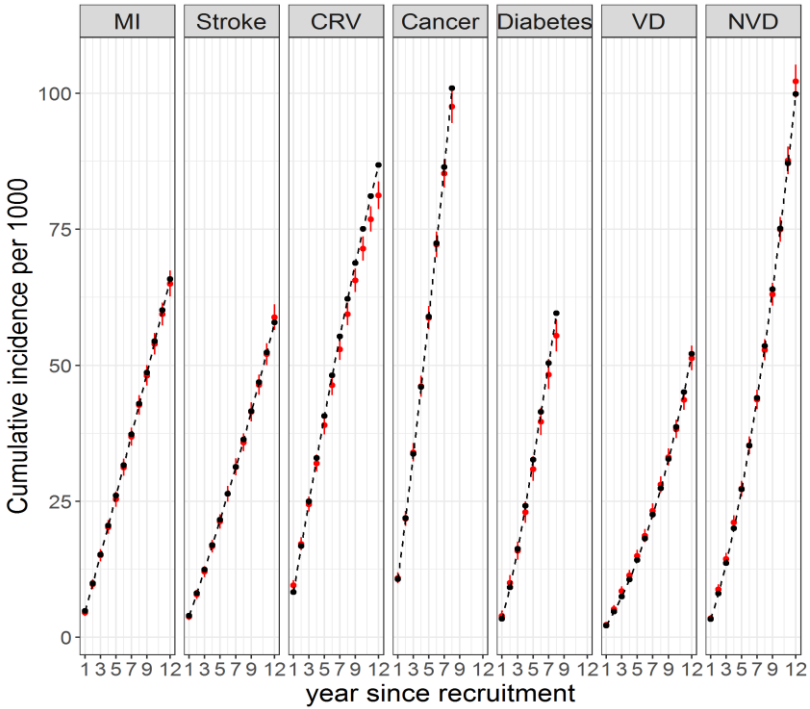


# Model fit in UKB following calibration

**Without previous CVD history**

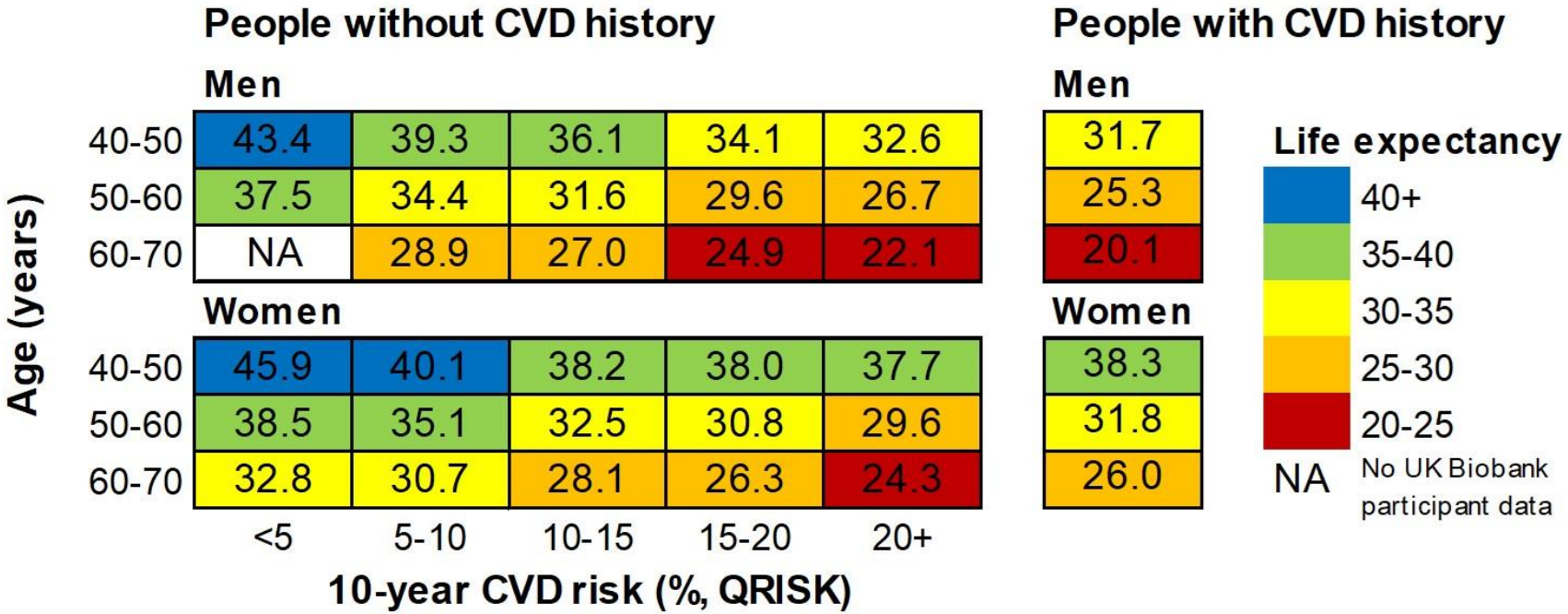



**With previous CVD history**




MI: myocardial infarction; CRV: coronary revascularisation; VD: vascular death; NVD: nonvascular death

# Model-predicted life expectancy, UKB (years)



# Summary

- As health data are increasingly available, transferring disease prediction models across populations is practically useful
- We present an approach of transferring a model between two large patient databases that differ in many characteristics
- The approach relies on proportional hazards assumption holding for associations between risk factors and endpoints
- The process includes extending model endpoints, modifying and adding covariates in calibration, and checking model fit

# Acknowledgement

This work is part of the NIHR HTA project 17/140/02 Cost-effectiveness of statin therapies evaluated using individual participant data from large randomised clinical trials.

The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.



# Thank you!

For further information please contact

Runguo Wu

r.wu@qmul.ac.uk

Queen Mary
University of London
Barts and The London