OXFORD POPULATION HEALTH

UNIVERSITY OF OXFORD

HERC

Estimating Costs Associated with Disease Model States Using Generalized Linear Models in R

Presented by **Junwen Zhou**
Senior Researcher in Health Economics
Health Economics Research Centre, University of Oxford
junwen.zhou@ndph.ox.ac.uk

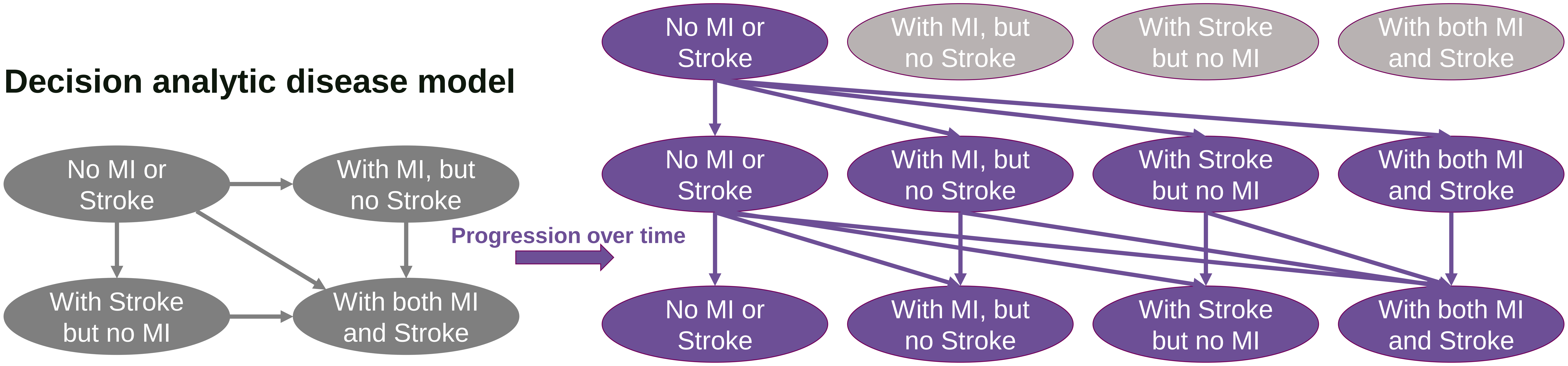
1 July 2024

R for HTA 2024

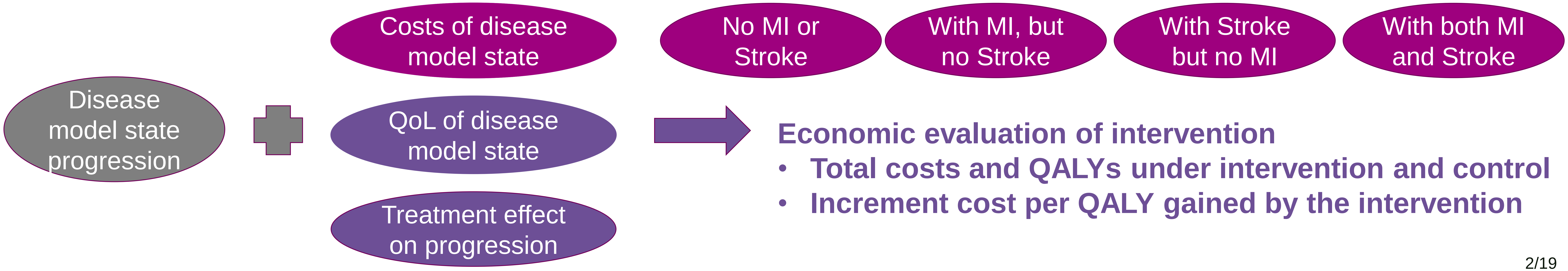
Background

Costs associated with disease model state

Decision analytic disease model



Additional information needed for economic evaluation


Disease model state progression
Costs of disease model state
QoL of disease model state
Treatment effect on progression
No MI or Stroke
With MI, but no Stroke
With Stroke but no MI
With both MI and Stroke

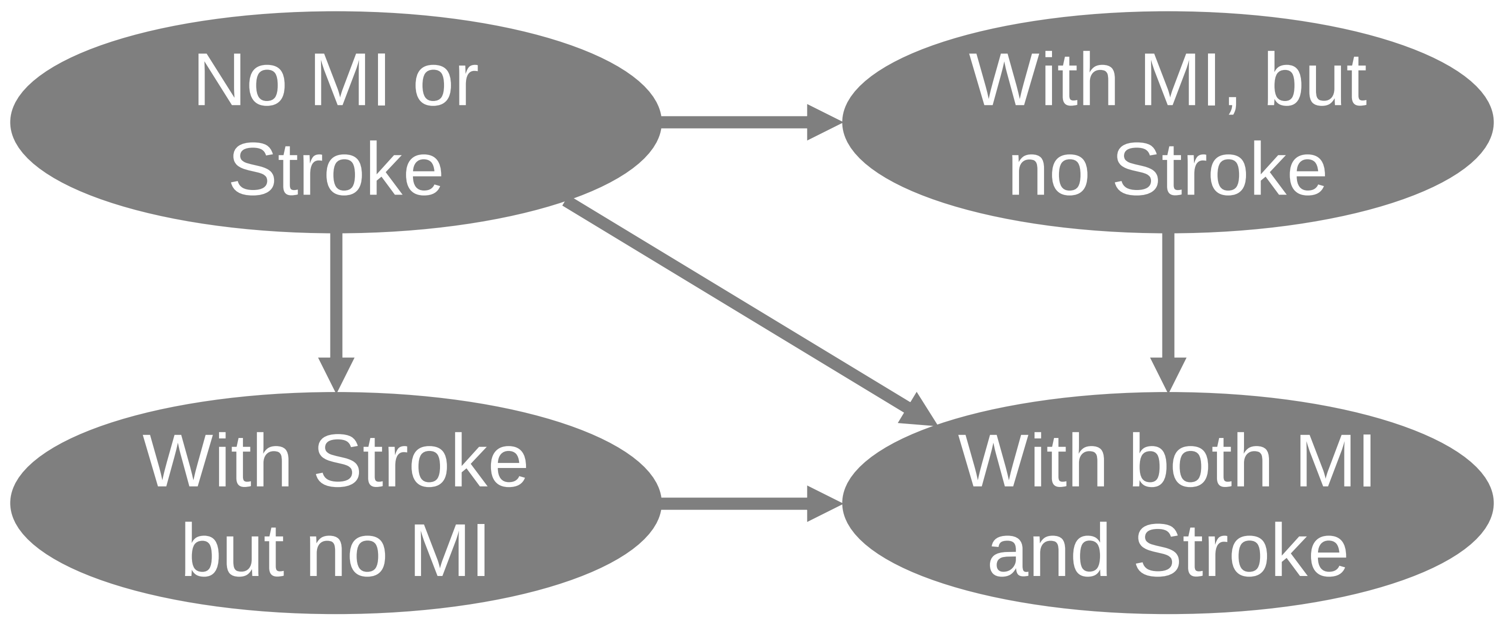
Economic evaluation of intervention

- **Total costs and QALYs under intervention and control**
- **Increment cost per QALY gained by the intervention**

Background

Costs associated with disease model state

Heterogeneity costs of disease model state



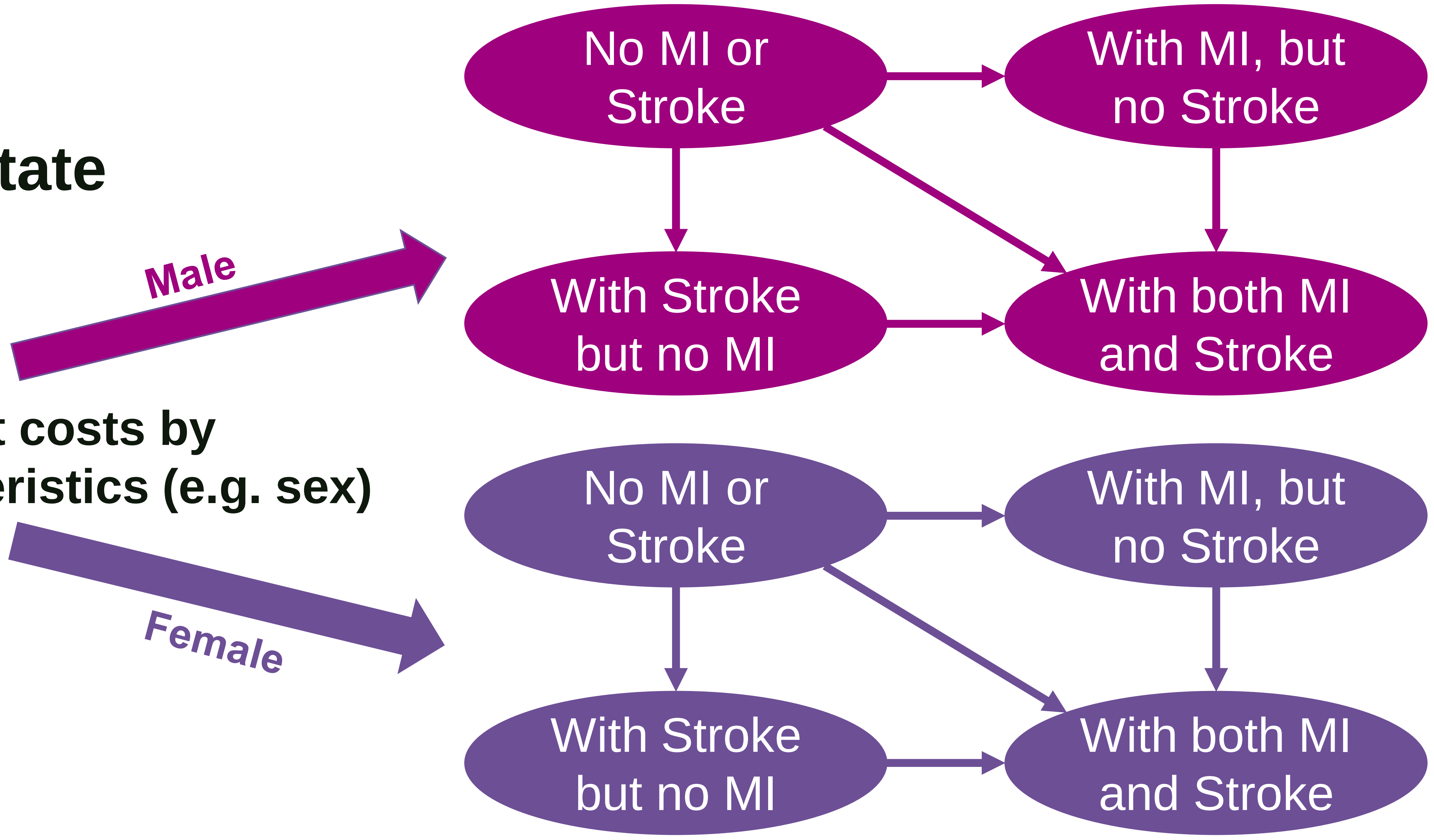
No MI or Stroke

With MI, but no Stroke

With Stroke but no MI

With both MI and Stroke

Different costs by characteristics (e.g. sex)



Male

No MI or Stroke

With MI, but no Stroke

With Stroke but no MI

With both MI and Stroke

Female

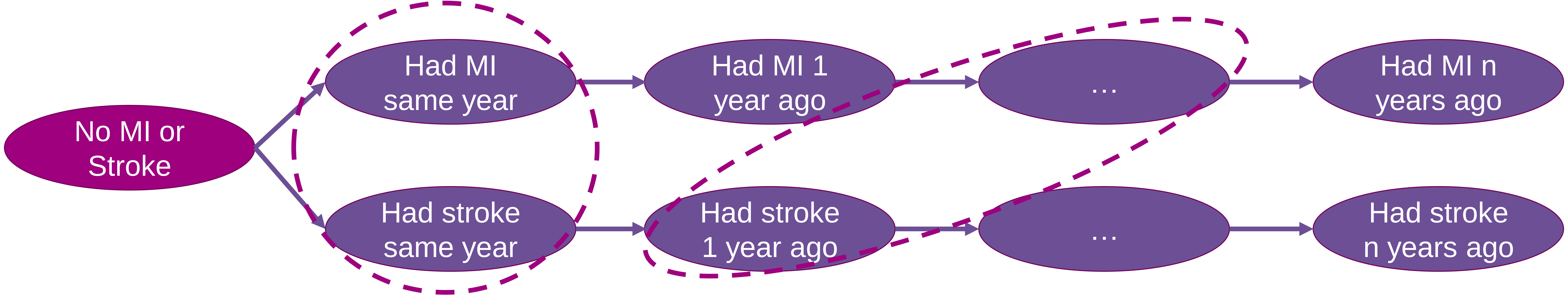
No MI or Stroke

With MI, but no Stroke

With Stroke but no MI

With both MI and Stroke

Different costs by complicate disease states



No MI or Stroke

Had MI same year

Had MI 1 year ago

…

Had MI n years ago

Had stroke same year

Had stroke 1 year ago

…

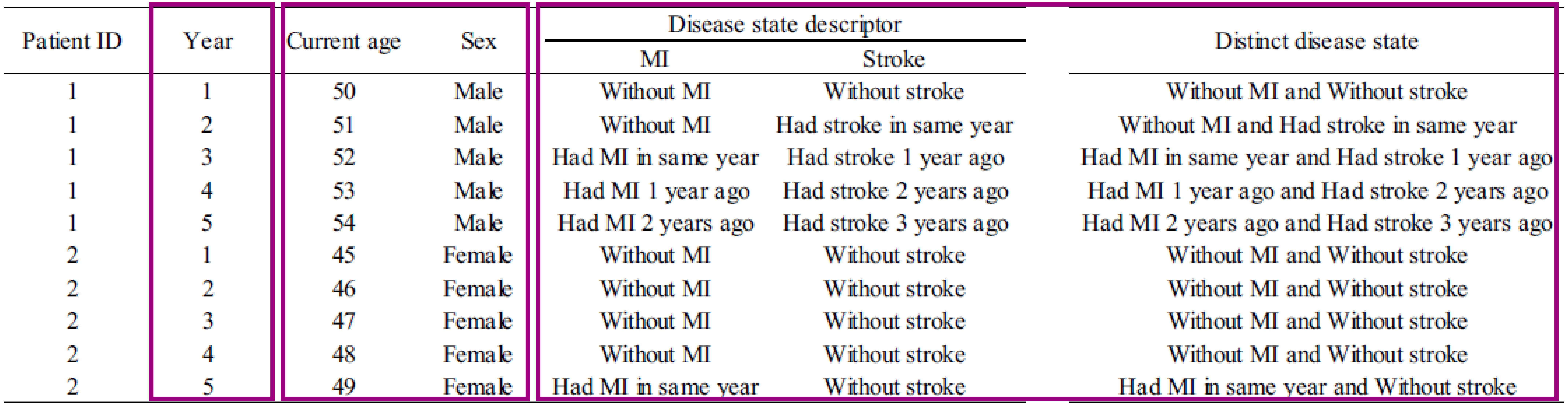
Had stroke n years ago

Background

Costs associated with disease model state

Specifically, we estimate

- costs over a fixed period
- associated with disease model state
- for an individual with specific characteristics



Patient ID	Year	Current age	Sex	Disease state descriptor		Distinct disease state
				MI	Stroke	
1	1	50	Male	Without MI	Without stroke	Without MI and Without stroke
1	2	51	Male	Without MI	Had stroke in same year	Without MI and Had stroke in same year
1	3	52	Male	Had MI in same year	Had stroke 1 year ago	Had MI in same year and Had stroke 1 year ago
1	4	53	Male	Had MI 1 year ago	Had stroke 2 years ago	Had MI 1 year ago and Had stroke 2 years ago
1	5	54	Male	Had MI 2 years ago	Had stroke 3 years ago	Had MI 2 years ago and Had stroke 3 years ago
2	1	45	Female	Without MI	Without stroke	Without MI and Without stroke
2	2	46	Female	Without MI	Without stroke	Without MI and Without stroke
2	3	47	Female	Without MI	Without stroke	Without MI and Without stroke
2	4	48	Female	Without MI	Without stroke	Without MI and Without stroke
2	5	49	Female	Had MI in same year	Without stroke	Had MI in same year and Without stroke

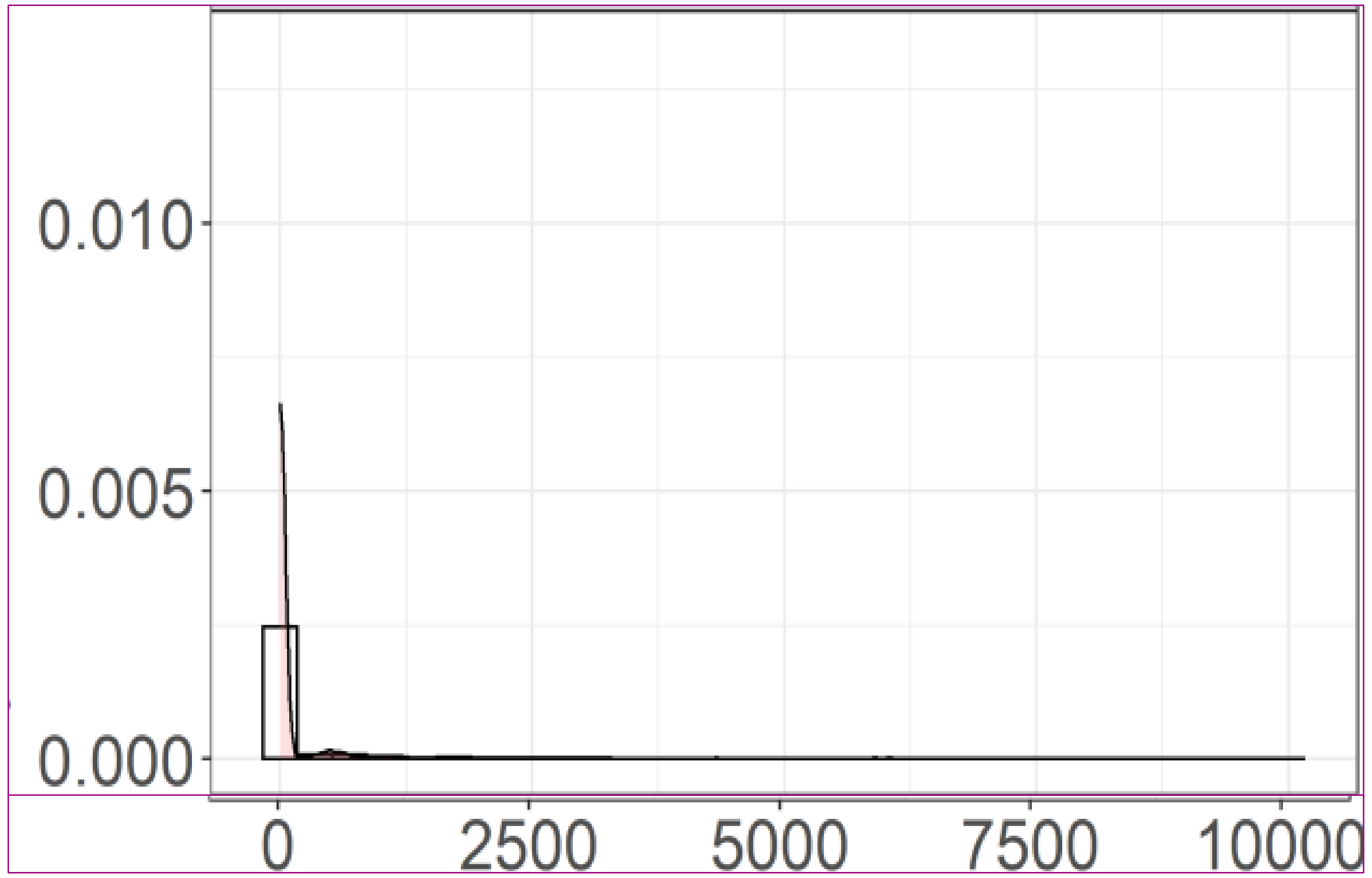
MI, myocardial infarction

Background

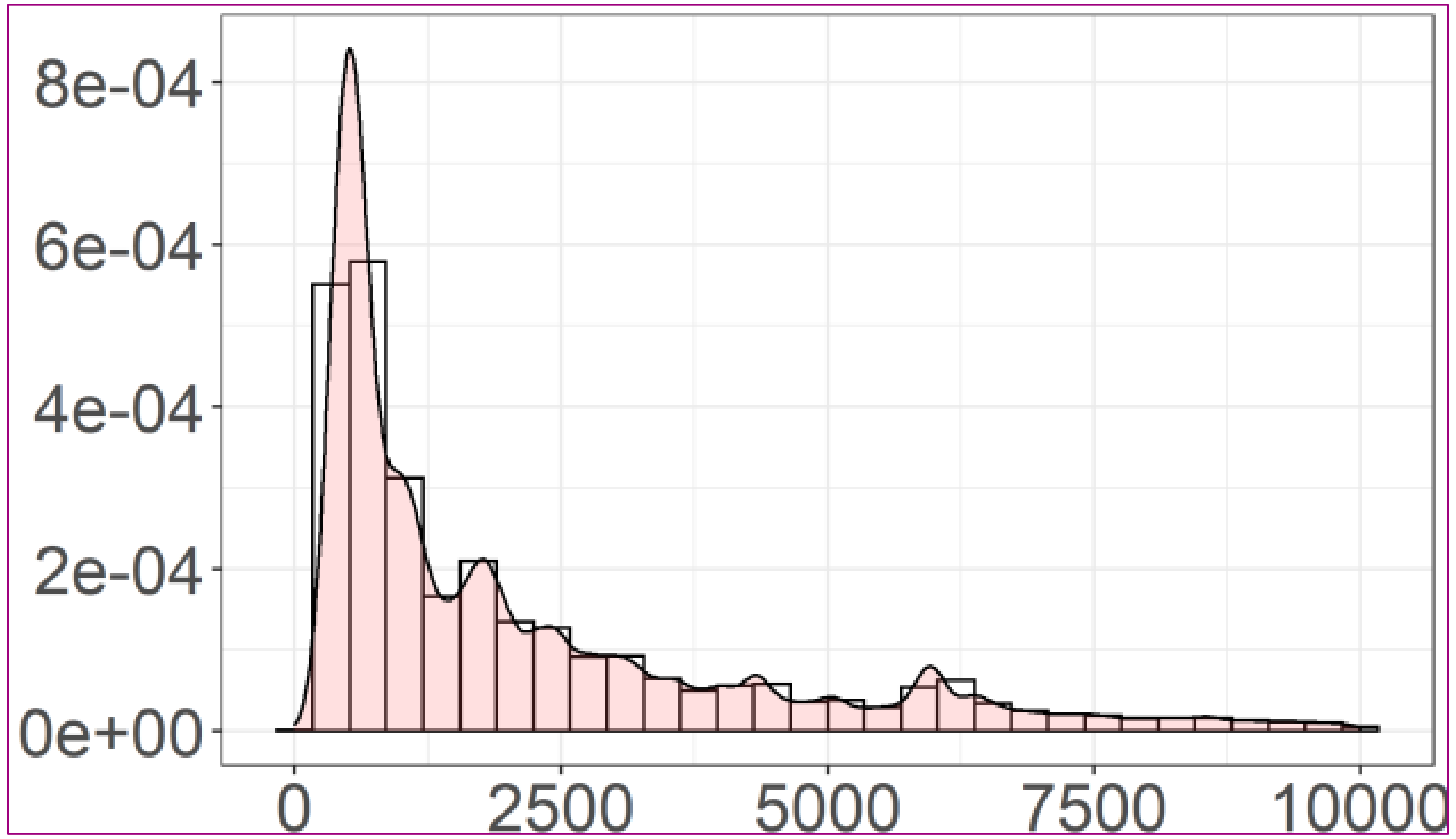
Costs associated with disease model state

Features of cost data

A large number of zero cost (e.g. hospital costs)



Heavy right-hand tailed distribution



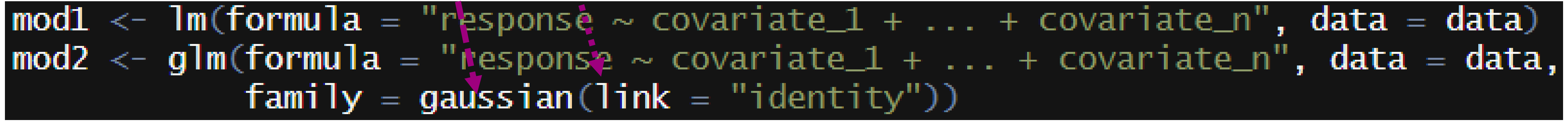
Background
Estimating costs using Generalized Linear Model (GLM)

General recommendation on estimating costs

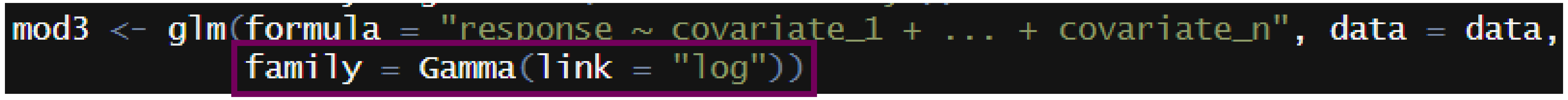
- simple methods when having large datasets
- address a small number of key data issue with smaller datasets

Features of GLM framework

- Address linearity issue between linear predictor and dependent variable
  - by fitting a **link function** between the linear predictor and outcome
- Accommodate the skewness in the distribution of the residual error
  - by fitting a **variance function**



```
mod1 <- lm(formula = "response ~ covariate_1 + ... + covariate_n", data = data)
mod2 <- glm(formula = "response ~ covariate_1 + ... + covariate_n", data = data,
            family = gaussian(link = "identity"))
```



```
mod3 <- glm(formula = "response ~ covariate_1 + ... + covariate_n", data = data,
            family = Gamma(link = "log"))
```

Key Steps of Statistical Modelling of Costs Associated with Disease Model State

1. Preparing the dataset for estimating costs of disease states

- Raw dataset generation
- Handling censored and missing data
- Covariate specification

2. Candidate statistical models

- Common candidate statistical models
- Initial set of covariates
- Tests to choose statistical model specification

3. Selecting the final model

- Covariate selection
- Final model selection
- Consideration of interactions

4. Use of the cost model

- Cost prediction given individual’s characteristics
- Effect of a disease state on costs

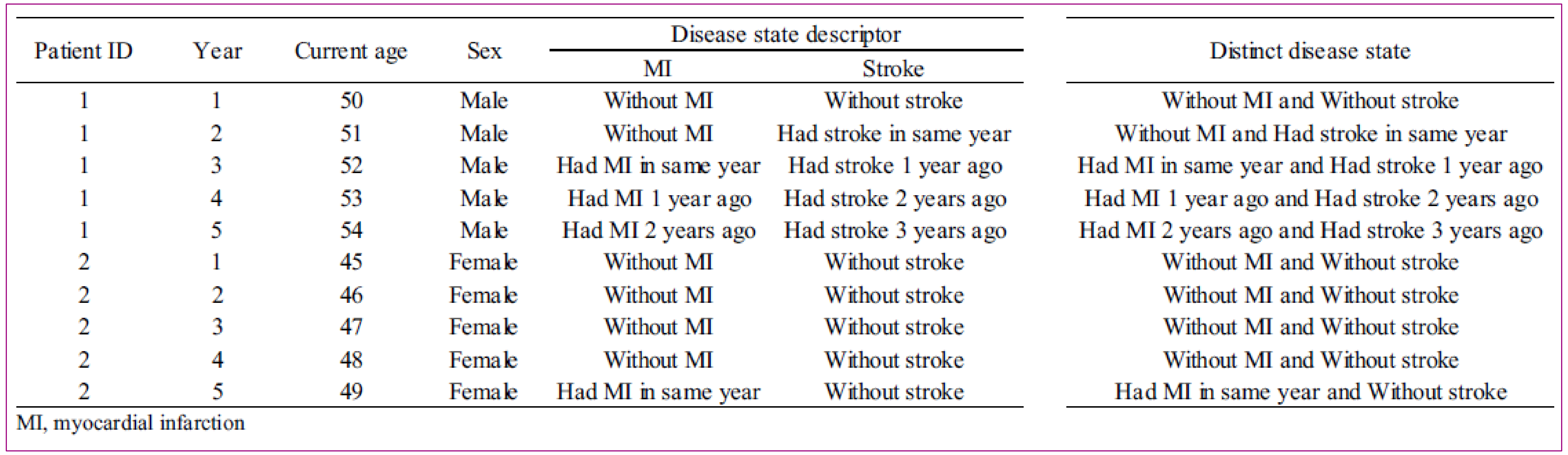
Illustrative Example - Modelling Hospital Costs Associated with Cardiovascular Events Using R

Research question

- costs over a fixed period: **annual hospital care costs**
- associated with disease model state: **associated with cardiovascular and mortality event**
  - **Myocardial infarction (MI): none, year of event, 1, 2, ≥3 years after event**
  - **Stroke: none, year of event, 1, 2, ≥3 years after event**
  - **Vascular death (VD): none, year of event**
  - **Non-vascular death (NVD): none, year of event**
- for an individual with specific characteristics: **a wide range of people without previous CVD**

Data used for the illustration

- A synthetic analytical dataset
  - 10,000 participants each with
  - 10 annual periods with columns
    - Response (costs in the year)
    - Covariates (state, characteristics)



Patient ID	Year	Current age	Sex	Disease state descriptor		Distinct disease state
				MI	Stroke	
1	1	50	Male	Without MI	Without stroke	Without MI and Without stroke
1	2	51	Male	Without MI	Had stroke in same year	Without MI and Had stroke in same year
1	3	52	Male	Had MI in same year	Had stroke 1 year ago	Had MI in same year and Had stroke 1 year ago
1	4	53	Male	Had MI 1 year ago	Had stroke 2 years ago	Had MI 1 year ago and Had stroke 2 years ago
1	5	54	Male	Had MI 2 years ago	Had stroke 3 years ago	Had MI 2 years ago and Had stroke 3 years ago
2	1	45	Female	Without MI	Without stroke	Without MI and Without stroke
2	2	46	Female	Without MI	Without stroke	Without MI and Without stroke
2	3	47	Female	Without MI	Without stroke	Without MI and Without stroke
2	4	48	Female	Without MI	Without stroke	Without MI and Without stroke
2	5	49	Female	Had MI in same year	Without stroke	Had MI in same year and Without stroke

MI, myocardial infarction

Illustrative Example

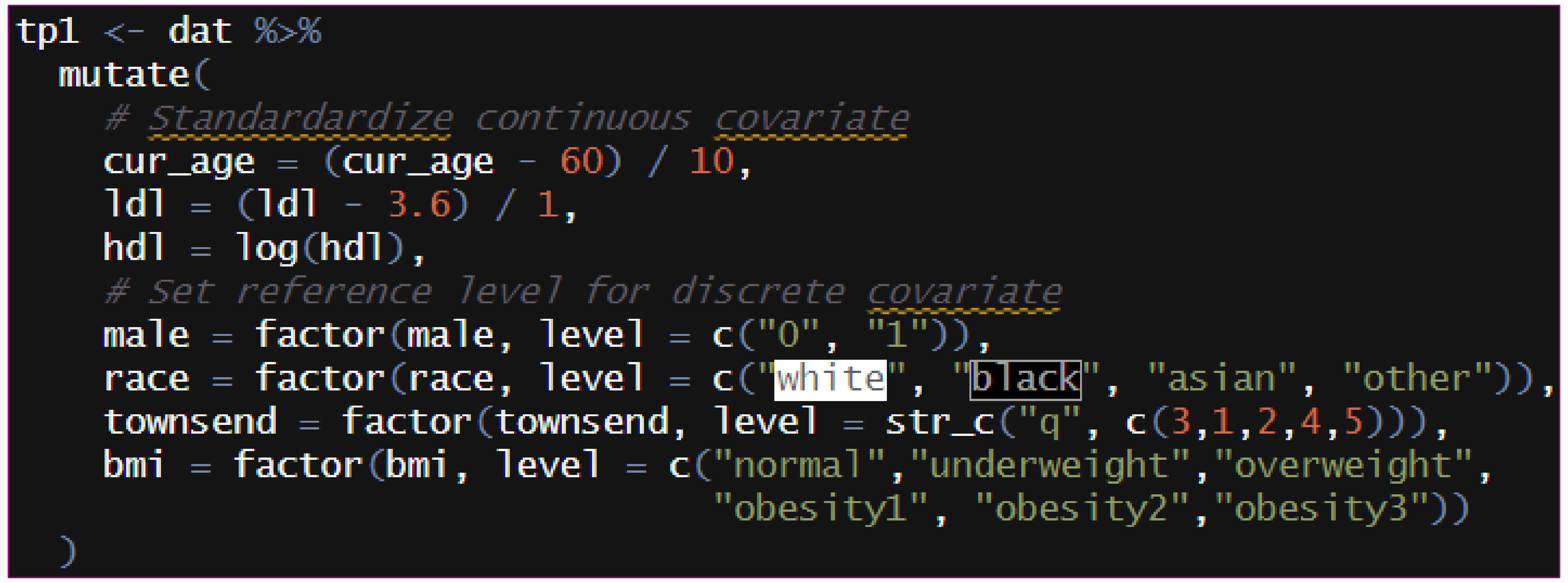
Step 1. Preparation of dataset – Covariate specification

Specify covariates to improve performance

- Continuous covariates: functional form
  - Known relationship: Z, $\ln(Z)$, $\sqrt{Z}$
  - Complex: spline, polynomial, categorization
- Discrete covariates: category combination

Specify covariates to facilitate interpretation

- **Continuous covariate: standardize**
- **Discrete covariate: set reference level**



```
tp1 <- dat %>%
  mutate(
    # Standardardize continuous covariate
    cur_age = (cur_age - 60) / 10,
    ldl = (ldl - 3.6) / 1,
    hdl = log(hdl),
    # Set reference level for discrete covariate
    male = factor(male, level = c("0", "1")),
    race = factor(race, level = c("white", "black", "asian", "other")),
    townsend = factor(townsend, level = str_c("q", c(3,1,2,4,5))),
    bmi = factor(bmi, level = c("normal","underweight","overweight",
                                "obesity1", "obesity2","obesity3"))
  )
```

Covariates	Raw dataset		Analytical dataset	
	Specification	Values	New specification	Values
Age (years)	Z	56 (8)	**(Z – 60)/10**	**-0.4 (0.8)**
LDL-C (mmol/L)	Z	3.6 (0.8)	**(Z - 3.6) / 1**	**0 (0.8)**
HDL-C (mmol/L)	Z	1.5 (0.4)	**Ln(Z)**	**0.4 (0.3)**
Sex	Female	5635 (56.4)	**Female (Ref)**	-
	Male	4365 (43.6)	Male	-
Ethnicity	White	9464 (94.6)	**White (Ref)**	-
	Black	179 (1.8)	Black	-
	South Asian	165 (1.7)	South Asian	-
	Others	192 (1.9)	Others	-
Townsend deprivation score, categorized into quintiles (Quintile 1: least deprived)	Quintile 1	3712 (37.1)	Quintile 1	-
	Quintile 2	1947 (19.5)	Quintile 2	-
	Quintile 3	1701 (17)	**Quintile 3 (Ref)**	-
	Quintile 4	1449 (14.5)	Quintile 4	-
	Quintile 5	1191 (11.9)	Quintile 5	-
Body mass index (BMI, kg/m^2), categorized	<18.5	53 (0.5)	<18.5	-
	≥18.5, <25	3295 (33)	**≥18.5, <25 (Ref)**	-
	≥25, <30	4334 (43.3)	≥25, <30	-
	≥30, <35	1682 (16.8)	≥30, <35	-
	≥35, <40	449 (4.5)	≥35, <40	-
	≥40	187 (1.9)	≥40	-

Illustrative Example

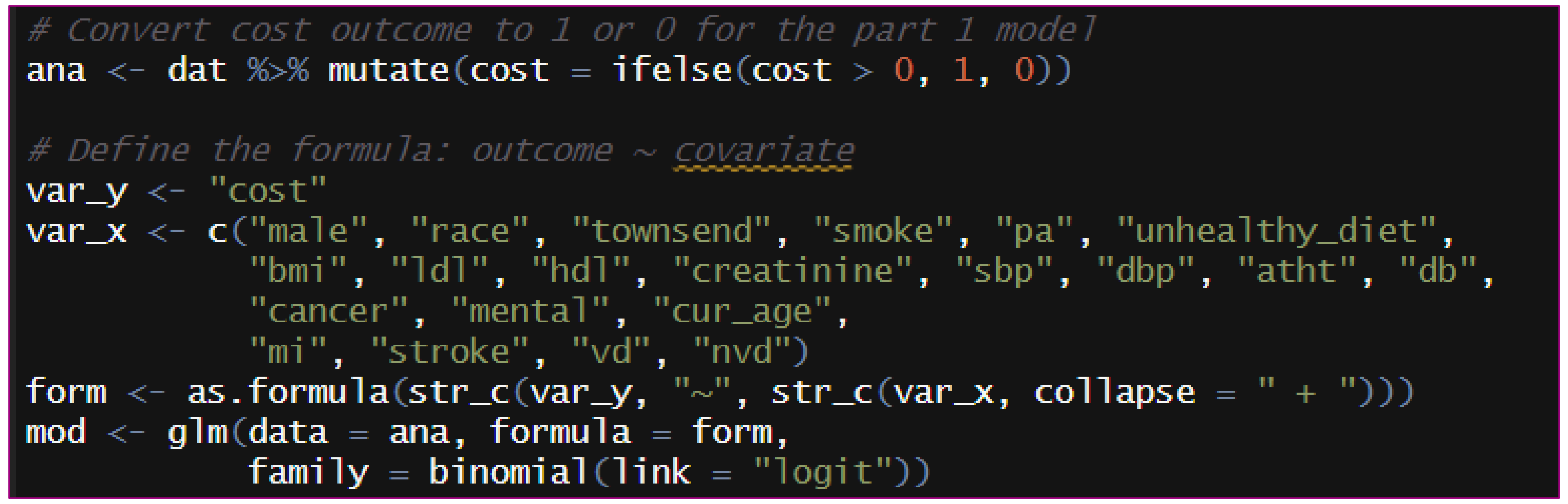
Step 2. Candidate statistical models

Six two-part models

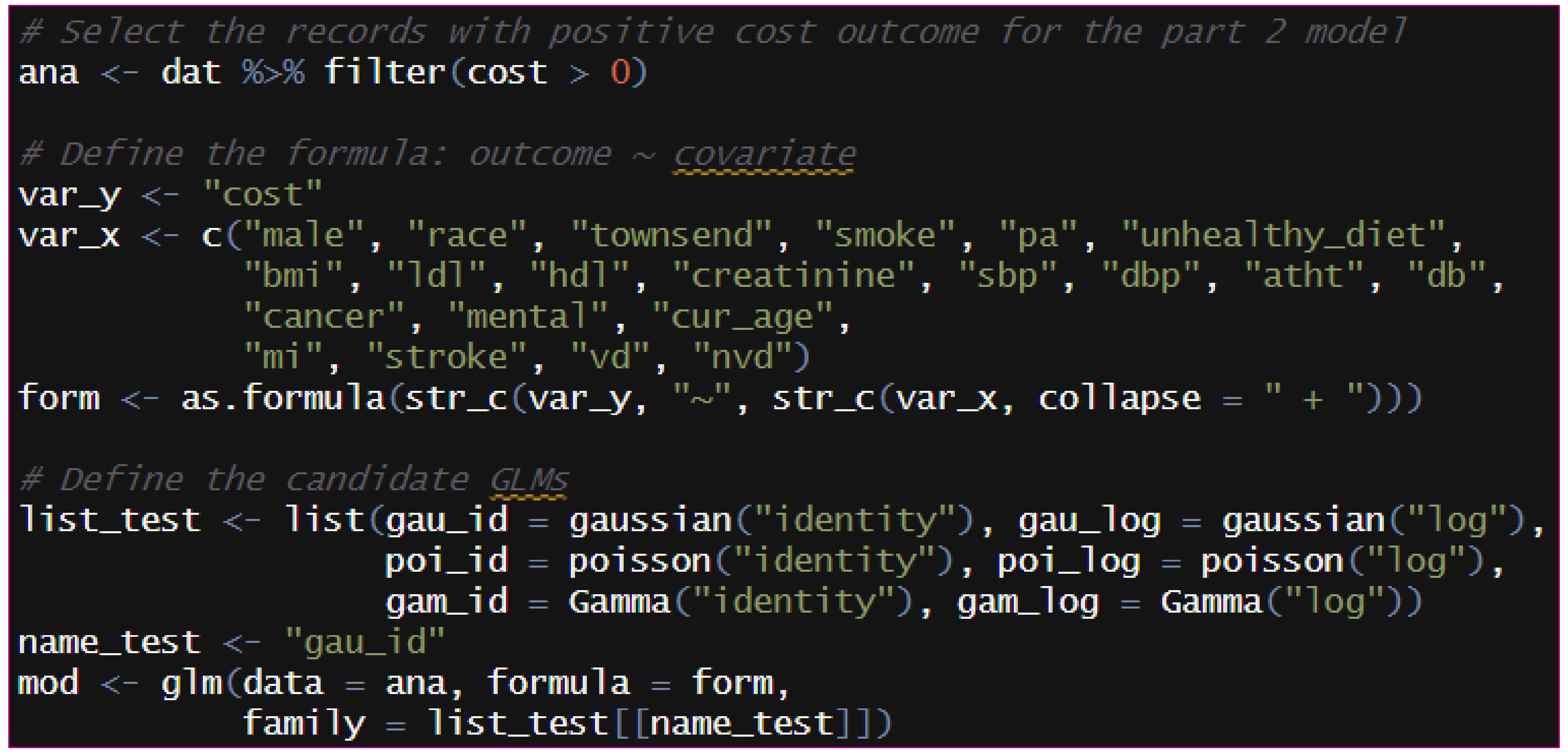
- First-part modelling the probability of incurring any costs
  - Same for all: Logistic regression
- Second-part modelling the costs conditional on any costs incurring
  - Six common GLMs
  - Gaussian-Identity, Gaussian-Log, Poisson-Identity, Poisson-Log, Gamma-Identity, Gamma-Log

One one-part model

- Gaussian-Identity GLM



```r
# Convert cost outcome to 1 or 0 for the part 1 model
ana <- dat %>% mutate(cost = ifelse(cost > 0, 1, 0))

# Define the formula: outcome ~ covariate
var_y <- "cost"
var_x <- c("male", "race", "townsend", "smoke", "pa", "unhealthy_diet",
           "bmi", "ldl", "hdl", "creatinine", "sbp", "dbp", "atht", "db",
           "cancer", "mental", "cur_age",
           "mi", "stroke", "vd", "nvd")
form <- as.formula(str_c(var_y, "~", str_c(var_x, collapse = " + ")))
mod <- glm(data = ana, formula = form,
           family = binomial(link = "logit"))
```



```r
# Select the records with positive cost outcome for the part 2 model
ana <- dat %>% filter(cost > 0)

# Define the formula: outcome ~ covariate
var_y <- "cost"
var_x <- c("male", "race", "townsend", "smoke", "pa", "unhealthy_diet",
           "bmi", "ldl", "hdl", "creatinine", "sbp", "dbp", "atht", "db",
           "cancer", "mental", "cur_age",
           "mi", "stroke", "vd", "nvd")
form <- as.formula(str_c(var_y, "~", str_c(var_x, collapse = " + ")))

# Define the candidate GLMs
list_test <- list(gau_id = gaussian("identity"), gau_log = gaussian("log"),
                  poi_id = poisson("identity"), poi_log = poisson("log"),
                  gam_id = Gamma("identity"), gam_log = Gamma("log"))
name_test <- "gau_id"
mod <- glm(data = ana, formula = form,
           family = list_test[[name_test]])
```

Illustrative Example

Step 2. Candidate statistical models – Specification test

Specification tests to choose promising candidate models

- Tests for link function
  - **Hosmer-Lemeshow test**: P<0.05 improper link
  - **Pregibon's test**: P<0.05 improper link
- Test for variance function
  - **Modified Park's tests**
    - Slope indicate proper distribution
    - 0: Gaussian; 1: Poisson; 2: Gamma; 3: Inverse Gaussian

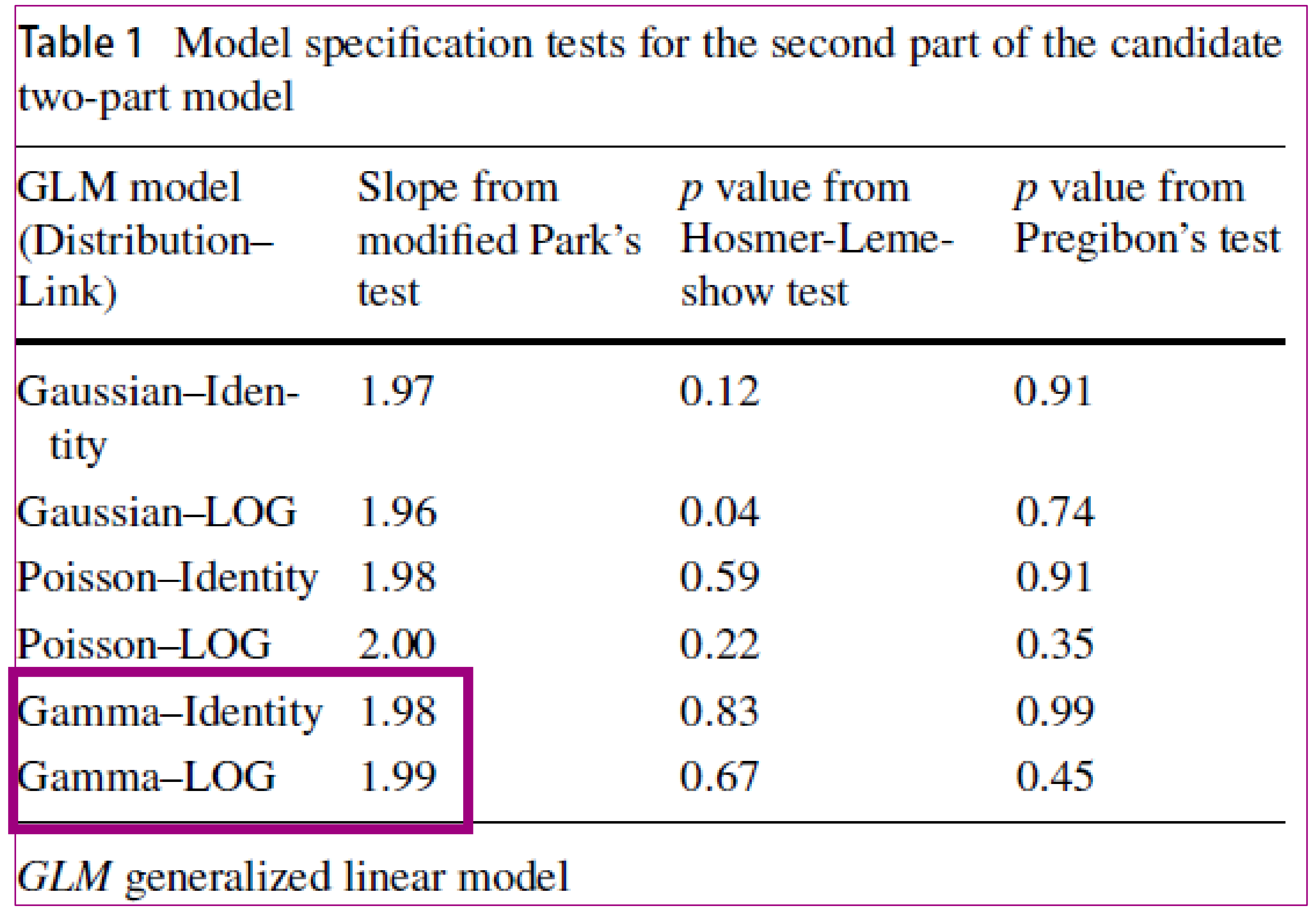
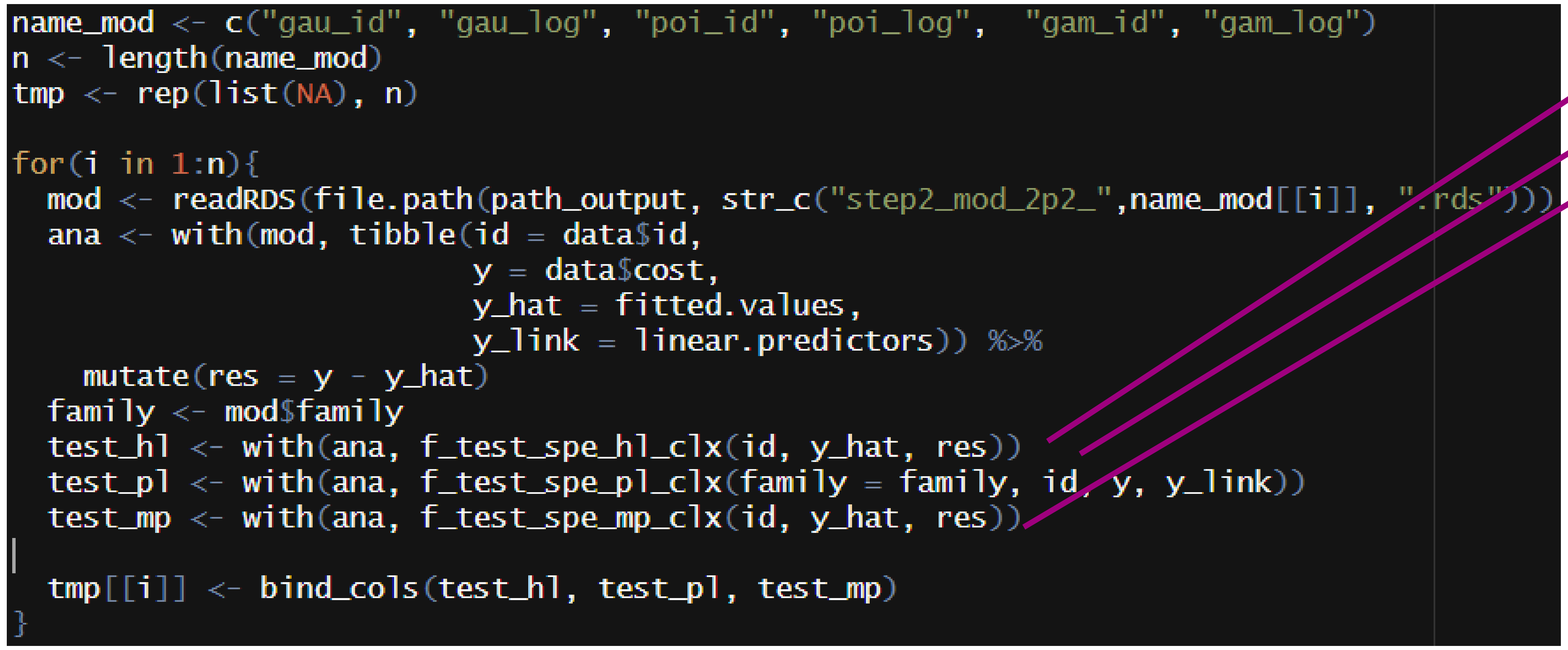


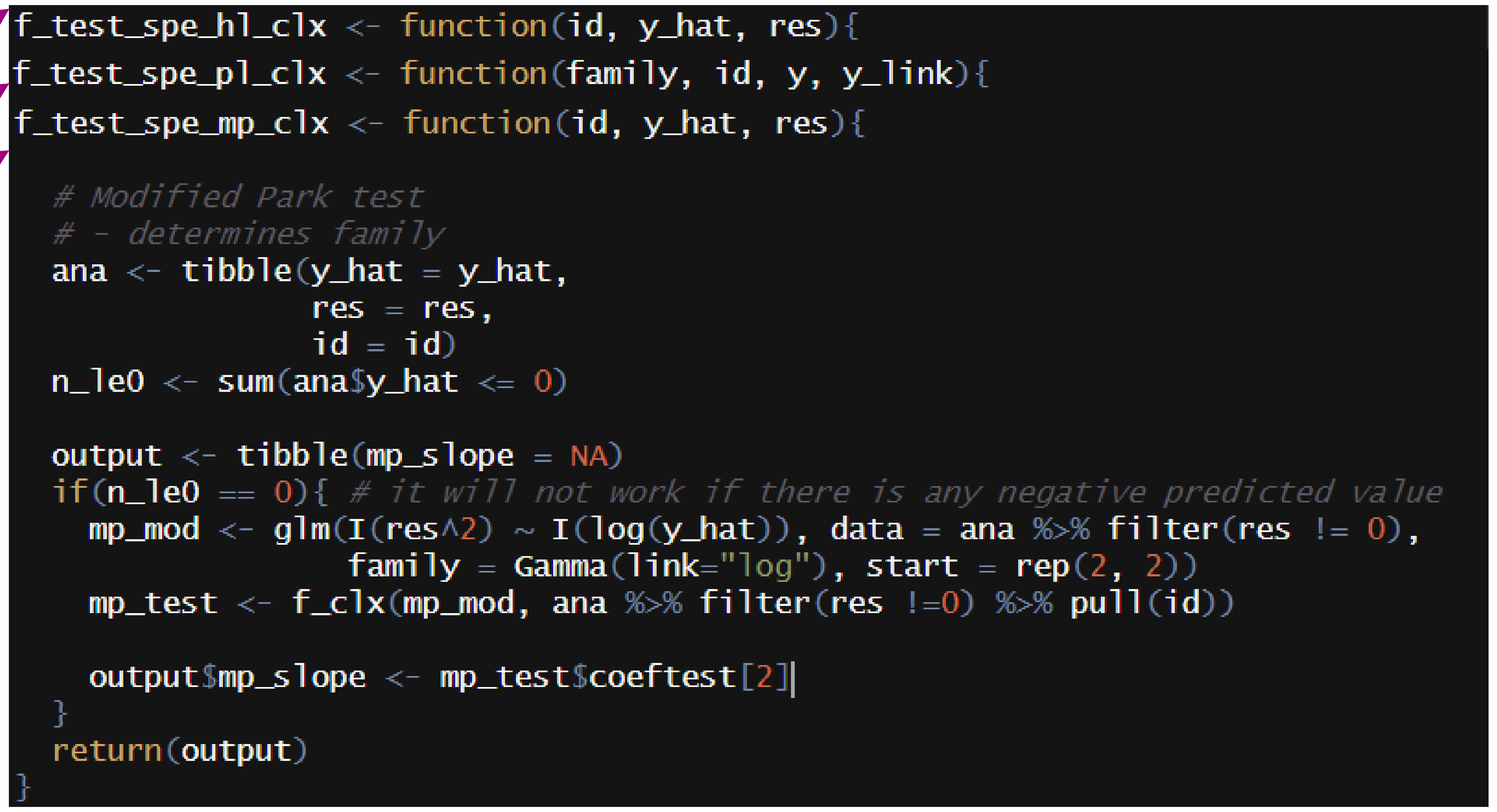
Table 1 Model specification tests for the second part of the candidate two-part model

GLM model (Distribution–Link)	Slope from modified Park's test	*p* value from Hosmer-Lemeshow test	*p* value from Pregibon's test
Gaussian–Identity	1.97	0.12	0.91
Gaussian–LOG	1.96	0.04	0.74
Poisson–Identity	1.98	0.59	0.91
Poisson–LOG	2.00	0.22	0.35
Gamma–Identity	1.98	0.83	0.99
Gamma–LOG	1.99	0.67	0.45

GLM generalized linear model



```
name_mod <- c("gau_id", "gau_log", "poi_id", "poi_log",  "gam_id", "gam_log")
n <- length(name_mod)
tmp <- rep(list(NA), n)

for(i in 1:n){
  mod <- readRDS(file.path(path_output, str_c("step2_mod_2p2_",name_mod[[i]], ".rds")))
  ana <- with(mod, tibble(id = data$id,
                          y = data$cost,
                          y_hat = fitted.values,
                          y_link = linear.predictors)) %>%
    mutate(res = y - y_hat)
  family <- mod$family
  test_hl <- with(ana, f_test_spe_hl_clx(id, y_hat, res))
  test_pl <- with(ana, f_test_spe_pl_clx(family = family, id, y, y_link))
  test_mp <- with(ana, f_test_spe_mp_clx(id, y_hat, res))

  tmp[[i]] <- bind_cols(test_hl, test_pl, test_mp)
}
```

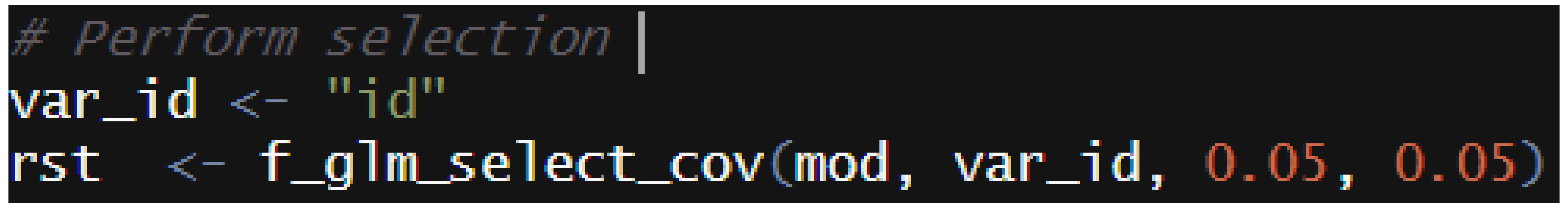


```
f_test_spe_hl_clx <- function(id, y_hat, res){
f_test_spe_pl_clx <- function(family, id, y, y_link){
f_test_spe_mp_clx <- function(id, y_hat, res){

  # Modified Park test
  # - determines family
  ana <- tibble(y_hat = y_hat,
                res = res,
                id = id)
  n_le0 <- sum(ana$y_hat <= 0)

  output <- tibble(mp_slope = NA)
  if(n_le0 == 0){ # it will not work if there is any negative predicted value
    mp_mod <- glm(I(res^2) ~ I(log(y_hat)), data = ana %>% filter(res != 0),
                  family = Gamma(link="log"), start = rep(2, 2))
    mp_test <- f_clx(mp_mod, ana %>% filter(res !=0) %>% pull(id))

    output$mp_slope <- mp_test$coeftest[2]
  }
  return(output)
}
```

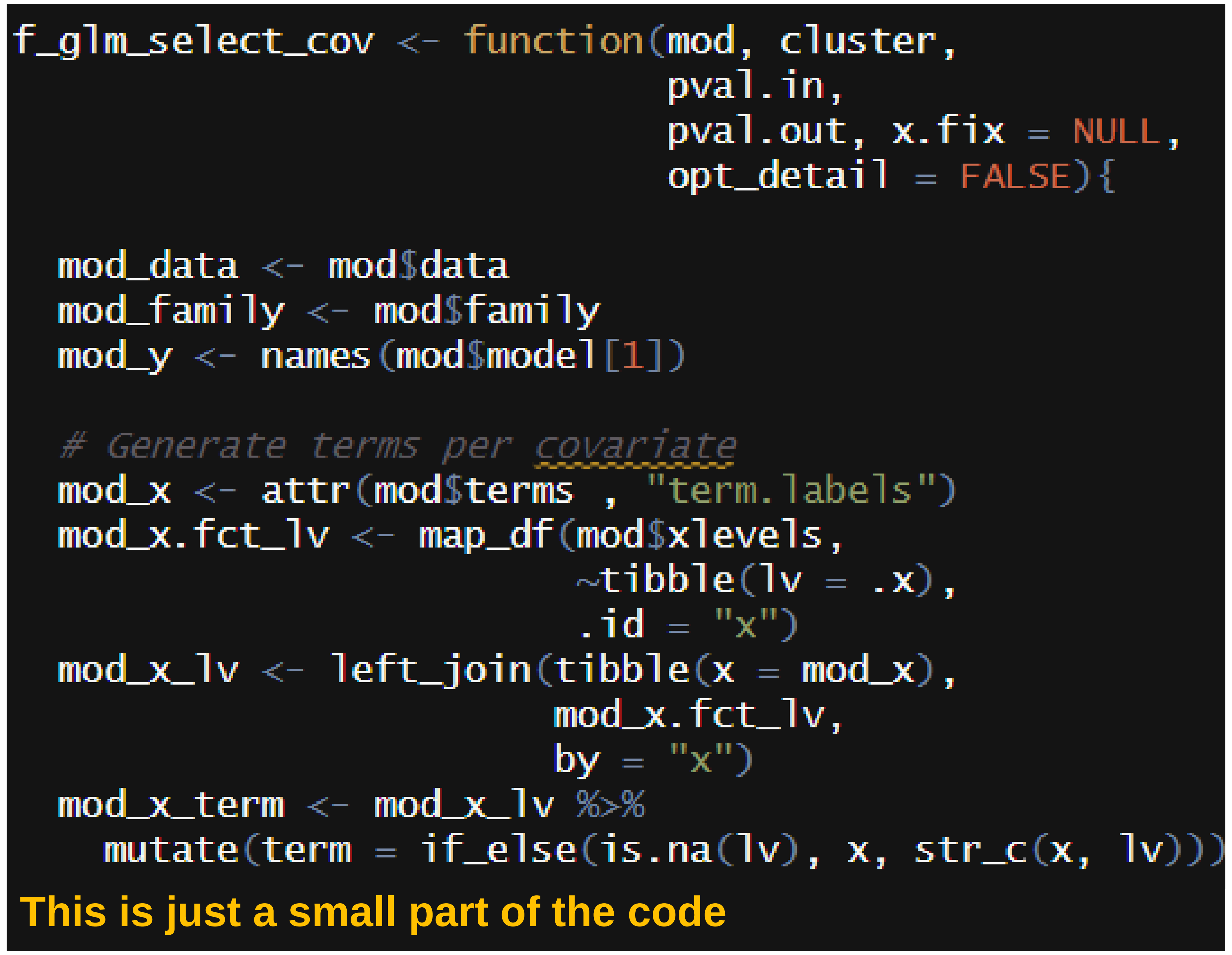
Illustrative Example

Step 3. Model selection – Covariate selection

Select the covariates reliably predicting response (example using stepwise backward selection)



```
# Perform selection
var_id <- "id"
rst  <- f_glm_select_cov(mod, var_id, 0.05, 0.05)
```



```
f_glm_select_cov <- function(mod, cluster,
                             pval.in,
                             pval.out, x.fix = NULL,
                             opt_detail = FALSE){

  mod_data <- mod$data
  mod_family <- mod$family
  mod_y <- names(mod$model[1])

  # Generate terms per covariate
  mod_x <- attr(mod$terms , "term.labels")
  mod_x.fct_lv <- map_df(mod$xlevels,
                         ~tibble(lv = .x),
                         .id = "x")
  mod_x_lv <- left_join(tibble(x = mod_x),
                        mod_x.fct_lv,
                        by = "x")
  mod_x_term <- mod_x_lv %>%
    mutate(term = if_else(is.na(lv), x, str_c(x, lv)))
```

This is just a small part of the code

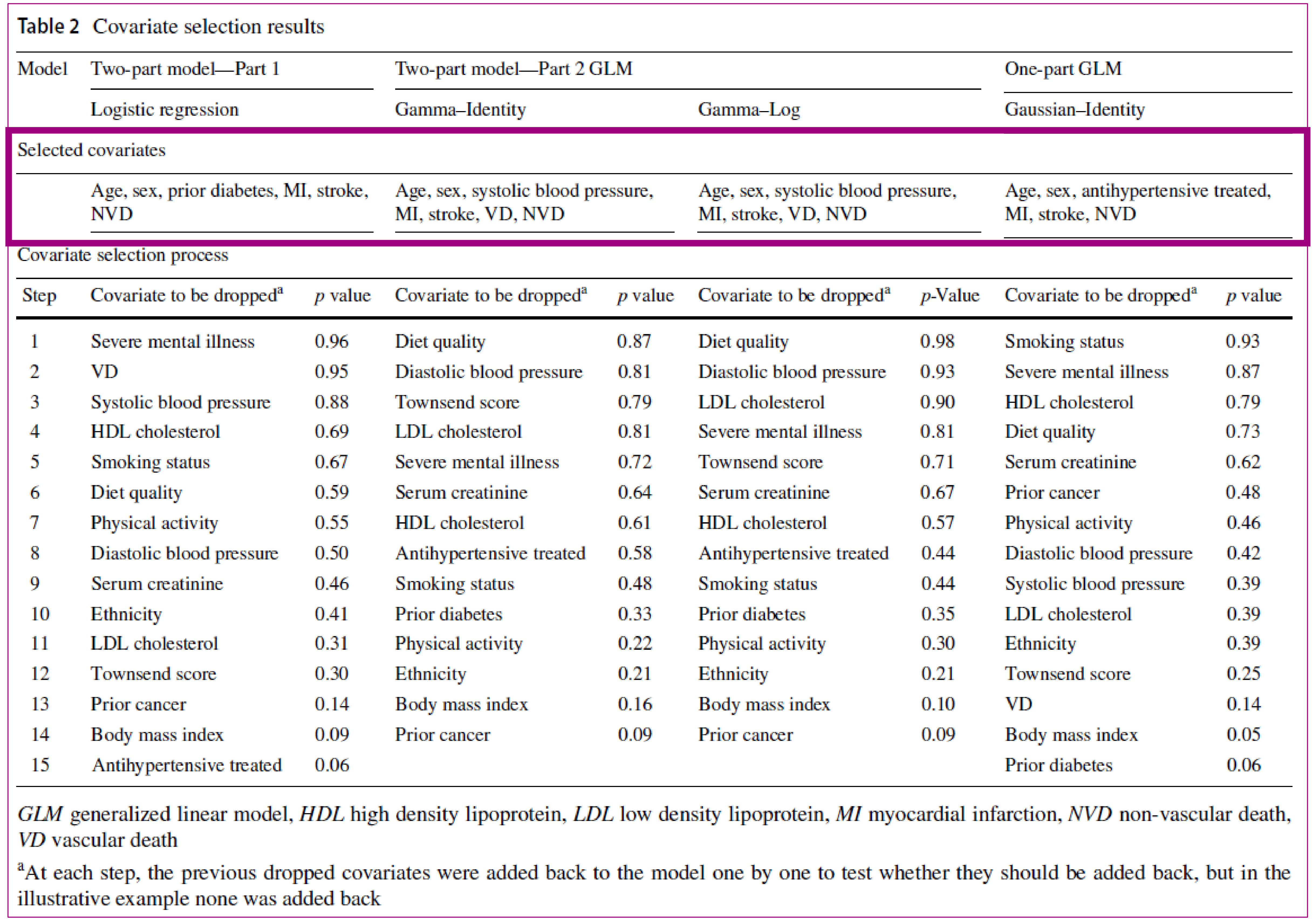


Table 2 Covariate selection results

Model	Two-part model—Part 1		Two-part model—Part 2 GLM				One-part GLM	
	Logistic regression		Gamma–Identity		Gamma–Log		Gaussian–Identity	
Selected covariates								
	Age, sex, prior diabetes, MI, stroke, NVD		Age, sex, systolic blood pressure, MI, stroke, VD, NVD		Age, sex, systolic blood pressure, MI, stroke, VD, NVD		Age, sex, antihypertensive treated, MI, stroke, NVD	
Covariate selection process								
Step	Covariate to be dropped[a]	*p* value	Covariate to be dropped[a]	*p* value	Covariate to be dropped[a]	*p*-Value	Covariate to be dropped[a]	*p* value
1	Severe mental illness	0.96	Diet quality	0.87	Diet quality	0.98	Smoking status	0.93
2	VD	0.95	Diastolic blood pressure	0.81	Diastolic blood pressure	0.93	Severe mental illness	0.87
3	Systolic blood pressure	0.88	Townsend score	0.79	LDL cholesterol	0.90	HDL cholesterol	0.79
4	HDL cholesterol	0.69	LDL cholesterol	0.81	Severe mental illness	0.81	Diet quality	0.73
5	Smoking status	0.67	Severe mental illness	0.72	Townsend score	0.71	Serum creatinine	0.62
6	Diet quality	0.59	Serum creatinine	0.64	Serum creatinine	0.67	Prior cancer	0.48
7	Physical activity	0.55	HDL cholesterol	0.61	HDL cholesterol	0.57	Physical activity	0.46
8	Diastolic blood pressure	0.50	Antihypertensive treated	0.58	Antihypertensive treated	0.44	Diastolic blood pressure	0.42
9	Serum creatinine	0.46	Smoking status	0.48	Smoking status	0.44	Systolic blood pressure	0.39
10	Ethnicity	0.41	Prior diabetes	0.33	Prior diabetes	0.35	LDL cholesterol	0.39
11	LDL cholesterol	0.31	Physical activity	0.22	Physical activity	0.30	Ethnicity	0.39
12	Townsend score	0.30	Ethnicity	0.21	Ethnicity	0.21	Townsend score	0.25
13	Prior cancer	0.14	Body mass index	0.16	Body mass index	0.10	VD	0.14
14	Body mass index	0.09	Prior cancer	0.09	Prior cancer	0.09	Body mass index	0.05
15	Antihypertensive treated	0.06					Prior diabetes	0.06

GLM generalized linear model, *HDL* high density lipoprotein, *LDL* low density lipoprotein, *MI* myocardial infarction, *NVD* non-vascular death, *VD* vascular death

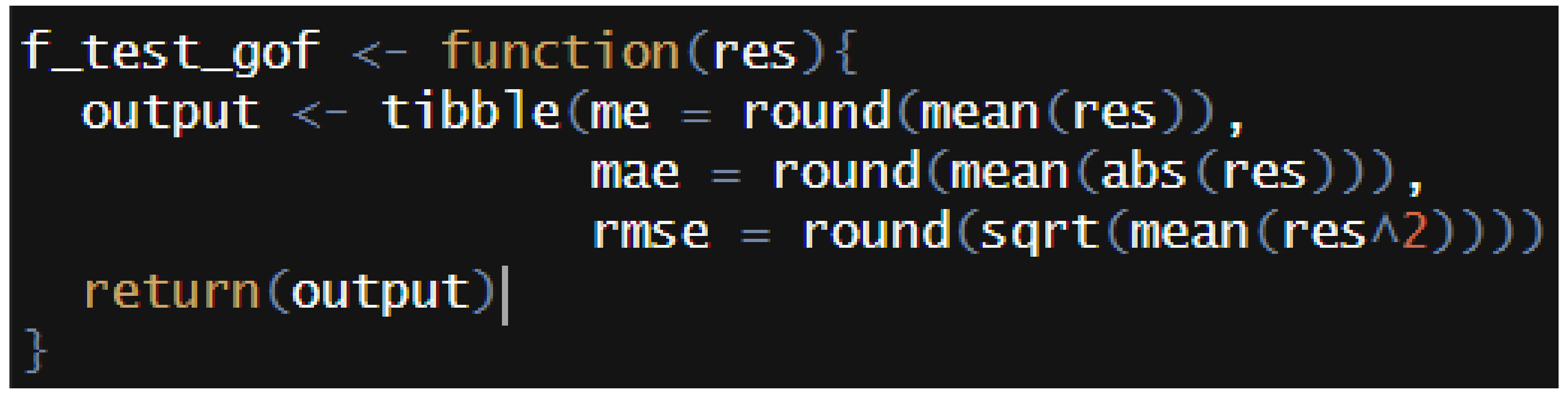
[a] At each step, the previous dropped covariates were added back to the model one by one to test whether they should be added back, but in the illustrative example none was added back

Illustrative Example

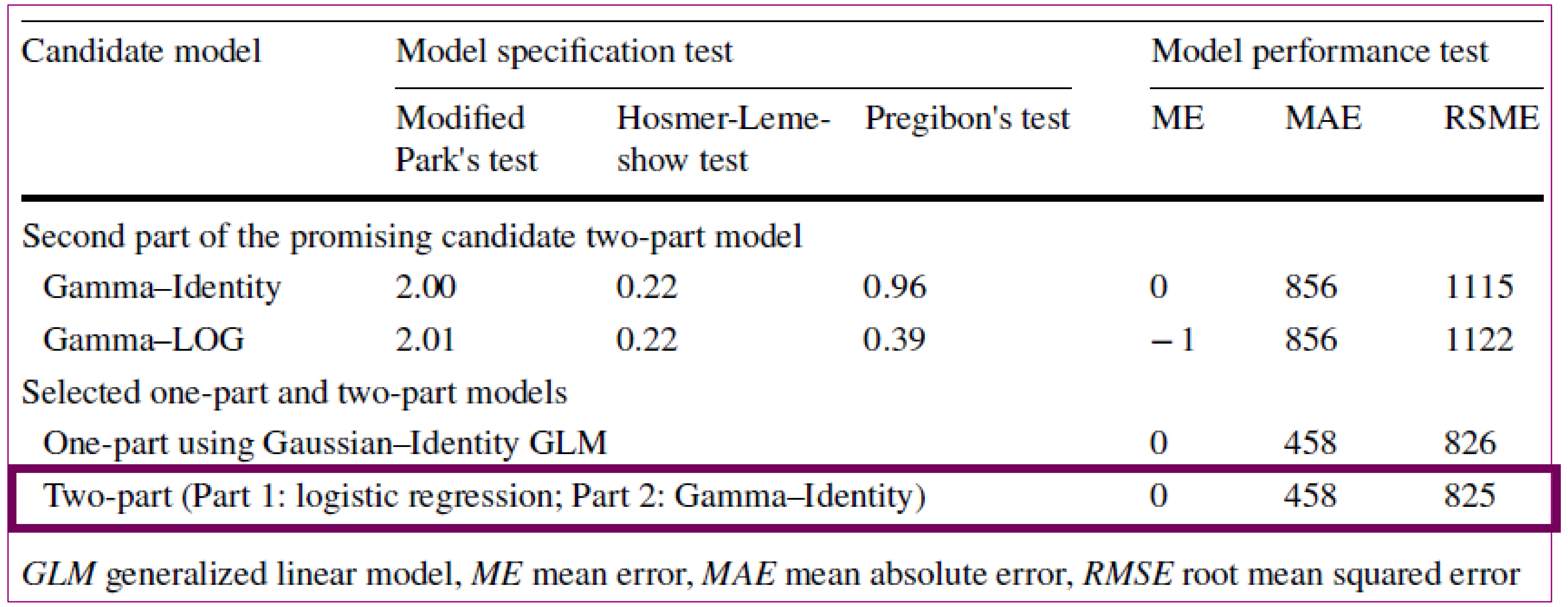
Step 3. Model selection – Performance tests

Model performance tests to check how well the model fit the data

- Mean error
- Mean absolute error
- Root squared mean error

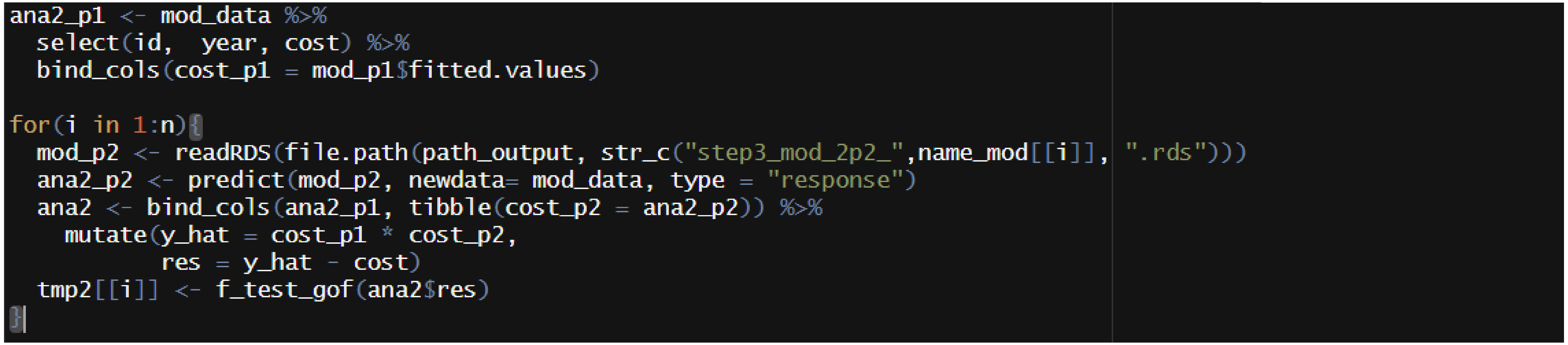


```
f_test_gof <- function(res){
  output <- tibble(me = round(mean(res)),
                   mae = round(mean(abs(res))),
                   rmse = round(sqrt(mean(res^2))))
  return(output)
}
```



Candidate model	Model specification test			Model performance test		
	Modified Park's test	Hosmer-Lemeshow test	Pregibon's test	ME	MAE	RSME
Second part of the promising candidate two-part model						
Gamma–Identity	2.00	0.22	0.96	0	856	1115
Gamma–LOG	2.01	0.22	0.39	– 1	856	1122
Selected one-part and two-part models						
One-part using Gaussian–Identity GLM				0	458	826
Two-part (Part 1: logistic regression; Part 2: Gamma–Identity)				0	458	825

GLM generalized linear model, *ME* mean error, *MAE* mean absolute error, *RMSE* root mean squared error

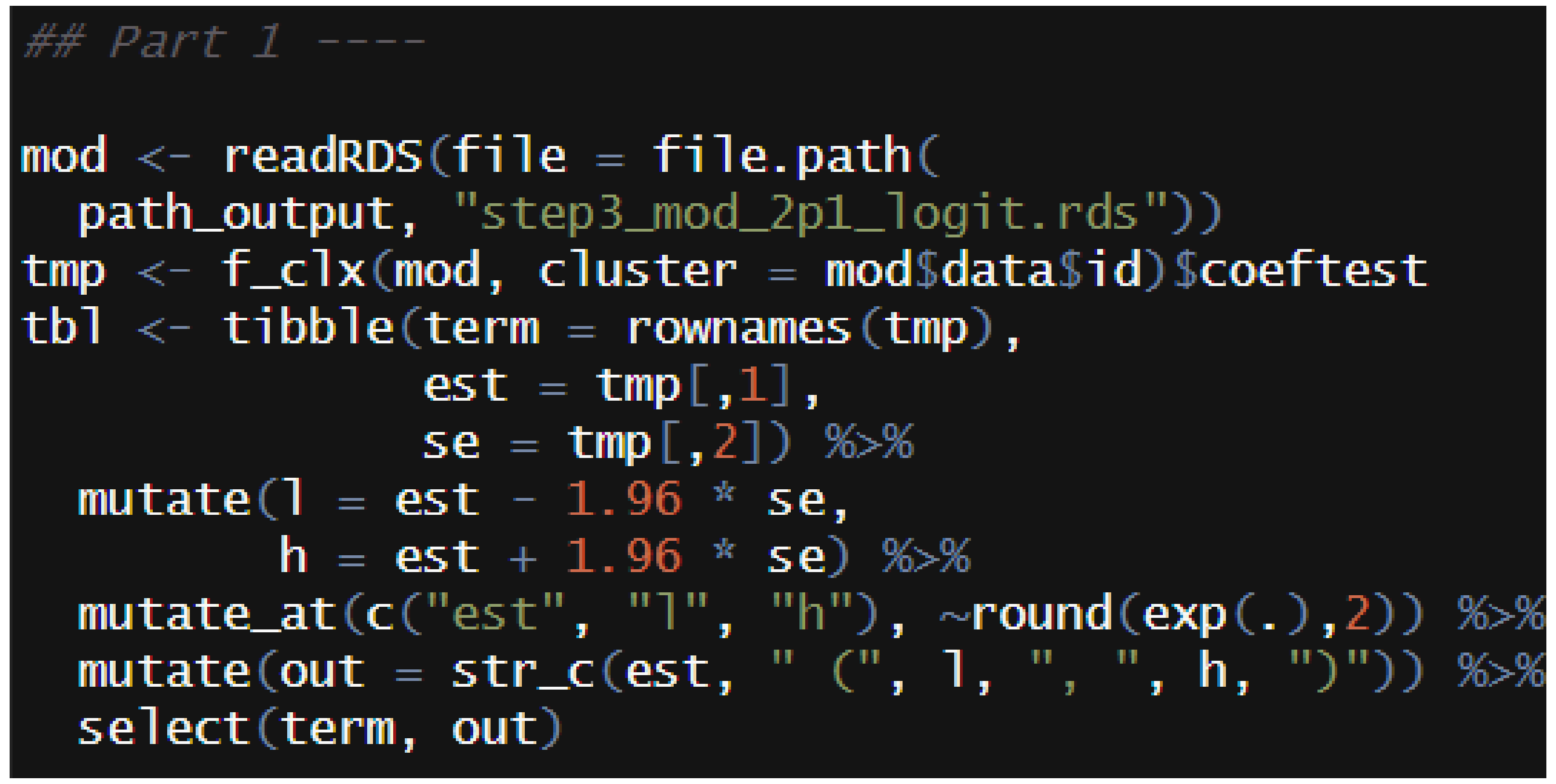


```
ana2_p1 <- mod_data %>%
  select(id,  year, cost) %>%
  bind_cols(cost_p1 = mod_p1$fitted.values)

for(i in 1:n){
  mod_p2 <- readRDS(file.path(path_output, str_c("step3_mod_2p2_",name_mod[[i]], ".rds")))
  ana2_p2 <- predict(mod_p2, newdata= mod_data, type = "response")
  ana2 <- bind_cols(ana2_p1, tibble(cost_p2 = ana2_p2)) %>%
    mutate(y_hat = cost_p1 * cost_p2,
               res = y_hat - cost)
  tmp2[[i]] <- f_test_gof(ana2$res)
}
```

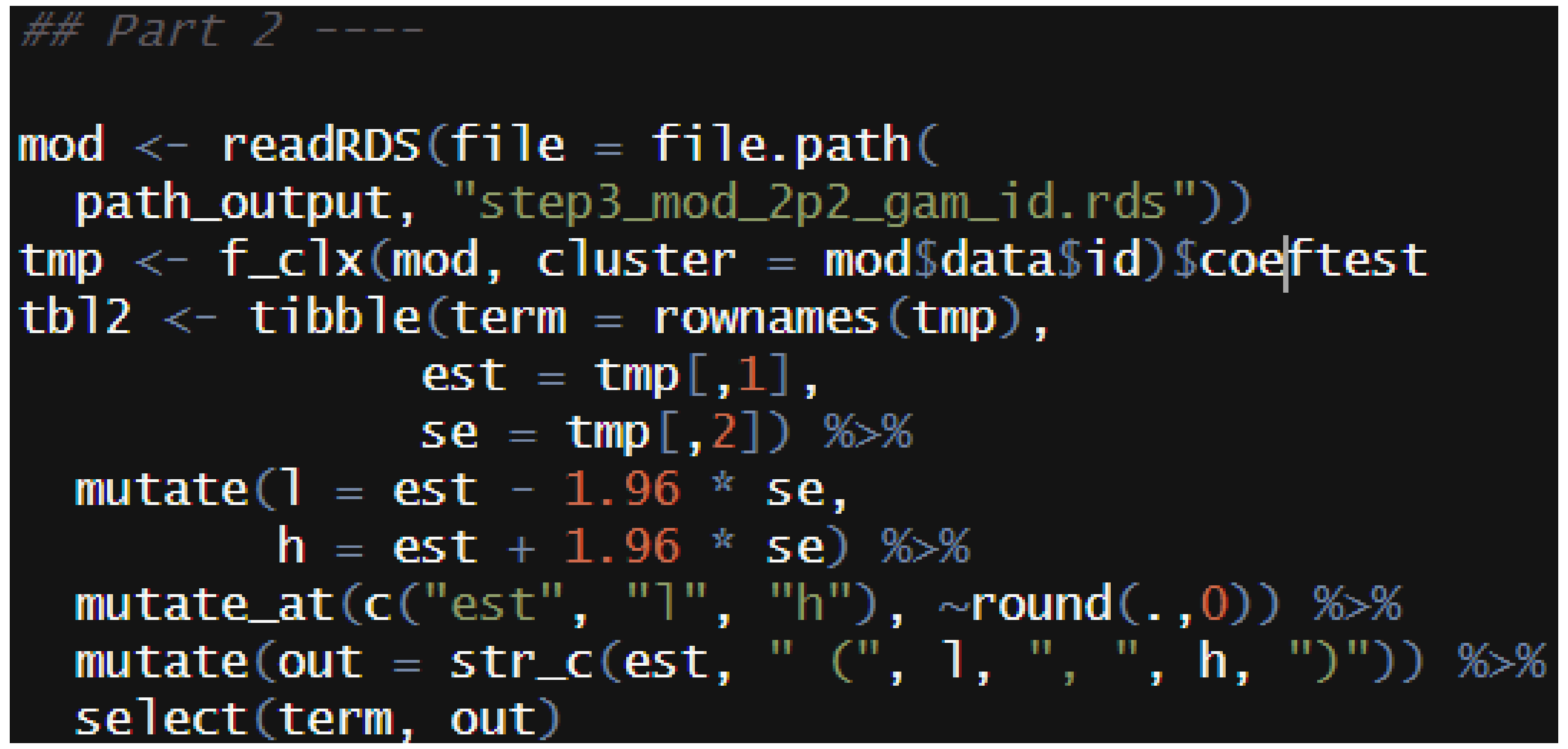
Illustrative Example

Step 3. Model selection – Final selected model

Tabulate model coefficients



```
## Part 1 ----

mod <- readRDS(file = file.path(
  path_output, "step3_mod_2p1_logit.rds"))
tmp <- f_clx(mod, cluster = mod$data$id)$coeftest
tbl <- tibble(term = rownames(tmp),
              est = tmp[,1],
              se = tmp[,2]) %>%
  mutate(l = est - 1.96 * se,
         h = est + 1.96 * se) %>%
  mutate_at(c("est", "l", "h"), ~round(exp(.),2)) %>%
  mutate(out = str_c(est, " (", l, ", ", h, ")")) %>%
  select(term, out)
```



```
## Part 2 ----

mod <- readRDS(file = file.path(
  path_output, "step3_mod_2p2_gam_id.rds"))
tmp <- f_clx(mod, cluster = mod$data$id)$coeftest
tbl2 <- tibble(term = rownames(tmp),
               est = tmp[,1],
               se = tmp[,2]) %>%
  mutate(l = est - 1.96 * se,
         h = est + 1.96 * se) %>%
  mutate_at(c("est", "l", "h"), ~round(.,0)) %>%
  mutate(out = str_c(est, " (", l, ", ", h, ")")) %>%
  select(term, out)
```

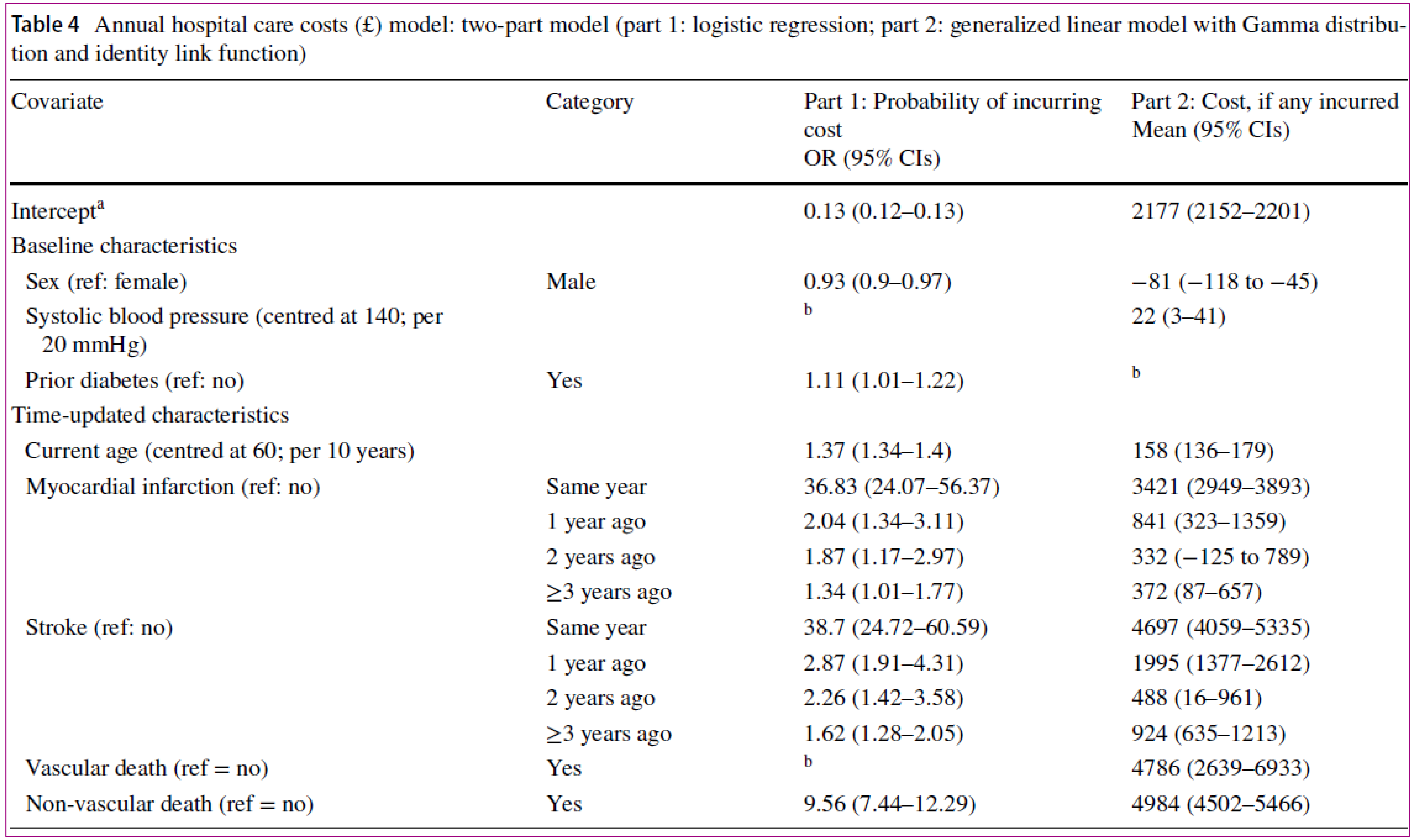


Table 4 Annual hospital care costs (£) model: two-part model (part 1: logistic regression; part 2: generalized linear model with Gamma distribution and identity link function)

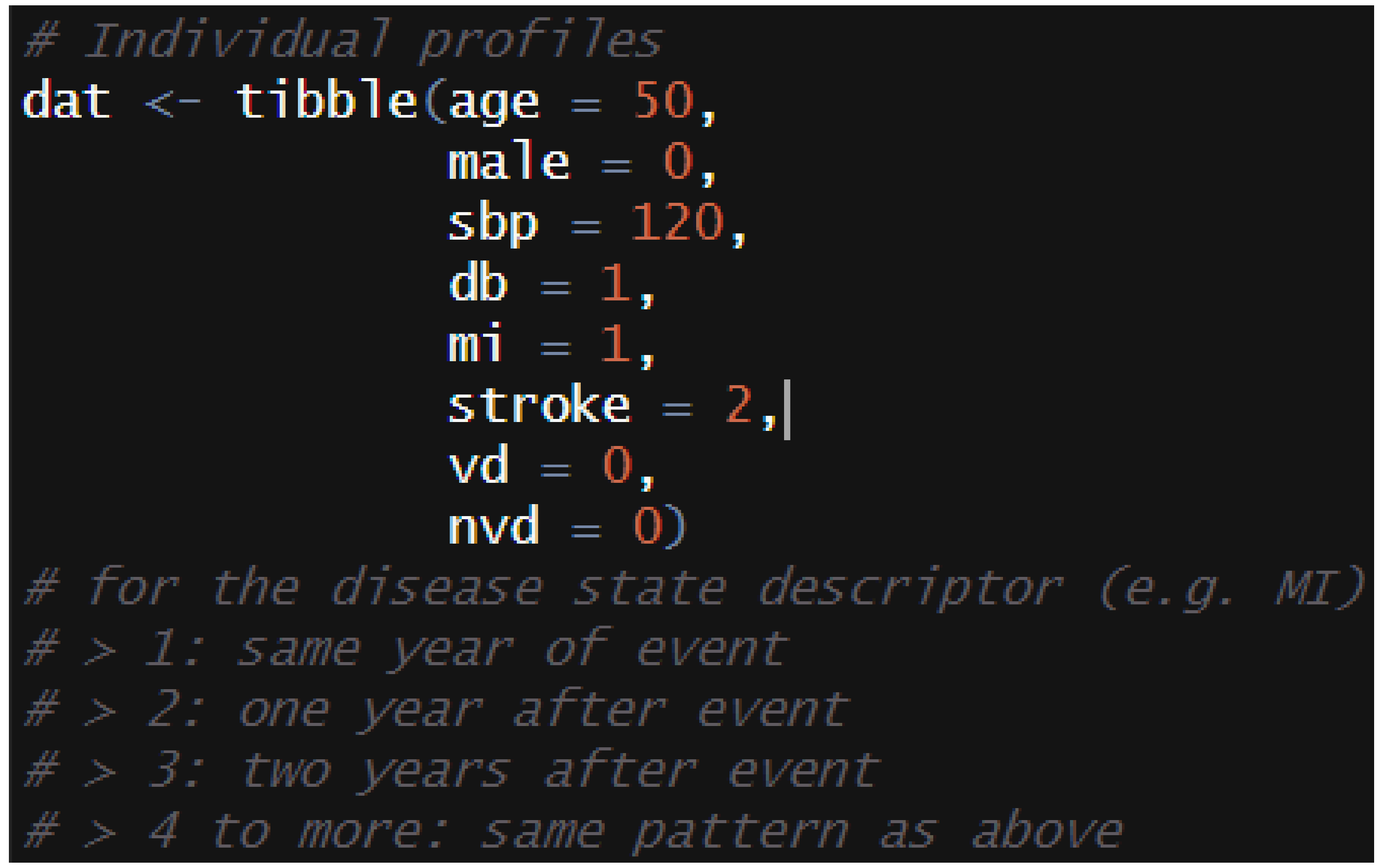
Covariate	Category	Part 1: Probability of incurring cost OR (95% CIs)	Part 2: Cost, if any incurred Mean (95% CIs)
Intercept[a]		0.13 (0.12–0.13)	2177 (2152–2201)
Baseline characteristics			
Sex (ref: female)	Male	0.93 (0.9–0.97)	−81 (−118 to −45)
Systolic blood pressure (centred at 140; per 20 mmHg)		[b]	22 (3–41)
Prior diabetes (ref: no)	Yes	1.11 (1.01–1.22)	[b]
Time-updated characteristics			
Current age (centred at 60; per 10 years)		1.37 (1.34–1.4)	158 (136–179)
Myocardial infarction (ref: no)	Same year	36.83 (24.07–56.37)	3421 (2949–3893)
	1 year ago	2.04 (1.34–3.11)	841 (323–1359)
	2 years ago	1.87 (1.17–2.97)	332 (−125 to 789)
	≥3 years ago	1.34 (1.01–1.77)	372 (87–657)
Stroke (ref: no)	Same year	38.7 (24.72–60.59)	4697 (4059–5335)
	1 year ago	2.87 (1.91–4.31)	1995 (1377–2612)
	2 years ago	2.26 (1.42–3.58)	488 (16–961)
	≥3 years ago	1.62 (1.28–2.05)	924 (635–1213)
Vascular death (ref = no)	Yes	[b]	4786 (2639–6933)
Non-vascular death (ref = no)	Yes	9.56 (7.44–12.29)	4984 (4502–5466)

Illustrative Example

Step 4. Use of developed model – Individual prediction

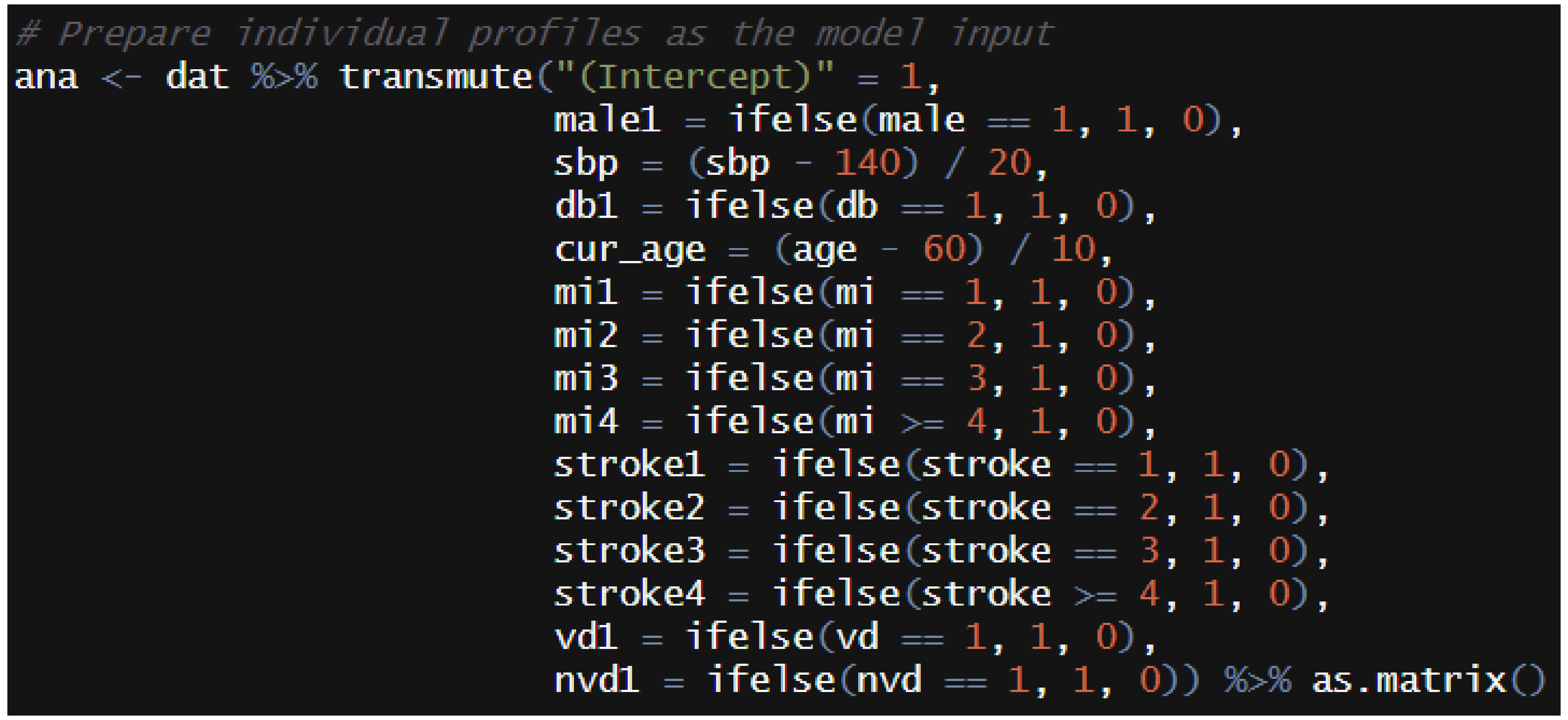
Prepare the profiles of the individual as the model input

- A 50-year old female, with a SBP of 120 mmHg, diagnosed with diabetes, had a MI in the year, a stroke 1 year ago, without other incident cardiovascular or other events modelled



```
# Individual profiles
dat <- tibble(age = 50,
              male = 0,
              sbp = 120,
              db = 1,
              mi = 1,
              stroke = 2,
              vd = 0,
              nvd = 0)
# for the disease state descriptor (e.g. MI)
# > 1: same year of event
# > 2: one year after event
# > 3: two years after event
# > 4 to more: same pattern as above
```

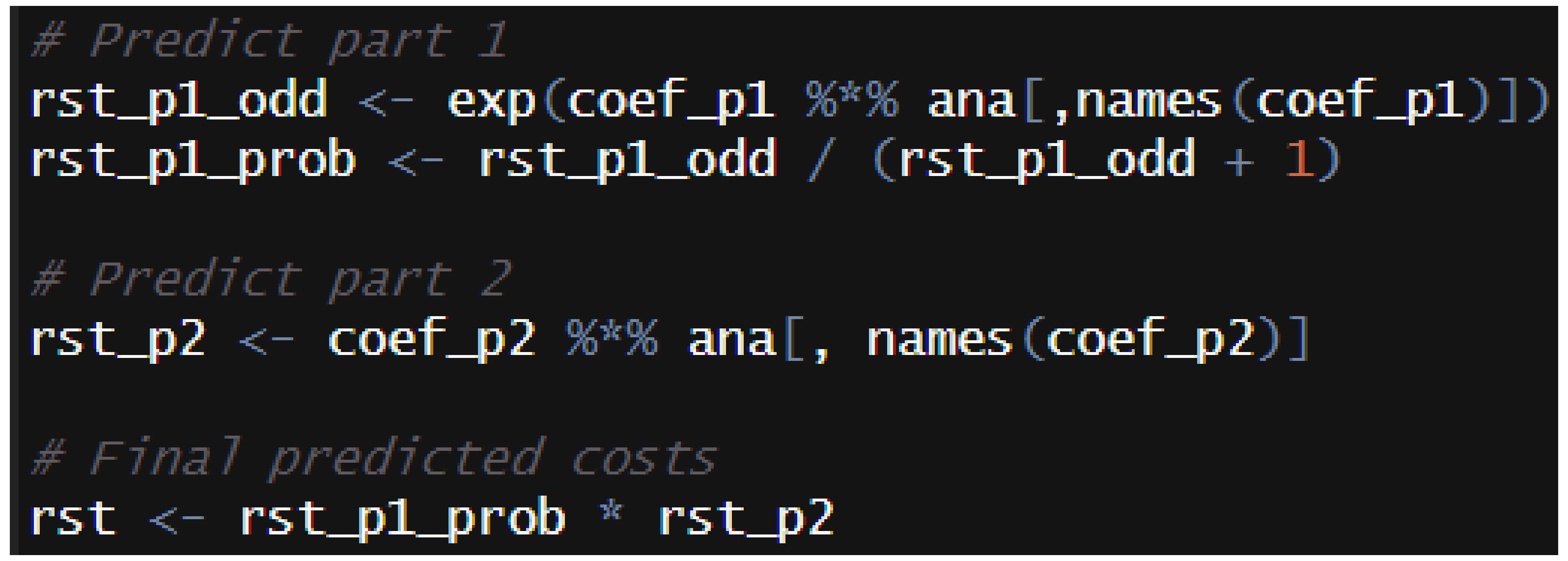
Prepare profiles



```
# Prepare individual profiles as the model input
ana <- dat %>% transmute("(Intercept)" = 1,
                         male1 = ifelse(male == 1, 1, 0),
                         sbp = (sbp - 140) / 20,
                         db1 = ifelse(db == 1, 1, 0),
                         cur_age = (age - 60) / 10,
                         mi1 = ifelse(mi == 1, 1, 0),
                         mi2 = ifelse(mi == 2, 1, 0),
                         mi3 = ifelse(mi == 3, 1, 0),
                         mi4 = ifelse(mi >= 4, 1, 0),
                         stroke1 = ifelse(stroke == 1, 1, 0),
                         stroke2 = ifelse(stroke == 2, 1, 0),
                         stroke3 = ifelse(stroke == 3, 1, 0),
                         stroke4 = ifelse(stroke >= 4, 1, 0),
                         vd1 = ifelse(vd == 1, 1, 0),
                         nvd1 = ifelse(nvd == 1, 1, 0)) %>% as.matrix()
```

Use the models to calculate the costs

- Part 1 – probability of any costs in the year = 0.92
- Part 2 – costs conditional on any incurring = 7413
- Predicted costs = 0.92 x 7413 = 6783



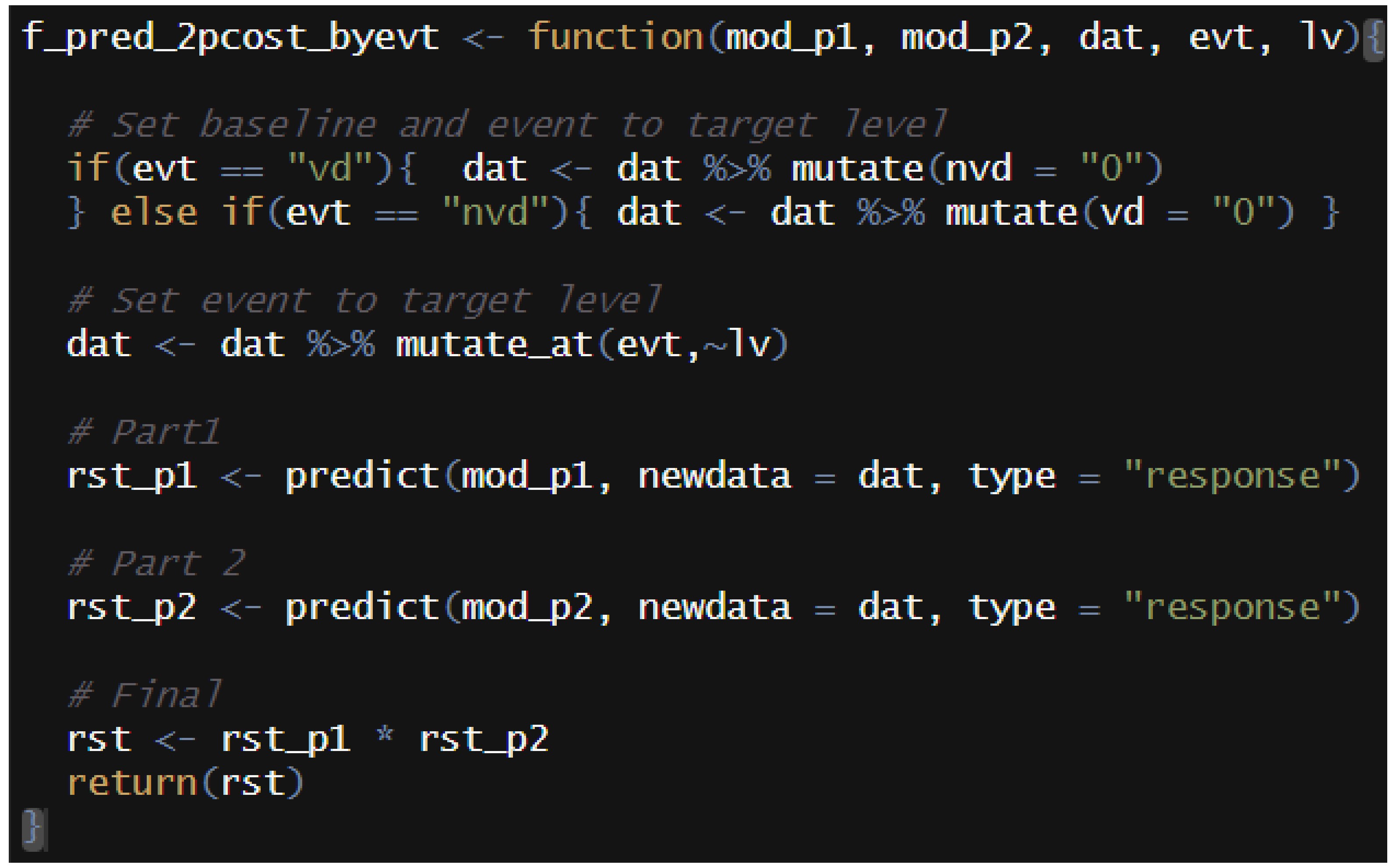
```
# Predict part 1
rst_p1_odd <- exp(coef_p1 %*% ana[,names(coef_p1)])
rst_p1_prob <- rst_p1_odd / (rst_p1_odd + 1)

# Predict part 2
rst_p2 <- coef_p2 %*% ana[, names(coef_p2)]

# Final predicted costs
rst <- rst_p1_prob * rst_p2
```

Illustrative Example

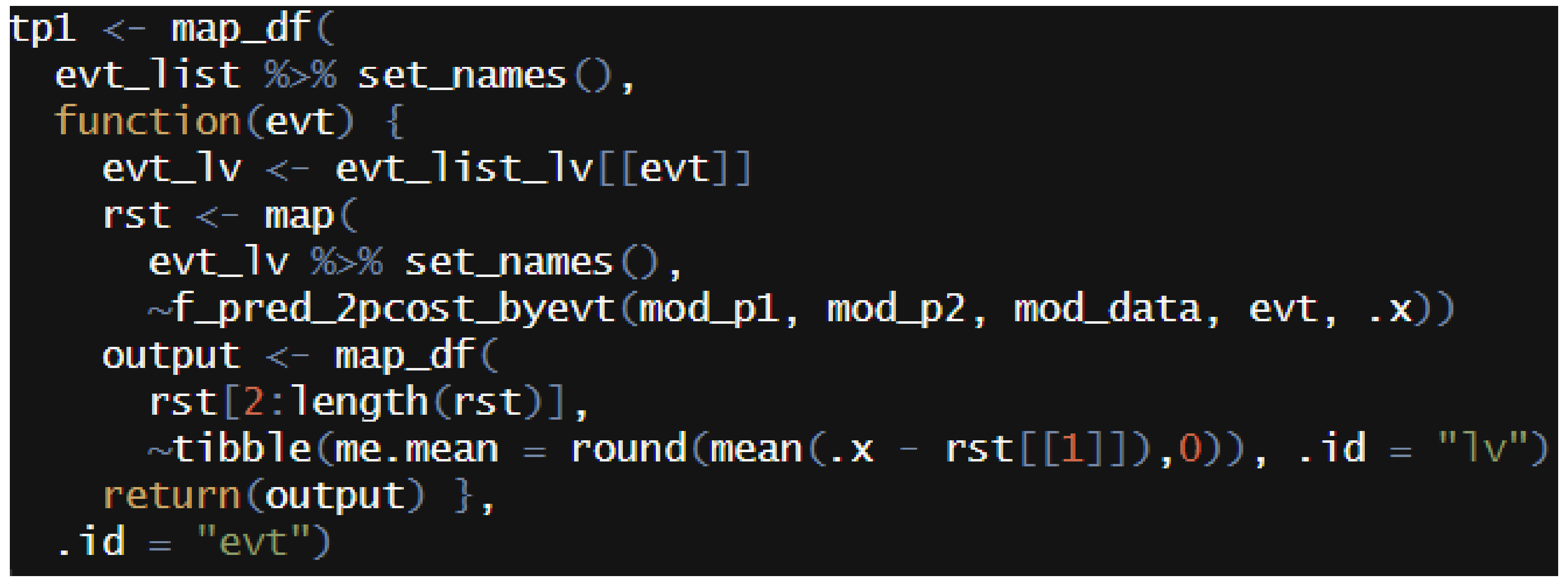
Step 4. Use of developed model – Marginal effect estimation

Marginal effect, or average costs associated with disease model state

- Mean A – Mean B (A assumes all have the condition; B assumes none has the condition)



```
f_pred_2pcost_byevt <- function(mod_p1, mod_p2, dat, evt, lv){

  # Set baseline and event to target level
  if(evt == "vd"){  dat <- dat %>% mutate(nvd = "0")
  } else if(evt == "nvd"){ dat <- dat %>% mutate(vd = "0") }

  # Set event to target level
  dat <- dat %>% mutate_at(evt,~lv)

  # Part1
  rst_p1 <- predict(mod_p1, newdata = dat, type = "response")

  # Part 2
  rst_p2 <- predict(mod_p2, newdata = dat, type = "response")

  # Final
  rst <- rst_p1 * rst_p2
  return(rst)
}
```



```
tp1 <- map_df(
  evt_list %>% set_names(),
  function(evt) {
    evt_lv <- evt_list_lv[[evt]]
    rst <- map(
      evt_lv %>% set_names(),
      ~f_pred_2pcost_byevt(mod_p1, mod_p2, mod_data, evt, .x))
    output <- map_df(
      rst[2:length(rst)],
      ~tibble(me.mean = round(mean(.x - rst[[1]]),0)), .id = "lv")
    return(output) },
  .id = "evt")
```

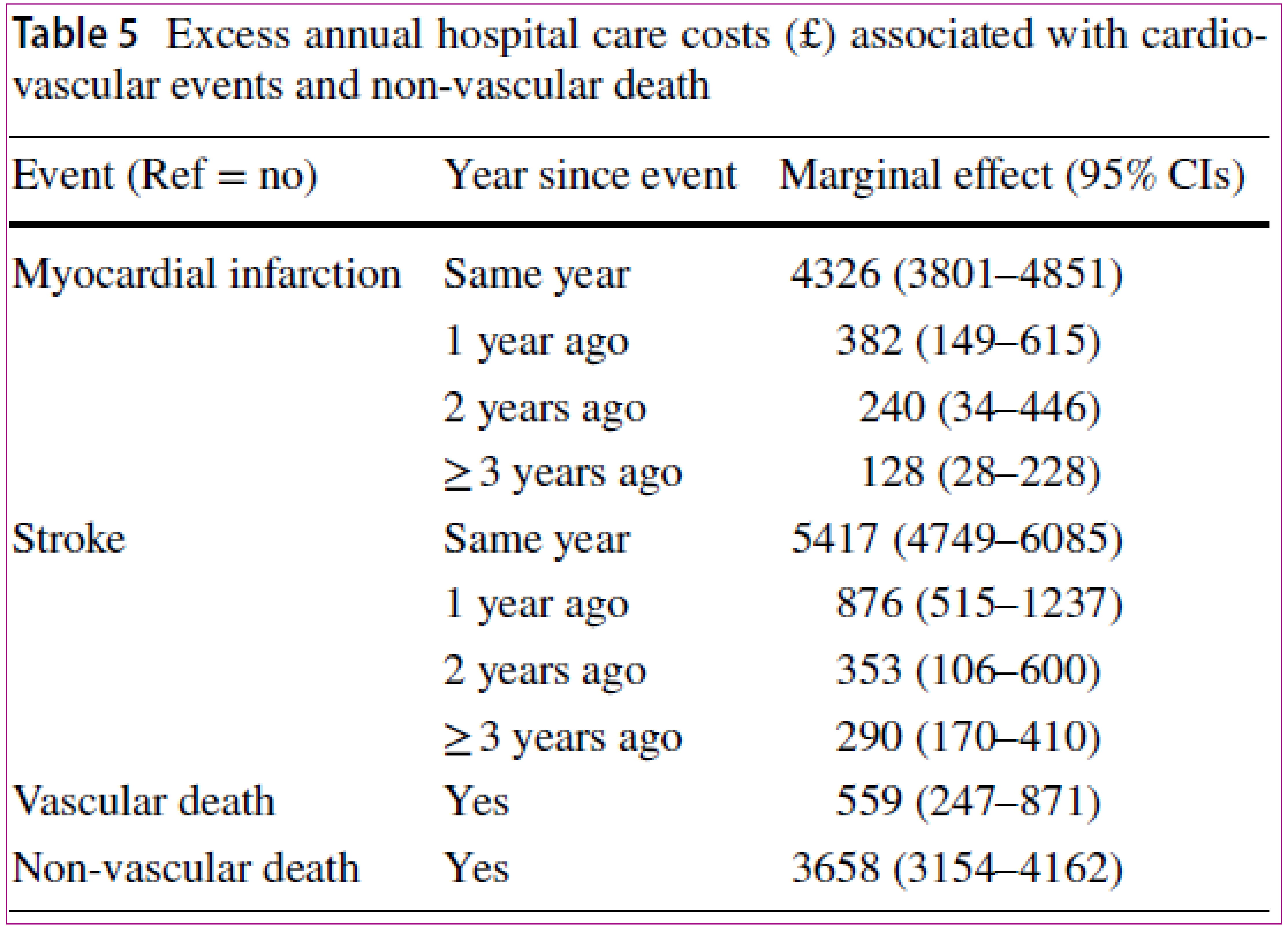
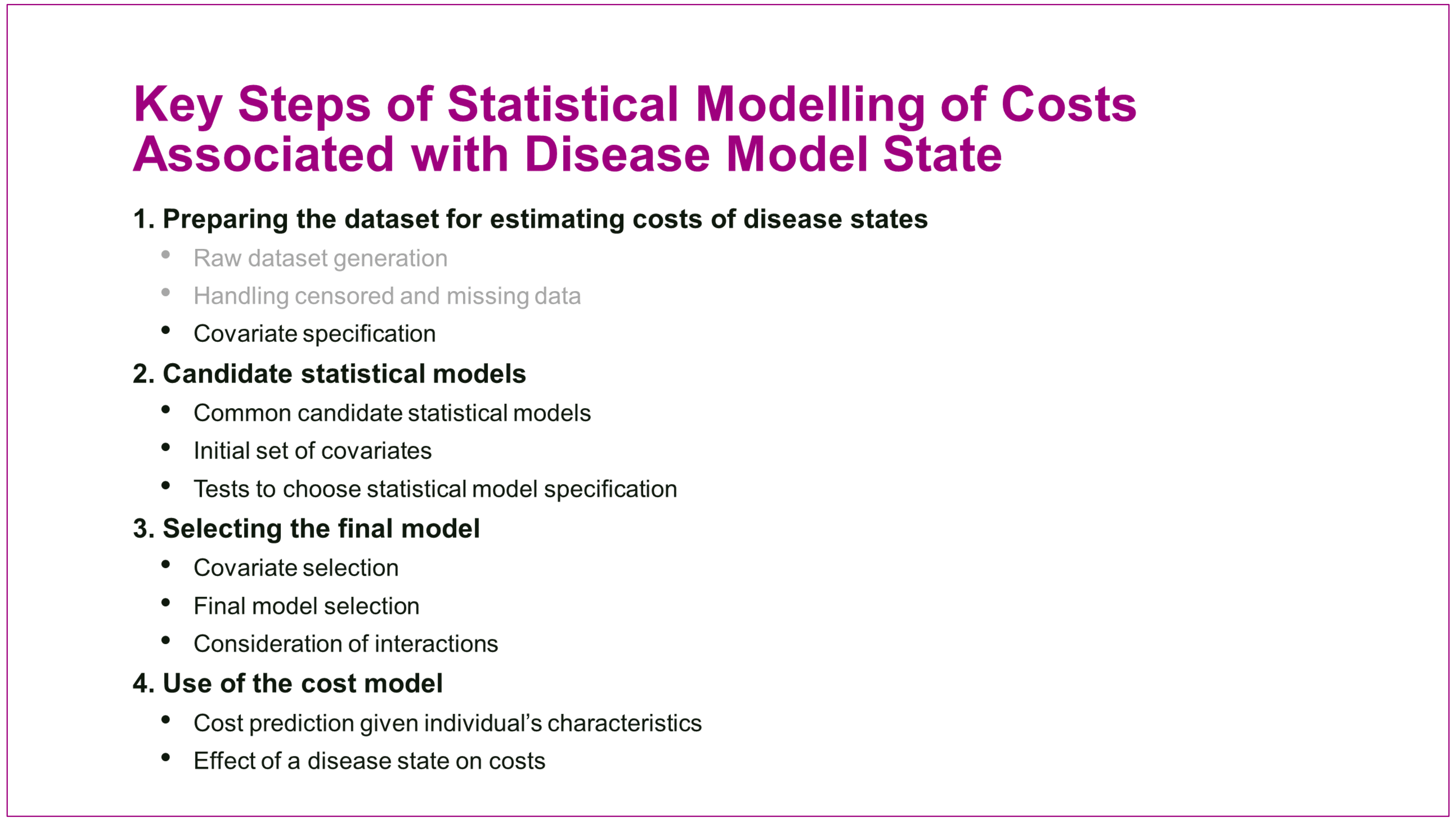


Table 5 Excess annual hospital care costs (£) associated with cardiovascular events and non-vascular death

Event (Ref = no)	Year since event	Marginal effect (95% CIs)
Myocardial infarction	Same year	4326 (3801–4851)
	1 year ago	382 (149–615)
	2 years ago	240 (34–446)
	≥3 years ago	128 (28–228)
Stroke	Same year	5417 (4749–6085)
	1 year ago	876 (515–1237)
	2 years ago	353 (106–600)
	≥3 years ago	290 (170–410)
Vascular death	Yes	559 (247–871)
Non-vascular death	Yes	3658 (3154–4162)

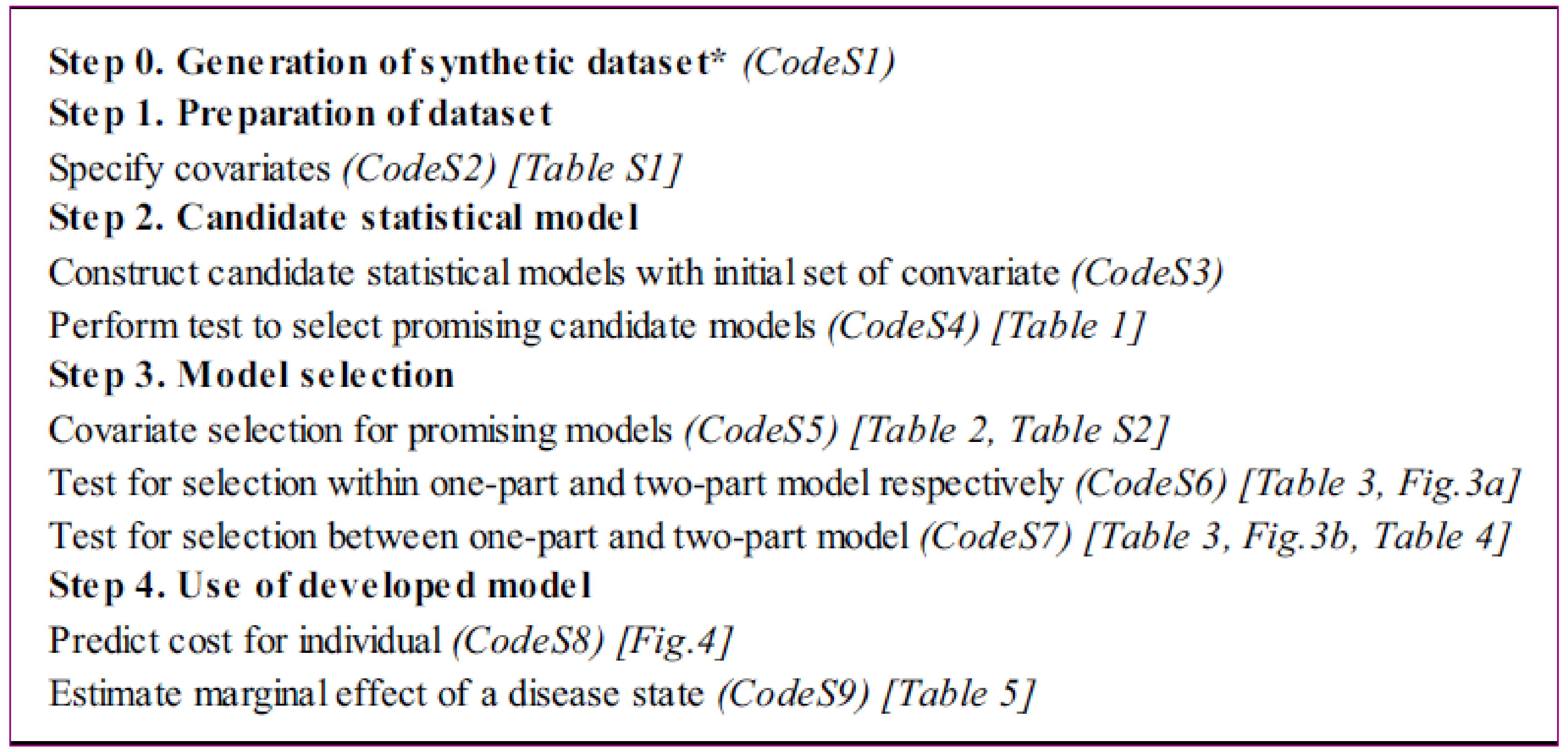
Further Information

Published tutorial paper



Key Steps of Statistical Modelling of Costs Associated with Disease Model State

1. Preparing the dataset for estimating costs of disease states
   - Raw dataset generation
   - Handling censored and missing data
   - Covariate specification
2. Candidate statistical models
   - Common candidate statistical models
   - Initial set of covariates
   - Tests to choose statistical model specification
3. Selecting the final model
   - Covariate selection
   - Final model selection
   - Consideration of interactions
4. Use of the cost model
   - Cost prediction given individual's characteristics
   - Effect of a disease state on costs



Step 0. Generation of synthetic dataset* *(CodeS1)*

Step 1. Preparation of dataset

Specify covariates *(CodeS2) [Table S1]*

Step 2. Candidate statistical model

Construct candidate statistical models with initial set of convariate *(CodeS3)*

Perform test to select promising candidate models *(CodeS4) [Table 1]*

Step 3. Model selection

Covariate selection for promising models *(CodeS5) [Table 2, Table S2]*

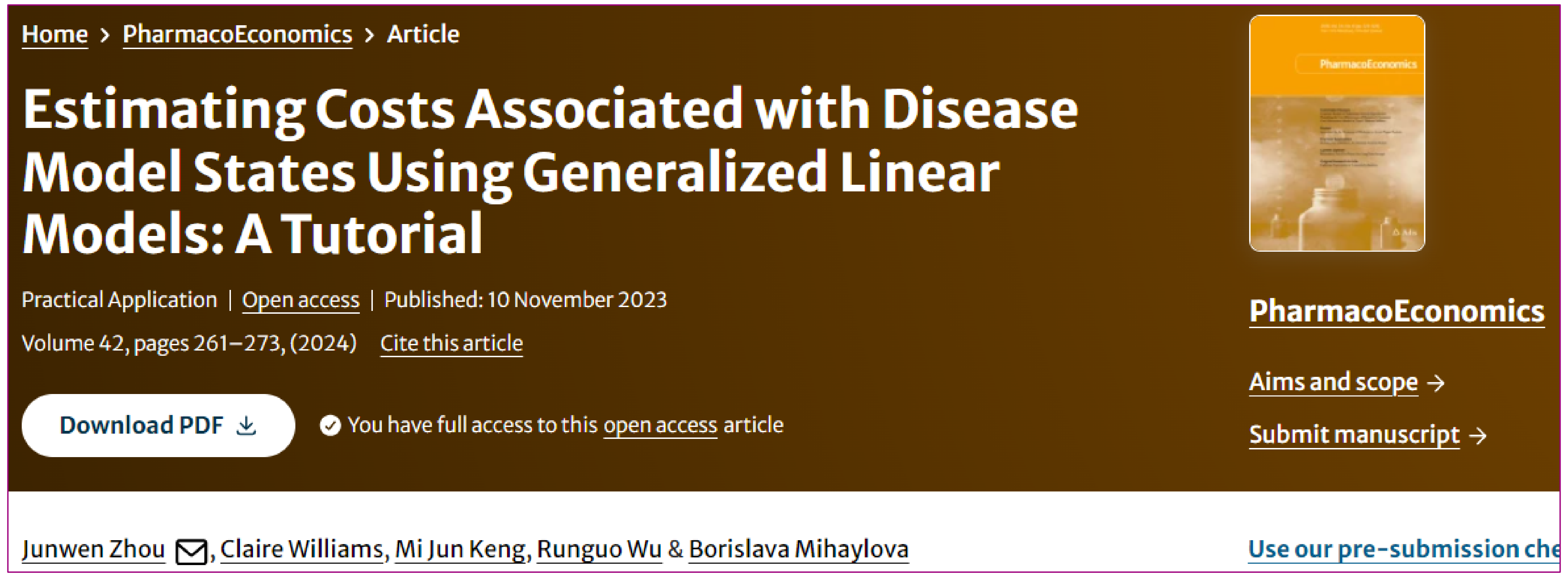
Test for selection within one-part and two-part model respectively *(CodeS6) [Table 3, Fig.3a]*

Test for selection between one-part and two-part model *(CodeS7) [Table 3, Fig.3b, Table 4]*

Step 4. Use of developed model

Predict cost for individual *(CodeS8) [Fig.4]*

Estimate marginal effect of a disease state *(CodeS9) [Table 5]*



Home > PharmacoEconomics > Article

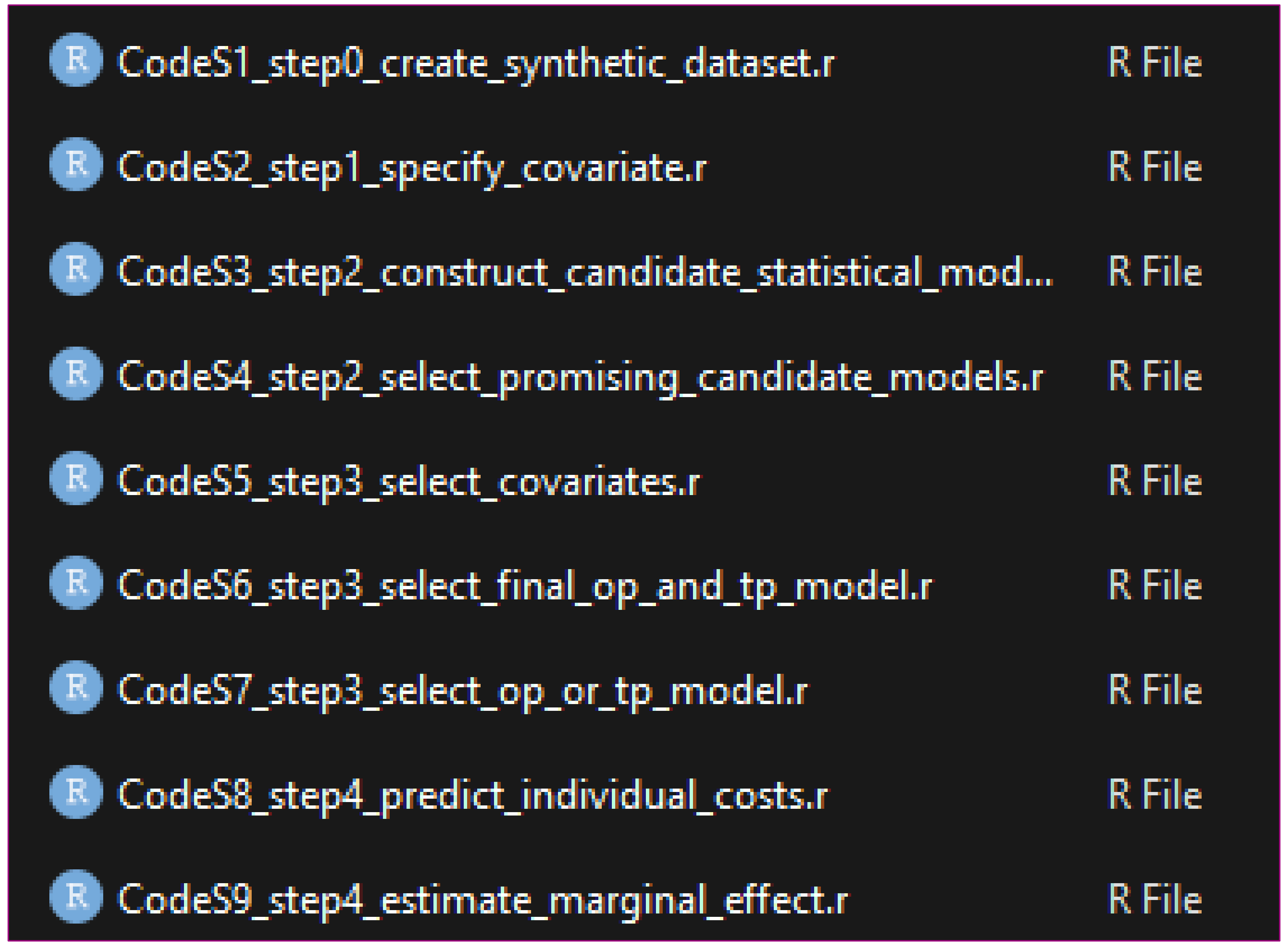
Estimating Costs Associated with Disease Model States Using Generalized Linear Models: A Tutorial

Practical Application | Open access | Published: 10 November 2023

Volume 42, pages 261–273, (2024) Cite this article

Download PDF
You have full access to this open access article
PharmacoEconomics
Aims and scope →
Submit manuscript →
Use our pre-submission che

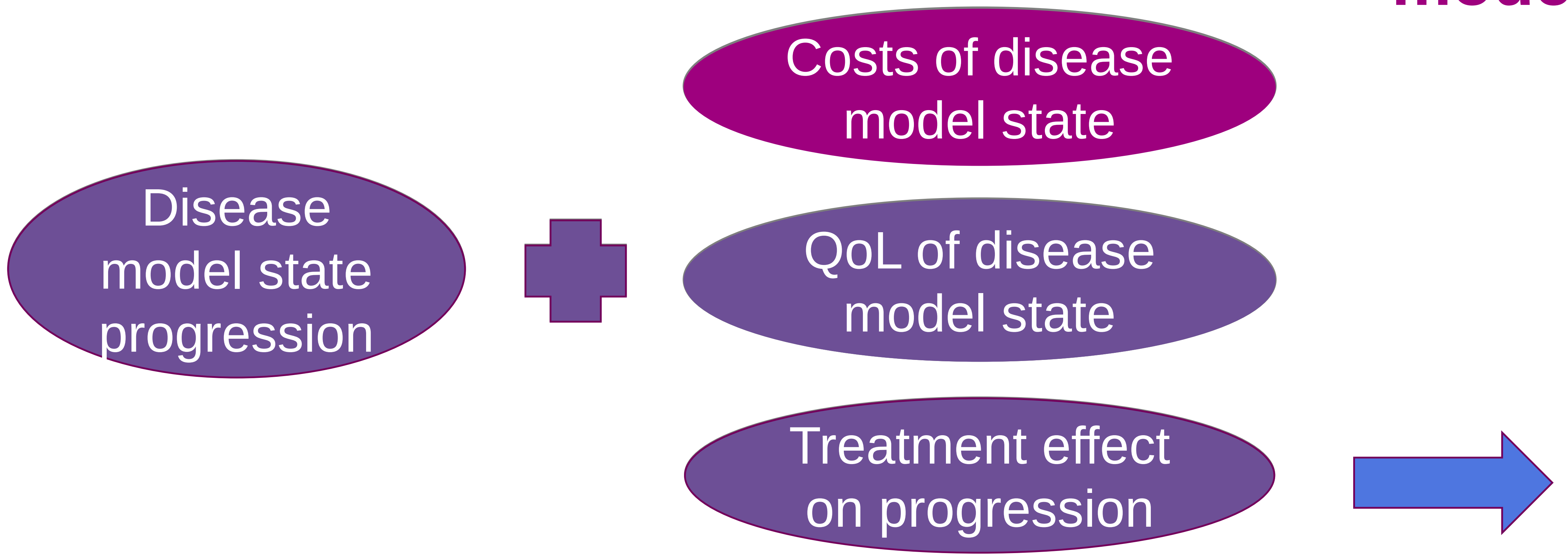
Junwen Zhou, Claire Williams, Mi Jun Keng, Runguo Wu & Borislava Mihaylova



```
CodeS1_step0_create_synthetic_dataset.r                R File
CodeS2_step1_specify_covariate.r                       R File
CodeS3_step2_construct_candidate_statistical_mod...    R File
CodeS4_step2_select_promising_candidate_models.r       R File
CodeS5_step3_select_covariates.r                       R File
CodeS6_step3_select_final_op_and_tp_model.r            R File
CodeS7_step3_select_op_or_tp_model.r                   R File
CodeS8_step4_predict_individual_costs.r                R File
CodeS9_step4_estimate_marginal_effect.r                R File
```

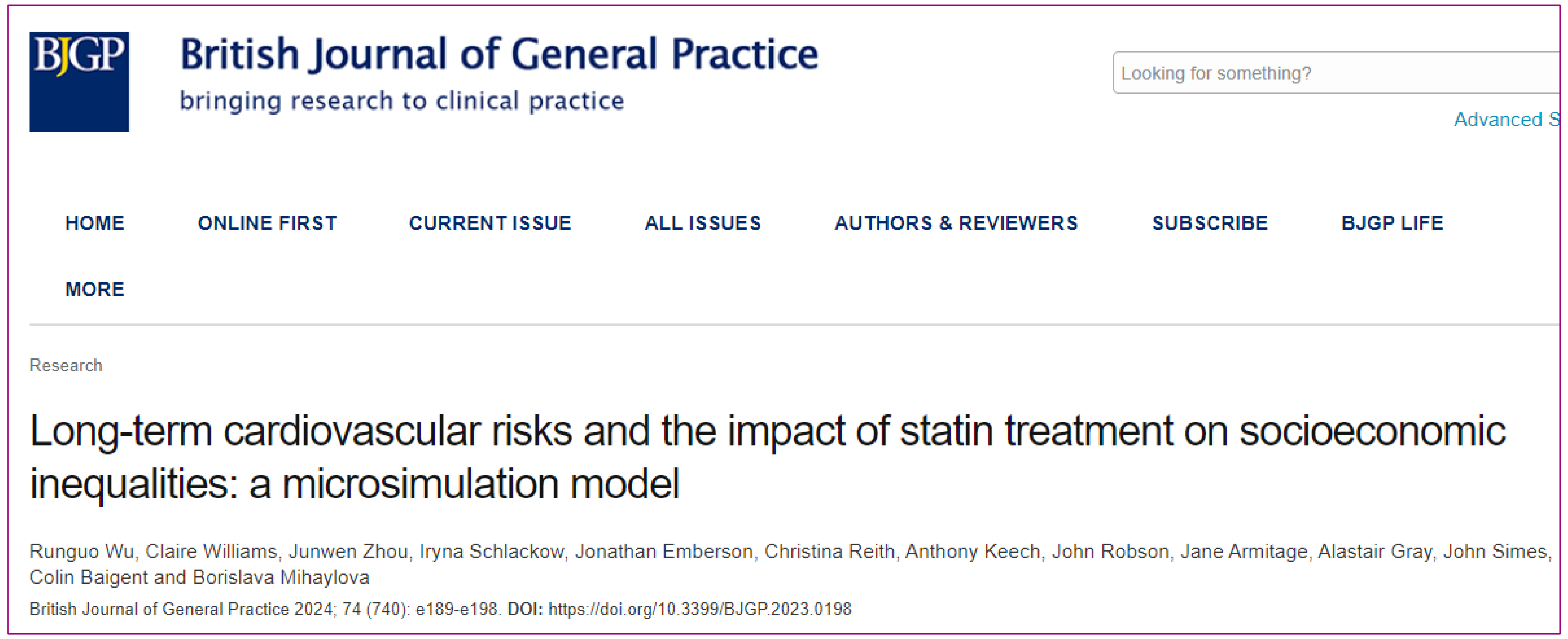
Further Information

Cost model supporting economic evaluation



Disease model state progression
Costs of disease model state
QoL of disease model state
Treatment effect on progression

Microsimulation & QoL model, treatment effect



Research

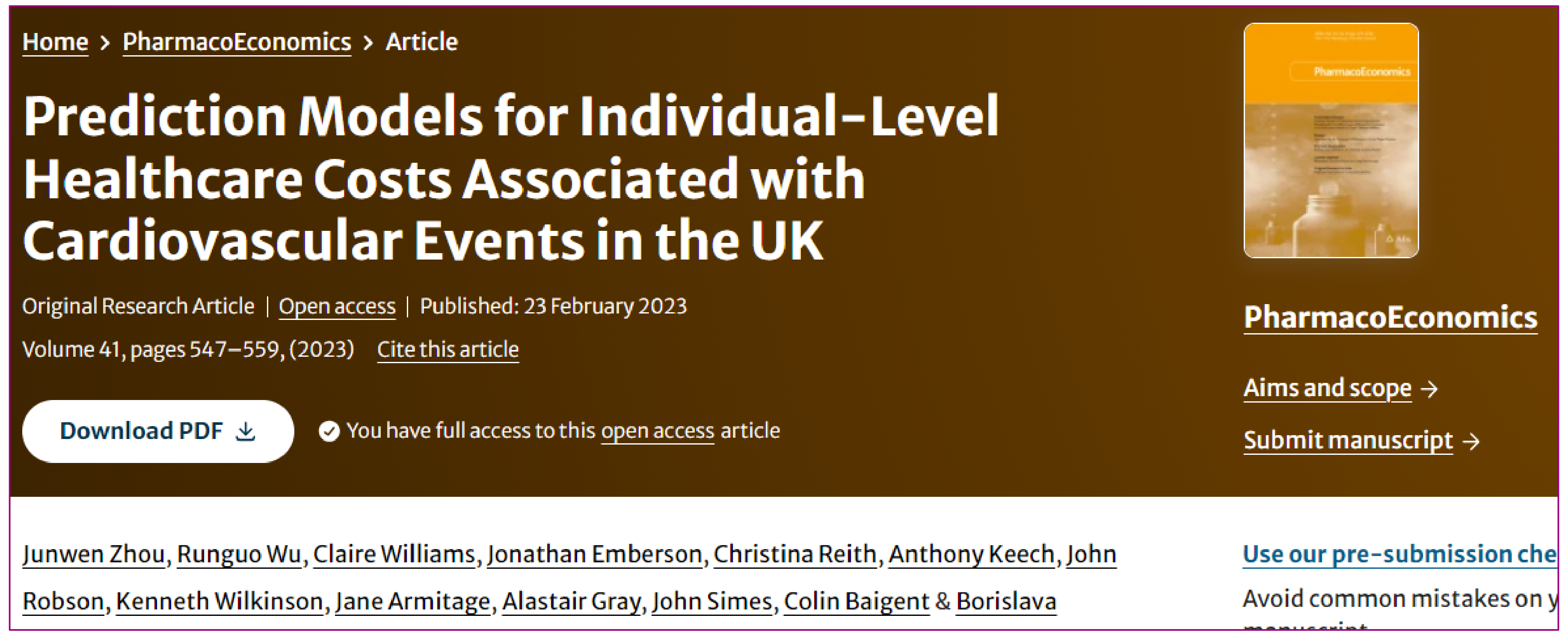
Long-term cardiovascular risks and the impact of statin treatment on socioeconomic inequalities: a microsimulation model

Runguo Wu, Claire Williams, Junwen Zhou, Iryna Schlackow, Jonathan Emberson, Christina Reith, Anthony Keech, John Robson, Jane Armitage, Alastair Gray, John Simes, Colin Baigent and Borislava Mihaylova

British Journal of General Practice 2024; 74 (740): e189-e198. DOI: https://doi.org/10.3399/BJGP.2023.0198

https://bjgp.org/content/74/740/e189

Cost model



Home › PharmacoEconomics › Article
Download PDF
PharmacoEconomics
Aims and scope →
Submit manuscript →
Use our pre-submission che
Avoid common mistakes on y

Prediction Models for Individual-Level Healthcare Costs Associated with Cardiovascular Events in the UK

Original Research Article | Open access | Published: 23 February 2023

Volume 41, pages 547–559, (2023) Cite this article

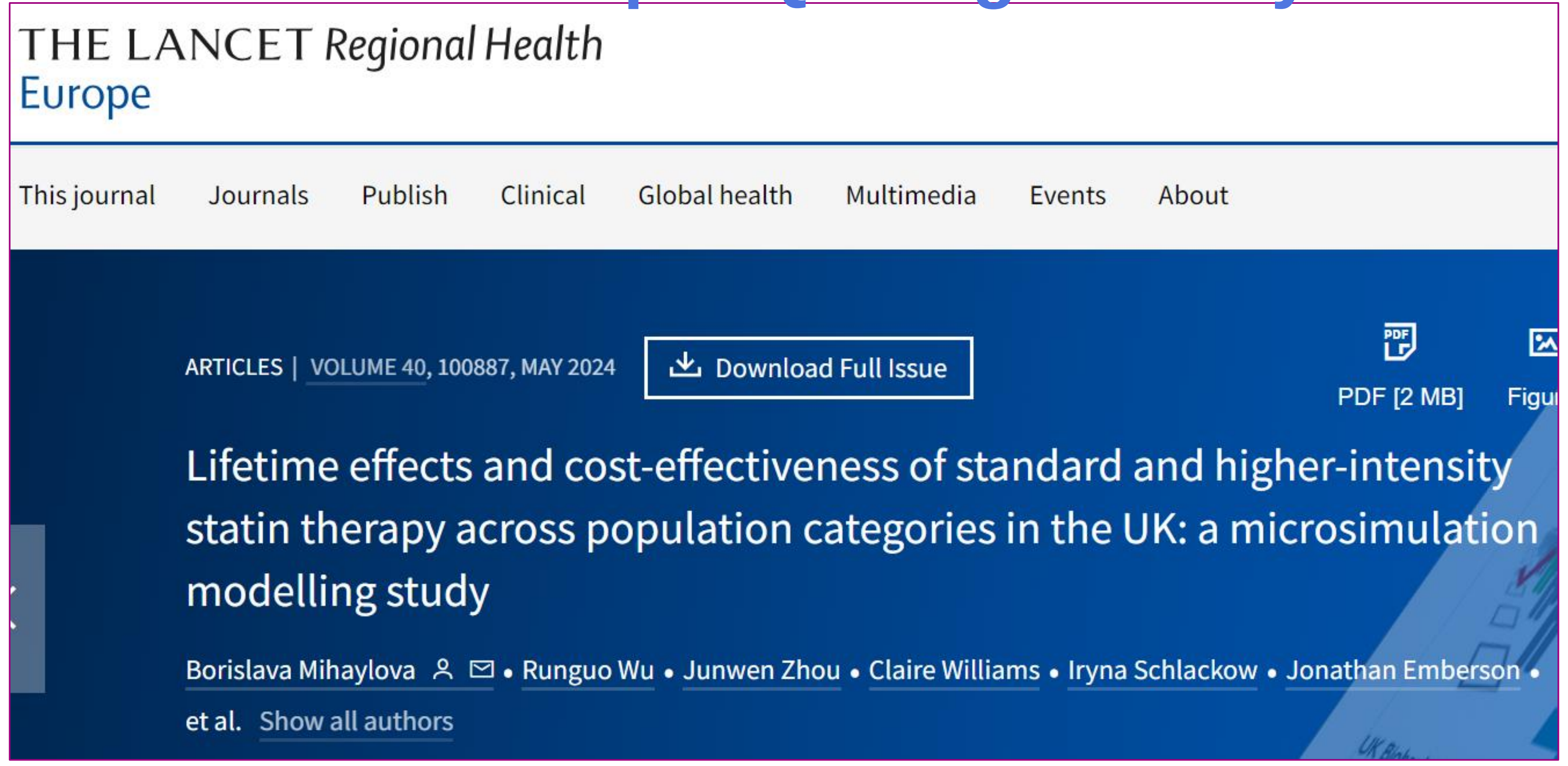
You have full access to this open access article

Junwen Zhou, Runguo Wu, Claire Williams, Jonathan Emberson, Christina Reith, Anthony Keech, John Robson, Kenneth Wilkinson, Jane Armitage, Alastair Gray, John Simes, Colin Baigent & Borislava

https://link.springer.com/article/10.1007/s40273-022-01219-6

Economic evaluation of intervention

- Total costs and QALYs under intervention and control
- Increment cost per QALY gained by the intervention



ARTICLES | VOLUME 40, 100887, MAY 2024

Lifetime effects and cost-effectiveness of standard and higher-intensity statin therapy across population categories in the UK: a microsimulation modelling study

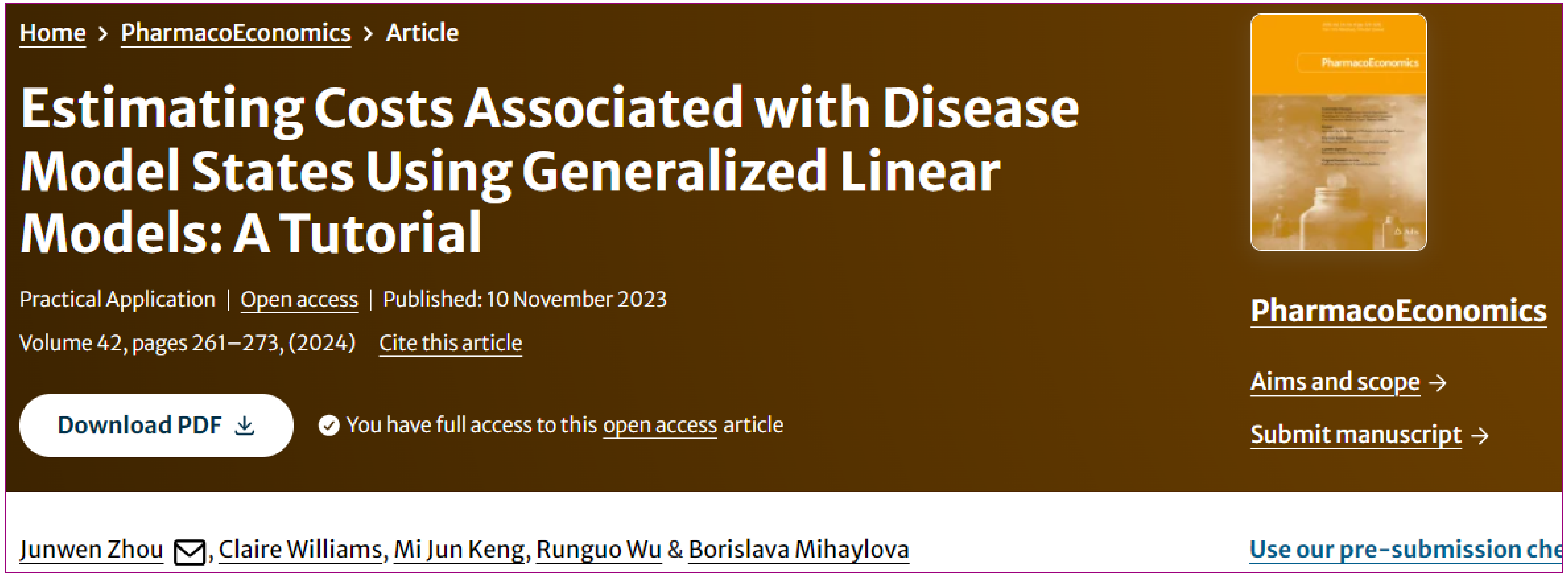
Borislava Mihaylova • Runguo Wu • Junwen Zhou • Claire Williams • Iryna Schlackow • Jonathan Emberson •

et al. Show all authors

https://doi.org/10.1016/j.lanepe.2024.100887

Final Remarks

- **Hope it is a useful starting point for researchers who plan to conduct costing analyses**



Home > PharmacoEconomics > Article

Estimating Costs Associated with Disease Model States Using Generalized Linear Models: A Tutorial

Practical Application | Open access | Published: 10 November 2023

Volume 42, pages 261–273, (2024) Cite this article

✓ You have full access to this open access article

Junwen Zhou, Claire Williams, Mi Jun Keng, Runguo Wu & Borislava Mihaylova

https://link.springer.com/article/10.1007/s40273-023-01319-x

- **Looking forward to more and more costing studies published to support HTA activities**