Quantifying excess mortality by cause of death against a pre-pandemic and socio-economic factor-dependent baseline

Katrien Antonio, Torsten Kleinow, Irene Simonetti,
Frank van Berkum & Michel Vellekoop

VSAE Actuarial Congress, 04 March 2025

University of Amsterdam

RCLR Research Centre for Longevity Risk

Outline

Introduction

Research objectives and approach

CBS microdata

Modelling cause-specific, pre-pandemic mortality with socio-economic factors
- Methodology
- Results

Quantifying and analyzing (excess) mortality by cause of death during 2020-2021
- Methodology
- Results

RCLR

Outline

Introduction

Research objectives and approach

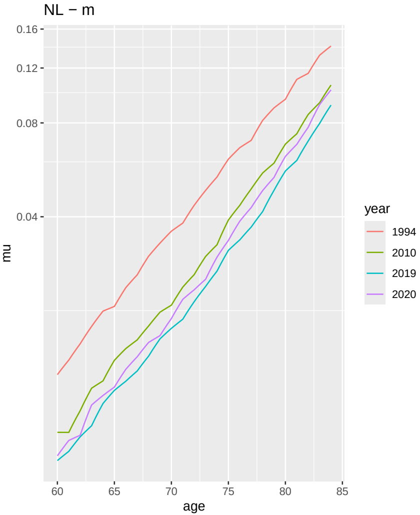
CBS microdata

Modelling cause-specific, pre-pandemic mortality with socio-economic factors
- Methodology
- Results

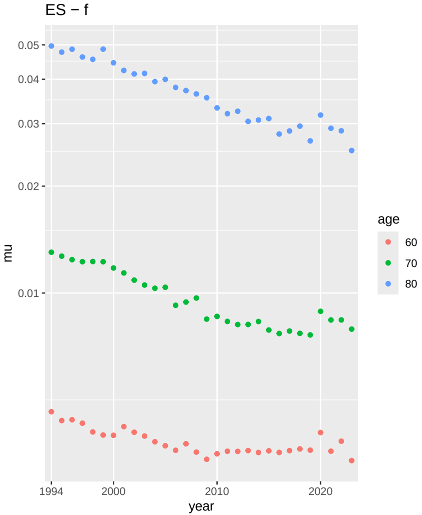
Quantifying and analyzing (excess) mortality by cause of death during 2020-2021
- Methodology
- Results

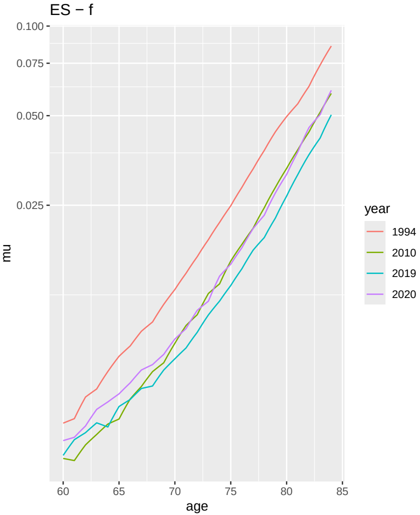
RCLR

Introduction



Introduction



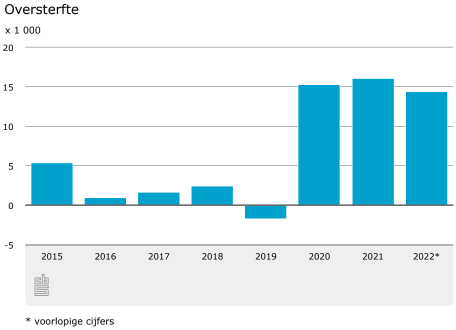


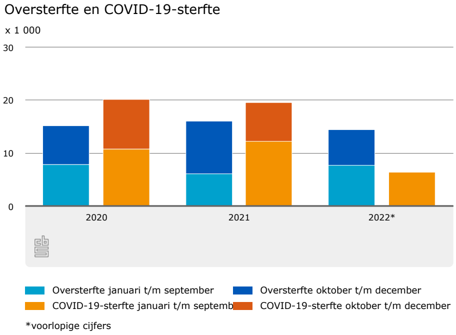
Introduction

- European countries lost about 10 years of mortality improvement.
- But in many countries, including the UK and the US, mortality slowdowns started earlier than 2020
- The challenge for setting mortality assumption for internal models is to project mortality
- ... and for that we need to deal with the effect of the pandemic
- In the Netherlands, CBS reported:
  - Excess mortality for the third consecutive year in 2022

RCLR

Introduction

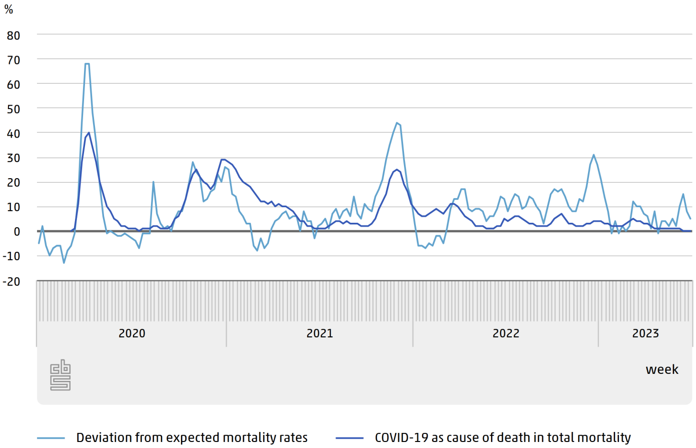




www.cbs.nl/en-gb/news/2023/04/excess-mortality-for-the-third-consecutive-year-in

Introduction

Deviation from expected mortality, and share of COVID-19 mortality



www.cbs.nl/en-gb/news/2023/50/covid-19-mortality-continues-to-decline-in-2023

Introduction

CBS:

- In 2023, excess mortality stood at 12.7 thousand (8 percent).
- In 2022, the figure was 14.7 thousand (9 percent).
- based on the COVID-19 mortality monitor
- CBS launched this monitor at the start of the pandemic, but stopped calculating possible excess mortality on a weekly basis at the end of 2023

www.cbs.nl/en-gb/news/2024/06/fewer-deaths-in-2023

RCLR

Introduction

- For calculating excess mortality and for taking the pandemic effect into account for mortality projections, **we need to construct a reliable baseline = expected mortality during 2020 - 2023**
- This is the aim of an RCLR research project funded by ZonMw

About RCLR: the Research Centre for Longevity Risk is a joint initiative of NN Group and the University of Amsterdam, with additional funding from the Dutch government's Public-Private Partnership (PPP) programme.

www.rclr.nl

Outline

Introduction

Research objectives and approach

CBS microdata

Modelling cause-specific, pre-pandemic mortality with socio-economic factors
- Methodology
- Results

Quantifying and analyzing (excess) mortality by cause of death during 2020-2021
- Methodology
- Results

Research objectives and approach

Overall objective: to quantify and to analyze excess mortality against a pre-pandemic baseline for expected cause-specific mortality in 2020 and 2021, taking socio-economic factors into account.

A phased approach, with intermediate research objectives:

1. to establish a cause-specific, pre-pandemic baseline mortality level, taking socio-economic factors into account

2.a. to quantify cause-specific excess mortality during the COVID waves in 2020 and 2021

2.b. to analyze to what extent excess mortality can be attributed to socio-economic risk factors and information on COVID-19 testing and vaccination.

RCLR

Outline

Introduction

Research objectives and approach

CBS microdata

Modelling cause-specific, pre-pandemic mortality with socio-economic factors
- Methodology
- Results

Quantifying and analyzing (excess) mortality by cause of death during 2020-2021
- Methodology
- Results

CBS microdata

Overview of datasets used

Dutch registry data

Made available via CBS microdata service

125 jaar Centraal Bureau voor de Statistiek

We combine data sources on:

- (per spell) start and end date of residence spells in the Netherlands
- (per event) cause, date and location of death + (during pandemic) vaccination uptake and COVID-19 tests
- (static) date of birth, gender, migration background, education
- (annual) property value, medical expenses, income and socio-economic status.

RCLR

CBS microdata

Construction of annual, pre-pandemic data

Merge the microdata to individual spells:

- time $t \in \mathcal{T} = \{2016, ..., 2019\}$
- CoDs $c \in \mathcal{C} = \{1, ..., C = 20\}$
- individuals $j \in \mathcal{J}_t = \{1, ..., J_t\}$
- individual-specific spells $i \in \mathcal{I}_{t,j} = \{1, ..., I_{t,j}\}$ with constant socio-economic factors

For each (t, j, i) combination we have:

- exposure-to-risk $\tau_{t,j,i}$
- death indicator $\delta_{t,j,i}$ and cause-specific death indicator $\delta^c_{t,j,i}$
- combination of constant risk factors.

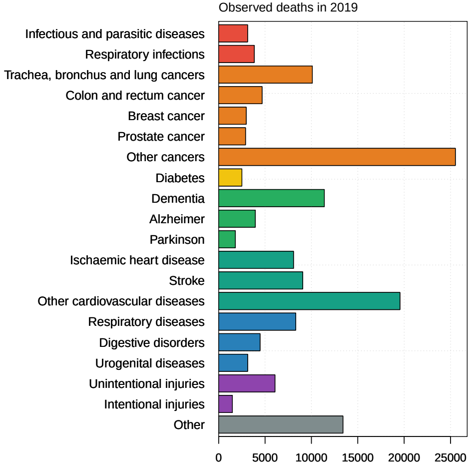
CBS microdata

Construction of annual, pre-pandemic data, CoDs under investigation

Mapping based on WHO, 2020:

1. Communicable diseases
2. Non-communicable diseases
3. Injuries

Final selection of 20 pre-COVID causes based on materiality.



RCLR

CBS microdata

Construction of weekly, pandemic data

Merge the microdata to weekly, individual spells:

- time $t \in \{2020,\ 2021\}$, the pandemic period
- ISO-8601 week definition
- additional CoD: COVID-19.

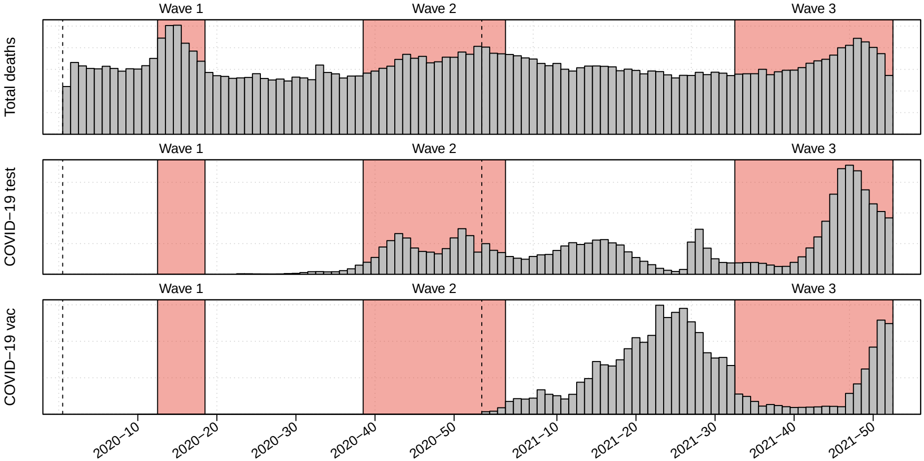
Extend available socio-economic factors, per week, with:

- time elapsed since last positive COVID infection test (in days and weeks)
- time elapsed since last COVID vaccination (in days and weeks)
- indicators for infection or vaccination in the given week.

RCLR

CBS microdata

Observed deaths and COVID-19 information



Outline

Introduction

Research objectives and approach

CBS microdata

Modelling cause-specific, pre-pandemic mortality with socio-economic factors
- Methodology
- Results

Quantifying and analyzing (excess) mortality by cause of death during 2020-2021
- Methodology
- Results

RCLR

Outline

Introduction

Research objectives and approach

CBS microdata

Modelling cause-specific, pre-pandemic mortality with socio-economic factors
- Methodology
- Results

Quantifying and analyzing (excess) mortality by cause of death during 2020-2021
- Methodology
- Results

Pre-pandemic mortality levels

Rationale: importance of the baseline mortality level when studying excess mortality during the pandemic.

Use statistical learning to calibrate

on the data from $\{2016, \ldots, 2019\}$

- $\mu_{t,j,i}$: the all cause force of mortality
- $\mu^{c}_{t,j,i}$: the cause c specific force of mortality,

applicable to individual j, in year t, with socio-economic factors specific to spell i,

including a time trend, an age and gender effect and the effect of selected socio-economic factors.

Pre-pandemic mortality levels

Quasi-Poisson regression with GAMs, all cause

From the (t, j, i) records of individual-specific spells we construct

$$\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{\delta}, \boldsymbol{\tau}) = \prod_{t \in \mathcal{T}} \prod_{j \in \mathcal{J}_t} \prod_{i \in \mathcal{I}_{t,j}} \exp\left[-\tau_{t,j,i} \cdot \mu_{t,j,i}(\boldsymbol{\theta})\right] \cdot \left(\mu_{t,j,i}(\boldsymbol{\theta})\right)^{\delta_{t,j,i}},$$

the survival likelihood, where the force of mortality depends on parameter vector $\boldsymbol{\theta}$.

This likelihood function is proportional a Poisson likelihood, see Brouhns et al., 2002. We switch to quasi-Poisson to allow for overdispersion.

We use Generalized Additive Models (GAMs) to calibrate an additive predictor (with smooth functions of covariates) for $\log \mu_{t,j,i}$, see van Berkum et al., 2020.

RCLR

Pre-pandemic mortality levels

Quasi-Poisson regression with GAMs, cause-specific

Analogously, for cause-specific mortality we construct

$$\mathcal{L}(\boldsymbol{\theta}^c|\boldsymbol{\delta}^c, \boldsymbol{\tau}) = \prod_{t\in\mathcal{T}} \prod_{j\in\mathcal{J}_t} \prod_{i\in\mathcal{I}_{t,j}} \exp\left[-\tau_{t,j,i} \cdot \mu^c_{t,j,i}(\boldsymbol{\theta}^c)\right] \cdot \left(\mu^c_{t,j,i}(\boldsymbol{\theta}^c)\right)^{\delta^c_{t,j,i}},$$

the cause-specific survival likelihood, where the force of mortality depends on parameter vector $\boldsymbol{\theta}^c$.

Again, we use GAMs with quasi-Poisson, log link and smooth functions of continuous covariates.

Pre-pandemic mortality levels

Lee-Carter structure

Recall the Lee-Carter structure:

$$\log \mu_{t,x} = \alpha_x + \beta_x \cdot \kappa_t.$$

We calibrate a smooth variant of the Lee-Carter model:

$$\log \mu_j = f_{g_j}(x_j) + h_{g_j}(x_j) \cdot (\bar{t} - t_j),$$

where only the gender g, the time t and age x of an individual j is used as covariate information. Hereby,

- $f_g(x)$ and $h_g(x)$ are gender specific smoothers of age x
- $\bar{t}$ is the average of t for $t \in \mathcal{T}$.

RCLR

Pre-pandemic mortality levels

Lee-Carter baseline structure plus socio-economic factors

To assess pandemic excess mortality we want to account for pre-pandemic existing differences in mortality among socio-economic groups.

Hereto, we examine model specifications built-up as follows:

$$
\begin{aligned}
\log \mu_i &= f_{g_i}(x_i) + h_{g_i}(x_i) \cdot (\bar{t} - t_i) \\
&+ s_{\text{ME}}(\log \text{ME}_i) \cdot \mathbb{I}(\text{ME}_i > 0) + \beta_{\text{ME\_0}} \cdot \mathbb{I}(\text{ME}_i = 0) \\
&+ s_{\text{Wealth}}(\text{Wealth}_i) \cdot \mathbb{I}(\text{Wealth}_i \text{ not missing}) \\
&+ \sum \beta_\ell \cdot \mathbb{I}(\text{PersIncSrc}_i = \ell) + \sum \beta_\ell \cdot \mathbb{I}(\text{HomeOwn}_i = \ell) \\
&+ \sum \beta_\ell \cdot \mathbb{I}(\text{Geo}_i = \ell),
\end{aligned}
$$

where i denotes a spell for individual j in year t during which the covariates do not change.

Pre-pandemic mortality levels

Lee-Carter baseline plus socio-economic factors

A closer look at the building blocks in the specification of μ_i:

$$\begin{aligned}
\log \mu_i &= f_{g_i}(x_i) + h_{g_i}(x_i) \cdot (\bar{t} - t_i) \\
&+ s_{\text{ME}}(\log \text{ME}_i) \cdot \mathbb{I}(\text{ME}_i > 0) + \beta_{\text{ME_0}} \cdot \mathbb{I}(\text{ME}_i = 0) \\
&+ s_{\text{Wealth}}(\text{Wealth}_{,i}) \cdot \mathbb{I}(\text{Wealth}_i \text{ not missing}) \\
&+ \sum \beta_\ell \cdot \mathbb{I}(\text{PersIncSrc}_i = \ell) + \sum \beta_\ell \cdot \mathbb{I}(\text{HomeOwn}_i = \ell) \\
&+ \sum \beta_\ell \cdot \mathbb{I}(\text{Geo}_i = \ell).
\end{aligned}$$

Based on pre-screening with single factor specifications (BIC + data availability): focus on: medical expenses + wealth (property value or income quantiles), source of personal income and home ownership + migration background.

Outline

Introduction

Research objectives and approach

CBS microdata

Modelling cause-specific, pre-pandemic mortality with socio-economic factors

- Methodology
- Results

Quantifying and analyzing (excess) mortality by cause of death during 2020-2021

- Methodology
- Results

RCLR

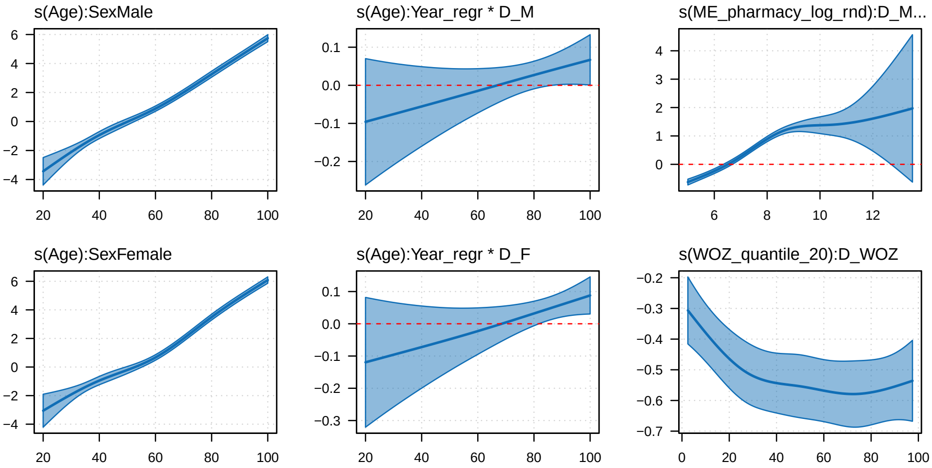
Example of model selection results

Alternatives considered: seven for Model 2 and nine for Model 3

Model for Diabetes	logL	EDF	BIC	ΔlogL	ΔEDF	ΔBIC
Model 1 (base)	-88.411	14,3	177.086			
Model 2 (pharmacy)	-83.840	19,5	168.041	4.571	5,2	-9.046
Model 2 (hospital)	-87.046	19,1	174.446	1.365	4,8	-2.641
...	...	...	...	...	...	...
Model 2 (chronic)	-85.682	18,9	171.715	2.729	4,7	-5.371
Model 2 (total expenses)	-84.954	20,3	170.283	3.457	6,0	-6.804
Model 2 (pharmacy)	-83.840	19,5	168.041			
Model 3 (property value)	-82.468	23,8	165.376	1.372	4,3	-2.665
Model 3 (property value + income source)	-82.088	27,4	164.684	1.752	7,9	-3.357
Model 3 (household income)	-82.403	23,5	165.241	1.437	4,0	-2.800
...	...	...	...	...	...	...
Model 3 (personal income by percentile and source)	-82.432	43,1	165.661	1.408	23,6	-2.380
Model 3 (property value + income source)	-82.088	27,4	164.684			
Model 4 (home ownership)	-81.720	32,5	164.040	368	5,1	-643
Model 4 (home ownership)	-81.720	32,5	164.040			
Model 5 (migration background)	-81.645	41,5	164.058	74	9,0	18

Example of calibrated risk factors

Other cardiovascular diseases



RCLR

Outline

Introduction

Research objectives and approach

CBS microdata

Modelling cause-specific, pre-pandemic mortality with socio-economic factors

- Methodology
- Results

Quantifying and analyzing (excess) mortality by cause of death during 2020-2021

- Methodology
- Results

Outline

Introduction

Research objectives and approach

CBS microdata

Modelling cause-specific, pre-pandemic mortality with socio-economic factors

- Methodology
- Results

Quantifying and analyzing (excess) mortality by cause of death during 2020-2021

- Methodology
- Results

From expected annual to weekly mortality

What we have at this point:

- the expected number of deaths (total or per cause) over an exposure period $\tau \in [0, 1]$ for individual j with characteristics i

  $$\tau \cdot \hat{\mu}_{tji} \text{ or } \tau \cdot \hat{\mu}^{c}_{tji} \text{ with } t \in \{2020, 2021\},$$

  where only the baseline Lee-Carter structure (t, x and g) or the LC structure + selected socio-economic factors can be included

However, this approach ignores the seasonality observed in weekly, empirical data on death counts (cfr. supra)!

RCLR

From expected annual to weekly mortality

Strategy to incorporate seasonal effect

Inspired by van Berkum et al., 2025 and Koninklijk Actuarieel Genootschap, 2022 we calibrate:

- week effects $\phi_{w,x,g}$ (all cause) and $\phi^{c}_{w,x,g}$ (cause specific) using cyclical cubic splines
- for males and females separately
- for all ages, but also for specific age groups $20-54$, $55-64$, $65-74$, $75-84$ and $85+$ separately
- on the prepandemic period 2016-2019, thus: no calibrated seasonal effect for COVID-19 CoD!

RCLR

Quantifying excess mortality in 2020 and 2021

As such, we obtain the expected number of deaths for every week w in $t = \{2020, 2021\}$, assuming pre-pandemic conditions:

$$\begin{array}{ll} \text{all cause} & \hat{d}_{t,w,j,i} = \tau_{t,w} \cdot \hat{\mu}_{tji} \cdot \hat{\phi}_{w,x,g} \\ \text{cause specific} & \hat{d}^{c}_{t,w,j,i} := \tau_{t,w} \cdot \hat{\mu}^{c}_{tji} \cdot \hat{\phi}^{c}_{w,x,g} \end{array}$$

We compare expected and observed weekly death counts in order to quantify excess mortality.

We quantify excess all-cause mortality on an annual basis (and similarly for individual causes) as:

- absolute excess mortality : $\sum_{w \in t,j,i} \{d^{\text{obs}}_{t,w,j,i} - \hat{d}_{t,w,j,i}\}$
- relative excess mortality : $\sum_{w \in t,j,i} d^{\text{obs}}_{t,w,j,i} / \sum_{w \in t,j,i} \hat{d}_{t,w,j,i} - 1$.

RCLR

Analyzing excess mortality in 2020 and 2021

Strategy

To dig deeper in the differences between expected and observed deaths during the pandemic, we explore

$$D_{t,w,j,i} \sim \text{POI}(\quad . \quad)$$

with a specification for the mean that is built up from

$$\underbrace{\tau_{t,w} \cdot \hat{\mu}_{tji} \cdot \hat{\phi}_{w,x}}_{\text{pre-pandemic}} \cdot \underbrace{\exp(\quad . \quad)}_{\text{socio-econ info}} \cdot \underbrace{\exp(\quad . \quad)}_{\text{infection + vacc info}},$$

where

- further adjustments based on socio-economic factors and information on COVID-19 infections and vaccinations.

RCLR

Outline

Introduction

Research objectives and approach

CBS microdata

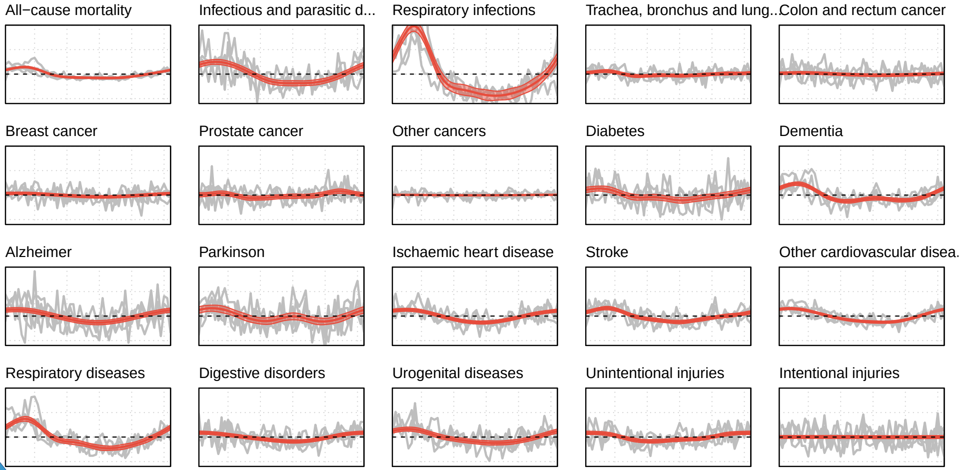
Modelling cause-specific, pre-pandemic mortality with socio-economic factors
- Methodology
- Results

Quantifying and analyzing (excess) mortality by cause of death during 2020-2021
- Methodology
- Results

RCLR

Quantifying excess mortality

Empirical and calibrated seasonal patterns



Quantifying excess mortality

Disclaimer on results

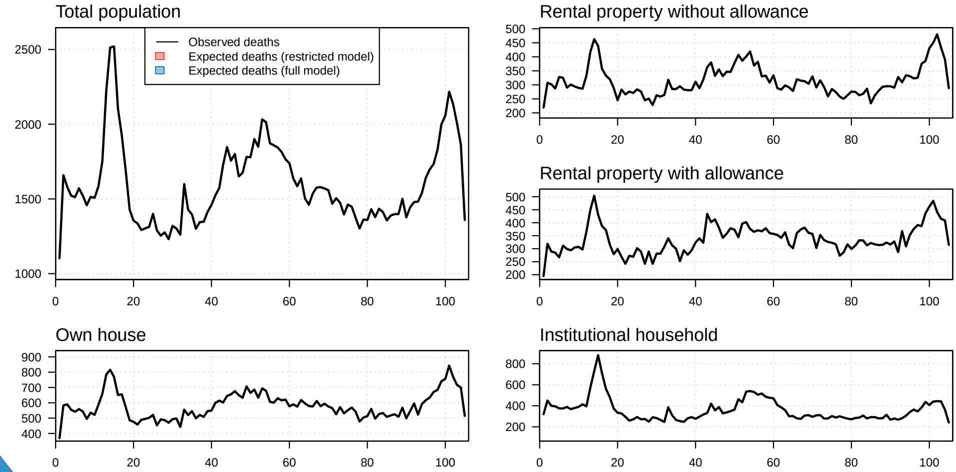
- The results for this phase are based on calibrations using 50% of the Dutch population because of computational challenges (even when using OSSC, the ODISSEI Secure Super Computer).

- The results on the following slides are preliminary; we have not yet been able to complete the analyses for all causes and for the separate COVID-waves.

RCLR

Quantifying excess mortality

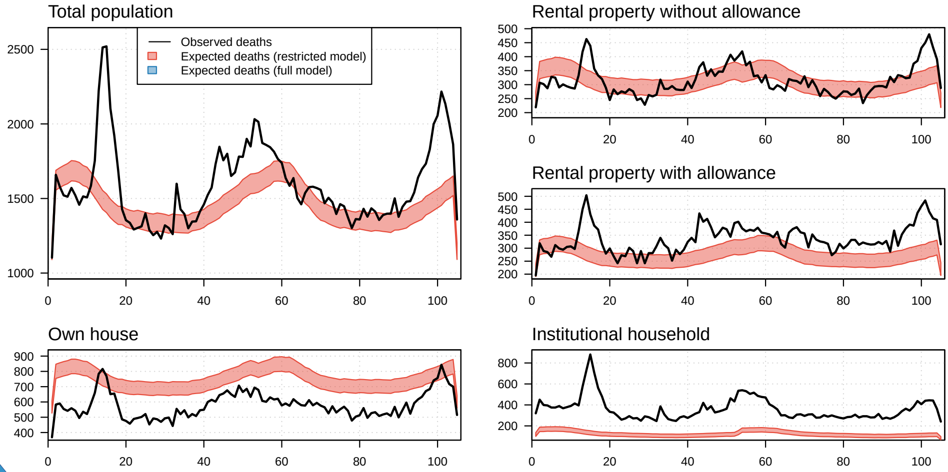
Illustrating the importance of a good baseline



RCLR

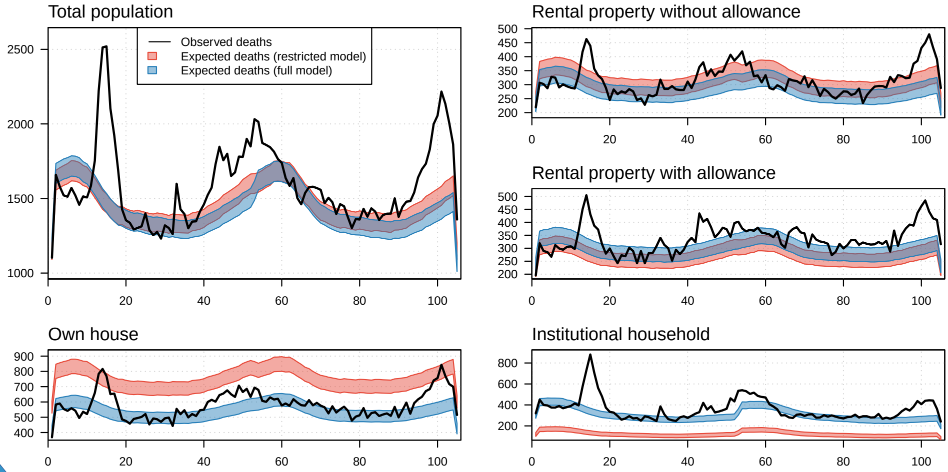
Quantifying excess mortality

Illustrating the importance of a good baseline



Quantifying excess mortality

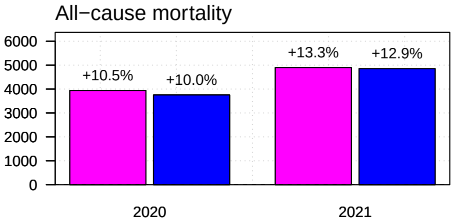
Illustrating the importance of a good baseline



RCLR

Quantifying excess mortality

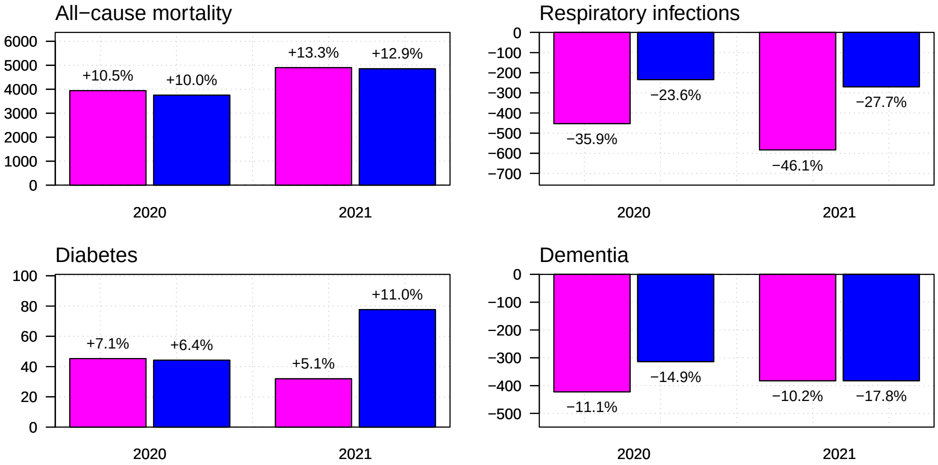
Excess mortality by cause (absolute and relative)



RCLR

Quantifying excess mortality

Excess mortality by cause (absolute and relative)



RCLR

Analyzing excess mortality

Diabetes regression results (baseline mortality based on restricted model)

Model	logL	EDF	BIC	Delta_logL	Delta_EDF	Delta_BIC
Baseline mortality (restricted model)	-36.726	1,0	73.473,1			
Baseline + home ownership	-36.150	7,0	72.443,0	576,3	6,0	-1.030,1
Baseline + property value + income source	-36.057	12,4	72.367,6	669,3	11,4	-1.105,5
Baseline + household income + income source	-36.104	12,2	72.457,7	622,1	11,2	-1.015,4
Baseline + personal income + income source	-36.147	16,6	72.632,4	579,2	15,6	-840,7
Baseline + medical expenses (pharmacy)	-35.197	7,0	70.537,1	1.529,0	6,0	-2.936,0
Baseline + migration background	-36.626	11,0	73.476,7	100,3	10,0	3,6
Baseline + COVID-19 infection	-36.710	4,1	73.503,7	16,5	3,1	30,6
Baseline + COVID-19 vaccination	-36.642	6,9	73.424,9	84,3	5,9	-48,2
Baseline + COVID-19 infection + vaccination	-36.652	5,9	73.424,8	74,5	4,9	-48,3

Table: logL = log-likelihood, EDF = number of parameters, BIC: lower is better

- Medical expenses seem to be the key predictor in explaining excess mortality
- COVID information on infections and testing seems to be of little importance

RCLR

Analyzing excess mortality

Diabetes regression results (baseline mortality based on full model)

Model	logL	EDF	BIC	Delta_logL	Delta_EDF	Delta_BIC
Baseline mortality (full)	-34.567	1,0	69.153,8			
Baseline + home ownership	-34.556	7,0	69.255,7	10,3	6,0	101,9
Baseline + property value + income source	-34.553	8,0	69.268,8	13,9	7,0	115,1
Baseline + household income + income source	-34.530	10,5	69.275,3	36,4	9,5	121,5
Baseline + personal income + income source	-34.529	13,5	69.333,8	37,2	12,5	180,0
Baseline + medical expenses (pharmacy)	-34.549	5,0	69.200,6	17,8	4,0	46,9
Baseline + migration background	-34.525	11,0	69.274,8	41,5	10,0	121,0
Baseline + COVID-19 infection	-34.559	3,0	69.178,4	8,1	2,0	24,6
Baseline + COVID-19 vaccination	-34.501	5,6	69.118,0	65,2	4,6	-35,7
Baseline + COVID-19 infection + vaccination	-34.506	5,4	69.122,0	61,0	4,4	-31,8

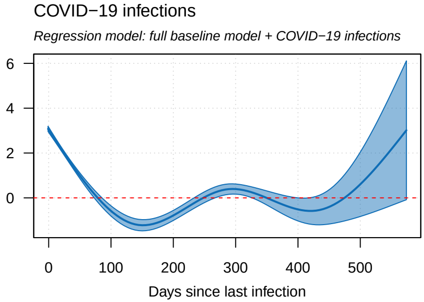
Table: logL = log-likelihood, EDF = number of parameters, BIC: lower is better

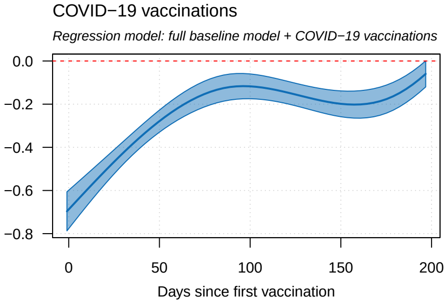
- When using a more refined baseline model, all socio-economic risk factors are no longer relevant in explaining excess mortality

RCLR

Analyzing excess mortality

All-cause mortality, estimated COVID-effects





- COVID-19 infections: increased risk of mortality immediately after infection, protection after some time but this protection wears off
- COVID-19 vaccinations: decreased risk of mortality immediately after vaccination, protection wears off over time

Key findings

- Mortality occurs most often at higher ages and risk factors that are available for older people are therefore likely to be more relevant
- Information on wealth / income and medical expenses are key in predicting the level of individuals' mortality
- Though on all-cause mortality there is substantial excess mortality in the period 2020-2021, for many individual causes observed mortality was lower than predicted (in line with previous CBS publications)
- When an appropriate baseline is used to define expected mortality, there seem to be limited differences in excess mortality between individuals with different socio-economic risk factors; COVID information is more relevant in explaining differences in excess mortality

RCLR

References

Natacha Brouhns, Michel Denuit, and Jeroen K. Vermunt. A poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and Economics*, 31(3):373–393, 2002.

Koninklijk Actuarieel Genootschap. Projections life table AG 2022, 2022.

Frank van Berkum, Katrien Antonio, and Michel Vellekoop. Quantifying Longevity Gaps Using Micro-Level Lifetime Data. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 184(2):548–570, 11 2020.

Frank van Berkum, Bertrand Melenberg, and Michel Vellekoop. Estimating the impact of COVID-19 on mortality using granular data. *Insurance: Mathematics and Economics*, 121:144–156, 2025.

WHO. WHO methods and data sources for country-level causes of death 2000-2019, 2020.

RCLR