



HAL
open science

Combining experience data of several Long-Term Care Insurance products with different disability definitions

Leonie Le Bastard, Stéphane Loisel, Adam W. Shao

► **To cite this version:**

Leonie Le Bastard, Stéphane Loisel, Adam W. Shao. Combining experience data of several Long-Term Care Insurance products with different disability definitions. 2023. hal-04333928v2

HAL Id: hal-04333928

<https://cnrs.hal.science/hal-04333928v2>

Preprint submitted on 18 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Combining experience data of several Long-Term Care Insurance products with different disability definitions

Léonie Le Bastard^{1,2}, Stéphane Loisel¹, and Adam W. Shao²

¹Université de Lyon, Université Claude Bernard Lyon 1, Laboratoire de Sciences Actuarielle et Financière, 50 Avenue Tony Garnier, F-69007 Lyon, France

²SCOR SE, 5 avenue Kléber, 75795 Paris Cedex 16, France

2024/02/18

Abstract

Long-term care (LTC) products cover the risk of permanent loss of autonomy. While the global definition of the loss of autonomy is the impossibility or difficulty of performing activities of daily living (ADL) alone, in the LTC insurance market, the exact definition of the health state leading to a claim varies across different markets and even within the same market. A difference in the disability definition implies a difference in the mortality rates of the autonomous and disabled policyholders. Insurers or reinsurers often have experience data coming from several long-term care products with differing definitions of risk. One solution is to separate the data to estimate mortality rates for each definition independently.

In this paper, we propose two methods to aggregate the experience data of two portfolios with different disability definitions to improve the estimations of the mortality. The mortality laws of the two products are modelled in a Poisson Generalized Linear Model framework. The first method uses a constrained optimization model and is solved by sequential quadratic programming. The second method uses the Penalized Composite Link Model (PCLM). These methods allow better and simultaneous estimation of mortality for both products by combining all available data.

Keywords: Long-Term Care Insurance; Actuarial modelling; Multiple definitions; Penalized Composite Link Model; Constrained optimization.

1 Introduction

Long-term care (LTC) costs are among the major risks faced by individuals in retirement. For example, Chapman (2012) reports that approximately two-thirds of people aged 65+ need LTC. By using the U.S. Health and Retirement Study data, Johnson (2019) estimates a large gap between the risk of needing vs. receiving LTC. As reported in Johnson (2019), 70% of individuals past age 65 need LTC, whereas less than 50% receive some form of LTC services. Similar levels of risk of needing LTC have been reported in many other studies (Shao et al., 2017; Kemper et al., 2005; Hurd et al., 2017).

LTC costs have been steadily increasing over the past decades in many countries, and this trend is projected to continue (Colombo et al., 2011; Shi and Zhang, 2013; H et al., 2023). A recent report by OECD (OECD, 2020) shows that on average, LTC costs were approximately USD 760 per capita in OECD countries in 2018, accounting for 1.5% of GDP. In some countries, such as Denmark, Norway and Sweden, LTC spending was as high as 3.5% of their GDP. In France, LTC spending was 2.5% of its GDP.

The fundamental source of funding for LTC costs in most countries is public programs. For example, in the U.S. LTC expenses are mainly funded by its public health program Medicaid (Colombo et al., 2008; Kaye et al., 2010); in France, LTC expenses are mainly funded by its national allowance program called Allocation Personnalisée d'Autonomie (APA) with a cap (Or and Penneau, 2021). Coupled with ageing populations, the increasing trend of LTC costs may place a large burden on public health programs in many countries. It has become very important in many countries to develop or to enhance the private insurance market to help fund LTC costs (Shao et al., 2019; Colombo et al., 2011; Productivity Commission of Australia, 2013).

A typical private LTC insurance policy entitles the policyholder to regular payments, such as cash benefits or on a reimbursement basis, when the policyholder loses autonomy according a certain definition (Haberman and Pitacco, 1999; Shao et al., 2017). The definitions of “losing autonomy” or “becoming disabled” vary across different markets, and these definitions are not uniform even within the same market.

In the U.S. market, the disability definition in a typical LTC insurance policy is the loss of independence in performing two or more activities of daily living (ADLs) and/or having cognitive impairment (Pritchard, 2006). In the French market, the disability definition is mostly based on the GIR (Groupe Iso-Ressources) assessment rules. Within the French LTC insurance market, different disability definitions can be used in different insurance policies. For example, some LTC insurance policies include a deferred period (such as 3 months, 6 months, or 9 months), and some start making payments at the date of the loss of autonomy.

Differing disability definitions may result in very different transition rates between health states (including disability rate and mortality rate) in the pricing model and can therefore have substantially different financial impacts. An insurer or reinsurer may have experience data for many different

portfolios of LTC insurance policies with differing disability definitions, where they typically analyse experience for each portfolio of LTC insurance policies. This separate modelling approach can result in information loss from a parameter estimation perspective. For example, two portfolios with different definitions, from the same insurer with similar claims management team and system, should have shared experience that can provide better estimations. This combined modelling approach can also help experience analysis for portfolios with limited experience or areas with limited experience data.

Let us consider an insured population such that the health state of each policyholder is observed with only one of the two definitions. Instead of separating the policyholders in two portfolios depending on the definition with which they are observed, we jointly estimate the mortality rates of each product to prevent information loss.

In this paper, we develop two methods that make the best use of data available by combining experience data of the two portfolios with different disability definitions. These two methods are the constrained optimization method and the Penalized Composite Link Model (PCLM) method. A common approach to estimate the parameters of a model is the maximum likelihood estimation. The first method proposed in this paper estimates parameters by maximizing the likelihood subject to some constraints linking biometric laws of the two portfolios. The second approach proposed in this paper uses the Penalized Composite Link Model introduced by Eilers (2007) as an extension of the Composite Link Model proposed by Thompson and Baker (1981). The Penalized Composite Link Model has already been used in the context of mortality modelling in Remund et al. (2018) for cause-of-death decomposition and in Camarda et al. (2016) by considering that the mortality curve is a sum of 3 smooth components. In the latter paper, the first component represents infant mortality, the second component captures mortality due to ageing, and the third component models the accident hump for early adult ages. The constrained optimization model and the Penalized Composite Link Model (PCLM) developed in this paper are embedded in the GLM framework. Finally, the two models are also compared with the separate modelling approach to show the benefits gained by combining experience data.

The two models developed in this paper use the P-splines smoothing method to avoid overfitting. This smoothing method introduced by Eilers and Marx (1996) is based on the idea of penalizing models with large variability between coefficients of adjacent splines.

The remainder of this paper is arranged as follows. Section 2 introduces the research question and modelling framework. In Section 3, we provide technical details on estimation methods, including P-splines smoothing, constrained optimization model, and the Penalized Composite Link Model (PCLM). These models are applied to a real-life problem in Section 4; in particular, the application is for the estimation results of two portfolios with and without the deferred period. Section 5 concludes.

2 Modelling of the multidefinition problem

In this paper, long-term care products are modelled by using the semi-Markov framework, where mortality in disability states depends on both attained age and time already spent in the disability state.

Let us consider two disability definitions, and let T_k denote the LTC state associated with definition $k \in \{1, 2\}$. Let $X_x^{(k)}$ denote the health status of a policyholder observed with definition k . The three possible health statuses in a long-term care insurance product are healthy, disabled and dead.

Let Ω_k denote the set of possible health statuses of a policyholder observed with disability definition k , then

$$\Omega_k = \{H_k, T_k, Death\},$$

where H_k denotes the healthy state of definition k .

Let F_α and F_β denote two levels of disability such that $F_\alpha \cap F_\beta = \emptyset$.

The two disability definitions are defined as follows:

- $T_1 = F_\alpha \cup F_\beta$, and
- $T_2 = F_\beta$,

such that T_2 is included in T_1 . Being disabled with definition 2 implies disability with definition 1.

By construction, the autonomous state of the second definition is $H_2 = H_1 \cup F_\alpha$. A policyholder being disabled with definition 1 is either in health state F_α or F_β . In this paper, we consider that for each disabled policyholder observed with definition 1 (i.e. $X_x^{(1)} = T_1$), its refined LTC state F_α or F_β is known.

Let us assume the following:

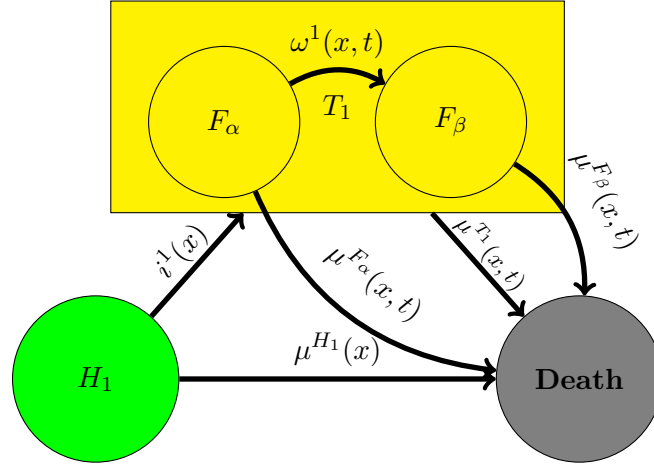
Assumption. 2.1

No return to a better health state is envisaged, i.e., recovery is assumed to be impossible.

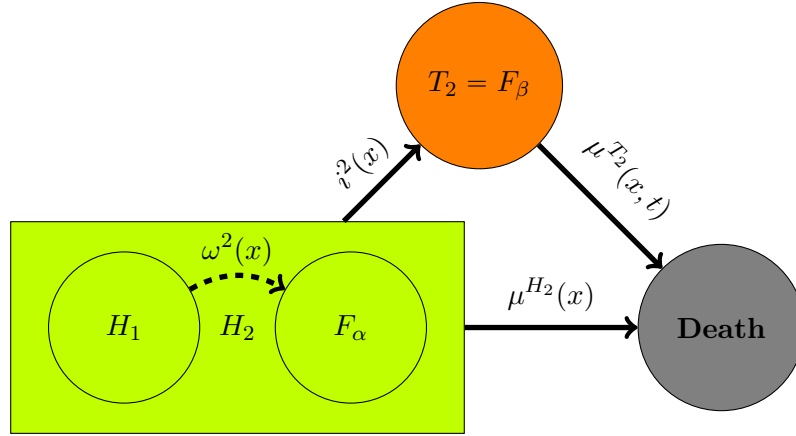
We then have transition diagrams for each of the two types of LTC insurance, as represented in Figure 1, where $i^k(x)$, $\mu^{H_k}(x)$ with $k \in \{1, 2\}$ and $\mu^{F_g}(x, t)$ with $g \in \{\alpha, \beta\}$ are the transition intensities.

The transition intensities are given by

- $i^k(x) = \lim_{h \rightarrow 0} \frac{\mathbb{P}(X_{x+h}^{(k)} = T_k | X_x^{(k)} = H_k)}{h}$,
- $\omega^2(x) = \lim_{h \rightarrow 0} \frac{\mathbb{P}(X_{x+h}^{(2)} = F_\alpha | X_x^{(2)} = H_1)}{h}$,



(a) Definition 1



(b) Definition 2

Figure 1: Modelling of the long-term care product with two disability definitions

- $\mu^{H_k}(x) = \lim_{h \rightarrow 0} \frac{\mathbb{P}(X_{x+h}^{(k)} = \text{Death} | X_x^{(k)} = H_k)}{h}$,
- $\mu^{F_g}(x, t) = \lim_{h \rightarrow 0} \frac{\mathbb{P}(X_{x+h}^{(1)} = \text{Death} | X_x^{(1)} = F_g, X_{(x-t)}^{(1)} = F_g, X_{(x-t)-}^{(1)} \neq F_g)}{h}$, and
- $\mu^{T_k}(x, t) = \lim_{h \rightarrow 0} \frac{\mathbb{P}(X_{x+h}^{(k)} = \text{Death} | X_x^{(k)} = T_k, X_{x-t}^{(k)} = T_k, X_{(x-t)-}^{(k)} = H_k)}{h}$.

The dashed arrow represents a transition that we cannot observe with disability definition 2. Transitions from F_α to F_β are assumed to be observed for individuals who are disabled according to disability definition 1.

Therefore, $\mu^{F_g}(x, t)$ denotes the mortality of a disabled of age x , who is in state F_g since t years, and $\mu^{T_k}(x, t)$ denotes the mortality of an individual of age x , disabled since t years according to the

definition k .

Each disability definition can be associated with a different long-term care product. In the remainder of the paper, the terms "product" and "disability definition" are equivalent.

We consider the following assumptions:

Assumption. 2.2

The policyholders are homogeneous and independent under the two types of insurance.

Assumption. 2.3

The interval of ages can be divided into subintervals such that mortality and incidence rates from healthy states are constant in each interval. The split points are denoted $\{x_1^H, x_2^H, \dots, x_{M^H}^H\}$, where M^H denotes the number of intervals of ages for healthy states.

Assumption. 2.4

The interval of ages and durations can be divided into subintervals such that mortality rates in states $F_g, g \in \{\alpha, \beta\}$ are constant in each interval. The split points of the age and duration dimensions are denoted $\{x_1^F, x_2^F, \dots, x_{M_x^F}^F\}$ and $\{t_1^g = 0, t_2^g, \dots, t_{M_t^g}^g\}$, respectively.

$$\mu^{F_g}(x, t) = \mu^{F_g}(x_p^F, t_q^g), \forall t | t_q^g \leq t < t_{q+1}^g, \forall x | x_p^F \leq x < x_{p+1}^F.$$

M_x^F and M_t^g denote the number of subdivisions of the intervals of ages and durations for the disabled state F_g , respectively.

The subdivision of ages can be different for healthy and disabled states. However, it is assumed in the following that the set of ages $\{x_1^F, x_2^F, \dots, x_{M_x^F}^F\}$ is included in $\{x_1^H, x_2^H, \dots, x_{M^H}^H\}$. Therefore, the subdivision of ages in the healthy state is a more refined subdivision of the interval of ages.

Let us note that the subdivision of the interval of ages is the same for the disabled states F_α and F_β , whereas the subdivision of the duration interval can differ.

As intensities $\mu^{H_k}()$ and $i^k()$, $k \in \{1, 2\}$ and $\mu^{F_g}()$, $g \in \{\alpha, \beta\}$ are piecewise constants (Assumption 2.3 and Assumption 2.4), we denote

- $\mu_p^{H_k} = \mu^{H_k}(x_p^H), \forall p \in \{1, \dots, M^H\}$,
- $i_p^k = i^k(x_p^H), \forall p \in \{1, \dots, M^H\}$,
- $\mu_{p,q}^{F_g} = \mu^{F_g}(x_p^F, t_q^g), \forall p \in \{1, \dots, M_x^F\}, q \in \{1, \dots, M_t^g\}$.

The total likelihood for all individuals from products 1 and 2 is given by

$$\begin{aligned}
 L_{tot} = & \prod_{p=1}^{M^H} \exp \left(- (i_p^1 + \mu_p^{H_1}) e^{H_1}(p) \right) (i_p^1)^{N^1(p)} (\mu_p^{H_1})^{D^{H_1}(p)} \\
 & \prod_{p=1}^{M_x^F} \prod_{q=1}^{M_t^\alpha} \exp \left(- \mu_{p,q}^{F_\alpha} e^\alpha(p, q) \right) (\mu_{p,q}^{F_\alpha})^{D^\alpha(p,q)} \times \\
 & \prod_{p=1}^{M^H} \exp \left(- (i_p^2 + \mu_p^{H_2}) e^{H_2}(p) \right) (i_p^2)^{N^2(p)} (\mu_p^{H_2})^{D^{H_2}(p)} \\
 & \prod_{p=1}^{M_x^F} \prod_{q=1}^{M_t^\beta} \exp \left(- \mu_{p,q}^{F_\beta} e^\beta(p, q) \right) (\mu_{p,q}^{F_\beta})^{D^\beta(p,q)}, \tag{1}
 \end{aligned}$$

where:

- $e^{H_k}(p)$ is the sum of the central exposures in state H_k of all policyholders from product k between integer ages x_p^H and x_{p+1}^H ,
- $e^\beta(p, q)$ is the sum of the central exposures of all disabled policyholders from both products between integer ages x_p^T and x_{p+1}^T and durations t_q^g and t_{q+1}^g ,
- $D^{H_1}(p)$ denotes the number of deaths from the healthy state between x_p^H and x_{p+1}^H in the first type of insurance,
- $D^{H_2}(p)$ denotes the number of deaths from the healthy state between x_p^H and x_{p+1}^H in the second type of insurance,
- $D^\alpha(p, q)$ denotes the number of deaths from the F_α disabled state with age at death between x_p^T and x_{p+1}^T and duration between t_q^α and t_{q+1}^α ,
- $D^{\beta,k}(p, q)$ denotes the number of deaths from the F_β disabled state for product k ,
- $D^\beta(p, q)$ is the total number of deaths of disabled policyholders with age at death between x_p^T and x_{p+1}^T and duration between t_q^β and t_{q+1}^β from both products.
- $N^1(p)$ denotes the number of transitions from H_1 to the disabled state T_1 between x_p^H and x_{p+1}^H , observed in product 1,
- $N^2(p)$ denotes the number of transitions from H_2 to the disabled state T_2 between x_p^H and x_{p+1}^H , observed in product 2.

The detailed proof of Equation 1 is given in Appendix A.

The last term of Equation 1 represents the likelihood of all observations from products 1 and 2 in state F_β .

The log likelihood function of the combined observations from products 1 and 2 is given by

$$\begin{aligned}
 l_{tot} = & \sum_{p=1}^{M^H} \left(- (i_p^1 + \mu_p^{H_1}) e^{H_1(p)} + N^1(p) \log (i_p^1) + D^{H_1}(p) \log (\mu_p^{H_1}) + \right. \\
 & \sum_{p=1}^{M_x^F} \sum_{q=1}^{M_t^\alpha} \left(-\mu_{p,q}^{F_\alpha} e^\alpha(p, q) + D^\alpha(p, q) \log (\mu_{p,q}^{F_\alpha}) + \right. \\
 & \sum_{p=1}^{M^H} \left(- (i_p^2 + \mu_p^{H_2}) e^{H_2(p)} + N^2(p) \log (i_p^2) + D^{H_2}(p) \log (\mu_p^{H_2}) + \right. \\
 & \left. \left. \sum_{p=1}^{M_x^F} \sum_{q=1}^{M_t^\beta} \left(-\mu_{p,q}^{F_\beta} e^\beta(p, q) + D^\beta(p, q) \log (\mu_{p,q}^{F_\beta}) \right) \right) \right). \tag{2}
 \end{aligned}$$

Equation 2 is equivalent to a sum of log-likelihoods of the Poisson distribution, where:

- $N^1(p) \sim \text{Poisson} (i_p^1 e^{H_1(p)})$,
- $N^2(p) \sim \text{Poisson} (i_p^2 e^{H_2(p)})$,
- $D^{H_1}(p) \sim \text{Poisson} (\mu_p^{H_1} e^{H_1(p)})$,
- $D^{H_2}(p) \sim \text{Poisson} (\mu_p^{H_2} e^{H_2(p)})$,
- $D^\alpha(p, q) \sim \text{Poisson} (\mu_{p,q}^{F_\alpha} e^\alpha(p, q))$, and
- $D^\beta(p, q) \sim \text{Poisson} (\mu_{p,q}^{F_\beta} e^\beta(p, q))$.

Therefore, we assume in the following that the counts of deaths and the counts of losses of autonomy have a Poisson distribution.

Moreover, Equation 2 shows that each transition rate can be estimated separately and independently. As we focus this research on mortality rates, incidence rates are considered constants in Equation 2. The problem is therefore simplified to

$$\max_{\mu^{H_1}, \mu^{H_2}, \mu^{F_\alpha}, \mu^{F_\beta}} l_{tot},$$

where l_{tot} is given by

$$\begin{aligned}
 l_{tot} = & \underbrace{\sum_{p=1}^{M^H} (-\mu_p^{H_1} e^{H_1(p)} + D^{H_1}(p) \log (\mu_p^{H_1}))}_{l^{H_1}} + \underbrace{\sum_{p=1}^{M_x^F} \sum_{q=1}^{M_t^\alpha} (-\mu_{p,q}^{F_\alpha} e^\alpha(p, q) + D^\alpha(p, q) \log (\mu_{p,q}^{F_\alpha}))}_{l^{F_\alpha}} + \\
 & \underbrace{\sum_{p=1}^{M^H} (-\mu_p^{H_2} e^{H_2(p)} + D^{H_2}(p) \log (\mu_p^{H_2}))}_{l^{F_\beta}} + \underbrace{\sum_{p=1}^{M_x^F} \sum_{q=1}^{M_t^\beta} (-\mu_{p,q}^{F_\beta} e^\beta(p, q) + D^\beta(p, q) \log (\mu_{p,q}^{F_\beta}))}_{l^{H_2}}. \tag{3}
 \end{aligned}$$

3 Methods

In this section, we introduce the two proposed methods to make better use of all the available information from observations with different disability definitions. In the first part of this section, we focus on the P-splines smoothing framework that is used in the two methods to prevent overfitting and to produce smooth estimated mortality laws. For the second part, we present the first method that uses the constrained optimization algorithm. The third and final part of this section is devoted to the presentation of the Penalized Composite Link Model, corresponding to the second method proposed in this paper. While the first method assumes that the mortality intensities in each state H_1 , H_2 , F_α and F_β can be expressed as the exponential of a combination of basis-splines, the second method uses the same assumption only on the three following states H_1 , F_α and F_β . No assumption on the shape of the mortality in H_2 is made in the second model, leading to fewer coefficients to estimate.

3.1 P-splines smoothing framework

Common parametric smoothing methods are introduced by Gompertz and Gompertz-Makeham, assuming that mortality increases exponentially with age, or as in Perks (1932), Beard (1959) allowing a deceleration of the mortality increase at old ages. Unlike these models, the P-splines smoothing method is a nonparametric smoothing technique. One of the strengths of these techniques is that they do not assume any particular shape of the mortality function.

A spline is a piecewise polynomial function that is continuous and has continuous derivatives up to a certain order. This type of function is commonly used for smoothing problems. The P-splines smoothing method, introduced in Eilers and Marx (1996), uses a B-spline basis with penalties to prevent overfitting. This smoothing method, also described in Marx and Eilers (1998), is applied to mortality estimation by using the Poisson-GLM framework in Currie and Durban (2002) and Macdonald et al. (2018).

First, let us consider the case of a one-dimensional mortality law, as the mortality in H_1 depends only on age. J^H splines are uniformly positioned over the entire interval of ages. Splines are therefore equidistant. An important property of the spline basis is that the sum of the B-splines equals 1 at any point on the support. Let $B^{H_1} \in \mathcal{M}_{M^H, J^{H_1}}$ denote the matrix of the spline basis for estimation of mortality in H_1 , where each column corresponds to a spline, and each row corresponds to an observation age ($x_p^H, p \in \{1, \dots, M^H\}$ in the case of mortality in healthy states). In the context of mortality estimation, the P-splines smoothing method requires the following assumption:

Assumption. 3.1

Mortality rates can be expressed as the exponential of a combination of basis-splines.

Therefore,

$$\mu^{H_1} = \exp(B^{H_1}\theta^{H_1}), \text{ where } \theta^{H_1} \in \mathbb{R}^{J^{H_1}}. \quad (4)$$

A term $\frac{1}{2}\rho\|D^{H_1}\theta^{H_1}\|_2^2$ is then added to the log-likelihood to penalize complex models with large variability between coefficients of adjacent splines. For simplicity, the penalty term can be written as

$$\frac{1}{2}(\theta^{H_1})^T P^{H_1}\theta^{H_1}, \quad (5)$$

where $P^{H_1} = \rho(D^{H_1})^T D^{H_1}$.

The problem is therefore

$$\max_{\theta^{H_1}} l^{H_1}(\theta^{H_1}) - \frac{1}{2}(\theta^{H_1})^T P^{H_1}\theta^{H_1}. \quad (6)$$

The output of this smoothing method depends on multiple hyperparameters listed below:

1. The number of splines distributed in the interval,
2. The degree of the splines,
3. The order of the penalty, denoted d ,
4. The smoothing parameter ρ is the weight of the penalty in the penalized log-likelihood .

As the extrapolation of the fitted mortality is mostly driven by the form of the penalty matrix, the choice of the order d and the smoothing weight ρ have a significant impact. In contrast, the number of splines and their degree are less critical, and have a rather limited impact on the fitted mortality laws. Ruppert (2002) and Eilers (2007) suggest that the following choice is often sufficient for these two parameters:

- Use cubic splines (degree 3),
- Fix a knot every 4 or 5 observations.

As recommended in Currie and Durban (2002), parameters ρ and d are often chosen to minimize the Bayesian Information Criterion (BIC), as defined in Schwarz (1978), given by

$$BIC = -2l^{H_1}(\theta^{H_1}) + \log(M^H) \times df,$$

where df is the degree of freedom of the model.

The one-dimensional P-splines smoothing method can be generalized to two-dimensional smoothing problems to consider the duration in the context of estimating mortality in disabled states. To this aim, two matrices of basis splines are needed. Let $g \in \{\alpha, \beta\}$. For each state F_g , let $B_x^g \in \mathcal{M}_{M_x^F, J_x^g}$ and $B_t^g \in \mathcal{M}_{M_t^g, J_t^g}$ denote the matrices of splines for the age and duration dimensions, respectively. J_t^g and J_x^g denote the number of splines for each dimension.

Notation: Let A be a matrix of dimensions $r \times c$, $A_{vec} = vec(A) = (A_{.1}^T, \dots, A_{.c}^T)^T$, where $A_{.k} \in \mathbb{R}^r$ is the k^{th} column of matrix A .

Then,

$$\mu_{vec}^{F_g} = \exp(B^g \theta^g), \quad (7)$$

with $B^g = B_t^g \otimes B_x^g \in \mathcal{M}_{M_x^F \times M_t^g, J_x^g \times J_t^g}$, where \otimes represents the Kronecker product.

The penalty matrix for two-dimensional P-splines smoothing is the sum of a penalty term on the age dimension and a penalty term on the duration dimension. The overall penalty matrix is given by the following equation:

$$P^g = (I_{J_t^g} \otimes P_x^g) + (P_t^g \otimes I_{J_x^g}) \in \mathcal{M}_{J_x^g \times J_t^g, J_x^g \times J_t^g}, \quad (8)$$

where P_x^g and P_t^g are the penalty matrix as described for the one-dimensional case and $I_{J_t^g}$ and $I_{J_x^g}$ are the identity matrices of dimensions J_x^g and J_t^g , respectively.

3.2 Optimization with constraint

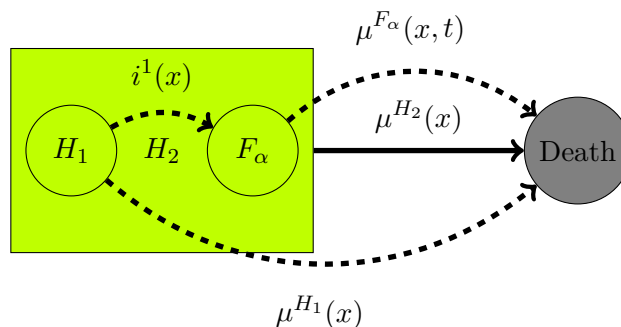


Figure 2: Zoom on transitions from state H_2 to Death

Figure 2 shows that $H_2 = H_1 \cup F_\alpha$. Therefore, at each age x , the mortality in state H_2 is a weighted average of the mortality in H_1 and the mortality in F_α . The 3 mortality laws μ^{H_2} , μ^{H_1} and μ^{F_α} , are linked by the following Equation (9)

$$\begin{aligned} \mu_p^{H_2} &= \lim_{h \rightarrow 0} \frac{\mathbb{P}(X_{x_p^H+h}^{(2)} = \text{Death} | X_{x_p^H}^{(2)} = H_2)}{h} \\ &= \frac{\mathbb{P}(X_x^{(1)} = H_1)}{\mathbb{P}(X_x^{(1)} = H_1 \cup X_x^{(1)} = F_\alpha)} \mu_p^{H_1} + \sum_{q=1}^{M_t^\alpha} \frac{\mathbb{P}(X_x^{(1)} = F_\alpha, X_{(x-t_q)}^{(1)} = F_\alpha, X_{(x-t_{q+1})}^{(1)} = H_1)}{\mathbb{P}(X_x^{(1)} = H_1 \cup X_x^{(1)} = F_\alpha)} \mu_{p,q}^{F_\alpha}. \end{aligned} \quad (9)$$

A detailed proof of Equation (9) is given in Appendix B.

By using the P-splines smoothing framework for each of the four mortality laws, the optimal

coefficients $\theta^{H_1}, \theta^{H_2}, \theta^\alpha$ and θ^β are obtained by solving the following constrained optimization problem

$$\begin{aligned} \max_{\theta^{H_1}, \theta^{H_2}, \theta^\alpha, \theta^\beta} \quad & l^{H_1}(\theta^{H_1}) - \frac{1}{2} (\theta^{H_1})^T P^{H_1} \theta^{H_1} + l^{H_2}(\theta^{H_2}) - \frac{1}{2} (\theta^{H_2})^T P^{H_2} \theta^{H_2} + \\ & l^\alpha(\theta^\alpha) - \frac{1}{2} (\theta^\alpha)^T P^\alpha \theta^\alpha + l^\beta(\theta^\beta) - \frac{1}{2} (\theta^\beta)^T P^\beta \theta^\beta, \end{aligned} \quad (10)$$

subject to

$$\begin{aligned} \mu_p^{H_2}(\theta^{H_2}) = & \frac{\mathbb{P}(X_x^{(1)} = H_1)}{\mathbb{P}(X_x^{(1)} = H_1 \cup X_x^{(1)} = F_\alpha)} \mu_p^{H_1}(\theta^{H_1}) + \\ & \sum_{q=1}^{M_t^\alpha} \frac{\mathbb{P}(X_x^{(1)} = F_\alpha, X_{(x-t_q^\alpha)}^{(1)} = F_\alpha, X_{(x-t_{q+1}^\alpha)}^{(1)} = H_1)}{\mathbb{P}(X_x^{(1)} = H_1 \cup X_x^{(1)} = F_\alpha)} \mu_{p,q}^{F_\alpha}(\theta^\alpha). \end{aligned} \quad (11)$$

The optimization problem is solved by using Sequential Quadratic Programming (SQP) as described in Kraft (1988) and Boggs and Tolle (1995). The SQP method is implemented on the statistical programming software **R** in the `slsqp()` function from the `nloptr` package. The SQP method replaces the original problem of optimization with a sequence of quadratic problems where the objectives are second-order approximations of the Lagrangian and the constraints are first-order approximations of the original constraints. One of the major advantages of the SQP methods is that the initial point does not need to satisfy all the constraints of the original problem. The documentation of the package can be found in Jelmer Ypma (2022).

3.3 Penalized Composite Link Model

The Penalized Composite Link Model was first proposed by Eilers (2007). This method is an extension of the Composite Link Model introduced by Thompson and Baker (1981), that uses the Generalized Linear Model framework (GLM) (Nelder and Wedderburn, 1972; McCullagh, 2019).

By assuming that the policyholders are homogeneous and independent under the two types of insurance, policyholders in state F_β from both products 1 and 2 can be aggregated.

As in Section 2, $D^{H_k} \in \mathbb{N}^{M^H}$ denotes the vector of counts of deaths in the healthy state H_k , $k \in \{1, 2\}$. $D^g \in \mathcal{M}_{M_x^F, M_t^g}(\mathbb{N})$ is the matrix of counts of deaths in state F_g . Since policyholders cannot be observed in state F_α with product 2, D^α is only composed of deaths of policyholders from product 1. However, since state F_β is observed for both products, $D^\beta \in \mathcal{M}_{M_x^F, M_t^\beta}(\mathbb{N})$ is the sum of the matrices of counts of deaths in F_β of the two products. $D_{i,j}^\beta$ denotes the number of deaths in F_β at age $x_i^F < x < x_{i+1}^F$ and duration $t_j^\beta < t < t_{j+1}^\beta$. Let D be the vector of counts of deaths such that $D = (D^{H_1}, D_{vec}^\alpha, D_{vec}^\beta, D^{H_2})^T \in \mathbb{N}^{2 \cdot M^H + M_x^F \cdot (M_t^\alpha + M_t^\beta)}$. Similar to the vector of counts of deaths, $e = (e^{H_1}, e_{vec}^\alpha, e_{vec}^\beta, e^{H_2})^T \in \mathbb{N}^{2 \cdot M^H + M_x^F \cdot (M_t^\alpha + M_t^\beta)}$ is the concatenation of the vectors of

central exposures in each state H_1, F_α, F_β , and H_2 .

Let us assume that the mortality rates in states H_1 and $F_g, g \in \{\alpha, \beta\}$ can be expressed as the exponential of a combination of basis splines as in Assumption 3.1. Let

- $\mu^{H_1} = (\mu^{H_1}(x_1^H), \dots, \mu^{H_1}(x_{M^H}^H))^T \in \mathbb{R}^{M^H}$, and
- $\mu_{vec}^{F_g} = vec(\mu^{F_g}) \in \mathbb{R}^{M_x^F \cdot M_t^g}$, where $\mu^{F_g} \in \mathcal{M}_{M_x^F, M_t^g}(\mathbb{R})$ is the matrix of the mortality intensities in state F_g such that

$$\mu_{p,q}^{F_g} = \mu^{F_g}(x_p^F, t_q^g) \forall p \in \{1, \dots, M_x^F\}, q \in \{1, \dots, M_t^g\}.$$

Then,

$$\mu^{H_1} = \exp(B^{H_1} \theta^{H_1}), \quad (12)$$

$$\mu_{vec}^{F_g} = \exp(B^g \theta^g), g \in \{\alpha, \beta\}, \quad (13)$$

where $B^{H_1} \in \mathcal{M}_{M^H, J^{H_1}}(\mathbb{R})$ and $B^g \in \mathcal{M}_{M_x^F \times M_t^g, J^g}(\mathbb{R})$ are the splines basis for states H_1 and F_g , respectively. J^{H_1} and $J^g, g \in \{\alpha, \beta\}$ denote the number of coefficients of splines needed to estimate the mortality laws in the healthy state and disabled states g , respectively. Using notations from Section 3.1, $J^g = J_x^g \times J_t^g$. Therefore, $\theta^{H_1} \in \mathbb{R}^{J^{H_1}}$ and $\theta^g \in \mathbb{R}^{J^g}$.

A difference with Section 3.2 is that no assumption on the mortality in state H_2 is made.

Based on Equation 2, it is assumed that the counts of deaths have a Poisson distribution with parameter $e \times \mu$, where e and μ denote the vectors of central exposures and transition intensities, respectively. This assumption is commonly used with Penalized Composite Link Models, as in Eilers (2007), Remund et al. (2017) or Rizzi et al. (2015). Therefore,

$$D^{\hat{H}_k} = \mathbb{E}[D^{H_k}] = e^{H_k} \times \mu^{H_k}, k \in \{1, 2\}, \quad (14)$$

$$\hat{D}^g = \mathbb{E}[D^g] = e^g \times \mu^{F_g}, g \in \{\alpha, \beta\}. \quad (15)$$

With observations from product 1, we can create a fictitious state H_2 composed of policyholders in H_1 and F_α from product 1. We denote H_2^{prod1} this fictitious state. Given Assumption 2.2, mortality in H_2^{prod1} is the same as in H_2 . Therefore,

$$\hat{D}^{H_2^{prod1}} = e^{H_2^{prod1}} \times \mu^{H_2}, \quad (16)$$

where $D^{H_2^{prod1}}$ and $e^{H_2^{prod1}}$ denote the vectors of counts of deaths and central exposures in the state $H_2^{prod1} = H_1 \cup F_\alpha$, respectively. The subdivision of the age interval of this state is the same as for states H_1 and H_2 .

Let $p \in \{1, \dots, M^H\}$. From Assumption 2.4, we know that the subdivision of the interval of ages in healthy states is either the same or more refined than is the subdivision of ages for the disabled

states. Moreover,

$$\forall p \in \{1, \dots, M^H\}, \exists \tilde{p} \in \{1, \dots, M_x^F\}, x_{\tilde{p}}^F \leq x_p^H < x_{\tilde{p}+1}^H \leq x_{\tilde{p}+1}^F.$$

Therefore, if not directly calculable from the database, the central exposure in state F_α between ages x_p^H and x_{p+1}^H (subdivision of the interval of ages used in states H_k) for duration between t_q^α and t_{q+1}^α can be estimated by

$$e_{sub_H}^\alpha(p, q) = e^\alpha(\tilde{p}, q) \frac{x_{p+1}^H - x_p^H}{x_{\tilde{p}+1}^F - x_{\tilde{p}}^F},$$

as illustrated in Figure 3.

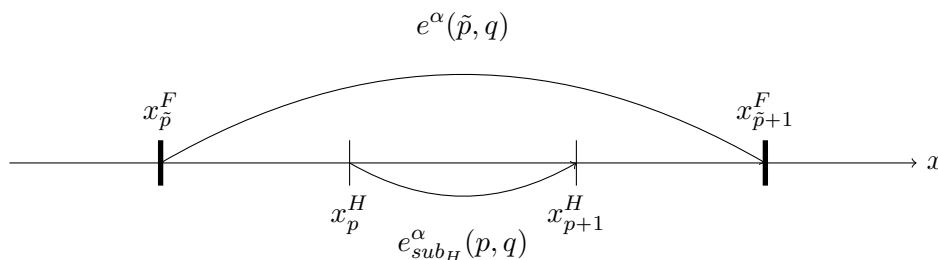


Figure 3: Subdivision of the age interval

A death in state H_2^{prod1} is either a death in H_1 or F_α . With product 1, we are able to decompose this count of deaths as a sum of deaths in H_1 and F_α . Therefore,

$$\hat{D}^{H_2^{prod1}}(p) = e^{H_1}(p) \times \mu_p^{H_1} + \sum_{q=1}^{M_t^\alpha} e_{sub_H}^\alpha(p, q) \times \mu^{F_\alpha}(x_p^H, t_q^\alpha). \quad (17)$$

Since $x_{\tilde{p}}^F \leq x_p^H < x_{\tilde{p}+1}^F$,

$$\mu^{F_\alpha}(x_p^H, t_q^\alpha) = \mu^{F_\alpha}(x_{\tilde{p}}^F, t_q^\alpha) = \mu_{\tilde{p}, q}^{F_\alpha}. \quad (18)$$

Therefore, Equation 17 becomes

$$\hat{D}^{H_2^{prod1}}(p) = e^{H_1}(p) \times \mu_p^{H_1} + \sum_{q=1}^{M_t^\alpha} e_{sub_H}^\alpha(p, q) \times \mu_{\tilde{p}, q}^{F_\alpha}. \quad (19)$$

Moreover, Equation 14 and Equation 16 lead to

$$D^{\hat{H}_2}(p) = e^{H_2}(p) \times \frac{\hat{D}^{H_2^{prod1}}(p)}{e^{H_2^{prod1}}(p)}. \quad (20)$$

Then,

$$D^{\hat{H}_2}(p) = e^{H_2}(p) \frac{e^{H_1}(p)}{e^{H_2^{prod1}}(p)} \times \mu_p^{H_1} + \sum_{q=1}^{M_t^\alpha} e^{H_2}(p) \frac{e_{sub_H}^\alpha(p, q)}{e^{H_2^{prod1}}(p)} \times \mu_{p,q}^{F_\alpha}. \quad (21)$$

Therefore, all the expected values of the Poisson variables $D^{H_k}(p)$, $k \in \{1, 2\}$ and $D^g(p, q)$, $g \in \{\alpha, \beta\}$ are linear combinations of the components of the vector $\Lambda = (\mu^{H_1}, \mu_{vec}^\alpha, \mu_{vec}^\beta)$.

We then write $D \sim Poisson(C\Lambda)$, where:

- $C \in \mathcal{M}_{(M^H+M^H+(M_t^\alpha+M_t^\beta), M_x^F), (M^H+(M_t^\alpha+M_t^\beta), M_x^F)}(\mathbb{R})$, and
- $\Lambda(\Theta) = \begin{bmatrix} \mu^{H_1}(\theta^{H_1}) \\ \mu_{vec}^{F_\alpha}(\theta^\alpha) \\ \mu_{vec}^{F_\beta}(\theta^\beta) \end{bmatrix} \in \mathbb{R}^{(M^H+(M_t^\alpha+M_t^\beta), M_x^F)}$, with $\Theta = (\theta^{H_1}, \theta^\alpha, \theta^\beta)^T \in \mathbb{R}^{J^{H_1}+J^\alpha+J^\beta}$.

Details for the structure of matrix C are given in Appendix C.

Therefore,

$$\hat{D} = C\Lambda(\Theta), \quad (22)$$

$$\Lambda(\Theta) = \exp(B\Theta), \quad (23)$$

where

$$B = \begin{bmatrix} B^{H_1} & 0 & 0 \\ 0 & B^\alpha & 0 \\ 0 & 0 & B^\beta \end{bmatrix} \in \mathcal{M}_{(M^H+(M_t^\alpha+M_t^\beta), M_x^F), (J^{H_1}+J^\alpha+J^\beta)}(\mathbb{R}). \quad (24)$$

To have smooth mortality rates, a penalty inspired from the P-splines smoothing method is added to the log-likelihood such that

$$l^{pen}(\Theta) = l(\Theta) - \frac{1}{2} \Theta^T P \Theta, \quad (25)$$

where:

- $l(\Theta)$ is the log-likelihood associated with the random vector $D = (D^{H_1}, D_{vec}^\alpha, D_{vec}^\beta, D^{H_2})^T$,

$$P = \begin{bmatrix} P^{H_1} & 0 & 0 \\ 0 & P^\alpha & 0 \\ 0 & 0 & P^\beta \end{bmatrix} + \left[\begin{array}{c|cc} 0 & 0 & 0 \\ \hline 0 & & \\ \hline 0 & P^{\alpha/\beta} & \end{array} \right] \in \mathcal{M}_{J^{H_1}+J^\alpha+J^\beta}(\mathbb{R}), \quad (26)$$

with $P^G = \rho^G (D^G)^T D^G$, $G \in \{H_1, \alpha, \beta, \alpha/\beta\}$, and

- ρ^G represents the weight given to the penalty, as in Section 3.1.

Penalty matrices $P^{H_1} \in \mathcal{M}_{J^{H_1}}(\mathbb{R})$, $P^\alpha \in \mathcal{M}_{J^\alpha}(\mathbb{R})$, and $P^\beta \in \mathcal{M}_{J^\beta}(\mathbb{R})$ are the typical penalty

matrices used in the P-splines smoothing method. Each state H_1 , F_α , and F_β has its own matrix, penalizing coefficients of adjacent splines to ensure that the fitted mortality laws are smooth.

One might want to add a penalty between coefficients of the states F_α and F_β . This is the aim of the penalty matrix $P^{\alpha/\beta} \in \mathcal{M}_{J^\alpha+J^\beta}(\mathbb{R})$. If not, $P^{\alpha/\beta} = 0$. Using vector notations, we have:

- $\mu_p^{H_1} = \Lambda_p, p \in \{1, \dots, M^H\}$,
- $\mu_{p,q}^{F_\alpha} = (\mu_{vec}^\alpha)_{(q-1).M_x^F+p} = \Lambda_{M^H+(q-1).M_x^F+p}, p \in \{1, \dots, M_x^F\}, q \in \{1, \dots, M_t^\alpha\}$,
- $\mu_{p,q}^{F_\beta} = (\mu_{vec}^\beta)_{(q-1).M_x^F+p} = \Lambda_{M^H+M_x^F.M_t^\alpha+(q-1).M_x^F+p}, p \in \{1, \dots, M_x^F\}, q \in \{1, \dots, M_t^\beta\}$, and
- $\mu_p^{H_2} = \Lambda_{M^H+M_x^F.(M_t^\alpha+M_t^\beta)+p}, p \in \{1, \dots, M^H\}$.

Therefore,

$$\begin{aligned}
l(\Theta) = & \sum_{p=1}^{M^H} -(\Lambda(\Theta) \times e)_p + D_p \log((\Lambda(\Theta) \times e)_p) + \\
& \sum_{p=1}^{M_x^F} \sum_{q=1}^{M_t^\alpha} -(\Lambda(\Theta) \times e)_{M^H+(q-1).M_x^F+p} + D_{M^H+(q-1).M_x^F+p} \log\left((\Lambda(\Theta) \times e)_{M^H+(q-1).M_x^F+p}\right) + \\
& \sum_{p=1}^{M_x^F} \sum_{q=1}^{M_t^\beta} -(\Lambda(\Theta) \times e)_{M^H+M_x^F.M_t^\alpha+(q-1).M_x^F+p} + \\
& D_{M^H+M_x^F.M_t^\alpha+(q-1).M_x^F+p} \log\left((\Lambda(\Theta) \times e)_{M^H+M_x^F.M_t^\alpha+(q-1).M_x^F+p}\right) + \\
& \sum_{p=1}^{M^H} -(\Lambda(\Theta) \times e)_{M^H+M_x^F.(M_t^\alpha+M_t^\beta)+p} + \\
& D_{M^H+M_x^F.(M_t^\alpha+M_t^\beta)+p} \log\left((\Lambda(\Theta) \times e)_{M^H+M_x^F.(M_t^\alpha+M_t^\beta)+p}\right). \tag{27}
\end{aligned}$$

The optimal coefficient $\hat{\Theta}$ is obtained by maximizing the penalized log-likelihood given by Equation 25, i.e.,

$$\hat{\Theta} = \arg \max_{\Theta} l^{pen}(\Theta). \tag{28}$$

We are then able to write the model as a Composite Link Model. This allows us to simultaneously estimate three smooth mortality laws while having a constraint allowing us to include observations from a portfolio with a different disability definition. Then, from Eilers (2007) and Remund et al. (2017), the coefficient $\hat{\Theta}$ is estimated by repeatedly solving the following system:

$$(\check{X}^T \tilde{W} \check{X} + P) \tilde{\Theta} = \check{X}^T (D - \tilde{D}) + \check{X}^T \tilde{W} \check{X} \tilde{\Theta}, \tag{29}$$

where $\check{X} = \tilde{W}^{-1} C \tilde{\Gamma} B$; $\tilde{W} = \text{diag}(\tilde{D})$; $\tilde{\Gamma} = \text{diag}(\Lambda)$; $\tilde{D} = C \Lambda(\tilde{\Theta})$.

A tilde indicates the current approximation at each iteration.

Starting values of the parameter $\tilde{\Theta}$ are needed. A convenient way is to start with the coefficient obtained by separately estimating the mortality rates in both healthy and disabled states (H_1 , F_α , and F_β) by using the P-splines smoothing method with observations from product 1 only.

4 Application to the problem of the deferred period

4.1 Introduction to the deferred period and the problem with the data

The deferred period is the minimum period that the loss of autonomy must last before the benefit begins. The length of the deferred period has a significant impact on the probability of the insurer paying the benefit to the policyholder. The premium is therefore lower as the length of the deferred period increases. The usual length in French long-term care contracts is 3 months. One of the advantages of the deferred period is that it reduces the number of short claims and, therefore, the management costs linked to the payment of the annuities.

Some long-term care products have a deferred period, while some do not. For contracts with a deferred period, as no annuity is paid to policyholders during the deferred period, the loss of autonomy is not reported in the database if the policyholder dies during this period, i.e., before the first annuity, as shown in Figure 4. This death is considered a death in autonomy. Only policyholders surviving until the end of the deferred period, at least until the payment of the first annuity, have a date of loss of autonomy reported in the database. Therefore, policyholders are considered healthy as long as they have not received any annuity. Therefore, the healthy group is composed of autonomous and newly disabled individuals, as shown in Figure 4.

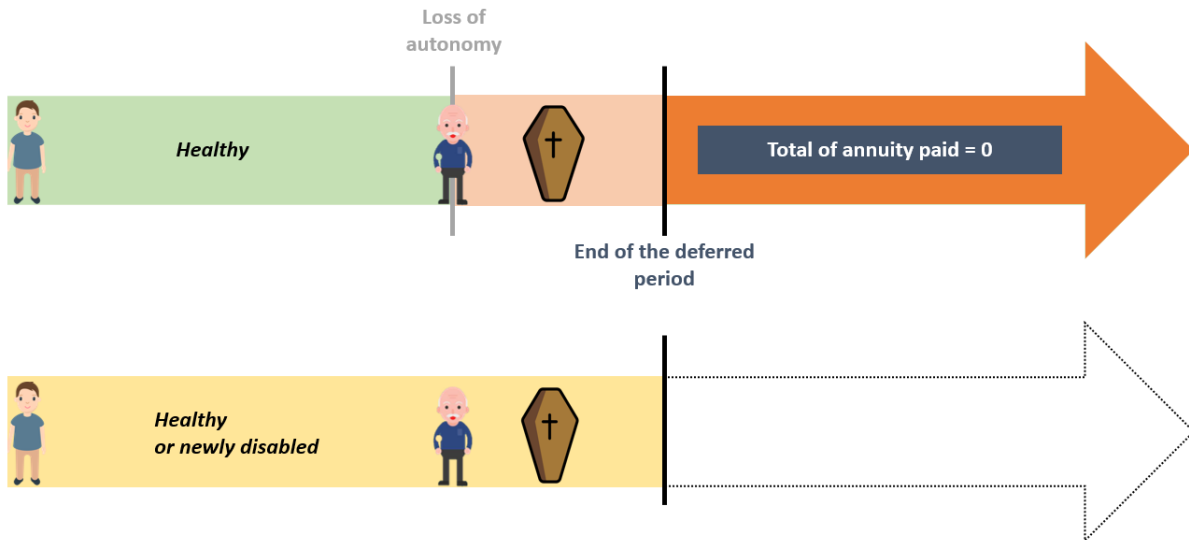


Figure 4: Illustration of the consequences of a deferred period on the data observed in the portfolio

Let us assume that the real date of loss of autonomy is available for all disabled policyholders who survived the deferred period. If not, the date of loss of autonomy can easily be estimated by

subtracting fr months from the date of the first annuity payment.

As the mortality during the first few months of disability is very high and the deaths occurring during the deferred period are reported as deaths in the healthy state, estimating mortality laws on such a database without considering this information can have multiple consequences:

- The mortality in the healthy state is overestimated.
- The mortality of the disabled policyholders during the deferred period is underestimated (only disabled individuals surviving the deferred period are considered exposed, but no deaths are reported during this period).
- The incidence is underestimated since the loss of autonomy is reported only if the newly disabled policyholder survives until the first annuity.

Let us assume that we have a database with some policyholders having a deferred period and some policyholders covered from the first day of the loss of autonomy (without any deferred period). This situation can also occur in the case of having several databases from different insurers. In this situation, one cannot simply aggregate the data basis without accounting for the deferred period of one product.

This situation corresponds to a problem of multidefinition as described in Section 2, with one definition included in the other one.

4.2 Modelling of the product

In this case, product 1 corresponds to the product without a deferred period, and product 2 corresponds to the product with a deferred period. Being disabled for product 2 implies being disabled for more than fr months. Therefore, a policyholder disabled with definition 2 (product 2) is disabled with definition 1 (product 1). Thus, $T_2 \subset T_1$.

These products can be represented as in Figure 1, where:

- H_1 denotes the state of healthy policyholders,
- F_α denotes the state of newly disabled policyholders. The loss of autonomy occurred less than fr months ago,
- F_β denotes the state of disabled policyholders who have lost their autonomy more than fr months ago, and
- H_2 denotes the state of healthy and newly disabled policyholders who have not yet received any annuity.

In this case,

$$\omega^1(x, t) = \begin{cases} 0, & \text{if } t \neq fr/12 \\ +\infty, & \text{if } t = fr/12. \end{cases} \quad (30)$$

The transition from F_α to F_β , if observed, systematically occurs at the end of the deferred period. It is natural to think that there is no jump in the mortality function before and after the deferred period. $D^\alpha(p, q), p \in \{1, \dots, M_x^F\}, q \in \{1, \dots, M_t^\alpha\}$ denotes the number of deaths occurring during the deferred period (only observed with product 1). As the maximum duration in F_α is $t = fr/12$, then $t_{M_t^\alpha+1}^\alpha = fr/12$.

$D^\beta(p, q), p \in \{1, \dots, M_x^F\}, q \in \{1, \dots, M_t^\alpha\}$ denotes the sum of counts of deaths occurring after the deferred period for products 1 and 2. As disabled policyholders enter state F_β at the end of the deferred period, the mortality before the end of the deferred period $\mu_{p, M_t^\alpha}^{F_\alpha}$ should be close to the mortality $\mu_{p,1}^{F_\beta}$ during the first subinterval of duration in F_β . It is therefore interesting to consider the mortality in F_α and F_β as only one smooth mortality law $\mu^F(x, t) = \mu^{T_1}(x, t)$, such that $F = \{F_\alpha \cup F_\beta\}$.

As introduced in Section 2,

$$\mu^F(x, t) = \lim_{h \rightarrow 0} \frac{\mathbb{P}(X_{x+h} = \text{Death} | X_x \in F, X_{x-t} \in F, X_{(x-t)-} = H_1)}{h}. \quad (31)$$

A policyholder who entered state F less than fr months ago is necessarily in state F_α . In contrast, a disabled policyholder who lost autonomy t years ago, with $t > fr/12$, is necessarily in state F_β for $t - fr/12$ years. Therefore,

$$\mu^F(x, t) = \begin{cases} \mu^{F_\alpha}(x, t), & \text{if } t < fr/12 \\ \mu^{F_\beta}\left(x, t - \frac{fr}{12}\right), & \text{if } t \geq fr/12 \end{cases}. \quad (32)$$

A more detailed proof of this equation (Equation 32) is available in Appendix D.

Note: The second parameter of $\mu^{F_\beta}(\cdot, \cdot)$ denotes the time since entry into state F_β . This corresponds to the time since the end of the deferred period. In contrast, the second parameter of $\mu^F(x, t)$ denotes the time since the loss of autonomy, corresponding to the time since entry in F_α .

Let $D^F = [D^\alpha : D^\beta] \in \mathcal{M}_{M_x^F, M_t^\alpha + M_t^\beta}$ be the augmented matrix of the counts of deaths. $D_{\cdot, M_t^\alpha}^F$ denotes the vector of counts of deaths in LTC during the last period before time $fr/12$. $D_{\cdot, M_t^\alpha+1}^F = D_{\cdot, 1}^\beta$ denotes the vector of counts of deaths in LTC occurring during the first period after the deferred period.

Therefore,

$$D_{vec}^F = (D_{vec}^\alpha, D_{vec}^\beta)^T. \quad (33)$$

As the subdivision of the interval of ages is the same for F_α and F_β , the subdivision of the interval of ages of F is $\{x_1^F, \dots, x_{M_x^F}^F\}$. By using the subdivision of the duration interval from F_α and F_β ,

the subdivision of the interval for the overall LTC state F is

$$\left\{t_1^F, \dots, t_{M_t^F+1}^F\right\} = \left\{t_1^\alpha, \dots, t_{M_t^\alpha+1}^\alpha = t_1^\beta + \frac{fr}{12}, \dots, t_{M_t^\beta+1}^\beta + \frac{fr}{12}\right\}, \quad (34)$$

where $M_t^F = M_t^\alpha + M_t^\beta$.

A common basis of B-splines is used for the age dimension for the 2 states. The matrix of splines for ages in disabled states is denoted $B_x^F \in \mathcal{M}_{M_x^F, J_x^F}$. Splines are positioned over the entire interval of duration $\left[t_1^F; t_{M_t^F+1}^F = t_{M_t^\beta+1}^\beta + \frac{fr}{12}\right]$. Some splines are common to F_α and F_β , as shown in Figure 5, which represents the splines basis on the overall duration interval, $J_t^F \leq J_t^\alpha + J_t^\beta$. In the example of Figure 5, 4 splines are common to F_α and F_β . Therefore, $J_t^F = J_t^\alpha + J_t^\beta - 4$.

The matrix of splines for the duration dimension is denoted $B_t^F \in \mathcal{M}_{M_t^\alpha+M_t^\beta, J_t^F}$.

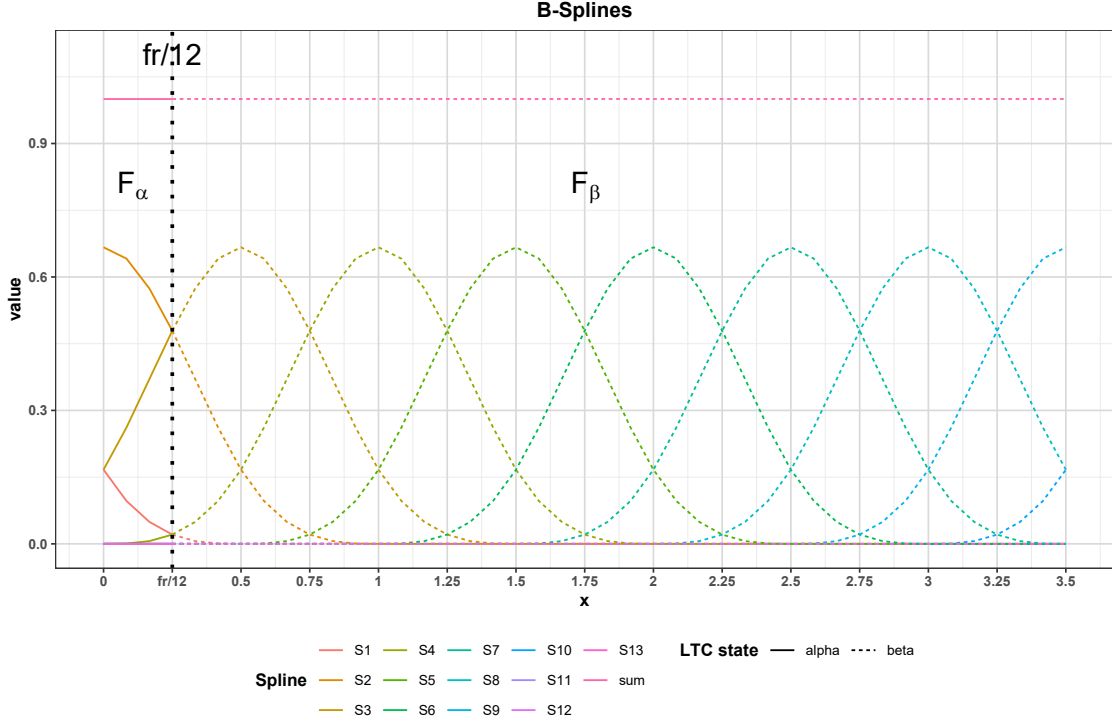


Figure 5: Basis of splines for the duration for states F_α and F_β

To ensure the same coefficients associated with the splines shared by F_α and F_β , $B^F = B_t^F \otimes B_x^F$ is of a slightly different structure than $\begin{bmatrix} B^\alpha & 0 \\ 0 & B^\beta \end{bmatrix}$.

More details about the structure of the matrices B_t^F and B^F are given in Appendix E.

4.2.1 Constrained optimization method applied to the deferred period

The method explained in Section 3.2 is applied to the problem of the deferred period. Having one single spline basis for state $F = F_\alpha \cup F_\beta$ has only a slight impact on the form of the objective function, which becomes

$$\begin{aligned} \max_{\theta^{H_1}, \theta^{H_2}, \theta^F} & l^{H_1}(\theta^{H_1}) - \frac{1}{2} (\theta^{H_1})^T P^{H_1} \theta^{H_1} + l^{H_2}(\theta^{H_2}) - \frac{1}{2} (\theta^{H_2})^T P^{H_2} \theta^{H_2} + \\ & l^\alpha(\theta^F) + l^\beta(\theta^F) - \frac{1}{2} (\theta^F)^T P^F \theta^F, \end{aligned} \quad (35)$$

subject to

$$\begin{aligned} \mu_p^{H_2}(\theta^{H_2}) &= \frac{\mathbb{P}(X_x = H_1)}{\mathbb{P}(X_x = H_1 \cup X_x = F_\alpha)} \mu_p^{H_1}(\theta^{H_1}) + \\ & \sum_{q=1}^{M_t^\alpha} \frac{\mathbb{P}(X_x = F_\alpha, X_{(x-t_q^\alpha)} = F_\alpha, X_{(x-t_{q+1}^\alpha)} = H_1)}{\mathbb{P}(X_x = H_1 \cup X_x = F_\alpha)} \mu_{p,q}^{F_\alpha}(\theta^F). \end{aligned} \quad (36)$$

4.2.2 PCLM method applied to the deferred period

The submatrix $\begin{bmatrix} B^\alpha & 0 \\ 0 & B^\beta \end{bmatrix}$ from matrix B in Equation 24 is replaced by a single matrix B^F . This slight modification does not affect the remaining formulas or the method for the estimation of the optimal coefficients.

With the same reasoning, the submatrix $\begin{bmatrix} P^\alpha & 0 \\ 0 & P^\beta \end{bmatrix}$ from matrix P in Equation 26 is replaced by a single matrix P^F . This allows us to add constraints between splines before and after the deferred period to ensure smoothness of the mortality law on the axis of the duration.

$$\text{Therefore, } B = \left[\begin{array}{c|cc} B^{H_1} & 0 & 0 \\ \hline 0 & & \\ 0 & & B^F \end{array} \right], \text{ and } P = \left[\begin{array}{c|cc} P^{H_1} & 0 & 0 \\ \hline 0 & & \\ 0 & & P^F \end{array} \right].$$

4.3 Data: Application to a single portfolio by recreating a fictitious deferred period

4.3.1 Presentation of the data set

We rely on data from a large French LTC portfolio. This portfolio does not have any deferred period. As the incidence and mortality laws greatly differ for males and females, biometric functions have to be estimated separately for each gender. We focus the application only on females. In this application, we consider mild and severe LTC, with the GIR1234 definition from the AGGIR grid used by the French government for the attribution of public aid and described in Dupourqué (2012). In this portfolio, 1 388 deaths are observed in the autonomous state (H_1), versus 832 in LTC

$(T_1 = F_\alpha \cup F_\beta)$.

To ensure that Assumption 2.2 of homogeneity and independence is met, this portfolio is divided into two data sets. The first one, denoted DB_1 , corresponds to the definition 1, without a deferred period. This portfolio is only a subset of the initial portfolio. Policyholders not represented in DB_1 are selected in the second portfolio DB_2 . This second database is then modified to fictitiously create a deferred period by postponing the eventual date of loss of autonomy by $fr = 3$ months.

Mortality rates are piecewise constant for the duration. However, since the mortality is very high at the date of occurrence of the loss of autonomy and decreases substantially during the first year, it is common in long-term care modelling to fix smaller steps on the subdivision of the interval of duration during the first year. In the following application, we assume a constant mortality rate by month during the first year of loss of autonomy. Starting from the second year, mortality rates are assumed to be yearly constants.

For the age dimension, mortality rates are assumed to be constant between two integer ages.

After cleaning the original data at the individual granularity, observations are aggregated to obtain exposures and counts of deaths for each subdivision of ages and durations on which mortality rates are assumed to be constants.

Observations from DB_1 are:

- the vector of central exposures in H_1 : e^{H_1} ,
- the vector of counts of deaths in H_1 : D^{H_1} ,
- the matrix of central exposures in F_α : e^α ,
- the matrix of counts of deaths in F_α : D^α ,
- the matrix of central exposures in F_β (only individuals in DB_1): $e^{\beta,1}$, and
- the matrix of counts of deaths in F_β (only individuals in DB_1): $D^{\beta,1}$.

Observations from DB_2 are:

- the vector of central exposures in H_2 : e^{H_2} ,
- the vector of counts of deaths in H_2 : D^{H_2} ,
- the matrix of central exposures in F_β (only individuals in DB_2): $e^{\beta,2}$, and
- the matrix of counts of deaths in F_β (only individuals in DB_2): $D^{\beta,2}$.

4.3.2 Results

In the case of the deferred period, insurers are mostly interested in mortality in H_1 and $F = F_\alpha \cup F_\beta$. Observations from H_2 are used only to improve the estimations of the two other mortality laws.

The mortality for H_2 has a rather limited interest for insurers.

In this section, we compare the results of estimations of the 2 mortality laws μ^{H_1} and μ^F with 3 different methods:

1. Independent estimations of each mortality law with the P-splines smoothing method (Without constr.),
2. Constrained optimization (Constr.) from Section 3.2,
3. Penalized Composite Link Model (PCLM) from Section 3.3.

The smoothing parameter ρ in each of these methods is selected to minimize the BIC.

Estimated mortality intensities in the healthy state and their associated confidence intervals are plotted in Figure 6. While estimated intensities are similar at young ages where data contain many observations, the mortality curve estimated with the PCLM method diverges from the two others after 85 years old. The estimated mortality function at old ages is higher with the PCLM method compared to that of the mortality obtained by only using information from the portfolio without a deferred period and the mortality estimated with the constrained optimization model. In fact, the optimal smoothing parameter ρ associated with mortality in H_1 minimizing the BIC is smaller with the PCLM method. Therefore, the linear smoothing penalty is weaker, and the fitted mortality function is more flexible than that with the two other methods. It gives more weight to the observations and less to the linear constraint, enabling the capture of more variance at old ages. The size of the confidence interval is smaller than that with the constrained optimization method than with the two other methods.

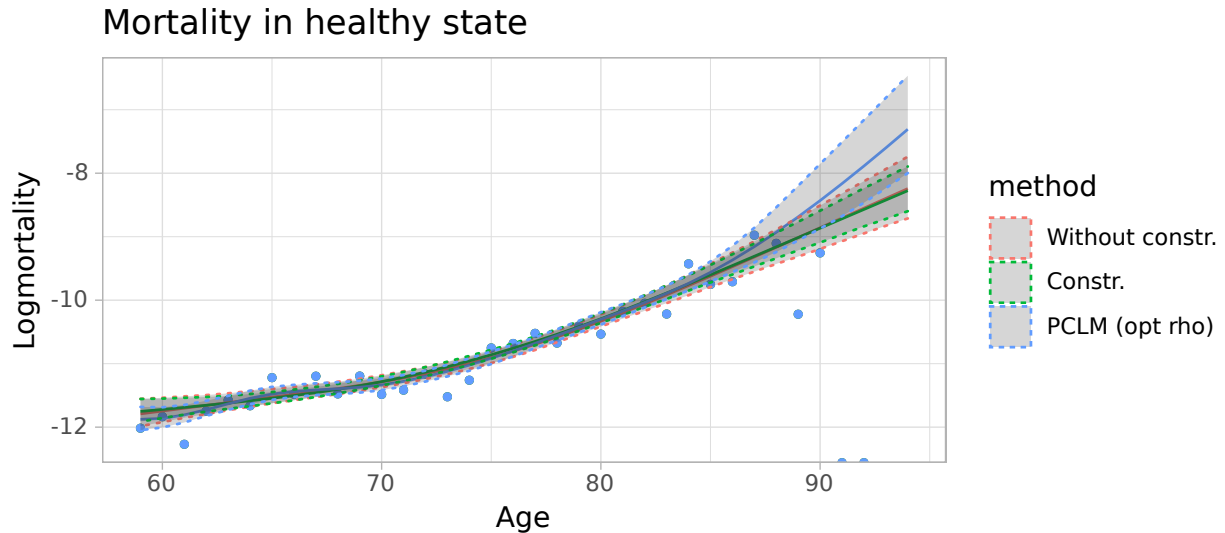


Figure 6: Mortality in the healthy state H_1

Mortality in LTC depends on 2 inputs, the attained age and the duration since the loss of autonomy.

Therefore, the log-mortality is represented as a surface instead of a curve for mortality in the healthy state. The surfaces of the log-mortality rates estimated with the three methods are plotted in Figure 7. The resulting mortality rates obtained either with the independent estimation or with the constrained optimization methods are similar. Adding the constraints to the maximization of the likelihood does not significantly change the optimum. However, the PCLM method is more different in the formulation of the problem since no coefficients are associated with mortality in state H_2 .

As a consequence, the maximum likelihood estimators of this method result in different mortality rates in LTC compared to those produced by the other two methods. As shown in Figure 7, the mortality function estimated with the PCLM method does not seem to depend greatly on the duration. Moreover, the pattern of this estimated mortality law shows that the mortality of disabled policyholders has a smile shape on the age dimension. For a fixed duration, the mortality is high at young and old ages and seems to reach the minimum at approximately 80 years old. This phenomenon can be explained by the prevalence of the pathologies affecting the disabled policyholders. This prevalence depends greatly on the attained age. Pathologies affecting young disabled policyholders are mostly diseases affecting mortality, such as cancer. In contrast, the most represented pathologies at approximately 80 years old are Alzheimer's disease and dementia. These pathologies are known to have a limited impact on mortality. Therefore, as explained in Biessy (2016), cancer has a high contribution to mortality in LTC at young ages, especially for low durations.

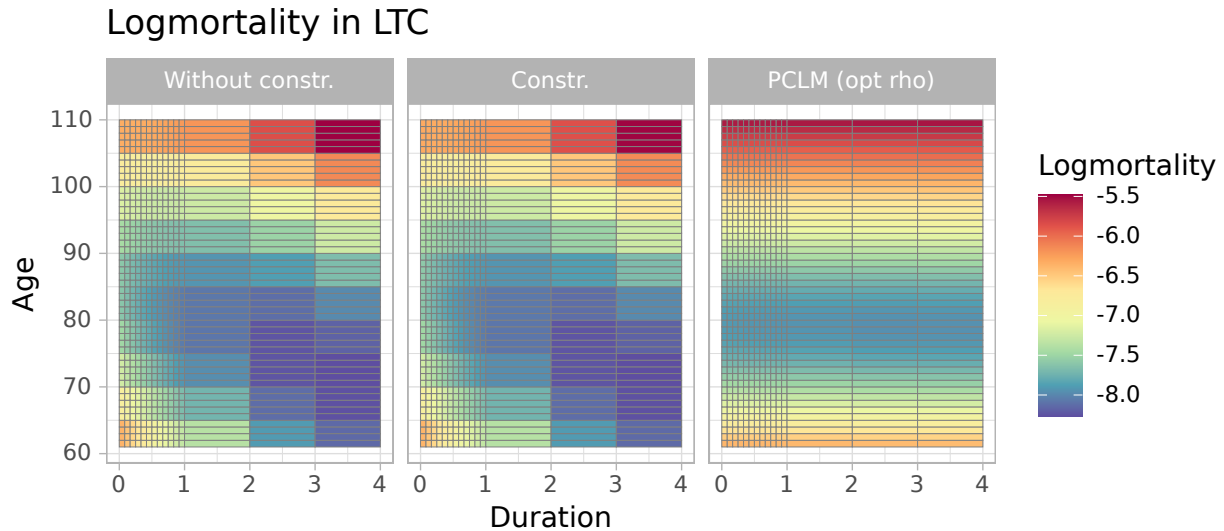


Figure 7: Mortality in Long-Term Care (state F)

Figure 8 compares the estimated mortality laws of disabled policyholders at 85 years old as a function of duration. For all methods, the order of the penalty is fixed to 2 on the age dimension and 1 on the duration dimension to prevent unreasonable divergence of the mortality with increasing duration. For a fixed attained age, the mortality with and without constraints has a smile shape, with mortality decreasing at low durations and increasing after the first year. In contrast, the mortality of disabled

policyholders aged 85 years old, estimated with the PCLM method, is essentially the same for all durations. The PCLM method gives more weight to the smoothing penalty ρ on the duration shape compared to that of the 2 other methods. Since this penalty tries to minimize the variations in the duration dimension, the mortality is more likely to be constant in this dimension. The confidence interval obtained with the PCLM method is narrower than those estimated with the two other methods.

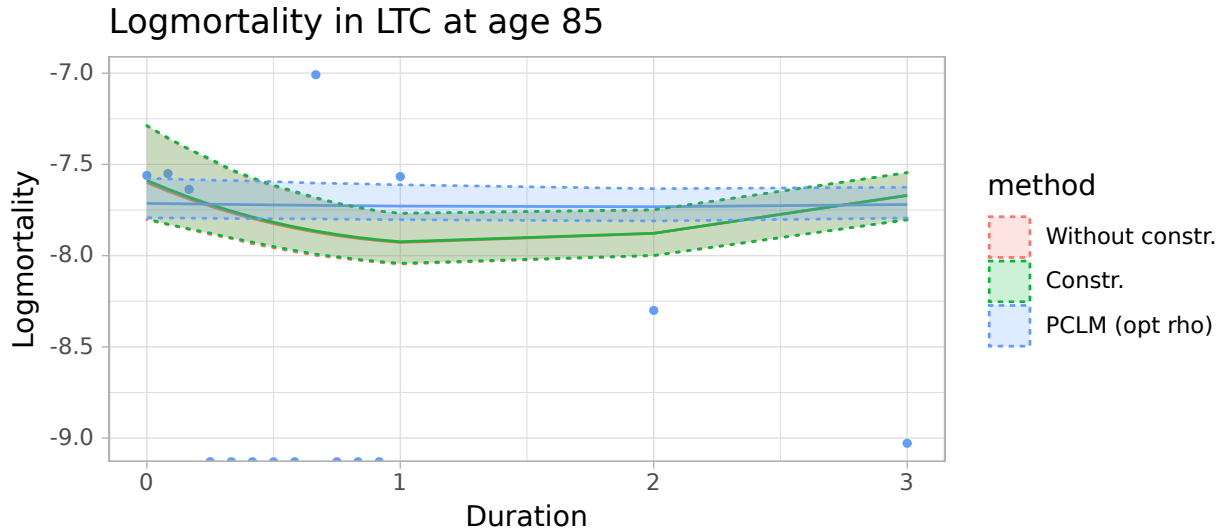


Figure 8: Mortality in Long-Term Care (state F) at age 85

One might want to evaluate the evolution of mortality with duration for a given entry age. Figure 9 shows the mortality rates of disabled policyholders losing their autonomy at 70 years old. For all methods, mortality decreased with time during the 3 first years. However, the decrease rate is smaller with the PCLM method. While the PCLM method gives the lowest mortality rates during the first three months, the estimated mortality for higher durations is higher than those with the two other methods.

Figure 10 shows the mortality intensities during the first month following the loss of autonomy. As anticipated with Figure 7, the mortality function has a smile shape on the age dimension. The confidence interval obtained with the PCLM method is narrower than those associated with the two other methods.

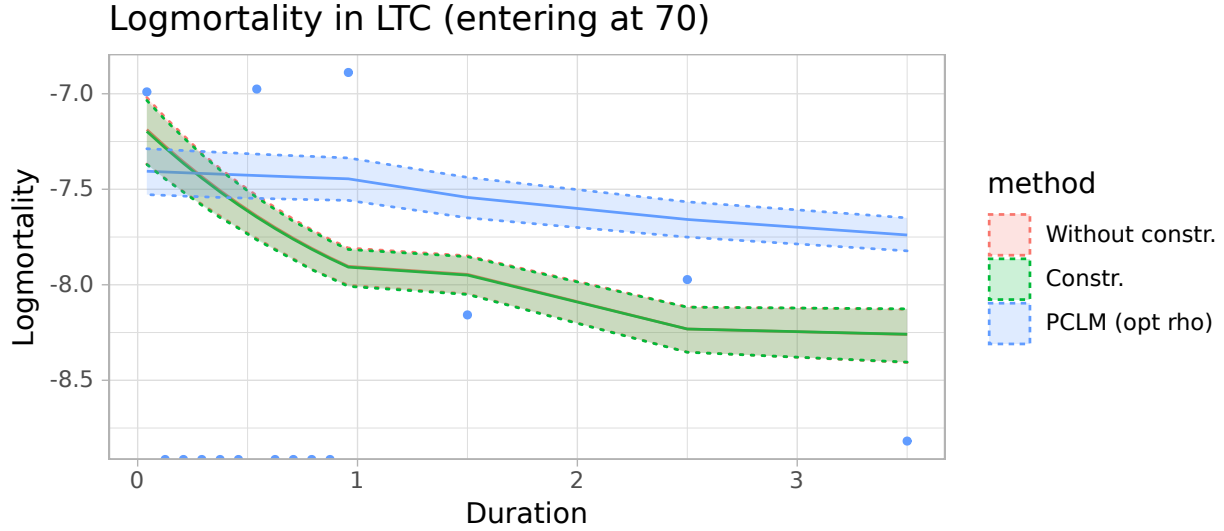


Figure 9: Mortality of a policyholder losing autonomy at age 70

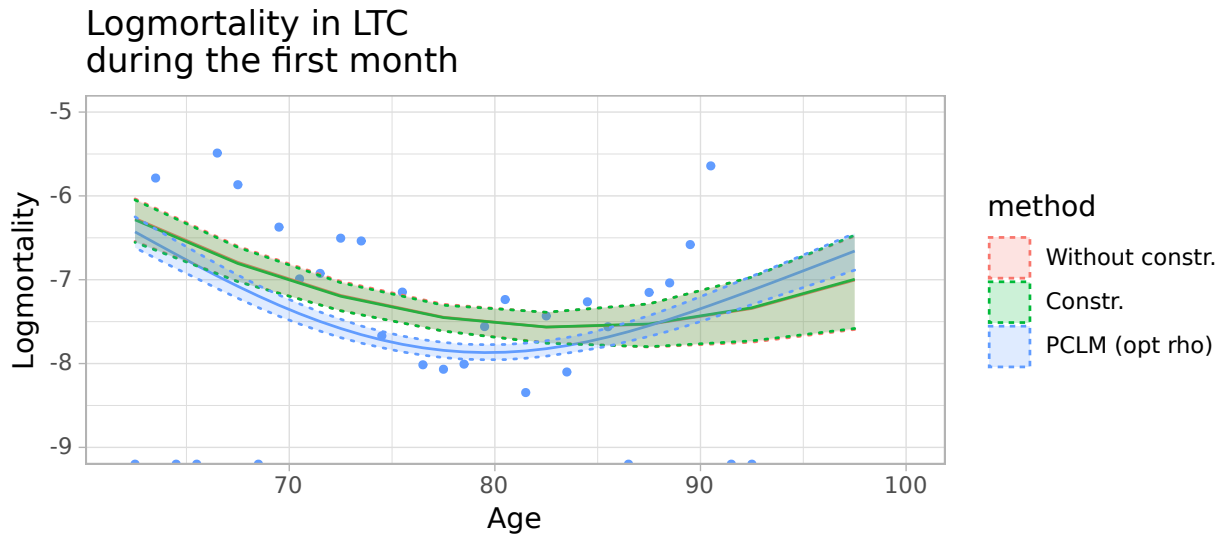


Figure 10: Mortality of disabled policyholders during the first month following the loss of autonomy

The deviance residuals in the healthy (H_1) and disabled (F) states are plotted in Figures 11 and 12, respectively. We use these graphics to analyse if the residuals are within the interval $[-2; 2]$ and if there exists a trend in the residuals (lower residuals at low ages and higher residuals at old ages, for example). The existence of a trend may indicate that the model does not well capture the effect of one of the dimensions (age and/or duration) on the mortality rates. In the healthy state, all models seem to perform well. In the disabled state, models with and without constraints seem to underestimate mortality during the first year compared to the PCLM model. In contrast, this model overestimates mortality during the second year. The residuals during the first year seem homogeneous and quite similar among the 3 models.

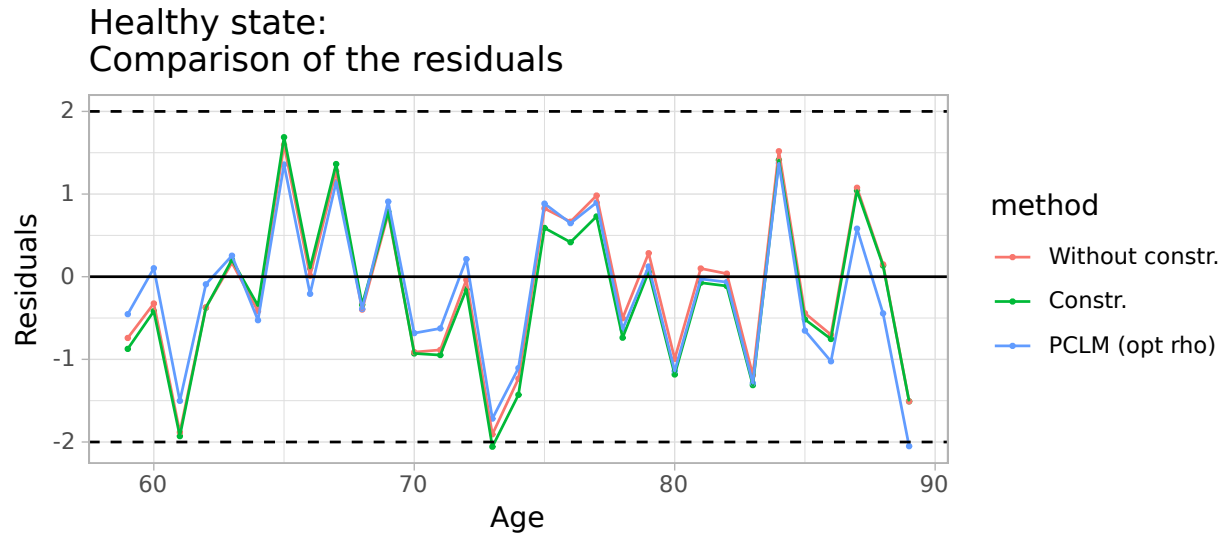
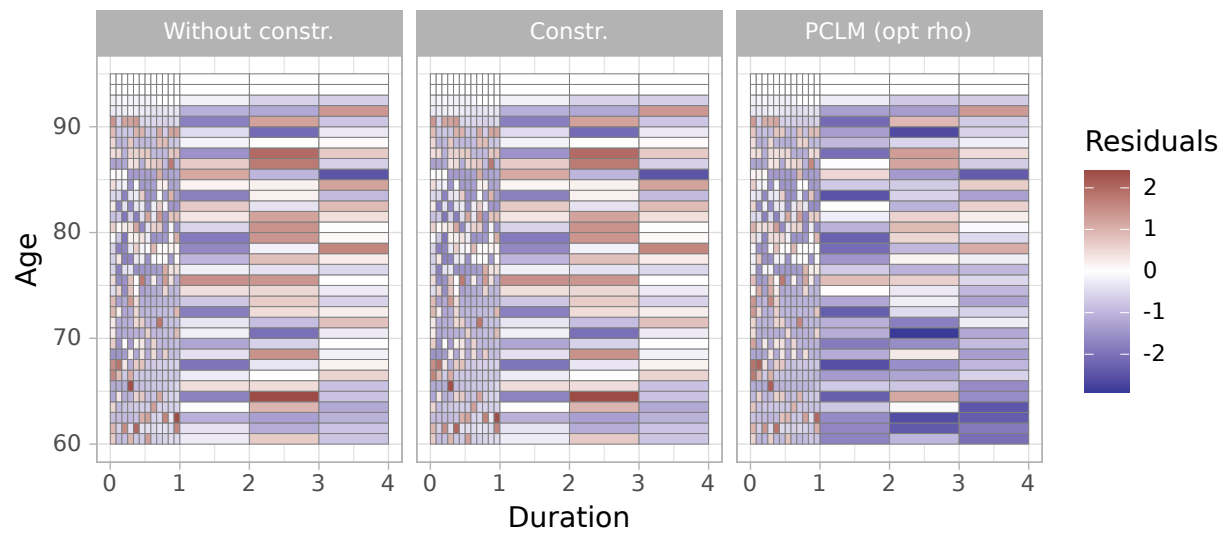
Figure 11: Deviance residuals in the healthy state H_1 Figure 12: Deviance residuals in the disabled state F

Figure 13 displays the first month of disability. The deviance residuals are within the specified interval $[-2; 2]$ and seem quite homogeneous among all ages.

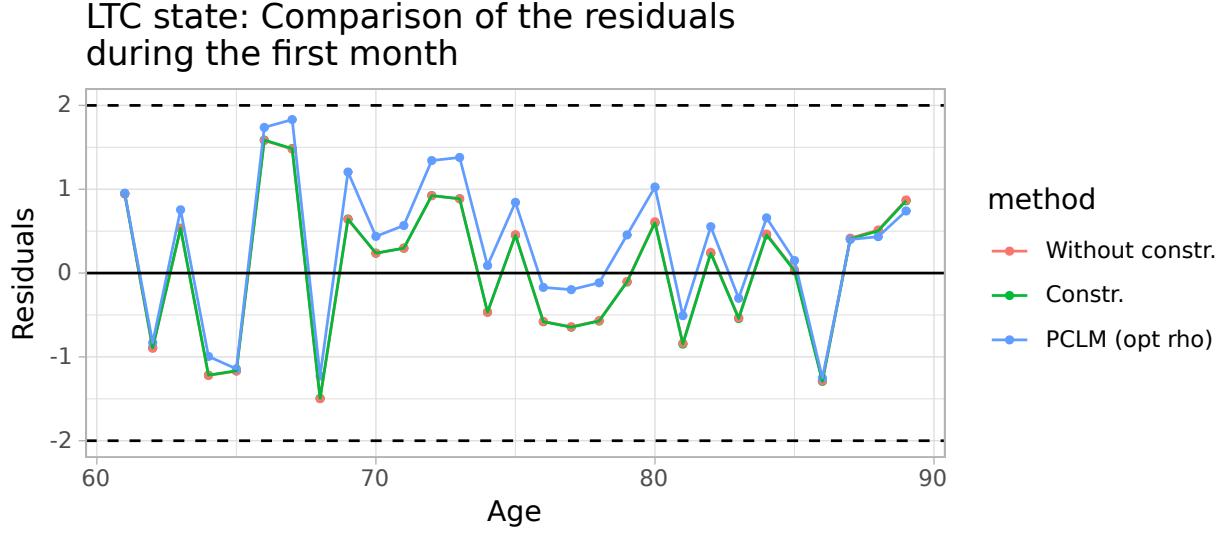


Figure 13: Deviance residuals in the disabled state F during the first year following the loss of autonomy

4.3.3 Model Performance

The predictive performances of the three methods are evaluated with cross-validation. The following metrics are analysed:

- the root mean square error (RMSE)

$$\frac{1}{M^G} \sum_{k=1}^{N^G} (D_k^G - \hat{D}_k^G)^2, G \in \{H_1, F\},$$

where $N^{H_1} = M^H$, $N^F = M_x^F (M_t^\alpha + M_t^\beta)$, and \hat{D}_k^G denotes the k^{th} term of the vector of expected counts of deaths in group G ,

- the root mean percentage error (RMPE)

$$\frac{1}{M^G} \sum_{k=1}^{N^G} \left(\frac{D_k^G - \hat{D}_k^G}{\hat{D}_k^G} \right)^2, G \in \{H_1, F\},$$

- the χ^2 statistic

$$\frac{1}{M^G} \sum_{k=1}^{N^G} \left(\frac{D_k^G - \hat{D}_k^G}{\sqrt{\hat{D}_k^G}} \right)^2, G \in \{H_1, F\},$$

- the deviance

$$\sum_{k=1}^{N^G} 2 \left(D_k^G \log \left(\frac{D_k^G}{\hat{D}_k^G} \right) - (D_k^G - \hat{D}_k^G) \right).$$

Model performance on the healthy state is evaluated on the age range [60, 90], and model performance on the disabled state is evaluated on the first year of disability for ages between 65 and 90 years old. When the metrics of the two methods presented in this paper are better than those of the method without constraints (independent estimation of each biometric law), then the metric is displayed in bold. The best value for each metric is written in red. The results are shown in Table 1 and Table 2.

Table 1 shows that adding a constraint to the P-splines smoothing method improves the Root Mean Square Error (RMSE), the Root Mean Percentage Error (RMPE), the deviance and the χ^2 statistics of mortality in the healthy state. This method also improves the metrics for the mortality of disabled policyholders, as shown in Table 2.

The PCLM method mostly improves the predictive performance of the mortality of disabled policyholders as shown in Table 2. The root mean percentage error and χ^2 are the lowest for this method compared to those of the two others. This method also improves the Root Mean Percentage Error (RMPE) on the prediction of the mortality of autonomous policyholders (Table 1).

Method	RMSE	RMPE	χ^2	Deviance
Constr.	2.19	0.60	173.32	204.82
Without Constr.	2.19	0.63	180.95	205.96
PCLM	2.21	0.57	176.70	215.38

Table 1: Cross validation: mortality in the healthy state

Method	RMSE	RMPE	χ^2	Deviance
Constr.	0.30	5.76	1 338,40	509.40
Without Constr.	0.30	5.80	1 344,35	509.52
PCLM	0.30	4.69	1 226,24	514.33

Table 2: Cross validation: mortality in disabled state

Additional plots of confidence intervals and analyses of Poisson residuals for each state H_1 , H_2 and F are available in Appendix F.

5 Discussion

In this paper, we present two methods to estimate transition rates by using combined experience data of two long-term care portfolios with differing disability definitions. In this paper, we focus on the case when one of the definitions is included in the other. In this situation, the healthy state (denoted H_2) of one of the products/definitions is a mixture of the healthy state (H_1) and the disabled state (T_1) of the other product, as described in Section 2. The disabled state T_1 is

divided in two distinct states, namely F_α and F_β , such that $F_\alpha = T_1 \setminus T_2$. F_β is the state containing policyholders disabled according to both definitions.

We focus on the estimation of mortality in healthy and disabled states by using combined experiences from both portfolios. Mortality in the healthy state is considered a function of age only; whereas, mortality in disabled states is assumed to be a function of age and duration since the loss of autonomy. In fact, mortality in the disabled state often depends greatly on the duration, with a higher mortality in the first months following entry into the disabled state.

Assuming that the population of both portfolios is homogeneous, the mortality of this healthy state H_2 is a mixture of the mortality in H_1 and F_α . The three mortality laws are then linked. To avoid overfitting, the two methods introduced in this paper rely on the P-splines smoothing method embedded in the Poisson Generalized Linear Model framework as presented in Section 3.1. With the Poisson assumption justified in Section 2, counts of deaths are assumed to have a Poisson distribution of parameters proportional to the central exposure and the mortality intensities.

The presented methods are based on the idea of maximizing the likelihood. The first approach described in Section 3.2 consists of maximizing the Poisson likelihood subject to multiple constraints linking the biometric functions of the two portfolios. In this method, the mortality laws (mortality in H_1 , H_2 , F_α and F_β) are assumed to be expressed as exponential of the combination of basis-splines. A set of spline coefficients must be estimated for each state. The second approach presented in Section 3.3 uses the Penalized Composite Link Model introduced by Eilers (2007). This method uses the fact that the expectancy of the counts of deaths in each state (H_1 , H_2 , F_α and F_β) can be expressed as a linear combination of the mortality in H_1 , F_α and F_β . Therefore, no coefficients are associated with mortality in H_2 . This method allows us to reduce the number of estimated coefficients compared to that of the first method.

The presented methods are then applied to solve the problem when having two portfolios with and without a deferred period. In this case, being disabled for the product with a deferred period implies having the disabled status for the product without a deferred period. This situation is then a special case of the problem solved in this paper, since one of the disability definitions is included in the other. To meet the assumption of homogeneity between the two portfolios, a single French long-term care portfolio is randomly divided into two. One of the data sets is then modified to include a fictitious deferred period. The two methods are then evaluated on these portfolios and compared with the separate modelling approach to show the benefits of combining information of the two portfolios despite the difference in the disability definitions.

We show that combining information improves the confidence intervals and the predictive performance. First, adding a constraint to the maximization of the P-splines likelihood helps to reduce the size of the confidence interval of the mortality in the healthy state, without having a large impact on the mortality estimated independently on one single portfolio. However, this first method has almost no impact on the estimated mortality in the disabled state and its associated confidence

interval. The results of the cross-validation show that this method has better predictive performance than that of the separate modelling approach. The PCLM method helps reduce the confidence interval of the estimated mortality in the disabled state, but it has an impact on the estimated mortality compared to that of the estimation without combining information of the portfolios. The cross-validation shows that this method gives better predictive performance on the RMPE (Root Mean Percentage Error) and χ^2 metrics. However, this method slightly increases the residual deviance of the model compared to that in the separate modelling approach.

The methods introduced in this paper can be extended to other cases where portfolios have different disability definitions. For example, some long-term care products offer the possibility for the policyholder to be covered for several levels of loss of autonomy, namely “Mild” and “Severe” disability. Policyholders only covered for severe disability are considered “autonomous” if they are only mildly disabled. The long-term care portfolio of the insurer is composed of policyholders with different disability definitions. This portfolio can therefore be divided into two different portfolios, one for each definition. The first covers both mild and severe disability, whereas the second covers only severe disability. In this situation, being disabled with the second definition implies being disabled for the first definition. It is therefore a special case of the problem of multiple definitions studied in this paper. The two methods that we introduced can be used to estimate simultaneously the biometric laws of both definitions by combining information of all policyholders, instead of separating policyholders covered for the mild disability from the other policyholders.

The methods introduced in this paper assume that the counts of deaths have a Poisson distribution. One of the properties of the Poisson distribution is that its mean equals variance. However, overdispersion is often observed in the context of mortality modelling. Accordingly, the Poisson assumption may be too restrictive. In future studies, a dispersion parameter can be introduced to improve the model. The models assume homogeneity of policyholders in the two portfolios. This assumption can be difficult to satisfy because of antiselection. In the example of the deferred period, the population of policyholders choosing a long-term care contract without a deferred period is likely different from the population of policyholders underwriting a contract with a deferred period. Finally, in future research the methods developed in this paper may be generalized to parametric forms of mortality.

A Proof of the total likelihood of the combined observations

Let j denote an individual. Each individual is observed with only one product (1 or 2). We introduce:

- $\overline{x^S}(j)$, the minimum age at which we observe individual j (age at underwriting, or age at the beginning of the observation period),
- $\overline{x^H}(j)$, the age of end of observation in healthy state (equal to the age of death, age of loss of autonomy, or age at censoring),
- $\overline{x^T}(j)$, the age of end of observation in disabled state T_k where k denotes the definition with which we observe individual j (age of death, or age at censoring if individual j is in state T_k at the end of observation period),
- $\overline{x^{F_g}}(j)$, the age of end of observation in disabled state F_g where $g \in \{\alpha, \beta\}$ (age of death, age of transition from F_α to F_β or age at censoring if individual j is in state F_g at the end of observation period),
- $c^H(j)$, the cause of exit of healthy state for individual j such that:
 - $c^H(j) = 0$ if individual j is still in the healthy state at the end of the observation (right censoring),
 - $c^H(j) = 1$ if individual j dies in healthy state,
 - $c^H(j) = 2$ if the cause of exit of the healthy state is the loss of autonomy (entry in state T_k if the health status of individual j is observed with definition k),
- $c^T(j)$, the cause of exit of disabled state ($T_k, k \in \{1, 2\}$) for individual j such that:
 - $c^T(j) = 0$ if individual j is in state T_k at the end of the observation (right censoring),
 - $c^T(j) = 1$ if individual j dies in the disabled state (T_1 or T_2 depending on the definition with which we observe individual j)
- $c^{F_g}(j)$, the cause of exit of disabled state ($F_g, g \in \{\alpha, \beta\}$) for individual j such that:
 - $c^{F_g}(j) = 1$ if individual j dies in disabled state F_g ,
 - $c^{F_g}(j) = 0$ otherwise.

If individual j does not lose autonomy during the observation period, then $\overline{x^T}(j) = \overline{x^H}(j)$. If individual j is observed with definition 2, then $\overline{x^{F_\alpha}}(j) = \overline{x^H}(j)$. If individual j dies in state F_α then $\overline{x^{F_\beta}}(j) = \overline{x^{F_\alpha}}(j) = \overline{x^T}(j)$.

The likelihood associated with one individual observed with definition k is given by

$$\begin{aligned}
 L_j = & \exp \left(- \int_{\overline{x^S(j)}}^{\overline{x^H(j)}} \left(i^k(x) + \mu^{H_k}(x) \right) dx \right) \left(i^k \left(\overline{x^H(j)} \right) \right)^{\mathbb{1}_{\{c^H(j)=2\}}} \left(\mu^{H_k} \left(\overline{x^H(j)} \right) \right)^{\mathbb{1}_{\{c^H(j)=1\}}} \\
 & \exp \left(- \int_0^{\overline{x^{F_\alpha}(j)} - \overline{x^H(j)}} \mu^{F_\alpha} \left(\overline{x^H(j)} + t, t \right) dt \right) \left(\mu^{F_\alpha} \left(\overline{x^{F_\alpha}(j)}, \overline{x^{F_\alpha}(j)} - \overline{x^H(j)} \right) \right)^{\mathbb{1}_{\{c^{F_\alpha}(j)=1\}}} \\
 & \exp \left(- \int_0^{\overline{x^{F_\beta}(j)} - \overline{x^{F_\alpha}(j)}} \mu^{F_\beta} \left(\overline{x^{F_\alpha}(j)} + t, t \right) dt \right) \left(\mu^{F_\beta} \left(\overline{x^{F_\beta}(j)}, \overline{x^{F_\beta}(j)} - \overline{x^{F_\alpha}(j)} \right) \right)^{\mathbb{1}_{\{c^{F_\beta}(j)=1\}}} .
 \end{aligned} \tag{37}$$

Therefore, the total likelihood for all individuals from products 1 and 2 is given by

$$\begin{aligned}
 L_{tot} = & \prod_{j=1}^{n_1} \exp \left(- \int_{\overline{x^S(j)}}^{\overline{x^H(j)}} \left(i^1(x) + \mu^{H_1}(x) \right) dx \right) \left(i^1 \left(\overline{x^H(j)} \right) \right)^{\mathbb{1}_{\{c^H(j)=2\}}} \left(\mu^{H_1} \left(\overline{x^H(j)} \right) \right)^{\mathbb{1}_{\{c^H(j)=1\}}} \\
 & \exp \left(- \int_0^{\overline{x^{F_\alpha}(j)} - \overline{x^H(j)}} \mu^{F_\alpha} \left(\overline{x^H(j)} + t, t \right) dt \right) \left(\mu^{F_\alpha} \left(\overline{x^{F_\alpha}(j)}, \overline{x^{F_\alpha}(j)} - \overline{x^H(j)} \right) \right)^{\mathbb{1}_{\{c^{F_\alpha}(j)=1\}}} \\
 & \exp \left(- \int_0^{\overline{x^{F_\beta}(j)} - \overline{x^{F_\alpha}(j)}} \mu^{F_\beta} \left(\overline{x^{F_\alpha}(j)} + t, t \right) dt \right) \left(\mu^{F_\beta} \left(\overline{x^{F_\beta}(j)}, \overline{x^{F_\beta}(j)} - \overline{x^{F_\alpha}(j)} \right) \right)^{\mathbb{1}_{\{c^{F_\beta}(j)=1\}}} \times \\
 & \prod_{j=1}^{n_2} \exp \left(- \int_{\overline{x^S(j)}}^{\overline{x^H(j)}} \left(i^2(x) + \mu^{H_2}(x) \right) dx \right) \left(i^2 \left(\overline{x^H(j)} \right) \right)^{\mathbb{1}_{\{c^H(j)=2\}}} \left(\mu^{H_2} \left(\overline{x^H(j)} \right) \right)^{\mathbb{1}_{\{c^H(j)=1\}}} \\
 & \exp \left(- \int_0^{\overline{x^{F_\beta}(j)} - \overline{x^H(j)}} \mu^{F_\beta} \left(\overline{x^H(j)} + t, t \right) dt \right) \left(\mu^{F_\beta} \left(\overline{x^{F_\beta}(j)}, \overline{x^{F_\beta}(j)} - \overline{x^H(j)} \right) \right)^{\mathbb{1}_{\{c^{F_\beta}(j)=1\}}} , \tag{38}
 \end{aligned}$$

where $n_k, k \in \{1, 2\}$ denotes the number of observed policyholders from product k .

Let $e_j^{H_k}(p)$ denote the central exposure to risk of individual j in state H_k between integer ages x_p^H and x_{p+1}^H . Let $e_j^g(p, q)$ denote the exposure to risk of individual j in state F_g between integer ages x_p^T and x_{p+1}^T and durations t_q^g and t_{q+1}^g . Let $e^{H_k}(p) = \sum_{j=1}^{n_k} e_j^{H_k}(p)$ be the sum of the central

exposures in state H_k of all policyholders from product k . Let $e^\alpha(p, q) = \sum_{j=1}^{n_1} e_j^\alpha(p, q)$. F_β is the

only state observed in both products. Therefore, let $e^{\beta, k}(p, q) = \sum_{j=1}^{n_k} e_j^\beta(p, q)$ denote the sum of the central exposures in state F_β for product k . We denote $e^\beta(p, q) = e^{\beta, 1}(p, q) + e^{\beta, 2}(p, q)$ as the sum of the central exposures of both products.

Let $N^1(p)$ denote the number of transitions from H_1 to T_1 between x_p^H and x_{p+1}^H . Let $N^2(p)$ denote the number of transitions from H_2 to T_2 between x_p^H and x_{p+1}^H .

Let $D^{H_1}(p)$ denote the number of deaths from the healthy state between x_p^H and x_{p+1}^H in the first type of insurance. Let $D^{H_2}(p)$ denote the number of deaths from the healthy state between x_p^H and x_{p+1}^H in the second type of insurance. Let $D^\alpha(p, q)$ denote the number of deaths from the F_α disabled state with age at death between x_p^T and x_{p+1}^T and duration between t_q^α and t_{q+1}^α . For the central exposure, let $D^{\beta,k}(p, q)$ denote the number of deaths from the F_β disabled state with age at death between x_p^T and x_{p+1}^T and duration between t_q^β and t_{q+1}^β for product k . We denote $D^\beta(p, q) = D^{\beta,1}(p, q) + D^{\beta,2}(p, q)$.

As intensities $\mu^{H_k}(\cdot)$ and $i^k(\cdot)$, $k \in \{1, 2\}$ and $\mu^{F_g}(\cdot)$, $g \in \{\alpha, \beta\}$ are piecewise constants (Assumption 2.3 and Assumption 2.4), we denote

- $\mu_p^{H_k} = \mu^{H_k}(x_p^H)$, $\forall p \in \{1, \dots, M^H\}$,
- $i_p^k = i^k(x_p^H)$, $\forall p \in \{1, \dots, M^H\}$, and
- $\mu_{p,q}^{F_g} = \mu^{F_g}(x_p^F, t_q^g)$, $\forall p \in \{1, \dots, M_x^F\}$, $q \in \{1, \dots, M_t^g\}$,

and Equation 38 becomes

$$\begin{aligned}
 L_{tot} = & \prod_{p=1}^{M^H} \exp(- (i_p^1 + \mu_p^{H_1}) e^{H_1}(p)) (i_p^1)^{N^1(p)} (\mu_p^{H_1})^{D^{H_1}(p)} \\
 & \prod_{p=1}^{M_x^F} \prod_{q=1}^{M_t^\alpha} \exp(-\mu_{p,q}^{F_\alpha} e^\alpha(p, q)) (\mu_{p,q}^{F_\alpha})^{D^\alpha(p,q)} \times \\
 & \prod_{p=1}^{M^H} \exp(- (i_p^2 + \mu_p^{H_2}) e^{H_2}(p)) (i_p^1)^{N^2(p)} (\mu_p^{H_2})^{D^{H_2}(p)} \\
 & \prod_{p=1}^{M_x^F} \prod_{q=1}^{M_t^\beta} \exp(-\mu_{p,q}^{F_\beta} e^\beta(p, q)) (\mu_{p,q}^{F_\beta})^{D^\beta(p,q)}. \tag{39}
 \end{aligned}$$

The last term of Equation 39 represents the likelihood of all observations from products 1 and 2 in state F_β .

The log likelihood function of the combined observations from products 1 and 2 is given by

$$\begin{aligned}
 l_{tot} = & \sum_{p=1}^{M^H} \left(- (i_p^1 + \mu_p^{H_1}) e^{H_1(p)} + N^1(p) \log(i_p^1) + D^{H_1}(p) \log(\mu_p^{H_1}) + \right. \\
 & \sum_{p=1}^{M_x^F} \sum_{q=1}^{M_t^\alpha} \left(-\mu_{p,q}^{F_\alpha} e^\alpha(p, q) + D^\alpha(p, q) \log(\mu_{p,q}^{F_\alpha}) + \right. \\
 & \sum_{p=1}^{M^H} \left(- (i_p^2 + \mu_p^{H_2}) e^{H_2(p)} + N^2(p) \log(i_p^1) + D^{H_2}(p) \log(\mu_p^{H_2}) + \right. \\
 & \left. \left. \sum_{p=1}^{M_x^F} \sum_{q=1}^{M_t^\beta} \left(-\mu_{p,q}^{F_\beta} e^\beta(p, q) + D^\beta(p, q) \log(\mu_{p,q}^{F_\beta}) \right) \right) \right). \tag{40}
 \end{aligned}$$

B Proof of the constraint

Let us prove the constraint given by Equation (9) in Section 3.2.

$$\mu_p^{H_2} = \mu^{H_2}(x_p^H) = \lim_{h \rightarrow 0} \frac{\mathbb{P}(X_{x_p^H+h}^{(2)} = \text{Death} | X_{x_p^H}^{(2)} = H_2)}{h}. \quad (41)$$

$$\begin{aligned} \mathbb{P}(X_{x+h}^{(2)} = \text{Death} | X_x^{(2)} = H_2) &= \mathbb{P}(X_{x+h}^{(1)} = \text{Death} | X_x^{(1)} = H_1 \cup X_x^{(1)} = F_\alpha) \\ &= \frac{\mathbb{P}(X_{x+h}^{(1)} = \text{Death}, \{X_x^{(1)} = H_1 \cup X_x^{(1)} = F_\alpha\})}{\mathbb{P}(X_x^{(1)} = H_1 \cup X_x^{(1)} = F_\alpha)} \\ &= \frac{\mathbb{P}(X_{x+h}^{(1)} = \text{Death}, X_x^{(1)} = H_1) + \mathbb{P}(X_{x+h}^{(1)} = \text{Death}, X_x^{(1)} = F_\alpha)}{\mathbb{P}(X_x^{(1)} = H_1 \cup X_x^{(1)} = F_\alpha)} \\ &= \frac{\mathbb{P}(X_{x+h}^{(1)} = \text{Death} | X_x^{(1)} = H_1) \mathbb{P}(X_x^{(1)} = H_1)}{\mathbb{P}(X_x^{(1)} = H_1 \cup X_x^{(1)} = F_\alpha)} + \\ &\quad \sum_{q=1}^{M_t^\alpha} \frac{\mathbb{P}(X_{x+h}^{(1)} = \text{Death} | \{X_x^{(1)} = F_\alpha, X_{(x-t_q^\alpha)}^{(1)} = F_\alpha, X_{(x-t_{q+1}^\alpha)}^{(1)} = H_1\})}{\mathbb{P}(X_x^{(1)} = H_1 \cup X_x^{(1)} = F_\alpha)} \times \\ &\quad \mathbb{P}(X_x^{(1)} = F_\alpha, X_{(x-t_q^\alpha)}^{(1)} = F_\alpha, X_{(x-t_{q+1}^\alpha)}^{(1)} = H_1) \\ &= \frac{\mathbb{P}(X_{x+h}^{(1)} = \text{Death} | X_x^{(1)} = H_1) \mathbb{P}(X_x^{(1)} = H_1)}{\mathbb{P}(X_x^{(1)} = H_1 \cup X_x^{(1)} = F_\alpha)} + \\ &\quad \sum_{q=1}^{M_t^\alpha} \frac{\mathbb{P}(X_{x+h}^{(1)} = \text{Death} | \{X_x^{(1)} = F_\alpha, X_{(x-t_q^\alpha)}^{(1)} = F_\alpha, X_{(x-t_q^\alpha)^-} = H_1\})}{\mathbb{P}(X_x^{(1)} = H_1 \cup X_x^{(1)} = F_\alpha)} \times \\ &\quad \mathbb{P}(X_x^{(1)} = F_\alpha, X_{(x-t_q^\alpha)}^{(1)} = F_\alpha, X_{(x-t_{q+1}^\alpha)}^{(1)} = H_1). \end{aligned}$$

Therefore, Equation (41) becomes

$$\begin{aligned}
\mu_p^{H_2} &= \lim_{h \rightarrow 0} \frac{\mathbb{P}(X_{x_p^H+h}^{(2)} = \text{Death} | X_{x_p^H}^{(2)} = H_2)}{h} \\
&= \frac{\mathbb{P}(X_x^{(1)} = H_1)}{\mathbb{P}(X_x^{(1)} = H_1 \cup X_x^{(1)} = F_\alpha)} \lim_{h \rightarrow 0} \frac{\mathbb{P}(X_{x_p^H+h}^{(1)} = \text{Death} | X_{x_p^H}^{(1)} = H_1)}{h} + \\
&\quad \sum_{q=1}^{M_t^\alpha} \frac{\mathbb{P}(X_x^{(1)} = F_\alpha, X_{(x-t_q^\alpha)}^{(1)} = F_\alpha, X_{(x-t_{q+1}^\alpha)}^{(1)} = H_1)}{\mathbb{P}(X_x^{(1)} = H_1 \cup X_x^{(1)} = F_\alpha)} \times \\
&\quad \lim_{h \rightarrow 0} \frac{\mathbb{P}(X_{x+h}^{(1)} = \text{Death} | \{X_x^{(1)} = F_\alpha, X_{(x-t_q^\alpha)}^{(1)} = F_\alpha, X_{(x-t_q^\alpha)-}^{(1)} = H_1\})}{h} \\
&= \frac{\mathbb{P}(X_x^{(1)} = H_1)}{\mathbb{P}(X_x^{(1)} = H_1 \cup X_x^{(1)} = F_\alpha)} \mu_p^{H_1} + \sum_{q=1}^{M_t^\alpha} \frac{\mathbb{P}(X_x^{(1)} = F_\alpha, X_{(x-t_q^\alpha)}^{(1)} = F_\alpha, X_{(x-t_{q+1}^\alpha)}^{(1)} = H_1)}{\mathbb{P}(X_x^{(1)} = H_1 \cup X_x^{(1)} = F_\alpha)} \mu_{p,q}^{F_\alpha}. \quad (42)
\end{aligned}$$

C Structure of matrix C

Let E^{H_1} and $E^{H_2} \in \mathcal{M}_{M^H}(\mathbb{R})$ be the diagonal matrices of the central exposures in states H_1 and H_2 , respectively. Let $E^\alpha \in \mathcal{M}_{M_x^F, M_t^\alpha}(\mathbb{R})$ and $E^\beta \in \mathcal{M}_{M_x^F, M_t^\beta}(\mathbb{R})$ be the matrix of the sum of the central exposures in F_α and F_β , respectively. We denote ${}^v E_{.,k}^g = \text{diag}(E_{.,k}^g) \in \mathcal{M}_{M_x^F}(\mathbb{R})$, $g \in \{\alpha, \beta\}$ as the diagonal matrix of the central exposures in state F_g for the k^{th} duration.

$$\bullet C = \begin{bmatrix} C_0 \\ C_1 \end{bmatrix} \in \mathcal{M}_{(M^H + M^H + (M_t^\alpha + M_t^\beta) \cdot M_x^F), (M^H + (M_t^\alpha + M_t^\beta) \cdot M_x^F)}(\mathbb{R}),$$

$$\bullet C_0 = \begin{bmatrix} E^{H_1} & 0 & \dots & \dots & \dots & \dots & \dots & \dots & 0 \\ 0 & {}^v E_{.,1}^\alpha & 0 & \dots & \dots & \dots & \dots & \dots & 0 \\ 0 & 0 & \ddots & 0 & \dots & \dots & \dots & \dots & 0 \\ 0 & \dots & 0 & \ddots & 0 & \dots & \dots & \dots & 0 \\ 0 & \dots & \dots & 0 & {}^v E_{.,M_t^\alpha}^\alpha & 0 & \dots & \dots & 0 \\ 0 & \dots & \dots & \dots & 0 & {}^v E_{.,1}^\beta & 0 & \dots & 0 \\ 0 & \dots & \dots & \dots & \dots & 0 & \ddots & 0 & 0 \\ 0 & \dots & \dots & \dots & \dots & \dots & \dots & \ddots & 0 \\ 0 & \dots & \dots & \dots & \dots & \dots & \dots & 0 & {}^v E_{.,M_t^\beta}^\beta \end{bmatrix} \in \mathcal{M}_{(M^H + (M_t^\alpha + M_t^\beta) \cdot M_x^F)}(\mathbb{R}),$$

$$\bullet C_1 = \begin{bmatrix} \frac{E^{H_2} \cdot E^{H_1}}{E^{H_1} + \sum_{q=1}^{M_t^\alpha} {}^v E_{.,q}^\alpha} & \frac{E^{H_2} \cdot {}^v E_{.,1}^\alpha}{E^{H_1} + \sum_{q=1}^{M_t^\alpha} {}^v E_{.,q}^\alpha} & \dots & \frac{E^{H_2} \cdot {}^v E_{.,M_t^\alpha}^\alpha}{E^{H_1} + \sum_{q=1}^{M_t^\alpha} {}^v E_{.,q}^\alpha} & 0 & \dots & 0 \end{bmatrix}$$

$$C_1 = (E^{H_1} + \sum_{q=1}^{M_t^\alpha} {}^v E_{.,q}^\alpha)^{-1} \cdot E^{H_2} \cdot \begin{bmatrix} E^{H_1} & {}^v E_{.,1}^\alpha & \dots & {}^v E_{.,M_t^\alpha}^\alpha & 0 & \dots & 0 \end{bmatrix}.$$

The matrix $C_1 \in \mathcal{M}_{M^H, (M^H + (M_t^\alpha + M_t^\beta) \cdot M_x^F)}(\mathbb{R})$ is composed of "concatenation" of $M_t^\beta + 1$ diagonal matrices of dimension M^H , and the matrix is completed with columns of 0.

D Proof of the mortality rates in state F in the application to the problem of the deferred period

Let $X_t^{(1)} \in \{H_1, F_\alpha, F_\beta, Death\}$ denote the health state at time t observed with LTC product 1, as modelled with the diagram in Figure 1a. Let $F = \{F_\alpha \cup F_\beta\}$

Let $X_t^{(2)} \in \{H_2, F_\beta, Death\}$ denote the health state at time t observed with LTC product 1, as modelled with the diagram in Figure 1a.

Let $\tau_k = \min\{u > \tau_{k-1} | X_u \neq X_{\tau_{k-1}}\}$ denote the sequence of jump times, with $\tau_0 = 0$. Let $Z_k = X_{\tau_k}$ be the sequence of visiting states.

$$\begin{aligned}
 \mu^F(x, t) &= \lim_{h \rightarrow 0} \frac{1}{h} \mathbb{P} \left(X_{x+h}^{(1)} = Death | X_x^{(1)} \in F, X_{x-t}^{(1)} \in F, X_{(x-t)^-}^{(1)} = H_1 \right) \\
 &= \lim_{h \rightarrow 0} \mathbb{P} \left(X_{x+h}^{(1)} = Death | X_x^{(1)} \in F, X_{\tau_1}^{(1)} \in F, \tau_1 = (x-t), X_0 = H_1 \right) \\
 &= \lim_{h \rightarrow 0} \frac{\mathbb{P} \left(X_{x+h}^{(1)} = Death, X_x^{(1)} \in F | X_{\tau_1}^{(1)} \in F, \tau_1 = (x-t), X_0 = H_1 \right)}{\mathbb{P} \left(X_x^{(1)} \in F | X_{\tau_1}^{(1)} \in F, \tau_1 = (x-t), X_0 = H_1 \right)} \\
 &= \lim_{h \rightarrow 0} \frac{\mathbb{P} \left(X_{x+h}^{(1)} = Death, X_x^{(1)} \in F | X_{\tau_1}^{(1)} \in F, \tau_1 = (x-t), X_0 = H_1 \right)}{\mathbb{P} \left(X_x^{(1)} \in F | X_{\tau_1}^{(1)} \in F, \tau_1 = (x-t), X_0 = H_1 \right)} \\
 &= \lim_{h \rightarrow 0} \frac{\mathbb{P} \left(X_{x+h}^{(1)} = Death, X_x^{(1)} = F_\alpha | X_{\tau_1}^{(1)} \in F, \tau_1 = (x-t), X_0 = H_1 \right) + \mathbb{P} \left(X_{x+h}^{(1)} = Death, X_x^{(1)} = F_\beta | X_{\tau_1}^{(1)} \in F, \tau_1 = (x-t), X_0 = H_1 \right)}{\mathbb{P} \left(X_x^{(1)} = F_\alpha | X_{\tau_1}^{(1)} \in F, \tau_1 = (x-t), X_0 = H_1 \right) + \mathbb{P} \left(X_x^{(1)} = F_\beta | X_{\tau_1}^{(1)} \in F, \tau_1 = (x-t), X_0 = H_1 \right)} \\
 &= \lim_{h \rightarrow 0} \frac{\mathbb{P} \left(X_{x+h}^{(1)} = Death, X_x^{(1)} = F_\beta | X_{\tau_1}^{(1)} \in F, \tau_1 = (x-t), X_0 = H_1 \right)}{\mathbb{P} \left(X_x^{(1)} = F_\alpha | X_{\tau_1}^{(1)} \in F, \tau_1 = (x-t), X_0 = H_1 \right) + \mathbb{P} \left(X_x^{(1)} = F_\beta | X_{\tau_1}^{(1)} \in F, \tau_1 = (x-t), X_0 = H_1 \right)}. \tag{43}
 \end{aligned}$$

In the case of the deferred period, the transition rate from H_1 to F_β is equal to 0. Therefore,

$$\{X_{\tau_1}^{(1)} \in F, \tau_1 = (x-t), X_0 = H_1\} = \{X_{\tau_1}^{(1)} = F_\alpha, \tau_1 = (x-t), X_0 = H_1\}.$$

Additionally, as no return to a better health state is possible,

$$\mathbb{P} \left(X_x^{(1)} = F_\alpha | X_{\tau_1}^{(1)} = F_\alpha, \tau_1 = (x-t), X_0 = H_1 \right) = \mathbb{P} \left(\tau_2 > x | X_{\tau_1}^{(1)} = F_\alpha, \tau_1 = (x-t), X_0 = H_1 \right).$$

Moreover, from Equation 30

$$\omega^1(x, t) = \begin{cases} 0, & \text{if } t \neq fr/12 \\ +\infty, & \text{if } t = fr/12. \end{cases}$$

Therefore,

$$\begin{aligned}
 \mathbb{P}\left(\tau_2 > x | X_{\tau_1}^{(1)} = F_\alpha, \tau_1 = (x-t), X_0 = H_1\right) &= \exp\left(-\int_{x-t}^x (\mu^{F_\alpha}(u, u-(x-t)) + \omega^1(u, u-(x-t))) du\right) \\
 &= \exp\left(-\int_0^t (\mu^{F_\alpha}(x-t+y, y) + \omega^1(x-t+y, y)) dy\right) \\
 &= \begin{cases} 0 & \text{if } t \geq fr/12 \\ \exp\left(-\int_0^t \mu^{F_\alpha}(x-t+y, y) du\right) & \text{if } t < fr/12. \end{cases}
 \end{aligned} \tag{44}$$

Another consequence of the no return to a better health state is that

$$\begin{aligned}
 \mathbb{P}\left(X_x^{(1)} = F_\beta | X_{\tau_1}^{(1)} = F_\alpha, \tau_1 = (x-t), X_0 = H_1\right) &\leq \mathbb{P}\left(\tau_2 \leq x, X_{\tau_2}^{(1)} = F_\beta | X_{\tau_1}^{(1)} = F_\alpha, \tau_1 = (x-t), X_0 = H_1\right) \\
 &= \int_0^t \mathbb{P}\left(\tau_2 \geq x-t+y | X_{\tau_1}^{(1)} = F_\alpha, \tau_1 = (x-t), X_0 = H_1\right) \times \\
 &\quad \delta_{fr/12}(y) dy,
 \end{aligned}$$

$$\text{where } \delta_{fr/12}(y) = \begin{cases} +\infty & \text{if } t = fr/12 \\ 0 & \text{otherwise.} \end{cases}$$

Therefore,

$$\mathbb{P}\left(X_x^{(1)} = F_\beta | X_{\tau_1}^{(1)} = F_\alpha, \tau_1 = (x-t), X_0 = H_1\right) = 0, \forall t < fr/12. \tag{45}$$

Then,

- if $t < fr/12$, from Equation 43 and Equation 45

$$\begin{aligned}
 \mu^F(x, t) &= \lim_{h \rightarrow 0} \frac{1}{h} \frac{\mathbb{P}\left(X_{x+h}^{(1)} = Death, X_x^{(1)} = F_\alpha | X_{\tau_1}^{(1)} = F_\alpha, \tau_1 = (x-t), X_0 = H_1\right)}{\mathbb{P}\left(X_x^{(1)} = F_\alpha | X_{\tau_1}^{(1)} = F_\alpha, \tau_1 = (x-t), X_0 = H_1\right)} \\
 &= \lim_{h \rightarrow 0} \frac{1}{h} \mathbb{P}\left(X_{x+h}^{(1)} = Death | X_x^{(1)} = F_\alpha, X_{\tau_1}^{(1)} = F_\alpha, \tau_1 = (x-t), X_0 = H_1\right) \\
 &= \mu^{F_\alpha}(x, t),
 \end{aligned}$$

- if $t \geq fr/12$, from Equation 43 and Equation 44

$$\begin{aligned}
 \mu^F(x, t) &= \lim_{h \rightarrow 0} \frac{1}{h} \frac{\mathbb{P}\left(X_{x+h}^{(1)} = Death, X_x^{(1)} = F_\beta | X_{\tau_1}^{(1)} = F_\alpha, \tau_1 = (x-t), X_0 = H_1\right)}{\mathbb{P}\left(X_x^{(1)} = F_\beta | X_{\tau_1}^{(1)} = F_\alpha, \tau_1 = (x-t), X_0 = H_1\right)} \\
 &= \lim_{h \rightarrow 0} \frac{1}{h} \mathbb{P}\left(X_{x+h}^{(1)} = Death | X_x^{(1)} = F_\beta, X_{\tau_1}^{(1)} = F_\alpha, \tau_1 = (x-t), X_0 = H_1\right).
 \end{aligned}$$

As the transition from F_α to F_β can occur only fr months after entry into F_α , and no return to a better health state is possible, then

$$\begin{aligned} & \mathbb{P} \left(X_{x+h}^{(1)} = Death | X_x^{(1)} = F_\beta, X_{\tau_1}^{(1)} = F_\alpha, \tau_1 = (x-t), X_0 = H_1 \right) = \\ & \mathbb{P} \left(X_{x+h}^{(1)} = Death | X_x^{(1)} = F_\beta, X_{\tau_2}^{(1)} = F_\beta, \tau_2 = (x-t+fr/12), X_{\tau_1}^{(1)} = F_\alpha, \tau_1 = (x-t), X_0 = H_1 \right). \end{aligned}$$

Therefore,

$$\begin{aligned} \mu^F(x, t) &= \lim_{h \rightarrow 0} \frac{1}{h} \mathbb{P} \left(X_{x+h}^{(1)} = Death | X_x^{(1)} = F_\beta, X_{\tau_2}^{(1)} = F_\beta, \tau_2 = (x-t+fr/12) \right) \\ &= \mu^{F_\beta}(x, t - fr/12). \end{aligned}$$

E Details on the structure of the matrix B^F for the application to the deferred period

Let $B_t^\alpha \in \mathcal{M}_{M_t^\alpha, J_t^\alpha}$ be the matrix of the values of the splines at each subdivision point before the deferred period ($\{t_1^\alpha, \dots, t_{M_t^\alpha+1}^\alpha\}$). J_t^α denotes the number of splines from B_t^α having a non-null value on the interval $\left[t_1^\alpha; t_{M_t^\alpha}^\alpha\right]$.

Let $B_t^\beta \in \mathcal{M}_{M_t^\beta, J_t^\beta}$ be the matrix of the values of the splines at each subdivision point after the deferred period ($\left\{t_1^\beta + \frac{fr}{12}, \dots, t_{M_t^\beta+1}^\beta + \frac{fr}{12}\right\}$). J_t^β denotes the number of splines from B_t^β having a non-null value on the interval $\left[t_1^\beta + \frac{fr}{12}; t_{M_t^\beta+1}^\beta + \frac{fr}{12}\right]$.

Let s denote the number of splines shared by states F_α and F_β ($s = 4$ in Figure 5). The first s columns of B_t^β are the continuous extension of the last s columns of B_t^α . Let $B_t^g(s), g \in \{\alpha, \beta\}$ denote the submatrix of B_t^g composed of only splines shared by F_α and F_β . Let $B_t^g(-s), g \in \{\alpha, \beta\}$ denote the submatrix of B_t^g composed of only splines not shared by F_α and F_β .

$$\text{Then, } B_t^F = \left[\begin{array}{c|c|c} B_t^\alpha(-s) & B_t^\alpha(s) & 0 \\ 0 & B_t^\beta(s) & B_t^\beta(-s) \end{array} \right].$$

Therefore, from Section 3.1, the matrix B^F for the estimation of the mortality in two dimensions in LTC is given by

$$B^F = B_t^F \otimes B_x^F \tag{46}$$

$$\left[\begin{array}{c|c|c} B_t^\alpha(-s) \otimes B_x^F & B_t^\alpha(s) \otimes B_x^F & 0 \\ 0 & B_t^\beta(s) \otimes B_x^F & B_t^\beta(-s) \otimes B_x^F \end{array} \right]. \tag{47}$$

F Additional plots

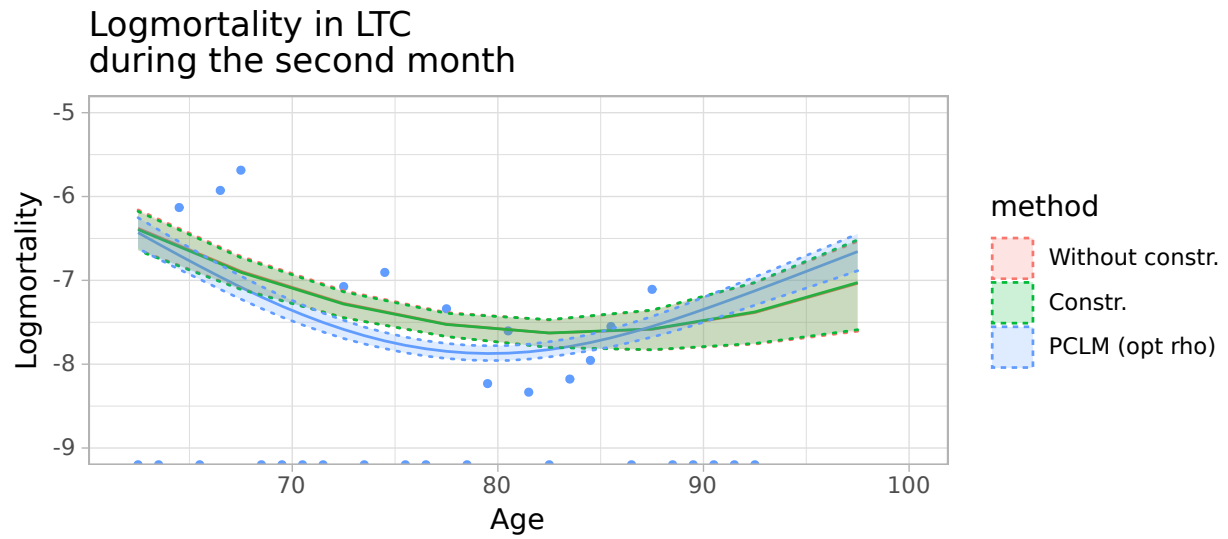


Figure 14: Mortality of disabled policyholders during the second month following the loss of autonomy

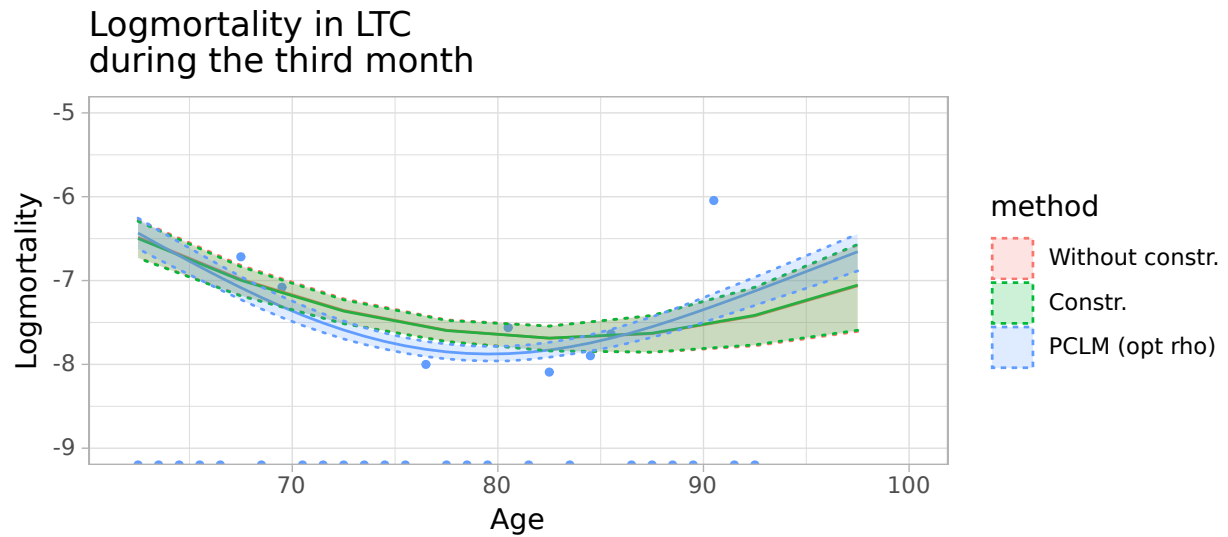


Figure 15: Mortality of disabled policyholders during the third month following the loss of autonomy

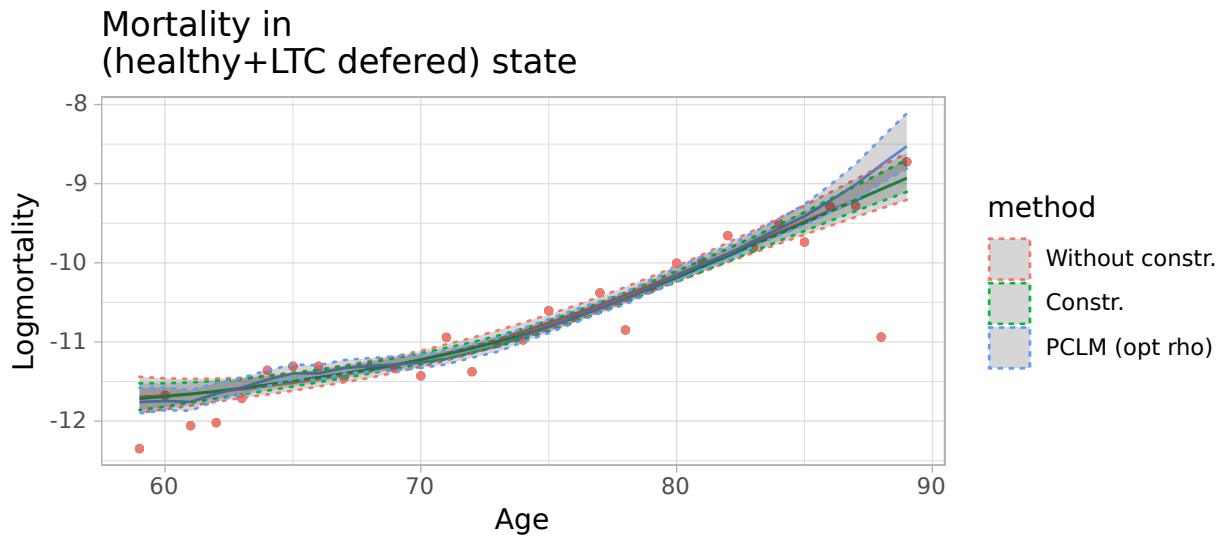


Figure 16: Mortality in state H_2

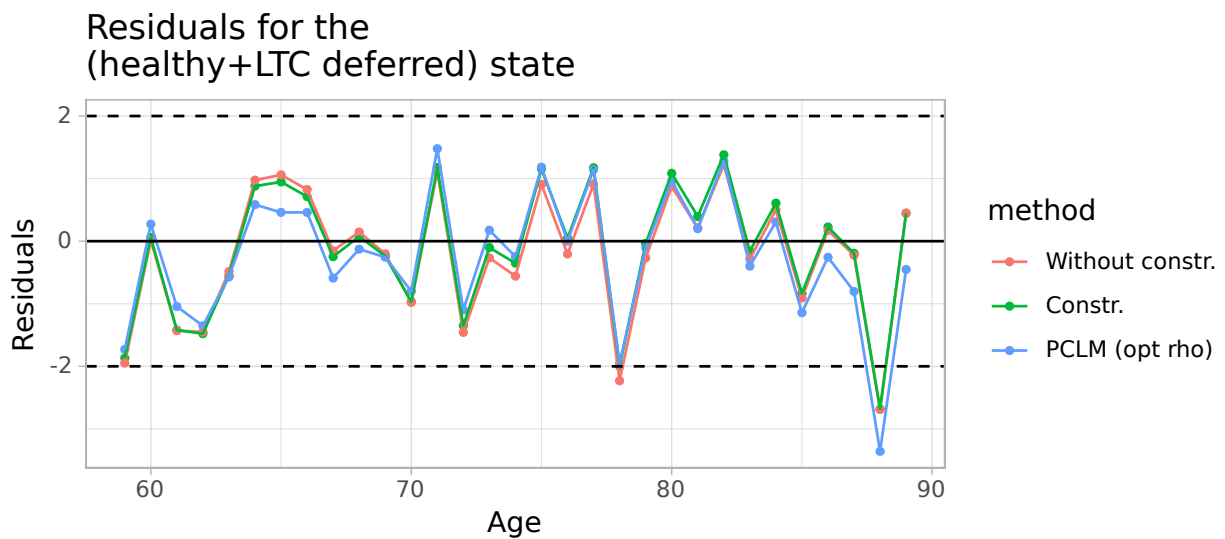


Figure 17: Deviance residuals in the state H_2

References

- Beard, R. E. (1959). *Appendix: Note on Some Mathematical Mortality Models*, pp. 302–311. John Wiley & Sons, Ltd.
- Biessy, G. (2016, August). A semi-Markov model with pathologies for Long-Term Care Insurance. working paper or preprint.
- Boggs, P. T. and J. W. Tolle (1995). Sequential quadratic programming. *Acta numerica* 4, 1–51.
- Camarda, C. G., P. H. C. Eilers, and J. Gampe (2016). Sums of smooth exponentials to decompose complex series of counts. *Statistical Modelling* 16, 279 – 296.
- Chapman, C. (2012). A conversation on responsibility. *The Actuary, Society of Actuaries* 8(6), 8–10.
- Colombo, F., A. Llana-Nozal, J. Mercier, and F. Tjadens (2008). Medicaid eligibility issues for long-term care insurance partnership programs. *Center for Health Care Strategies Long-term Care Partnership Expansion Project March*.
- Colombo, F., A. Llana-Nozal, J. Mercier, and F. Tjadens (2011). Help wanted? providing and paying for long-term care. *OECD Health Policy Studies*. OECD Publishing, Paris, France.
- Currie, I. D. and M. Durban (2002). Flexible smoothing with P-splines: A unified approach. *Statistical Modelling* 2(4), 333–349.
- Dupourqué, E. (2012). AGGIR, the work of grids. *Long-Term Care News* 32.
- Eilers, P. (2007, 10). Ill-posed problems with counts, the composite link model, and penalized likelihood. *Statistical Modelling* 7.
- Eilers, P. H. C. and B. D. Marx (1996). Flexible smoothing with B-splines and penalties. *Statistical Science* 11(2), 89 – 121.
- H, J., S. Y, P. Y, P. S, C. X, L. X, W. D, L. Y, L. H, and W. C (2023). Projecting long-term care costs for home and community-based services in china from 2005 to 2050. *Journal of the American Medical Directors Association* 24(2), 228–234.
- Haberman, S. and E. Pitacco (1999). *Actuarial models for disability insurance*. FL: Chapmanand Hall/CRC Press.
- Hurd, M. D., P. C. Michaud, and S. Rohwedder (2017). Distribution of lifetime nursing home use and of out-of-pocket spending. *Proceedings of the National Academy of Sciences of the United States of America* 114(37), 9838–9842.
- Jelmer Ypma, S. G. J. (2022). R interface to nlopt. *CRAN*.
- Johnson, R. W. (2019). What is the lifetime risk of needing and receiving long-term services and supports? *Assistant Secretary for Planning and Evaluation Reports April*.

-
- Kaye, H. S., C. Harrington, and M. P. LaPlante (2010). Long-term care: who gets it, who provides it, who pays, and how much? *Health affairs (Project Hope)* 29(1), 11–21.
- Kemper, P., H. L. Komisar, and L. Alecxih (2005). Long-term care over an uncertain future: what can current retirees expect? *Inquiry: a journal of medical care organization, provision and financing* 42(4), 335–350.
- Kraft, D. (1988). *A Software Package for Sequential Quadratic Programming*. Deutsche Forschungs- und Versuchsanstalt für Luft- und Raumfahrt Köln: Forschungsbericht. Wiss. Berichtswesen d. DFVLR.
- Macdonald, A. S., S. J. Richards, and I. D. Currie (2018). *Modelling Mortality With Actuarial Applications*. Cambridge University Press.
- Marx, B. D. and P. H. Eilers (1998). Direct generalized additive modeling with penalized likelihood. *Computational Statistics & Data Analysis* 28(2), 193 – 209.
- McCullagh, P. (2019). *Generalized linear models*. Routledge.
- Nelder, J. A. and R. W. M. Wedderburn (1972). Generalized linear models. *Journal of the Royal Statistical Society. Series A (General)* 135(3), 370–384.
- OECD (2020). Spending on long-term care. *OECD Report Nov*.
- Or, Z. and A. Penneau (2021). *Pricing long-term care for older persons*, Chapter Case Study - France. WHO Centre for Health Development.
- Perks, W. (1932). On some experiments in the graduation of mortality statistics. *Journal of the Institute of Actuaries* 63, 12–57.
- Pritchard, D. J. (2006). Modeling disability in long-term care insurance. *North American Actuarial Journal* 10(4), 48–75.
- Productivity Commission of Australia (2013). An ageing australia: Preparing for the future. *Commission Research Paper*.
- Remund, A., C. Camarda, and T. Riffe (2018, June). A cause-of-death decomposition of young adult excess mortality. *Demography* 55(3), 957–978.
- Remund, A., T. Riffe, and C. Camarda (2017, 01). A cause-of-death decomposition of the young adult mortality hump. *Demography* 55.
- Rizzi, S., J. Gampe, and P. H. C. Eilers (2015, 06). Efficient Estimation of Smooth Distributions From Coarsely Grouped Data. *American Journal of Epidemiology* 182(2), 138–147.
- Ruppert, D. (2002). Selecting the number of knots for penalized splines. *Journal of Computational and Graphical Statistics* 11(4), 735–757.
- Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, 461–464.
-

- Shao, A. W., H. Chen, and M. Sherris (2019). To borrow or insure? long term care costs and the impact of housing. *Insurance: Mathematics and Economics* 85, 15–34.
- Shao, A. W., M. Sherris, and J. H. Fong (2017). Product pricing and solvency capital requirements for long-term care insurance. *Scandinavian Actuarial Journal* 2017(2), 175–208.
- Shi, P. and W. Zhang (2013). Managed Care and Health Care Utilization: Specification of Bivariate Models Using Copulas. *North American Actuarial Journal* 17(4), 306–324.
- Thompson, R. and R. J. Baker (1981). Composite link functions in generalized linear models. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 30(2), 125–131.