A semi-Markov model with pathologies for Long-Term Care Insurance

Guillaume Biessy

A semi-Markov model with pathologies for Long-Term Care Insurance

Guillaume BIESSY[1]

SCOR Global Life [2] - LaMME[3]

Abstract

Most Long-Term Care (LTC) Insurance products rely on definitions for functional disability based on the Activities of Daily Living (ADL). While functional disability may reflect the level of care required by the insured life, it is not on its own a good predictor of lifespan in LTC, which strongly depends on the underlying pathology responsible for disability. Indeed, cancer and respiratory diseases are associated with extremely short lifespan while dementia and neurological diseases make for much longer claims. Pathologies are therefore responsible for heterogeneity in the data, which makes estimation of mortality in LTC a difficult issue. As a consequence the associated literature is still scarce.

In this paper, we study the mortality in LTC associated with 4 different groups of pathologies: *cancer*, *dementia*, *neurological diseases* and *other causes* based on data from a French LTC portfolio. We consider a semi-Markov framework, where mortality in LTC depends on both age at claim inception and time already spent in LTC. We first derive the incidence rate in LTC and mortality rate associated with each group of pathologies and for both males and females. To do so, we rely on local likelihood methods that we apply directly to transition intensities of the model. We then combine those transition intensities to get a second-step estimator of the overall mortality in LTC, which proves more accurate than a direct estimate regardless of the pathology. Finally our results indicates that the peak of mortality following entry in LTC observed in the data is mostly due to the *cancer* group.

Keywords: Long-Term Care Insurance, continuous time semi-Markov process, competing risks, local likelihood

1 Introduction

Long-Term Care (LTC) Insurance providers often base their guarantees on the notion of Activities of Daily Living to describe the level of functional disability reached by their insured lives. The main advantages of this approach is that it is easy to understand for the insured life, relatively easy to assess for the insurer and it has strong links with the level of care required by the insured life. Therefore, to better describe the process of LTC, a natural approach is to consider models that take into account several levels of functional disability. Multi-states models have been identified early as an adequate tool to study the risk, as in Haberman and Pitacco (1998) or Denuit and Robert (2007). However, LTC products are relatively recent, and the available data is still scarce. Alternatively, Czado and Rudolph (2002) provide an interesting study of the impact of covariates such as gender and level of care on the risk based on a German LTC portfolio and using Cox proportional hazard model (as in Cox, 1992). While Markov models allow for an impact of the age of the insured life on the risk, they generally ignore the impact of the duration in the LTC state. Nevertheless, Helms et al. (2005) partially overcome this limit by considering number of years already spent in LTC as a covariate in a Markov framework using the Aalen-Johansen estimator (one can refer to Aalen and Johansen, 1978, for a description of the estimator) also based on a German insurance portfolio. The semi-Markov framework, where mortality in LTC depends on both age of entry and duration in LTC, has only been applied very recently to LTC. Using data from the French public aid for LTC, Lepez et al. (2013) and Biessy (2015b) rely on a multi-state model for the mortality in LTC with 4 states of functional disability. However, the short observation period and specificities of the methodology used to gather the data makes estimation very difficult.

---

[1]Contact: gbiessy@scor.com.

[2]SCOR Global Life SE - 5 avenue Kléber, 75795 Paris Cedex 16, France.

[3]Laboratoire de Mathématiques et Modélisation d'Évry - Université d'Évry Val d'Essonne, UMR CNRS 8071, USC INRA, Évry, France.

Functional disability is only a descriptive indicator of the insured life health status while LTC has a variety of causes, the most frequent being cancer, dementia, neurological diseases, cardiovascular diseases, muscular and skeletal diseases. Among less frequent causes one can find respiratory diseases, blindness, depression, accidents, diabetes and cirrhosis. This results in a large amount of heterogeneity within the disabled population and lifespan in LTC strongly depends on the underlying cause. Therefore causes should be included in LTC models when data is available. Guibert and Planchet (2014) study the incidence rates associated with 4 groups of pathology, based on a French LTC portfolio. Using the same data, Guibert and Planchet (2015) computes non-parametric estimate of mortality rates in LTC for each group of pathology using the Aalen-Johansen estimator. Data containing information about pathologies is however extremely rare. When this information is missing, Biessy (2015a) suggests to introduce nonetheless a mixture model and derives a parametric model for the mortality in LTC which proves a reasonable fit to the data considered.

In this article, we propose to estimate the incidence rates and the mortality in LTC associated with 4 different groups of pathology: *cancer*, *dementia*, *neurological diseases* and *other causes*. We then recombine the results in order to get a second-step estimator of the global mortality in LTC. We rely on data from a French LTC portfolio with close to 20,000 observed claims. We consider the framework of the illness death model with no return (see Pitacco, 2014, for an extensive study of the illness-death model and its uses in health and life insurance) and consider that mortality in LTC is a function of both age at entry and time spent in LTC, as in Biessy (2015a). To estimate the mortality in LTC, we propose to rely on local likelihood methods first introduced in Tibshirani and Hastie (1987). The interested reader can refer to the excellent book from Loader (1999) for an exhaustive description of those methods and their various applications. An application to the estimation of mortality in LTC is presented in Tomas and Planchet (2013), who compare local likelihood methods with adaptive bandwidth to non-adaptive local likelihood or B-splines methods (one can refer to Eilers and Marx, 1996, for a description of B-splines). In this article, we propose to apply local likelihood methods for the direct estimation of continuous time transition intensities rather than for the smoothing of empirical estimates. While, this approach is partially described in Chapter 7 of Loader (1999), it has received very little interest and to the best of our knowledge has not yet been used for insurance portfolio studies. A strong constraint of this approach is that an access to individual data is required. Also, a small extension of the original model provided by Loader (1999) is required to handle left-truncated data.

Section 2 of the article provides a general description of the data at hand, and introduces the framework for the estimation of incidence and mortality in LTC. We first consider a multi-state model with autonomy, death and 4 states of LTC corresponding to the 4 groups of pathologies available in the data and provide a formula to compute the overall mortality in LTC from incidence and mortality rates for each group. Section 3 introduces the local likelihood method and apply it for the estimation of autonomous mortality, incidence in LTC and the 2-dimension surface of mortality in LTC. Section 4 highlights some of the results obtained about the impact of causes on the risk and results on the second-step estimate of mortality we get by recombining all causes. Lastly, section 5 summarizes the results of the article and discusses limits and potential improvements of our approach.

2 Data and model

2.1 Data at hand

The data used in this article comes from a French LTC portfolio. The associated policy is triggered when insured lives have lost the ability to perform on their own at least 3 of the 4 Activities of Daily Living as given in the product definition: eating, bathing, clothing and functional mobility. We consider an observation period from 1998 to 2015 included. For each individual, the following information is available: date of birth, date of subscription, date of entry in LTC, cause of entry in LTC, date of death and gender. Causes of entry in LTC are clustered in 4 groups. The first group contains cases of dementia, mainly caused by Alzheimer disease. The second group gathers neurological disease, among them Parkinson disease as well as sclerosis. The third group contains cases of cancer: tumours and lymphocyte. Lastly, the fourth group gathers claims that do not fit in the previous 3 categories. Major causes for LTC that have not been mentioned previously include cardiovascular diseases such as infarction or stroke, muscular and skeletal diseases such as arthrosis or arthritis. Cases where the need for LTC arises from multiple pathologies also fall in this category. Other minor causes for LTC include accidents, depression, blindness, cirrhosis, HIV, diabetes. We therefore expect a lot of residual heterogeneity in this last group. The distribution of pathologies in claims can be found in Table 1. *Dementia* appears to

be the most frequent cause of claim for both genders followed by *neurological diseases* and *other causes.* *Cancer* is less frequent but the amount of available data is still large with more than 900 uncensored claims for each gender.

Pathology	Number of claims		Censoring rate	
	male	female	male	female
Dementia	2,584	7,007	23.5 %	37.5 %
Neurological	1,362	2,194	19.9 %	30.3 %
Cancer	982	1,024	4.5 %	6.3 %
Other	1,207	2,752	17.1 %	28.2 %
Total	**14,213**	**6,775**	**18,9 %**	**33,3 %**

Table 1: Distribution of claims.

2.2 Notation

In this section we first introduce the 3-state illness-death model used to describe the LTC process. We then consider a multi state process with p states of LTC corresponding to p different causes. The application is based on the previously introduced data which corresponds to the case $p = 4$. Finally we provide a way to derive the all-causes mortality in LTC in the former model from estimates of transitions intensities in the latter.

The illness-death model

For $x_0 \geq 0$, let us consider a continuous-time process $(Z_x)_{x \geq x_0}$ with values in the 3-state set $E = \{A, I, D\}$ of autonomy, LTC (or "illness"), death respectively. Let us assume that Z is *càd-làg* and that $Z_{x_0} = A$.

We now assume that $(Z_x)_{x \geq x_0}$ is a non-homogeneous semi-Markov process, that death is an absorbing state and that there is no transition allowed from LTC to autonomy. We introduce the transition intensities, also called instantaneous transition probabilities

$$\mu_a(x) = \lim_{h \to 0} \frac{1}{h} P\left(Z_{x+h} = D | Z_x = A\right),$$
$$\lambda(x) = \lim_{h \to 0} \frac{1}{h} P\left(Z_{x+h} = I | Z_x = A\right),$$
$$\mu_i(x, t) = \lim_{h \to 0} \frac{1}{h} P\left(Z_{x+t+h} = D | Z_{x^-} = A, Z_x = I, Z_{x+t} = I\right).$$

Those intensities are called respectively intensity of incidence in LTC, intensity of autonomous mortality and intensity of mortality in LTC, with the latter intensity depending on both the age at onset of LTC (also called age at claim) and time spent in LTC. Figure 1 provides a representation of the model. The interested reader can refer to Biessy (2015a) for a more detailed study of the model.
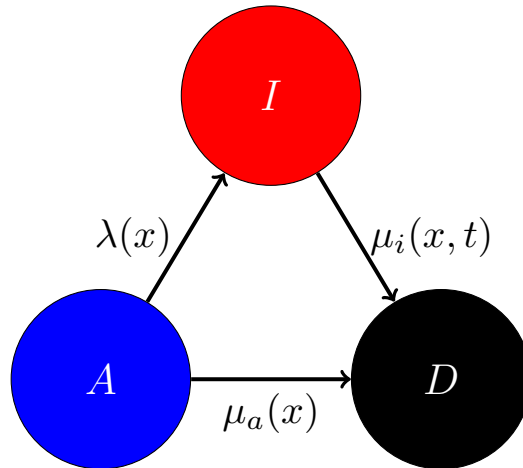

Figure 1: The 3 states continuous-time semi-Markov model with its transition intensities.

Multi-state model with pathologies

For $x_0 \geq 0$, let us consider another continuous-time process $(Z_x)_{x \geq x_0}$ with values in the discrete set $E = \{A, I_1, \ldots, I_p, D\}$ of autonomy, LTC (or "illness") with cause $1, \ldots, p$ at claim and death respectively. Let us assume that Z is *càd-làg* and that $Z_{x_0} = A$.

We now assume that $(Z_x)_{x \geq x_0}$ is a non-homogeneous semi-Markov process and consider that death is an absorbing state and that there is no transition allowed from LTC to autonomy. Besides, let us notice that each state of LTC corresponds to a different group for the cause of claim. Therefore there is no transition between states of LTC. We introduce the transition intensities of the model

$$\mu_a(x) = \lim_{h \to 0} \frac{1}{h} P\left(Z_{x+h} = D | Z_x = A\right),$$

$$\lambda_k(x) = \lim_{h \to 0} \frac{1}{h} P\left(Z_{x+h} = I_k | Z_x = A\right),$$

$$\mu_{i,k}(x,t) = \lim_{h \to 0} \frac{1}{h} P\left(Z_{x+t+h} = D | Z_{x^-} = A, Z_x = I_k, Z_{x+t} = I_k\right).$$

for $k \in \{1, \ldots, p\}$.

Those intensities are called respectively intensity of entry in LTC with cause k, intensity of autonomous mortality and intensity of mortality in LTC for cause k. A representation of the model is given by Figure 2.
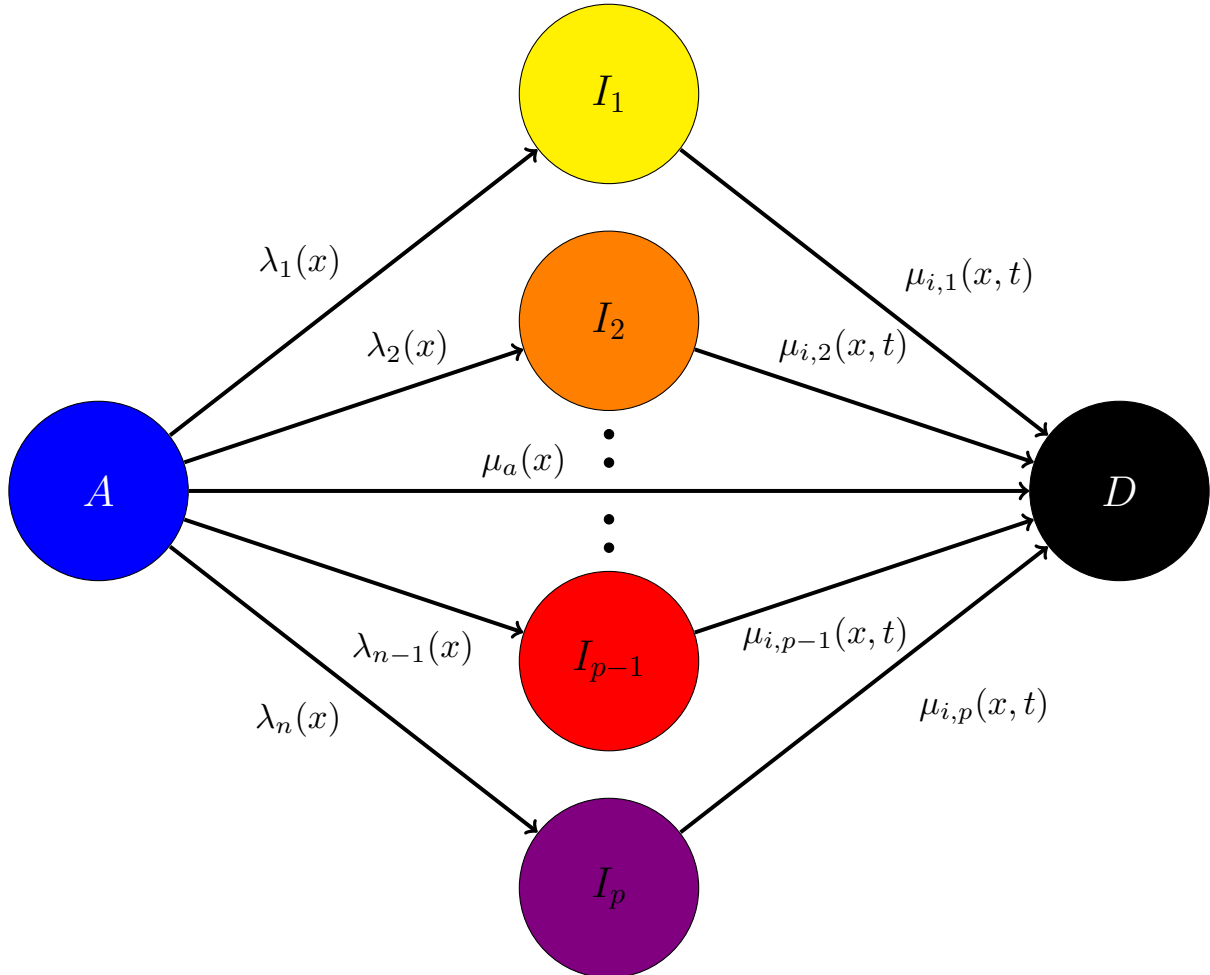

Figure 2: The continuous-time semi-Markov model with several groups of pathologies.

Link between both models

In this section, we introduce a relation which links the mortality in LTC of the previous 3-state model to intensities of transition of the multi-state model with pathologies.

Lemma 1. *Let us consider the two models defined above such that under the same notation $I = I_1 \cup \cdots \cup I_p$ and for $(j,k) \in \{1,\ldots,p\}$ such that $j \neq k$, $I_j \cap I_k = \emptyset$. Then to ensure the compatibility between both models, the following relations must be satisfied for all $x > x_0$, $t \geq 0$*

$$\lambda(x) = \sum_{k=1}^{p} \lambda_k(x)$$

$$\mu_i(x,t) = \sum_{k=1}^{p} \eta_k(x,t) \mu_{i,k}(x,t),$$

where for $x \geq x_0$, $t \geq 0$, $k \in \{1,\ldots,p\}$

$$\eta_k(x,t) = \frac{\lambda_k(x) \exp\left(-\int\limits_0^t \mu_{i,k}(x,u)\,du\right)}{\sum\limits_{j=1}^{p} \lambda_j(x) \exp\left(-\int\limits_0^t \mu_{i,j}(x,u)\,du\right)}$$

3 Local likelihood

In this section, we present the local likelihood methods that we use to build smooth estimates for each of the transition intensity in the previous models. In a nutshell, for a given function of interest and a given set of points, local likelihood approximates the function in the neighbourhood of each point by a different polynomial function. Coefficients are inferred using a localized version of the maximum likelihood estimator (MLE), which differs from the original maximum likelihood estimator through the addition of a kernel function whose value decreases with the distance to the fitting point. Local likelihood possesses some of the strengths associated with a parametric approach, while being more flexible. Local likelihood was initially introduced in Tibshirani and Hastie (1987) and extended to density estimation by Loader (1996). Chapter 7 of Loader (1999) briefly introduces the estimation of hazard rates using local likelihood, which is also implemented in the **locfit** package from the same author, on the statistical programming software **R** (R Core Team, 2016). While hazard rate estimation is very similar to density estimation, we believe that providing explicit formulas may prove very useful nonetheless.

3.1 Uni-dimensional local likelihood for right-censored left-truncated data

In this section we are interested in the estimation of a hazard rate function $\mu(x)$ which only depends on the current age x of the individual. In our model, $\mu(x)$ may be the autonomous mortality or the incidence rate in LTC. Each trajectory consists of x_i the age at which the individual enters the observation, y_i the age at which the individual leaves the observation and c_i an indicator of whether the trajectory is censored before the event of interest occurs (i.e. $c_i = 1$ in case of right-censoring and 0 otherwise). As in most life insurance portfolio studies, observations are left-truncated and right-censored. Let us notice that the **locfit** package does not handle left-truncated observations.

Principle of the estimation

For an age x and for u close to x, $\log \mu(u)$ may be locally approximated by a polynomial function i.e. for $d \in \mathbb{N}$ there exists $a = (a_0, a_1, \ldots, a_d)^T$ such that

$$\log \mu(u) = a_0 + a_1(u-x) + \cdots + a_d(u-x)^d + o((u-x)^d)$$
$$= \langle a, A(u-x) \rangle + o((u-x)^d)$$

where $A(u) = (1, u, \ldots, u^d)^T$.

We first introduce the local log-likelihood function

$$\mathcal{L}_x(a) = \sum_{i=1}^{n} (1-c_i) W\left(\frac{y_i - x}{h(x)}\right) \langle a, A(y_i - x) \rangle - \int\limits_{x_{min}}^{y_{max}} N(u) W\left(\frac{u-x}{h(x)}\right) e^{\langle a, A(u-x) \rangle}\,du \qquad (1)$$

where

$$N(u) = \sum_{i=1}^{n} \mathbb{I}\{x_i < u \leq y_i\}$$

5

is the number of individuals at risk at age u, W is a kernel function, h a bandwidth function, $x_{min} = \min\limits_{i \in \{1,...,n\}} x_i$ and $y_{max} = \max\limits_{i \in \{1,...,n\}} y_i$.

An estimate $\hat{a} = (\hat{a}_0, \ldots, \hat{a}_d)^T$ of a is then obtained by maximizing the local likelihood function

$$(\hat{a}_0, \ldots, \hat{a}_d)^T = \underset{a}{\operatorname{argmax}} \; \mathcal{L}_x(a).$$

Finally an estimate $\hat{\mu}(x)$ of $\mu(x)$ is given by

$$\hat{\mu}(x) = \exp(\hat{a}_0)$$

Local likelihood methods require the setting of several parameters

- The degree d of the polynomial function fitted locally,

- The kernel function W used in the localized MLE,

- The bandwidth function h.

The degree of the polynomial function used for the fit has a significant impact on the estimation. Higher degrees provide more flexible fits, the downside being that the volatility of such fit is higher. Local constant fit corresponding to $d = 0$ usually leads to poor results in the tail of the distribution, while there is usually no reason to use fits of degree greater than 3. Compared to the other two components, the choice of the kernel has a rather limited impact on the estimation, as long as the kernel belongs to one of the common families (one can refer to Wand and Jones, 1995, for an introduction on those kernels functions). In what follows we use the Epanechikov kernel defined as

$$W(u) = I\{|u| < 1\} \left(1 - u^2\right)^2.$$

Epanechikov kernel function is optimal in the sense that any weight function producing the same asymptotic bias has larger asymptotic variance. One may refer to Epanechnikov (1969) for a proof of this result. At last, the choice of the bandwidth is of utmost importance and several methodologies exist. Therefore it deserves its own section.

Bandwidth choice

For an age x, the bandwidth function $h(x)$ determines the number of trajectories used in the estimation of $\mu(x)$. The most obvious choice of bandwidth is a constant bandwidth independent from x. This approach is however very limited, as intuitively in regions where data is scarce, a larger bandwidth should be selected to limit the variance of $\mu(x)$. At the other end of the scope of complexity, one may find more sophisticated adaptive bandwidths methods as the intersection of confidence intervals methodology presented in Chichignoud (2010) or methods which try to minimize a local validation criterion as presented in Chapter 10 of Loader (1999).

The bandwidth that we use in what follows is a slightly modified version of the nearest neighbours bandwidth presented in Chapter 2 of Loader (1999). In the initial method, for a given α the bandwidth $h_\alpha(x)$ is obtained by sorting the vector of distances between x and each of the y_i by increasing order and selecting the j-th element of the vector where $j = \lceil n\alpha \rceil$. We bring a slight modification to this method by only considering distances from x to uncensored observations. This way we make sure that for low values of α, there is still a reasonable number of non-zero components in the left term of the localized likelihood.

Parameter selection

For each value of the degree of the fit d and for each fraction of nearest neighbours used for bandwidth selection α we have a local likelihood estimator $\hat{\mu}_{d,\alpha}$ of the hazard rate μ. In this section we introduce tools to compare those estimators and select adequate values for smoothing parameters.

We denote for $k \in \mathbb{N}$

$$M_k(x, d, \alpha) = - \int_{x_{min}}^{y_{max}} N(u) \left[W\left(\frac{u-x}{h_\alpha(x)}\right) \right]^k A(u-x)A(u-x)^T \exp\left(-\hat{\mu}_{d,\alpha}(u)\right) \mathrm{d}u.$$

While there is no guarantee that $M_k(x, d, \alpha)$ may be inverted from a theoretical point of view, in practice except for very low values of α it always proved to be the case.

We define the influence function

$$\text{infl}_{d,\alpha}(x) = W(0)e_1^T M_1^{-1}(x, d, \alpha)e_1$$

as well as the degrees of freedom

$$\nu(d, \alpha) = \sum_{i=1}^{n}(1 - c_i)\text{infl}_{d,\alpha}(y_i)$$

where for $j \in \{1, _d ots, d+1\}$, e_j is a vector of size $d+1$ with its j-th component equal to 1 and other components equal to 0.

For $i \in \{1, \ldots, n\}$, $(1 - c_i)\text{infl}_{d,\alpha}(y_i)$ may be interpreted as the sensitivity of the estimate $\hat{\mu}_{d,\alpha}(y_i)$ to the event, observed or not, (y_i, c_i). Let us notice that while censored events do not contribute to the estimation, the remain of the trajectory still plays a role through the number of individuals at risk $N(u)$. The degrees of freedom measure the flexibility of the fit. They may therefore be seen as an equivalent of the number of parameters in parametric models.

We already have a measure of the quality of the fit through the log-likelihood function

$$l(\mu) = \sum_{i=1}^{n}(1 - c_i)\log\mu(y_i) - \int\limits_{x_{min}}^{y_{max}} N(u)\mu(u)\mathrm{d}u$$

and a measure of the complexity through the degrees of freedom. Models with higher degrees of freedom are likely to better replicate the features observed in data, some of which may not correspond to real phenomena. The selected model should therefore provide a good compromise between quality of the fit and number of degrees of freedom. A natural approach for model selection is then to consider a linear combination of those two components which gives the Akaike information criterion (AIC)

$$\text{AIC}(d, \alpha) = -2l(\hat{\mu}_{d,\alpha}) + 2\nu(d, \alpha).$$

Model selection is then performed by minimizing the AIC. For a given degree d, we may therefore compute the value of the AIC on a grid of values for α and represent the value of the AIC as a function of ν as recommendeds in Chapter 10 of Loader (1999). In cases where several models exhibit very close values of AIC, Loader (1999) recommend to select the model with the lowest degree of freedom.

Pseudo-residuals

A sizeable difficulty while working with hazard rates is that we do not actually observe the hazard rate. Thus defining residuals and running diagnoses may prove difficult. One way to solve this problem would be to build empirical estimates of hazard rate, for example by dividing the number of events during intervals of the form $[x+1)$ by the number of years lived by the portfolio over the same interval. However, such estimates are very volatile on the tail of the distribution, where few data is available. Instead, an alternative solution consists of replacing empirical values by the output of a fit with a much smaller bandwidth. Such a fit should be well defined, have a bias which could be neglected and have a correct performance on the tail of the distribution. A local constant fit is not suited for this purpose as performance in the tail is very poor. We suggest to use a local linear fit corresponding to $d = 1$ and select a bandwidth using the α nearest neighbours methodology with $\alpha = 0.05$.

We define Pearson residuals as follows

$$r(x) = \frac{\hat{\mu}_{d,\alpha}(x) - \hat{\mu}_{1,0.05}(x)}{\sqrt{\text{var}(\hat{\mu}_{1,0.05}(x))}}$$

where

$$\text{var}(\hat{\mu}_{d,\alpha}(x)) = e_1^T M_1^{-1}(x, d, \alpha)M_2(x, d, \alpha)M_1^{-1}(x, d, \alpha)e_1$$

Those residuals may help us assess the quality of the fit. A good fit should present residuals with uniformly distributed signs as well as absolute values inferior to 2 most of the time.

Computational aspects

This section briefly discusses the implementation of the local likelihood method. As the implementation is roughly the same as in the **locfit** package in **R**, the interested reader will find a more detailed discussion in Chapter 12 of Loader (1999).

Estimating $\mu(x)$ for some x using local likelihood is computationally expensive. Indeed it requires to find the maximum of the localized likelihood function. To do so, one often searches for the point where the first order derivative is equal to 0. Except for the local constant fit ($d = 0$) for which a closed formula yields the solution, one has to rely on numerical methods. Our implementation is based on Newton-Raphson algorithm. At each step of the algorithm, the likelihood and its first order and second order derivatives are computed, which requires numerical approximation of the integral. To compute the likelihood of equation 1, one has to apply local likelihood as many times as the number of uncensored observations in the data, which proves very difficult to manage in a reasonable amount of time. Instead, we only use the local likelihood on a limited number of carefully chosen points, and then use an interpolation method to approximate $\mu(x)$ at any point of interest.

To define the points where we actually use local likelihood, we grow an adaptive tree using the method described in Chapter 12 of Loader (1999). We start with two boundaries that include all the observations, such as x_{min} and y_{max} then we create new points by splitting in half the intervals between two consecutive points. In each case, the point $(x + y)/2$ is added to the list if $|y - x| > c \min(h(x), h(y))$ where c is a tuning parameter. After some back-testing on the data we decide to set c at 0.01.

Once we have our grid of points, we implement Newton-Raphson algorithm to maximise the localized likelihood. First and second order of the derivatives are given by the formulas

$$\frac{\partial \mathcal{L}_x}{\partial a}(a) = \sum_{i=1}^{n}(1 - c_i)W\left(\frac{y_i - x}{h(x)}\right)A(y_i - x) - \int_{x_{min}}^{y_{max}} N(u)W\left(\frac{u - x}{h(x)}\right)A(u - x)e^{\langle a, A(u-x)\rangle}\mathrm{d}u$$

and

$$\frac{\partial^2 \mathcal{L}_x}{\partial a^2}(a) = -\int_{x_{min}}^{y_{max}} N(u)W\left(\frac{u - x}{h(x)}\right)A(u - x)A(u - x)^T e^{\langle a, A(u-x)\rangle}\mathrm{d}u.$$

At each step of the algorithm, we compute the localized likelihood and its first and second order derivatives. In Newton-Raphson algorithm, the next-step estimate $k + 1$ is obtained from the estimate at step k using the following formula

$$a^{(k+1)} = a^{(k)} - \left(\frac{\partial^2 \mathcal{L}_x}{\partial a^2}(a)\right)^{-1}\frac{\partial \mathcal{L}_x}{\partial a}(a).$$

To ensure convergence, we instead use a so-called damped version of the original algorithm

$$a^{(k+1)} = a^{(k)} - \frac{1}{2^j}\left(\frac{\partial^2 \mathcal{L}_x}{\partial a^2}(a)\right)^{-1}\frac{\partial \mathcal{L}_x}{\partial a}(a)$$

where j is the smallest natural number such that

$$\mathcal{L}_x(a^{(k)}) \leq \mathcal{L}_x\left(a^{(k)} - \frac{1}{2^j}\left(\frac{\partial^2 \mathcal{L}_x}{\partial a^2}(a)\right)^{-1}\frac{\partial \mathcal{L}_x}{\partial a}(a)\right).$$

Finally, the initial value of the parameter vector is set as the solution for the constant local fit

$$a^{(0)} = \frac{\sum_{i=1}^{n}(1 - c_i)W\left(\frac{y_i - x}{h(x)}\right)}{\int_{x_{min}}^{y_{max}} N(u)W\left(\frac{u - x}{h(x)}\right)\mathrm{d}u}e_1.$$

For the interpolation, we rely on a cubic interpolation method. For two fitting points v_0 and v_1, the interpolated value of $f(x)$ for $v_0 \leq x \leq v_1$ is given by

$$f(x) = \phi_0(\lambda)f(v_0) + \phi_1(\lambda)f(v_1) + (v_1 - v_0)\left[\psi_0(\lambda)f^{'}(v_0) + \psi_1(\lambda)f^{'}(v_1)\right]$$

where

$$\lambda = \frac{x - v_0}{v_1 - v_0}$$
$$\phi_0(\lambda) = (1 - \lambda)^2(1 + 2\lambda)$$
$$\phi_1(\lambda) = \lambda^2(3 - 2\lambda)$$
$$\psi_0(\lambda) = \lambda(1 - \lambda)^2$$
$$\psi_1(\lambda) = \lambda^2(1 - \lambda).$$

Interpolation is used not only for the transition intensity $\hat{\mu}$ but also for the influence function $\mathrm{infl}(x)$ and the variance of the estimate $\mathrm{var}(\hat{\mu}(x))$. Derivative of those functions may be approximated by their local slopes

$$\hat{\mu}'(x) \simeq \hat{a}_1 \exp(\hat{a}_0)$$
$$\mathrm{infl}'(x) \simeq W(0)e_1^T M_1^{-1}(x)e_2$$
$$\mathrm{var}'(\hat{\mu}(x)) \simeq \hat{\mu}^2(x)e_1^T M_1^{-1}(x)M_2(x)M_1^{-1}(x)e_2$$

Let us notice that linear interpolation should still be used in the case of local constant fitting ($d = 0$) as derivatives are not available in that case.

3.2 Bi-dimensional local likelihood for right-censored data

In this section we focus on the mortality in LTC which depends on both the age at claim inception x and the time already spent in the LTC state t. As in the uni-dimensional case, each trajectory consists of x_i the age at which the individual enters the observation, y_i the age at which the individual leaves the observation and c_i an indicator of whether the trajectory is censored before the event of interest occurs. We also denote by $t_i = y_i - x_i$ the time spent in LTC when the individual leaves the observation sample.

For the sake of clarity, let us focus on the log-quadratic case which corresponds to $d = 2$. For each couple (x, t) and for (u, v) close to (x, t), $\log \mu(u, v)$ may be locally approximated by a polynomial function i.e. for $d \in \mathbb{N}$ there is $a = (a_0, a_1, \ldots, a_5)^T$ such that

$$\log \mu(u, v) = a_0 + a_1(u - x) + a_2(v - t) + a_3(u - x)^2 + a_4(u - x)(v - t) + a_5(v - t)^2 + \mathrm{o}((u - x)^2 + (v - t)^2)$$
$$= \langle a, A(u - x, v - t) \rangle + \mathrm{o}((u - x)^2 + (v - t)^2)$$

where $A(u, v) = (1, u, v, u^2, uv, v^2)^T$.

We first introduce the local log-likelihood function

$$\mathcal{L}_{x,t}(a) = \sum_{i=1}^{n} (1 - c_i) \, W\left(\frac{\rho(x_i - x, t_i - t)}{h(x, t)}\right) \langle a, A(x_i - x, t_i - t) \rangle$$
$$- \sum_{i=1}^{n} \int_0^{t_i} W\left(\frac{\rho(x_i - x, u - t)}{h(x, t)}\right) e^{\langle a, A(x_i - x, u - t) \rangle} \mathrm{d}u$$

where ρ is a distance function, W is a kernel function and h is a bandwidth function.

An estimate $\hat{a} = (\hat{a}_0, \ldots, \hat{a}_5)^T$ of a is then obtained by maximizing the local likelihood function so that

$$(\hat{a}_0, \ldots, \hat{a}_5)^T = \underset{a}{\mathrm{argmax}} \, \mathcal{L}_{t,x}(a).$$

Finally an estimate $\hat{\mu}(x, t)$ of $\mu(x, t)$ is given by

$$\hat{\mu}(x, t) = \exp(\hat{a}_0)$$

There are a few noticeable differences with the uni-dimensional case. As h depends on both the age at claim inception x and the time spent in LTC t, we have to define the distance between two points (x, t) and (u, v). A natural choice would be the euclidean distance. However, both coordinates do not play a similar role and different weights should be considered. We normalize each coordinate by dividing it by the empirical standard deviation for this component observed on the sample and define the distance ρ as

$$\rho(u - x, v - t) = \sqrt{\left(\frac{u - x}{\hat{\sigma}(X)}\right)^2 + \left(\frac{v - t}{\hat{\sigma}(T)}\right)^2}$$

where $X = \{x_1, \ldots, x_n\}$, $T = \{t_1, \ldots, t_n\}$ and $\hat{\sigma}$ is the estimated standard deviation on the database.

Another sizeable difference is of computational nature. In the uni-dimensional case, it was possible to account for the number of individuals at risk through the addition of $N(u)$ and compute numerically a single integral. In the bi-dimensional case, age at claim inception impacts mortality and we must compute one integral for each trajectory. Regarding other aspects, the bi-dimensional case is very similar to the unidimensional case. Results obtained in the next two sections are based on the implementation of local likelihood from the **locfit** package.

Finally, as in the uni-dimensional case, we define pseudo-residuals thanks to a reference. In the bi-dimensional case, we decide to use a local constant fit and consider the distance to the 100-th nearest neighbour as the local bandwidth, which corresponds to $d = 0$ and $\alpha = 100/n$. This choice of bandwidth, while providing a fit with a low amount of smoothing, ensures that there is enough observations within the bandwidth to create regular mortality surfaces and avoid visual disturbances.

4 Inference of transition probabilities

In this section, we apply the local likelihood methods defined previously to infer the transition probabilities in the model. Autonomous mortality, incidence in LTC for each cause and overall incidence in LTC are inferred using the uni-dimensional method with our own implementation of the algorithm while mortality in LTC for each cause and overall mortality in LTC is computed using the algorithm implemented in the **locfit** package on **R**.

4.1 Autonomous mortality

Figure 3 represents the intensity of autonomous mortality smoothed using local likelihood with a logarithmic scale for ages between 60 and 95. Information about the selected smoothing parameters may be found in Table 2. The autonomous mortality is constantly higher for males than females and for both genders it increases exponentially with respect to age. Fits of degrees 2 and 3 with a large bandwidth were selected which indicates that the shape of the curve is quite complex but also regular. Figure 4 represents the residuals for the fit. Except at the border of the graphs where data is scarce, residuals have uniformly distributed signs and their absolute value rarely exceeds 2. From a graphical point of view, the fit does not appear to over-smooth the data.
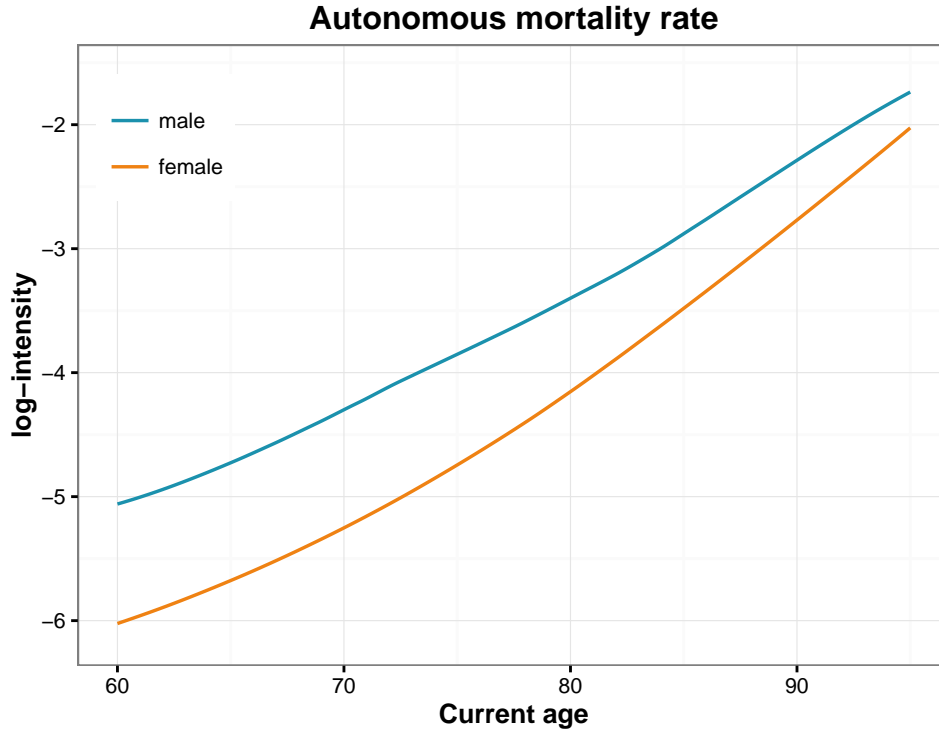

Figure 3: Intensity of autonomous mortality estimated using local likelihood: logarithmic scale.

Gender	d	α	ν	AIC
male	3	0.85	6.76	155,532.4
female	2	0.95	4.36	181,422.4

Table 2: Smoothing parameters selected for autonomous mortality.

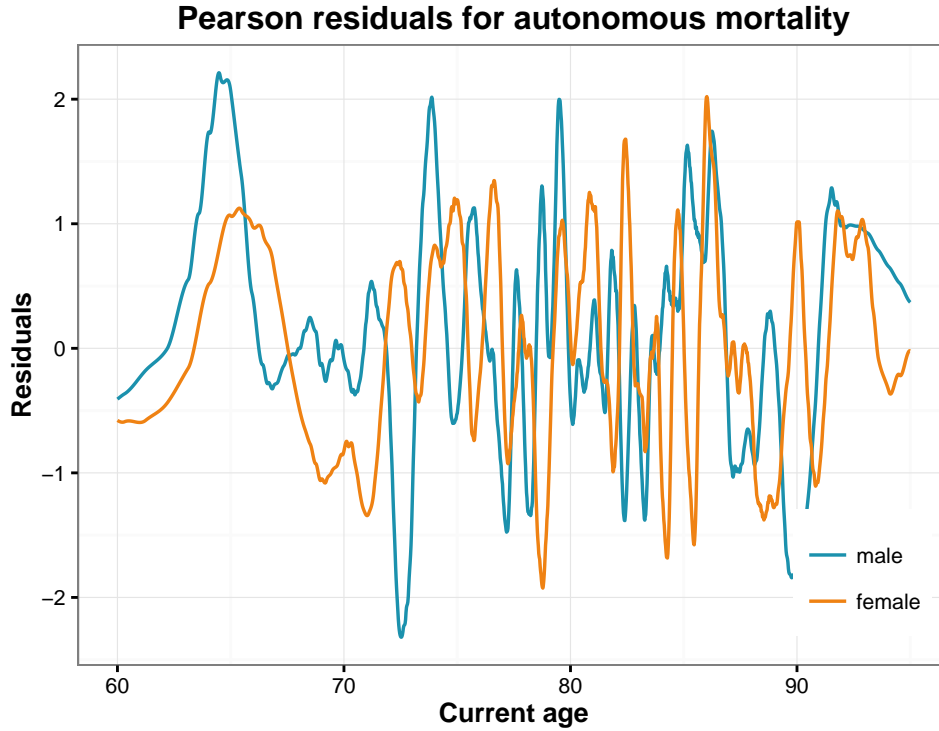**Pearson residuals for autonomous mortality**

Figure 4: Pearson residuals for autonomous mortality.

4.2 Overall incidence in LTC

Figure 5 represents the intensity of incidence in LTC smoothed using local likelihood with a logarithmic scale for ages between 60 and 95 as well. Information about the selected smoothing parameters may be found in Table 3. Incidence in LTC is initially higher for males than for females but the situation is reversed for ages above 82. Degree 3 was selected for both genders, with an average bandwidth for males, which results in a complex curve with many degrees of freedom, while for females a large bandwidth was selected and the incidence curve appears very smooth. Figure 8 represents the residuals for the fit. Once again, except at the borders where data is scarce, residuals have uniformly distributed signs and their absolute value rarely go above 2.

Gender	d	α	ν	AIC
male	3	0.55	9.67	60,036.5
female	3	1	4.82	121,155.5

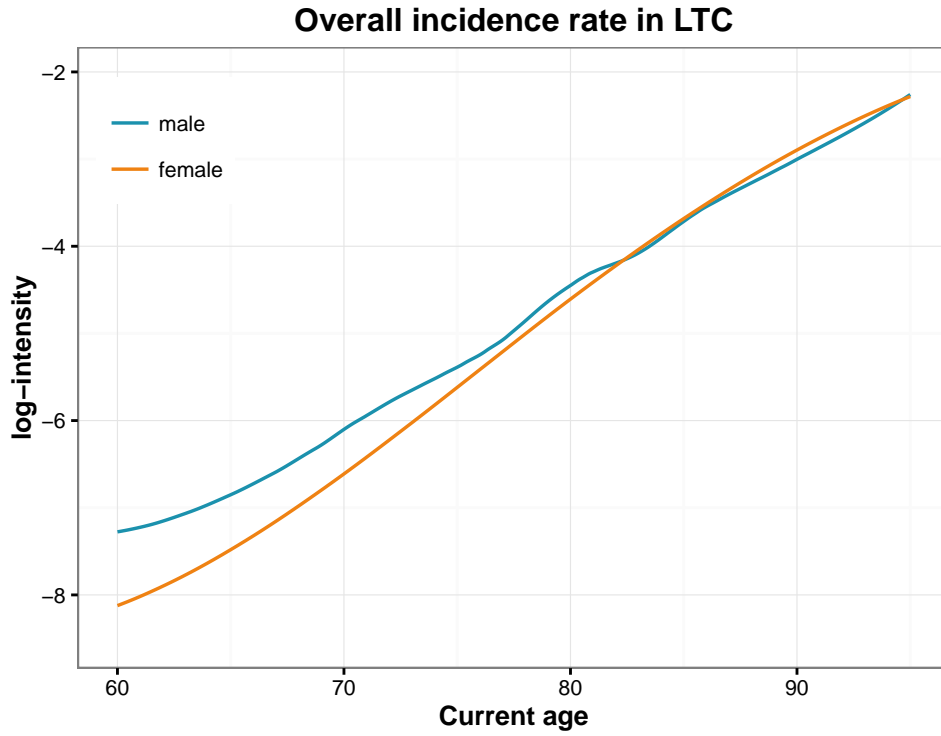Table 3: Smoothing parameters selected for overall incidence in LTC.

Figure 5: Intensity of overall incidence in LTC, estimated using local likelihood: logarithmic scale.
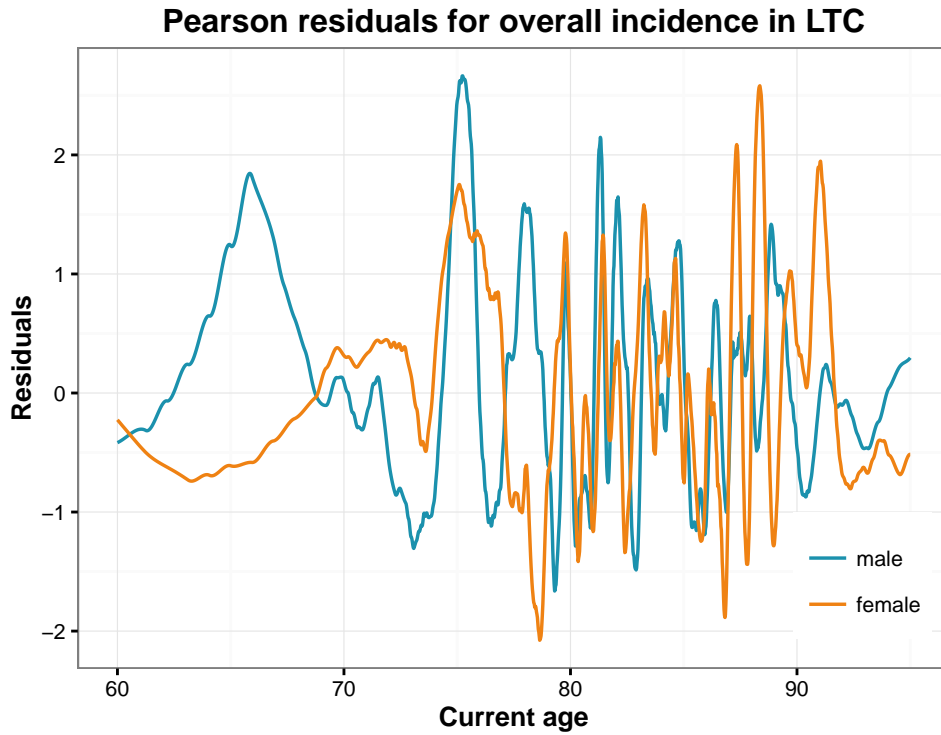

Figure 6: Pearson residuals for overall incidence in LTC.

4.3 Incidence in LTC by group

Figure 7 represents the intensity of incidence in LTC for each of the 4 groups of pathologies present in the data. Information about the selected smoothing parameters may be found in Table 4. For both genders, *cancer* is the most frequent disease at age 60, while *dementia* and *other diseases* have the higher incidence rates from age 80 onward. Males exhibit higher incidence rates for *cancer* and *neurological*

diseases. Incidence rates for *dementia* and *other diseases* are lower for females than males at age 60 but higher from age 80 onward. Figure 6 represents the residuals for the fit. Once again, except for lower ages where data is scarce, signs are uniformly distributed and absolute value of the residuals rarely go above 2.
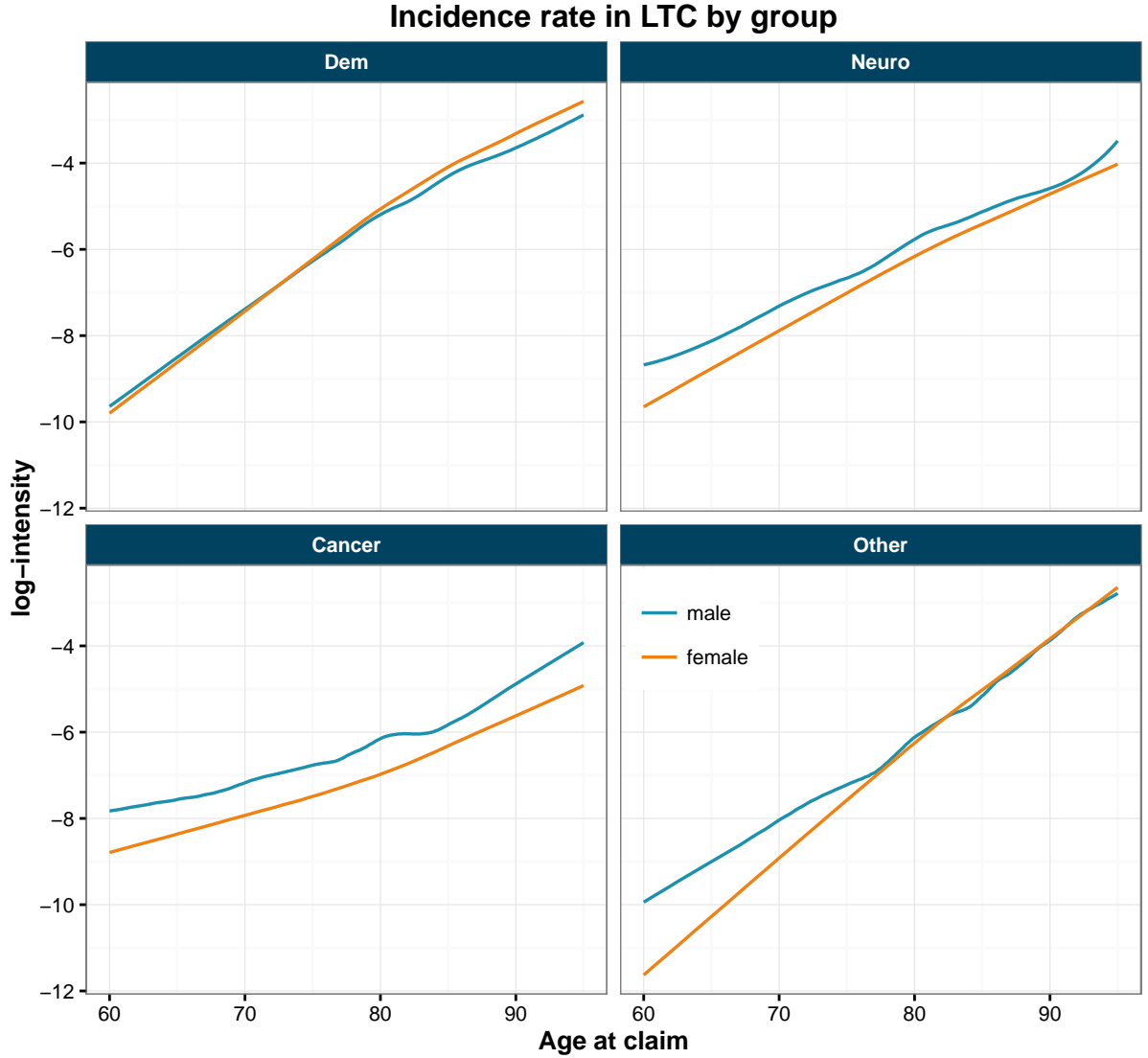

Figure 7: Intensity of incidence in LTC by group, estimated using local likelihood: logarithmic scale.

Gender	Group	d	α	ν	AIC
male	dementia	1	0.4	6.33	32,468.2
	neurological	3	0.65	8.85	19,576.1
	cancer	1	0.35	6.92	14,792.0
	other	1	0.25	8.96	17,110.1
female	dementia	1	0.5	5.30	82,897.1
	neurological	1	0.9	3.09	32,077.8
	cancer	1	0.85	3.26	17,077.3
	other	1	0.8	3.51	37,620.1

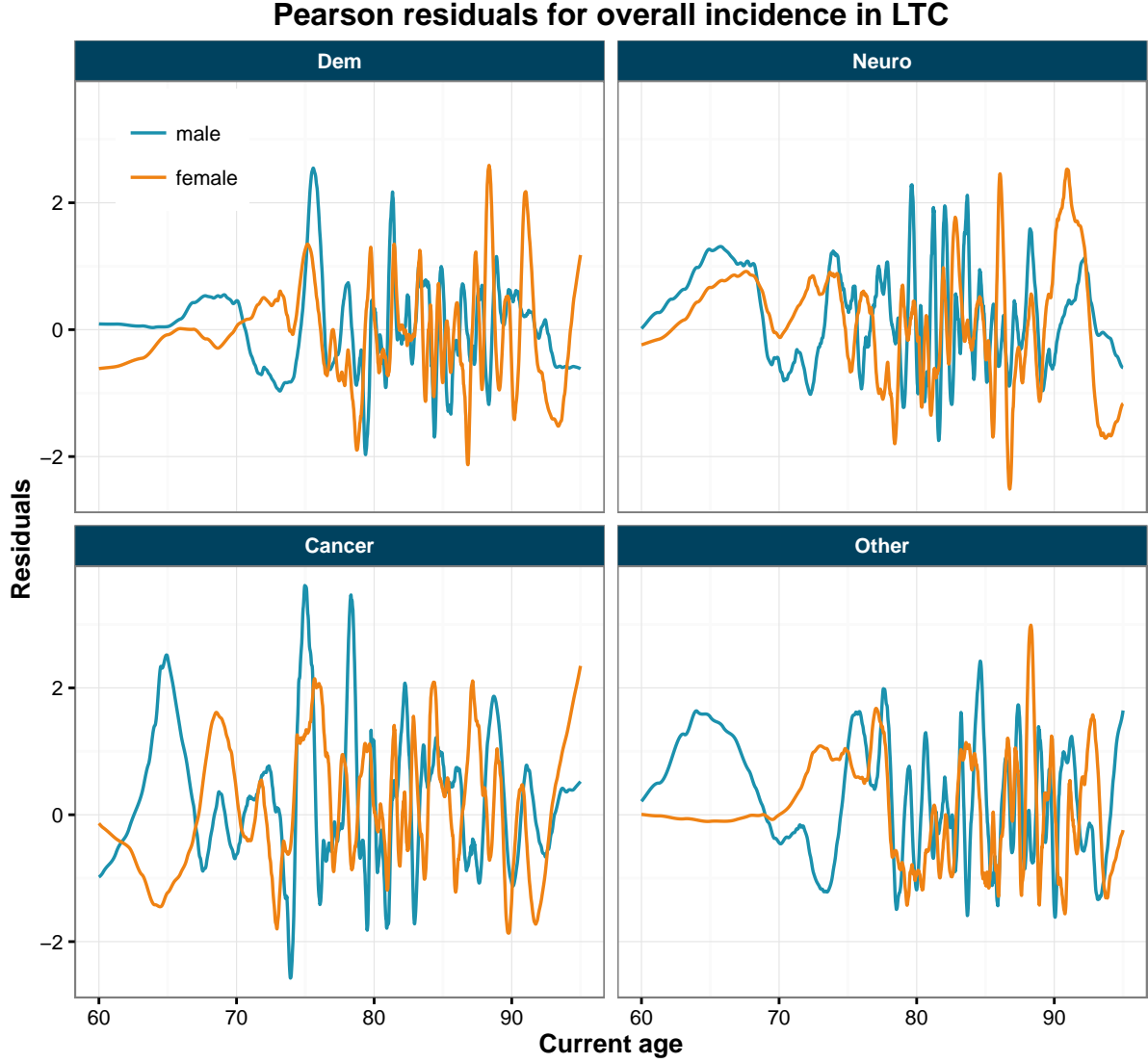Table 4: Smoothing parameters selected for incidence in LTC by group.

Figure 8: Pearson residuals for incidence in LTC by group.

4.4 Overall mortality in LTC

Figure 9 represents the mortality surface obtained by using the reference fit defined previously for x between 70 and 95 and for durations t lower than 8 years.

Mortality surfaces for males and females appear to present very similar features despite the level of mortality being constantly lower for females than for males. The initial level of mortality at claim inception is very high, and the mortality decreases very quickly with respect to duration t in LTC then increases slowly for higher durations. The impact of age at claim x on mortality seems non trivial.

Figure 10 represents the fit obtained using selected optimal parameters that may be found in Table 5. Compared to the reference fit, the mortality surfaces appear over-smoothed, especially for low durations. This is confirmed by Figure 11 which represents pseudo-residuals associated with the optimal fit. Indeed, those residuals show a lot of structure for low durations. The initial mortality appears to have been underestimated for durations below 1 and overestimated for durations between 1 and 2.

Gender	d	α	ν	AIC
male	1	0.6	7.71	6,801.3
female	1	0.5	8.80	22,502.5

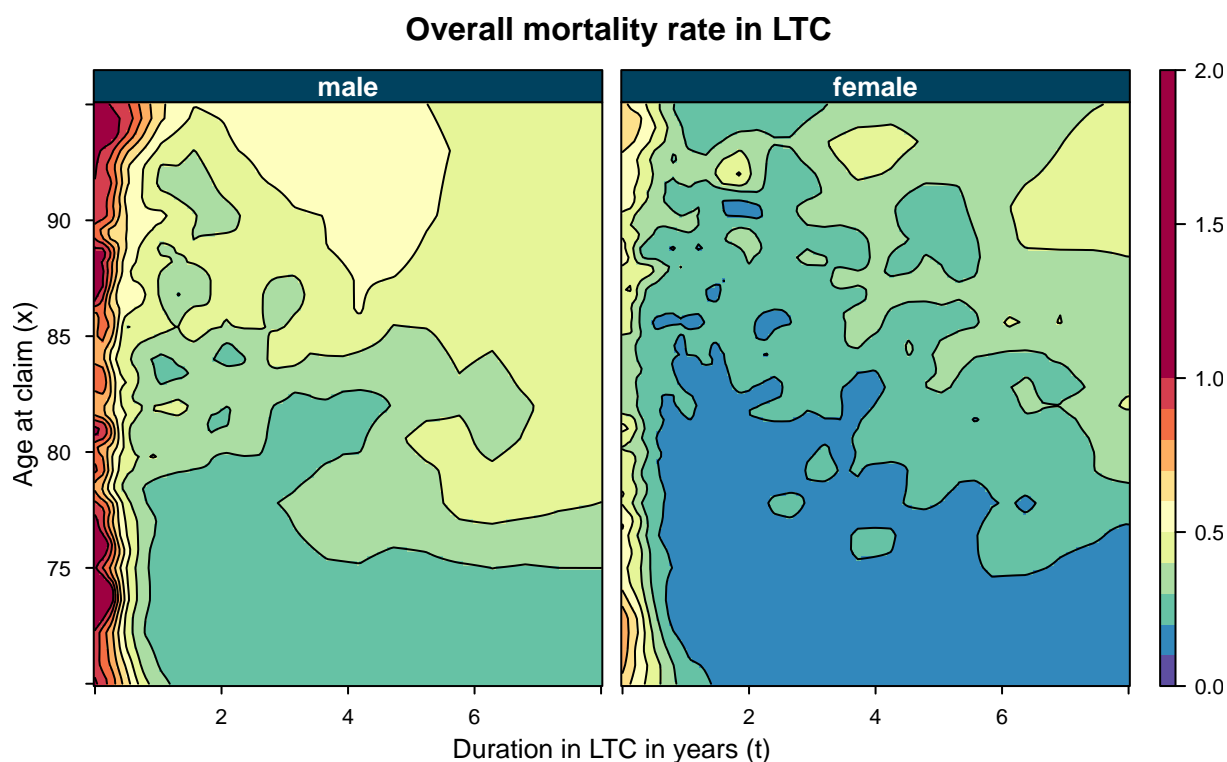Table 5: Smoothing parameters selected for overall mortality in LTC.

Figure 9: Intensity of mortality in LTC obtained by using constant local fitting and distance to the 100-th nearest neighbours as bandwidth.
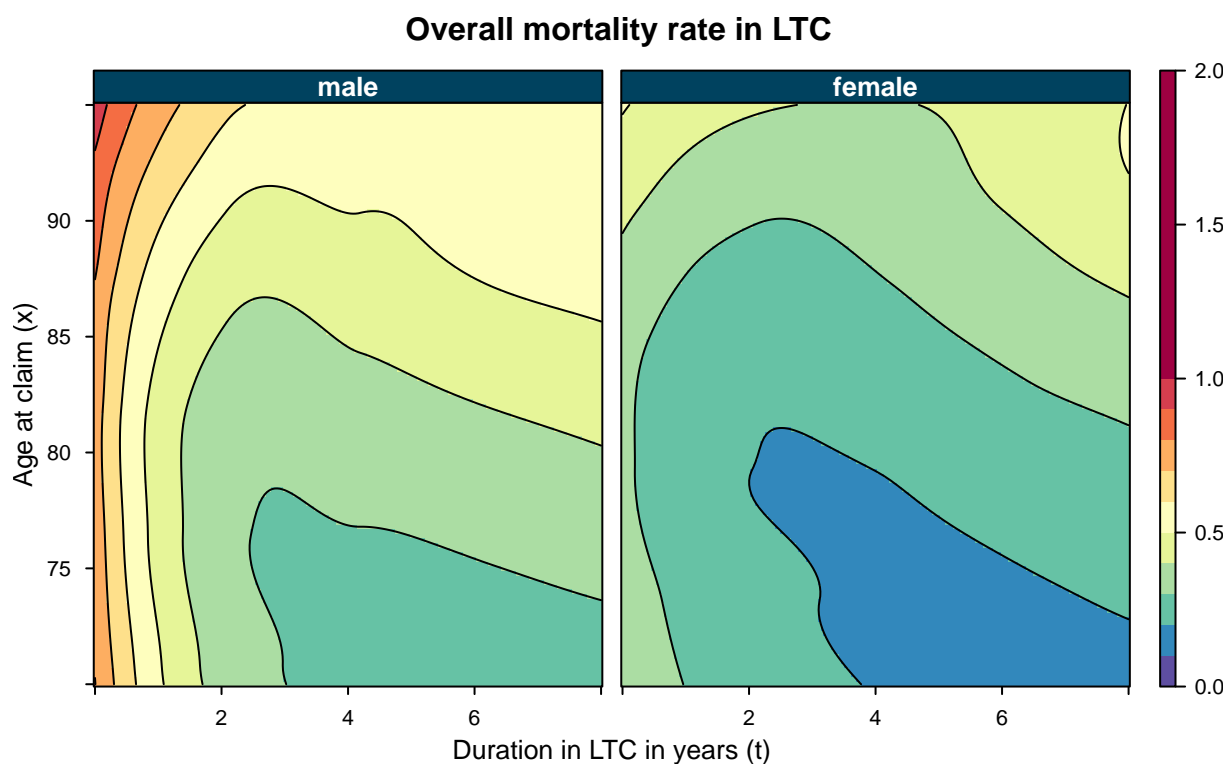

Figure 10: Intensity of mortality in LTC obtained using local likelihood with optimal smoothing parameters.
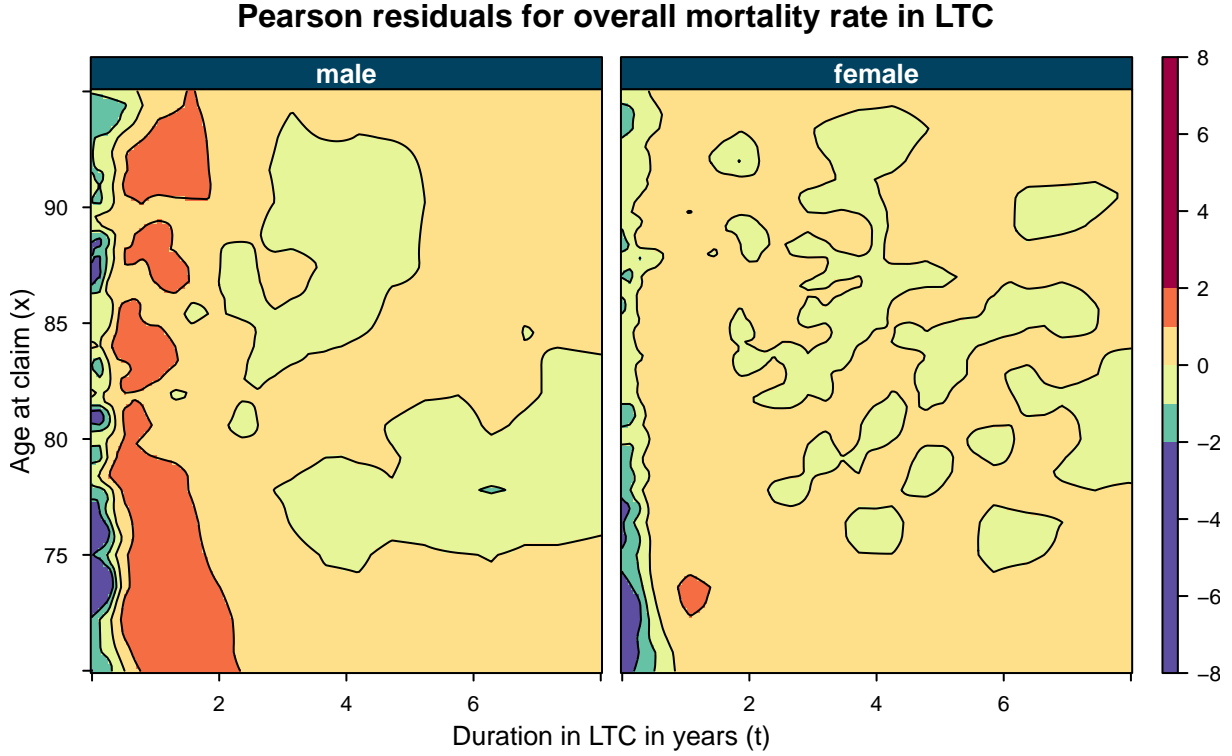
Pearson residuals for overall mortality rate in LTC

Figure 11: Pearson residuals associated with overall mortality in LTC.

4.5 Mortality in LTC by group

Figure 12 represents the surface of mortality in LTC for each group of pathology, obtained by applying a local constant fit with the 100-th nearest neighbours bandwidth. We observe significant differences between groups. First of all, mortality for *cancer* is extremely high during the first few months following claim inception. It is also the only group of pathology for which mortality does not increase with respect to age at claim. For *dementia*, mortality increases with respect to both age at claim and duration, while for the two remaining causes, mortality increases with age at claim but decreases with respect to duration. Let us notice that for other causes, especially for males, we observe a very high mortality for ages at claim inception beyond 85. Such decline may result from heterogeneity in the data. Similarly, we may believe that there is still heterogeneity in the category *other causes*. Our guess would be that infarction and stroke are responsible for the increased mortality near claim inception, which could explain why higher ages and males are mostly concerned. At last, for higher duration, mortality for *neurological diseases* and *other causes* starts to increase again which may correspond to the impact of the ageing process.

Gender	Group	d	α	ν	AIC
male	dementia	1	0.95	4.55	4,089.4
	neurological	1	0.75	6.26	2,397.5
	cancer	0	0.15	13.27	- 2,150.1
	other	1	0.45	9.30	1,013.3
female	dementia	1	0.9	5.06	13,300.2
	neurological	1	0.5	8.72	4,317.1
	cancer	0	0.15	13.00	- 1,679.5
	other	1	0.5	8.69	4,299.1

Table 6: Smoothing parameters selected for mortality in LTC by group.

Figure 13 represents the surface of mortality for each group obtained using selected optimal parameters that may be found in Table 6. Features of the initial surfaces seem to have been very well preserved. Fits of degree 1 have been selected as well as very large bandwidth except for the *cancer* group. Mortality surface obtained therefore appear very regular. Figure 19 represents the residuals for the fit. Those residuals do not present a lot of structure, except once again for *cancer*.
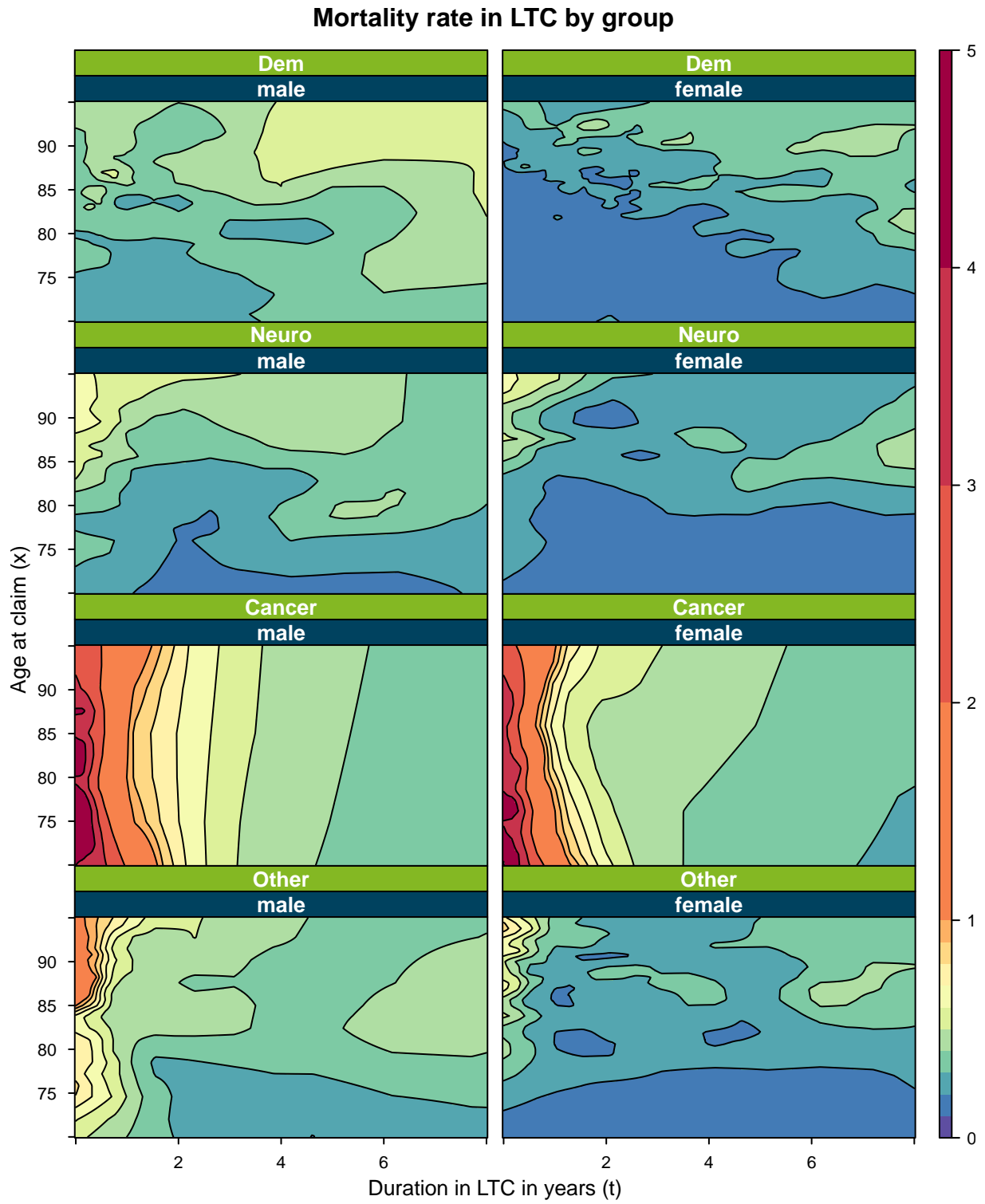
Figure 12: Cause-specific intensity of mortality in LTC obtained by using constant local fitting and distance to the 100-th nearest neighbours as bandwidth.
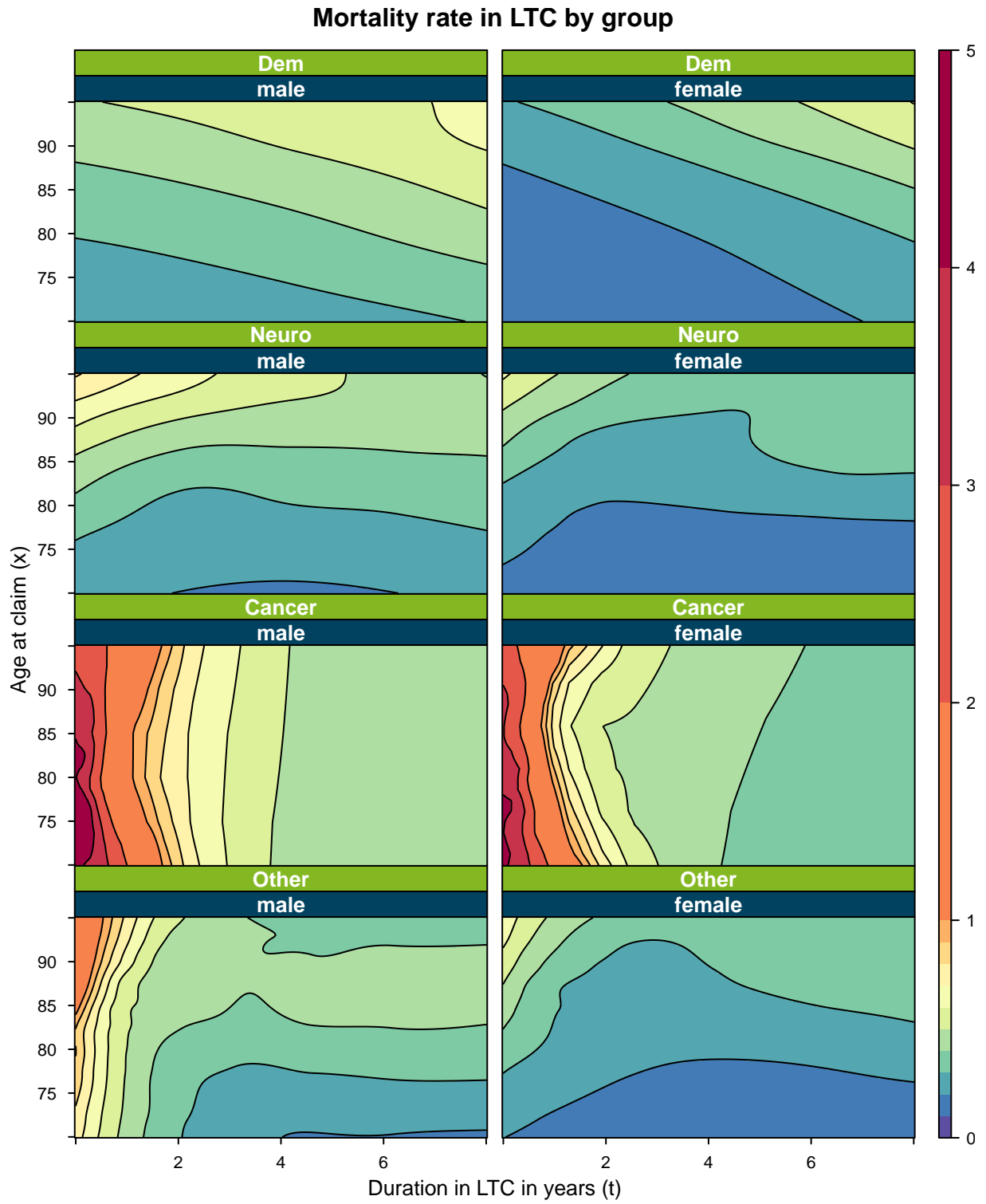
Figure 13: Cause-specific intensity of mortality in LTC obtained using local likelihood with optimal smoothing parameters.
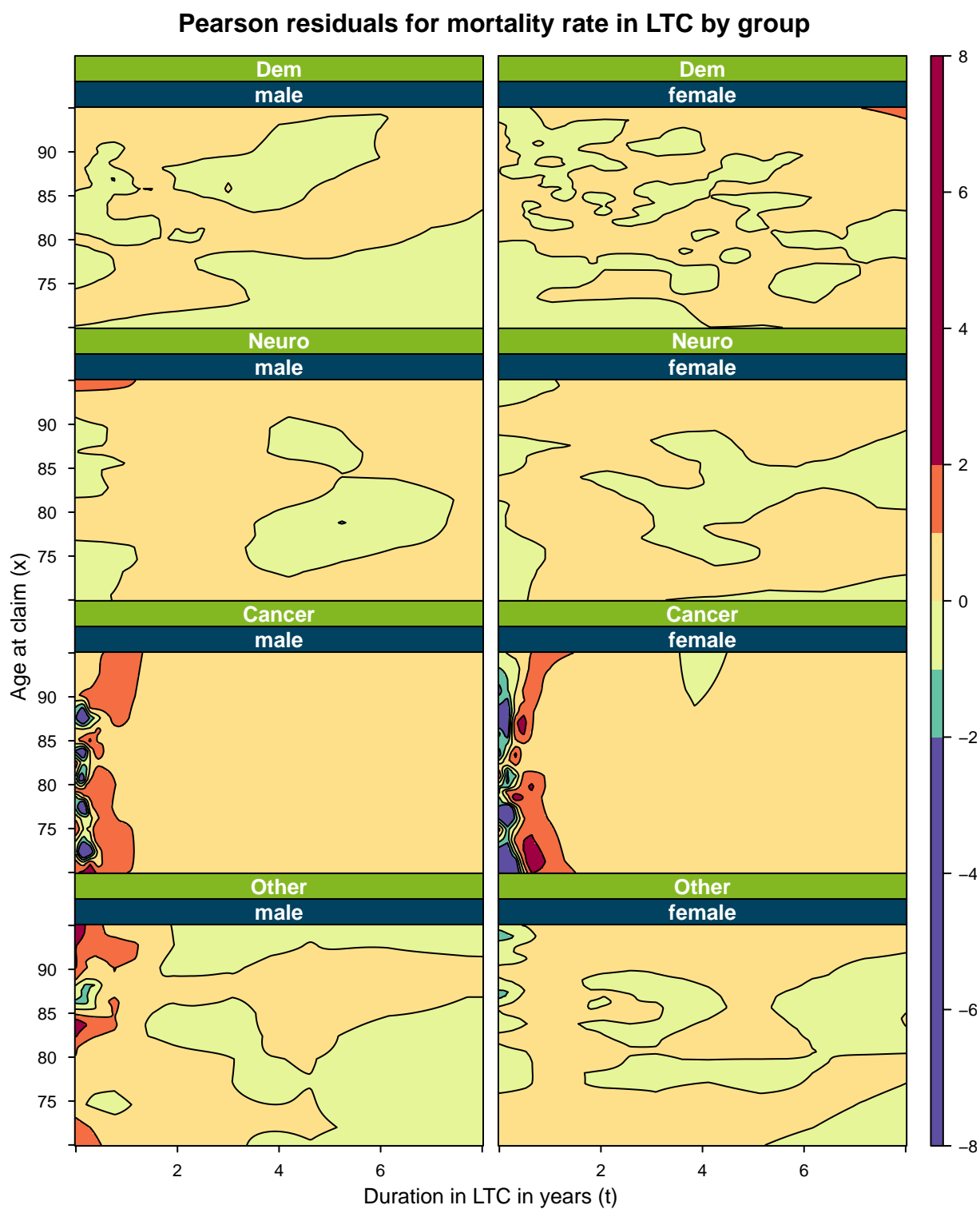
Figure 14: Pearson residuals associated with mortality in LTC by group.

5 Consequences for the LTC insurer

Using local likelihood, we were able to estimate the different transition intensities for both the illness-death model of Figure 1 and the multi-state model with several groups of pathology of Figure 2. In this section we first derive results relative to each group of pathologies and then use Lemma 1 to compute a second-step estimate of overall mortality in LTC. Lastly we look at the weight of each group in the disabled population and the contribution of those groups to overall mortality in LTC.

5.1 Results relative to individual groups of pathologies

Let us set $x_0 = 50$ and introduce the following notations

$$A(x) = \exp\left(-\int_{x_0}^{x} [\mu_a(u) + \lambda(u)]\, \mathrm{d}u\right),$$

$$I_{x,k}(t) = \exp\left(-\int_0^t \mu_{i,k}(x,u)\mathrm{d}u\right)$$

corresponding to the probability of remaining autonomous between ages x_0 and x and the probability of surviving while in LTC for an age of entry x and a pathology of group k between durations 0 and t, for $x_0 \le x$, $0 \le t$ and $k \in \{1, \ldots, p\}$.

Figure 15 represents the expected duration of claims $\mathrm{ED}_k(x)$ according to age at claim, gender and pathology group

$$\mathrm{ED}_k(x) = \int_0^\infty I_{x,k}(u)\mathrm{d}u.$$

Cancer is an outlier, with an associated duration of claim far below the level of other groups. It is also the only group of pathology for which expected duration of claim increases with respect to age at claim, when other groups exhibit a sharp decrease. Expected duration of claim is constantly lower for males than females for every group of pathology no matter the age at claim inception.
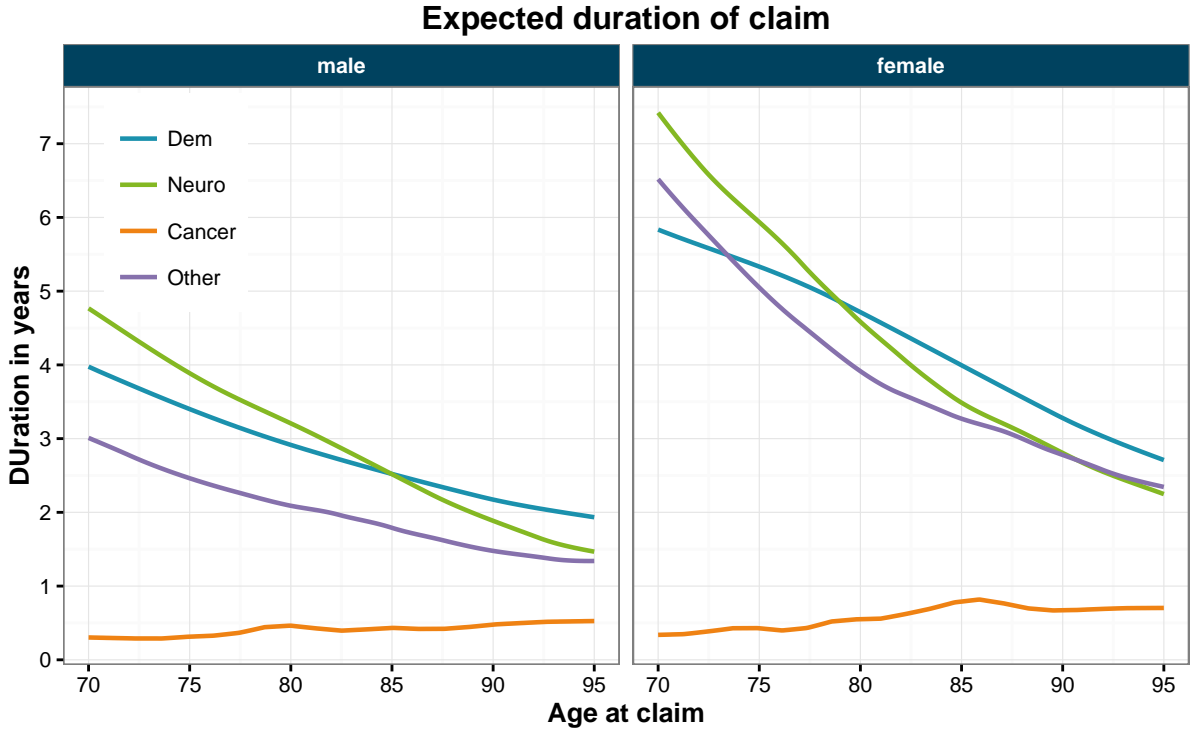

Figure 15: Expected duration of claim by age at claim inception, gender and pathology group.

Using autonomous mortality in addition to the specific incidence and mortality rates by group, we are able to project the evolution of a population of initially autonomous insured lives from age 50 onwards.

Figure 16 represents the number of open claims $\mathrm{NC}_k(x)$ as a percentage of the initial autonomous population at age 50

$$\mathrm{NC}_k(x) = \int\limits_{x_0}^{x} \lambda_k(u)A(u)I_{u,k}(x-u)\mathrm{d}u.$$

For an insurer, the area under each curve directly relates to the cost associated with each pathology, although the actual cost also depends on the technical rate used by the insurer. For both genders, cost related to *cancer* may be neglected. *Dementia* is the more expensive cause for both males and females followed by *other causes*. Finally *neurological diseases* correspond to a higher fraction of total claim cost in males than in females. Maximum amount of open claims is reached at age 90 for males and 92 for females.
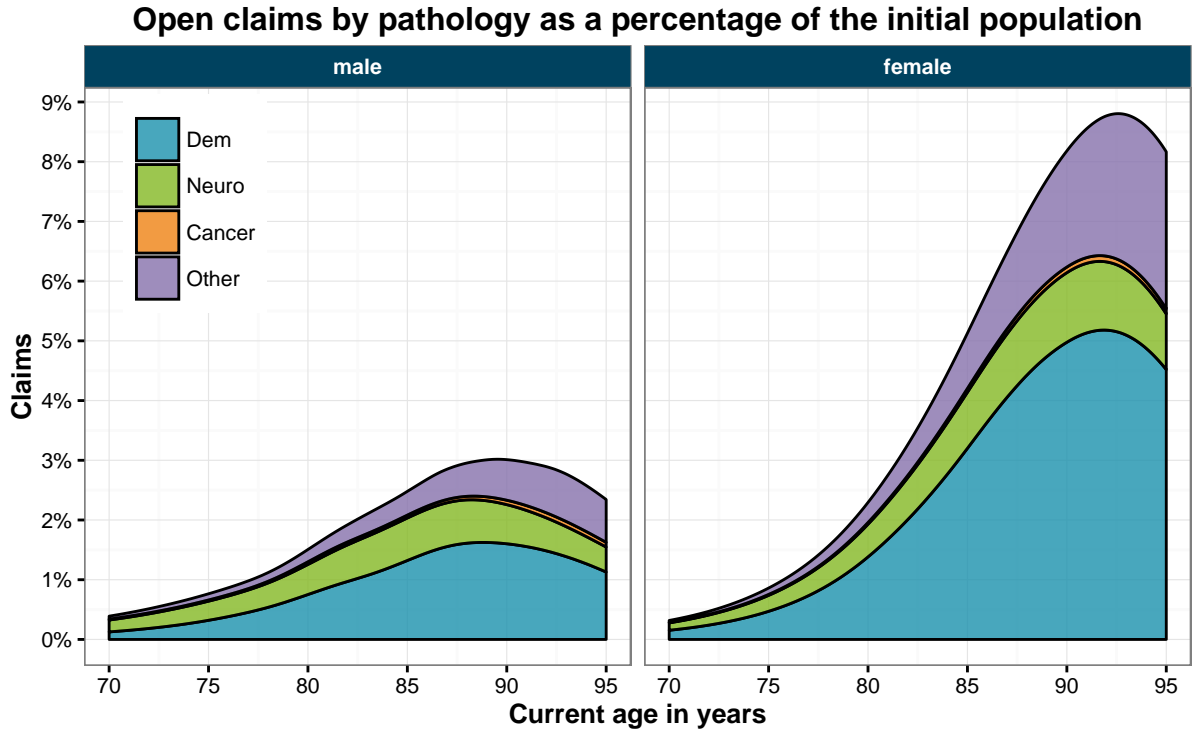

Figure 16: Projected open claims by pathology as a percentage of the initial population at age 50.

Figure 17 represents the prevalence $\mathrm{P}_k(x)$ of each pathology group in the portfolio

$$\mathrm{P}_k(x) = \frac{\mathrm{NC}_k(x)}{A(x) + \sum\limits_{j=1}^{p} \mathrm{NC}_j(x)}.$$

Prevalence of cancer remains very low, which should come at no surprise given the high mortality rates associated with it. Prevalence of *neurological diseases* starts higher than other causes at age 70 but increases more slowly. This prevalence is about the same for males and females. On the other hand, prevalence of *dementia* and *other causes* in females is much higher than prevalence in males. This is mostly due to the lower mortality observed for females in LTC.
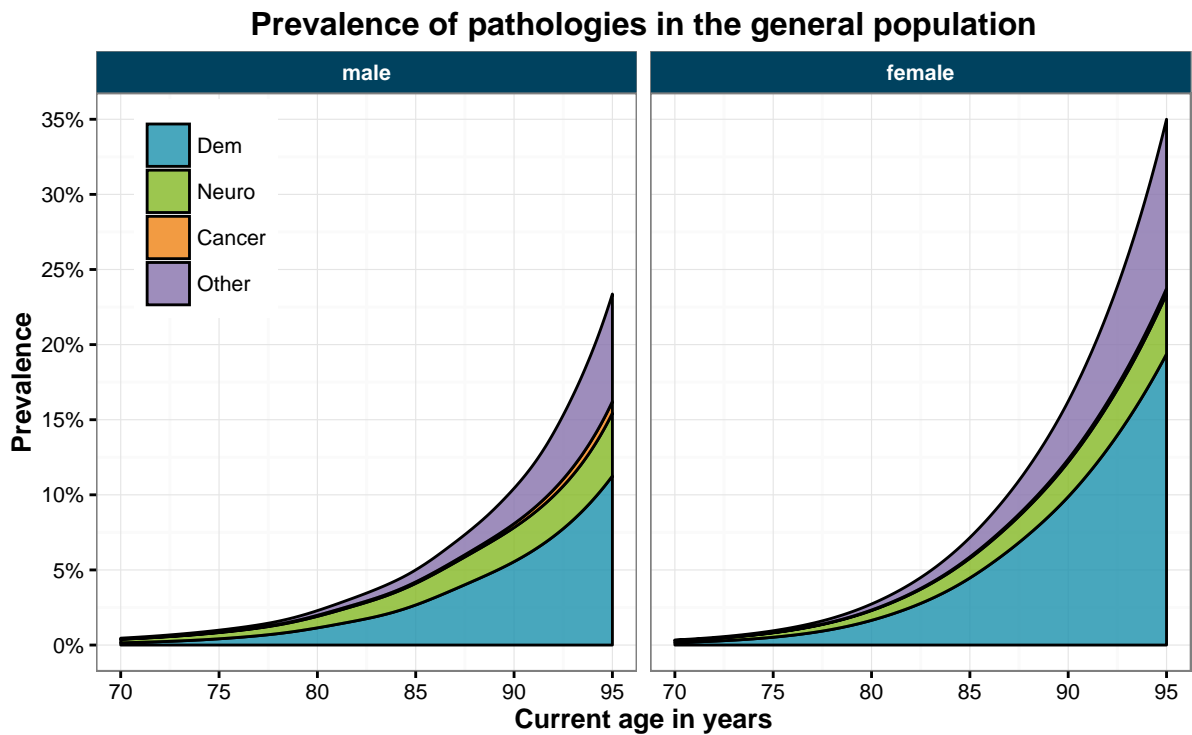
Prevalence of pathologies in the general population

Figure 17: Projected prevalence of pathologies in the portfolio.

5.2 A second-step estimate of mortality in LTC
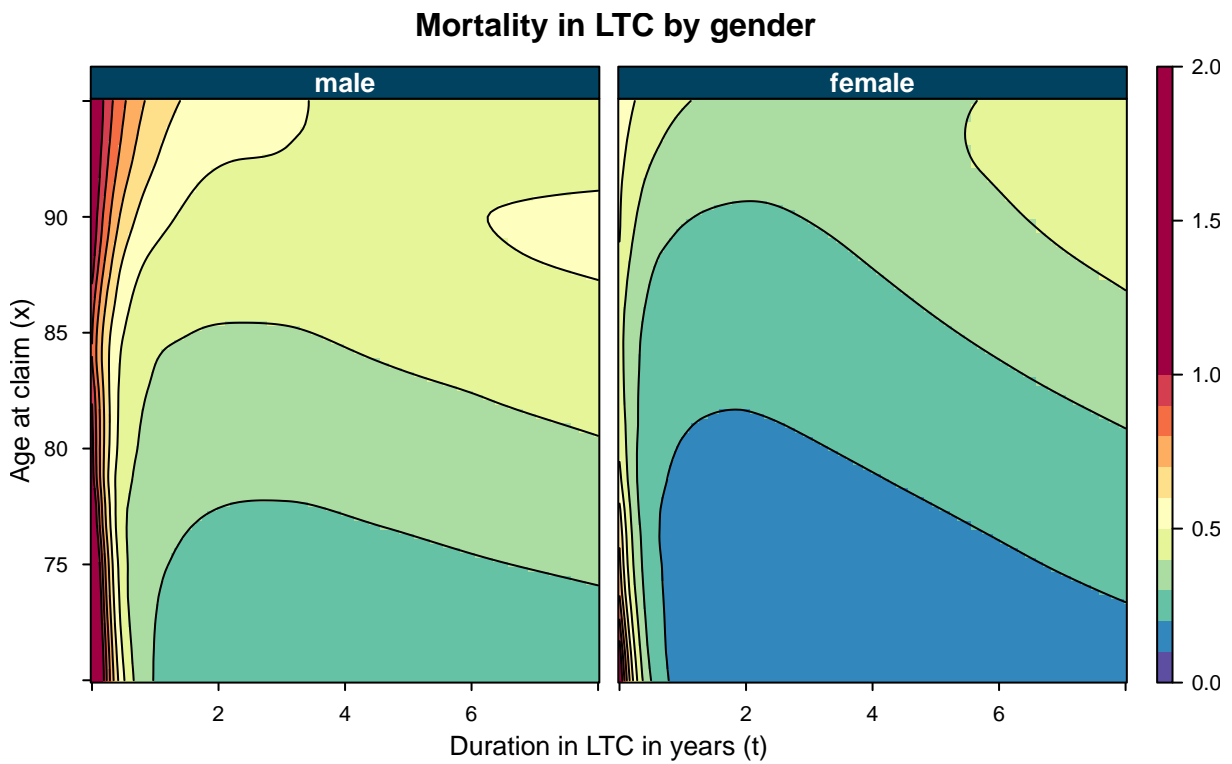
Mortality in LTC by gender

Figure 18: Intensity of mortality in LTC obtained by combining incidence and mortality for each group of pathologies.

Finally, we combine the incidence and mortality for each group to get an overall mortality surface thanks to Lemma 1. Figure 18 represents the mortality surface obtained by combining the smoothed incidence rates in LTC for each group and the associated smoothed mortality rates in LTC. This second-step estimate reproduces the features of the reference fit with 100 nearest neighbours more fairly than the first-step estimate obtained by smoothing directly the overall mortality. Indeed the high mortality rates following claim inception are preserved and contour curves are matched more closely. If we take a look at the residuals represented on Figure 19, we see that with the exception of some features in the region of low age at claim and low duration, unlike the first-step estimate they no longer seem to present any structure.
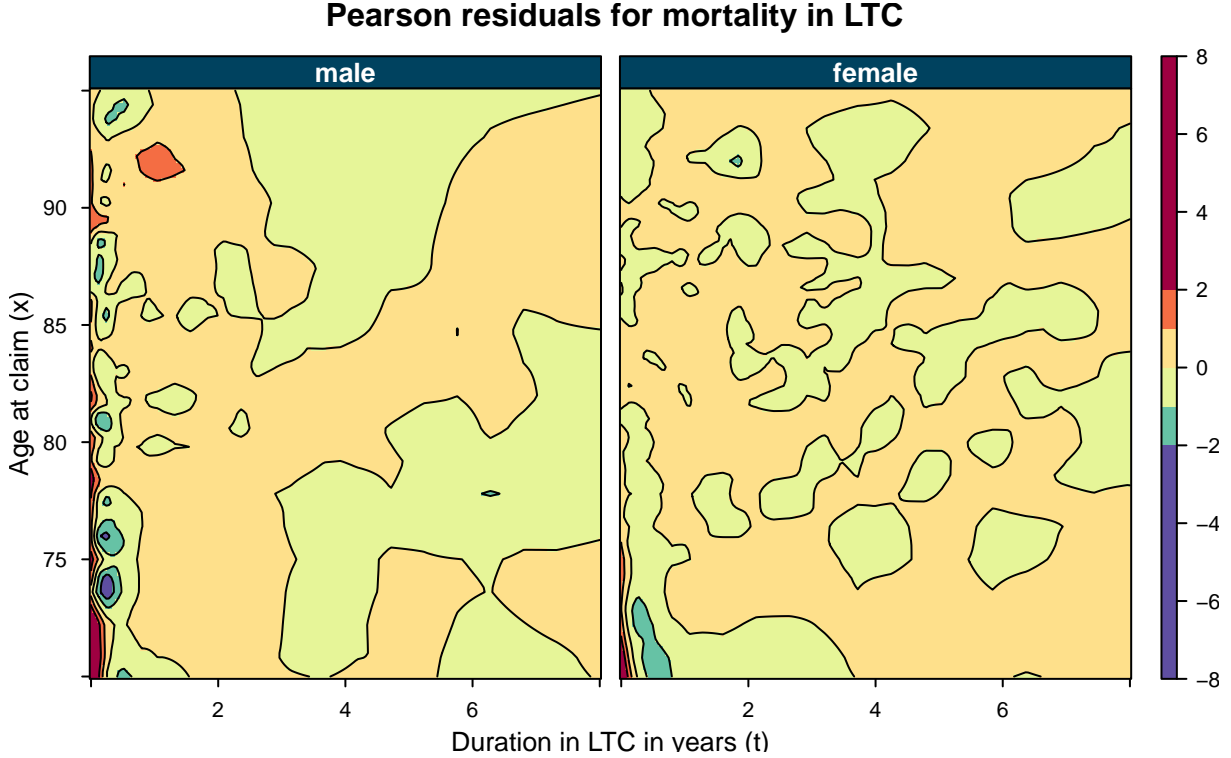

Figure 19: Pearson residuals associated with combined mortality in LTC.

5.3 Weight in the disabled population and contribution to mortality in LTC

The shape of the curve of mortality in LTC may be explained by the distribution of groups within the disabled population. Figure 20 represents the evolution of each group weight in this distribution as duration goes by, for several ages at claim. The fraction of disabled for a pathology group k corresponds to the quantity $\eta_k(x,t)$ introduced in Section 2. Due to the high mortality associated with it, *cancer* vanishes quickly from the disabled population. The proportion of *dementia* slightly declines with duration, especially for males, but remains significant. Proportions of *neurological diseases* and *other causes* tend to increase with duration, *neurological diseases* being more frequent for lower ages and *other causes* for higher ages at claim inception. Figure 21 represents the contribution of each pathology to the mortality in LTC $\eta_k(x,t)\mu_{i,k}(x,t)$. At duration 0, overall mortality in LTC is very high, due to the high mortality associated with *cancer*. This phenomenon is more substantial for males and for lower age at claim inception as the initial weight of *cancer* in the population is higher. At higher durations, contribution to mortality converges toward distribution of pathologies in the population.
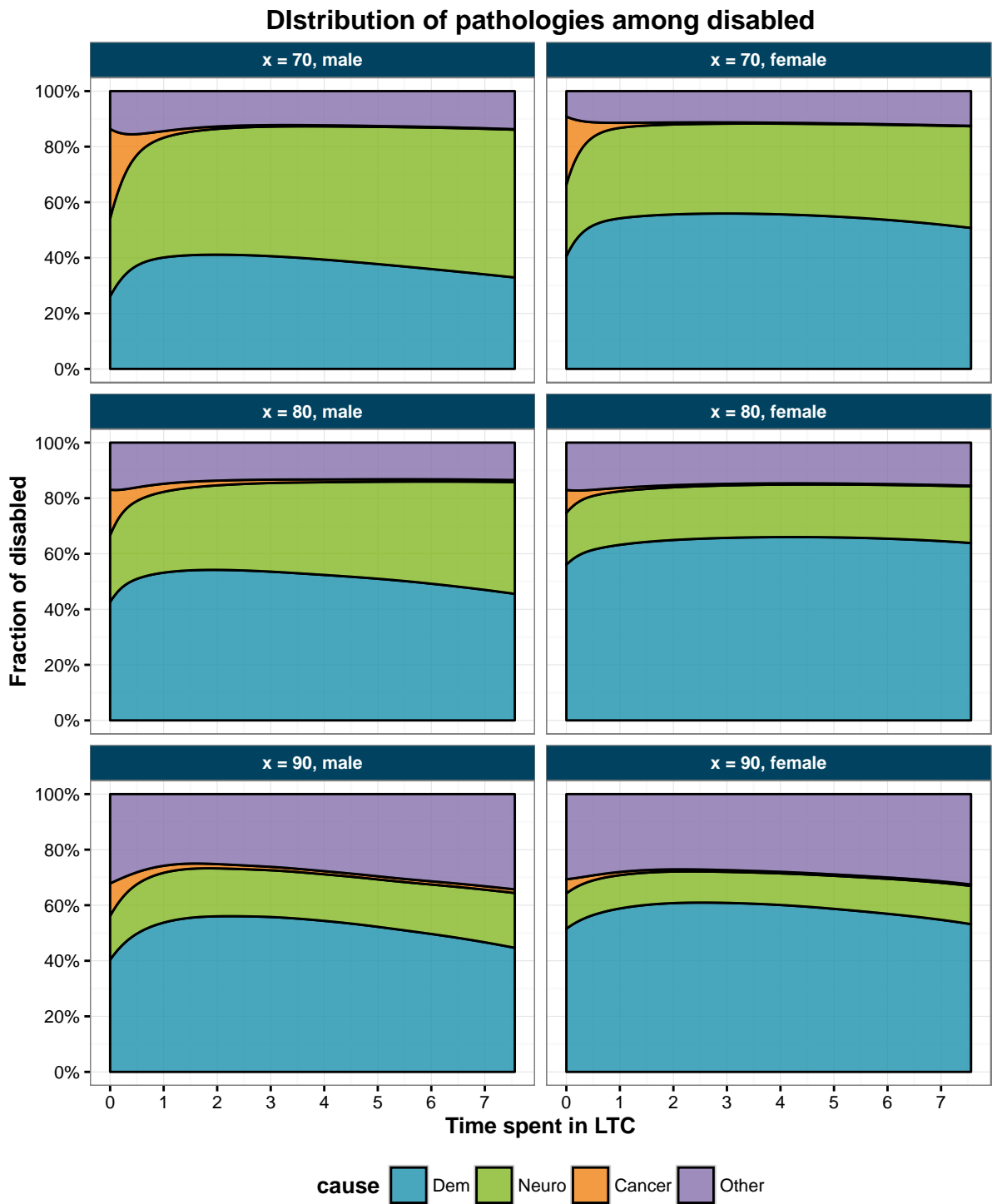
DIstribution of pathologies among disabled

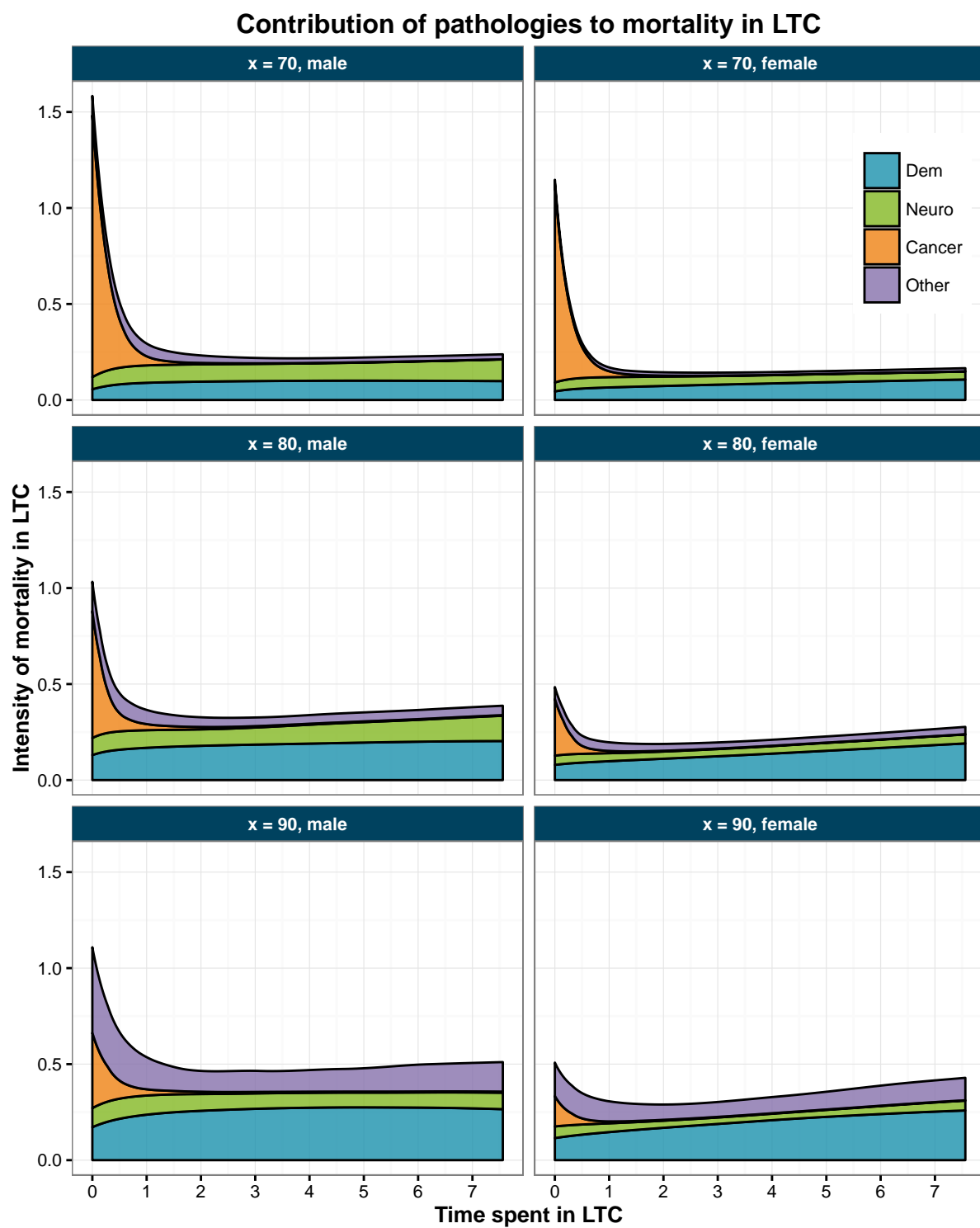

Figure 20: Distribution of pathologies among disabled.

Figure 21: Contribution of pathologies to overall mortality in LTC.

6 Discussion

In this article, we study the mortality in Long Term Care (LTC) associated with 4 different groups of pathologies: *cancer*, *dementia*, *neurological diseases* and *other causes*. We rely on data from a French LTC portfolio. We introduce a continuous time semi-Markov framework where the mortality in LTC depends on both the age at claim inception and the time spent in the LTC state. We consider two models an illness-death model with 3 states autonomy, LTC and death and a multi-state model with autonomy, death and 4 states of LTC which correspond to the 4 groups of pathologies observed in the data. Both models are defined using transition intensities: autonomous mortality, incidence in LTC and mortality in LTC (for all causes or for a given pathology group). We provide an expression of the overall mortality in LTC which depends on the incidence and mortality rates in LTC for each group thus linking both models. We introduce local likelihood methods that we use for direct inference of transition intensities in the model. Local likelihood methods are presented in the very didactic book from Loader (1999) and is implemented in the **locfit** package on the statistical software **R** (R Core Team, 2016). We rely on this implementation for the inference of mortality in LTC. Regarding the estimation of autonomous mortality and incidence rates in LTC, however, formulas need to be extended to account for left-truncated data and we rely on our own implementation of the method in that case. Local likelihood requires the specification of smoothing parameters such as the degree of the local fit and the bandwidth function. We rely on an adaptive bandwidth method where the bandwidth is defined as the distance to the k-th nearest neighbours. We test multiple combination of smoothing parameters and perform model selection based on Akaike information criterion (AIC). Once the inference of transition intensities has been completed, we derive a second-step estimate of overall mortality in LTC by combining incidence and mortality in LTC for each group of pathologies.

In the light of our results, impact of pathology groups on mortality in LTC proves substantial. Due to a sizeable difference in the initial level of mortality for *cancer* compared to other groups, a direct smoothing of the all-causes mortality in LTC leads to over-smooth the mortality surface in the region of low durations in LTC. On the other hand, smoothing separately the mortality in LTC associated with each group of pathology and combining the transition intensities results in a second-step estimate of the overall mortality in LTC allows us to overcome this difficulty. Studying mortality associated with each group of pathologies also provides useful information for the insurer, such as average duration of claim and cost associated with each group. At last, our results indicates that the singular shape of the curve for mortality in LTC with the high mortality right after claim inception is mostly due to the *cancer* group. This strongly advocates for estimating separately the mortality in LTC due to *cancer* and the mortality for other groups. When information about pathology is not available, one may still introduce a mixture model as in Biessy (2015a).

In this study, pathologies are gathered into 4 groups. Access to data with a higher level of details would allow to isolate some relatively frequent causes for LTC that are currently in the group of *other causes* such as cardiovascular diseases or muscular or skeletal diseases. Similarly, information about the type of cancer might be helpful to better model the mortality in LTC in the first few months following claim inception. Besides, the local likelihood estimates we provided could be used as a preliminary step in the setting-up of a parametric model. Indeed, while the local-likelihood estimate we obtain is useful on its own, a parametric model may yield better results when applied to a smaller portfolio with limited data. To this extent, our local-likelihood estimate makes for an interesting reference for comparison purposes. In addition, the degrees of freedom indicate how many parameters should ideally be used in the parametric model, though it may prove difficult due to the high amount of degrees of freedom in the selected fits. At last, we used the nearest neighbour methodology to select the bandwidth. While being an adaptive bandwidth method, it is still very basic and more complex methods among those presented in Tomas and Planchet (2013) may perform better.

---

[4] CNRS: Centre National de la Recherche Scientifique, France's largest public organism for scientific research.

References

Aalen, O. O. and S. Johansen (1978). An empirical transition matrix for non-homogeneous markov chains based on censored observations. *Scandinavian Journal of Statistics*, 141–150.

Biessy, G. (2015a). Continuous time semi-Markov inference of biometric laws associated with a Long-Term Care Insurance portfolio. Technical report, hal-01220564.

Biessy, G. (2015b). Long-term care insurance: A multi-state semi-Markov model to describe the dependency process in elderly people. *Bulletin Français d'Actuariat 15(29)*, 41–73.

Chichignoud, M. (2010). *Performances statistiques d?estimateurs non-linéaires*. Ph. D. thesis, Citeseer.

Cox, D. R. (1992). Regression models and life-tables. In *Breakthroughs in statistics*, pp. 527–541. Springer.

Czado, C. and F. Rudolph (2002). Application of survival analysis methods to long-term care insurance. *Insurance: Mathematics and Economics 31*(3), 395–413.

Denuit, M. and C. Robert (2007). *Actuariat des Assurances de Personnes*. Economica.

Eilers, P. H. and B. D. Marx (1996). Flexible smoothing with b-splines and penalties. *Statistical science*, 89–102.

Epanechnikov, V. A. (1969). Non-parametric estimation of a multivariate probability density. *Theory of Probability & Its Applications 14*(1), 153–158.

Guibert, Q. and F. Planchet (2014). Construction de lois d'expérience en présence d'évènements concurrents: Application à l'estimation des lois d'incidence d'un contrat dépendance. *Bulletin Français d'Actuariat 14*(27), 5–28.

Guibert, Q. and F. Planchet (2015). Non-parametric inference of transition probabilities based on aalen-johansen integral estimators for semi-competing risks data: application to ltc insurance. In *Conference of the LIFE Section of the International Actuarial Association*.

Haberman, S. and E. Pitacco (1998). *Actuarial Models for Disability Insurance*. Chapman and Hall/CRC, 1st edition.

Helms, F., C. Czado, and S. Gschlößl (2005). Calculation of ltc premiums based on direct estimates of transition probabilities. *Astin Bulletin 35*(02), 455–469.

Lepez, V., S. Roganova, and A. Flahault (2013). A semi-Markov model to investigate the different transitions between states of dependency in elderly people. In: Colloquium of the International Actuarial Association, Lyon.

Loader, C. (1999). *Local regression and likelihood*. Statistics and Computing. Springer-Verlag, New York.

Loader, C. R. (1996, 08). Local likelihood density estimation. *Ann. Statist. 24*(4), 1602–1618.

Pitacco, E. (2014). *Health Insurance - Basic Actuarial Models*. Springer International Publishing.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Tibshirani, R. and T. Hastie (1987). Local likelihood estimation. *Journal of the American Statistical Association 82*(398), 559–567.

Tomas, J. and F. Planchet (2013). Multidimensional smoothing by adaptive local kernel-weighted log-likelihood: Application to long-term care insurance. *Insurance: Mathematics and Economics 52*(3), 573–589.

Wand, M. P. and M. C. Jones (1995). *Kernel smoothing*, Volume 60 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, Ltd., London.