



Modelling the burden of long-term care for institutionalised elderly based on care duration and intensity

Martin Bladt¹, Michel Fuino¹, Aleksandr Shemendyuk¹  and Joël Wagner^{1,2*} 

¹Department of Actuarial Science, Faculty of Business and Economics (HEC Lausanne), University of Lausanne, Extranef, 1015 Lausanne, Switzerland; and ²Swiss Finance Institute, University of Lausanne, 1015 Lausanne, Switzerland

*Corresponding author. E-mail: joel.wagner@unil.ch

(Received 20 December 2021; revised 12 July 2022; accepted 17 July 2022; first published online 15 August 2022)

Abstract

The financing of long-term care and the planning of care capacity are of increasing interest due to demographic changes and the ageing population in many countries. Since many care-intensive conditions begin to manifest at higher ages, a better understanding and assessment of the expected costs, required infrastructure, and number of qualified personnel are essential. To evaluate the overall burden of institutional care, we derive a model based on the duration of stay in dependence and the intensity of help provided to elderly individuals. This article aims to model both aspects using novel longitudinal data from nursing homes in the canton of Geneva in Switzerland. Our data contain comprehensive health and care information, including medical diagnoses, levels of dependence, and physical and psychological impairments on 21,758 individuals. We build an accelerated failure time model to study the influence of selected factors on the duration of care and a beta regression model to describe the intensity of care. We show that apart from age and gender, the duration of stay before death is mainly affected by the underlying diseases and the number of different diagnoses. Simultaneously, care intensity is driven by the individual level of dependence and specific limitations. Using both evaluations, we approximate the overall care severity for individual profiles. Our study sheds light on the relevant medical, physical, and psychological health indicators that need to be accounted for, not only by care providers but also by policy-makers and insurers.

Keywords: Long-term care; Institutional care; Accelerated failure time; Beta regression; Empirical data

1. Introduction

Population ageing is one of the major challenges faced by the society of most developed countries. As longevity improves, pathologies and dependence that appear at higher ages put a strain on the old-age care systems, their organisation, and financial planning (Hirsch, 2005; Okma & Gusmano, 2020; Waitzberg *et al.*, 2020). At higher ages, it is common that elderly individuals present difficulties in performing activities that are part of their daily lives (see, e.g. Fuino & Wagner, 2018; Vanella *et al.*, 2020). In that sense, care delivered to maintain functional abilities is identified under the name of old-age long-term care (LTC). In most developed countries, the provision of care to elderly people and its financing are issues present in social policy discussions (Karlsson *et al.*, 2006; Le Corre, 2012; Duell *et al.*, 2019). Many studies have evidenced problems related to handling LTC needs (see, e.g. Pang & Warshawsky, 2010; Shao *et al.*, 2015, 2019). As critical points, they specify that current schemes will soon face a lack of financing (Kitchener *et al.*, 2006; Brown

& Finkelstein, 2009) in available care infrastructure (Katz, 2011; Cosandey, 2016) and in professional caregivers (Nichols *et al.*, 2010; Colombo *et al.*, 2011). This stresses the relevance of proper planning and integrates the societal and political evolutions of the scope of handling LTC, both in institutions and at home. Therefore, the emergence of high demand for LTC requires evaluating the capacity of the infrastructure, in particular, the availability of beds and the number of qualified caregivers in specialised institutions. The total care needs set the target for the governments that are ultimately responsible for providing elderly individuals with decent care. Defining an appropriate monitoring for determining the amount of care required is fundamental. The estimation of the overall amount of care also is crucial to secure the financing of care, a concern that involves the government, the insurers, and elderly individuals themselves.

In this paper, we study the factors that drive the total care burden of institutionalised elderly individuals. We express the overall burden in terms of the number of hours of care received while living in an institution. We call the *severity of care* the total number of hours. It can be modelled by multiplying the *duration of stay*, that is, the time spent in dependence in an institution, with the *intensity of care*, that is, the amount of help, expressed in time units, received per period. The duration of stay indicates the occupancy period of a bed in an LTC institution, while the intensity of care is associated with utilisation of medical and care resources, in particular, the number of minutes of help from nurses. We study both components and investigate the determinants that explain the duration and intensity. Building on longitudinal data, we first derive a model to estimate the number of months that elderly individuals stay in an institution. A second model evaluates the number of minutes of care that persons receive each week along the most influential covariates. Combining both pieces of information, we evaluate the overall care severity per elderly individual for different profiles.

The overall burden or severity of LTC can be determined in several ways. On the one hand, the overall institutions' costs stem from the duration in dependence. In fact, factors including the health state, socio-demographic characteristics, and physical and psychological impairments determine the time spent in institutional care (see, e.g. Hedinger *et al.*, 2015; Moore *et al.*, 2019). Moreover, different family situations affect the duration of stay in the institution, as well as the amount of care provided; see, for example, Pinquart & Sörensen (2011) and Mommaerts (2020). Mortality is another cause that shortens the duration. For example, patients with schizophrenia, mental disorders, tumours, and cognitive impairments suffer from higher mortality rates; see Campbell *et al.* (1985), Davidson *et al.* (1988) and Pack (2009). On the other hand, measuring the LTC severity also requires a metric for the intensity of care delivered by professionals (Carrino *et al.*, 2018). Indeed, cost analysis highlights that most expenses are devoted to the time that caregivers have to spend on each resident (Hu, 1986; Dorr *et al.*, 2005) to such an extent that, within OECD countries (i.e., member countries of the Organisation for Economic Co-operation and Development), patient profiles are defined in relation to the minutes of care required (Muir, 2017). From a medical perspective, doctors evaluate LTC severity based on medical metrics such as the level of functional limitations and cognitive troubles. Many researchers state that mental and physical diseases can affect functional abilities and therefore entail LTC needs (Anderson *et al.*, 1993; Guibert & Planchet, 2018; Fuino & Wagner, 2020). In this context, the most well-documented pathologies regarding LTC dependence are dementia and Alzheimer's (Arrighi *et al.*, 2010; Koroukian *et al.*, 2016; Farias *et al.*, 2017), heart failure, and high blood pressure (Kuo *et al.*, 2005; Sinclair *et al.*, 2008; Lesman-Leegte *et al.*, 2009), and to some extent cancer (Goodwin, 1991; Avis & Deimling, 2008; Chavan *et al.*, 2017). Therefore, individuals' pathology profiles affect the severity of the dependence and lead to diverse health outcomes and mortality patterns (Tomas & Planchet, 2013; Albarrán *et al.*, 2019; Jennings *et al.*, 2020). Finally, from a patient perspective, becoming care dependent means a change in lifestyle that is difficult to accept. When moving into an institutional setup, lifestyle habits are drastically altered, and elderly individuals feel a loss of intimacy (Hyer *et al.*, 2005). This is reflected in practice by the higher prevalence of depression syndromes among institutionalised elderly individuals (Boyle, 2005; Thakur & Blazer, 2008).

Our study is based on a longitudinal data set covering the whole institutionalised population of the Geneva canton in Switzerland over a 22-year period, from 1996 to 2018. We build our analysis on the records of 21,758 individuals filled out using the Canadian monitoring method “PLAISIR”. The method records a person’s pathologies, physical limitations, psychological and sensory impairments, the amount of care provided, and further personal information such as gender, date of birth, and date of death. Accounting for the right-censored nature of the data, we model the duration in institutional care using the survival approach, namely, an accelerated failure time (AFT) model with Weibull distribution, while the intensity of care is modelled using a beta regression.

We find that, after age and gender, the pathologies are the key drivers of the duration of stay. However, despite significantly affecting the intensity of care, diseases do not affect the amount of care provided to elderly individuals as much as dependence-related limitations and physical and psychological impairments. The latter two are the key indicators to describe the intensity of care required by an elderly individual. In contrast, physical limitations are less relevant in describing the duration of stay in dependence. Introducing profiles that relate to different types of health conditions, we find that, in general, women come with the highest total care burden. Furthermore, we show that mental and osteoarticular diseases lead to the highest overall care severity, a finding that results from the higher duration of stay. In contrast, elderly individuals with tumours have the lowest overall care burden, which can be explained by the higher mortality rate, and thus a lower expected duration of stay. Finally, given that the variations in care intensity are relatively small in our data, we find that the overall care severity is mainly shaped by the duration of stay.

The remainder of this paper is organised as follows. In Section 2, we lay out the research framework: we review the financing and monitoring of LTC costs in Switzerland and discuss the literature on the variables associated with care duration and intensity. In Section 3, we introduce the available data set, describe the variables, and present descriptive statistics. In Section 4, we introduce the model framework, including the AFT model that lays the basis for the duration analysis and the beta regression model used for the study of the intensity of care. We present and discuss the model results in Section 5. For selected profiles of elderly individuals, we evaluate how certain variables affect the overall care severity of an institutionalised elderly individuals. We conclude the paper in Section 6.

2. Research Framework

Our objective is to model the total care burden in terms of the total number of hours of care an institutionalised elderly individual receives. Therefore, our model relies on the time spent in dependence, the amount of help received, and their respective determinants. On the one hand, the duration of stay defines the time of occupancy of a bed in an LTC institution. The contributions of the government directly relate to it, disregarding the specific pathology. On the other hand, the intensity of care relates to the usage of labour from nurses. It strongly depends on the medical condition of elderly individuals. Payments of health insurance are typically associated with the required number of minutes of care. Since our study uses data from Switzerland, we lay out the costs and the financing of the Swiss care system for elderly dependents in this section. Indeed, the costs are directly linked to the duration of stay and the intensity of care. In Section 2.1, we describe the organisation and cost monitoring of LTC institutions. In Section 2.2, we review the literature on the drivers influencing the care burden.

2.1 Financing and monitoring of LTC costs in Switzerland

Switzerland is a federal state consisting of 26 cantons distributed among the German, French, and Italian linguistic regions. Rules for the care of elderly individuals and its financing are defined at both the federal and cantonal levels. The federal base framework is tailored along cantonal

rules accounting for specific situations. In its broadest definition, LTC denotes care delivered to elderly individuals having difficulties performing daily life activities, often identified through the number of limitations in “activities of daily living” (ADL) and “instrumental activities of daily living” (IADL, see, e.g. Kempen *et al.*, 1995). While this definition, at least in theory, appears valid in most developed countries (see, e.g. Fuino *et al.*, 2020), getting an appropriate definition of LTC in practice is more controversial, in particular, when such a definition influences political and budget decisions and ultimately the amount of care provided and financed by insurers. While in countries such as the US, more than ten LTC systems are listed (Seematter-Bagnoud *et al.*, 2012), LTC in Switzerland is either provided at home or in an institution. Furthermore, we observe three categories of curative LTC treatments: nursing, personal hygiene, and ergotherapy (Home Care Association of the Canton of Vaud, 2020). While the first two categories relate to the ability to perform (I)ADL, the third category emphasises the importance of social aspects beyond the ability to perform specific activities.

The financing of LTC relies on social health insurance, the state government, and out-of-pocket payments by the dependent elderly individual. Swiss mandatory health insurance covers all costs defined by the health care benefits ordinance (Swiss Federal Department of Home Affairs, 2021). The contribution to the care costs is defined on a scale along the required number of minutes of care¹. The state government also participates in funding LTC costs. All Swiss residents aged 65+ years in need of LTC are eligible to receive an allowance regulated under the old-age and survivor’s insurance law². The pre-defined amount increases with the acuity level. The state government also indirectly participates in the financing of institutional care, for example, by constructing new infrastructure and providing further means-tested allowances for those who cannot afford the costs. Finally, households are responsible for a set of non-covered care mostly related to accommodation expenses (lodging, feeding, and laundry) in institutional care. Furthermore, since 2011 and to limit the increase in health insurance premiums, copayments of up to 20% of the costs are required from residents (Swiss Federal Social Insurance Office, 2010). In 2016, the overall monthly cost of a stay in a Swiss institution is estimated to be CHF 9,652 (Social Insurance and Accommodation Service of the Canton of Vaud, 2016).

The development of the elderly population in recent decades has increased LTC costs. Against this background, it is essential to assess the overall dynamic to predict the needs of nursing homes and qualified personnel in the future. Based on the findings from Donabedian (1973), an appropriate monitoring method for nursing homes accounts for three dimensions, namely, the patient health condition, the type of service required, and the resources needed (Roussel & Tilquin, 1993). Under these conditions, two types of monitoring have been implemented in Switzerland since the 1990s. The cantons of Vaud, Geneva, Neuchâtel, and Jura have implemented the Canadian monitoring method “PLAISIR”, which stands for *Planification Informatisée des Soins Infirmiers Requis*³, while the other cantons have implemented the American “Resource Utilization Groups” monitoring method⁴. Even though significant differences appear in how data are gathered, the results of both methods remain comparable and allow the development of care plans based on estimates of the needs for nursing care and assistance. Indeed, the classifications are articulated around the patient health conditions, the type of care service required and the staff resources needed. Finally, we note that an assessment of the costs of care includes knowledge about the duration of stay and the intensity of care provided.

¹Contributions increase along a scale with 12 levels. The cover amounts to CHF 9.60 per day for up to 20 minutes of daily care and increases by the same amount for each increment of 20 minutes up to a maximum contribution of CHF 115.20 for treatments exceeding 220 minutes/day.

²See www.ahv-iv.ch/en/Social-insurances/Old-age-and-survivorss-insurance-OASI.

³See the Commission intercantonale PLAISIR at www.ctplaisir.ch.

⁴See the Centers for Medicare & Medicaid Services at www.cms.gov.

2.2 Review of the determinants of the duration of stay and intensity of care

In our modelling, we separately assess the time spent in dependence and the amount of help received per period. Both dimensions can be investigated through demographic variables and linked to medical diagnoses, the inability to perform (I)ADLs, limitations in physical and cognitive activities, and impairments of psychological and sensory functions. In the following, we discuss a selection of relevant variables outlined in the extant literature.

A vast amount of literature agrees that age and gender are both relevant determinants of the duration in dependency (see Mathers, 1996; Germain *et al.*, 2016; Fong *et al.*, 2017; Fuino & Wagner, 2020) and the intensity of care (see, e.g. de Meijer *et al.*, 2011; Xue, 2011). For example, elderly individuals at high ages are more susceptible to developing multiple types of diseases (van den Akker *et al.*, 1998) that lead to higher mortality rates (Menotti *et al.*, 2001) and reduce the duration of stay in an institution. At the same time, Deeg *et al.* (2002) find that women have a higher life expectancy than men, regardless of the multimorbidity profile. Moreover, Rickayzen & Walsh (2002) observe that women are more likely than men to develop a higher dependency from external help across all ages and thus require a higher intensity of care.

The World Health Organization (1980) and the Institute of Medicine (1991) introduce the relationships among pathologies, impairments of organ systems, and the resulting limitations or disabilities. In general, a disease causes certain organ malfunctions that result in a loss or abnormality of mental, emotional, or physiological structures. Furthermore, these impairments lead to functional limitations, that is, lack of ability to perform an action or activity in a manner considered normal, and overall disability, that is, limitation in performing socially defined activities and roles. As a consequence, different pathologies affect both mortality and the dependence, directly affecting the duration of stay, and the intensity of care, respectively. Furthermore, the influence of pathologies on LTC needs is widely studied by, for example Boulton *et al.* (1994), Guccione *et al.* (1994), Tomiak *et al.* (2000), Pritchard (2006), Callahan *et al.* (2012), Biessy (2017) and Rudnitskyi & Wagner (2019). Finally, research suggests accounting for the number of diseases, that is, multimorbidity (Marengoni *et al.*, 2011; Barnett *et al.*, 2012).

Geriatric syndromes are conditions commonly experienced by older individuals (see Inouye *et al.*, 2007), including visual and hearing impairments, depressive symptoms, low cognitive performance, persistent dizziness, or lightheadedness. Branch & Jette (1982) show that similar impairments do not significantly influence the decision to enter a nursing home. However, Koroukian *et al.* (2016) find that, when explaining the health status of a person, accounting for the co-occurrence of functional limitations and geriatric syndromes is more informative than considering chronic conditions alone. This means that impairments in psychological and sensory functions should add to the explanation of the duration of stay and the intensity of care. Further works on the influence of impairments, limitations, and disabilities and LTC needs are those of Branch & Jette (1982), Miller & Weissert (2000), Tomiak *et al.* (2000), Rickayzen & Walsh (2002), and de Meijer *et al.* (2011).

3. Data on Institutional LTC and Descriptive Statistics

In the following, we describe the data used in our study. In Section 3.1, we provide information on the available data and define the variables used in the analysis. In Section 3.2, we report descriptive statistics on the duration of stay in LTC and care intensity. Exploratory data analysis allows us to substantiate the choice of the methodological approach for modelling.

3.1 Available data and description of variables

Our study is based on a private data set containing observations on elderly individuals who have received institutional care in the canton of Geneva in Switzerland during the period from 1996 to

Table 1. Description of the variables.

Variable	Description	Values
Duration and intensity of care		
<i>D</i>	Duration of stay in the institution	Positive (in months)
<i>T</i>	Intensity of care provided per week	Positive from 0 to 10,080 (in minutes)
<i>C</i>	Right-censoring indicator	Yes, no
Demographic variables		
<i>AG</i>	Age at entry in the institution	65, 66, 67, . . . years
<i>GE</i>	Gender	Male, female
Medical diagnoses		
<i>D1</i>	Main diagnosis	Mental, cerebrovascular, nervous, osteo-articular, heart, tumour, other problem
<i>D2</i>	Secondary diagnosis	See <i>D1</i> , plus “none”
<i>ND</i>	Number of additional diagnoses	0, 1, . . . , 7
Levels of dependence		
<i>DP</i>	Dependence in ADL	Four levels: 1–6, 7, 8, 9
<i>PM</i>	Physical mobility limitations	Five levels: 1–5, 6, 7, 8, 9
<i>OR</i>	Orientation problems	Five levels: 1–4, 5, 6, 7, 8+
<i>OC</i>	Occupational limitations	Five levels: 1–5, 6, 7, 8, 9
<i>SI</i>	Social integration limitations	Five levels: 1–4, 5, 6, 7, 8+
Impairments of psychological and sensory functions		
<i>RM</i>	Recent memory impairment	Adequate, mild, moderate, severe
<i>PA</i>	Perception and attention impairments	Adequate, mild, moderate, severe
<i>IM</i>	Impairment in impulses	Adequate, mild, moderate, severe
<i>WM</i>	Will and motivation impairment	Adequate, mild, moderate, severe
<i>BH</i>	Behavioural impairment	Adequate, mild, moderate, severe
<i>VS</i>	Visual impairment	Adequate, mild, moderate, severe

2018. The anonymous individual data stem from the evaluation tool used to assess the care needs (Republic and Canton of Geneva, General Directorate for Health, 2019). The data contain information on individuals, their medical diagnoses and comprehensive details on their limitations and impairments. Furthermore, it provides information on the date of entry into the care institution, the number of minutes of help provided each week and, if applicable, the date of death. For this study, we retain a set characterising 21,758 individuals, focusing on elderly individuals aged 65 years or older⁵. After entering an institution, every person passes an initial medical screening. Various tests examine the overall state of health, pathologies, and physical and mental health disorders. During an observation period, the intensity of care required by the person and expressed as the number of minutes of care per week is recorded. While such tests are typically repeated approximately every 1–2 years, we focus on the first evaluation made at entry. Of the available information (for the data collection methodology, see Roussel & Tilquin, 1993), and following the literature review in Section 2.2, we consider 19 key variables that characterise each record in our data (see Table 1).

⁵The set of data used in this study covers 93.3% of the original data retrieved from the evaluation tool. Thereby, we removed individuals who departed from the institution before death (3.82%), who were younger than 65 years (2.67%), or who presented empty records or corrupt data (0.21%).

Duration and intensity of care. The original data contain the date of admission in the institution and the date of death if applicable. If the person died before the end of the observation period, the difference between the date of death and the date of entry allowed us to calculate the duration of stay D in months⁶. If the date of death was empty, the person was still alive at the date of data extraction. In this case, we calculate the duration D as the difference between the date of entry and the latest date available in the data set, which is August 21st, 2018. Additionally, we create a separate indicator variable C telling us if the person is alive or dead, that is, indicating if the data are right-censored or not.

The intensity of care is recorded as the total number of minutes of care T provided to the person per week, that is, a number between 0 and 10,080, where the upper bound corresponds to the total number of minutes in 1 week. This number includes the time of care given for respiratory help (respiratory exercises, chest physical therapy), eating and drinking (providing vitamins, verifying the diet), elimination (giving, removing, and emptying urinal, maintaining hygiene and skin integrity), hygiene (personal, hair, and beauty care), mobility (pushing wheelchair, getting up, or lying down), communication (supportive communication, teaching, group activities), medication needs, intravenous therapy, and other treatments (dressing wounds, blood pressure). The full list and descriptions are available from Roussel & Tilquin (1993).

Demographic variables. Using the date of birth and the date of admission, we calculate the age at entry into institution AG by taking the number of full years that have passed since birth until entry. In our data, the youngest age at entry is 65 years, while the oldest person has entered institutional care at the age of 106. Furthermore, information on the gender GE of the individual is available in the data. It is recorded as a binary factor with levels “male” and “female”.

Medical diagnoses. Each observation carries information about the person’s disorders. Individuals may have received several medical diagnoses and up to nine are recorded in the data. They are ranked by importance by a doctor. We label the most important (“first”) diagnosis as $D1$ and explicitly consider one secondary diagnosis $D2$. The variable ND indicates the number of additional diagnoses beyond the first two. $D1$ and $D2$, with ND , allow us to account for the top two medical conditions and possible interactions, as well as to have information on the number of additional health problems that relate to the overall severity.

Diagnoses are encoded using the International Classification of Diseases⁷. We reduce the number of unique diagnoses by aggregating the diseases into six groups, namely, mental, cerebrovascular, nervous, osteoarticular, heart, and tumour diseases. All other diagnoses are grouped in a category labelled “other”⁸. The main diagnosis $D1$ is a factor variable with seven levels. The

⁶In approximately 20% of the records, typically, in older observations, our data do not include the exact date of death. In such cases, we rely on information on the period of the year when death has occurred. Indeed, the time of death is available through the period of the year coded through “January–March”, “April–May”, “June–August”, “September–October”, and “November–December”. This information enables us to approximate the date of death and the duration of stay using the middle point of the period.

⁷See www.who.int/standards/classifications/classification-of-diseases. Since the data collection dates back to 1996, earlier diagnoses have been encoded using the old ICD-9 standard. Since 2007, the contemporary ICD-10 classification has been used. To avoid inconsistencies, we map all ICD-9 encodings into ICD-10 using the conversion table provided by the Centers for Medicare & Medicaid Services (2014).

⁸The group of mental diagnoses includes dementia, schizophrenia, depressive disorders, mental retardation, and neurotic disorders. Next, cerebrovascular diagnoses include subarachnoid or intracerebral haemorrhage, strokes, cerebral infarction, and other sequelae of cerebrovascular diseases. The group of diagnoses labelled “nervous” include Alzheimer’s disease, brain degeneration, epilepsy, extrapyramidal syndromes and movement disorders (e.g., Parkinson’s disease), and sclerosis. Osteoarticular diagnoses consist of arthritis, scoliosis, osteochondritis, damage to cervical disks, deformation of the limbs, and osteoporosis. Heart diseases include hypertension, ischaemic heart diseases, myocardial infarction, cardiac arrhythmias, heart failure, diseases of the blood transportation system, and others. The group of tumour diagnoses combines all types of tumours, for example, breast tumours, digestive system tumours, skin tumours, and tumours of the respiratory organs. Finally, the group of other diagnoses comprises endocrine, nutritional, metabolic (including obesity and diabetes), and respiratory diseases.

secondary diagnosis *D2* is a factor variable with eight levels since we include the possibility of no second diagnosis labelled “none”. Note that the same disease group may appear several times, for example, when a person has several disorders from the same disease group.

Level of dependence. The level of dependence is measured along five dimensions: limitations with ADL, physical mobility, orientation, occupation, and social integration. As with the World Health Organization (1980), the recorded limitations are measured on ordered nine-level scales, where each level corresponds to a particular severity. In Table 2, we describe the limitations for the different levels in each dimension (see also Roussel & Tilquin, 1993). Since individuals entering institutional care are mostly moderately or severely dependent, we observe very few records showing lower levels of dependence. Therefore, we aggregate the levels so that they constitute a share of at least approximately 10%–15% of the data (see Table 1 and the descriptive statistics in Section 3.2). This allows for a lower number of categories in the further modelling. For example, for the social integration limitations variable *SI*, we consolidate the levels from 1 to 4 into “1–4”, yielding 8.9% of the individuals.

First, the level of dependence in ADL (*DP*) considers the physical dependence in performing ADL. It refers to the individual’s ability to complete, independently, the basic ADL (e.g., personal hygiene, eating, dressing) and the IADL (e.g. housekeeping, cooking). The assessment of a person’s abilities does not consider the institutional environment, that is, it compares the individual’s potential to perform (I)ADL to a usual healthy person of the same age and gender. Next, the variable *PM* measures the limitations of physical mobility, that is, the ability to move effectively in the surroundings. The evaluation considers the independent use of mechanical aids (e.g., prosthesis, wheelchair, cane) but not the aid given by other persons. The principal indicator is the distance to which the person can move away from the bed or the chair. The capacity for orientation and interactions with the environment is coded in the variable *OR*. This concept includes the reception of signals from the environment, their assimilation and the formulation of a response. Living and occupations throughout the day are assessed through the limitations in time and type in *OC*. This concept refers to the person’s ability to occupy the time in a manner customary for the age and gender group within the institutional environment. Here, all activities related to employment, recreation, education, creation, and customary everyday tasks are included. The difference between the levels “no occupation” and “inappropriate occupations” stems from the ability to perform activities where the first refers to persons who are incapable of sustaining any form of activities, while the second refers to those who do activities without a defined goal. Finally, the social relations and their limitations are recorded in *SI*. This concept refers to the person’s ability to participate in social activities and maintain adequate social relations, considering life in an institutional setting.

Impairments of psychological and sensory functions. Roussel & Tilquin (1993) propose an ordered four-level scale, with the levels adequate, mild, moderate, and severe, to measure the severity of 16 psychological and sensory function impairments⁹. The evaluation considers any compensation used by the person (glasses, medication that corrects psychological impairments) and compares his or her performance to the average performance of a healthy individual of the same age and gender. For certain functions such as recent memory or sight, it is possible to describe precise definitions for the four levels. For other functions, the person’s state is assessed more qualitatively.

The impairments are closely related to the above-discussed medical diagnoses and levels of dependence. In the forthcoming models, we reduce complexity and keep only those impairments that help explain institutional LTC. Following our model selection (see Section 5), we

⁹Recent memory, long-term memory, thinking (content, speed), perception and attention, consciousness and wakefulness, orientation (time/person/space), decision-making, impulses (drives), will and motivation, emotions (including feelings and mood), behaviour, language, sight, hearing, making self understood, and understanding others.

Table 2. Measurement of the level of dependence.

Level	Dependence in ADL, <i>DP</i>	Physical mobility limitations, <i>PM</i>	Orientation problems, <i>OR</i>	Occupational limitations, <i>OC</i>	Social integration limitations, <i>SI</i>
1	Independence	Full mobility	Full orientation	Appropriate occupations	Socially integrated
2	Independence subject to mechanical assistance	Occasionally limited mobility	Fully compensated orientation problems	Intermittent occupations	Inhibited participation (discomfort, shyness)
3	Independence subject to adaptation of environment	Deficient mobility (slowness)	Intermittent orientation disturbances	Occup. limited in scope	Limited participation (type of social activities)
4	Predictable need for help in certain situations	Limited mobility in general	Partially compensated orientation disturbances	Adjusted occupations	Limited relations (only primary and secondary contacts)
5	Predictable need for help up to once a day	Limited to the neighbourhood	Moderate disturbances of orientation	Occup. limited in time	Poor relations (difficulties with secondary contacts)
6	Predictable need for help more than once a day	Limited to the institution	Severe disturbances of orientation	Occup. limited in type	Reduced relations (only primary contacts)
7	Unpredictable or quasi-permanent need for help	Limited to the floor	Deprivation of orientation	Occup. limited in time and type	Disturbed relations (difficulties with primary contacts)
8	Need for help with most ADL	Limited to the room	Disorientation	No occupation	Non-existent social relations (no contacts due to incapacity)
9	Need for help with all ADL	No mobility (restricted to bed or chair)	Coma or vegetative state	Inappropriate occupations	Social isolation (cut off from the outside)

retain six of the 16 impairments available in the data: Limitations in the recent or short-term memory *RM* refer to the individuals' ability to store new information. The "adequate" level is assigned to those who have no memory problems, while "severe" refers to those who can name up to one of three objects mentioned or shown 5 minutes earlier. Perception and attention *PA* refers to the functions that allow an individual to receive information, process it, and concentrate on certain aspects. Impairments of perception include disturbances of the perception of one's own body, time, place, hallucinations, and difficulties in differentiating fantasy from reality. Impairments of attention include inattentiveness, distractibility, and inability to change the focus of attention. Impairments in impulses or drives *IM* refers to the increase, decrease, and change of form of different behaviours related to basic physiological needs or instincts (e.g., anorexia, bulimia, dependence on alcohol, or tobacco). Will and motivation impairments *WM* refer to disturbances in the ability to orient one's behaviour, control one's actions, and pursue a goal. The evaluation considers, for example, a lack of initiative, overcompliance, excessive cooperation, and compulsion. Behavioural impairments *BH* refer to patterns of behaviour that interfere with social adjustment and functioning. These patterns may be present since adolescence and throughout adult life or may appear due to neurological or mental illness. They mainly manifest themselves as accentuated character traits (e.g. suspiciousness, excessive shyness, worrying, self-destruction, indecisiveness). Visual impairments *VS* refer to the person's ability to see and are assessed considering corrected eyesight, for example, with eyeglasses. The "mild" level corresponds to a person who cannot read regular print but can read large prints. The "moderate" level person is unable to read but can follow an object with the eyes. Finally, the "severe" level relates to blindness.

3.2 Descriptive statistics

In the following, we present the descriptive statistics for the duration of stay D and the intensity of care T . Recall that the available data cover a fixed period that terminates at the date of data extraction. The data include $n = 21,758$ persons who entered a care institution: 17,919 (82.4%) of them died in the observation period, while 3,839 (17.6%) were still alive at the time of data extraction. Due to this right censoring, we cannot directly calculate the mean duration of stay D , and thus, we use survival analysis techniques.

The standard way to obtain median estimates of the duration of stay D is to apply the Kaplan–Meier product-limit estimator, a non-parametric estimator based on the survival curve proposed by Kaplan & Meier (1958). Indeed, the Kaplan–Meier estimate allows for right censoring and left truncation in seriatim data. To report the median duration of stay D_{med} across multiple factors, we apply Kaplan–Meier estimates on subsets of the data. In Table 3, we present the median duration of stay D_{med} and the mean intensity of care T_{avg} for the different variables' categories¹⁰.

Demographic variables. We divide the age at entry AG into six classes to illustrate the underlying distribution. We observe that the Kaplan–Meier estimates of the median duration of stay D_{med} are decreasing with the age at entry, which is due to increasing mortality rates at higher ages. At the same time, the mean intensity of care T_{avg} provided to the person fluctuates around the same value of approximately 16 hours/week. In groups of persons aged 100 years or more at entry, the intensity of care increases to 20.5 hours/week.

Most of the elderly individuals in our data are women. Men constitute just above a quarter (27.5%) of the data. We note that men stay dependent for a shorter amount of time than women, with their median duration being almost 15 months lower. Furthermore, men require on average 1.5 hours of care more per week. The prevalence and higher median duration of stay of women

¹⁰For the lower levels of *DP*, *PM*, *OR*, *OC*, and *SI* that contain very few observations associated with similar values of D_{med} and T_{avg} , we merge the first four, five, or six levels together until the cumulative share reaches approximately 10% to 15% of the data.

Table 3. Descriptive statistics on the median duration of stay D_{med} (in months) and the mean intensity of care T_{avg} (in minutes/week).

	<i>n</i>	%	D_{med}	T_{avg}		<i>n</i>	%	D_{med}	T_{avg}
Right censoring <i>C</i>					Orientation problems <i>OR</i>				
Yes	3,839	17.6	–	–	1–4	3,574	16.4	52.4	530
No	17,919	82.4	–	–	5	7,524	34.6	37.3	792
Age at entry <i>AG</i>					6	5,837	26.8	31.2	1,098
65–69	647	3.0	68.6	931	7	2,414	11.1	30.9	1,276
70–79	3,887	17.9	48.0	984	8+	2,409	11.1	28.6	1,469
80–89	11,323	52.0	38.3	945	Occupational limitations <i>OC</i>				
90–99	5,745	26.4	27.0	969	1–5	2,062	9.5	61.0	435
100+	156	0.7	15.9	1,231	6	5,933	27.3	40.9	706
Gender <i>GE</i>					7	10,226	47.0	31.5	1,068
Female	15,776	72.5	40.2	937	8	2,748	12.6	27.0	1,392
Male	5,982	27.5	25.9	1,021	9	789	3.6	35.2	1,333
Main diagnosis <i>D1</i>					Social integration limitations <i>SI</i>				
Mental	7,463	34.3	38.6	958	1–4	1,937	8.9	54.6	516
Cerebrovascular	1,267	5.8	35.2	1,139	5	5,221	24.0	42.8	689
Nervous	3,974	18.3	36.0	1,125	6	7,666	35.2	33.5	950
Osteoarticular	1,567	7.2	44.8	800	7	5,385	24.7	29.7	1,235
Heart	1,860	8.5	35.2	757	8+	1,549	7.1	29.9	1,519
Tumour	390	1.8	8.7	989	Recent memory impairment <i>RM</i>				
Other	5,237	24.1	33.4	910	Adequate	1,554	7.1	59.4	561
Secondary diagnosis <i>D2</i>					Mild	5,974	27.5	41.1	718
None	379	1.7	62.2	820	Moderate	8,347	38.4	33.5	955
Mental	5,584	25.7	37.6	988	Severe	5,883	27.0	30.2	1,317
Cerebrovascular	713	3.3	33.3	1,087	Perception and attention imp. <i>PA</i>				
Nervous	1,532	7.0	36.3	1,144	Adequate	1,831	8.4	56.5	528
Osteoarticular	2,191	10.1	43.9	872	Mild	7,207	33.1	40.5	738
Heart	3,570	16.4	34.5	866	Moderate	11,710	53.8	30.9	1,116
Tumour	345	1.6	12.2	1,012	Severe	1,010	4.6	37.5	1,512
Other	7,444	34.2	34.1	964	Impulse impairment <i>IM</i>				
Number of additional diagnoses <i>ND</i>					Adequate	2,608	12.0	53.2	538
0	1,468	6.7	54.5	856	Mild	8,168	37.5	39.0	783
1	2,202	10.1	48.6	876	Moderate	9,922	45.6	30.6	1,147
2	3,007	13.8	43.5	912	Severe	1,060	4.9	25.0	1,612
3	3,068	14.1	37.4	943	Will and motivation imp. <i>WM</i>				
4	2,991	13.7	33.9	956	Adequate	1,559	7.2	55.7	505
5	2,454	11.3	33.5	979	Mild	6,172	28.4	43.1	682
6	2,122	9.8	29.6	999	Moderate	12,039	55.3	32.1	1,068
7	4,446	20.4	26.1	1,053	Severe	1,988	9.1	27.8	1,524
Dependence in ADL <i>DP</i>					Behavioural impairment <i>BH</i>				
1–6	3,643	16.7	53.7	414	Adequate	1,388	6.4	52.9	542
7	8,856	40.7	40.6	780	Mild	6,783	31.2	39.9	735
8	7,623	35.0	27.9	1,274	Moderate	12,734	58.5	32.2	1,089
9	1,636	7.5	17.9	1,686	Severe	853	3.9	41.0	1,502

Table 3. Continued.

	<i>n</i>	%	<i>D</i> _{med}	<i>T</i> _{avg}		<i>n</i>	%	<i>D</i> _{med}	<i>T</i> _{avg}
Physical mobility limitations <i>PM</i>					Visual impairment <i>VS</i>				
1-5	2,256	10.4	61.1	421	Adequate	5,099	23.4	47.5	773
6	6,315	29.0	44.1	689	Mild	11,445	52.6	34.4	975
7	5,534	25.4	34.2	970	Moderate	4,631	21.3	31.2	1,093
8	2,195	10.1	28.5	1,052	Severe	583	2.7	28.9	1,229
9	5,458	25.1	24.1	1,449					

can be explained by their higher life expectancy; see, e.g., Mathers *et al.* (2001), Fong *et al.* (2017), Schünemann *et al.* (2017), and Fuino & Wagner (2018).

Medical diagnoses. Mental diagnoses are those with the highest prevalence in persons entering institutional care. While they rank first in the main diagnosis *D1*, mental ranks second in the secondary diagnosis *D2*, after the group of other diagnoses. Pathologies of the nervous system and heart problems also show a high prevalence in both the main and secondary diagnoses. We observed similar values for the median duration of stay across the different main diagnoses, except for persons with osteoarticular problems and tumours. Indeed, osteoarticular pathologies are associated with higher median durations of stay by more than half a year (total 44.8 months), while half of the tumour patients die after 8.7 months. In contrast, only the groups with cerebrovascular and nervous pathologies increased the mean intensity of care. Those with heart diseases as the main diagnosis require the least amount of help during the week. It is remarkable that if a person has one sole diagnosis, that is, the secondary diagnosis *D2* is “none”, the median duration of stay in the institution almost doubles, being slightly above 5 years (62.2 months).

Considering the main diagnosis *D1*, Figure 1 extends the results of the Kaplan–Meier estimates of the duration of stay *D* and of the intensity of care *T* distribution from the descriptive statistics in Table 3. From Figure 1(a), we see that the duration of stay *D* is drastically reduced by a tumour diagnosis, while osteoarticular pathologies come with higher survival rates. The kernel density estimation in Figure 1(b) sheds light on the influence of the main diagnosis on the intensity of care¹¹. Overall, we note that the distribution of *T* is bimodal, with peaks at approximately 450 minutes (7.5 hours) and 1,200 minutes (20 hours) of care per week. We see that the intensity of care in the nervous and cerebrovascular pathologies is left-skewed, which results in higher mean values. The other diagnoses yield distinct bimodal distributions, which raises the hypothesis that diagnoses, on their own, are insufficient to explain the intensity provided to a person.

Most dependent persons have multiple diseases with several additional diagnoses *ND*, with the highest prevalence being found at three and seven additional diagnoses, respectively. The individuals who have seven additional diagnoses are characterised by the lowest median duration of stay (26.1 months) and the highest mean intensity of care (1,053 minutes/week). With an increasing number of additional diagnoses, the median duration of stay decreases and the mean intensity of care slightly increases.

Level of dependence. From Table 3, we see that all variables representing the dependence (i.e., dependence in ADL *DP*, physical mobility limitations *PM*, orientation problems *OR*, occupational limitations *OC*, social integration limitations *SI*) follow the intuition that the higher the dependence level is, the lower the median duration of stay and the higher the mean intensity of care are. All variables representing limitations affect the median duration of stay *D*_{med} and the mean

¹¹We use the kernel density estimator given by $\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K(\frac{x-T_i}{h})$, where T_i is the intensity of care of the individual i , $n = 21,758$ is the number of individuals, the kernel $K(z) = (2\pi)^{-1/2} \exp(-z^2/2)$ is the standard normal density function, and $h = 0.9 \min(\hat{\sigma}, \text{IQR}/1.34) n^{-0.2}$ is a smoothing parameter with sample standard deviation $\hat{\sigma}$ and interquartile range IQR Silverman (1986), p. 45.

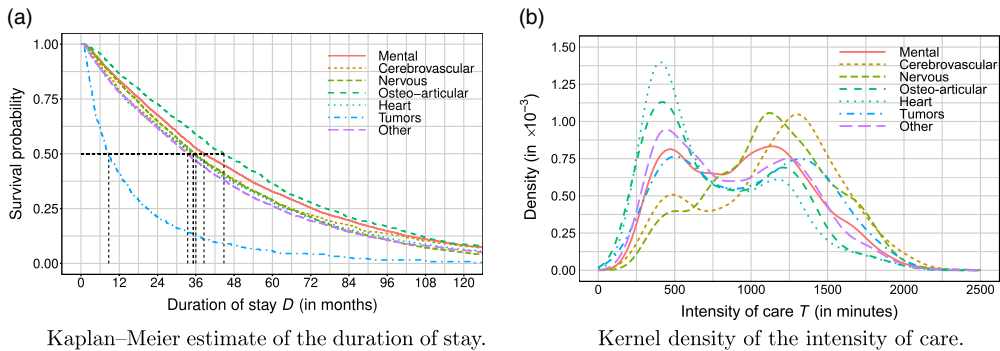


Figure 1. Kaplan–Meier estimation of the duration of stay D (in months) and kernel density estimation of the intensity of care per week T (in minutes) across main diagnoses $D1$.

intensity of care T_{avg} . The spread between the lowest and the highest levels is up to 3 years of stay (from 61.1 to 24.1 months in PM) and more than 20 hours of care per week (from 414 to 1,686 minutes in DP).

In particular, we note that 35% of the institutionalised elderly individuals (Level 8 in dependence in ADL DP , see Table 2) require help for most of their daily needs, and 7.5% require constant aid from the personnel (Level 9 in DP). Almost a quarter of elderly individuals are restricted to a bed or chair in terms of mobility (Level 9 in PM), 10.1% are limited to their room (Level 8 in PM), while the rest can move around in the institution. For both dependence in ADL DP and physical mobility limitations PM , we see a clear distinction in the mean intensity of care between the different severity levels. Simultaneously, we observe 22.2% suffering from severe impairment of orientation or complete disorientation (Levels 7 and 8+ in orientation problems OR). Those individuals, on average, require a higher intensity of care, comparable to those who are highly dependent on ADL or highly limited in physical mobility. The most prevalent level of occupational limitations OC is 7, that is, occupations limited in time and type, supporting the decision to institutionalise for these elderly individuals. Furthermore, most of the persons socialise only with primary contacts (Level 6 of social integration limitations SI). Indeed, the prevalence of the isolated elderly individuals living in nursing homes is at least twice as high as that of the community-dwelling population (Victor, 2012).

We extend the results obtained in Table 3 by plotting the Kaplan–Meier and kernel density estimates for the dependent variables in Figure 2. The graphs demonstrate clear differences in the survival probability along D and in the density of T across different levels of dependence. While many curves can be well separated, the representing curves of some levels are intertwined, as with the intensity of care across levels of OC in Figure 2(h), or are close to each other, as with the curves representing the intensity of care in Levels 7 and 8 (floor and room limitations, respectively) of physical mobility PM .

Impairments of psychological and sensory functions. We observe a similar impact of the impairments (recent memory impairments RM , perception and attention impairments PA , impulse impairments IM , will and motivation impairments WM , behavioural impairments BH , visual impairments VS). As expected, we observe that when the level of psychological or sensory function impairment worsens, the median duration of stay D_{med} decreases, and the mean intensity of care T_{avg} increases. Regarding dependence, impairments importantly affect the median duration of stay D_{med} and the mean intensity of care T_{avg} , with spread of approximately two and a half years (from 59.4 to 30.2 months for RM) and approximately 18 hours or care per week (from 538 to 1,612 minutes in IM). For all variables, except visual impairments VS , the most prevalent level of psychological and sensory function impairments is the moderate level.

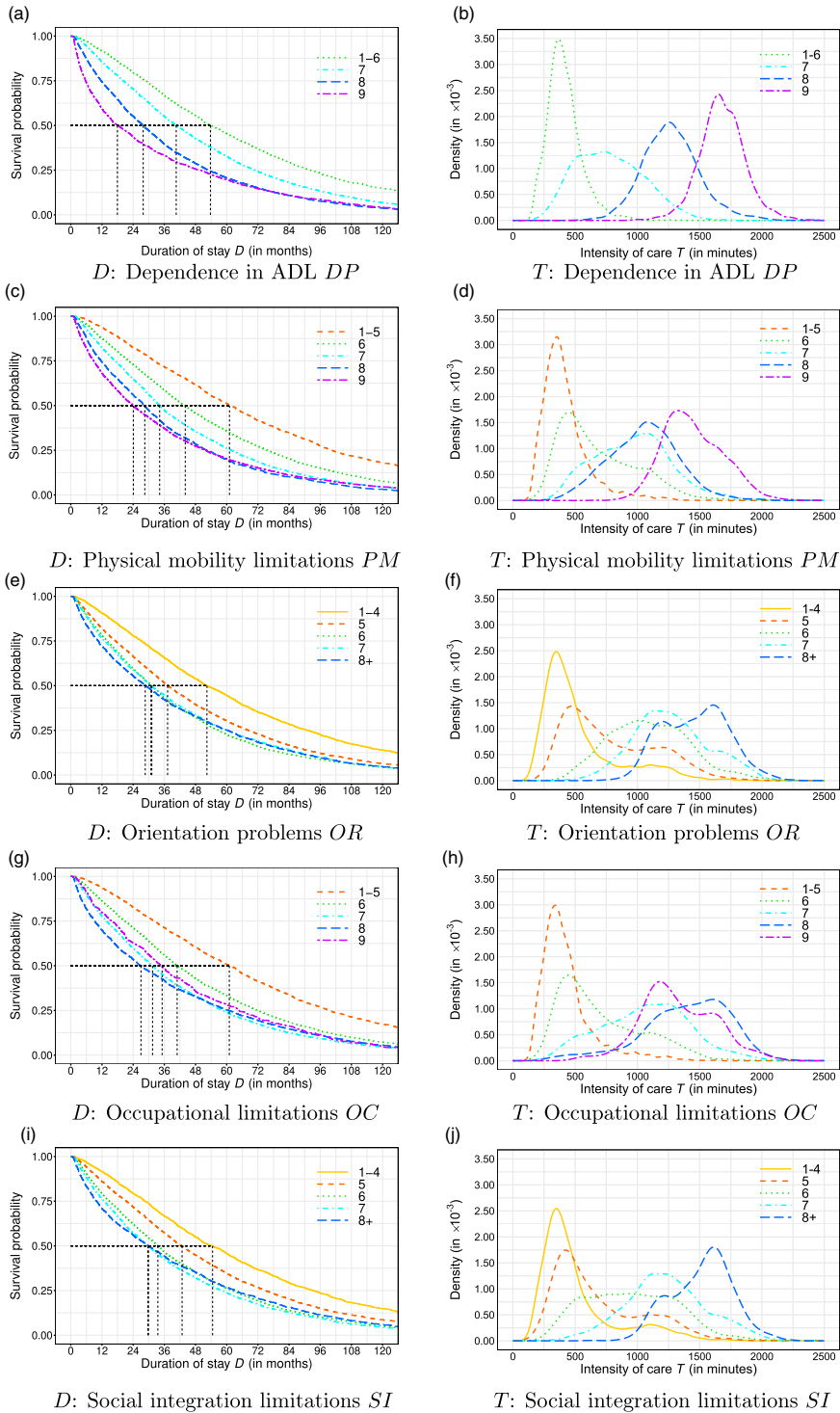


Figure 2. Kaplan–Meier estimation of the duration of stay D (in months) and density estimation of the intensity of care per week T (in minutes) across dependence factors.

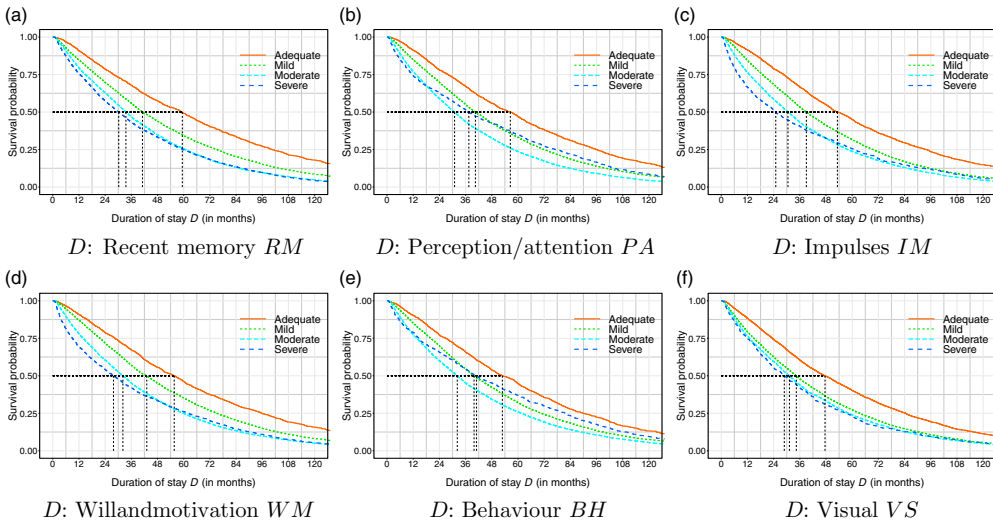


Figure 3. Kaplan–Meier estimation of the duration of stay D (in months) across psychological and sensory function impairments.

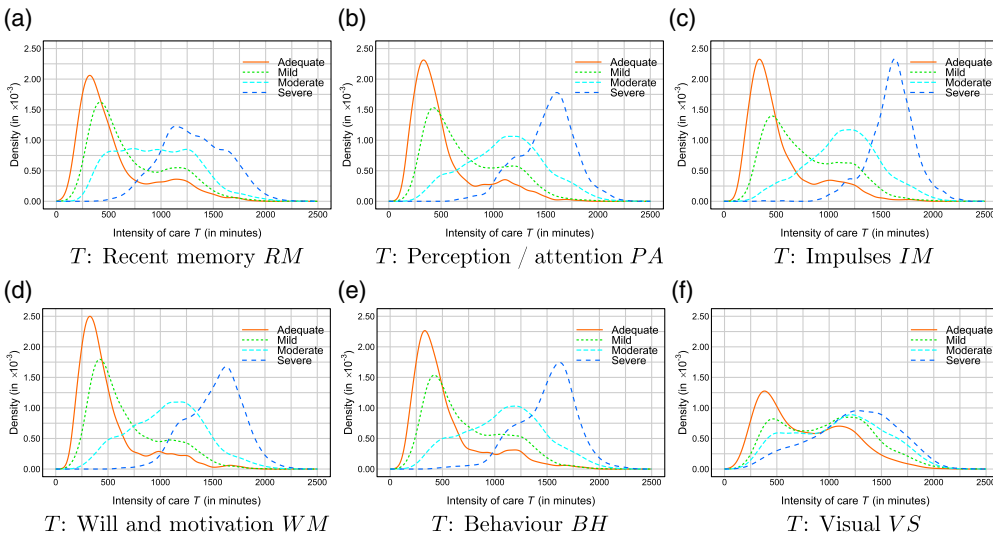


Figure 4. Density estimation of the intensity of care per week T (in minutes) across psychological and sensory function impairments.

We present the Kaplan–Meier and density estimates of the duration of stay D and the intensity of care T , respectively, in Figures 3 and 4. The “adequate” level of all psychological and sensory impairments is associated with the highest survival probabilities, while the other levels intersect each other. At the same time, for all impairments, except VS , we note a good separation of the distribution of T across the levels. In Figure 4(f), we observe that all distributions of T for visual impairments VS are intertwined. Although the moderate and severe levels of visual impairments are left-skewed, there is no clear separation of the distributions such as, for instance, in Figure 4(c) for impulse impairments IM .

4. Modelling Framework

Our objective is to fit an explanatory model to calculate the foreseeable need for LTC of a person who has just entered the institution. The overall care need or severity S corresponds to the total number of hours of care elderly individuals receive while sojourning in an institution. Thus, we quantify the overall care severity through the product $S = D \cdot T$, where D stands for the duration of stay and T for the intensity of care¹². This decomposition allows us to separately investigate the time spent in the institution, that is, the period where a bed is occupied, and the amount of care received, that is, the usage of caregivers’ resources. Indeed, as discussed in Section 2, we expect to find different influencing factors for the two dimensions. The overall mean care severity is then given by $\mathbb{E}[S] = \mathbb{E}[D \cdot T]$. In our modelling, we consider the simplifying assumption that T and D are independent. Such an approach is commonly accepted, for example, in the practice of actuarial calculations in non-life insurance¹³. We approximate the mean care severity \hat{S} by

$$\hat{S} = \mathbb{E}[S] = \mathbb{E}[D] \cdot \mathbb{E}[T] = \hat{D} \cdot \hat{T}, \tag{1}$$

where $\hat{D} = \mathbb{E}[D]$ and $\hat{T} = \mathbb{E}[T]$ are estimated with two separate models.

In Section 4.1, we present a model for the duration of stay D , expressing the number of months an elderly individual spends in an institution until death. In Section 4.2, we lay out a model for the intensity of care T , corresponding to the number of minutes of care received per week based on the individual’s evaluation at entry.

4.1 Duration of stay model

To model the duration of stay D in the institution, we first assess its distribution by fitting the observed durations with distributions commonly used in survival models. We find that the formal hazard proportionality test described in Grambsch & Therneau (1994) is not passed, although the assumption of proportional hazards visually holds when plotting the Cox–Snell residuals. This is commonplace when faced with a large amount of data, and thus, we proceed by considering a different type of model, the AFT model; see Collett (2015). The two model classes intersect when the underlying distribution is Weibull.

To select the underlying distribution for the AFT model, we use the Bayesian information criterion (BIC) scores of the fitted models with all variables (see Table 1) for the exponential, Weibull, Gaussian, logistic, log-normal, and log-logistic distributions of D_i . The lowest BIC score is obtained using the Weibull distribution. Thus, we apply the Weibull AFT model to our data.

A standard log-linear form of the AFT model with Weibull distribution is

$$\log D_i = x_i^T \lambda + \sigma \varepsilon_i, \quad i = 1, \dots, n, \tag{2}$$

where $\lambda = (\lambda_0, \lambda_1, \dots, \lambda_k)^T$ is a vector of unknown regression parameters, $x_i = (x_{i0}, x_{i1}, \dots, x_{ik})$ are the observations of known covariates, and n stands for the number of observations. Here, λ_0 corresponds to the intersection term, and hence we have $x_{i0} = 1$, for $i = 1, \dots, n$. In this form, ε_i does in fact have the standard Gumbel distribution¹⁴.

¹²In the original data and in the numerical results the duration of stay, D is expressed in months and the intensity of care T is expressed in minutes per week (see Table 1). In Section 5.5, we express the numerical results for the care severity S in hours. Therefore, we need to express D in weeks by multiplying with 4.345, the average number of weeks in a months, and T in hours per week by dividing by 60. Thus, taking units into account, the severity S is obtained from $S = D \cdot 4.345 \cdot T/60$.

¹³This assumption dates back to Lundberg (1903). The analysis of possible “correlations” between D and T is beyond the scope of the present study. Such analysis could build on a Copula method (see, e.g., Czado *et al.*, 2012) adapted to the case of right-censored variables.

¹⁴The standard Gumbel distribution with the location parameter 0 and the scale parameter 1 has the survival function $S_{\varepsilon_i}(x) = \exp[-\exp(x)]$. Therefore, the random variable $D_i = \exp(x_i^T \lambda + \sigma \varepsilon_i)$ has the survival function $S_{D_i}(t) = \exp[-\exp(-x_i^T \lambda / \sigma) t^{1/\sigma}]$, which is the Weibull survival function with scale parameter $\exp(x_i^T \lambda)$ and shape parameter $1/\sigma$.

In our case, D_i follows a Weibull distribution with scale parameter $\exp(x_i^T \lambda)$ and shape parameter $1/\sigma$. The mean duration is then expressed as

$$\mathbb{E}[D_i] = \exp(x_i^T \lambda) \Gamma(1 + \sigma), \quad (3)$$

where Γ stands for the gamma function, $\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx$.

AFT models account for right censoring, and the coefficients λ are obtained by maximising the likelihood function (see Klein & Moeschberger, 1997, Chapter 3.5),

$$\mathcal{L} = \prod_{i=1}^n [f_W(D_i)]^{\delta_i} [S_W(D_i)]^{1-\delta_i}, \quad (4)$$

where f_W is the Weibull density function, S_W is the Weibull survival function, and $\delta_i = 0$ if the data are right-censored and $\delta_i = 1$ otherwise. To fit the model, we applied the `survreg` function from the survival package in R; see Therneau (2021).

4.2 Care intensity model

The intensity of care T provided to a person each week is bounded and takes values in the interval $(0, 10,080)$. The upper bound is derived from the total number of minutes in 1 week. Usual practice performs a regression analysis on a transformation of the data so that the modified response variable, say \tilde{T} , takes values in the whole real line. Such an approach is in general disadvantageous since we can interpret the resulting analysis only in terms of the mean of \tilde{T} , while our interest is on the mean of T . However, a simple linear transformation $\tilde{T} = T/10,080$ does not obstruct our intentions. Moreover, we see from Figures 1, 2, and 4 that the data seem to be heteroskedastic. Indeed, not only do the means depend on the levels of the predictors, but the variance also changes.

An interpretative model in terms of the mean of T based on the beta distribution, hence, called the beta regression model, has been proposed by Ferrari & Cribari-Neto (2004). Later, Simas *et al.* (2010) provided an extension of the beta regression model that allows for non-linearity and variable dispersion. In the latter model, the standard beta density function

$$f(\tilde{T}; p, q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \tilde{T}^{p-1} (1-\tilde{T})^{q-1}, \quad 0 < \tilde{T} < 1, \quad (5)$$

is parameterised by the mean $\mu = p/(p+q) \in (0, 1)$ and the precision parameter $\phi = p+q > 0$, yielding

$$f(\tilde{T}; \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} \tilde{T}^{\mu\phi-1} (1-\tilde{T})^{(1-\mu)\phi-1}, \quad 0 < \tilde{T} < 1. \quad (6)$$

While the mean μ relates to the mean of \tilde{T} , the numerical value of the precision parameter ϕ does not have a simple interpretation. However, its estimate provides information on the variance of \tilde{T} . Indeed, by definition, the variance of a random variable with a beta distribution is $\text{var}(\tilde{T}) = V(\mu)/(1+\phi)$, where $V(\mu) = \mu/(1-\mu)$. Thus, for a given mean μ , the larger the value of ϕ is, the smaller the variance of \tilde{T} and T .

Let \tilde{T}_i , $i = 1, \dots, n$, follow the beta distribution with the above density function. We assume that the mean μ and the precision parameter ϕ characterising \tilde{T}_i satisfy the following functional relations:

$$g_1(\mu_i) = x_i^T \beta, \quad \text{and} \quad g_2(\phi_i) = x_i^T \theta, \quad (7)$$

where $\beta = (\beta_0, \beta_1, \dots, \beta_k)^T$ and $\theta = (\theta_0, \theta_1, \dots, \theta_k)^T$ are vectors of unknown regression parameters, and $2(k+1) < n$. Here, β_0 and θ_0 correspond to the intersection terms, and hence, we have $x_{i0} = 1, \forall i = 1, \dots, n$.

The resulting log-likelihood function has the following form:

$$\ell(\beta, \theta) = \sum_{i=1}^n \left[\log \Gamma(\phi_i) - \log \Gamma(\mu_i \phi_i) - \log \Gamma((1 - \mu_i)\phi_i) + (\mu_i \phi_i - 1) \log \tilde{T}_i + ((1 - \mu_i)\phi_i - 1) \log (1 - \tilde{T}_i) \right],$$

where $\mu_i = g_1^{-1}(x_i^T \beta)$ and $\phi_i = g_2^{-1}(x_i^T \theta)$ are defined in equation (7). There are various approaches to choose the link functions such that $g_1^{-1} : \mathbb{R} \rightarrow (0, 1)$ and $g_2^{-1} : \mathbb{R} \rightarrow \mathbb{R}_+$. It is the best practice to use interpretable link functions, as opposed to data-driven approaches, since only in the former case will the standard errors of the resulting parameter estimates be truthful. In our numerical implementation, we choose the standard transformations

$$g_1(\mu) = \log \left(\frac{\mu}{1 - \mu} \right), \quad \text{and} \quad g_2(\phi) = \log(\phi), \tag{8}$$

and, therefore, we have, for $i = 1, \dots, n$,

$$\mu_i = \frac{1}{1 + \exp(-x_i^T \beta)}, \quad \text{and} \quad \phi_i = \exp(x_i^T \theta). \tag{9}$$

To fit this model, we use the R package `betareg`. Details can be found in the original work by Cribari-Neto & Zeileis (2010), and the extended work by Grün *et al.* (2012).

5. Results and Discussion

In this section, we apply the econometric models introduced in Section 4. In Section 5.1, we discuss the specification of the models. We introduce the variable transformations and discuss the selection of the psychological and sensory function impairments, the inclusion of interaction terms between age at entry and gender, and how the variable importance is measured. We also show how the coefficients of the models are interpreted. We present the results of the duration of stay and intensity of care models, respectively, in Table 4 and analyse and interpret them in Sections 5.2 and 5.3. Next, we assess the goodness-of-fit of both models in Section 5.4, and finally, we highlight model estimates of the duration of stay, the intensity of care, and the overall care severity for selected profiles of the institutionalised elderly individuals in Section 5.5.

5.1 Specification of the models and results

In Table 4, we report the regression results for both the duration of stay (equation (2)) and intensity of care (equation (6)) models. In each model and for each variable, we report the estimates for the regression coefficients with the standard deviation and the significance level. Recall that the original data contain 16 variables related to psychological and sensory function impairments (see Section 3.1). To reduce the complexity and improve the scores of the models, we reduce the number of these factor variables using a variable selection procedure based on the BIC. Overall, we retain six variables associated with psychological and sensory function impairments. Three variables (*PA*, *BH*, *VS*) appear in the duration of stay model, and four (*RM*, *IM*, *WM*, *VS*) appear in the intensity of care model. All the other covariates summarised in Table 1 are included in both models.

Most of the covariates in our data are categorical variables (see Table 1). From the descriptive statistics in Table 3, we observe that women (72.5%) are more prevalent than men in institutional care, and, therefore, we choose “female” as the baseline for the gender *GE* variable. For the main diagnosis *D1*, we choose the group of mental pathologies as the baseline since it has the highest

Table 4. Model results for the duration of stay D and the intensity of care T models.

	Duration of stay D					Intensity of care T							
	$\hat{\lambda}$	$\sigma_{\hat{\lambda}}$	Sig.	Effect	Imp. (rank)	$\hat{\beta}$	$\sigma_{\hat{\beta}}$	Sig.	Effect	$\hat{\theta}$	$\sigma_{\hat{\theta}}$	Sig.	Imp. (rank)
Intercept	4.9862	(0.0498)	***	135.0		-3.4955	(0.0166)	***	296.8	5.6313	(0.0872)	***	
$\ln \sigma$	-0.2702	(0.0058)	***										
AG	-0.0354	(0.0009)	***		1,527.5 (1)	-0.0014	(0.0002)	***		0.0114	(0.0015)	***	75.2 (7)
Gender GE (baseline: Female)					795.3 (2)								404.5 (3)
Male	-0.3884	(0.0131)	***	-43.4		0.0684	(0.0033)	***	20.3	-0.0793	(0.0220)	***	
Main diagnosis $D1$ (baseline: Mental)					304.8 (4)								5.8 (12)
Cerebrovascular	-0.0111	(0.0266)		-1.5		0.0336	(0.0061)	***	9.8	-0.0811	(0.0441)	.	
Nervous	-0.0821	(0.0171)	***	-10.6		0.0332	(0.0041)	***	9.7	-0.0244	(0.0285)		
Osteoarticular	0.0276	(0.0241)		3.8		0.0229	(0.0067)	***	6.7	-0.1177	(0.0411)	**	
Heart	-0.1391	(0.0226)	***	-17.5		0.0057	(0.0062)		1.7	-0.0385	(0.0385)		
Tumour	-0.8512	(0.0437)	***	-77.4		0.0328	(0.0112)	**	9.6	-0.0171	(0.0766)		
Other	-0.1266	(0.0162)	***	-16.1		0.0358	(0.0041)	***	10.5	-0.0864	(0.0270)	**	
Secondary diagnosis $D2$ (baseline: None)					41.4 (8)								-82.9 (14)
Mental	0.0099	(0.0445)		1.3		-0.0030	(0.0135)		0.9	0.1425	(0.0774)	.	
Cerebrovascular	-0.0327	(0.0530)		-4.3		0.0048	(0.0152)		1.4	0.0952	(0.0917)		
Nervous	0.0006	(0.0482)		0.1		0.0176	(0.0141)		5.1	0.1789	(0.0836)	*	
Osteoarticular	0.0761	(0.0464)		10.7		-0.0091	(0.0140)		-2.6	0.1996	(0.0807)	*	
Heart	-0.0747	(0.0448)	.	-9.7		-0.0042	(0.0137)		-1.2	0.1869	(0.0782)	*	
Tumour	-0.3899	(0.0618)	***	-43.6		0.0206	(0.0175)		6.0	0.1868	(0.1087)	.	
Other	-0.0283	(0.0440)		-3.8		0.0142	(0.0135)		4.1	0.1122	(0.0767)		
ND	-0.0522	(0.0028)	***		340.1 (3)	0.0130	(0.0007)	***		-0.0062	(0.0046)		321.1 (5)

Table 4. Continued.

	Duration of stay <i>D</i>				Intensity of care <i>T</i>								
	$\hat{\lambda}$	$\sigma_{\hat{\lambda}}$	Sig.	Effect	Imp. (rank)	$\hat{\beta}$	$\sigma_{\hat{\beta}}$	Sig.	Effect	$\hat{\theta}$	$\sigma_{\hat{\theta}}$	Sig.	Imp. (rank)
Dependence in ADL <i>DP</i> (baseline: 1–6)					182.6 (5)								6,489.1 (1)
7	−0.0750	(0.0211)	***	−9.7		0.4308	(0.0066)	***	152.7	−1.0442	(0.0339)	***	
8	−0.2532	(0.0275)	***	−30.2		0.6946	(0.0078)	***	280.6	−0.7241	(0.0451)	***	
9	−0.5687	(0.0407)	***	−58.5		0.9084	(0.0097)	***	408.6	−0.7383	(0.0703)	***	
Physical mobility limitations <i>PM</i> (baseline: 1–5)					77.1 (6)								4,010.2 (2)
6	−0.1407	(0.0232)	***	−17.7		0.1756	(0.0082)	***	55.0	−0.1266	(0.0387)	**	
7	−0.2299	(0.0263)	***	−27.7		0.3212	(0.0089)	***	107.9	−0.0973	(0.0436)	*	
8	−0.2930	(0.0308)	***	−34.3		0.3838	(0.0098)	***	132.9	−0.0077	(0.0509)		
9	−0.2842	(0.0301)	***	−33.4		0.5207	(0.0094)	***	192.9	0.4890	(0.0510)	***	
Orientation problems <i>OR</i> (baseline: 1–4)					−17.6 (11)								363.4 (4)
5	−0.0227	(0.0247)		−3.0		0.0549	(0.0086)	***	16.2	0.0448	(0.0432)		
6	0.0125	(0.0328)		1.7		0.1570	(0.0108)	***	48.7	0.2902	(0.0587)	***	
7	0.0460	(0.0409)		6.4		0.1699	(0.0125)	***	53.1	0.2507	(0.0753)	***	
8+	−0.0980	(0.0486)	*	−12.6		0.2090	(0.0139)	***	66.5	0.2728	(0.0891)	**	
Occupational limitations <i>OC</i> (baseline: 1–5)					−26.2 (12)								279.8 (6)
6	−0.0736	(0.0248)	**	−9.6		0.0917	(0.0087)	***	27.6	−0.0689	(0.0419)		
7	−0.0993	(0.0278)	***	−12.8		0.1235	(0.0095)	***	37.7	−0.0924	(0.0470)	*	
8	−0.0767	(0.0341)	*	−10.0		0.0999	(0.0104)	***	30.2	−0.1684	(0.0586)	**	
9	−0.0790	(0.0436)	.	−10.2		0.2076	(0.0126)	***	66.0	−0.3851	(0.0742)	***	
Social integration limitations <i>SI</i> (baseline: 1–4)					−33.9 (13)								−2.6 (13)
5	0.0212	(0.0284)		2.9		−0.0332	(0.0096)	***	−9.4	−0.0170	(0.0470)		
6	−0.0141	(0.0335)		−1.9		−0.0381	(0.0109)	***	−10.8	−0.0405	(0.0553)		
7	−0.0427	(0.0401)		−5.6		−0.0196	(0.0121)		−5.6	−0.2120	(0.0661)	**	
8+	−0.0132	(0.0556)		−1.8		−0.0446	(0.0144)	**	−12.6	0.0801	(0.0931)		

Table 4. Continued.

	Duration of stay <i>D</i>					Intensity of care <i>T</i>								
	$\hat{\lambda}$	$\sigma_{\hat{\lambda}}$	Sig.	Effect	Imp. (rank)	$\hat{\beta}$	$\sigma_{\hat{\beta}}$	Sig.	Effect	$\hat{\theta}$	$\sigma_{\hat{\theta}}$	Sig.	Imp. (rank)	
Recent memory <i>RM</i> (baseline: Adequate)														25.5 (10)
Mild						0.0100	(0.0098)		2.6	0.1643	(0.0459)	***		
Moderate						0.0300	(0.0110)	**	-9.8	0.2677	(0.0539)	***		
Severe						0.0300	(0.0122)	**	-9.2	0.3737	(0.0675)	***		
Perception and attention <i>PA</i> (baseline: Adequate)					16 (9)									
Mild	-0.0175	(0.0268)		-2.3										
Moderate	-0.0783	(0.0322)	*	-10.2										
Severe	0.1257	(0.0473)	**	18.1										
Impairment in impulses <i>IM</i> (baseline: Adequate)														26.8 (9)
Mild						0.0469	(0.0088)	***	13.8	0.0601	(0.0432)			
Moderate						0.0671	(0.0098)	***	19.9	0.0825	(0.0515)			
Severe						0.0415	(0.0124)	***	12.2	0.2106	(0.0856)	*		
Will and motivation <i>WM</i> (baseline: Adequate)														15.2 (11)
Mild						0.0195	(0.0105)	.	5.7	0.0037	(0.0504)			
Moderate						0.0640	(0.0117)	***	19.0	-0.0138	(0.0582)			
Severe						0.0629	(0.0131)	***	18.7	-0.0779	(0.0776)			
Behavioural impairment <i>BH</i> (baseline: Adequate)					64.6 (7)									
Mild	-0.0455	(0.0291)		-6.0										
Moderate	-0.0369	(0.0326)		-4.9										
Severe	0.2986	(0.0483)	***	47.0										
Visual impairment <i>VS</i> (baseline: Adequate)					11.2 (10)									52.3 (8)
Mild	-0.0706	(0.0147)	***	-9.2		0.0313	(0.0042)	***	9.2	0.0932	(0.0250)	***		
Moderate	-0.0212	(0.0181)		-2.8		0.0291	(0.0049)	***	8.5	0.1275	(0.0309)	***		
Severe	0.0834	(0.0373)	*	11.7		0.0625	(0.0090)	***	18.5	-0.0502	(0.0640)			

Note: Significance levels in column “Sig.” are reported as follows: *p*-value < 0.1., < 0.05 *, < 0.01 **, < 0.001 ***. The column “Effect” reports the effect related to a category deviating from the baseline (all other parameters remaining at the baseline). The column “Imp. (rank)” reports the effect on the BIC and the corresponding rank of the variable.

prevalence (34.3%). The second diagnosis, $D2$, involves comorbidity and interactions with $D1$. To avoid ambiguous interpretations in the baseline, we assume that elderly individuals have no secondary diagnosis and choose “none” as the baseline for $D2$. Furthermore, for the variables that describe the levels of dependence, we use the group of lowest levels as a baseline. Thereby, we use the groups laid out in the descriptive statistics (see Table 3), that is, levels 1–6 for dependence in ADL DP , 1–5 for physical mobility limitations PM , 1–4 for orientation problems OR , 1–5 for occupational limitations OC , and 1–4 for social integration limitations SI . Finally, we use the “adequate” level (healthy) as the reference category in the impairments of psychological and sensory functions (recent memory impairments RM , perception and attention impairments PA , impulse impairments IM , will and motivation impairments WM , behavioural impairments BH , and visual impairments VS).

The set of covariates further includes two numerical variables: the age at entry AG and the number of additional diagnoses ND . While AG starts at 65 years, ND takes integer values from 0 to 7, where zero indicates that the person has no additional diagnoses, that is, only one or two diagnoses given by $D1$ and $D2$. As we have seen in Table 3, LTC prevalence rates and entrance into care institutions expand after the age of 80 years (see also, e.g., Colombo *et al.*, 2011). To account for this, we consider $AG = 80$ years as a reference point and subtract 80 from the AG variable when fitting the models, that is, we transform the predictor into $(AG - 80)$. Hence, an 80-year-old woman at entry with a mental main diagnosis, without secondary or additional diagnoses, the lowest dependence levels and no other impairments characterises the baseline health profile and corresponds to the intercept term, or baseline, in both models.

In Table 4, we also provide intuition for the interpretation of the obtained estimates in the column labelled “Effect”. For the categorical variables, the “Effect” measures the increase or decrease of D and T in absolute values when the variable switches from its baseline to the corresponding level, and all other parameters remain the same at the baseline. For numerical variables, we do not report the information. Indeed, since our models are non-linear, the corresponding effects must be evaluated through equations (3) and (9), respectively. If several parameters are changed at the same time, the effects also must be calculated using the original equations (see also the model estimations illustrated in Section 5.5).

Furthermore, to rank the importance of the variables, we remove each variable from the model and compute the resulting BIC score. Then, we subtract the BIC score of the full model and report the difference in the column labelled “Imp. (rank)” in Table 4. If the difference is positive, the reduced model would suffer from information loss, and the variable is important. In contrast, if the difference is negative, the reduced model would benefit from removing the variable. Since we report the results for all variables, some of them have negative importance values (e.g. OR , OC , SI in the duration of stay model), which means that these variables may be omitted. Finally, we order each difference in BIC scores and report the rank in parentheses.

5.2 Duration of stay

In the first part of Table 4 in the column labelled “Duration of stay D ”, we present the results of the AFT model (equation (2)) applied to our data set. The first columns present the fitted coefficients $\hat{\lambda}$ of the model and the corresponding standard deviations $\sigma_{\hat{\lambda}}$. The mean duration of stay for the baseline that can be read from the intercept row in the “Effect” column is 135 months.

Demographic variables. We observe a negative effect of the age at entry AG on the duration of stay with $\hat{\lambda}_{AG} = -0.0354$. This follows the intuition that the later a person enters institutional care, the less time he or she spends there. Our result substantiates the findings on increased mortality rates at higher ages (Mathers, 1996; Deeg *et al.*, 2002; Fong *et al.*, 2017) and, thus, a reduced duration of stay (Colombo *et al.*, 2011; Fuino & Wagner, 2020). At the same time, the coefficient $\hat{\lambda}_{GE} = -0.3884$ yields that males, on average and for the baseline profile, spend 43.4 fewer

months than women in the institution. Both variables share first and second place in the importance ranking, with the age at entry AG being almost twice as important as the gender GE in terms of BIC.

Medical diagnoses. As one can observe from the descriptive statistics in Table 3 and the survival curve pictured in Figure 1(a), pathologies are one of the key factors that shape the duration of stay D . In fact, the number of additional diagnoses ND is the third most important variable, followed by the main diagnosis $D1$ (rank four). The secondary diagnosis $D2$ takes eighth place. All main diagnoses significantly reduce the duration of stay compared to the group of mental diagnoses, except for cerebrovascular and osteoarticular pathologies, which appear to be not significant. A tumour in the main diagnosis significantly reduces the baseline duration by 77.4 months on average. In absolute value, the coefficient $\hat{\lambda}_{D1}^{\text{Tumour}} = -0.8512$ is the highest among the $D1$ levels. This result follows from the lower expected lifetime due to the often faster progression of the tumours, which more rapidly leads to death compared to other diseases (Guccione *et al.*, 1994). Next comes the group of heart-related diseases, which reduce the average duration of stay by 17.5 months.

The secondary diagnosis $D2$ does not significantly affect the duration of stay except for tumour diagnoses and, to a lesser extent, heart diseases (compared to the baseline with no secondary diagnosis). We observe that the duration of stay, on average, is reduced by an additional 43.6 months ($\hat{\lambda}_{D2}^{\text{Tumour}} = -0.3899$) regardless of the type of main diagnosis $D1$. Simultaneously, heart disease in $D2$ decreased the mean duration by 9.7 months. In general, the more pathologies a person has, the less time he or she spends in dependence ($\hat{\lambda}_{ND} = -0.0522$). Multi-morbidity comes with a higher chance of developing severe conditions along a pathology, which eventually increases the mortality rate, especially at higher ages (see, e.g., Menotti *et al.*, 2001; Deeg *et al.*, 2002; Byles *et al.*, 2005).

Level of dependence. Our regression results show significant differences between the different levels of DP , PM , and OC and thus support the clear distinction that we observed between the curves of the different levels in the graphs of Figure 2 (see Figures 2(a), (c), (g), respectively). All levels of dependence in ADL DP , the fifth most important variable, are highly significant and reduce the mean duration of stay by up to 58.5 months. Furthermore, physical mobility PM takes the sixth place in the importance ranking, with all coefficients being highly significant. Increased levels of PM monotonously shorten the duration of stay. Both DP and PM increase the mortality rates, which result in a shorter duration (Rickayzen & Walsh, 2002). Despite its rank at 12 (second to last in the ranking), the occupational limitations variable OC has all its coefficients significantly different from the baseline, although the p -value thresholds are different. For example, Level 7 (confined occupation in terms of time and type) reduces the mean duration of stay the most (12.8 months) with a three-star significance, while the most vulnerable group of unoccupiable elderly individuals (Level 9) yields the least significant coefficient (reduction of 10.2 months).

The overlapping curves for the levels in the variables OR and SI translate into regression coefficients that are not significant. The variable occupational problems OR has only one coefficient $\hat{\lambda}_{OR}^{8+} = -0.0980$ (disorientation or unconscious) that is one-star significant, which reduces the mean duration of stay by 12.6 months compared to the baseline.

Impairments of psychological and sensory functions. The BIC variable selection procedure leaves us with three factors related to the impairments that are included in the duration of stay model: perception and attention PA , behavioural BH , and visual impairments VS . Behaviour BH ranks seventh in the importance ranking, right after the dependence in ADL DP and physical mobility PM variables. Since behavioural impairments manifest themselves as accentuated character traits, elderly individuals with severe levels of BH are cared for more intensely (see Figure 4(e)), which leads to a higher mean duration of stay. The coefficient of the severe level ($\hat{\lambda}_{BH}^{\text{Severe}} = 0.2986$)

is the only significant coefficient. It increases the mean duration of stay by 47 months compared to the baseline.

The perception and attention variable *PA* takes the ninth rank in the importance list and is the second most important variable among psychological and sensory function impairments. We observe two coefficients significantly different from the baseline: $\hat{\lambda}_{PA}^{\text{Moderate}} = -0.0783$ (one star) and $\hat{\lambda}_{PA}^{\text{Severe}} = 0.1257$ (two stars). Surprisingly, they are of the opposite signs. Compared to the adequate level, moderate impairments decrease the mean duration of stay by 10.2 months, while severe impairments increase it by 18.1 months. From Figure 4(b), we deduce that the latter level corresponds to a much higher intensity of care, which through more attentive care results in slower health deterioration and, thus, a higher expected duration (Tombaugh & McIntyre, 1992).

Finally, the mild level in the visual impairments variable *VS* reduces the duration of stay by 9.2 months, which may be linked to the lower level of help and overconfidence of the person. Indeed, from Figure 4(f), we see that the adequate and mild (and, to some extent, the moderate) levels of *VS* receive similar intensities of care, while blind people (severe level) receive extra attention. We find that a severe level of sight impairment increases the mean duration of stay by 11.7 months.

5.3 Intensity of care

The results of the beta regression model (equation (6)) for the intensity of care are presented in the second part of Table 4 under the heading “Intensity of care *T*”. The first two columns provide the β and θ estimates corresponding to the mean and precision parameters of the beta distribution of \tilde{T} . The mean intensity of care *T* provided to the baseline profile is 296.8 minutes/week (see the column labelled “Effect”), corresponding to approximately 5 hours/week. The other values reported in the “Effect” column relate to increases and decreases in the intensity of care *T* in the different factors.

Demographic variables. We observe a significant negative, although small, effect of the age at entry *AG* on the mean intensity of care. The estimate $\hat{\beta}_{AG} = -0.0014$ indicates that older individuals receive, on average, slightly less help for their daily needs. At the same time, the positive three-star significant coefficient $\hat{\theta}_{AG} = 0.0114$ suggests that at higher ages, the variance of the intensity of care is lower compared to the younger elderly individuals, which may result from heterogeneous health conditions and an increase in disability (Fries, 1980; Olshansky *et al.*, 1991). The variable *AG* ranks seventh in importance. Furthermore, the coefficient $\hat{\beta}_{GE} = 0.0684$ indicates that males receive, on average, more help than females. For the baseline profile, the difference is 20.3 minutes/week. The negative coefficient $\hat{\theta}_{GE} = -0.0793$ indicates that the amount of help received by males has a higher variance, yielding more precise estimations for women than for men. Gender *GE* is one of the key determinants of the intensity of care, since it takes third place in importance.

Medical diagnoses. As we have observed in Figure 1(b), the main diagnosis *D1* is not a strong determinant factor for explaining the intensity of care. Indeed, *D1* takes only the 12th place in the importance ranking. Nevertheless, all diagnoses’ coefficients are significantly (three stars) different from the mental disease baseline, except for tumour (two-star significant) and heart diseases (not significant). Significant differences increase the mean intensity of care by approximately 10 minutes/week. The secondary diagnosis *D2* is the least important variable in our model. Since none of the $\hat{\beta}_{D2}$ coefficients is significant, there is no benefit in knowing the secondary diagnosis for explaining the intensity of care. However, the positive coefficients $\hat{\theta}_{D2}$ indicate that the variance of the intensity of care evaluation can be reduced by including *D2*. Finally, the number of additional diagnoses *ND* ranks fifth in importance. The coefficient $\hat{\beta}_{ND} = 0.0130$ is highly significant and has a positive effect on the mean intensity, that is, the more comorbidities a person has, the more care

he or she requires. This finding supplements the results of Barnett *et al.* (2012) when applied to the intensity of care provided to dependent elderly individuals in an institution. Indeed, the number of pathologies provides more relevant information than the underlying diseases themselves.

Level of dependence. Dependence in ADL DP is the most important variable explaining the intensity of care. Moreover, it comes with the highest values of $\hat{\beta}$ coefficients (all three-star significant) among all variables. Compared to the baseline (levels 1–6), levels 7, 8, and 9 increase the mean intensity of care by 152.7 minutes (2.5 hours), 280.6 minutes (4.7 hours), and 408.6 minutes (6.8 hours) per week, respectively. Thus, expectedly, ADL limitations are a key driver of care intensity. Simultaneously, the coefficients $\hat{\theta}_{DP}$ are negative, which indicates a higher variance of the intensity of care for different levels of DP compared to the baseline.

The second most important variable, PM , is associated with physical mobility limitations. All coefficients $\hat{\beta}_{PM}$ are three-star significant and positive, indicating an increasing mean intensity of care compared to the baseline. Moreover, the amount of help increases for the higher levels of PM by up to 192.9 minutes (approximately 3.2 hours) per week for the fully immobile elderly individuals (Level 9). At the same time, the precision coefficient $\hat{\theta}_{PM}^9 = 0.4890$ is three-star significant and positive, indicating a lower variance compared to the baseline. In contrast, Levels 6 and 7 lead to a higher variance, which may be explained by the fact that people who ordinarily get around independently inside the institution and their bedroom, respectively, are more exposed to a sudden and unexpected need for help when moving around. For example, when trying to get to another floor, being lost, or forgetting the place of their bedroom (Maresova *et al.*, 2019), it might be the case that these two levels should receive more help than they usually do to reduce the spread and make help requests more predictable.

The variables linked to orientation problems OR and occupational limitations OC take fourth and sixth place in the importance ranking. Like the dependence in ADL DP and physical mobility limitations PM , all $\hat{\beta}$ coefficients are three-star significant and positive, yielding an increased mean intensity of care for the higher levels of dependence. For both OR and OC , the maximum increase is approximately 1 hour/week (66.5 and 66.0 minutes, respectively). Finally, social integration SI is a much less relevant driver of the intensity of care. Indeed, limitations with physical activities drive the intensity of care much more than the needed help with social contacts. Surprisingly, a reduced level of communication abilities leads to a slightly lower amount of help, which may be explained by the solitary lifestyle of these persons (Simard & Volicer, 2020). Although the coefficients $\hat{\beta}_{SI}$ are small, the highest reduction in mean intensity of care (12.6 minutes/week) is found in the most alienated and abandoned individuals (levels 8+).

Impairments of psychological and sensory functions. Following the BIC variable selection procedure four variables are included in our model: recent memory RM , impairment in impulses IM , will and motivation WM , and visual impairment VS . Recent memory RM is the only factor that sees a slight decrease (approximately 10 minutes/week) in the intensity of care at moderate and severe levels (two stars). In the other variables (IM , WM , VS), all levels are significant and come with an increased intensity of care at the more severe levels. However, the increase remains relatively small and remains below 20 minutes/week. Indeed, in the impairment in impulses IM , the largest increase in the mean care intensity is 19.9 minutes/week at the moderate level. For will and motivation WM impairments, we observe an additional 19 minutes/week at the same level, while severe visual impairments require 18.5 minutes/week more.

5.4 Goodness-of-fit and alternative model specifications

Goodness-of-fit of the AFT model. In the previous sections, we applied the AFT and beta regression models to the data, and we reported the results for the duration of stay D and the intensity of care T . To assess the quality of the fit of the AFT model (equation (2)) with Weibull distribution,

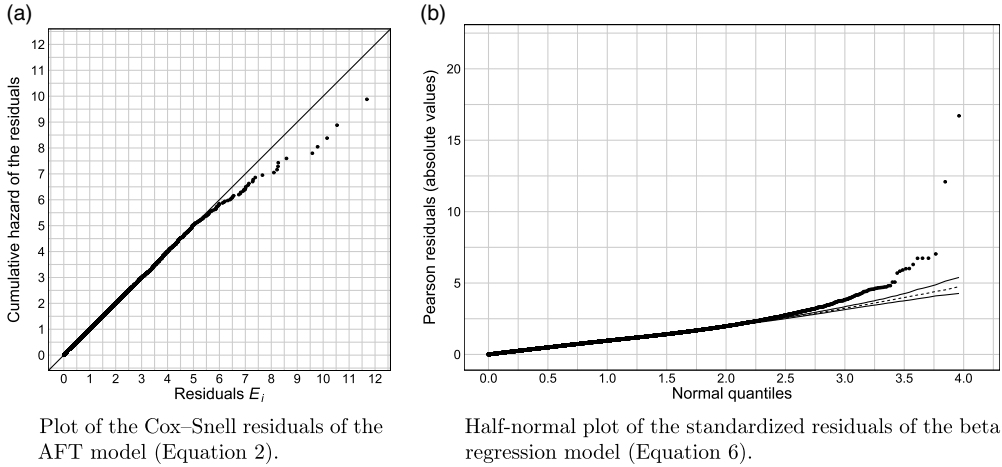


Figure 5. Illustration of the goodness of fit.

we plot the Cox–Snell residuals (see Klein & Moeschberger, 1997, Chapter 11.2). Therefore, we calculate the cumulative distribution function of the Weibull distribution at the observed durations of stay,

$$U_i \equiv F_W(D_i; \lambda, \sigma) = 1 - \exp \left(- \left(\frac{D_i}{\exp(x_i^T \hat{\lambda})} \right)^{1/\hat{\sigma}} \right), \tag{10}$$

using the coefficient estimates $\hat{\lambda}$ and $\hat{\sigma}$ reported in Table 4. Under the correct model, the values U_i should be uniformly distributed on the unit interval. However, recall that 17.6% of the observations are censored. To overcome this, we apply the transform $E_i = -\log(1 - U_i)$, and now the Cox–Snell residuals E_i are expected to constitute a right-censored sample that follows the exponential distribution with unit rate. We compute the Neslon–Aalen estimator of the cumulative hazard rate of E_i without any predictors, which should be approximately equal to the cumulative hazard rate of the unit exponential, that is, the identity function under the null hypothesis.

We report both quantities in Figure 5(a) and we confirm the required behaviour. Indeed, we see that most of the residuals are concentrated on the diagonal, while only the right tail is slightly below. Note that these are few observations from the whole data set, which contains 21,758 records. However, standard formal statistical tests reject the null hypothesis, which is not surprising for a data set of this size. Presently, we are satisfied with the visual agreement of the curve depicted in Figure 5(a). Different model specifications could correct the slight deviation, but at the potential cost of loss of interpretability.

Goodness-of-fit of the beta regression model. To assess the goodness-of-fit of the beta regression model (equation (6)), we proceed by calculating the pseudo R^2 as a global measure of explained variance. It is defined as the square of the sample correlation coefficient between the observed intensity of care and the fitted intensity of care. The resulting pseudo R^2 of our model is 79.31%, which in practice is considered relatively high.

We further follow the approach proposed by Ferrari & Cribari-Neto (2004) to extend our analysis. We define the standardised residuals, also called Pearson residuals, as

$$r_i = \left(\tilde{T}_i - \hat{\mu}_i \right) / \sqrt{\widehat{\text{var}}(\tilde{T}_i)}, \tag{11}$$

where $\widehat{\text{var}}(\tilde{T}_i) = [\hat{\mu}_i(1 - \hat{\mu}_i)] / (1 + \hat{\phi}_i)$, and $\hat{\mu}$ and $\hat{\phi}$ are the estimates of μ and ϕ , respectively. Since the exact distribution of these residuals is not known, half-normal plots with simulated

envelopes are a useful diagnostic tool (see Atkinson, 1985; Kutner *et al.*, 2005). The idea is to enhance the standard half-normal plot by adding a simulated envelope that can be used to decide whether the observed residuals are consistent with the fitted model. The construction of a half-normal envelope for the beta regression model is fully described in Ferrari & Cribari-Neto (2004). Under the null hypothesis, the observations should lie inside the envelope a prespecified percentage of times. In Figure 5(b) we draw the envelope (area between both plain curves) and the Pearson residuals using 500 simulations and a 95% confidence interval. We observe three outliers at the top right corner and some observations that lie outside the envelope (see the records starting from 2.2 on the horizontal axis). Overall, 76.3% of the data lie inside the confidence interval. Note that given the large number of observations, the boundaries of the simulated envelope are rather narrow. Thus, we conclude that the intensity of care model achieves satisfactory concordance with the data.

Alternative models. To further investigate the results, we also consider models where the age at entry AG and the number of additional diagnoses ND are coerced into categorical variables, separately and simultaneously, with the levels used in the descriptive statistics reported in Table 3. It appears that the resulting BIC values increase importantly, which is due, to a large extent, to the increase in the number of degrees of freedom. Thus, we confirm that AG and ND should be kept as numerical variables. Furthermore, in the extant literature, the interaction term between AG and gender GE is proven to significantly affect LTC usage (see, e.g. von Strauss *et al.*, 2003; Crimmins *et al.*, 2010; Fuino *et al.*, 2020). Indeed, at increasing ages, the duration of stay in an institution and the intensity of care are different for males and females. In our models, however, adding this interaction term yields higher BIC values, and moreover, if added, the coefficients of the interaction terms in both models are not significant for both numerical and categorical definitions of AG .

5.5 Estimation of the overall care severity

Using selected profiles of the institutionalised elderly, we estimate the duration of stay, the intensity of care, and the overall care severity. Thereby, the overall mean care severity is estimated by $\hat{S} = \hat{D} \cdot \hat{T}$; see equation (1), where where $\hat{D} = \mathbb{E}[D]$ and $\hat{T} = \mathbb{E}[T]$ are estimated through equations (3) and (9), respectively.

Effect of the age at entry and the gender. Our first analysis looks at the care requirements along the age at entry AG and the gender GE . To present the results, we introduce a profile that we call “modal”, where for each characteristic other than AG and GE , we take the most prevalent value, that is, the mode value (see Table 3). Hence, the modal profile characterises an individual with a mental main diagnosis, a secondary diagnosis grouped under “other” and having seven additional diagnoses. The dependence of the individual is such that he or she needs help with ADL at unpredictable times ($DP = 7$) and has mobility limited to the institution ($PM = 6$), moderate disturbances of orientation ($OR = 5$), occupational limitations in time and type ($OC = 7$), and social limitations with only primary contacts ($SI = 6$). Furthermore, the individual has moderate impairments in recent memory RM , perception and attention PA , impulses IM , will and motivation WM , behaviour BH , and, finally, a mild visual impairment VS .

In the graphs of Figure 6, we present the estimated duration of stay \hat{D} , intensity of care \hat{T} , and overall care severity \hat{S} across the age at entry and both genders for the modal profile. We note that throughout all ages, women stay longer in the institution, and they require less help per week. Consequently, the overall care severity \hat{S} of women is larger. In line with their higher life expectancy, for any age at entry, women stay longer in an institution than men (Mathers *et al.*, 2001; Fong *et al.*, 2017; Schünemann *et al.*, 2017). The graph in Figure 6(a) shows that the mean duration of stay when entering an institution at 65 years is 90.5 and 61.3 months for women and

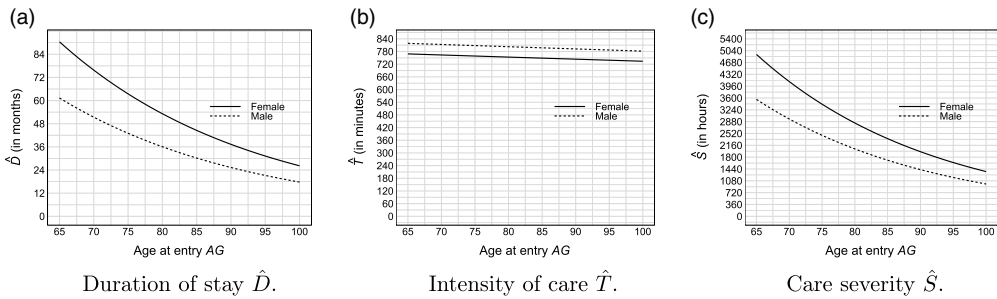


Figure 6. Estimated duration of stay, intensity of care and overall care severity for the modal profile.

men, respectively. At this age, the difference (29.2 months) between the genders is the largest. It decreases as the age at entry AG increases, and the duration of stay shortens. For example, men (women) entering at the age of 80 years are estimated to stay 36 (53) months. These values are comparable with the range of observations reported in Fuino & Wagner (2020), Tables 11 and 12. Figure 6(b) indicates that there is only a small difference in the intensity of care between the genders, the intensity yielding 750 (800) minutes/week for men (women). Indeed, we have found that the age at entry is not one of the key determinants of the intensity of care (see Table 4). Furthermore, we observe that, for any gender, the difference between the age of 65 and 100 years is only 36 minutes.

With the estimates for the overall care severity \hat{S} in Figure 6(c), we can observe the total burden of institutional care that comes with men and women entering an institution at different ages. Depending on the age at entry and gender, the care severity can change by a factor of five. Indeed, the estimated values for the modal profile range from 4,925 hours of care for women institutionalised at the age of 65 years to 984 hours of care for men entering the institution at the age of 100 years. We observe a decreasing trend of \hat{S} across the ages at entry that is in line with the trend observed for the duration of stay \hat{D} . While the shapes for both genders resemble each other, the care severity for women is approximately 40% higher than that for men (2,840 versus 2,050 hours at $AG = 80$) but decreases faster. The difference of approximately 1,350 hours at the age of 65 years is only approximately 380 hours at the age of 100 years.

Effect of the main and secondary diagnoses. Considering the modal profile for an 80-year-old woman at entry, we present in Figure 7 the estimates for different combinations of the main $D1$ and secondary $D2$ diagnoses, as all other characteristics remain the same. From our regression analysis (see Table 4), we found that both the main and secondary diagnoses greatly influence the duration of stay, while they have only a small effect on the intensity of care. By estimating the overall care severity, we integrate the effects on both the duration and intensity. We find that the combination of mental and osteoarticular diseases (in any order) leads to the longest estimated duration of stay of approximately 62 months, also triggering the highest overall care severity of approximately 3,650 hours. We observe that a tumour as the main diagnosis is care intensive, with an intensity ranging from 845 minutes/week, in combination with an osteoarticular secondary diagnosis, to 868 minutes/week with a secondary tumour diagnosis. However, the group with a main tumour diagnosis, $D1$, showed the shortest duration of stay (from 17 to 27 months), leading to the lowest overall severity (from 1,040 to 1,620 hours). As with tumours, heart-related diagnoses show a lower estimated duration in both main (from 35 to 55 months) and secondary (from 23 to 56 months) diagnoses. At the same time, they yield one of the lowest intensities of care (between 825 and 845 minutes/week when considering the main diagnosis) after the group of mental diagnoses. Again, the shape of the overall care severity is driven by the duration of stay because pathologies affect the intensity of care less.

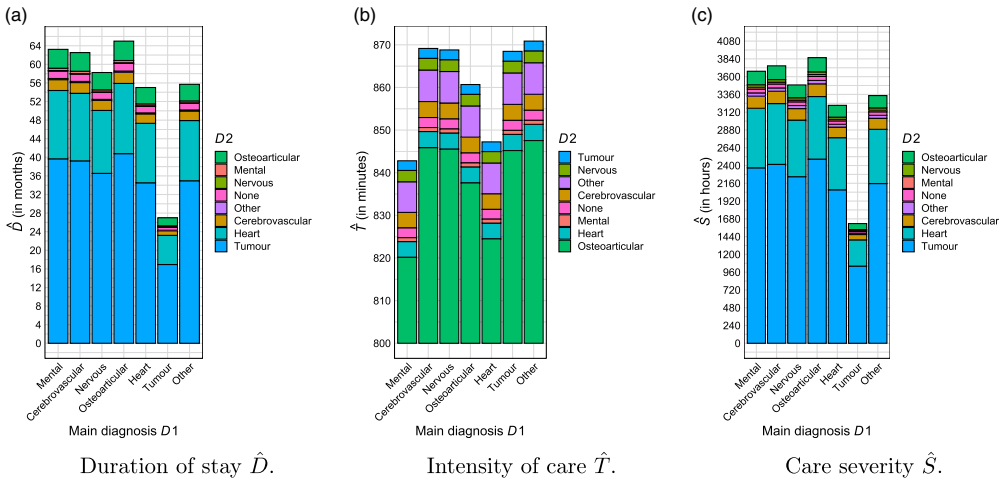


Figure 7. Estimated duration of stay, intensity of care and overall care severity for the modal profile with different main and secondary diagnoses.

Effect of the number of medical diagnoses. To investigate the effect of the number of additional diagnoses ND , we consider three profiles of an 80-year-old woman at entry with a mental main diagnosis and an “other” secondary diagnosis. First, in the “reference” profile, the considered elderly individuals have the lowest dependence levels and no other impairments. Second, in the “modal” profile, dependence and impairments take the mode values as laid out above. Third, in the “severe” profile, all dependence levels and impairments take the highest possible values (i.e., dependence Level 9 and severe impairment). Additionally, we consider a fourth profile focusing on tumours: the “tumour” profile is based on the modal profile but with main and secondary tumour diagnoses. This additional profile is motivated by the observation that a tumour diagnosis noticeably influences the duration of stay (see Figure 1(a) and Table 4).

In Figure 8, we present the estimated duration of stay \hat{D} , intensity of care \hat{T} , and overall care severity \hat{S} for the four profiles across the number of additional diagnoses ND . We observe that with more diagnoses, the duration decreases and the intensity increases for all profiles, which is intuitive. However, each profile has its own character. For example, individuals with the reference profile require the least amount of help with 300 (330) minutes/week for $ND = 0$ ($ND = 7$). However, they stay longer in the institution with a mean duration of 130 (90) months when $ND = 0$ ($ND = 7$). This leads to an overall care severity between 2,800 and 2,120 hours. Individuals with the modal profile, on average, stay in the institution from 82 to 57 months with zero or seven additional diagnoses, respectively. At the same time, the intensity of care ranges from 770 to 840 minutes/week, leading to an overall care severity ranging from 4,480 to 3,380 hours and decreasing with ND .

Individuals with the severe profile stay in the institution slightly less long than the modal profile (range from 77 to 53 months). However, in contrast to the modal profile, individuals with the severe profile are the most demanding for help in their daily needs, with the care intensity ranging from 1,770 to 1,910 minutes/week. Thus, the resulting care severity is almost doubled compared to the modal profile (range from 9,640 to 7,200 hours). Finally, from Figure 7(b), we note that individuals with the tumour profile require, on average, 30 minutes of care per week less than elderly individuals with the modal profile. However, the duration of stay of these persons is only 24 months if they do not have any additional diagnoses ($ND = 0$). This duration even reduces to 17 months when $ND = 7$. As a result, the overall care severity of the tumour profile is the lowest, ranging from 1,380 to 1,040 hours.

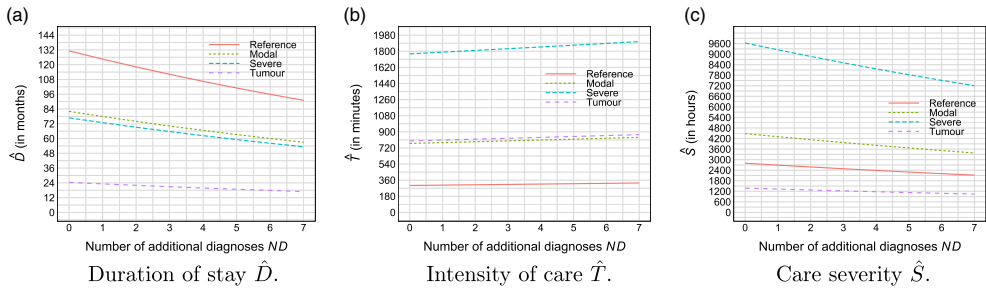


Figure 8. Estimated duration of stay, intensity of care and overall care severity along the number of additional diagnoses for different profiles.

6. Conclusion

In this paper, we propose a model to assess the total LTC burden based on the time spent in institutionalised care and the intensity of care provided to elderly individuals. We empirically estimate the model parameters using a private dataset comprising approximately 21,758 individuals from nursing homes in Switzerland. Due to the longitudinal nature of the right-censored data, we relied on survival analysis methods and used an AFT model for the duration of stay. After the age at entry and gender, we found that the pathologies were the key factors affecting the duration of stay in the institution before death. More precisely, individuals with mental and osteoarticular diseases stay the longest, while elderly people with tumours have the shortest duration. Further determinants of the duration are the dependence on ADL and physical mobility limitations. Finally, we find that some mild and moderate levels of physical and sensory function impairments are likely to receive less attention from nurses, which is linked to a reduced duration of stay.

We parameterise a beta regression model for the intensity of care to estimate the mean number of minutes of care per week that a person receives. While we show that pathologies have a minor effect on the amount of help, the key determinants are different limitations in ADL, physical mobility, orientation in space, and occupation. Our results further suggest that psychological and sensory function impairments shape only the distribution of the intensity of care, rather than considerably affecting the mean. Finally, we explore the results by studying selected profiles of the institutionalised elderly individuals. Thereby, we illustrate the overall care severity as the product of the duration of stay and intensity of care. We find that women entering an institution have a higher total care burden than men. We also show that mental and osteoarticular diseases lead to the highest overall severity, mainly resulting from the higher duration of stay. At the same time, tumours lead to the lowest duration of stay, resulting in the lowest care burden.

Our results are directly relevant for policy-makers and the planning of care capacity in institutions. The decomposition of the total number of hours of care in duration and intensity, or, in other words, in the occupancy period of a bed and in the utilisation of care resources, is relevant for deriving actionable policies. On the one hand, the length of stay in an institution drives the number of required places for LTC and the construction of infrastructure. On the other hand, the amount of help needed directly relates to the number of personnel and the education of caregivers. It is also a useful indicator to measure the efficiency of prevention measures, for example, to slow down the aggravation of medical conditions and to contain the increase in required care. Planning the future capacity of institutions and the handling of dependence at large must combine both dimensions of occupancy and prevention while supporting the efforts with policies promoting ageing at home associated with an efficient use of care resources.

In our work, we link demographic factors and medical diagnoses to the care requirements of elderly individuals. The care severity estimate gives an indication of the number of hours of care needed. Hence, it provides an estimate for the future care costs of an elderly individual upon entering an institution. The indicator can help to make the pricing of LTC coverage more precise. For

example, innovative insurance contracts could evaluate the one-time premium to guarantee a life annuity underwritten at the moment of entering a care institution. While a large part of the care costs are financed today by health insurers, copayments are required by the elderly individuals to finance their stay in the institution. These may be more important depending on the type of institution (e.g., private and public facilities) and specific types of care (e.g. medical nursing and social company). While the numerical results may vary across countries and populations, we trust that our findings provide relevant indicators when assessing the LTC burden in institutions beyond Switzerland. Indeed, care severity is based on health conditions rather than on country-specific variables.

We provide robust results on institutional LTC estimates solely using information on elderly individuals from the moment when they enter an institution, that is, in this work, we consider only the first individual health evaluation. During the time in the institution, the health status evolves, and new evaluations are conducted every 1 or 2 years. With typically deteriorating health conditions, more information can provide updates on the remaining duration of stay and, more importantly, on the intensity of care, which, on average, increases. Therefore, since our model does not consider the care intensity dynamics, our estimates can potentially be enhanced. Furthermore, the study of the development of the different medical conditions and limitations over time could provide further insight and extend this work. Additionally, it is of interest to derive patterns from the data and to establish key profiles of elderly individuals to further understand the main drivers of the need for care and to better assess the corresponding workload for the personnel. Finally, in this paper, we have approximated the total care burden considering the product of duration and intensity. This approximation holds best if both factors are independent. Further research could study the dependence structure between the two dimensions.

Acknowledgment. Michel Fuino and Joël Wagner acknowledge financial support from the Swiss National Science Foundation (grant no. 100018_169662). All authors are thankful to the General Directorate for Health of the Republic and Canton of Geneva for providing the data used in this study.

References

- Albarrán, I., Alonso-González, P.J., Arribas-Gil, A. & Grané, A. (2019). How functional data can enhance the estimation of health expectancy: the case of disabled Spanish population. *ASTIN Bulletin*, **49**(1), 57–84.
- Anderson, J.J., Anthony, J.M., Zhang, Y., Wilson, P.W.F., Kelly-hayes, M., Wolf, P.A., Kreger, B.E. & Kannel, W.B. (1993). The effects of specific medical conditions on the functional limitations of elders in the Framingham study. *American Public Health Association*, **84**(3), 351–358.
- Arrighi, H.M., McLaughlin, T. & Leibman, C. (2010). Prevalence and impact of dementia-related functional limitations in the United States, 2001 to 2005. *Alzheimer Disease & Associated Disorders*, **24**(1), 72–78.
- Atkinson, A. (1985). *Plots, Transformations and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis*. Oxford Science Publications. Oxford University Press, Oxford.
- Avis, N.E. & Deimling, G.T. (2008). Cancer survivorship and aging. *Cancer*, **113**(12), 3519–3529.
- Barnett, K., Mercer, S.W., Norbury, M., Watt, G., Wyke, S. & Guthrie, B. (2012). Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. *The Lancet*, **380**(9836), 37–43.
- Biessy, G. (2017). Continuous-time semi-markov inference of biometric laws associated with a long-term care insurance portfolio. *ASTIN Bulletin*, **47**(2), 527–561.
- Boult, C., Kane, R.L., Louis, T.A., Boult, L. & McCaffrey, D. (1994). Chronic conditions that lead to functional limitation in the elderly. *Journal of Gerontology*, **49**(1), M28–M36.
- Boyle, G. (2005). The role of autonomy in explaining mental ill-health and depression among older people in long-term care settings. *Ageing and Society*, **25**(5), 731–748.
- Branch, L.G. & Jette, A.M. (1982). A prospective study of long-term care institutionalization among the aged. *American Journal of Public Health*, **72**(12), 1373–1379. PMID: [6814269](https://pubmed.ncbi.nlm.nih.gov/6814269/).
- Brown, J.R. & Finkelstein, A. (2009). The private market for long-term care insurance in the united states: a review of the evidence. *Journal of Risk and Insurance*, **76**(1), 5–29.

- Byles, J.E., D'Este, C., Parkinson, L., O'Connell, R. & Treloar, C. (2005). Single index of multimorbidity did not predict multiple outcomes. *Journal of Clinical Epidemiology*, **58**(10), 997–1005.
- Callahan, C.M., Arling, G., Tu, W., Rosenman, M.B., Counsell, S.R., Stump, T.E. & Hendrie, H.C. (2012). Transitions in care for older adults with and without dementia. *Journal of the American Geriatrics Society*, **60**(5), 813–820.
- Campbell, A.J., Diep, C., Reinken, J. & McCosh, L. (1985). Factors predicting mortality in a total population sample of the elderly. *Journal of Epidemiology & Community Health*, **39**(4), 337–342.
- Carrino, L., Orso, C.E. & Pasini, G. (2018). Demand of long-term care and benefit eligibility across european countries. *Health Economics*, **27**(8), 1175–1188.
- Centers for Medicare & Medicaid Services (2014). General Equivalence Mappings (GEMs) – Procedure Codes and Guide [accessed 11-April-2021].
- Chavan, P.P., Kedia, S.K. & Yu, X. (2017). Physical and functional limitations in US older cancer survivors. *Journal of Palliative Care & Medicine*, **7**(4), 1–8.
- Collett, D. (2015). *Modelling Survival Data in Medical Research. Chapman & Hall/CRC Texts in Statistical Science*. CRC Press, Boca Raton, Florida.
- Colombo, F., Llena-Nozal, A., Mercier, J. & Tjadens, F. (2011). *Help Wanted? Providing and Paying for Long-Term Care. OECD Health Policy Studies*. OECD Publishing, Paris.
- Cosandey, J. (2016). De nouvelles mesures pour les soins aux personnes âgées, *Avenir Suisse, Zurich*.
- Cribari-Neto, F. & Zeileis, A. (2010). Beta regression in R. *Journal of Statistical Software, Articles*, **34**(2), 1–24.
- Crimmins, E.M., Kim, J.K. & Solé-Auró, A. (2010). Gender differences in health: results from SHARE, ELSA and HRS. *European Journal of Public Health*, **21**(1), 81–91.
- Czado, C., Kastenmeier, R., Brechmann, E.C. & Min, A. (2012). A mixed copula model for insurance claims and claim sizes. *Scandinavian Actuarial Journal*, **2012**(4), 278–305.
- Davidson, I.A., Dewey, M.E. & Copeland, J.R.M. (1988). The relationship between mortality and mental disorder: evidence from the liverpool longitudinal study. *International Journal of Geriatric Psychiatry*, **3**(2), 95–98.
- Deeg, D.J.H., Portrait, F. & Lindeboom, M. (2002). Health profiles and profile-specific health expectancies of older women and men: the netherlands. *Journal of Women & Aging*, **14**(1–2), 27–46. PMID: [12537278](https://pubmed.ncbi.nlm.nih.gov/12537278/).
- de Meijer, C., Koopmanschap, M., d' Uva, T.B. & van Doorslaer, E. (2011). Determinants of long-term care spending: age, time to death or disability? *Journal of Health Economics*, **30**(2), 425–438.
- Donabedian, A. (1973). *Aspects of Medical Care Administration*. Harvard University Press, Cambridge.
- Dorr, D.A., Horn, S.D. & Smout, R.J. (2005). Cost analysis of nursing home registered nurse staffing times. *Journal of the American Geriatrics Society*, **53**(5), 840–845.
- Duell, D., Lindeboom, M., Koolman, X. & Portrait, F. (2019). Practice variation in long-term care access and use: the role of the ability to pay. *Health Economics*, **28**(11), 1277–1292.
- Farias, S.T., Lau, K., Harvey, D., Denny, K.G., Barba, C. & Mefford, A.N. (2017). Early functional limitations in cognitively normal older adults predict diagnostic conversion to mild cognitive impairment. *Journal of the American Geriatrics Society*, **65**(6), 1152–1158.
- Ferrari, S. & Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, **31**(7), 799–815.
- Fong, J.H., Sherris, M. & Yap, J. (2017). Forecasting disability: application of a frailty model. *Scandinavian Actuarial Journal*, **2017**(2), 125–147.
- Fries, J.F. (1980). Aging, natural death, and the compression of morbidity. *The New England Journal of Medicine*, **303**(3), 130–135.
- Fuino, M., Rudnytskyi, I. & Wagner, J. (2020). On the characteristics of reporting ADL limitations and formal LTC usage across Europe. *European Actuarial Journal*, **10**, 557–597.
- Fuino, M. & Wagner, J. (2018). Long-term care models and dependence probability tables by acuity level: new empirical evidence from Switzerland. *Insurance: Mathematics and Economics*, **81**, 51–70.
- Fuino, M. & Wagner, J. (2020). Duration of long-term care: socio-economic factors, type of care interactions and evolution. *Insurance: Mathematics and Economics*, **90**, 151–168.
- Germain, C.M., Vasquez, E., Batsis, J.A. & McQuoid, D.R. (2016). Sex, race and age differences in muscle strength and limitations in community dwelling older adults: data from the health and retirement survey (hrs). *Archives of Gerontology and Geriatrics*, **65**, 98–103.
- Goodwin, J.S. (1991). A population-based study of functional status and social support networks of elderly patients newly diagnosed with cancer. *Archives of Internal Medicine*, **151**(2), 366.
- Grambsch, P.M. & Therneau, T.M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, **81**(3), 515–526.
- Grün, B., Kosmidis, I. & Zeileis, A. (2012). Extended beta regression in r: shaken, stirred, mixed, and partitioned. *Journal of Statistical Software, Articles*, **48**(11), 1–25.

- Guccione, A.A., Felson, D.T., Anderson, J.J., Anthony, J.M., Zhang, Y., Wilson, P.W., Kelly-Hayes, M., Wolf, P.A., Kreger, B.E. & Kannel, W.B. (1994). The effects of specific medical conditions on the functional limitations of elders in the framingham study. *American Journal of Public Health*, **84**(3), 351–358. PMID: [8129049](#).
- Guibert, Q. & Planchet, F. (2018). Non-parametric inference of transition probabilities based on Aalen–Johansen integral estimators for acyclic multi-state models: application to LTC insurance. *Insurance: Mathematics and Economics*, **82**, 21–36.
- Hedinger, D., Hämmig, O. & Bopp, M. (2015). Social determinants of duration of last nursing home stay at the end of life in Switzerland: a retrospective cohort study Health services research. *BMC Geriatrics*, **15**(1), 1–8.
- Hirsch, D. (2005). *Facing the Cost of Long-term Care: Towards a Sustainable Funding System*. Joseph Rowntree Foundation, York.
- Home Care Association of the Canton of Vaud (2020). *Les Tarifs*.
- Hu, T. (1986). The economic impact of urinary incontinence. *Clinics in Geriatric Medicine*, **2**(4), 673–687.
- Hyer, L., Carpenter, B., Bishmann, D. & Wu, H.-S. (2005). Depression in long-term care. *Clinical Psychology: Science and Practice*, **12**(3), 280–299.
- Inouye, S.K., Studenski, S., Tinetti, M.E. & Kuchel, G.A. (2007). Geriatric syndromes: clinical, research, and policy implications of a core geriatric concept. *Journal of the American Geriatrics Society*, **55**(5), 780–791.
- Institute of Medicine (1991). *Disability in America: Toward a National Agenda for Prevention*. The National Academies Press, Washington, DC.
- Jennings, L.A., Hollands, S., Keeler, E., Wenger, N.S. & Reuben, D.B. (2020). The effects of dementia care co-management on acute care, hospice, and long-term care utilization. *Journal of the American Geriatrics Society*, p. jgs.16667.
- Kaplan, E.L. & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, **53**(282), 457–481.
- Karlsson, M., Mayhew, L., Plumb, R. & Rickayzen, B. (2006). Future costs for long-term care: cost projections for long-term care for older people in the United Kingdom, *Health Policy*, **75**(2), 187–213.
- Katz, P.R. (2011). An international perspective on long term care: focus on nursing homes. *Journal of the American Medical Directors Association*, **12**(7), 487–492.e1.
- Kempen, G.I., Myers, A.M. & Powell, L.E. (1995). Hierarchical structure in ADL and IADL: analytical assumptions and applications for clinicians and researchers. *Journal of Clinical Epidemiology*, **48**(11), 1299–1305.
- Kitchener, M., Ng, T., Miller, N. & Harrington, C. (2006). Institutional and community—based long-term care. *Journal of Health & Social Policy*, **22**(2), 31–50.
- Klein, J.D. & Moeschberger, M.L. (1997). *Survival Analysis: Techniques for Censored and Truncated Data*. *Statistics for Biology and Health*. Springer-Verlag, New York.
- Koroukian, S.M., Schiltz, N., Warner, D.F., Sun, J., Bakaki, P.M., Smyth, K.A., Stange, K.C. & Given, C.W. (2016). Combinations of chronic conditions, functional limitations, and geriatric syndromes that predict health outcomes. *Journal of General Internal Medicine*, **31**(6), 630–637.
- Kuo, H.K., Jones, R.N., Milberg, W.P., Tennstedt, S., Talbot, L., Morris, J.N. & Lipsitz, L.A. (2005). Effect of blood pressure and diabetes mellitus on cognitive and physical functions in older adults: a longitudinal analysis of the advanced cognitive training for independent and vital elderly cohort. *Journal of the American Geriatrics Society*, **53**(7), 1154–1161.
- Kutner, M.H., Nachtshiem, C.J., Neter, J. & Li, W. (2005). *Applied Linear Regression Models*, 5 edition. McGraw-Hill/Irwin, New York.
- Le Corre, P.-Y. (2012). *Long-Term Care Insurance: Building a Successful Development* (pp. 53–72). Palgrave Macmillan, UK, London.
- Lesman-Leegte, I., Jaarsma, T., Coyne, J.C., Hillege, H.L., Van Veldhuisen, D.J. & Sanderman, R. (2009). Quality of life and depressive symptoms in the elderly: a comparison between patients with heart failure and age- and gender-matched community controls. *Journal of Cardiac Failure*, **15**(1), 17–23.
- Lundberg, F. (1903). *Approximerad framställning af sannolikhetsfunktioner. Återförsäkring af kollektivrisiker*. PhD thesis, University of Uppsala.
- Marengoni, A., Angleman, S., Melis, R., Mangialasche, F., Karp, A., Garmen, A., Meinow, B. & Fratiglioni, L. (2011). Aging with multimorbidity: a systematic review of the literature. *Ageing Research Reviews*, **10**(4), 430–439.
- Maresova, P., Javanmardi, E., Barakovic, S., Barakovic Husic, J., Tomson, S., Krejcar, O. & Kuca, K. (2019). Consequences of chronic diseases and other limitations associated with old age – a scoping review. *BMC Public Health*, **19**(1), 1431.
- Mathers, C. (1996). Trends in health expectancies in Australia 1981–1993. *Journal of the Australian Population Association*, **13**, 1–15.
- Mathers, C.D., Vos, E.T., Stevenson, C.E. & Begg, S.J. (2001). The burden of disease and injury in Australia. *Bulletin of the World Health Organization*, **79**(11), 1076–1084.
- Menotti, A., Mulder, I., Nissinen, A., Giampaoli, S., Feskens, E.J. & Kromhout, D. (2001). Prevalence of morbidity and multimorbidity in elderly male populations and their impact on 10-year all-cause mortality: the fine study (finland, italy, netherlands, elderly). *Journal of Clinical Epidemiology*, **54**(7), 680–686.
- Miller, E.A. & Weissert, W.G. (2000). Predicting elderly people’s risk for nursing home placement, hospitalization, functional impairment, and mortality: a synthesis. *Medical Care Research and Review*, **57**(3), 259–297. PMID: [10981186](#).

- Mommaerts, C.** (2020). Long-term care insurance and the family, Working Paper, University of Wisconsin.
- Moore, D.C., Keegan, T.J., Dunleavy, L. & Froggatt, K.** (2019). Factors associated with length of stay in care homes: a systematic review of international literature. *Systematic Reviews*, *8*(1), 56.
- Muir, T.** (2017). Measuring social protection for long-term care, OECD Health Working Papers, (93).
- Nichols, B.L., Davis, C.R. & Richardson, D.R.** (2010). An integrative review of global nursing workforce issues. *Annual Review of Nursing Research*, *28*(1), 113–132.
- Okma, K. & Gusmano, M.K.** (2020). Aging, pensions and long-term care: what, why, who, how? comment on “financing long-term care: lessons from Japan”. *International Journal of Health Policy and Management*, *9*(5), 218–221.
- Olshansky, S.J., Rudberg, M.A., Carnes, B.A., Cassel, C.K. & Brody, J.A.** (1991). Trading off longer life for worsening health: the expansion of morbidity hypothesis. *Journal of Aging and Health*, *3*(2), 194–216.
- Pack, S.** (2009). Poor physical health and mortality in patients with schizophrenia. *Nursing Standard*, *23*(21), 41–45.
- Pang, G. & Warshawsky, M.** (2010). Optimizing the equity-bond-annuity portfolio in retirement: the impact of uncertain health expenses. *Insurance: Mathematics and Economics*, *46*(1), 198–209.
- Pinquart, M. & Sörensen, S.** (2011). Spouses, adult children, and children-in-law as caregivers of older adults: a meta-analytic comparison. *Psychology and Aging*, *26*(1), 1–14.
- Pritchard, D.J.** (2006). Modeling disability in long-term care insurance. *North American Actuarial Journal*, *10*(4), 48–75.
- Republic and Canton of Geneva**, General Directorate for Health (2019). EROS/PLAISIR data from 1996–2018.
- Rickayzen, B.D. & Walsh, D.E.P.** (2002). A multi-state model of disability for the united kingdom: implications for future need for long-term care for the elderly. *British Actuarial Journal*, *8*(2), 341–393.
- Roussel, B. & Tilquin, C.** (1993). *Bio-psycho-social profile of clients in long term care and planning of nursing care required*. Équipe de Recherche Opérationnelle en Santé. Reference manual.
- Rudnytskyi, I. & Wagner, J.** (2019). Drivers of old-age dependence and long-term care usage in Switzerland—a structural equation model approach. *Risks*, *7*(3), 92.
- Schünemann, J., Strulik, H. & Trimborn, T.** (2017). The gender gap in mortality: how much is explained by behavior? *Journal of Health Economics*, *54*, 79–90.
- Seematter-Bagnoud, L., Fustinoni, S., Meylan, L., Monod, S., Junod, A., Büla, C. & Santos-Eggimann, B.** (2012). *Le Resident Assessment Instrument (RAI) et Autres Instruments d'Évaluation, d'Orientation et de Communication Relatifs aux Personnes Âgées dans le Réseau de Soins*. Centre d'Observation et d'Analyse du Vieillessement, Lausanne.
- Shao, A.W., Chen, H. & Sherris, M.** (2019). To borrow or insure? Long term care costs and the impact of housing. *Insurance: Mathematics and Economics*, *85*, 15–34.
- Shao, A.W., Hanewald, K. & Sherris, M.** (2015). Reverse mortgage pricing and risk analysis allowing for idiosyncratic house price risk and longevity risk. *Insurance: Mathematics and Economics*, *63*, 76–90.
- Silverman, B.W.** (1986). *Density Estimation for Statistics and Data Analysis*. Chapman & Hall/CRC, London.
- Simard, J. & Volicer, L.** (2020). Loneliness and isolation in long-term care and the Covid-19 pandemic. *Journal of the American Medical Directors Association*, *21*(7), 966–967.
- Simas, A.B., Barreto-Souza, W. & Rocha, A.V.** (2010). Improved estimators for a general class of beta regression models. *Computational Statistics & Data Analysis*, *54*(2), 348–366.
- Sinclair, A.J., Conroy, S.P. & Bayer, A.J.** (2008). Impact of diabetes on physical function in older people. *Diabetes Care*, *31*(2), 233–235.
- Social Insurance and Accommodation Service of the Canton of Vaud** (2016). *La Facture de l'EMS*.
- Swiss Federal Department of Home Affairs** (2021). *Ordonnance du DFI sur les Prestations dans l'Assurance Obligatoire des Soins en Cas de Maladie (OPAS, 832.112.31)*.
- Swiss Federal Social Insurance Office** (2010). *Nouveau Régime de Financement des Soins*.
- Thakur, M. & Blazer, D.G.** (2008). Depression in long-term care. *Journal of the American Medical Directors Association*, *9*(2), 82–87.
- Therneau, T.M.** (2021). *A Package for Survival Analysis in R*. R package version 3.2-10.
- Tomas, J. & Planchet, F.** (2013). Multidimensional smoothing by adaptive local kernel-weighted log-likelihood: application to long-term care insurance. *Insurance: Mathematics and Economics*, *52*(3), 573–589.
- Tombaugh, T.N. & McIntyre, N.J.** (1992). The mini-mental state examination: a comprehensive review. *Journal of the American Geriatrics Society*, *40*(9), 922–935.
- Tomiaik, M., Berthelot, J.-M., Guimond, E. & Mustard, C.A.** (2000). Factors associated with nursing-home entry for elders in Manitoba, Canada. *The Journals of Gerontology: Series A*, *55*(5), M279–M287.
- van den Akker, M., Buntinx, F., Metsemakers, J.F., Roos, S. & Knottnerus, J.** (1998). Multimorbidity in general practice: prevalence, incidence, and determinants of co-occurring chronic and recurrent diseases. *Journal of Clinical Epidemiology*, *51*(5), 367–375.
- Vanella, P., Heß, M. & Wilke, C.B.** (2020). A probabilistic projection of beneficiaries of long-term care insurance in Germany by severity of disability. *Quality & Quantity*, *54*, 943–974.
- Victor, C.R.** (2012). Loneliness in care homes: a neglected area of research?. *Aging Health*, *8*(6), 637–646.

- von Strauss, E., Agüero-Torres, H., Kåreholt, I., Winblad, B. & Fratiglioni, L. (2003). Women are more disabled in basic activities of daily living than men only in very advanced ages: a study on disability, morbidity, and mortality from the kungsholmen project. *Journal of Clinical Epidemiology*, *56*(7), 669–677.
- Waitzberg, R., Schmidt, A.E., Blümel, M., Penneau, A., Farmakas, A., Ljungvall, Å., Barbabella, F., Augusto, G.F., Marchildon, G.P., Saunes, I.S., Vočanec, D., Miloš, I., Contel, J.C., Murauskiene, L., Kroneman, M., Tambor, M., Hroboň, P., Wittenberg, R., Allin, S. & Or, Z. (2020). Mapping variability in allocation of long-term care funds across payer agencies in OECD countries, *Health Policy*, *124*(5), 491–500.
- World Health Organization (1980). International classification of impairments, disabilities, and handicaps: a manual of classification relating to the consequences of disease, published in accordance with resolution WHA29.35 of the Twenty-ninth World Health Assembly, May 1976.
- Xue, Q.-L. (2011). The frailty syndrome: definition and natural history. *Clinics in Geriatric Medicine*, *27*(1), 1–15. Frailty.