WIFO ■ WORKING PAPERS
707/2025

# Predicting the Uptake of Long-Term Care Benefits in Austria

Ulrike Famira-Mühlberger
Klaus Nowotny

# Predicting the Uptake of Long-Term Care Benefits in Austria

Ulrike Famira-Mühlberger, Klaus Nowotny

Research assistance: Lydia Grandner, Manuel Zerobin

WIFO Working Papers 707/2025
June 2025

## Abstract

We use administrative microdata and statistical learning methods to analyse how personal characteristics and the consumption of healthcare services help predict the first-time receipt of "long-term care allowance" (LTCA), a needs-tested cash-for-care benefit in Austria. Our findings suggest that short-term information from the health-care sector, particularly in the quarter prior to LTCA enrolment, provides substantial explanatory power. Apart from old age, the most influential predictors include the frequency of doctor visits and hospital stays as well as diagnoses such as dementia, cerebral infarction, and hypertension. Our findings emphasise the importance of data-driven approaches in anticipating the uptake of long-term care benefits and informing policy, especially against the background of the demographic transition.

E-Mail: ulrike.famira-muehlberger@wifo.ac.at, klaus.nowotny@wifo.ac.at

2025/1/W/0



Media owner (publisher), producer: Austrian Institute of Economic Research
1030 Vienna, Arsenal, Objekt 20 | Tel. (43 1) 798 26 01 0 | https://www.wifo.ac.at
Place of publishing and production: Vienna

# Predicting the uptake of long-term care benefits in Austria

Ulrike Famira-Mühlberger[1] and Klaus Nowotny[2,1]

[1]Austrian Institute of Economic Research (WIFO), Vienna, Austria
[2]Paris Lodron University Salzburg, Austria

June, 2025

**Abstract**

We use administrative microdata and statistical learning methods to analyze how personal characteristics and the consumption of healthcare services help predict the first-time receipt of "long-term care allowance" (LTCA), a needs-tested cash-for-care benefit in Austria. Our findings suggest that short-term information from the healthcare sector, particularly in the quarter prior to LTCA enrollment, provides substantial explanatory power. Apart from old age, the most influential predictors include the frequency of doctor visits and hospital stays as well as diagnoses such as dementia, cerebral infarction, and hypertension. Our findings emphasize the importance of data-driven approaches in anticipating the uptake of long-term care benefits and informing policy, especially against the background of the demographic transition.

**JEL classification:** C53, H51, I18, J14

Based on research for the Federation of Austrian Social Insurances. We thank Lydia Grandner and Manuel Zerobin for valuable research assistance and are grateful for helpful comments from Thomas Leoni and René Böheim as well as conference participants at the Annual Conference 2022 of the International Long-term Care Policy Network, the Annual Conference 2022 of the Austrian Health Economics Association, the Annual Conference 2024 of the Austrian ESPANet and the Data Science Interest Group of the Global Observatory of Long-Term Care.

# 1 Introduction

Europe's demographic landscape is undergoing a profound transformation. By 2070, over 30 % of the EU population will be 65 years of age or older, with the number of people requiring long-term care (LTC) projected to increase by 23 % by 2070 compared to 2022 (European Commission, Directorate General for Economic and Financial Affairs, 2024). Austria exemplifies this trend: 22 % of citizens aged 65 or older currently receive "Long-Term Care Allowance" (LTCA), a cash benefit based on needs, but not means tested for individuals in need of long-term care. This figure is expected to grow by 50 % by 2050 (Famira-Mühlberger, 2024). The demographic shift intersects with declining family care capacity (Verbakel, 2018), workforce shortages (OECD, 2020), and fiscal strain (European Commission, Directorate General for Economic and Financial Affairs, 2024). We address two important questions: 1) Can we predict the uptake of LTCA using administrative healthcare data? and 2) what acute health events precipitate care dependency?

These questions have gained growing relevance in light of the increasing demands on long-term care systems in aging societies and the parallel need for evidence-based policy responses that promote both sustainability and equity in care provision. Improving the accuracy of LTC-need forecasts is highly valuable: if routine health-utilization data (recent hospital admissions, doctor visits, prescriptions, etc.) can reliably signal who will soon require care, policymakers could target early interventions (medical, rehabilitative, or social support) to delay dependency and reduce costs. Likewise, identifying which acute health events precipitate loss of independence is critical for designing prevention measures. For example, severe neurological episodes (strokes, the onset of dementia) often herald the entry into care. Determining whether and how administrative healthcare records can predict entry into the LTC allowance system and which health shocks precede that entry is key to enabling more timely, cost-effective care planning.

In Austria, the LTCA functions as a central support mechanism for individuals requiring sustained care, though the empirical relationship between healthcare service utilization and LTCA receipt remains underexplored. We address this gap by analyzing how personal characteristics and healthcare service usage patterns predict subsequent LTCA enrollment through advanced statistical learning techniques applied to comprehensive administrative microdata. Employing cross-validated regularization methods on high-dimensional datasets containing thousands of potential variables, we identify key predictors of first-time LTCA entry while maintaining rigorous predictive performance.

Our primary contribution lies in its novel integration of detailed administrative health records with LTCA allocation data to develop a robust predictive model within Austria's universal tax-funded care system. By combining these datasets through machine learning approaches, we reveal previously unidentified interactions between healthcare utilization patterns and the uptake of long-term care benefits, offering policymakers insights for resource allocation and early intervention strategies in social care systems.

Our analysis reveals that short-term healthcare utilization—particularly in the quarter immediately preceding LTCA receipt—is highly predictive for receiving LTCA, highlighting the acute health shocks that often precipitate care needs. While adding more detailed medical information, such as specialist visits, diagnoses, and drug dispensings, yields only marginal improvements and increases the risk of overfitting, regularization methods maintain both model interpretability and performance. Age is as the strongest predictor, alongside intensified healthcare use and specific diagnoses such as dementia, cerebral infarction, and hypertension. Although these models demonstrate strong overall predictive accuracy, the low incidence of LTCA entry at the individual level currently limits the development of reliable short-term leading indicators, suggesting that further methodological advances are needed. Nevertheless, aggregating individual-level predictions could offer significant value for policymakers, enabling more effective (regional) resource allocation and improved systemic capacity planning at broader levels, and ensuring that LTC budgeting decisions are grounded in robust empirical evidence.

## 2 The predictors of long-term care benefits

The literature shows that predictors of long-term care needs are predominantly health-related. For example, evidence demonstrates that chronic conditions like hypertension, heart disease and neurological disorders exponentially increase LTC utilization likelihood compared to age-matched controls (Liu et al., 2013). Mental health conditions, particularly depression and cognitive impairments such as dementia, also significantly impact the need for long-term care: dementia increases the likelihood of institutionalization dramatically, especially at advanced stages, thus accounting for a high share of LTC costs in European countries (Bu & Rutherford, 2019; sm-Rahman et al., 2021). The importance of health-related factors for LTC utilization is amplified if health problems become more severe and/or numerous (multimorbidity): the severity of health problems is directly related to the likelihood of institutionalization, as individuals with more severe health impairments are more likely to require formal care (Schulz et al., 2020).

The importance of health-related factors mostly stems from their influence on functional limitations in activities of daily living (ADL) and instrumental activities of daily living (IADL), which also serve as universal gateways to long-term care benefits across international systems. Here, health-related predictors affect three critical dimensions: 1) physical capacity (mobility restrictions, chronic disease progression), 2) cognitive functions (dementia severity, executive dysfunctions) and 3) complex care needs (multimorbidity, frailty, neurological disorders).

Negative health shocks are also related to frailty, another important predictor of the need for LTC. Schulz et al. (2020) have shown that frail older people are significantly more likely to require long-term care services than their non-frail counterparts, and frailty increases vulnerability to adverse health outcomes, including dependency and mortality (Reeves et al., 2018).

In the absence of direct measures of (I)ADL limitations or frailty, insurance and healthcare claims data can be used to predict future LTC needs. Sato et al. (2022) develop and validate machine learning models to predict individual-level future demand for LTC in Japan. Key predictors identified include current eligibility level, age, indicators of deteriorating health (such as dementia diagnosis and frequency of medical consultations), and engagement in preventive care services (like rehabilitation and training). The authors demonstrate that claims-based models can robustly forecast changes in LTC needs, with implications for both resource planning and early intervention.

Apart from health-related predictors demographic factors, educational inequality, economic factors, and policy-related factors also play an important role. Among the *demographic factors*, age is a key predictor of receiving long-term care benefits. Women generally have a higher propensity to use long-term care benefits, reflecting their longer life expectancy and higher incidence of chronic conditions (Stolz et al., 2019). Being unmarried or living alone increases the likelihood of utilizing formal long-term care services, as these may lack informal care support systems (Larsson et al., 2014; Liu et al., 2013; Steinbeisser et al., 2018).

*Educational inequalities* in the need for long-term care are also well established. Individuals with a higher level of education have a lower proportion of years spent in need of care, an association that can be explained by various mechanisms, such as a higher level of health education, a healthier lifestyle and less physically demanding working conditions (Frangos et al., 2023; Grigoriev & Doblhammer, 2019; Stolz et al., 2019). Dynamic microsimulations of the future uptake of long-term care benefits show a strong moderating effect of the educational expansion of recent decades (Famira-Mühlberger et al., 2025; Warum et al., 2025).

*Economic factors*, including income and wealth, play an important role in determining access to LTC services. Higher income levels are associated with increased access to private LTC insurance and services, while lower income individuals often rely on public support programs (Da Roit & Le Bihan, 2010; Drzazga, 2021). The availability of cash-for-care schemes, which provide financial support rather than direct services, has been introduced in several European countries to increase choice and accessibility for families (Da Roit & Le Bihan, 2010). However, disparities in income and wealth can lead to inequitable access to long-term care services, and policy interventions are needed to address these gaps (Drzazga, 2021; Floridi et al., 2021).

*Policy-related factors*, including the structure and financing of LTC systems, have a significant impact on utilization patterns (Billings, 2013; Drzazga, 2021). Countries with comprehensive LTC policies that promote equity and accessibility tend to have better outcomes in terms of service utilization (Drzazga, 2021; Leichsenring et al., 2013). Furthermore, the role of migrant care workers in filling care gaps has become increasingly important in many European countries, reflecting the changing dynamics of LTC provision (Frimmel et al., 2023; Horn et al., 2019; Leiblfinger et al., 2021).

A multitude of factors is thus pivotal in determining LTC dependency and the use of state-provided LTC benefits. In turn, LTC benefits can also influence levels of dependency among recipients. International research highlights that LTC benefits can significantly reduce hospital admissions by improving chronic disease management and preventing avoidable hospitalizations (e.g., Costa-Font et al., 2018; Guo et al., 2015; Xu et al., 2022). Additionally, these benefits lessen reliance on institutional care by enabling individuals to remain at home with appropriate support (Han et al., 2019; Rice et al., 2018). They also alleviate the burden on informal caregivers by reducing caregiving demands and improving caregivers' quality of life (Kim & Lim, 2015; Løken et al., 2017). Furthermore, LTC benefits have been shown to decrease the demand for medical services, such as emergency visits and inpatient stays, thereby easing pressure on healthcare systems (Bakx et al., 2020; Rice et al., 2018). While these interventions have the potential to generate cost savings—such as reducing hospitalizations during nursing home stays (Han et al., 2019; Serrano-Alarcón et al., 2022)—they often face challenges in reaching intended target populations. For instance, non-take up rates and inequitable access remain persistent issues (Pennerstorfer & Österle, 2023; Vidiella-Martin et al., 2024). This body of evidence underscores the critical role of LTC benefits in mitigating healthcare system strain and supporting families, though further efforts are needed to improve access and equity in certain contexts.

In short, the predictors of long-term care (LTC) use are complex and interrelated, encompassing demographic, health, economic, and policy factors. To deepen our understanding of the factors driving the demand for LTC and LTC benefits, it is essential to develop a nuanced understanding of these predictors.

## 3 Long-term care allowance in Austria

In many countries, people in need of long-term care are supported by the public sector through cash benefits and benefits in kind (nursing homes, residential care facilities, home care services). In 1993, Austria introduced a needs-based long-term care allowance ("Pflegegeld"). Eligible Austrian citizens are legally entitled to this allowance, regardless of income and assets and regardless of the cause of the need for long-term care.

To determine eligibility, persons who apply for LTCA are assessed by trained and qualified experts such as doctors or nurses, who follow a fixed assessment scheme based on legal and regulatory guidelines.[1] The assessment focuses mainly on limitations in activities of daily living (ADL) and instrumental activities of daily living (IADL).[2] LTCA applicants are assessed and granted an allowance based on their level of dependency, measured as the number of hours of care required per month. Benefit levels range from a minimum of 65 hours per month (level 1) to more than 180 hours (levels 5–7, Table 1). In addition, specific diagnoses such as dementia trigger an automatic 25-hour care premium in the assessment

[1] These include the Classification Regulation for the Federal Care Allowance Act (BGBl. II Nr. 37/1999) and the Guidelines for the uniform application of the Federal Care Allowance Act (RPGG 2012).

[2] For example, applicants are assessed on their ability to perform specific tasks such as getting dressed, walking, getting up, eating, using the toilet, and shopping.

process (Trukeschitz et al., 2022) and certain medical conditions—including blindness and paraplegia—qualify for predefined benefit levels, creating standardized support pathways for these populations. The allowance can be used to (partly) cover various care services, including in-home care, professional care services, and assistance from family members, although it does not fully cover the costs of the care required on average (Nagl-Cupal et al., 2018). There is no obligation to document the use of the LTCA.

Table 1: Long-term care allowance in Austria (2025)

| LTCA level | Criteria | Amount of LTC allowance p.m. in € |
|---|---|---|
| 1 | > 65 hours of care required p.m. | 200.80 |
| 2 | > 95 hours of care required p.m. | 370.30 |
| 3 | > 120 hours of care required p.m | 577.00 |
| 4 | > 160 hours of care required p.m. | 865.10 |
| 5 | > 180 hours of care p.m. if an extraordinary amount of care is required | 1,175.20 |
| 6 | > 180 hours of care p.m. if uncoordinated care measures are required, and these must be provided regularly during the day and night, or the permanent presence of a carer is required during the day and night because there is a likelihood of danger to oneself or others | 1,641.10 |
| 7 | > 180 hours of care p.m. if no purposeful movements of the four extremities with functional realization are possible or a condition requiring equal attention is present | 2,156.60 |

Source: BMSGPK (2025). The LTCA is yearly adjusted to inflation.

The granting of care allowance or its specific categorization has far-reaching consequences for other benefits in the care sector. Financial support for live-in care is only available to people receiving at least care allowance level 3.[3] The benefit for care-providing relatives (*"Angehörigenbonus"*) introduced in 2024 is linked to a minimum classification in LTCA level 4.[4] The subsidy for the use of mobile care services is regulated differently in the individual federal states, but in most federal states the amount of the subsidy is linked to the care allowance classification. Admission to a nursing home is also generally linked to a minimum classification in the care allowance system (usually care allowance level 4). Subsidized carer's leave is only possible if the person being cared for is classified at level 3. Finally, social insurance support for carers can only be claimed from care allowance level 3.

Approximately half of all LTCA beneficiaries in Austria are classified in care levels 1 and 2. In 2023, an average of around 476,000 individuals—representing 5.2 % of the total Austrian population—received the long-term care allowance (BMSGPK, 2024). This proportion increases significantly with age, reaching 22 % among those aged 65 and older,

[3]This financial support for live-in care is a government subsidy for round-the-clock home care for individuals in need of continuous assistance who hire a migrant care worker, who is usually self-employed. The subsidy is usually € 800 per month.

[4]That is a tax-free monthly payment (€ 130.80 as of 2025) for caregivers who reduce their employment to provide long-term home care for a close relative.

and 53 % among those aged 80 and above. The LTCA is funded through general federal tax revenues and accounts for approximately 0.65 % of Austria's GDP.

It is important to note that the receipt of LTCA is not equivalent to the need for long-term care, and that we can only analyze the former and not the latter. This distinction is especially relevant given anecdotal evidence of imperfect LTCA take-up among some strata of the population, especially foreigners living in Austria (Pennerstorfer & Österle, 2023). Furthermore, the LTCA requires a minimum of 65 hours of care per month, and persons who require a lower amount are not included in the system. Nevertheless, it can be assumed that the vast majority of potentially eligible persons do apply for LTCA, especially since it is not means-tested, so that the receipt of LTCA is a good proxy for the need for long-term care.

# 4 Data and methods

To analyze which factors are most strongly related to the first-time receipt of LTCA, we use two sets of pseudonymized individual-level administrative data provided by the Austrian Federation of Social Insurances, the umbrella organization of the statutory social insurance institutions of Austria.

The first set is our "case sample", an (unbalanced) monthly panel that includes 550,409 individuals aged 60 and over who received any level of the LTCA at least once between January 1, 2016 and December 31, 2018, including 388,024 individuals who already were LTCA recipients in January 2016.[5] The remaining 162,385 individuals received the LTCA for the first time between February 2016 and December 2018.[6] Thus, on average 4,640 individuals enter the LTCA system each month, 77.8 % of whom (about 3,611 per month) are between 60 and 85 years of age. The mean age of first-time receivers in our data set is 78.2 years, and more than half (50.7 %) enter the system at the lowest allowance level 1.[7] 52.7 % of first-time receivers are female, which is slightly lower than the population share of women among persons age 60 and over in Austria (55.7 %).

The second data set is our "control sample", an (unbalanced) monthly panel of 435,332 individuals between 60 and 85 years of age who did not receive the long-term care allowance

[5]Mortaility is the main reason why the panel is unbalanced. In sum, 168,151 individuals in our dataset pass away between January 2016 and December 2018.

[6]In our data, we measure the receipt of LTCA by an indicator variable that is either zero (for non-receivers) or between 1 and 7, representing the respective level of LTCA received at time $t$. We identify new recipients by a switch in this indicator from zero to a value between 1 and 7 from one month to the next. Since there is no information on the level of LTCA in December 2015, our data does not allow us to distinguish between those who received LTCA for the first time in January 2016 and those who were already LTCA recipients in December 2015.

[7]Another 20.4 % of first-time receivers enter at level 2, 14.3 % at level 3, 8.1 % at level 4 and 4.6 % at level 5. It is rare for first-time receivers to enter at levels 6 (1.3 % of all first-time recipients) or 7 (0.5 %). Thus, in general, first-time receivers enter at relatively low levels of the LTCA. Changes in the LTCA level are, however, quite frequent, and we observe 391,919 level changes, of which 96.5 % are increases, mostly by one level (52.6 %).

during the same period. These individuals were randomly chosen based on a stratified random sample from the universe of all individuals in the public health system such that for each individual receiving LTCA between January 1, 2016 and December 31, 2018 there are two individuals in the same stratum who did not receive LTCA. Together with the case sample, the control sample will be used to classify the first-time receipt of the LTCA (see below). Strata were defined as the interactions of region of residence (at the NUTS-2 level), sex (female and male) and 5-year age brackets. As LTCA receipt increases strongly with age, only one person in the same stratum could be chosen for individuals in the age bracket between 81 and 85 years. Data for older cohorts are not available for analysis, but given that the majority of first-time recipients are between 60 and 85 years of age, the control sample has sufficient common support to classify the vast majority of first-time receivers in the case sample.

In addition to details about the LTCA status and level, both data sets contain not only information about personal characteristics such as age, sex, and place of residence (at the NUTS-3 level, i. e. districts or groups of municipalities), but also about the consumption of health services, such as detailed information on the frequency of contact with general practitioners, specialist doctors and other contractual partners (bandage specialists, emergency services, etc., see appendix Table A.2 for a full list). Consultations of general practitioners and specialist doctors who have a contract with at least one public health insurance provider are, in general, free of charge for patients in the Austrian public health insurance system, and doctors directly invoice the patient's insurance provider for each visit.[8] Every time a patient consults a doctor, a "contact" is registered in our data and this information is aggregated at the monthly level.

Furthermore, the data sets contain information on the drugs dispensed according to the "Anatomical-Therapeutic-Chemical Classification" (ATC). Unfortunately, information on the quantity or the recommended frequency of consumption is not included; instead, we proxy drug consumption using indicator variables that indicate whether a drug with a specific ATC code was dispensed in a given month (1, zero else). Finally, they contain data from the hospital sector. The latter provide information on the frequency and duration of inpatient hospital stays (aggregated at the monthly level) as well as the principal and (if given) secondary diagnoses (according to the "International Statistical Classification of Diseases and Health Problems", ICD-10) and the individual medical services provided. Both diagnoses and medical services are measured using indicator variables.

In our exploratory empirical analysis, personal characteristics and the use or frequency of use of healthcare services during a defined "observation period" $[t-a, t)$ are used to model the occurrence of first-time receipt of long-term care allowance in a subsequent "event period" $[t, t+b)$.[9] For this purpose, we apply statistical learning methods. These

[8] According to the Austrian Medical Association, 59.7 % of all general practitioners with a doctor's office have had contracts with at least one public health insurance provider in 2023, see ÖAK, 2024.

[9] Following the conventional interval notation, square brackets indicate that an endpoint is included in the interval while parentheses indicate that an endpoint is excluded. The interval $[01/2017, 01/2018)$ for example includes January 1, 2017, but ends on December 31, 2017.

are suitable for filtering out those variables from a large number of potential explanatory factors that, ceteris paribus, have the strongest statistical correlation with the future occurrence of the event in question. This is especially relevant in this context because it allows us to make use of the detailed information available in the administrative data, which differentiates between 58 types of medical specialists and doctors, 483 Level 4 ATC codes, 1,483 3-digit ICD-10 codes and a total of 1,553 different individual medical services, from medical tests to operations. Consequently, the results of the exploratory data analysis should not be understood as causal relationships, but rather as statistical correlations (or precedents) between the use of healthcare services and changes in the use of long-term care benefits.

Specifically, we use logistic regressions with regularization to avoid overfitting and to help with variable selection. Three common regularization methods are used: Lasso, Ridge Regression and Elastic Net (Hoerl & Kennard, 1970; Tibshirani, 1996; Zou & Hastie, 2005). Lasso and Elastic Net are particularly suitable as they shrink the parameters of variables with low explanatory value to zero. This leads to a data-driven selection of variables, which improves the interpretability of the results (Tibshirani, 1996).

To select the strength of the regularization, the data sets are randomly divided into training and test data (Hastie et al., 2009). The model is first optimized using the training data (75 % of all observations) using fivefold cross-validation and using the area under the "receiver operating characteristic" (ROC) curve (AUC) as the evaluation metric.[10] Then, the quality of the model is assessed using the test data (25 % of all observations). As these have not been used to estimate and optimize the model ("out-of-sample" prediction), this approach is suitable for determining how well the estimated model can predict events in other, yet "unknown" data.

Throughout our analysis, the "observation" and "event" periods will pivot around January 2018 ($t = 01/2018$). This allows us to consider event periods of up to 12 months (January – December 2018) as well as preceding observation periods of up to 24 months (January 2016–December 2017). The summary statistics in Table 2 therefore show information about personal characteristics in January 2018 as well as aggregated health service consumption in the observation period January 2016–December 2017 for individuals in the case sample who were first-time receivers of LTCA in the event period January–December 2018 and individuals in the control sample. Both groups include only those who were between 60 and 85 years of age to make the "case" and "control" samples comparable in this respect.[11] These observations will be used to classify the first-time receipt of LTCA.

Table 2 shows that first-time receivers in the event period [01/2018, 01/2019) are on average 3.1 years older than those in the control sample. As mentioned above, this is

[10] ROC curves compare the true-positive rate of a classification method (sensitivity) with its false-positive rate (failure rate or 1-specificity). In general, models with the highest possible sensitivity at the lowest possible failure rate are desirable. Such models have an AUC close to 1. Models with an AUC between 0.8 and 0.9 can be said to have "excellent" discriminatory power (see Hosmer & Lemeshow, 2000, p. 162).

[11] Individuals who died before January 2018 are not included.

Table 2: Summary statistics for first-time receivers of LTCA in [01/2018, 01/2019) and control sample

| | Control sample | | Case sample, First-time receivers | | | |
|---|---|---|---|---|---|---|
| | Mean | S. D. | Mean | S. D. | Difference in means | Std. Error |
| Female (=1) | 0.566 | 0.496 | 0.575 | 0.494 | 0.009*** | 0.003 |
| Age | 73.163 | 6.144 | 76.303 | 6.470 | 3.140*** | 0.033 |
| General practitioner contacts | 23.348 | 20.521 | 35.797 | 27.301 | 12.449*** | 0.139 |
| Specialist doctor contacts | 15.059 | 15.052 | 16.893 | 17.637 | 1.834*** | 0.091 |
| Other contractor contacts | 6.651 | 13.526 | 10.689 | 21.822 | 4.038*** | 0.110 |
| Hospital inpatient stays (days) | 2.919 | 8.024 | 9.960 | 18.942 | 7.041*** | 0.095 |
| Observations | 433,448 | | 40,534 | | | |

Note: Event period [01/2018, 01/2019). General practitioner, specialist doctor and other contractor contacts as well as hospital inpatient stays measured during observation period [01/2016, 01/2018). Includes only individuals between 60 and 85 years of age. * significant at 10 %, ** 5 %, *** 1 % level.

to be expected and highlights the importance of controlling for age in the regressions. Furthermore, first-time receivers had significantly more contacts with general practitioners, specialist doctors and other contractual partners in the Austrian social security system in the preceding observation period [01/2016, 01/2018).[12] In addition, they had more inpatient stays in the observation period than those in the control sample. This suggests that data on medical service consumption may substantially improve the prediction of the first-time receipt of LTCA. There is also a small but insubstantial deviation in the percentage of women between the case and control samples.

# 5 Empirical analysis

## 5.1 Logit regressions

To provide a first impression of the correlations between the consumption of health services and the first-time receipt of LTCA, Table 3 shows the results of basic Logit regressions on the training data using observation and event periods of one year: age increases the likelihood of LTCA uptake while women are, given age and region of residence, less likely to start receiving LTCA in the event period.[13] As expected, data on medical service consumption improves the prediction, as highlighted by an increase in the AUC from column (1) to column (2). The number of general practitioner visits and the length of hospital inpatient stays in 2017 are positively correlated with receiving LTCA for the first

[12]For general practitioners, specialist doctors, and other contractors, each visit counts as a "contact". "Other contractors" include providers of medical products such as pharmacies, medical stores, optometrists, etc., but also medical laboratories, medical service providers such as physiotherapists, dental laboratories, orthopedic technicians, etc. Detailed information about the type of contractor will be used in the analysis below.

[13]While the region of residence is not of special interest to us, dummy variables for the NUTS-3 region of residence are included because they may reflect differences between rural and urban regions, for example regarding the supply of health services or income, wealth and educational distributions.

time in 2018. Given the number of general practitioner visits, the number of specialist visits has a negative correlation, while more frequent visits to other contractors in the observation period are again positively correlated with receiving LTCA in the event period.

Table 3: Logit regressions, $y_i$ = first-time receipt of LTC allowance in [01/2018, 01/2019).

| | (1) | (2) |
|---|---|---|
| Female (= 1) | −0.063*** | −0.028** |
| | (0.012) | (0.013) |
| Age | 0.094*** | 0.087*** |
| | (0.001) | (0.001) |
| General practitioner contacts | | 0.031*** |
| | | (0.001) |
| Specialist doctor contacts | | −0.011*** |
| | | (0.001) |
| Other contractor contacts | | 0.011*** |
| | | (0.001) |
| Hospital inpatient stays (days) | | 0.049*** |
| | | (0.001) |
| Constant | −10.145*** | −10.186*** |
| | (0.131) | (0.137) |
| NUTS-3 fixed effects (= 1) | Yes | Yes |
| Observations (training data) | 355,486 | 355,486 |
| Observations (test data) | 118,496 | 118,496 |
| AUC (training data) | 0.651 | 0.733 |
| AUC (test data) | 0.651 | 0.732 |
| Log Likelihood | -100,073.899 | -93,598.770 |

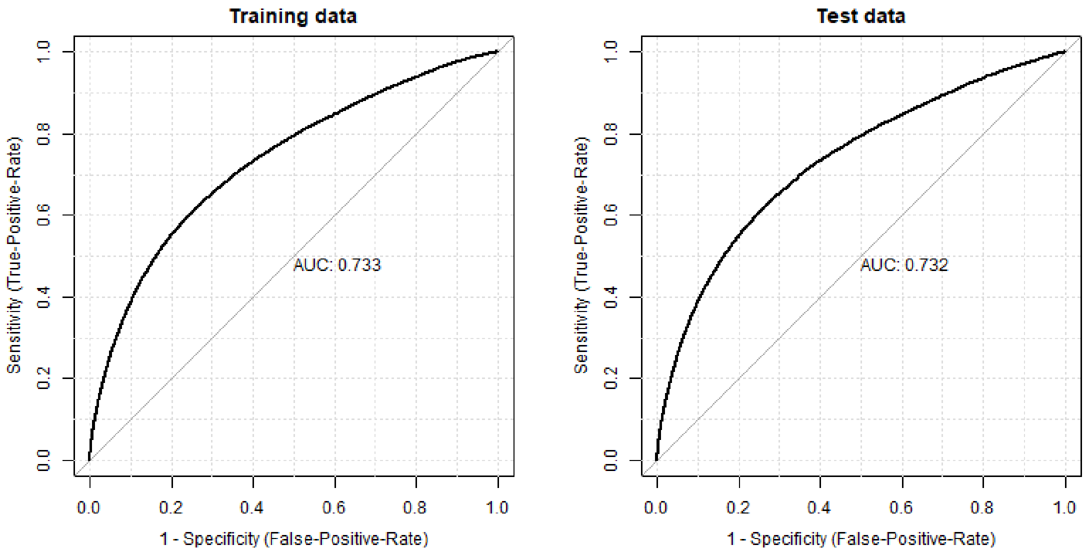
Notes: Robust standard errors in parentheses. Model estimated on training data (75 % random sample of all data). Dependent variable: $y_i = 1$ if person $i$ received LTC allowance for the first time in event period [01/2018, 01/2019), zero else. Observation period: [01/2017, 01/2018). Includes only individuals between 60 and 85 years of age. Coefficients for 34 regional dummies at NUTS-3 level estimated but not shown. AUC: Area under ROC curve. *significant at 10 %, **5 %, ***1 % level.

A comparison of the areas under the ROC curves between the training data (on which the model was estimated) and the test data (the remaining 25 % of all data) shows that there is hardly any difference between the two metrics. This suggests that overfitting is not a problem with such a restricted set of variables. However, it also shows that the predictive power of such a model is relatively limited. For example, as Figure 1 shows, at a sensitivity (true-positive-rate) of 0.8, where 80 % of the true first-time receivers are correctly classified, the false-positive-rate is as high as 51.1 (50.8) % in the test (training) data using model (2) of Table 3: thus, more than half of those who did not receive LTCA in the event period [01/2018, 01/2019) would be incorrectly classified as first-time receivers.[14]

Adding more information regarding the consumption of health services in the observation period improves the predictive power of the model. Table 4 shows the results of a Logit model that includes the variables used to estimate the model in Table 3 but replaces the specialist doctor and other contractor contact variables with 57 variables measuring the frequency of contacts to each of the specific specialist doctors and other contractors (SDOCs) listed in Appendix Table A.2. In addition, we include dummy variables for

[14]The choice of using a sensitivity of 0.8 to evaluate the false-positive-rate is arbitrary, but will be followed throughout.

Figure 1: ROC curves for model (2) in Table 3 evaluated on the training and test data.




Notes: ROC: Receiver operating characteristic; AUC: Area under the ROC curve.

individual medical services (IMS, "Medizinische Einzelleistungen") performed during (inpatient and outpatient) hospital visits listed in the 22 chapters of the catalog of services of the Austrian hospital sector (see BMSGPK, 2020b, and Appendix Table A.1). These dummy variables take on a value of one if a service from the chapter was performed at least once during the observation period.[15] Dummy variables for each of the 22 ICD-10 blocks that equal one if at least one diagnosis from the block was made during these hospital visits (zero else) are also included.[16] Finally, we also add dummy variables for all 14 ATC groups (Appendix Table A.4) to the specification that take on the value one if at least one drug from the respective group was dispensed at least once during the observation period, zero else.

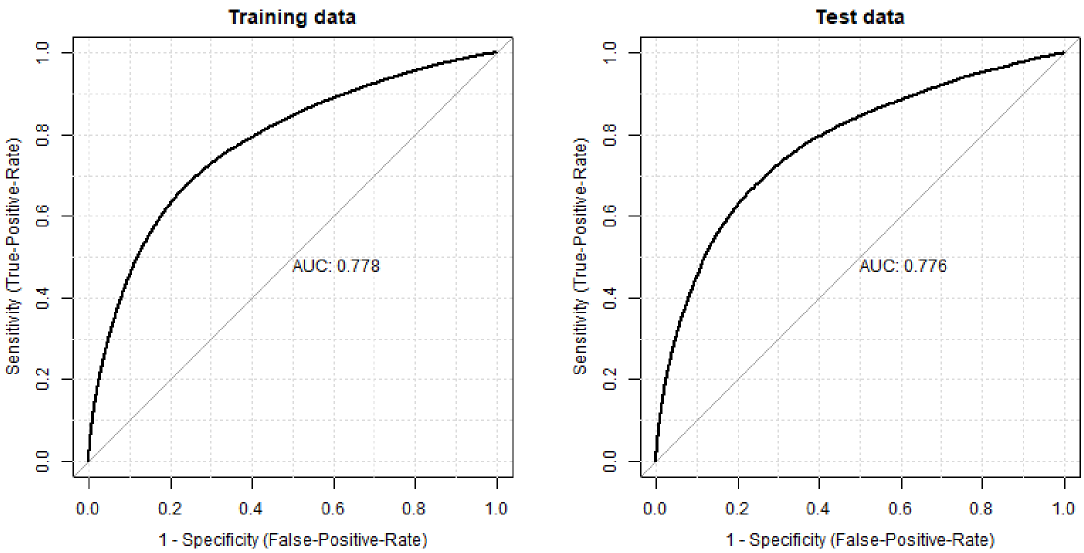
The resulting model has an AUC of 0.776 (0.778) in the test (training) data, and at a sensitivity of 0.8 there is a substantial reduction in the false-positive rate to 40.9 % in the test data (Figure 2). As Table 4 shows, many specialist doctor and other contractor contacts as well as medical services are, ceteris paribus, negatively correlated with the event. Whether this suggests that, all else equal, visiting a specialist can be seen as a preventive measure or whether this is a selection effect (such that, for example, those who

[15] This national catalog is used to document the medical services delivered in the Austrian hospital sector. In 2020, it contained 2,007 different 5-digit codes organized into 22 chapters. For example, one of the most common medical services, "CAT scan of head and neck" (code ZA010), is contained in chapter "Diagnostic and Interventional Imaging" (IMS chapter 12) that, inter alia, also includes MRI and X-ray imaging. Since there are some codes that are not included in the 22 chapters, Table 4 also contains a 23rd dummy variable "IMS chapter NA" for these services.

[16] The ICD blocks "1" and "9" are not officially part of the ICD-10, but Austrian national special codes that represent additional diagnoses, see BMSGPK, 2020a.

are able to visit an ophthalmologist or a dentist are in otherwise better physical condition), cannot be answered with the data and methods at hand. Among the disease diagnoses and drug classifications most statistically significant correlations are, however, positive.

Figure 2: ROC curves for model in Table 4 evaluated on the training and test data.




Notes: AUC: Area under the ROC curve. ROC: Receiver operating characteristic.

To assess the ability of the model to predict the first-time receipt of LTCA over different time horizons, we test four different lengths of the observation period—two years (January 1, 2016 to December 31, 2017), one year (January 1, 2017 to December 31, 2017), half a year (July 1, 2017 to December 31, 2017) and, finally, one quarter (October 1, 2017 to December 31, 2017)—and four different lengths of the event periods: one month (January 1, 2018 to January 31, 2018), one quarter (January 1, 2018 to March 31, 2018), half a year (January 1, 2018 to June 30, 2018) and one year (January 1, 2018 to December 31, 2018).

The results in Table 4 show that the first-time receipt of LTCA can be predicted relatively well in the short term, while models that use longer event periods and/or longer observation periods have lower predictive abilities.[17] For example, as shown in the upper panel of Table 5, a Logit model estimating the probability of receiving care allowance for the first time in January 2018 based on healthcare services received in the third quarter of 2017 achieves an area under the ROC curve of up to 0.883 in the test data. Accordingly, the false-positive-rate also decreases substantially, for example to 18.8 % at a sensitivity (true-positive-rate)

[17]The size of the training and test data vary by the length of the event period because each month new individuals enter the case sample (as first-time receivers of LTCA) while persons who die exit the case and control samples. For example, a person who receives LTCA in February 2018 is not in the case sample for the event period [01/2018, 02/2018) but would be included in the case sample for the event period [01/2018, 04/2018). Similarly, a person from the control sample who died in February 2018 would be included in the control sample for the event period [01/2018, 02/2018), but not for the event period [01/2018, 04/2018).

Table 4: Logit regression, $y_i$ = first-time receipt of LTC allowance in [01/2018, 01/2019).

| Variable | Coef. | S. E. | Variable | Coef. | S. E. |
|---|---|---|---|---|---|
| Female (=1) | 0.001 | (0.015) | IMS chapter 3 | 0.029 | (0.033) |
| Age | 0.092*** | (0.002) | IMS chapter 4 | −0.124* | (0.068) |
| Gen. pract. contacts | 0.015*** | (0.001) | IMS chapter 5 | −0.034 | (0.022) |
| SDOC code 2 | 0.132 | (0.372) | IMS chapter 6 | −0.355*** | (0.137) |
| SDOC code 3 | −0.022*** | (0.005) | IMS chapter 7 | −0.034* | (0.019) |
| SDOC code 4 | −0.015* | (0.009) | IMS chapter 8 | −0.123*** | (0.024) |
| SDOC code 5 | −0.017*** | (0.005) | IMS chapter 9 | −0.099*** | (0.027) |
| SDOC code 6 | −0.126*** | (0.016) | IMS chapter 10 | −0.114*** | (0.024) |
| SDOC code 7 | 0.002 | (0.003) | IMS chapter 11 | −0.921* | (0.509) |
| SDOC code 8 | −0.213 | (0.438) | IMS chapter 12 | 0.073*** | (0.006) |
| SDOC code 9 | −0.055*** | (0.008) | IMS chapter 13 | −0.009 | (0.019) |
| SDOC code 10 | 0.014** | (0.007) | IMS chapter 14 | 0.023 | (0.034) |
| SDOC code 11 | 0.046*** | (0.008) | IMS chapter 15 | 0.016 | (0.014) |
| SDOC code 12 | −0.002 | (0.002) | IMS chapter 16 | −0.073*** | (0.014) |
| SDOC code 13 | −0.010* | (0.005) | IMS chapter 17 | 0.079* | (0.044) |
| SDOC code 14 | −0.113*** | (0.008) | IMS chapter 18 | 0.061*** | (0.010) |
| SDOC code 15 | −0.073*** | (0.027) | IMS chapter 20 | −0.090** | (0.045) |
| SDOC code 16 | −0.039*** | (0.006) | IMS chapter 21 | 0.038*** | (0.006) |
| SDOC code 17 | −0.055*** | (0.004) | IMS chapter 22 | 0.871** | (0.339) |
| SDOC code 18 | 0.010 | (0.033) | IMS chapter NA | −0.064** | (0.032) |
| SDOC code 19 | 0.081*** | (0.007) | ICD block 1 | 0.279* | (0.157) |
| SDOC code 20 | 0.043*** | (0.010) | ICD block 9 | −0.041 | (0.073) |
| SDOC code 21 | −0.071 | (0.120) | ICD block A | 0.113* | (0.065) |
| SDOC code 24 | −1.120 | (0.986) | ICD block B | −0.025 | (0.073) |
| SDOC code 27 | −0.047*** | (0.006) | ICD block C | 0.725*** | (0.035) |
| SDOC code 29 | −0.094 | (0.078) | ICD block D | 0.180*** | (0.036) |
| SDOC code 40 | 0.017 | (0.018) | ICD block E | 0.078*** | (0.029) |
| SDOC code 42 | 0.003 | (0.002) | ICD block F | 0.555*** | (0.040) |
| SDOC code 43 | 0.078 | (0.253) | ICD block G | 0.119*** | (0.036) |
| SDOC code 50 | −0.009** | (0.004) | ICD block H | −0.031 | (0.035) |
| SDOC code 52 | −0.184** | (0.090) | ICD block I | 0.043 | (0.026) |
| SDOC code 53 | −0.170*** | (0.021) | ICD block J | 0.285*** | (0.035) |
| SDOC code 55 | −0.045 | (0.055) | ICD block K | −0.103*** | (0.031) |
| SDOC code 61 | 0.074*** | (0.005) | ICD block L | 0.293*** | (0.061) |
| SDOC code 62 | −0.046 | (0.123) | ICD block M | −0.278*** | (0.030) |
| SDOC code 63 | 0.008* | (0.004) | ICD block N | 0.119*** | (0.032) |
| SDOC code 64 | 0.017 | (0.036) | ICD block Q | −0.065 | (0.115) |
| SDOC code 65 | 0.015*** | (0.001) | ICD block R | 0.119*** | (0.032) |
| SDOC code 66 | 0.021 | (0.017) | ICD block S | 0.398*** | (0.075) |
| SDOC code 67 | −0.009 | (0.043) | ICD block T | 0.033 | (0.055) |
| SDOC code 68 | −0.037 | (0.028) | ICD block U | 0.218 | (1.379) |
| SDOC code 69 | 0.081*** | (0.014) | ICD block Z | −0.025 | (0.037) |
| SDOC code 72 | 0.034 | (0.031) | ATC group A | 0.012*** | (0.001) |
| SDOC code 73 | 0.485*** | (0.090) | ATC group B | 0.013*** | (0.002) |
| SDOC code 75 | 0.021 | (0.015) | ATC group C | −0.000 | (0.001) |
| SDOC code 80 | −0.007 | (0.007) | ATC group D | −0.001 | (0.004) |
| SDOC code 81 | 0.123*** | (0.027) | ATC group G | −0.001 | (0.002) |
| SDOC code 84 | 0.100*** | (0.011) | ATC group H | 0.009** | (0.004) |
| SDOC code 85 | −0.007*** | (0.002) | ATC group J | 0.001 | (0.005) |
| SDOC code 86 | −0.002 | (0.008) | ATC group L | 0.031*** | (0.003) |
| SDOC code 87 | 0.064*** | (0.020) | ATC group M | 0.016*** | (0.002) |
| SDOC code 91 | −0.008*** | (0.002) | ATC group N | 0.026*** | (0.001) |
| SDOC code 92 | 0.012 | (0.011) | ATC group P | −0.079 | (0.055) |
| SDOC code 99 | 0.020** | (0.008) | ATC group R | 0.023*** | (0.001) |
| Hosp. inp. stays (days) | 0.024*** | (0.002) | ATC group S | 0.000 | (0.002) |
| IMS chapter 1 | −0.025 | (0.024) | ATC group V | 0.041** | (0.016) |
| IMS chapter 2 | 0.046*** | (0.012) | Constant | −10.548*** | (0.141) |
| NUTS-3 fixed effects (= 1) | Yes | | AUC (training data) | 0.778 | |
| Observations (training data) | 355,486 | | AUC (test data) | 0.776 | |
| Observations (test data) | 118,496 | | Log Likelihood | -89,060.100 | |

Notes: Robust standard errors in parentheses. Model estimated on training data (75 % random sample of all data). Dependent variable: $y_i = 1$ if person $i$ received LTC allowance for the first time in event period [01/2018, 01/2019), zero else. Observation period: [01/2017, 01/2018). Includes only individuals between 60 and 85 years of age. Coefficients for 34 regional dummies at NUTS-3 level estimated but not shown. SDOC: specialist doctor or other contractor, IMS: individual medical service, ATC: Anatomical-Theurapeutic-Chemical Classification, ICD: International Classification of Diseases, AUC: Area under ROC curve. *significant at 10 %, **5 %, ***1 % level.

of 80 %. The importance of short-run information becomes even more apparent if the consumption of health services in the month preceding the event period (i. e., December 2017) is excluded from the analysis (lower panel of Table 5), which causes a substantial decline in our evaluation metric. And while the AUC drops from 0.883 points to 0.847 (−0.36 points) with an extended observation period of two years, it hardly changes with the length of observation period if the month prior to the event period is omitted, which shows that short-run information is crucial.

This suggests that an attempt to predict first-time LTCA receivers is most promising if it 1) uses short-term data on the consumption of health services, while healthcare information that dates further back is less relevant, and 2) tries to predict the event under investigation within a relatively short time frame of, for example, one month. For this reason, the remaining discussion will focus on predicting the first-time receipt of LTCA in the event period $[01/2018, 02/2018)$ based on the observation period $[10/2017, 01/2018)$.

Table 5: AUC based on the test data for models with different observation and event periods.

| Observation period | Event periods<br>[01/2018, 02/2018) | <br>[01/2018, 04/2018) | <br>[01/2018, 07/2018) | <br>[01/2018, 01/2019) |
|---|---|---|---|---|
| $[10/2017, 01/2018)$ | 0.883 | 0.830 | 0.799 | 0.772 |
| $[07/2017, 01/2018)$ | 0.877 | 0.829 | 0.800 | 0.775 |
| $[01/2017, 01/2018)$ | 0.862 | 0.822 | 0.797 | 0.776 |
| $[01/2016, 01/2018)$ | 0.847 | 0.812 | 0.791 | 0.776 |
| | Event periods<br>[01/2018, 02/2018) | <br>[01/2018, 04/2018) | <br>[01/2018, 07/2018) | <br>[01/2018, 01/2019) |
| $[10/2017, 12/2017)$ | 0.818 | 0.787 | 0.773 | 0.754 |
| $[07/2017, 12/2017)$ | 0.817 | 0.791 | 0.777 | 0.760 |
| $[01/2017, 12/2017)$ | 0.806 | 0.787 | 0.777 | 0.763 |
| $[01/2016, 12/2017)$ | 0.797 | 0.782 | 0.774 | 0.765 |
| Frequency of $y_i = 1$ (training data) | 0.0074 | 0.026 | 0.051 | 0.086 |
| Frequency of $y_i = 1$ (test data) | 0.0074 | 0.026 | 0.051 | 0.085 |
| Observations (training data) | 349,012 | 348,996 | 348,977 | 355,486 |
| Observations (test data) | 116,337 | 116,332 | 116,326 | 118,496 |

Notes: Model estimated on training data (75 % of all data) using specification from Table 4. Lower panel shows estimations excluding the month immediately prior to the event period (December 2017). AUC: Area under the ROC curve. ROC: Receiver operating characteristic.

## 5.2 Statistical learning models

To harness the full potential of our data we can use the most detailed levels of the health consumption variables available: instead of 22 IMS chapters, we can use the 1,553 IMS codes for all individual medical services in our data, instead of 22 ICD-10 blocks we can use 1,483 3-digit ICD-10 codes and instead of 14 ATC groups we can use 483 Level 4

ATC codes. Using such high-dimensional data, however, increases the risk of "overfitting" the model to the training data based on the particular idiosyncrasies of this dataset. An overfitted model may then perform worse on the test data than a model that is slightly less-well fitted to the training data ("bias–variance trade-off", see Hastie et al., 2009, p. 223).

This also appears to be the case in our data: a Logit model fitted on the training data that includes the detailed health consumption information outlined above reaches an AUC of 0.909 (Table 6). When applied to the test data, however, the AUC drops considerably to 0.745 with a corresponding failure rate (at a sensitivity of 80 %) of almost 60 % false-positives. The out-of-sample performance of the model using the fully detailed health data is thus worse than the performance of the model in Table 4 although the latter is based on fewer details.[18]

This calls for a supervised learning approach that takes this trade-off into account. We therefore extend the logistic regression applied above by introducing regularization methods. These methods add a "penalty term" to the log-likelihood function maximized in logistic regression that penalizes either the sum of the squared coefficients (Ridge regression) or the sum of the absolute values of the coefficients (Lasso) in a model. More specifically, the regularized log-likelihood of the intercept $\beta_0$ and the other estimated parameters $\beta$ given the dependent variable $y$ and the $k = \{1, \ldots, K\}$ explanatory variables $X$ is given by (Hastie et al., 2009, p. 125):

$$LL^r(\beta_0, \beta | y, X) = \sum_{i=1}^{N} \left\{ y_i \left( X_i^T \beta + \beta_0 \right) - \log \left[ 1 + \exp \left( X_i^T \beta + \beta_0 \right) \right] \right\} - \lambda \ell_j, \tag{1}$$

where $\ell_j$ is the $\ell_1$ norm in case of the Lasso or the $\ell_2$ norm in case of the Ridge regression over all $K$ coefficients of the explanatory variables (except for the intercept $\beta_0$):

$$\begin{aligned} \ell_1 &= \|\beta\|_1 = \sum_{k=1}^{K} |\beta_k| \\ \ell_2 &= \|\beta\|_2^2 = \sum_{k=1}^{K} \beta_k^2. \end{aligned} \tag{2}$$

The hyperparameter $\lambda$ determines the strength of the penalty. It is not estimated with the Logit model, but optimized using fivefold cross-validation on the training data. Given the

[18] The number of active explanatory variables used in the estimation is 2,651, which is lower than the theoretically possible 3,611 explanatory variables. The reason for this is that many individual medical services, ICD-10 3-digit diagnoses and Level 4 ATC drug codes were not observed in the training data during the observation period. Furthermore, only variables with non-zero variance among both first-time recipients and non-recipients can be included in the model. For example, if a specific medical service was provided to only one person in the observation period and that person later received LTCA for the first time in the event period, this medical service would perfectly predict the dependent variable (because $y_i$ is always 1 if this service is provided), causing complete separation and potential problems for regression parameters and convergence (Mansournia et al., 2018).

penalty term $-\lambda \ell_j$, maximizing the log-likelihood may involve a "shrinkage" of (a large number of) large coefficients that reflect idiosyncrasies of the training data towards zero. This introduces a bias in the estimated coefficients, but may reduce the variance of the model predictions when applied to the test data (James et al., 2013, p. 214). Since Lasso can shrink the coefficients of unimportant variables to exactly zero, it can also be seen as a variable selection tool and is thus especially well suited to our analysis because it increases the interpretability of the classification model. Finally, we also apply an Elastic Net model where the penalty term is given by:

$$-\lambda \left[\alpha \ell_1 + (1-\alpha)\ell_2\right]$$

with $\alpha \in [0,1]$. The Elastic Net is thus a mixture of the Lasso and Ridge regression, and we set the combination ratio to $\alpha = 0.5$.

Table 6: Logit, Lasso, Ridge regression, and Elastic Net models, $y_i$ = first-time receipt of LTCA in [01/2018, 02/2018), observation period [10/2017, 01/2018).

| Model | Logit | | Lasso | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Data | Training | Test | Training | Test |
| Frequency of $y_i = 1$ | 0.0074 | 0.0074 | 0.0074 | 0.0074 |
| AUC | 0.909 | 0.745 | 0.891 | 0.888 |
| Failure rate (sensitivity = 0.8) | 0.128 | 0.594 | 0.154 | 0.164 |
| Active variables | 2,651 | 2,651 | 2,651 | 461 |
| Observations | 349,223 | 116,126 | 349,223 | 116,126 |

| Model | Ridge Regression | | Elastic Net | |
|---|---|---|---|---|
| | (5) | (6) | (7) | (8) |
| Data | Training | Test | Training | Test |
| Frequency of $y_i = 1$ | 0.0074 | 0.0074 | 0.0074 | 0.0074 |
| AUC | 0.896 | 0.889 | 0.894 | 0.889 |
| Failure rate (sensitivity = 0.8) | 0.173 | 0.185 | 0.151 | 0.161 |
| Active variables | 2,651 | 2,651 | 2,651 | 513 |
| Observations | 349,223 | 116,126 | 349,223 | 116,126 |

Notes: Model estimated on training data (75 % of all data). Regularization strength $\lambda$ optimized using fivefold cross-validation. Dependent variable: $y_i = 1$ if person $i$ received LTCA for the first time in event period [01/2018, 02/2018), zero else. Observation period: [10/2017, 01/2018). Includes only individuals between 60 and 85 years of age. AUC: Area under ROC curve. ROC: Receiver operating characteristic.

Columns (3)–(8) of Table 6 show the results of the three regularization models applied to both the training and test data. All three regularization methods yield relatively similar AUC values when evaluated on the test data (0.888–0.889) and perform substantially better on this metric than unregularized Logit regression (0.745). In addition, the Lasso and Elastic Net regularizations have a lower failure (false-positive) rate than Ridge Regression at a sensitivity of 0.8. Compared to the simpler model estimated using the same observation and event period, but less detailed information shown in Table 5, the gain from using the disaggregated drug, diagnosis and individual medical service variables is, however, marginal: AUC evaluated on the test data rises from 0.883 to, at best, a maximum of 0.889, calling into question the usefulness of the more detailed data. Furthermore, the coefficients of 461 (513) variables are shrunk to zero in the Lasso (Elastic Net) regressions. These results are also robust to a shift in the event period.[19]

As a final check of the predictive power of our approach the data and models in Table 6 are expanded with information on the consumption of health services not only in the quarter before the event period, but also during three lagged observation periods: (i) [07/2017, 10/2017), (ii) [01/2017, 07/2017) and (iii) [01/2016, 01/2017). These additional variables will allow us to check whether medical procedures performed longer ago can improve the predictive power of our statistical learning models. In sum, 10,810 predictor variables will be used in the analysis, and it can be expected that models without regularization will suffer from serious overfitting given such a large number of explanatory variables.

Table 7: Logit, Lasso, Ridge regression, and Elastic Net models, $y_i$ = first-time receipt of LTCA in [01/2018, 02/2018), observation periods [10/2017, 01/2018), [01/2017, 07/2017) and [01/2016, 01/2017).

| Model | Logit | | Lasso | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Data | Training | Test | Training | Test |
| Frequency of $y_i = 1$ | 0.0074 | 0.0074 | 0.0074 | 0.0074 |
| AUC | 0.960 | 0.611 | 0.895 | 0.889 |
| Failure rate (sensitivity = 0.8) | 0.032 | 0.891 | 0.142 | 0.162 |
| Active variables | 10,810 | 10,810 | 10,810 | 535 |
| Observations | 349,223 | 116,126 | 349,223 | 116,126 |

| Model | Ridge Regression | | Elastic Net | |
|---|---|---|---|---|
| | (5) | (6) | (7) | (8) |
| Data | Training | Test | Training | Test |
| Frequency of $y_i = 1$ | 0.0074 | 0.0074 | 0.0074 | 0.0074 |
| AUC | 0.912 | 0.881 | 0.898 | 0.891 |
| Failure rate (sensitivity = 0.8) | 0.132 | 0.202 | 0.139 | 0.159 |
| Active variables | 10,810 | 10,810 | 10,810 | 632 |
| Observations | 349,223 | 116,126 | 349,223 | 116,126 |

Notes: Model estimated on training data (75 % of all data). Regularization strength $\lambda$ optimized using fivefold cross-validation. Dependent variable: $y_i = 1$ if person $i$ received LTCA for the first time in event period [01/2018, 02/2018), zero else. Observation periods: [10/2018, 01/2018), [07/2017, 10/2017), [01/2017, 07/2017) and [01/2016, 01/2017). Includes only individuals between 60 and 85 years of age. AUC: Area under ROC curve. ROC: Receiver operating characteristic.

Indeed, although the AUC of the Logit model without regularization increases considerably in the training data (from 0.909 to 0.960), in the test data it is even lower than before (0.611) with a failure rate of almost 90 % at a sensitivity of 0.8 (Table 7), supporting the choice of using regularization for classification. But even for the Logit models with regularization, our main classification metric improves only marginally in the test data. The best performing model is the Elastic Net, with an AUC in the test data of 0.891 in Table 7, a modest increase compared to the model that uses only the most recent health information (Table 6). This is also highlighted by the relatively small increase in the

[19]Results are available from the authors upon request.

number of active variables (i. e., variables with non-zero coefficients) which rises from 513 to only 632.

This supports the conclusion that the recent health information is most important for classifying the first time receipt of LTCA. More specifically, among the 632 active variables in the best performing model (Elastic Net), 61.6 % (389) represent personal characteristics (such as age) or health services consumed in the quarter immediately preceding the event period. Only 16.6 % (105) are health services consumed in the third quarter of 2017, 13.1 % (83) are health services observed in the first half of 2017 and 8.7 % (55) represent health services consumed in 2016.

### 5.3 Most important predictors

As mentioned above, an attractive feature of the Lasso and Elastic Net regressions is that they shrink the parameters of less relevant variables to zero. The remaining "active" variables in the (best-performing) Elastic Net model can therefore deliver insights into the most important short-term predictors of the first-time receipt of LTCA.[20] Among the personal characteristics, only age remains as an active variable. The coefficients of sex or region of residence are shrunk to zero, and thus have no additional predictive power given the other variables included in the model. All other active characteristics, including inpatient hospital stays and general practitioner visits, represent health variables. The majority of the active variables (97.3 %) indicate individual medical services performed (279 active variables, 44.1 % of all active variables) and diagnoses made during hospital visits (266 variables, 42.1 %) or ATC codes (70 variables, 11.1 %). This is, however, not surprising given that IMS as well as ICD and ATC codes account for 97.7 % of all variables considered in the model.

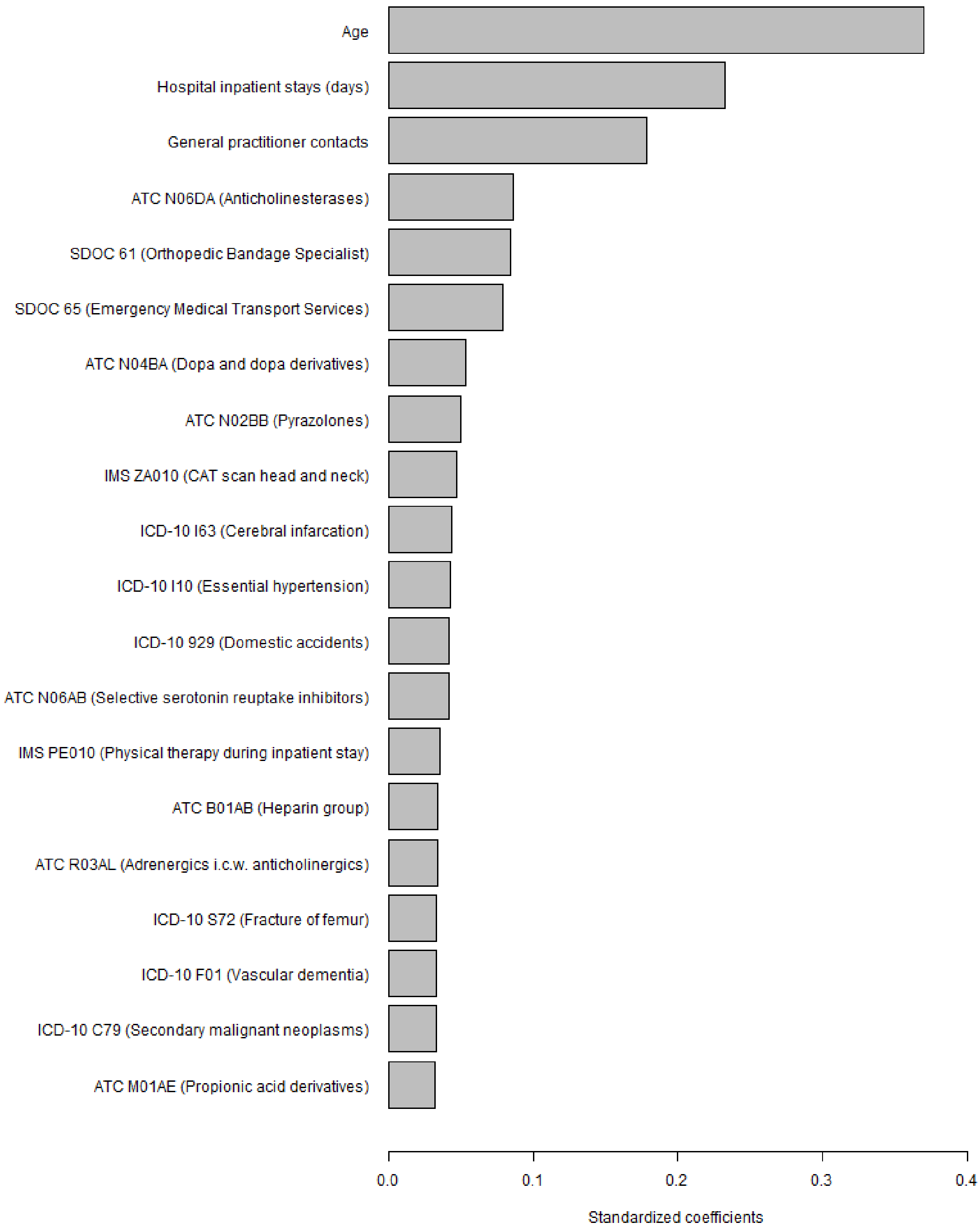
To help assess the relative importance of the various factors, Figure 3 lists the 20 largest standardized coefficients of the Elastic Net model of Table 7. These "beta" coefficients measure the (standard deviation) change in the dependent variable following a one standard deviation change in the explanatory variables. Standardization of the coefficients not only eliminates differences in the scale of the explanatory variables, thus making the effects of variables with different units of measurement comparable, it also relates the estimated effects to the actual variation in the explanatory variables in the sample: variables with large coefficients but small variation between observations therefore have smaller standardized coefficients than otherwise similar variables with larger variation between individuals.

All variables related to the consumption of health services in Figure 3 are observed in the quarter before the event period.[21] The most important predictor variables for the first-

[20] Again, one has to bear in mind that these factors are not to be interpreted causally. In addition, the $\ell_1$ norm also tends to shrink the coefficient of collinear variables to zero. Therefore, some predictors of LTCA receipt may be missing from the final set of "active" variables if they are highly correlated with other active variables.

[21] Among the 50 variables related to the consumption of health services with the largest standardized coefficients, 88 % are observed in the period [10/2017, 01/2018).

Figure 3: 20 variables with largest standardized coefficients of the Elastic Net model of Table 7.




Notes: All listed variables related to the consumption of health services observed in observation period [10/2017, 01/2018). SDOC: specialist doctor or other contractor, IMS: individual medical service, ATC: Anatomical-Theurapeutic-Chemical Classification, ICD: International Classification of Diseases.

time receipt of LTCA in the short term are age, inpatient hospital stays (in days) and the number of contacts to general practitioners: a one-standard-deviation increase in these variables raises the dependent variable by 0.370, 0.233 and 0.179 standard deviations, respectively. There is also a noticeable accumulation of drugs from ATC group "N" among the most important predictors. These drugs are related to the nervous system and include anti-dementia drugs (N06DA), anti-Parkinson drugs (N04BA), antidepressants (N06AB) as well as Pyrazolones (N02BB) that belong to the subgroup of analgesics (N02). Computer tomography scans of the head and neck or diagnoses such as "cerebral infarction", "essential hypertension"[22] or domestic accidents also show a clear precedence for the first-time receipt of LTCA in the short term. The relative importance of orthopedic bandage specialists and emergency medical transport services may mask some of the underlying medical reasons for visiting orthopedic bandage specialists (e. g. the diagnosis on which the prescription of an orthopedic bandages was based on) or the medical emergency that triggered an ambulance ride. Again, these variables cannot be interpreted as suggesting causality (e. g., more frequent doctor visits or ambulance rides in one period causally increasing the probability of receiving LTCA in a later period), but rather as a precedence: all else equal, we observe more frequent doctor visits and ambulance rides in the months preceding the first-time receipt of long-term care allowance, so that these variables ceteris paribus imply that the person is more likely to become a receiver of LTCA in the immediate future.

# 6 Discussion

Our empirical analysis provides a comprehensive assessment of the predictors for first-time receipt of the Austrian long-term care allowance (LTCA) using rich administrative data and statistical learning methods. Several key findings emerge from the results.

The analysis demonstrates that short-term healthcare data are particularly powerful for predicting LTCA entry. Models that use health service consumption in the quarter immediately preceding the event period achieve the highest predictive accuracy, while extending the observation window back in time diminishes predictive performance. This highlights the acute nature of health shocks or rapid declines that often trigger LTCA applications, emphasizing the value of timely data for early identification of individuals at risk.

Expanding the set of predictors to include detailed information on specialist visits, individual medical services, diagnoses (ICD-10 codes), and drug dispensings (ATC codes) improves model performance only marginally. While high-dimensional models can capture more nuanced risk profiles, the risk of overfitting increases, as evidenced by the drop in out-of-sample performance for the most granular models. The use of regularization techniques such as Lasso, Ridge, and Elastic Net mitigates this risk, maintaining strong predictive power while improving model interpretability by selecting the most relevant variables.

[22] High blood pressure that is not associated with any other disease.

Age stands out as the most significant predictor for entering LTCA, which aligns with the expectation that increasing age is closely associated with declining health and functional capacity. In addition to age, healthcare utilization patterns—notably, frequent contacts with general practitioners, higher numbers of hospital inpatient days, and increased use of other medical services—are strongly correlated with subsequent LTCA receipt. These findings suggest that deteriorating health, as reflected in intensified healthcare interactions, often precedes the transition into the long-term care allowance system. Among medical predictors, dementia-related diagnoses and dispensings of neuroactive medications (ATC group N) are especially prominent, corroborating the established link between cognitive decline and the need for care.

Other strong predictors include diagnoses such as cerebral infarction and essential hypertension, as well as certain medical procedures like computer tomography scans. Interestingly, some elective procedures (e. g., joint replacements) are negatively correlated with LTCA entry, possibly reflecting that such interventions are more likely among healthier individuals.

Can the model results be used to develop a short-term leading indicator for future receipt of LTCA at the individual level? We argue that this remains methodologically challenging, as the events analyzed are not evenly distributed in the available data and the low baseline incidence of LTCA entry. Between 2016 and 2018, an average of just over 3,600 first-time recipients of long-term care benefits per month were observed in the age group 60 to 85, which comprises about 1.9 million people. Even if those already receiving long-term care benefits are excluded from this figure, there would still be a high number of false positive classifications even with a very low relative error rate, i. e., the model would incorrectly assign too many people to the group of first-time long-term care allowance beneficiaries. Despite the already good predictive performance of the methods used here, the development of a leading indicator at the individual level therefore seems to require further research to evaluate alternative advanced classification methods. However, aggregating individual-level predictions offers significant value for more detailed systemic capacity planning, for example at the regional level.

# 7 Conclusions

We address the pressing question of how well long-term care allowance (LTCA) uptake can be predicted using recent healthcare utilization data. We use Austrian administrative data to systematically investigate the empirical relationship between individuals' health service consumption and their subsequent entry into the LTCA system. This question is of growing importance given the rapidly increasing demands on long-term care systems in aging societies, as well as the need for effective, data-driven policy responses to ensure sustainability and equity in care provision.

The empirical analysis confirms that LTCA entry can be effectively anticipated using recent patterns in healthcare utilization. Age is the most influential predictor, reflecting the cumulative effects of functional and cognitive decline that often accompany advanced years. Hospitalizations, frequent general practitioner (GP) contacts, and dispensings for neuroactive medications also emerge as strong short-term predictors, highlighting the role of acute health events and deteriorating health status in triggering care needs. The predictive power of our models is highest when relying on healthcare data from the months immediately preceding LTCA uptake, underscoring the importance of timely information for early intervention. Our findings are consistent with previous research indicating the critical role of healthcare interactions in transitions to formal care systems, and they provide a foundation for data-driven improvements in policy and practice.

While regularization techniques such as Lasso, Elastic Net, and Ridge regression help manage the complexity of high-dimensional data and prevent overfitting, the challenge of developing a reliable individual-level early warning indicator remains. This is due to the relatively low incidence of new LTCA recipients and the risk of false positives. Nevertheless, the findings provide actionable insights for policymakers. Integrating health and long-term care data can help identify high-risk groups, inform resource allocation, and support the design of targeted preventive interventions.

The results have several important policy implications. First, they underscore the need for age-sensitive policies that address not only chronological age but also the associated declines in function and cognition. Preventive geriatric assessments and home adaptation programs could help delay institutionalization and improve quality of life for older adults. The strong association between dementia diagnoses and LTCA entry highlights the urgent need to expand community-based dementia care and early intervention strategies focused on cognitive health. Frequent GP visits are a clear signal of escalating care needs, positioning primary care providers as key actors in the early identification and management of conditions that may lead to long-term care dependency. Strengthening the role of GPs through structured care pathways and incentives for proactive frailty management could have substantial preventive impact.

Despite the robustness of our findings, the study is limited by the lack of socio-economic and psychological variables in the administrative data. These factors—such as wealth, (pension) income, family status, educational background, and psychological aspects—are known to affect both care needs and access (Casanova et al., 2023; Hoang et al., 2023; McMaughan et al., 2020; Muir, 2017; OECD, 2024). Their exclusion thus constrains the depth of our analysis. However, since a merged dataset encompassing all of these characteristics is not currently available for Austria, integrating data on long-term care and health services remains the most feasible option to show the health risks that precede the uptake of LTCA.

This paper contributes to the literature by highlighting the complex interplay of medical, demographic, and systemic factors in determining long-term care needs. The results advocate for a multifaceted approach to policy formulation, emphasizing preventive healthcare, regional equity, and the integration of medical and long-term care strategies. Future research should continue to refine predictive models and explore innovative solutions to meet the evolving challenges of long-term care provision in an aging society. For instance, one could explore hybrid approaches combining these predictive models with qualitative assessments of social support networks or caregiver availability. In addition, future research could include information about the level of LTCA as an outcome variable or the health reasons for transitions from one level to another, eventually higher level among those who receive LTCA.

## References


Bakx, P., Wouterse, B., Van Doorslaer, E., & Wong, A. (2020). Better Off at Home? Effects of Nursing Home Eligibility on Costs, Hospitalizations and Survival. *Journal of Health Economics*, *73*, 102354. https://doi.org/10.1016/j.jhealeco.2020.102354

Billings, J. (2013). The INTERLINKS framework for long-term care of older people in Europe. *Journal of Integrated Care*, *21*(3), 126–138. https://doi.org/10.1108/JICA-02-2013-0007

BMSGPK. (2020a). *ICD-10 BMASGK 2020 – Systematisches Verzeichnis*. Austrian Federal Ministry for Social Affairs, Health, Care and Consumer Protection.

BMSGPK. (2020b). *Leistungskatalog BMASGK 2020*. Austrian Federal Ministry for Social Affairs, Health, Care and Consumer Protection.

BMSGPK. (2024). *Österreichischer Pflegevorsorgebericht 2023*. https://www.sozialministerium.at/dam/jcr:9023f2c0-047c-41a8-84c6-498004be4646/Pflegevorsorgebericht2023.pdf

BMSGPK. (2025). *Höhe des Pflegegeldes* [accessed on 2025-06-12]. https://www.oesterreich.gv.at/themen/pflege/4/Seite.360516.html

Bu, F., & Rutherford, A. (2019). Dementia, home care and institutionalisation from hospitals in older people. *European Journal of Ageing*, *16*(3), 283–291. https://doi.org/10.1007/s10433-018-0493-0

Casanova, G., Martarelli, R., Belletti, F., Moreno-Castro, C., & Lamura, G. (2023). The Impact of Long-Term Care Needs on the Socioeconomic Deprivation of Older People and Their Families: Results from Mixed-Methods Scoping Review. *Healthcare*, *11*(18), 2593. https://doi.org/10.3390/healthcare11182593

Costa-Font, J., Jimenez-Martin, S., & Vilaplana, C. (2018). Does Long-Term Care Subsidization Reduce Hospital Admissions and Utilization? *Journal of Health Economics*, *58*, 43–66. https://doi.org/10.1016/j.jhealeco.2018.01.002

Da Roit, B., & Le Bihan, B. (2010). Similar and Yet So Different: Cash-for-Care in Six European Countries' Long-Term Care Policies. *The Milbank Quarterly*, *88*(3), 286–309. https://doi.org/10.1111/j.1468-0009.2010.00601.x

Drzazga, B. (2021). Identification of Long-term Care Challenges in Europe. *Social Entrepreneurship Review*, *2*, 7–22. https://doi.org/10.15678/SER.2021.2.01

European Commission, Directorate General for Economic and Financial Affairs. (2024). *2024 ageing report: Economic & budgetary projections for the EU Member States (2022 2070)*. Publications Office. https://data.europa.eu/doi/10.2765/022983

Famira-Mühlberger, U. (2024). Projektionen des öffentlichen Pflegeaufwands bis 2050. *WIFO Studies*.

Famira-Mühlberger, U., Horvath, T., Leoni, T., Spielauer, M., Szenkurök, V., & Warum, P. (2025). Demographic Change and the Future of Austria's Long-Term Care Allowance: A Dynamic Microsimulation Study. *WIFO Working Papers*, (705).

Floridi, G., Carrino, L., & Glaser, K. (2021). Socioeconomic Inequalities in Home-Care Use Across Regional Long-Term Care Systems in Europe. *The Journals of Gerontology: Series B*, *76*(1), 121–132. https://doi.org/10.1093/geronb/gbaa139

Frangos, E., Barriguete-Mélendez, A., Debré, P., Gutiérrez Robledo, L. M., Parodi, A.-L., & Michel, J.-P. (2023). Life Course Education, Health, and Ageing Well: A Short Inter-Academic Report. *Gerontology*, *69*(7), 799–808. https://doi.org/10.1159/000529869

Frimmel, W., Halla, M., Paetzold, J., & Schmieder, J. (2023). Health of Parents, Their Children's Labor Supply, and the Role of Migrant Care Workers. *Journal of Labor Economics*, 729102. https://doi.org/10.1086/729102

Grigoriev, O., & Doblhammer, G. (2019). Changing educational gradient in long-term care-free life expectancy among German men, 1997-2012 (Y. D. Kwon, Ed.). *PLOS ONE*, *14*(9), e0222842. https://doi.org/10.1371/journal.pone.0222842

Guo, J., Konetzka, R. T., & Manning, W. G. (2015). The Causal Effects of Home Care Use on Institutional Long-Term Care Utilization and Expenditures. *Health Economics*, *24*(S1), 4–17. https://doi.org/10.1002/hec.3155

Han, E., Quek, R. Y. C., Tan, S. M., Singh, S. R., Shiraz, F., Gea-Sánchez, M., & Legido-Quigley, H. (2019). The role of community-based nursing interventions in improving outcomes for individuals with cardiovascular disease: A systematic review. *International Journal of Nursing Studies*, *100*, 103415. https://doi.org/10.1016/j.ijnurstu.2019.103415

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer New York. https://doi.org/10.1007/978-0-387-84858-7

Hoang, M. T., Kåreholt, I., Schön, P., Von Koch, L., Xu, H., Tan, E. C., Johnell, K., Eriksdotter, M., & Garcia-Ptacek, S. (2023). The Impact of Educational Attainment and Income on Long-Term Care for Persons with Alzheimer's Disease and Other Dementias: A Swedish Nationwide Study. *Journal of Alzheimer's Disease*, *96*(2), 789–800. https://doi.org/10.3233/JAD-230388

Hoerl, A. E., & Kennard, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, *12*(1), 55–67. https://doi.org/10.1080/00401706.1970.10488634

Horn, V., Schweppe, C., Böcker, A., & Bruquetas Callejo, M. (2019). Live-in migrant care worker arrangements in Germany and the Netherlands. Motivations and justifications in family decision-making. *International Journal of Ageing and Later Life*, *13*(2), 83–113. https://doi.org/10.3384/ijal.1652-8670.18410

Hosmer, D. W. J., & Lemeshow, S. (2000). *Applied logistic regression*. John Wiley & Sons.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning* (7th ed.). Springer.

Kim, H. B., & Lim, W. (2015). Long-term care insurance, informal care, and medical expenditures. *Journal of Public Economics*, *125*, 128–142. https://doi.org/10.1016/j.jpubeco.2014.12.004

Larsson, K., Kåreholt, I., & Thorslund, M. (2014). Care utilisation in the last years of life in Sweden: The effects of gender and marital status differ by type of care. *European Journal of Ageing*, *11*(4), 349–359. https://doi.org/10.1007/s10433-014-0320-1

Leiblfinger, M., Prieler, V., Rogoz, M., & Sekulová, M. (2021). Confronted with COVID-19: Migrant live-in care during the pandemic. *Global Social Policy*, *21*(3), 490–507. https://doi.org/10.1177/14680181211008340

Leichsenring, K., Billings, J., Nies, H., & Kümpers, S. (2013). Integrating long-term care in Europe – Improving policy and practice. *International Journal of Integrated Care*, *13*(5). https://doi.org/https://doi.org/10.5334/ijic.1249

Liu, N., Zeng, L., Li, Z., & Wang, J. (2013). Health-related quality of life and long-term care needs among elderly individuals living alone: A cross-sectional study in rural areas of Shaanxi Province, China. *BMC Public Health*, *13*(1), 313. https://doi.org/10.1186/1471-2458-13-313

Løken, K. V., Lundberg, S., & Riise, J. (2017). Lifting the Burden: Formal Care of the Elderly and Labor Supply of Adult Children. *Journal of Human Resources*, *52*(1), 247–271. https://doi.org/10.3368/jhr.52.1.0614-6447R1

Mansournia, M. A., Geroldinger, A., Greenland, S., & Heinze, G. (2018). Separation in logistic regression: Causes, consequences, and control. *American Journal of Epidemiology*, *187*(4), 864–870. https://doi.org/10.1093/aje/kwx299

McMaughan, D. J., Oloruntoba, O., & Smith, M. L. (2020). Socioeconomic Status and Access to Healthcare: Interrelated Drivers for Healthy Aging. *Frontiers in Public Health*, *8*, 231. https://doi.org/10.3389/fpubh.2020.00231

Muir, T. (2017, March). *Measuring social protection for long-term care* (OECD Health Working Papers No. 93). https://doi.org/10.1787/a411500a-en

Nagl-Cupal, M., Kolland, F., Zartler, U., Mayer, H., Bittner, M., Koller, M. M., Parisot, V., & Stöhr, D. (2018). *Angehörigenpflege in österreich. Einsicht in die Situation pflegender Angehöriger und in die Entwicklung informeller Pflegenetzwerke*. Bundesministerium für Arbeit, Soziales, Gesundheit und Konsumentenschutz (BMASGK). https://ucrisportal.univie.ac.at/de/publications/angeh%C3%B6rigenpflege-in-%C3%B6sterreich-einsicht-in-die-situation-pflegen

ÖAK. (2024). *Ärztestatistik für Österreich*. Österreichische Ärztekammer.

OECD. (2020, June 22). *Who Cares? Attracting and Retaining Elderly Care Workers*. https://doi.org/10.1787/92c0ef68-en

OECD. (2024, October 29). *Is Care Affordable for Older People?* https://doi.org/10.1787/450ea778-en

Pennerstorfer, A., & Österle, A. (2023). Take-Up and Distribution of a Universal Cash Benefit: The Case of the Austrian Long-Term Care Allowance. *Journal of Social Policy*, 1–18. https://doi.org/10.1017/S0047279423000375

Reeves, D., Pye, S., Ashcroft, D. M., Clegg, A., Kontopantelis, E., Blakeman, T., & van Marwijk, H. (2018). The challenge of ageing populations and patient frailty: Can primary care adapt? *BMJ*, *362*, k3349. https://doi.org/10.1136/bmj.k3349

Rice, H., Say, R., & Betihavas, V. (2018). The effect of nurse-led education on hospitalisation, readmission, quality of life and cost in adults with heart failure. A systematic review. *Patient Education and Counseling*, *101*(3), 363–374. https://doi.org/10.1016/j.pec.2017.10.002

Sato, J., Mitsutake, N., Kitsuregawa, M., Ishikawa, T., & Goda, K. (2022). Predicting demand for long-term care using Japanese healthcare insurance claims data. *Environmental Health and Preventive Medicine*, *27*, 42–42. https://doi.org/10.1265/ehpm.22-00084

Schulz, M., Czwikla, J., Tsiasioti, C., Schwinger, A., Gand, D., Schmiemann, G., Schmidt, A., Wolf-Ostermann, K., Kloep, S., Heinze, F., & Rothgang, H. (2020). Differences in medical specialist utilization among older people in need of long-term care – results from German health claims data. *International Journal for Equity in Health*, *19*(1), 22. https://doi.org/10.1186/s12939-020-1130-z

Serrano-Alarcón, M., Hernández-Pizarro, H., López-Casasnovas, G., & Nicodemo, C. (2022). Effects of long-term care benefits on healthcare utilization in Catalonia. *Journal of Health Economics*, *84*, 102645. https://doi.org/10.1016/j.jhealeco.2022.102645

sm-Rahman, A., Hydén, L.-C., & Kelfve, S. (2021). Eldercare services for people with and without a dementia diagnosis: An analysis of Swedish registry data. *BMC Health Services Research*, *21*(1), 893. https://doi.org/10.1186/s12913-021-06891-6

Steinbeisser, K., Grill, E., Holle, R., Peters, A., & Seidl, H. (2018). Determinants for utilization and transitions of long-term care in adults 65+ in Germany: Results from the longitudinal KORA-Age study. *BMC Geriatrics*, *18*. https://doi.org/10.1186/s12877-018-0860-x

Stolz, E., Mayerl, H., Rásky, É., & Freidl, W. (2019). Individual and country-level determinants of nursing home admission in the last year of life in Europe. *PLOS ONE*, *14*(3), e0213787. https://doi.org/10.1371/journal.pone.0213787

Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Trukeschitz, B., Österle, A., & Schneider, U. (2022). Austria's Long-Term Care System: Challenges and Policy Responses. *Journal of Long Term Care*, 88–101. https://doi.org/10.31389/jltc.112

Verbakel, E. (2018). How to understand informal caregiving patterns in Europe? The role of formal long-term care provisions and family care norms. *Scandinavian Journal of Public Health*, *46*(4), 436–447. https://doi.org/10.1177/1403494817726197

Vidiella-Martin, J., Hernández-Pizarro, H. M., García-Gómez, P., & López-Casasnovas, G. (2024). Unravelling hidden inequities in a universal public long-term care system. *The Journal of the Economics of Ageing*, *29*, 100527. https://doi.org/10.1016/j.jeoa.2024.100527

Warum, P., Culotta, F., Famira-Mühlberger, U., Horvath, T., Leoni, T., & Spielauer, M. (2025). Challenges in Long-term Care: Modelling and Quantifying Future Care Needs, Arrangements and Gaps in Ageing, Low-fertility Societies. A Novel Comparative Approach Applied in Microsimulation Projections for Austria and Italy. *WIFO Working Papers*. https://www.wifo.ac.at/publication/pid/56896344

WHO. (2025a). *Anatomical Therapeutic Chemical (ATC) Classification* [accessed on 2025-06-03]. https://www.who.int/tools/atc-ddd-toolkit/atc-classification

WHO. (2025b). *ICD-10 Version 2019* [accessed on 2025-06-03]. https://icd.who.int/browse10/2019/en

Xu, Y., Koh, X. H., Chua, Y. T. S., Tan, C. G. I., Aloweni, F. A. B., Yap, B. E. J., Tan, P. C., Chua, X., Lim, Y. K. S., Oh, H. C., Teo, S. H. S., & Lim, S. F. (2022). The Impact of Community Nursing Program on Healthcare Utilization: A Program Evaluation. *Geriatric Nursing*, *46*, 69–79. https://doi.org/10.1016/j.gerinurse.2022.04.024

Zou, H., & Hastie, T. (2005). Regularization and Variable Selection Via the Elastic Net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *67*(2), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x

# A Appendix

Table A.1: IMS Chapter Codes

| Code | IMS Chapter |
|---|---|
| IMS chapter 1 | Nervous system, psyche, cranial vault, spine |
| IMS chapter 2 | Eyes and orbit |
| IMS chapter 3 | Ears, nose, oral cavity, pharynx, face, facial skull, neck |
| IMS chapter 4 | Respiratory system, thorax |
| IMS chapter 5 | Heart and circulatory system |
| IMS chapter 6 | Endocrine glands |
| IMS chapter 7 | Digestive tract, abdomen |
| IMS chapter 8 | Urogenital tract, obstetrics |
| IMS chapter 9 | Skin and appendages |
| IMS chapter 10 | Musculoskeletal system |
| IMS chapter 11 | Organ transplantations |
| IMS chapter 12 | Diagnostic imaging and interventions |
| IMS chapter 13 | Radiotherapy |
| IMS chapter 14 | Nuclear medical diagnostics and therapy |
| IMS chapter 15 | Endoscopic diagnostics and therapy |
| IMS chapter 16 | Invasive cardiological diagnostics and therapy |
| IMS chapter 17 | Dialysis procedures |
| IMS chapter 18 | Diagnostic and therapeutic procedures |
| IMS chapter 19 | Neonatal/pediatric intensive care services |
| IMS chapter 20 | Therapy in specialized departments |
| IMS chapter 21 | Oncological therapy and other pharmacotherapy |
| IMS chapter 22 | New diagnostic and treatment methods |

Source: BMSGPK (2020b), own translation.

Table A.2: SDOC Codes

| Code | Description |
|---|---|
| SDOC code 1 | General Practitioner |
| SDOC code 2 | Specialist in Anesthesiology and Intensive Care Medicine |
| SDOC code 3 | Specialist in Ophthalmology and Optometry |
| SDOC code 4 | Specialist in Surgery |
| SDOC code 5 | Specialist in Dermatology and Venereology |
| SDOC code 6 | Specialist in Gynecology and Obstetrics |
| SDOC code 7 | Specialist in Internal Medicine |
| SDOC code 8 | Specialist in Pediatrics and Adolescent Medicine |
| SDOC code 9 | Specialist in Ear, Nose, and Throat Diseases |
| SDOC code 10 | Specialist in Pulmonary Diseases |
| SDOC code 11 | Specialist in Neurology and Psychiatry |
| SDOC code 12 | Specialist in Orthopedics and Orthopedic Surgery |
| SDOC code 13 | Specialist in Physical and Rehabilitative Medicine |
| SDOC code 14 | Specialist in Radiology |
| SDOC code 15 | Specialist in Trauma Surgery |
| SDOC code 16 | Specialist in Urology |
| SDOC code 17 | Specialist in Dentistry, Oral and Maxillofacial Medicine |
| SDOC code 18 | Specialist in Neurosurgery |
| SDOC code 19 | Specialist in Neurology |
| SDOC code 20 | Specialist in Psychiatry |
| SDOC code 21 | Specialist in Plastic Surgery |
| SDOC code 24 | Specialist in Nuclear Medicine |
| SDOC code 27 | Doctor of Dental Medicine |
| SDOC code 29 | Specialist in Immunology |
| SDOC code 30 | Orthodontist |
| SDOC code 40 | Hearing Aid Acoustician |
| SDOC code 41 | Supplier of Hearing and Speech Aids |
| SDOC code 42 | Supplier of Medical Aids and Assistive Devices |
| SDOC code 43 | Ophthalmic Prosthetist |
| SDOC code 50 | Specialist in Medical and Chemical Laboratory Diagnostics |
| SDOC code 52 | Cytodiagnostic Laboratory |
| SDOC code 53 | Specialist in Pathology |
| SDOC code 55 | Specialist in Hygiene and Microbiology / Microbiological-Serological Laboratory Diagnostics |
| SDOC code 60 | Public Pharmacy |
| SDOC code 61 | Orthopedic Bandage Specialist |
| SDOC code 62 | Licensed Dentist / General Dental Practitioner |
| SDOC code 63 | Certified Physiotherapist |
| SDOC code 64 | Certified Massage Therapist |
| SDOC code 65 | Emergency Medical Transport Services |
| SDOC code 66 | Certified Speech Therapist |
| SDOC code 67 | Optician / Contact Lens Optician |
| SDOC code 68 | Orthopedic Shoemaker |
| SDOC code 69 | Orthopedic Technician |
| SDOC code 70 | Midwife |
| SDOC code 71 | Certified Nurse |
| SDOC code 72 | Psychotherapist |
| SDOC code 73 | Clinical Psychologist |
| SDOC code 75 | Certified Occupational Therapist |
| SDOC code 80 | Inpatient Medical Facility |
| SDOC code 81 | Outpatient Medical Facility |
| SDOC code 84 | CT, MRI, and Other Diagnostic Services |
| SDOC code 85 | Independent Outpatient Clinic |
| SDOC code 86 | Independent Outpatient Clinic for Dental, Oral, and Maxillofacial Medicine |
| SDOC code 87 | Nursing Home for the Chronically Ill |
| SDOC code 89 | Outpatient Rehabilitation Facility |
| SDOC code 91 | Independent Outpatient Clinic for Physical Medicine |
| SDOC code 92 | Home Nursing Organization |
| SDOC code 99 | Other Contractual Partners |

Source: Austrian Federation of Social Insurances, own translation.

Table A.3: ICD-10 Chapters

| ICD-10 block | ICD-10 Chapter |
|---|---|
| ICD-10 block 1 | Austrian Group 1: Nervous system, skull, spine |
| ICD-10 block 9 | Austrian Group 9: Skin and appendages |
| ICD-10 block A | Chapter I: Certain infectious and parasitic diseases |
| ICD-10 block B | Chapter I: Certain infectious and parasitic diseases |
| ICD-10 block C | Chapter II: Neoplasms |
| ICD-10 block D | Chapter II: Neoplasms and Chapter III: Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism |
| ICD-10 block E | Chapter IV: Endocrine, nutritional and metabolic diseases |
| ICD-10 block F | Chapter V: Mental and behavioural disorders |
| ICD-10 block G | Chapter VI: Diseases of the nervous system |
| ICD-10 block H | Chapter VII: Diseases of the eye and adnexa and Chapter VIII: Diseases of the ear and mastoid process |
| ICD-10 block I | Chapter IX: Diseases of the circulatory system |
| ICD-10 block J | Chapter X: Diseases of the respiratory system |
| ICD-10 block K | Chapter XI: Diseases of the digestive system |
| ICD-10 block L | Chapter XII: Diseases of the skin and subcutaneous tissue |
| ICD-10 block M | Chapter XIII: Diseases of the musculoskeletal system and connective tissue |
| ICD-10 block N | Chapter XIV: Diseases of the genitourinary system |
| ICD-10 block Q | Chapter XVII: Congenital malformations, deformations and chromosomal abnormalities |
| ICD-10 block R | Chapter XVIII: Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified |
| ICD-10 block S | Chapter XIX: Injury, poisoning and certain other consequences of external causes |
| ICD-10 block T | Chapter XIX: Injury, poisoning and certain other consequences of external causes |
| ICD-10 block U | Chapter XXII: Codes for special purposes |
| ICD-10 block Z | Chapter XXI: Factors influencing health status and contact with health services |

Source: WHO (2025b), blocks 1 and 9 refer to the Austrian classification as used in 2020.

Table A.4: ATC Codes

| Code | Description |
|---|---|
| ATC Code A | Alimentary tract and metabolism |
| ATC Code B | Blood and blood forming organs |
| ATC Code C | Cardiovascular system |
| ATC Code D | Dermatologicals |
| ATC Code G | Genito urinary system and sex hormones |
| ATC Code H | Systemic hormonal preparations, excl. sex hormones and insulins |
| ATC Code J | Antiinfectives for systemic use |
| ATC Code L | Antineoplastic and immunomodulating agents |
| ATC Code M | Musculo-skeletal system |
| ATC Code N | Nervous system |
| ATC Code P | Antiparasitic products, insecticides and repellents |
| ATC Code R | Respiratory system |
| ATC Code S | Sensory organs |
| ATC Code V | Various |

Source: WHO (2025a).