ARC Centre of Excellence in Population Ageing Research

Working Paper 2024/13

Length of stay in residential aged care: patterns and determinants from a population-based cohort study

Mengyi Xu and Gaoyun Yan

Length of stay in residential aged care: patterns and determinants from a population-based cohort study

Mengyi Xu*[1,2] and Gaoyun Yan †[2]

[1]*Department of Statistics and Department of Mathematics, Purdue University, United States*
[2]*ARC Centre of Excellence in Population Ageing Research, UNSW Sydney, Australia*

2nd August 2024

Abstract

The length of stay in permanent residential care is a crucial metric for evaluating the utilization of institutional care and informing sustainable aged care policies. Understanding this metric is especially relevant in Australia, where the decision on how to pay the substantial nursing home accommodation costs must be made shortly after admission and is heavily influenced by the expected duration of stay. We investigate the length of stay in long-term institutional care by analyzing a cohort of older Australians first admitted to permanent residential care in 2008. By employing survival analysis that captures time-varying covariates, we find that, in addition to demographic factors like age and gender, the organization type of nursing homes and their service size significantly influence the length of stay. Failing to account for potential changes due to transfers between nursing homes can lead to a significant underestimation of the impact of organization type and service size.

Keywords: length of stay, nursing homes, AFT model, survival analysis, prospective cohort study

*Correspondence to: 150 N. University St. West Lafayette, IN, 47907 U.S. Email: xumy@purdue.edu
†Email: g.yan@unsw.edu.au

1 Introduction

Population aging, driven by declining fertility rates and rising life expectancy, poses significant challenges in healthcare, social services, and family dynamics. A major concern is the care of older adults who struggle with activities of daily living, as the demand for aged care has surged, particularly among the aging baby boomer generation. Among OECD countries, Spain has seen the most substantial increase in this demand; the number of seniors over age 65 using long-term care services tripled from 2009 to 2022, reaching nearly 1.2 million (OECD, 2023). In Australia, the number of seniors requiring care doubled from 2006 to 2021, surpassing 600,000. This challenge is further compounded by shrinking family sizes, which reduce the availability of informal care, and by the growing prevalence of diseases such as dementia, which require specialized care.

Among various types of aged care, residential care is considered the most expensive due to its high labor and capital requirements. Residents typically require extensive care and assistance, necessitating facilities equipped with special amenities. In countries like Australia, the high costs are driven by needs for round-the-clock care, specialized services, and tailor-made accommodation. To reduce care costs and meet public preferences for aging in place, governments worldwide have been promoting in-home and community-based care options. For instance, Australia's Living Longer Living Better (LLLB) reform in 2012 has significantly enhanced the availability of home care packages and expanded their scope to include intensive care and medical support. This type of policy shift may be effective as a decline in the incidence rate of accessing permanent residential aged care was observed from 2008-09 to 2015-16 (Khadka et al., 2019). On the other hand, the promising policy effect may be temporary as the demographic trends toward an aging population could potentially increase the demand for institutional care (Alders and Schut, 2019). Therefore, alongside the evolution of policy, research on key aspects of care utilization and costs, particularly the length of stay in residential care, is of great importance.

Understanding the length of stay in residential care is crucial for all stakeholders, as it is a primary determinant of overall expenditures and provides insights into care utilization. For residents and their families, knowledge about the length of stay informs financial planning and life-stage preparations. For example, anticipating the need for residential care may encourage retirees to increase their precautionary savings, which could influence their consumption patterns during retirement (Alonso-García et al., 2022). From the government's perspective, detailed knowledge of these durations can enhance aged care system management by facilitating more effective resource allocation, ensuring that services align with the needs and preferences of the population (Fuino and Wagner, 2020; Hoben et al., 2019). For service providers, understanding the duration of stays is essential, as high resident turnover can lead to increased administrative costs and lower occupancy rates, affecting their operational efficiency (Liu, 1996).

Studying the length of stay in permanent residential aged care in Australia is particularly important due to the high accommodation costs and their strong association with the expected length of stay. The LLLB reform expanded payment options for permanent care residents from a lump-sum deposit to three choices: 1) a refundable lump-sum accommodation deposit (RAD), 2) a rental-style payment known as the daily accommodation payment (DAP), or 3) a combination of both RAD and DAP. The decision, which must be made within 28 days after admission, often involves hundreds of thousands of dollars. Therefore, selecting a payment method requires careful consideration by residents and their families, with the expected length of stay being a critical factor. Those expecting a shorter stay may prefer the DAP to avoid the need to quickly secure a large sum of money, whereas those anticipating a longer stay might choose the RAD to preserve bequests and avoid investment risks.

It is worth noting that in many countries, including the U.S. and the U.K., the term residential care distinguishes between care homes, which offer only assisted living, and nursing homes, which provide comprehensive care for daily activities and 24-hour on-site medical care by qualified nurses. In contrast, Australia does not make this distinction; all services requiring a transition from a family home to a facility are categorized as residential aged care, regardless of the recipient's medical needs. Prior to the 2021 Royal Commission into Quality of Aged Care, Australian service providers were not required to have qualified nurses on-site. Following the Commission's recommendations, this became a mandatory requirement for all residential care facilities. As a result, current residential aged care in Australia includes both daily life assistance and medical care, staffed by qualified nurses. In our paper, we use the terms nursing homes, care homes, institutional care, and residential care interchangeably.

It is also important to note that nursing homes in countries such as the U.S., U.K., Canada, and Australia typically provide both short-term and long-term care services. Short-term services, like respite care and transition care following hospital discharge, are designed to be temporary, with durations often capped by regulations. In contrast, our study focuses on long-term care, which involves extended stays that are crucial for understanding the broader implications of residential care. This emphasis is reflected in our research methodology and data sample selection, which target permanent residential care for the elderly.

Research on the length of stay in residential care from first admission to death is sparse due to several factors. First, older people in long-term institutional care are frequently excluded from survey samples (Moore et al., 2019). When data are available, they often conflate short-term and long-term residents (see e.g., Kelly et al., 2010). Since the lengths of stay for short-term residents are usually capped, this pooling can obscure the true distribution of long-term stays and lead to skewed conclusions about patterns and determinants. Although census-based surveys can be a solution, their implementation is costly, with Connolly et al. (2014) being the only example to date.

Second, when administrative data are used for length of stay analysis, some studies rely only on data from the last episode rather than the complete record (Zhang et al., 2023; Hedinger et al., 2015). This approach fails to account for transfers between facilities, which we will demonstrate is important to consider. Additionally, some studies focus on specific regions (Hoben et al., 2019;

Steventon and Roberts, 2012; Schön et al., 2016) or particular population groups (Welberry et al., 2020), limiting the generalizability of their findings. While population-based research could mitigate these demographic limitations, such studies often do not connect length of stay with associated characteristics, instead focusing on estimating the distributions of length of stay (Liu and Manton, 1983; Martikainen et al., 2014; Schön et al., 2016). Furthermore, existing regression-based research typically focuses on health conditions, often neglecting socioeconomic and institutional factors.

Overall, there is a lack of population-based studies on length of stay that also examine related social and facility characteristics, with notable exceptions in Switzerland (Fuino and Wagner, 2020) and the U.S. (Spector et al., 1998). To address these gaps, we will analyze individual-level administrative data from Australia, aiming to provide detailed insights into the length of stay in long-term institutional care and the associated characteristics.

Our dataset includes a cohort of residents first admitted to permanent residential care in 2008. We obtained their complete admission records from the first admission to their last discharge or up to June 30, 2022. The data was sourced from the Australian Institute of Health and Welfare. We define the length of stay as the total duration spent in a residential aged care facility from the first admission to the final discharge. This definition represents the lifetime use of nursing home care, aligning with those used in other studies (see e.g., Hoben et al., 2019; Kemper and Murtaugh, 1991). Unlike measures that consider only individual episodes of care, this approach captures the overall utilization of institutional care. It therefore better informs policy development and accurately reflects the long-term nature of planning for nursing home stays. The length of stay variable is subject to right censoring, which occurs if the final discharge is not due to death, or if the resident remains in care at the end of the observation period. To address this, we apply survival analysis techniques to adjust for right censoring.

We use the accelerated failure time model to analyze the length of stay, distinguishing our approach from existing studies in several key ways. First, unlike previous research that often limits analysis to one- or two-parameter distributions like exponential and Weibull (Fuino and Wagner, 2020), we explore families of distributions including the generalized gamma and generalized F distributions. The generalized F distribution, in particular, provides the best fit due to its flexibility in capturing the force of mortality for both short- and long-stay residents. Second, we apply the model to the full sample rather than to sub-samples, as done in studies such as Liu (1996) and Zhang et al. (2023). This approach improves the reliability of parameter estimation and enables us to test the significance of each covariate. Our findings reveal that, in addition to common demographic factors, the organization type and service size of the nursing home significantly impact the length of stay. Furthermore, we incorporate time-dependent covariates in our regression, contrasting with prior research that typically uses covariates measured at the first admission (see e.g., Fuino and Wagner, 2020; Hoben et al., 2019). We discover that failing to account for changes due to transfers between nursing homes can lead to an underestimation of the impact of the organization type and service size of nursing homes.

To our knowledge, this paper is the first to present a prospective cohort study of the overall length of stay in permanent residential aged care within Australia with a comprehensive analysis of the length of stay patterns and associated characteristics. The extensive duration of our data allows us to accurately capture the complete record of long-term institutional care with minimal information loss. As a result, this study offers a detailed portrayal of long-term institutional aged care in Australia. Internationally, it stands out as one of the few population-based studies capturing demographic, socioeconomic, and institutional characteristics in residential aged care. The inclusion of these explanatory variables is particularly significant due to the more socially oriented nature of institutional long-term care, which contrasts with the medical focus typically associated with hospital stays and transitional care in nursing homes. Furthermore, we apply a novel modeling framework that significantly extends existing literature, revealing new determinants of length of stay.

The remainder of the paper is organized as follows. Section 2 provides a brief overview of the aged care system in Australia. Section 3 describes the data used for our analysis, including exploratory data analysis. Section 4 outlines the modeling framework. Section 5 discusses the estimation results. Section 6 concludes with a brief discussion on policy implications and future research.

2 Aged care system in Australia

In Australia, aged care services are designed to assist older individuals in maintaining their daily routines. To qualify for government-subsidized aged care services, individuals must be at least 65 years old, or 50 for Aboriginal or Torres Strait Islander people. The Aged Care Assessment Team (ACAT), funded by the Australian government, serves as the primary evaluator. Its responsibilities include assessing individuals' care needs, determining eligibility for services, and providing recommendations on the appropriate type and level of care.

Depending on the recipient's condition and needs, the aged care service can be provided in short-term or long-term care. Short-term care refers to non-permanent arrangements lasting no more than 12 weeks, including transition care that aids recovery after a hospital stay and respite care that provides temporary relief for primary caregivers. Long-term care, in contrast, is usually on an ongoing basis, and can be delivered to older people's homes or delivered in facilities that accommodate service recipients on-site. Both home care and residential care are available at multiple levels to meet varying care needs. Home care providers offer packages ranging from entry-level support (through the Commonwealth Home Support Programme) to more intensive levels (available through Home Care Packages). Residential care, on the other hand, involves relocating to a nursing home where round-the-clock care is provided.

It is important to note that the transition from home care to permanent residential care is typically irreversible. In fact, in our dataset, less than 2% of residents were discharged to return to family or home, based on the last discharge record of each resident. After all, the primary determinant for the ACAT team to recommend the appropriate care setting—whether at home or in a nursing home—is the recipient's ability to safely maintain living at home with assistance. In other words, home care services support the aging-in-place paradigm by providing domestic assistance, social support, personal care, and transportation. However, if an individual's mental or physical condition deteriorates beyond what home services can accommodate, the ACAT will recommend admission to a nursing home for permanent care, following a request for assessment from the care recipients or their families.

In terms of expenditure in Australian nursing homes, costs fall into four main categories: the basic daily fee, the means-tested care fee, the means-tested accommodation costs, and fees for additional services. The basic daily fee, which covers everyday living expenses, is not subsidized but is capped at 85% of the single basic age pension. Similarly, no government subsidy is applied to cover fees for extra services. The accommodation cost constitutes a significant portion of the overall expenditures in nursing homes, with total amounts often reaching hundreds of thousands of dollars. While the government subsidizes care and accommodation costs based on residents' means test results, the primary responsibility for covering these expenses still falls on the individual (Sherris, 2021).

3 Prospective cohort data

There are three typical ways of analyzing the length of stay: retrospectively at discharge, by a cross-section of residents, and prospectively at admission (Keeler et al., 1981). The retrospective method is heavily influenced by variations in the size and composition of all past admission cohorts, and the cross-sectional method is biased towards long stayers even though their length of stay is censored. By contrast, the prospective method provides an unbiased measure of the length of stay (Keeler et al., 1981). Consequently, we use the prospective method for our analysis. We describe the dataset, including its size and variables, in Section 3.1. This is followed by an explanatory data analysis on admission, discharge, and length of stay records in Section 3.2.

3.1 Data description

We acquired the de-identified individual-level permanent residential care admission records from the Australian Institute of Health and Welfare (AIHW). We obtained the complete admission records of a cohort of residents who were first admitted to nursing homes in 2008. Our dataset covers the whole period from residents' first admission (in 2008) to their last discharge or up to June 30, 2022.

The dataset is organized as a table, where each row represents an episode of care, and each column represents a variable. An episode of care is defined as a period of consecutive stay in a residential aged care facility, starting when a resident is admitted into a nursing home and ending when the resident is discharged. Since a resident can freely leave a nursing home and re-enter as long as a spot is available, there can be multiple episodes of care for one individual.

The raw dataset obtained from AIHW initially includes 67,552 episodes of care for 52,658 individuals. Individuals associated with episodes of care that had a zero length of stay were removed from the sample, resulting in 52,619 individuals and 67,461 episode records. Considering the eligibility age for government-subsidized aged care services, as discussed in Section 2, we further restricted our sample to those who were at least 50 years old at the first admission. The final dataset used for analysis includes 51,738 individuals and 65,989 episodes. The number of observations removed from the raw dataset is insignificant (less than 2.5%).

Table 1 displays the full list of variables along with their descriptions, covering characteristics of individual residents and aged care facilities. These variables can either be time-dependent or time-independent. For time-varying variables, information is recorded at the start of each episode. Individual characteristics include demographic factors such as age, gender, indigenous status, country of birth, and preferred language. The data on residential aged care facilities includes ownership structure (i.e., not-for-profit, government, private), service size (grouped in 20-bed intervals), and location details (covering Aged Care Planning Region, geographic remoteness, and state/territory). Additionally, the duration of each episode is recorded in days. We convert it to months by dividing $\frac{365}{12}$.

Table 1. Variable names and their explanations in the dataset.

Variable name	Explanation
Individual characteristics	
DEIDENT_INDIV_CODE	Unique identifier assigned to each resident.
AGE_GROUP	5-year age group at the time of admission.
SEX	Gender of a resident.
INDIGENOUS_INDICATOR	Indicator for Aboriginal and Torres Strait Islander peoples.
COUNTRY_OF_BIRTH	Country of birth of a resident.
PREFERRED_LANGUAGE	Preferred language spoken by a resident.
Residential aged care facility characteristics	
ORGANISATION_TYPE*	Type of ownership structure of the provider organization managing a residential aged care facility.
SERVICE_SIZE	Number of beds in a residential aged care facility.
ACPR_CODE†	Code of the Aged Care Planning Region (ACPR) based on the location of the service delivered.
ACPR_NAME	Name of the ACPR based on the location of the service delivered.
REMOTENESS‡	Geographical classification indicating the remoteness of a residential aged care facility.
STATE§	State/territory where a residential aged care facility is located.
Admission and discharge	
ADMISSION_YEAR	Year in which a resident entered a residential aged care facility for the current episode of care.
DISCHARGE_YEAR	Year in which a resident left a residential aged care facility for the current episode of care.
DISCHARGE_REASON	Reason why a resident left a residential aged care facility for the current episode of care.
LOS_DAYS	Number of days a resident stayed in a residential aged care facility for the current episode of care.

* The organization types are classified into three categories: not-for-profit (including charities, religious organizations, and community-based organizations), government (encompassing state government, territory government, and local government organizations), and private (which includes publicly listed companies and organizations registered as private companies) (AIHW, 2023).

† ACPRs are the regions in Australia where aged care services are funded and delivered. We use the 2018 ACPRs, which are based on Statistical Area Level 2 (SA2) boundaries from the Australian Bureau of Statistics Australian Statistical Geography Standard 2016.

‡ Geographic remoteness is classified into five categories across Australia according to the Australian Statistical Geography Standard: Major Cities, Inner Regional, Outer Regional, Remote, and Very Remote.

§ Australia contains six states—New South Wales (NSW), Victoria (Vic), Queensland (Qld), South Australia (SA), Western Australia (WA), and Tasmania (Tas)—and two territories—Northern Territory (NT) and Australian Capital Territory (ACT).

Prior studies have highlighted the influence of socioeconomic status (SES) on the duration of nursing home stays (see e.g., Hedinger et al., 2015). As individual-level SES data (e.g., income or education) are unavailable in our dataset, we use area-level measures as proxies, a common approach to mitigate data scarcity (Moss et al., 2021). Specifically, we employ the Index of Relative Socio-economic Disadvantage (IRSD) to assign scores to each Aged Care Planning Region (ACPR). The IRSD is part of the Socio-Economic Indexes for Areas (SEIFA) produced by the Australian Bureau of Statistics. The methodology for calculating these scores is detailed in Appendix A. ACPRs are then ranked by these scores, with the lowest 20% categorized as SES Q1 and the highest 20% as SES Q5. Figure 1 depicts this distribution, showing that higher SES areas are predominantly located in coastal regions near capital cities across various states and territories.

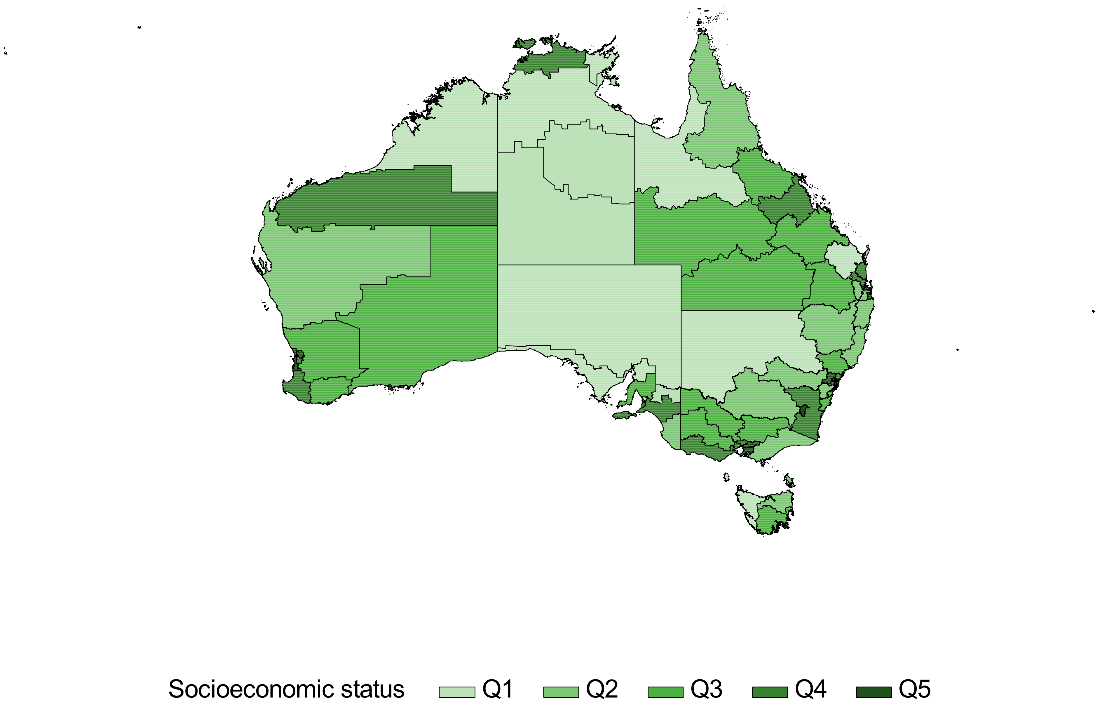

Figure 1. The Aged Care Planning Region (ACPR) by socioeconomic status, with Q1 representing the lowest and Q5 representing the highest.

3.2 Exploratory data analysis

3.2.1 Admission, discharge and transfer

Table 2 presents the age distribution at the first admission, categorized by several demographic factors, including gender, indigenous status, country of birth, and preferred language. To facilitate comparisons, we calculate a weighted average age at the first admission for each group, the weights being the proportion of residents in each age group. Since ages are grouped in five-year intervals, we use the midpoint of each range; for the age group of 100+, we use 102 as the midpoint for these calculations.

Table 2. The distribution of age at the first admission by different demographic factors.

	Gender		Indigenous Indicator		Country of Birth		Preferred Language	
	Female	Male	No	Yes	Australia	Other	English	Other
Age at the first admission								
50-54	0.4%	0.7%	0.4%	7.4%	0.6%	0.2%	0.5%	0.3%
55-59	0.7%	1.5%	1.0%	7.4%	1.1%	0.9%	1.0%	0.9%
60-64	1.2%	2.5%	1.7%	7.4%	1.8%	1.6%	1.7%	1.7%
65-69	2.4%	4.8%	3.3%	10.4%	3.3%	3.5%	3.3%	3.4%
70-74	5.4%	8.4%	6.4%	14.4%	6.3%	7.0%	6.4%	7.5%
75-79	12.4%	15.2%	13.4%	18.8%	13.1%	14.4%	13.2%	16.4%
80-84	25.0%	25.4%	25.2%	15.4%	24.4%	27.0%	24.7%	29.0%
85-89	29.2%	25.8%	28.0%	12.4%	27.9%	27.9%	28.2%	25.4%
90-94	18.0%	12.6%	16.0%	4.4%	17.0%	13.1%	16.4%	11.7%
95-99	4.8%	2.9%	4.1%	1.3%	4.1%	4.0%	4.1%	3.5%
100+	0.6%	0.3%	0.5%	0.7%	0.4%	0.5%	0.5%	0.3%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Average age at the first admission†								
	84.02	81.65	83.19	73.91	83.23	82.87	83.22	82.31

† It is determined by taking the midpoint of each age group and calculating a weighted average, with the weights based on the proportion of residents in each age group. For the age group of 100+, 102 is used as the midpoint in the calculation.

Table 2 shows that men are admitted to aged care facilities at a younger age than women, with a greater concentration in the male age groups under 80. Indigenous individuals enter these facilities at younger ages, attributable to their earlier eligibility for government-subsidized aged care services. Additionally, elders who are Australian-born and English-speaking tend to enter care at an older age compared to those born overseas or who prefer speaking a foreign language, though the difference is minor (less than one year).

The primary reason for discharge in our dataset is death. Among 51,738 residents, there were 49,081 deaths recorded in nursing homes. At the end of the observation period, 583 residents were still in care. Other common reasons for discharge include returning to family or home and being transferred to a hospital. These observations, along with those who were still in care, represent slightly over 5% of the total records. Figure 2 presents a breakdown of the deceased and censored observations by the year of final discharge. It shows that the vast majority of residents remained in permanent residential care until the end of their lives.

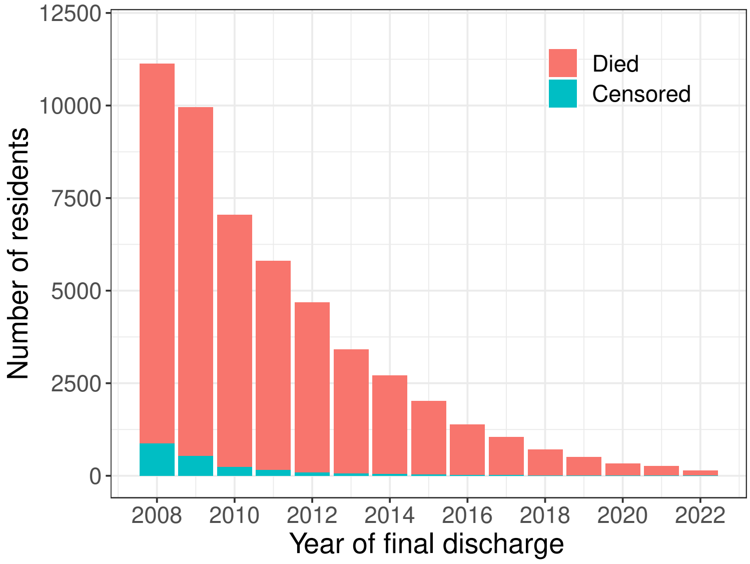

Figure 2. The number of residents who left the residential aged care facility for the last time between 2008 and 2022.

More than one-fifth of the residents have multiple admission records (Table 3). A significant majority of these residents were discharged and re-admitted within the same year. Although our dataset does not identify specific nursing home facilities, it enables us to track changes in location and hence the SES quintile, organization type, and service size. Figure 3 and Figure 4 illustrate these transitions. Changes in location are minimal, with no inter-state transfers and largely consistent remoteness areas and SES quintiles (Figure 3). In addition, over half of the re-admission records retain the same organization types as previous episodes (left panel of Figure 4). By contrast, there is a notable tendency for residents with multiple admissions to move to larger facilities, particularly those with more than 100 beds (right panel of Figure 4).

Table 3. Distribution of the number of admission records.

	Number of admissions			
	1	2	3+	Total
Frequency	39,914	9,892	1,932	51,738
Percentage (%)	77.1%	19.1%	3.7%	100.0%

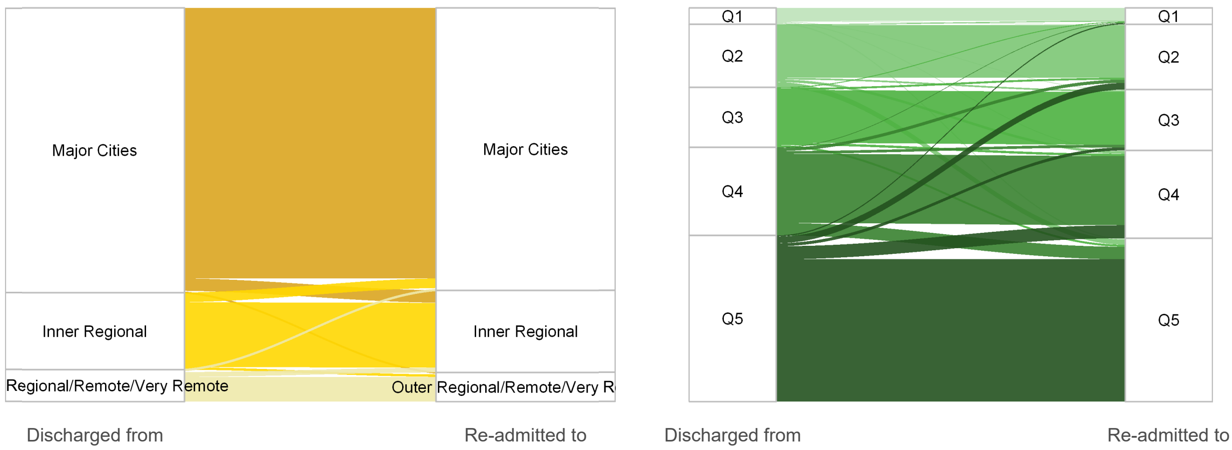

Figure 3. Transitions in (Left Panel) remoteness and (Right Panel) socioeconomic status for residents with more than one admission record. The widths of the flows are proportional to the number of transitions.

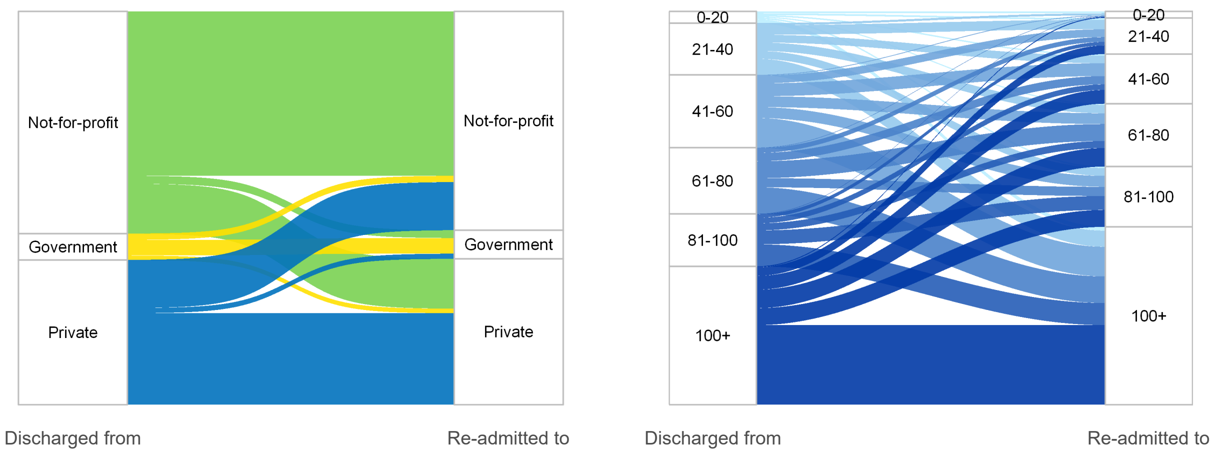

Figure 4. Transitions in (Left Panel) organization type and (Right Panel) service sizes for residents with more than one admission record. The widths of the flows are proportional to the number of transitions.

3.2.2 Length of stay

Figure 5 shows the histogram of the length of stay, which features a long tail. The right skewness is a common characteristic of the length of stay distributions, typically resulting from higher mortality rates among newly admitted residents (Connolly et al., 2014; Kelly et al., 2010). While the majority of residents spent only a few years in nursing homes, it is not uncommon for the length of stay to exceed 10 years. A notable bump occurs at around 168 months, coinciding with our investigation period of 14 years (or 168 months). Most of these observations correspond to residents who were still in care at the end of the period.

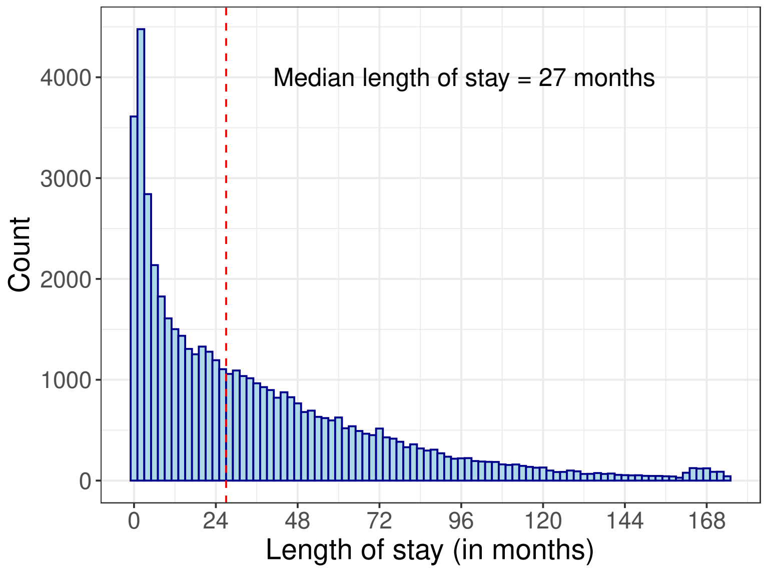

Figure 5. Histogram of length of stay. The red dashed line represents the median length of stay using the Kaplan-Meier estimate.

Given the long-tail distribution of length of stay, the median is a more meaningful summary statistic than the average. Moreover, calculating the median without considering the right censoring will underestimate the value. We therefore use the Kaplan-Meier estimator (Kaplan and Meier, 1958) to calculate the median length of stay. The overall median length of stay is 27 months, comparable to the international experiences of around two years (Allers and Hoffmann, 2018; McCann et al., 2009; Vossius et al., 2018). The slight discrepancy can be attributed to our study's exclusive focus on permanent stays, whereas prior studies often pool data from both short-stay and permanent care residents.

Table 4 presents the median length of stay and its 95% confidence interval for various groups of residents, along with the proportion of residents within each group. The group with the highest proportion will be used as the baseline in subsequent regression analyses. The table indicates that age significantly influences the duration of stay in permanent aged care: the younger the age at the first admission, the longer the duration in nursing homes, with the exception of the youngest group. Gender also plays a role, as women typically have longer stays than men, attributed to their longer life expectancy. This is despite men generally entering permanent care at a younger age (see Table 2). Indigenous residents, despite their shorter life expectancy

(Zhao et al., 2022), tend to spend more time in nursing homes, which is attributed to their younger age at the first admission. Differences based on country of birth and preferred language are relatively minor and correlate with variations in age at the first admission.

Table 4. Descriptive statistics on the length of stay (in months): median and its 95% confidence interval (CI). The last column reports the relative frequencies in percentage.

	Median	95% CI	%
Overall	27.0	(26.6, 27.4)	100.0
Age at the first admission			
50-54	47.3	(35.1, 66.5)	0.5
55-59	55.5	(46.1, 67.6)	1.0
60-64	52.4	(45.2, 57.2)	1.7
65-69	36.4	(33.3, 40.1)	3.3
70-74	34.2	(32.1, 36.8)	6.5
75-79	31.3	(30.3, 32.6)	13.5
80-84	29.1	(28.2, 29.9)	25.1
85-89	25.9	(25.2, 26.5)	27.9
90-94	21.9	(21.1, 22.7)	15.9
95-99	17.9	(16.5, 19.1)	4.1
100+	10.2	(8.6, 13.2)	0.5
Gender			
Female	32.6	(32.1, 33.1)	62.5
Male	18.7	(18.1, 19.2)	37.5
Indigenous status			
Non-indigenous	27.0	(26.5, 27.4)	99.4
Indigenous	33.2	(27.2, 41.8)	0.6
Country of birth			
Australia	26.8	(26.3, 27.3)	71.8
Other	27.5	(26.7, 28.4)	28.2
Preferred language			
English	26.8	(26.4, 27.3)	90.6
Other	28.7	(27.3, 29.9)	9.4
Socioeconomic status at the first admission			
Q1	24.8	(23.3, 26.8)	5.2
Q2	26.7	(25.8, 27.6)	16.8
Q3	28.1	(27.0, 29.0)	15.3
Q4	26.5	(25.6, 27.4)	24.0
Q5	27.3	(26.6, 27.9)	38.7
Organization type at the first admission			
Not-for-profit	30.5	(30.0, 31.1)	54.7
Government	24.9	(23.4, 27.1)	5.3
Private	22.5	(22.0, 23.1)	40.0

Continued on next page

Table 4 – continued from previous page

	Median	95% CI	%
Service size at the first admission			
0-20	31.6	(29.1, 34.8)	1.7
21-40	30.8	(29.6, 32.4)	8.7
41-60	28.4	(27.4, 29.4)	15.7
61-80	25.7	(24.7, 26.9)	16.8
81-100	25.7	(24.8, 26.7)	15.6
100+	26.4	(25.8, 27.0)	41.6
Remoteness at the first admission			
Major Cities	27.1	(26.6, 27.6)	69.9
Inner Regional	26.2	(25.3, 27.1)	21.7
Outer Regional/Remote/Very Remote	28.0	(26.6, 30.1)	8.4
State/Territory			
New South Wales	26.5	(25.8, 27.3)	34.4
Victoria	27.6	(26.9, 28.5)	25.7
Queensland	27.5	(26.5, 28.5)	17.6
Western Australia	27.5	(26.2, 28.8)	8.4
South Australia	26.4	(25.2, 27.8)	9.9
Tasmania	23.9	(22.3, 25.7)	2.8
Australian Capital Territory	26.8	(24.0, 30.7)	1.1
Northern Territory	31.8	(24.0, 49.6)	0.2

The length of stay varies significantly with the type of organization. Residents in nursing homes owned by not-for-profit organizations have a median stay that is eight months longer than those in privately owned facilities. Service size also appears to influence the length of stay; facilities with fewer than 60 beds exhibit significantly longer stays compared to those with more than 60 beds. However, the median length of stay is relatively consistent across different areas defined by remoteness. Across states and territories, the duration of stays is generally comparable, except in Tasmania, where it is significantly shorter than the national average, and in the Northern Territory, where it is noticeably longer, primarily due to its significant indigenous population. Additionally, the Northern Territory shows a wider confidence interval, which can be attributed to its smaller overall population size.

Transfers between nursing home facilities can result in changes in organization type, service size, and remoteness, as shown in Figures 3 and 4. When analyzing how length of stay varies with these factors, it is crucial to consider their potential changes over time to accurately reflect their impact. Treating these variables as time-independent, by only using their values at the first admission, not only wastes data but also incorrectly extrapolates their effects. Table 5 compares the differences between the time-independent and time-dependent approaches, the latter using values at each admission. The differences between the two methods correlate with the proportion of changes depicted in Figures 3 and Figure 4. If the majority of transfers involve no changes in the value of the factor (e.g., remoteness), the distortion is relatively small. However, changes in service size are non-negligible. As a result, the time-independent approach consistently underestimates the median length of stay for small facilities (with 60 or fewer beds).

Table 5. Compare Kaplan-Meier estimates stratified by time-dependent and time-independent variables: median length of stay and its 95% confidence interval (CI).

	Time-dependent		Time-independent	
	Median	95% CI	Median	95% CI
Socioeconomic status				
Q1	24.8	(23.3, 26.9)	24.8	(23.3, 26.8)
Q2	26.5	(25.6, 27.4)	26.7	(25.8, 27.6)
Q3	28.0	(26.9, 28.9)	28.1	(27.0, 29.0)
Q4	26.5	(25.6, 27.4)	26.5	(25.6, 27.4)
Q5	27.5	(26.7, 28.1)	27.3	(26.6, 27.9)
Organisation type				
Not-for-profit	30.9	(30.4, 31.5)	30.5	(30.0, 31.1)
Government	24.8	(22.9, 26.4)	24.9	(23.4, 27.1)
Private	22.1	(21.6, 22.7)	22.5	(22.0, 23.1)
Service size				
0-20	32.9	(30.1, 36.8)	31.6	(29.1, 34.8)
21-40	31.0	(29.8, 32.8)	30.8	(29.6, 32.4)
41-60	29.2	(28.2, 30.6)	28.4	(27.4, 29.4)
61-80	25.6	(24.7, 26.7)	25.7	(24.7, 26.9)
81-100	26.0	(25.2, 27.1)	25.7	(24.8, 26.7)
100+	26.1	(25.5, 26.7)	26.4	(25.8, 27.0)
Remoteness				
Major Cities	27.2	(26.7, 27.7)	27.1	(26.6, 27.6)
Inner Regional	26.2	(25.2, 27.0)	26.2	(25.3, 27.1)
Outer Regional/Remote/Very Remote	28.0	(26.5, 29.8)	28.0	(26.6, 30.1)

The median length of stay, estimated by the Kaplan-Meier method, provides an initial indication of the impact of each covariate on residential care durations. To test the significance of these covariates, both time-varying and time-independent, a more formal statistical model is required.

4 Modeling determinants of length of stay

We adopt regression models in survival analysis to assess how various factors influence the length of stay while controlling for other covariates. The Cox proportional hazards model is widely used for such analyses; however, statistical tests based on Schoenfeld residuals (Grambsch and Therneau, 1994) indicate that the proportional hazards assumption is violated for almost all of the variables, as shown in Table 6.

The accelerated failure time (AFT) model offers an alternative when the proportional hazards assumption fails. This model is based on the survival curve and assumes that a covariate has a time-consistent multiplicative effect on survival time. To evaluate the validity of this

Table 6. Test the proportional hazard assumption in Cox models based on the Schoenfeld residuals.

Covariate	Degrees of freedom	Test statistics
Age at the first admission	10	637.07***
Gender	1	303.70***
Indigenous indicator	1	0.48
Country of birth	1	4.66**
Preferred language	1	4.36**
Socioeconomic status	4	32.72***
Organisation type	2	181.71***
Service size	5	10.47*
Remoteness	2	13.40***
State	7	42.46***

Note: Significance levels $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

assumption, quantile-quantile (Q-Q) plots of the survival function are often used. Ideally, these plots should display a straight line through the origin if the assumption holds. We estimate the survival function using the Kaplan-Meier method and generate Q-Q plots for each covariate listed in Table 6. These plots compare the survival function quantiles of each group against those of their respective baseline group. The results, displayed in Appendix B, confirm that the AFT model assumption is appropriate for our dataset.

Estimating an AFT model involves two main steps. The first step is to determine the appropriate distribution for the survival time, in our case, the length of stay. We introduce candidate distributions in Section 4.1. The second step involves selecting explanatory variables. We use the backward stepwise selection method based on the Akaike Information Criterion (AIC). This method requires a likelihood function, which we discuss in Section 4.2.

4.1 Candidate distributions for length of stay

When fitting the observed length of stay, we consider not only commonly used distributions in survival analysis—log-normal, log-logistic, exponential, Weibull, and gamma—but also two families of distributions: generalized gamma and generalized F. The generalized gamma distribution is a three-parameter generalization of the gamma distribution and includes the log-normal, exponential, Weibull, and gamma distributions, but not the log-logistic (Stacy, 1962; Stacy and Mihram, 1965). On the other hand, the generalized F distribution is a four-parameter family of distributions that include both the generalized gamma and the log-logistic distributions (Ciampi et al., 1986).

A random variable T follows the generalized F distribution if the transformed random variable $(e^{-\mu}T)^{1/\sigma}$ follows the F distribution for certain parameters μ and $\sigma > 0$. We use the parameterization proposed in Prentice (1975), which has become the standard due to its well-behaved log-likelihood function, particularly in the limiting case of the generalized gamma distribution.

The probability density function of the generalized F distribution is given by

$$f(t \mid \mu, \sigma, Q, P) = \frac{\delta(m_1/m_2)^{m_1} e^{m_1 w}}{\sigma\, t(1 + m_1 e^{w}/m_2)^{m_1+m_2} B(m_1, m_2)}, \quad \sigma > 0, P > 0, \tag{1}$$

where

$$m_1 = 2\left(Q^2 + 2P + Q\delta\right)^{-1},\ m_2 = 2\left(Q^2 + 2P - Q\delta\right)^{-1},\ \delta = \left(Q^2 + 2P\right)^{1/2},\ w = \frac{(\ln t - \mu)\delta}{\sigma},$$

and $B(m_1, m_2)$ is the beta function evaluated at m_1, m_2. Cox (2008) describes how the generalized F distribution given by Equation (1) is related to other distributions in the generalized gamma family and the log-logistic distribution.

4.2 Accelerated failure time model

A standard accelerated failure time model with time-independent covariates is given by

$$\ln T = -\boldsymbol{\beta}^\top \boldsymbol{X} + \epsilon,$$

where T is the survival time, or the length of stay in our case, $\boldsymbol{\beta}$ represents unknown parameters, $\boldsymbol{X}$ is a vector of explanatory variables, and ϵ is a measurement error independent of $\boldsymbol{X}$. Let S_0 and f_0 denote the survival function and density function, respectively, of e^ϵ. For an individual with a set of covariates $\boldsymbol{X}$, the survival function can be expressed as

$$S_1(t) = S_0\left(t\, e^{\boldsymbol{\beta}^\top \boldsymbol{X}}\right). \tag{2}$$

In Equation (2), e^{β_j} measures the extent to which the time-to-failure is accelerated by the j^{th} covariate compared to the baseline, and is thus interpreted as the time acceleration factor. More intuitively, the expression $100(e^{-\beta_j} - 1)$ calculates the percentage change in the median (or any other quantile) survival time relative to the baseline, providing a direct measure of the covariate's impact on survival time.

To allow for time-dependent covariates, we consider the following extension introduced in Cox and Oakes (1984, p. 67)

$$e^\epsilon = \int_0^T e^{\boldsymbol{\beta}^\top \boldsymbol{X}(t)}\, \mathrm{d}t.$$

Let $S_{T|X}(t)$ and $f_{T|X}(t)$ represent the conditional survival and density functions, respectively, of T given $\boldsymbol{X}$. We have

$$S_{T|X}(t) = \Pr\left(e^\epsilon > \int_0^t e^{\boldsymbol{\beta}^\top \boldsymbol{X}(s)}\, \mathrm{d}s\right) = S_0\left(\int_0^t e^{\boldsymbol{\beta}^\top \boldsymbol{X}(s)}\, \mathrm{d}s\right),$$

and

$$f_{T|X}(t) = f_0\left(\int_0^t e^{\boldsymbol{\beta}^\top \boldsymbol{X}(s)}\, \mathrm{d}s\right) e^{\boldsymbol{\beta}^\top \boldsymbol{X}(t)}.$$

For a random sample of n individuals, the data consists of $(Y_i, \Delta_i, \boldsymbol{X}_i(\cdot)), i = 1, \ldots, n$, where $Y_i = \min(T_i, C_i)$, $\Delta_i = I(T_i \leq C_i)$, and $I(\cdot)$ is the indicator function. We use the maximum likelihood estimation method to estimate the parameters. The likelihood function is given by

$$\prod_{i=1}^{n} \left[f_{T|X_i}(Y_i)\right]^{\Delta_i} \left[S_{T|X_i}(Y_i)\right]^{1-\Delta_i}. \tag{3}$$

In our study, the time-dependent covariates are considered piecewise constant because each transfer of a resident is counted as a new episode, thereby generating a new record. Consequently, the integrals in Equation (3) are straightforward to evaluate.

5 Estimating accelerated failure time models

5.1 Distribution fitting

We first find a suitable distribution to model the length of stay. Table 7 shows the generalized F distribution provides the best goodness-of-fit among all the distributions considered. Hence, it will be the distribution we use in our analysis.

Table 7. Distribution fitting to the length of stay data: number of parameters (nPars), the maximized value of the log-likelihood function (Loglik), Akaike information criterion (AIC), and Bayesian information criterion (BIC) values. The lowest AIC and BIC values are in bold.

Distribution	nPars	Loglik	AIC	BIC
Log-normal	2	-232,457	464,918	464,935
Log-logistic	2	-231,774	463,553	463,570
Exponential	1	-227,755	455,513	455,522
Weibull	2	-226,947	453,898	453,916
Gamma	2	-226,671	453,345	453,363
Generalized gamma	3	-226,292	452,589	452,616
Generalized F	4	-226,278	**452,564**	**452,599**

To investigate why the generalized F distribution outperforms other models, we compare hazard rates estimated from the three parametric models with the lowest AIC values against those obtained via the kernel-based method introduced in Muller and Wang (1994). These comparisons are depicted in Figure 6. The two-parameter gamma distribution appears to lack the flexibility needed to accurately model hazard rates over varying durations: it tends to overestimate the hazard rate in the short to medium term and underestimate it in the long term. Moreover, the generalized gamma and the generalized F distributions exhibit similar performance during the first ten years. Beyond this period, however, the generalized gamma distribution starts to overestimate the hazard rate. The comparison indicates that the generalized F distribution provides the best goodness-of-fit due to its ability to capture the hazard rate dynamics for varying lengths of stay.

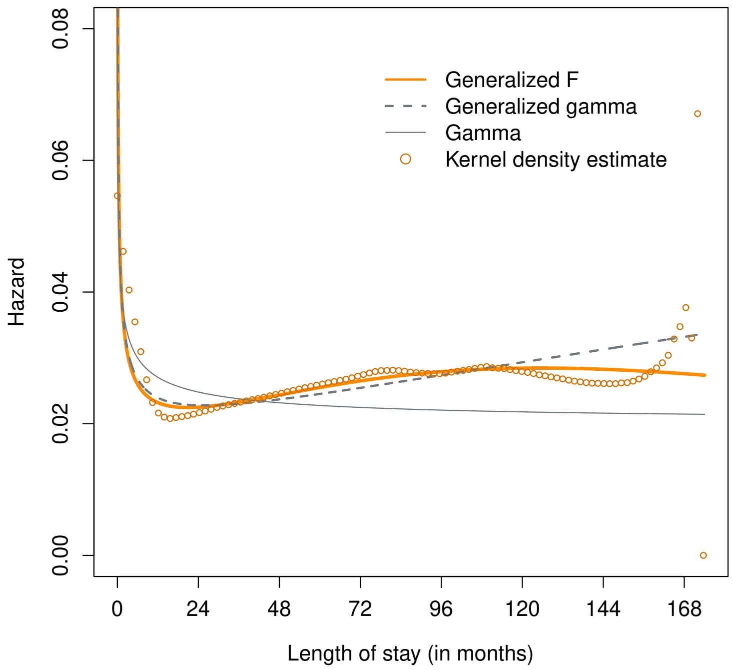

Figure 6. Hazard fitted from alternative parametric models and kernel density estimates.

5.2 Explanatory variables

The process of selecting explanatory variables reveals that the final model includes all covariates listed in Table 6 except for the indigenous indicator. This exclusion is likely due to the very limited presence of indigenous people in our sample, comprising fewer than 300 out of 51,738 residents. Table 8 displays the estimated parameter values and their standard errors. Results from the exploratory data analysis, presented in Table 5, indicate the differences between using time-dependent and time-independent variables. To further evaluate these distinctions, we estimate two models: a time-dependent model using time-varying covariates for SES, organization type, service size, and remoteness recorded at the start of each episode, and a time-independent model using covariate records from the first admission. The coefficients from the two models vary little for time-constant variables. The main differences in estimation results lie in variables that can change over time, which we will discuss in detail later in this section.

Table 8. Estimated parameter values with standard errors in parentheses.

	Time-dependent		Time-independent	
Generalized F distribution				
μ	4.0084***	(0.0153)	4.0112***	(0.0153)
σ	0.8000***	(0.0124)	0.7970***	(0.0125)
Q	1.7949***	(0.0251)	1.8038***	(0.0251)
P	0.5074***	(0.0781)	0.5073***	(0.0793)
Age at the first admission (baseline: 85-89)				
50-54	-1.0267***	(0.0721)	-1.0337***	(0.0718)
55-59	-0.9718***	(0.0470)	-0.9771***	(0.0469)
60-64	-0.8859***	(0.0360)	-0.8890***	(0.0360)
65-69	-0.6577***	(0.0253)	-0.6552***	(0.0252)
70-74	-0.4305***	(0.0181)	-0.4304***	(0.0181)
75-79	-0.2775***	(0.0135)	-0.2779***	(0.0135)
80-84	-0.1381***	(0.0111)	-0.1387***	(0.0110)
90-94	0.1576***	(0.0126)	0.1570***	(0.0125)
95-99	0.3578***	(0.0212)	0.3554***	(0.0211)
100+	0.6737***	(0.0613)	0.6690***	(0.0609)
Gender (baseline: female)				
Male	0.3626***	(0.0089)	0.3616***	(0.0089)
Country of birth (baseline: Australia)				
Other	-0.0258**	(0.0105)	-0.0271***	(0.0105)
Preferred language (baseline: English)				
Other	-0.0256	(0.0163)	-0.0298*	(0.0163)
Socioeconomic status (baseline: Q5)				
Q1	0.0266	(0.0224)	0.0364	(0.0223)
Q2	-0.0123	(0.0149)	-0.0188	(0.0148)
Q3	0.0210	(0.0156)	0.0128	(0.0156)
Q4	-0.0082	(0.0111)	-0.0090	(0.0110)
Organization type (baseline: Not-for-profit)				
Government	0.2214***	(0.0204)	0.1556***	(0.0202)
Private	0.1692***	(0.0091)	0.1490***	(0.0091)
Service size (baseline: 100+)				
0-20	-0.1994***	(0.0403)	-0.0913***	(0.0341)
21-40	-0.1107***	(0.0176)	-0.0431***	(0.0160)
41-60	-0.0716***	(0.0130)	-0.0210*	(0.0123)
61-80	0.0015	(0.0120)	0.0124	(0.0119)
81-100	0.0010	(0.0120)	0.0132	(0.0121)
Remoteness (baseline: Major Cities)				
Inner Regional	0.0566***	(0.0134)	0.0491***	(0.0134)
Outer Regional/Remote/Very Remote	0.0234	(0.0187)	0.0138	(0.0185)
State/Territory (baseline: New South Wales)				
Victoria	-0.0384***	(0.0114)	-0.0339***	(0.0114)
Queensland	-0.0188	(0.0122)	-0.0177	(0.0121)
Western Australia	0.0224	(0.0164)	0.0142	(0.0163)
South Australia	0.0144	(0.0156)	0.0086	(0.0155)
Tasmania	0.0928***	(0.0264)	0.0898***	(0.0263)
Australian Capital Territory	0.0141	(0.0407)	0.0089	(0.0407)
Northern Territory	-0.0077	(0.0960)	-0.0117	(0.0965)

Note: Significance levels $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

For a more intuitive explanation of the estimate parameter values, we plot the estimated acceleration factors, $\exp(\hat{\beta}_j)$, along with their 95% confidence intervals from Figure 7 to Figure 10. These plots are based on the estimation results from the time-dependent model unless otherwise stated. An acceleration factor greater than one indicates that the effect of the covariate acts to shorten the length of stay compared to the baseline.

5.2.1 Individual characteristics

Age at the first admission has a strong influence on the length of stay, with older residents tending to have shorter stays. This result is statistically significant across all age groups, ranging from the youngest at 50-54 to the oldest at 100+. Despite the Kaplan-Meier estimates in Table 4 suggesting that residents in the 50-54 age group have a shorter median length of stay than those in the 55-59 and 60-64 groups, this discrepancy is likely due to an interaction effect that the Kaplan-Meier method cannot adequately control. The age effect aligns with expectations as older ages are typically associated with higher mortality rates. Similar trends have been observed in other countries (see e.g., Fuino and Wagner, 2020; Hoben et al., 2019), as well as in Australia using an older dataset (Liu, 1996).

Figure 7 indicates that among the youngest three age groups considered, the differences in length of stay are minor, as evidenced by overlapping 95% confidence intervals of the acceleration factors. However, beginning with the 65-69 age group, the age differences become significant. Given that age 65 is the eligibility threshold for government-subsidized aged care services for non-indigenous Australians, our findings highlight the need for age-specific planning in anticipating the requirements for permanent residential care.

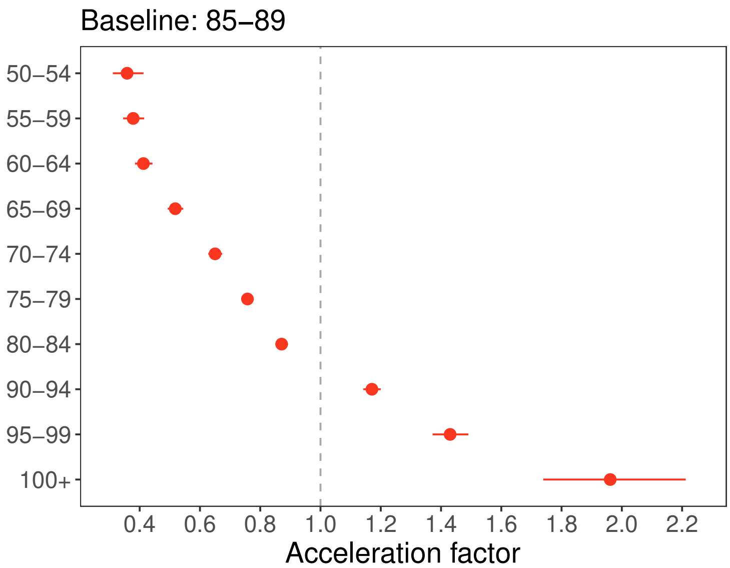

Figure 7. The estimated acceleration factor for the covariate of age at the first admission. The point represents the mean estimate, and the line represents the corresponding 95% confidence interval.

The impact of gender on the length of stay in nursing homes is well-documented in the literature (see e.g., Breuer et al., 1998; Fuino and Wagner, 2020; Hoben et al., 2019; Kelly et al., 2010; Liu and Manton, 1983). Consistent with previous studies, our results indicate that men typically have shorter stays than women, likely due to differences in mortality rates between genders. Additionally, our analysis reveals that foreign-born residents tend to have longer stays than their Australian-born counterparts. This observation aligns with findings from Canadian nursing home data, which also report longer lengths of stay for foreign-born individuals (Hoben et al., 2019). Such differences can be attributed to mortality differentials, with immigrants in Australia generally exhibiting longer life expectancy than the native-born population (Huang et al., 2023).

However, when examining the impact of language preferences, we find that the difference in length of stay between English and non-English speakers was not statistically significant at the 5% level. The left panel of Figure 8 indicates that while the point estimates for country of birth and preferred language are similar, the confidence interval for language preference is wider. The difference can be explained by more imbalanced data: although more than a quarter of residents were born overseas, less than a tenth preferred not to speak English.

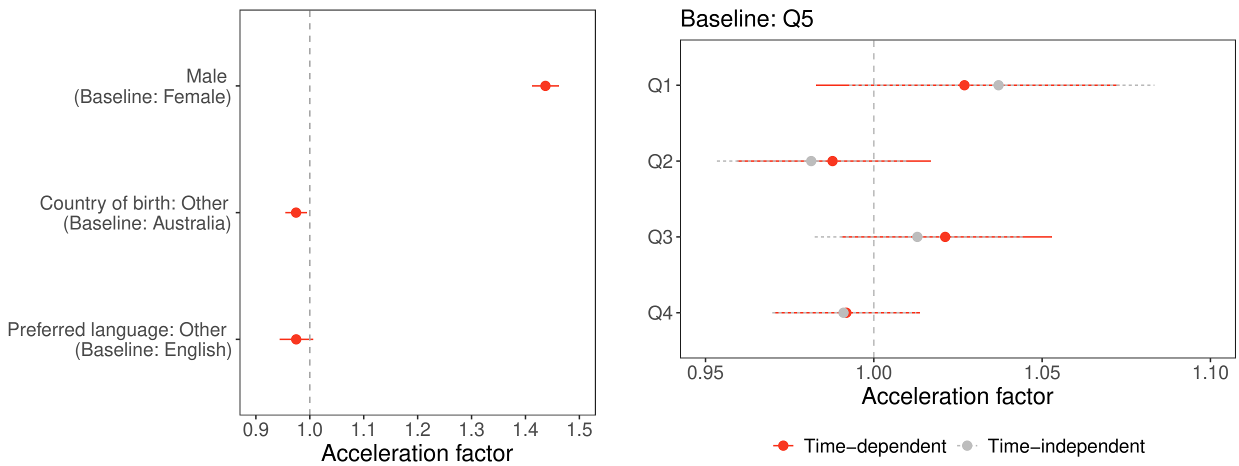

Figure 8. The estimated acceleration factor for the covariates: (Left Panel) gender, country of birth, preferred language, and (Right Panel) socioeconomic status. The point represents the mean estimate, and the line represents the corresponding 95% confidence interval.

We find no statistically significant impact of SES on the length of stay, a result that is not unexpected given our reliance on area-level rather than individual-level SES measures. Prior research demonstrating significant impacts typically utilizes individual-level measures, such as pre-retirement income (Fuino and Wagner, 2020) or household net worth (Kelly et al., 2010). Area-level SES proxies are known for their interpretive challenges and are less effective at capturing individual socioeconomic characteristics (Geronimus et al., 1996). For instance, using Australian SEIFA data, similar to our data source, Walker and Becker (2005) find that these proxies have considerably lower explanatory power compared to individual-based SES indicators. Furthermore, area-level SES measures represent the collective socioeconomic characteristics of populations within specific regions, which may not accurately reflect the demographics of the older age groups of interest in our study. Such mismatches could introduce biases that lead to seemingly unexpected results in our analysis.

5.2.2 Residential aged care facility characteristics

We find that facility characteristics such as the organization type and service size significantly impact the length of stay in nursing homes. Importantly, neglecting to account for changes in these characteristics can lead to underestimations of their effects, as demonstrated in Figure 9. Residents in government- and privately-owned nursing homes typically experience shorter stays than those in not-for-profit organizations, with the difference being especially pronounced in government-owned facilities. This pattern aligns with findings from the U.S., where non-profit facilities are associated with lower mortality and fewer infections compared to for-profit facilities (Spector et al., 1998). Moreover, overlooking changes in organization type can underestimate the difference in median length of stay between not-for-profit and government-owned facilities by approximately eight percentage points, reducing it from 24.8% to 16.8%.

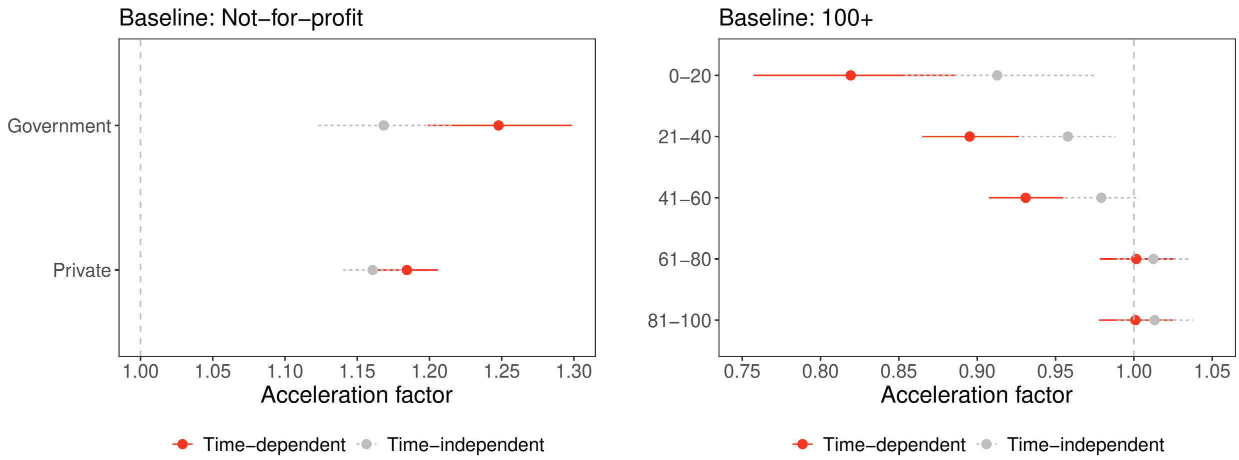

Figure 9. The estimated acceleration factor for the covariates: (Left Panel) organization type, and (Right Panel) service size. The point represents the mean estimate, and the line represents the corresponding 95% confidence interval.

In terms of service size, residents in smaller facilities, particularly those with fewer than 60 beds, tend to experience longer stays. In contrast, differences in length of stay among medium-to-large-size facilities are not statistically significant. Ignoring changes in service size leads to an underestimation of these extended stays, which are most notable in the smallest facilities with fewer than 20 beds. Most research on nursing home sizes indicates that smaller facilities often provide better resident outcomes and quality of care (Baldwin et al., 2017). Consistent with these findings, our study shows that smaller facilities are associated with longer lengths of stay, possibly reflecting a more favorable living environment.

The location of nursing homes has some impact on the length of stay. Facilities outside Major Cities typically host their residents for a shorter duration, with the difference being statistically significant for Inner Regional Australia. Since changes in remoteness are uncommon, neglecting such changes results in only a minor impact, as illustrated in the left panel of Figure 10. In terms of state and territory variations, residents in Tasmania usually have shorter stays than those in the rest of Australia, a result consistent with the Kaplan-Meier estimates presented in Table 4. In contrast, residents in Victoria tend to have slightly longer stays, which are statistically significant compared to those in New South Wales. Although the Kaplan-Meier estimates initially suggest the longest median length of stay in the Northern Territory, this observation changes once we control for other covariates such as age. However, the right panel of Figure 10 shows that the confidence interval remains large, reflecting the limited sample size from the Northern Territory.

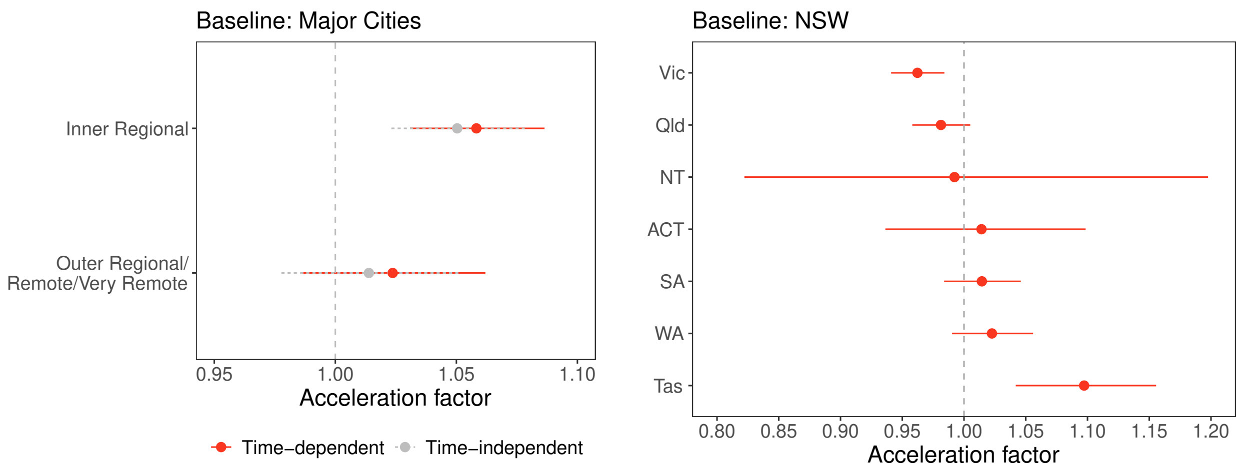

Figure 10. The estimated acceleration factor for the covariates: (Left Panel) remoteness, and (Right Panel) state/territory. The point represents the mean estimate, and the line represents the corresponding 95% confidence interval.

5.3 Prediction

Having discussed the estimated parameter values and acceleration factors, we further illustrate the effect of each covariate using the predicted length of stay for hypothetical individuals at their first admission to nursing homes. These individuals and the facilities they enter are characterized by various covariates included in our AFT model. We specifically alter the values of covariates that have shown a statistically significant impact in our regression analysis. By comparing the resulting lengths of stay, we can demonstrate the influence of each covariate on the duration of stay. Given that the indigenous indicator covariate is excluded from our analysis and the minimum age for receiving government-supported aged care is 65 for non-indigenous Australians, our analysis focuses on individuals aged 65 and above at their first admission. All predictions are based on the time-dependent estimation results shown in Table 8, unless otherwise stated.

5.3.1 Individual characteristics

Table 9 presents the median and mean lengths of stay along with their 95% prediction intervals. Considering age and gender as significant predictors, we analyze different combinations of these factors while keeping all other covariates at their baseline values. Both the mean and median are analytically calculated using the generalized F distribution. The prediction intervals are derived from simulations that assume each parameter (or its log transformation for σ and P) follows a normal distribution. We conduct these simulations 1,000 times. The simulation results closely align with the analytical solutions, confirming the validity of our method. Detailed comparison results can be found in Appendix C.

Table 9. Predicted length of stay for each gender and age combination: median, mean, and their corresponding 95% prediction intervals in parentheses. All other covariates are at baseline values.

Age Group	Female		Male		Female		Male	
	Median		Median		Mean		Mean	
65-69	56.7	(53.0, 60.4)	39.4	(37.0, 42.0)	76.7	(72.3, 81.3)	53.4	(50.3, 56.8)
70-74	45.2	(42.8, 47.6)	31.4	(29.7, 33.1)	61.1	(58.3, 64.2)	42.5	(40.5, 44.8)
75-79	38.8	(36.8, 40.8)	27.0	(25.6, 28.4)	52.5	(50.4, 55.0)	36.5	(34.9, 38.4)
80-84	33.7	(32.2, 35.3)	23.5	(22.3, 24.6)	45.6	(43.9, 47.7)	31.8	(30.5, 33.3)
85-89	29.4	(28.1, 30.6)	20.4	(19.5, 21.3)	39.8	(38.5, 41.2)	27.7	(26.6, 28.8)
90-94	25.1	(23.8, 26.3)	17.5	(16.6, 18.3)	34.0	(32.6, 35.5)	23.6	(22.6, 24.8)
95-99	20.5	(19.4, 21.8)	14.3	(13.4, 15.2)	27.8	(26.4, 29.4)	19.3	(18.3, 20.5)
100+	15.0	(13.2, 17.0)	10.4	(9.1, 11.8)	20.3	(17.9, 22.9)	14.1	(12.4, 16.0)

Table 9 reveals significant age and gender disparities in the predicted length of stay. Beginning with the 75-79 age group, advancing to each successive 5-year age bracket decreases the median length of stay by approximately four to five months for women and about three months for men. The difference in median length of stay between genders decreases almost linearly with age, dropping from 17.2 months in the 65-69 age group to just 4.6 months in the 100+ age group. Additionally, the mean length of stay is about 35% longer than the median, reflecting the long tail of the distribution.

The effect of country of birth is statistically significant, though the practical impact is minimal. Figure 11 compares the predicted median length of stay between Australian-born and foreign-born residents. The largest observed difference is less than 1.5 months, occurring among women in the 65-69 age group, while the smallest difference is under 0.3 months, noted among men in the 100+ age group.

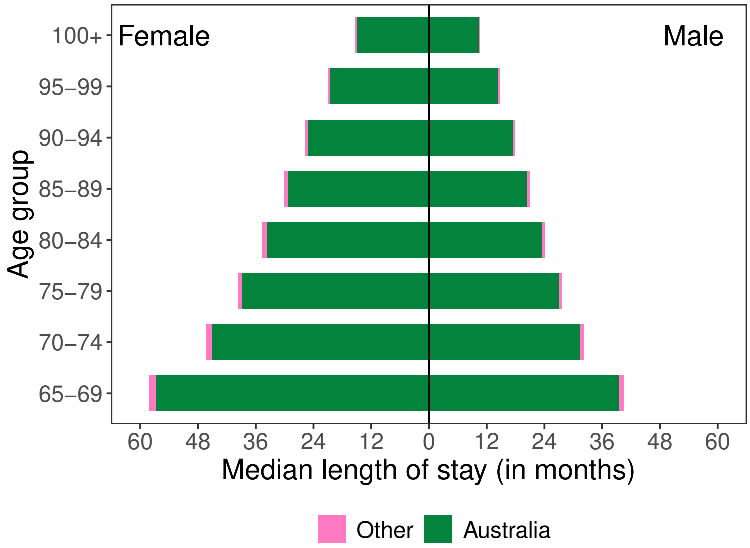

Figure 11. Predicted median length of stay by country of birth. All other covariates not displayed are at baseline values.

5.3.2 Residential aged care facility characteristics

We have found that the type of organization operating a nursing home significantly influences the length of stay. Not accounting for potential changes in organization type due to transfers between nursing homes can lead to an underestimation of this impact. Figure 12 illustrates these effects on the median length of stay. The left panel, which uses time-dependent estimation to consider changes over time, shows differences in median length of stay between not-for-profit and government-owned facilities ranging from 3.0 to 11.3 months for women. In contrast, the right panel, which uses time-independent estimation and does not account for such changes, shows narrower differences ranging from 2.2 to 8.2 months—an underestimation of one to three months. Similarly, the differences between not-for-profit and privately-owned facilities for women range from 2.3 to 8.8 months in the left panel and from 2.1 to 7.8 months in the right panel. Overall, the right panel consistently shows narrower gaps between different organization types across both genders and all age groups.

Figure 13 demonstrates the impact of service size on the predicted median length of stay. Similar to organization type, we present two plots to highlight the importance of accounting for changes in service size. Consider the baseline age group of 85-89, for instance. The left panel, using time-dependent estimation, shows that the median length of stay for women staying in facilities with fewer than 20 beds is close to 36 months, and for men, it exceeds 24 months. In contrast, the right panel, which employs time-independent estimation, shows shorter stays of 32 months for women and 22 months for men. This underestimation of stay length in smaller-sized facilities is more pronounced in younger age groups.

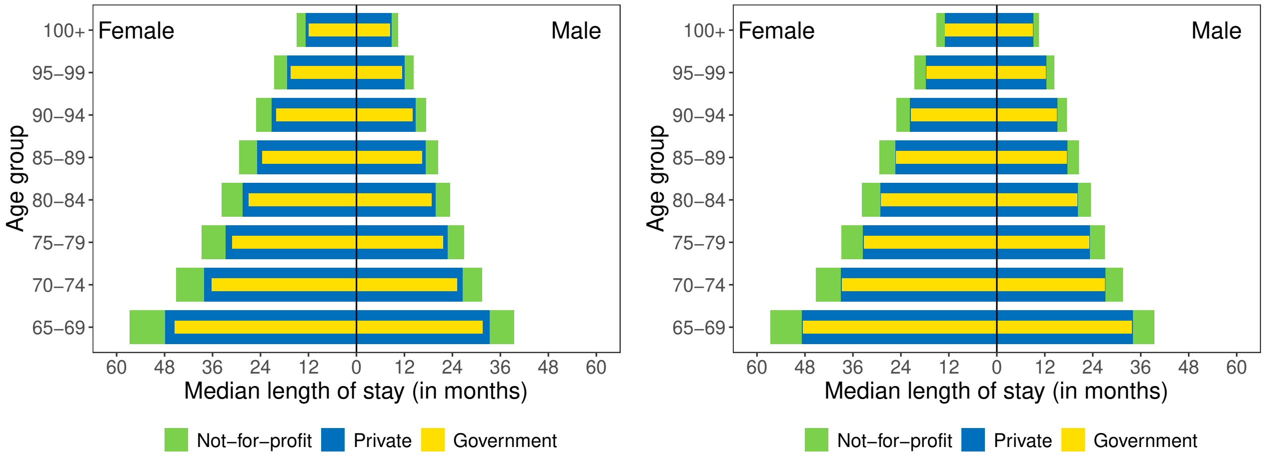

Figure 12. Predicted median length of stay by organization type: (Left Panel) using time-dependent estimation results; (Right Panel) using time-independent estimation results. All other covariates not displayed are at baseline values.

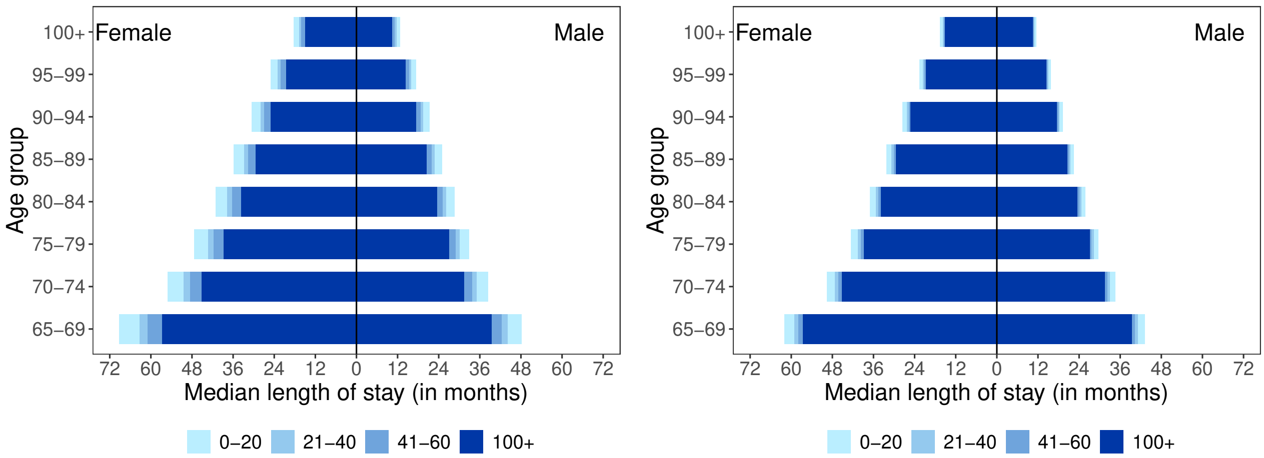

Figure 13. Predicted median length of stay by service size: (Left Panel) using time-dependent estimation results; (Right Panel) using time-independent estimation results. All other covariates not displayed are at baseline values.

Figure 14 illustrates the impact of nursing home locations on the predicted median length of stay. Residents in Inner Regional Australia are predicted to have a slightly shorter median length of stay compared to those in Major Cities, with differences ranging from 0.8 to 3.1 months for women and 0.6 to 2.2 months for men. State-wise, residents in Tasmania experience substantially shorter stays compared to New South Wales, with differences in the median length of stay being as much as 5.0 months for women and 3.5 months for men. Conversely, residents in Victoria are predicted to have median stays that are one to three months longer than their counterparts in New South Wales.

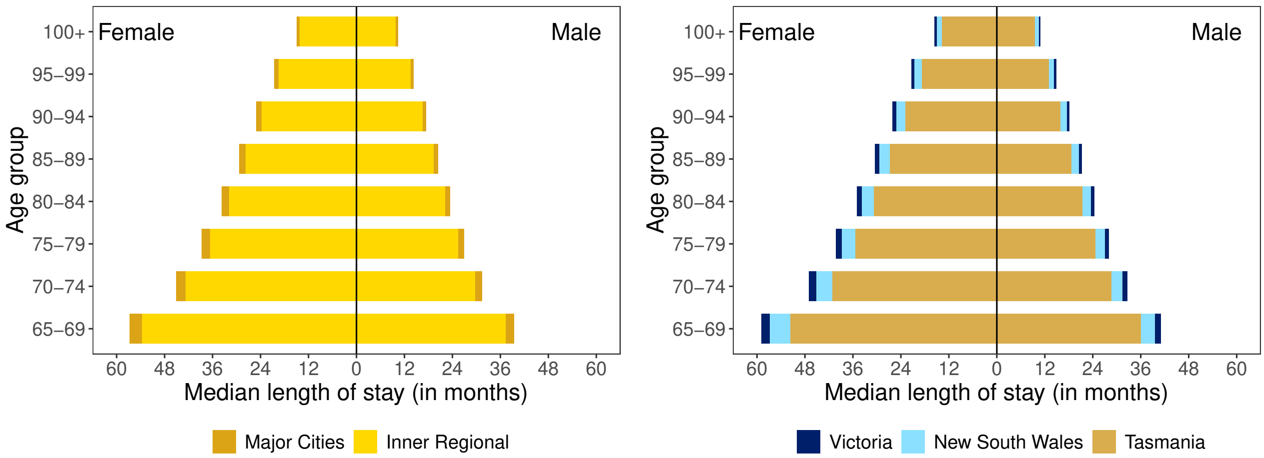

Figure 14. Predicted median length of stay by (Left Panel) remoteness and (Right Panel) state/territory. All other covariates not displayed are at baseline values.

6 Conclusion and discussion

We model the length of stay in permanent residential care for older adults in Australia, a timely assessment given the demographic shift toward an aging population. Our study adds the Australian experience to the existing literature, offering a basis for international comparisons and helping to inform accommodation payment decisions for older Australians and their families.

Using the framework of survival analysis, our approach includes several enhancements over similar applications. We incorporate time-dependent covariates to uncover their significance and use the full sample in our model estimation to enhance the reliability of the results. Additionally, we explore various families of distributions to model the length of stay, finding that the four-parameter generalized F distribution provides a superior fit compared to other distributions within the generalized F family.

Our study highlights the significant impact that organizational characteristics of nursing homes—specifically their type and service size—have on the duration of stay. These factors, alongside resident demographics, emerge as key determinants of length of stay. Crucially, our study also identifies a significant methodological gap: failing to account for transfers between nursing homes can lead to a substantial underestimation of the influence of these variables on the length of stay.

The risk associated with permanent residential care stems from the discrepancy between expected and actual lengths of stay. This difference arises from idiosyncratic risk, which can be diversified through risk pooling, and systematic risk, which is non-diversifiable. The government is usually better positioned to absorb systematic risks compared to the private market. However, with an aging population and a growing number of dependent elderly, the government is facing increasing pressure in financing related care. This limits its capacity to absorb systematic risks effectively. Consequently, there is a growing need for private long-term care insurance.

Developing private markets for long-term care insurance would help establish a multi-pillar aged care system, similar to multi-pillar pension systems, thereby enhancing the resilience and sustainability of aged care provision. In an ideal system, individuals with moderate wealth would likely exhibit a high demand for private long-term care insurance. In contrast, those with higher net worth, especially those with substantial home equity, could afford to self-insure. Meanwhile, individuals with limited financial resources could depend on the government's safety net. This diversified approach ensures that all segments of society have appropriate options for managing long-term care risks.

Insurance product designs depend on relevant data and an appropriate modeling framework. While Markov models are versatile for modeling multi-state health transitions, their application in long-term care insurance requires individual-level longitudinal data for the elderly population. Unfortunately, such data are often unavailable in many countries, including Australia. In this context, we demonstrate that the AFT model is a practical alternative when only administrative data are accessible. Using the AFT model, we illustrate how individual and nursing home heterogeneity impacts the length of stay, providing valuable insights for product designs.

Future research can greatly benefit from data refreshment and data linkage. Internationally, with the promotion of aging in place, trends indicate a reduction in the length of stay in nursing homes, as observed in Canada (Hoben et al., 2019) and Sweden (Schön et al., 2016). Recent findings by Fuino and Wagner (2020) suggest that individuals who received home care prior to nursing home admission tend to have shorter institutional stays. Given the expansion of home care packages following the 2012 LLLB reforms and the 2017 Increasing Choice in Home Care reforms in Australia, a shorter average length of stay in residential care is likely. Therefore, analyzing data from younger cohorts could help assess the impact of these policy reforms.

Regarding data linkage, our current study utilizes an area-level SES measure, which has not been a statistically significant determinant of length of stay. Employing individual-level SES measures could potentially yield different results. Furthermore, studies have indicated that the level of dependency at entry is closely associated with the length of stay in nursing homes (Hoben et al., 2019). Therefore, linking length of stay data with health assessment records could offer valuable insights into how entry conditions affect outcomes in residential care.

Acknowledgments

The authors acknowledge financial support from the Australian Research Council Centre of Excellence in Population Ageing Research (CEPAR) project number CE170100005.

References

AIHW (2023) Aged care Glossary. Web page. Canberra, Australia: Australian Institute of Health and Welfare. Available at: `https://www.aihw.gov.au/reports-data/health-welfare-services/aged-care/glossary` [Accessed 28/02/2024].

Alders, P. and Schut, F. T. (2019) Trends in ageing and ageing-in-place and the future market for institutional care: scenarios and policy implications. *Health Economics, Policy and Law*, **14**(1), 82–100.

Allers, K. and Hoffmann, F. (2018) Mortality and hospitalization at the end of life in newly admitted nursing home residents with and without dementia. *Social Psychiatry and Psychiatric Epidemiology*, **53**(8), 833–839.

Alonso-García, J., Bateman, H., Bonekamp, J., van Soest, A. and Stevens, R. (2022) Saving preferences after retirement. *Journal of Economic Behavior & Organization*, **198**, 409–433.

Australian Bureau of Statistics (2021) Socio-Economic Indexes for Areas (SEIFA), Australia methodology. Internet. Canberra, Australia: ABS. Available at: `https://www.abs.gov.au/methodologies/socio-economic-indexes-areas-seifa-australia-methodology/2021` [Accessed 31/05/2024].

Baldwin, R., Chenoweth, L., dela Rama, M. and Wang, A. Y. (2017) Does size matter in aged care facilities? A literature review of the relationship between the number of facility beds and quality. *Health Care Management Review*, **42**(4), 315–327. JSTOR: `48516383`.

Breuer, B., Wallenstein, S., Feinberg, G., Camargo, M.-J. F. and Libow, L. S. (1998) Assessing life expectancies of older nursing home residents. *Journal of the American Geriatrics Society*, **46**(8), 954–961.

Ciampi, A., Hogg, S. A. and Kates, L. (1986) Regression analysis of censored survival data with the generalized F family-An alternative to the proportional hazards model. *Statistics in Medicine*, **5**(1), 85–96.

Connolly, M. J., Broad, J. B., Boyd, M., Kerse, N. and Gott, M. (2014) Residential aged care: The de facto hospice for New Zealand's older people. *Australasian Journal on Ageing*, **33**(2), 114–120.

Cox, C. (2008) The generalized F distribution: An umbrella for parametric survival analysis. *Statistics in Medicine*, **27**(21), 4301–4312.

Cox, D. R. and Oakes, D. (1984) *Analysis of Survival Data*. London: Chapman and Hall.

Fuino, M. and Wagner, J. (2020) Duration of long-term care: Socio-economic factors, type of care interactions and evolution. *Insurance: Mathematics and Economics*, **90**, 151–168.

Geronimus, A. T., Bound, J. and Neidert, L. J. (1996) On the validity of using census geocode characteristics to proxy individual socioeconomic characteristics. *Journal of the American Statistical Association*, **91**(434), 529–537.

Grambsch, P. M. and Therneau, T. M. (1994) Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, **81**(3), 515–526. JSTOR: `2337123`.

Hedinger, D., Hämmig, O. and Bopp, M. (2015) Social determinants of duration of last nursing home stay at the end of life in Switzerland: a retrospective cohort study. *BMC Geriatrics*, **15**(1), 114.

Hoben, M., Chamberlain, S. A., Gruneir, A., Knopp-Sihota, J. A., Sutherland, J. M., Poss, J. W., Doupe, M. B., Bergstrom, V., Norton, P. G., Schalm, C., McCarthy, K., Kashuba, K., Ackah, F. and Estabrooks, C. A. (2019) Nursing home length of stay in 3 Canadian health regions: Temporal trends, jurisdictional differences, and associated factors. *Journal of the American Medical Directors Association*, **20**(9), 1121–1128.

Huang, G., Guo, F., Cheng, Z., Liu, L., Zimmermann, K. F., Taksa, L., Tani, M. and Franklin, M. (2023) Nativity in the healthy migrant effect: Evidence from Australia. *SSM - Population Health*, **23**, 101457.

Kaplan, E. L. and Meier, P. (1958) Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, **53**(282), 457–481.

Keeler, E. B., Kane, R. L. and Solomon, D. H. (1981) Short- and long-term residents of nursing homes. *Medical Care*, **19**(3), 363.

Kelly, A., Conell-Price, J., Covinsky, K., Cenzer, I. S., Chang, A., Boscardin, W. J. and Smith, A. K. (2010) Length of stay for older adults residing in nursing homes at the end of life. *Journal of the American Geriatrics Society*, **58**(9), 1701–1706.

Kemper, P. and Murtaugh, C. M. (1991) Lifetime use of nursing home care. *New England Journal of Medicine*, **324**(9), 595–600.

Khadka, J., Lang, C., Ratcliffe, J., Corlis, M., Wesselingh, S., Whitehead, C. and Inacio, M. (2019) Trends in the utilisation of aged care services in Australia, 2008–2016. *BMC Geriatrics*, **19**(1), 213.

Liu, K. and Manton, K. G. (1983) The length-of-stay pattern of nursing home admissions. *Medical Care*, **21**(12), 1211–1222. JSTOR: `3764553`.

Liu, Z. (1996) Length of stay in Australian nursing homes. Aged Care Series no. 1. Canberra: Australian Institute of Health and Welfare. Available at: `https://www.aihw.gov.au/reports/aged-care/length-of-stay-in-australian-nursing-homes/summary` [Accessed 17/06/2022].

Martikainen, P., Moustgaard, H., Einiö, E. and Murphy, M. (2014) Life expectancy in long-term institutional care by marital status: Multistate life table estimates for older Finnish men and women. *The Journals of Gerontology: Series B*, **69**(2), 303–310.

McCann, M., O'Reilly, D. and Cardwell, C. (2009) A census-based longitudinal study of variations in survival amongst residents of nursing and residential homes in Northern Ireland. *Age and Ageing*, **38**(6), 711–717.

Moore, D. C., Keegan, T. J., Dunleavy, L. and Froggatt, K. (2019) Factors associated with length of stay in care homes: a systematic review of international literature. *Systematic Reviews*, **8**(1), 56.

Moss, J. L., Johnson, N. J., Yu, M., Altekruse, S. F. and Cronin, K. A. (2021) Comparisons of individual- and area-level socioeconomic status as proxies for individual-level measures: evidence from the Mortality Disparities in American Communities study. *Population Health Metrics*, **19**(1), 1–10.

Muller, H.-G. and Wang, J.-L. (1994) Hazard rate estimation under random censoring with varying kernels and bandwidths. *Biometrics*, **50**(1), 61–76. JSTOR: 2533197.

OECD (2023) Long-term care resources and utilisation: Long-term care recipients. OECD Health Statistics 2023. Paris, France: Organisation for Economic Co-operation and Development. Available at: https://stats.oecd.org/index.aspx?queryid=30143 [Accessed 07/06/2024].

Prentice, R. L. (1975) Discrimination among some parametric models. *Biometrika*, **62**(3), 607–614.

Schön, P., Lagergren, M. and Kåreholt, I. (2016) Rapid decrease in length of stay in institutional care for older people in Sweden between 2006 and 2012: results from a population-based study. *Health & Social Care in the Community*, **24**(5), 631–638.

Sherris, M. (2021) On sustainable aged care financing in Australia. *Australian Economic Review*, **54**(2), 275–284.

Spector, W. D., Selden, T. M. and Cohen, J. W. (1998) The impact of ownership type on nursing home outcomes. *Health Economics*, **7**(7), 639–653.

Stacy, E. W. (1962) A generalization of the gamma distribution. *The Annals of Mathematical Statistics*, **33**(3), 1187–1192. JSTOR: 2237889.

Stacy, E. W. and Mihram, G. A. (1965) Parameter estimation for a generalized gamma distribution. *Technometrics*, **7**(3), 349–358.

Steventon, A. and Roberts, A. (2012) Estimating length of stay in publicly-funded residential and nursing care homes: a retrospective analysis using linked administrative data sets. *BMC Health Services Research*, **12**(1), 377.

Vossius, C., Selbæk, G., Benth, J. Š. and Bergh, S. (2018) Mortality in nursing home residents: A longitudinal study over three years. *PLOS ONE*, **13**(9), e0203480.

Walker, A. E. and Becker, N. G. (2005) Health inequalities across socio-economic groups: comparing geographic-area-based and individual-based indicators. *Public Health*, **119**(12), 1097–1104.

Welberry, H. J., Brodaty, H., Hsu, B., Barbieri, S. and Jorm, L. R. (2020) Impact of prior home care on length of stay in residential care for Australians with dementia. *Journal of the American Medical Directors Association*, **21**(6), 843–850.e5.

Zhang, J., Shi, Y. and Huang, G. (2023) Expected length of stay at residential aged care facilities in Australia: current and future. *Journal of Population Research*, **40**(4), 24.

Zhao, Y., Li, S. Q., Wilson, T. and Burgess, C. P. (2022) Improved life expectancy for indigenous and non-Indigenous people in the Northern Territory, 1999–2018: overall and by underlying cause of death. *Medical Journal of Australia*, **217**(1), 30–35.

Online Appendix

A Calculating the socioeconomic status score

We calculate a socioeconomic status (SES) score for each Aged Care Planning Region (ACPR) based on the Index of Relative Socio-economic Disadvantage (IRSD), which is one of the four indexes in the Socio-Economic Indexes for Areas (SEIFA). The SEIFA scores are initially calculated at the Statistical Area Level 1 (SA1) and then aggregated to higher levels such as Statistical Area Level 2 (SA2) using population-weighted averages of the SA1 scores (Australian Bureau of Statistics, 2021).

Our dataset contains the 2018 ACPRs, which are based on the SA2 boundaries from the 2016 Australian Statistical Geography Standard by the Australian Bureau of Statistics. Each ACPR comprises multiple SA2s, with each SA2 assigned an individual IRSD score. We aggregate these scores using population-weighted averages to derive a single score for each ACPR.

B Accelerated failure time model assumption

We evaluate the validity of the accelerated failure time (AFT) model assumption using quantile-quantile (Q-Q) plots of the survival function. For the assumption to hold, these plots should ideally show a straight line through the origin. Figure B.1 to Figure B.6 display these plots, where the survival functions are estimated using the Kaplan-Meier method. The red solid lines, representing least-square regression lines through the origin, closely trace the data points in the Q-Q plots. This alignment indicates that the AFT model assumption is generally appropriate for our dataset.

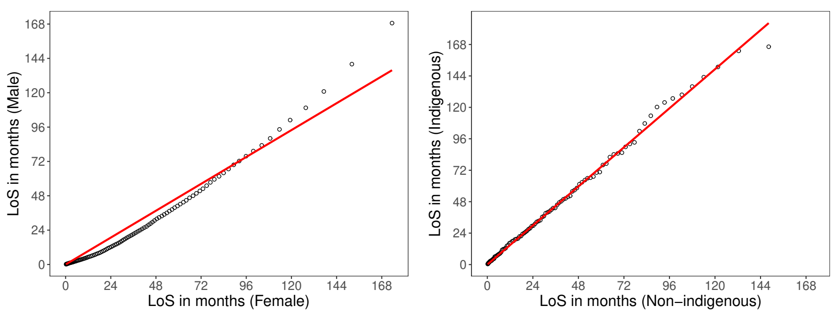

Figure B.1. Graphical test of accelerated failure time model assumption for (Left) gender and (Right) indigenous status: quantile-quantile plots for length of stay (LoS). The red solid line represents a least-square regression line through the origin.

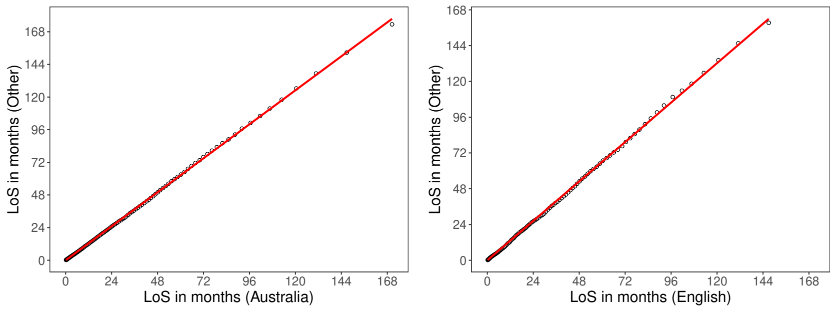

Figure B.2. Graphical test of accelerated failure time model assumption for (Left) country of birth and (Right) preferred language: quantile-quantile plots for length of stay (LoS). The red solid line represents a least-square regression line through the origin.

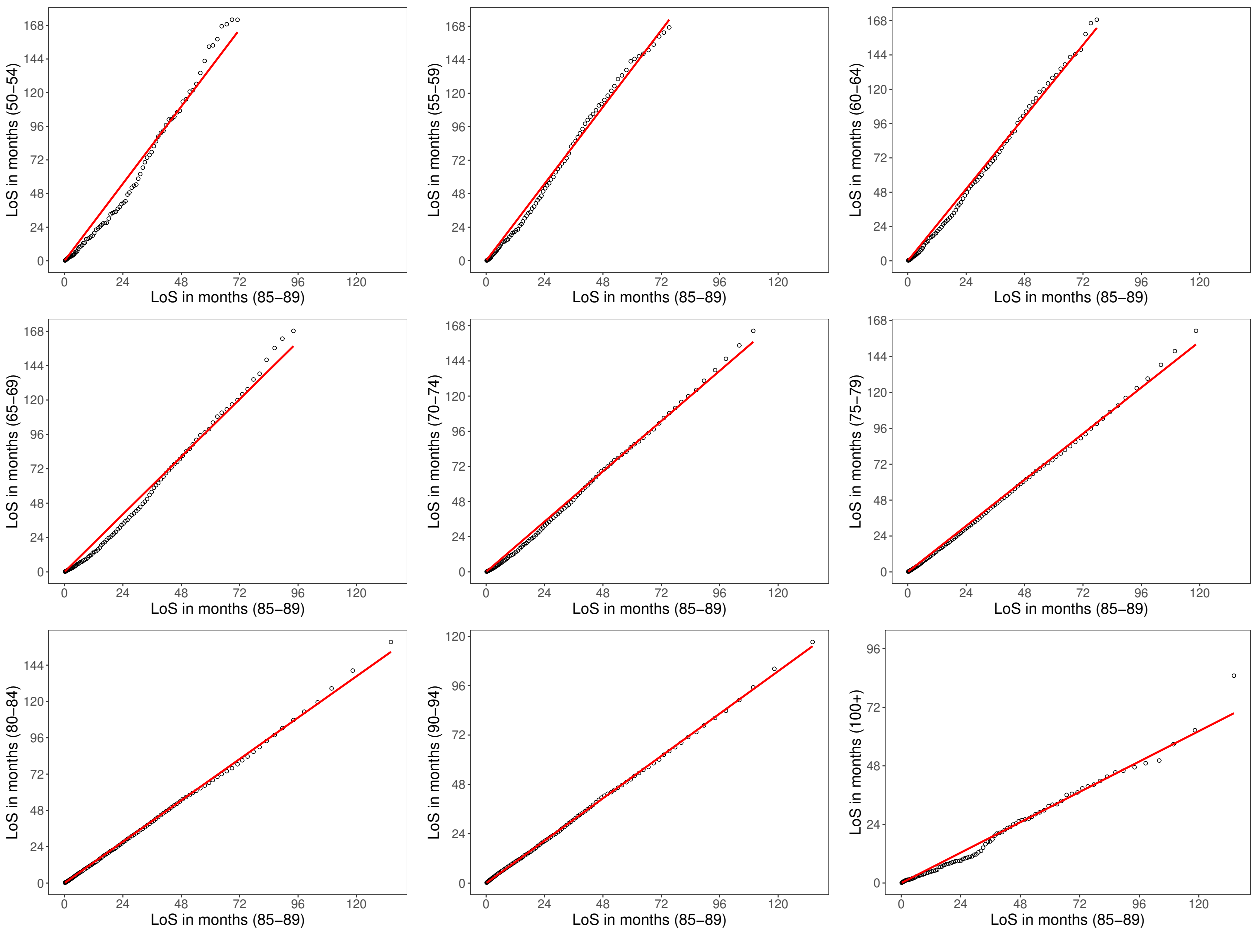

Figure B.3. Graphical test of accelerated failure time model assumption for age at the first admission: quantile-quantile plots for length of stay (LoS). The red solid line represents a least-square regression line through the origin.

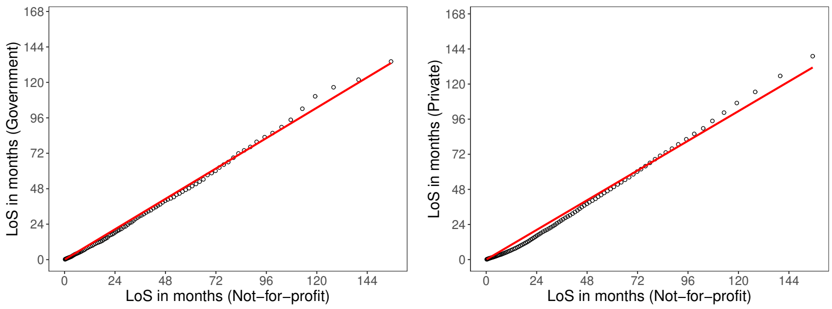

Figure B.4. Graphical test of accelerated failure time model assumption for organization type: quantile-quantile plots for length of stay (LoS). The red solid line represents a least-square regression line through the origin.

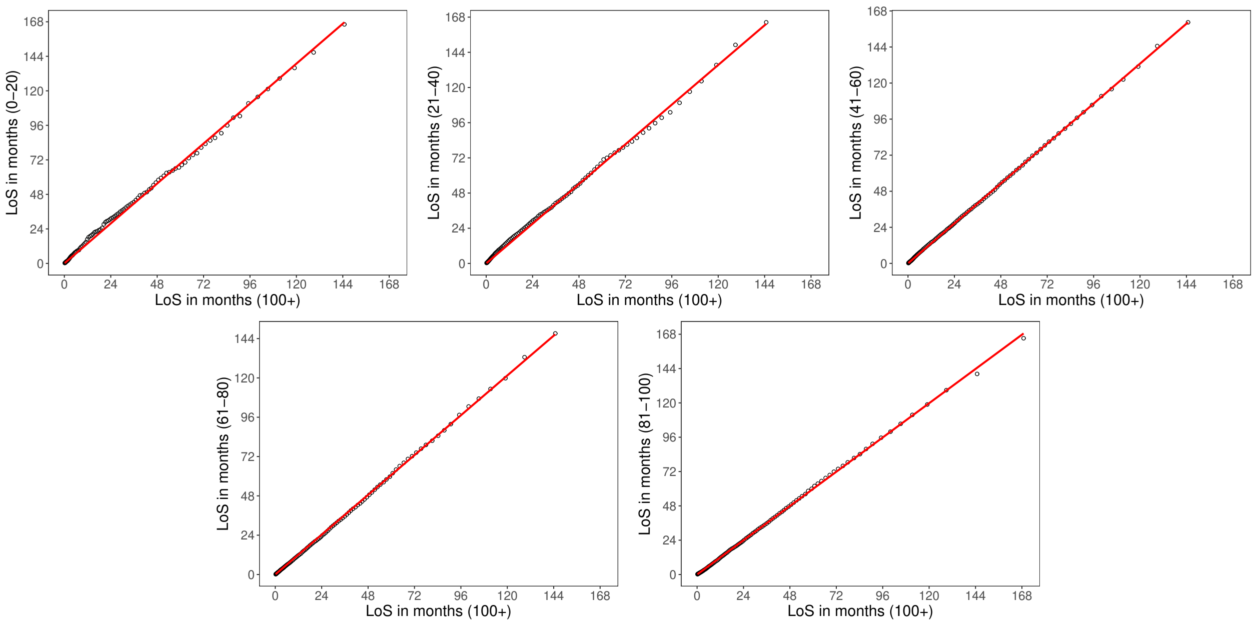

Figure B.5. Graphical test of accelerated failure time model assumption for service size: quantile-quantile plots for length of stay (LoS). The red solid line represents a least-square regression line through the origin.

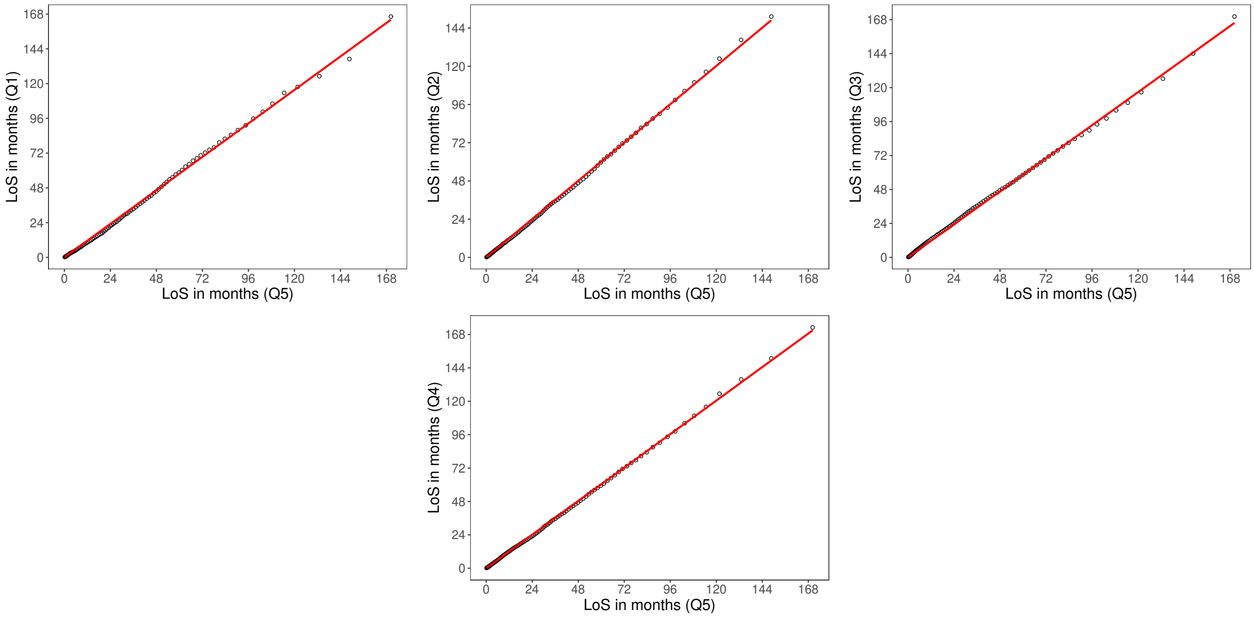

Figure B.6. Graphical test of accelerated failure time model assumption for socioeconomic status: quantile-quantile plots for length of stay (LoS). The red solid line represents a least-square regression line through the origin.

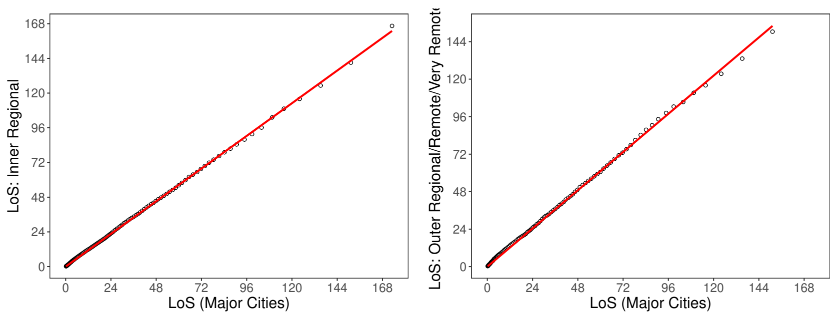

Figure B.7. Graphical test of accelerated failure time model assumption for remoteness: quantile-quantile plots for length of stay (LoS). The red solid line represents a least-square regression line through the origin.

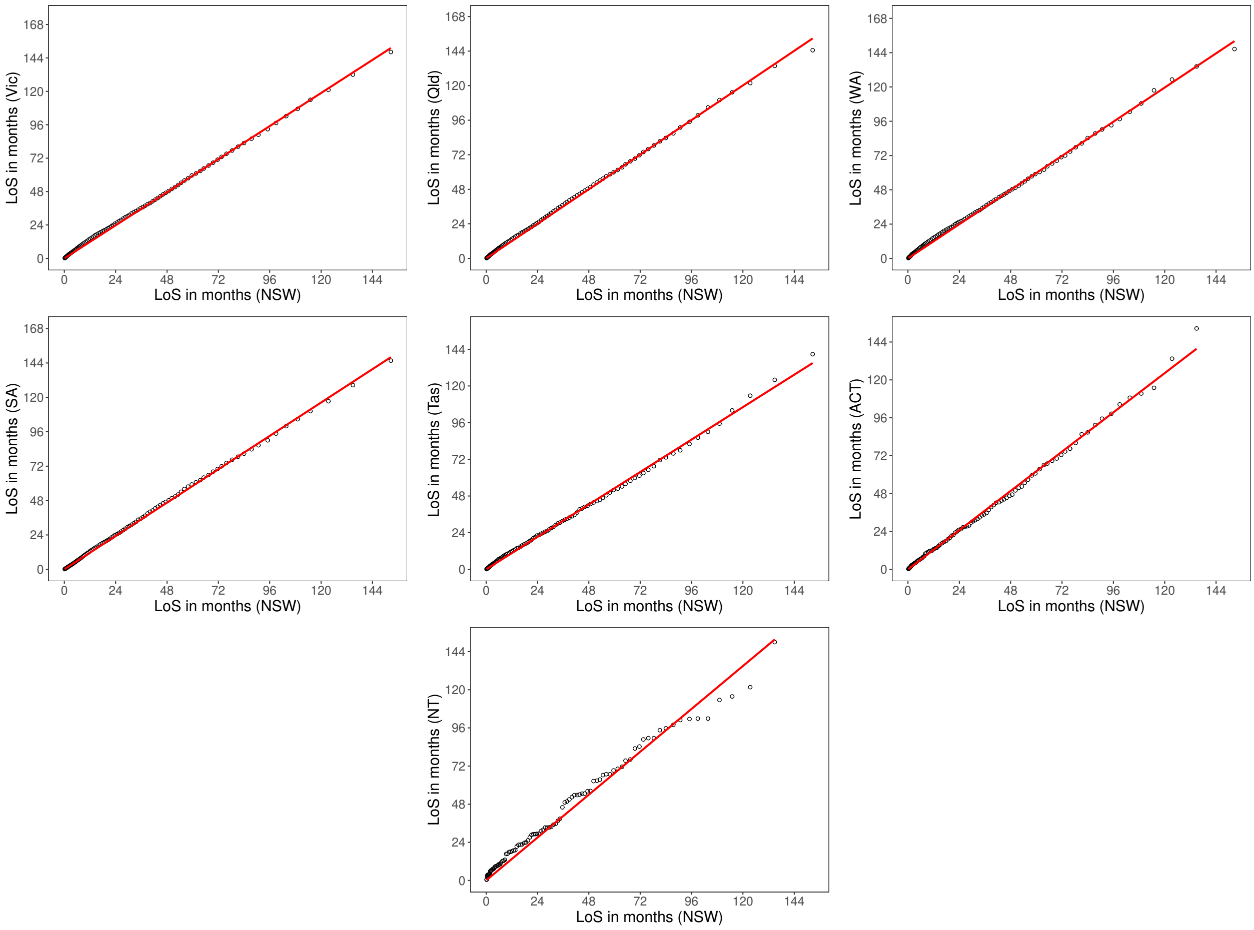

Figure B.8. Graphical test of accelerated failure time model assumption for state/territory: quantile-quantile plots for length of stay (LoS). The red solid line represents a least-square regression line through the origin. States/territories included: New South Wales (NSW), Victoria (Vic), Queensland (Qld), South Australia (SA), Western Australia (WA), Tasmania (Tas), Northern Territory (NT), and Australian Capital Territory (ACT).

C Validating the simulation method

We use the simulation method to derive prediction intervals, as discussed in Section 5.3. Figure C.1 and Figure C.2 display comparisons between the simulated median and mean lengths of stay and their analytical counterparts, using time-dependent and time-independent estimation results from Table 8, respectively. The close alignment of the simulation results with the analytical solutions across both figures confirms the validity of the simulation method.

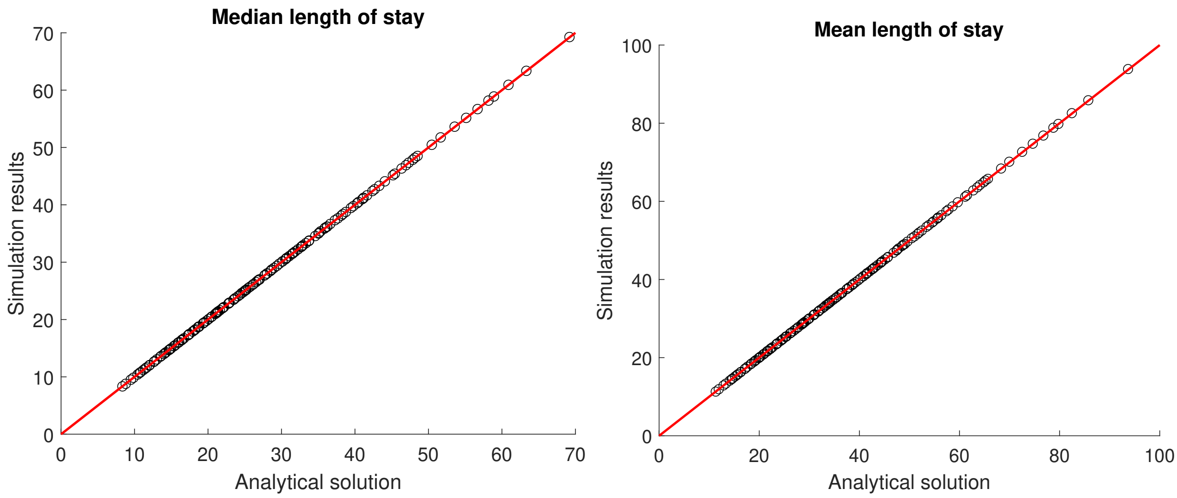

Figure C.1. Comparing simulated median and mean length of stay to their analytical solutions based on the time-dependent estimation results shown in Table 8. The red solid line is a 45-degree line through the origin.

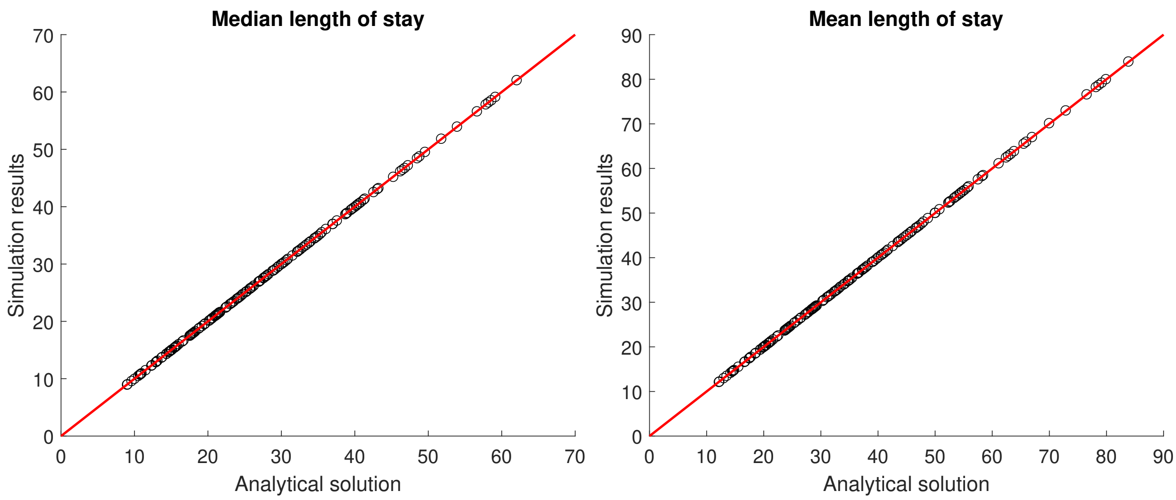

Figure C.2. Comparing simulated median and mean length of stay to their analytical solutions based on the time-independent estimation results shown in Table 8. The red solid line is a 45-degree line through the origin.