International Journal of Population Data Science

Journal Website: www.ijpds.org

Dementia Data for People Living in Care Homes: Cross-sectoral National Data Linkage Study of Care Home, Healthcare and Administrative Data Sources

Kalliopi Mavromati[1], Maria Drelciuc[1,2], Ellen Lynch[3], Sophie A. Hay[1,4], Peter Hanlon[5], Terry J. Quinn[1], and Jennifer Kirsty Burton[1,*]

Submission History

Submitted:	23/09/2025
Accepted:	20/02/2026
Published:	22/04/2026

[1] Academic Geriatric Medicine, School of Cardiovascular and Metabolic Health, College of Medical, Veterinary & Life Sciences, University of Glasgow, UK
[2] Liverpool University Hospitals NHS Foundation Trust, UK
[3] Social Care Analytical Unit, Scottish Government, UK
[4] Undergraduate Medical School, School of Medicine, Dentistry & Nursing, University of Glasgow, UK
[5] General Practice and Primary Care, School of Health and Wellbeing, College of Medical, Veterinary & Life Sciences, University of Glasgow, UK

Abstract

Introduction
Dementia is common among people living in care homes, but many residents lack a diagnosis. Quantifying the prevalence of dementia in care homes is therefore challenging. Using routinely collected data, from care homes, healthcare and administrative data sources offer a potential way to address this, but the utility of these data has not been evaluated in this population.

Objectives
To describe the dementia status of care home residents across multiple national datasets and describe the overlaps and conflicts between sources.

Methods
A retrospective cohort study was undertaken using national care home data from the Scottish Care Home Census. Records for individuals in the cohort were obtained from healthcare (community prescribing, general hospital and psychiatric hospital inpatient diagnoses) and administrative data (death records). These were further categorised for their dementia diagnosis information, including specific dementia subtypes.

Results
A cohort of 63,308 adults living in 1,231 care homes in Scotland between 01/04/2012-31/03/2016 was created. They were 66.8% female and 86.6% died by follow-up, May 2020. Care home data identified 36,275 (57.3%) with dementia. In total, 14,269 (22.5%) had a prior prescription for a dementia medication, 13,920 (22.0%) had a general hospital discharge diagnosis of dementia, 2,750 (4.3%) had a psychiatric hospital discharge diagnosis of dementia and 31,825 (50.3%) had dementia recorded on their death certificate. We found 30.2% of the cohort had no dementia data in any national data sources and 25.3% had dementia data in all three sources and 5.1% only had dementia included in death data.

Conclusions
This study demonstrates the feasibility and utility of linking care home, healthcare and administrative data to better understand dementia in the population living in care homes. The contributions made by different data sources are of interest to those creating population-level datasets to understand dementia epidemiology across the population for researchers, clinicians and policymakers.

Key words
dementia; care home; data linkage; reliability; clinical coding

*Corresponding Author:
Email Address: jenni.burton@glasgow.ac.uk (Jennifer Kirsty Burton)

https://doi.org/10.23889/ijpds.v11i1.3359
April 2026 © The Authors. Open Access under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/deed.en)

Introduction

Dementia is a significant public health challenge [1]. The global population living with dementia is anticipated to rise significantly, due to population ageing and the impact of modifiable risk factors [2]. People living with dementia experience poor health outcomes, such as more frequent hospitalisations [3], accelerated functional decline and reduced quality of life [4]. Over four-fifths are living with multiple long-term conditions [5], which increases the risk of adverse health outcomes and dependency [4, 6]

Care homes play a crucial role supporting people living with dementia [7, 8] However, we do not know the true prevalence of dementia among those living in care homes in the UK. International systematic review data estimates a pooled dementia prevalence of 57.8% among adults living in long-term care settings [9]. However, there is significant heterogeneity in access, population and terminology across care home settings internationally [10, 11]. In addition, it has long been recognised that there is a gap between the numbers of people living in care homes who have received a dementia diagnosis and the numbers living with the condition [12, 13].

Data driven research to improve understanding of dementia among care home residents in the UK is not uncomplicated. There are challenges undertaking care home research, such as staff workload pressures, the availability of proxies and the safe and adequate inclusion of adults with impaired capacity for research participation [14, 15]. Making more effective use of data offers a potential novel approach, however care home residents are under-represented in longitudinal studies of ageing including those monitoring cognition [16]. Identifying care home residents within routinely collected national data is also challenging [17]. Routine data from hospital records, prescribing, primary care data and death registrations can all be used to estimate dementia prevalence [18]. However, the accuracy of dementia diagnoses within routine data is uncertain [19], with limitations such as incomplete recording resulting in underestimation of prevalence [20]. There is also no gold standard method for dementia ascertainment [20].

The Scottish Care Home Census (SCHC) was a national annual data collection open to all care homes in Scotland (irrespective of ownership and funding), which included information about dementia [21, 22]. Care home staff collected individual-level resident data and shared these for inclusion in the SCHC [21]. The utility of dementia data in the SCHC has not been established, nor has it been compared to other sources of dementia diagnosis information. This is an important consideration in the wider national efforts to generate aggregate data on the population living with dementia in Scotland, in a Dementia Index [23], in a difficult to research, but critical population.

To address this research gap, the aims of this study were to:

(i) Describe care home residents recorded in routinely collected national care home data and their dementia status

(ii) Compare the numbers of care home residents with dementia identified from multiple routinely collected national health and administrative data sets (inpatient general hospital and psychiatry data, community prescribing and death records)

(iii) Describe the overlaps and conflicts between routinely collected health, care home and administrative datasets in their dementia information to make recommendations for research and practice

Methods

The study has been reported using REporting of studies Conducted using Observational Routinely collected Data (RECORD) guidance [24], where applicable.

Study Setting/Practice Context

This study is set in Scotland where the majority of in hospital care and prescribing is delivered by the publicly funded National Health System (NHS). Community prescribing data are nationally available and capture records where prescribed medications are available to be dispensed without charge [25]. National death records are comprehensive across the population. Care home stays are funded by individuals and local authorities and inclusion in the care home data is not dependent on funding.

Indexing and Linkage

An in-depth data resource profile on the Scottish Care Home Census (SCHC), including full results of the indexing process are reported elsewhere [26]. In summary, SCHC records needed to be indexed using available personal identifiers to match an individuals' Community Health Index (CHI) number (*national personal identifier variable*) to their SCHC record to enable linkage to other national datasets [27]. If records contained insufficient personal identifiers for this process, they could not be indexed and could not be included in further analyses.

All available financial years of data from 2010/11 to 2015/16 were considered for inclusion. The first two years of data achieved rates of indexing of $<$80%, therefore the decision was made to use data from 2012/13 onwards, for data quality purposes. Indexing rates increased over time, through improved collection of personal identifiers, from 81.9% in 2012/13 to 95.3% in 2015/16 (see Supplementary Table 1).

Data Sources

The following national data sources were used:

- **Community Health Index** (CHI) population spine – based on primary care registration
- **Scottish Care Home Census** (SCHC) – from 2010/11 to 2015/16
- **Scottish Morbidity Records (01/50)** (referred to throughout as SMR01) – General/Acute Inpatient and Day Cases & Geriatric Long Stay – date of hospital admission from 01/04/1981 to 31/03/2016
- **Scottish Morbidity Records (04)** (SMR04) – Inpatient & Day Case Psychiatry – date of hospital admission from 01/04/1981 to 31/03/2016

- **Prescribing Information System** (PIS) – community prescribing – dispensed date from April 2010 to March 2016
- **National Records of Scotland** (NRS) – mortality data – date of death from 01/04/2010 to 31/05/2020

The variables extracted from each of these data sources are described below and in Table 1.

National Care Home Data (Scottish Care Home Census)

The SCHC data were collected annually by care home staff and submitted electronically with the national care regulator, the Care Inspectorate. All care home services were invited to participate aggregate data, about the care home service and individual-level resident data, about long stay residents (intended six weeks or longer). The data were collated and published as a national statistics publication, by Public Health Scotland, based on the routinely collected care home data [21]. Since commencing this research a decision has been taken to pause collection of SCHC data.

Participants

Our complete cohort are adults (aged $\geq$18 years at start of census year) living in a care home in Scotland in each of the defined financial years of interest. This is based on them having a linkable record in the SCHC. Age eligibility was assessed using adjusted date of birth (day of the month removed) compared to age at start of the census year (based on financial year starting on 1st April). Individuals could appear in multiple years of the SCHC and so the latest SCHC record per person was used for analysis.

Our specific interest was in exploring their dementia diagnosis status across the included routinely collected national health, administrative and care home data sources.

Variables

We included variables to describe the care home services, individuals' characteristics and detailed dementia information from each dataset (Table 1).

Care home subtypes are based on registration with the Care Inspectorate [28]. Area-based characteristics are included to describe the distribution of care home services including the national measure for area-based deprivation, the Scottish Index of Multiple Deprivation (SIMD) [29] and the Scottish Government urban/rural classification, based on density of population and accessibility of services [30]. Both use postcode to assign their categories. SIMD rank for each care home was provided in the dataset. This was converted to SIMD quintile using the lookup-key provided. There were a small number of cases (less than five) where SIMD 2012 rank was missing. If available, the 2016 rank was used instead to assign a rank. If neither rank was available, the case was assigned to quintile 3 (mid-point).

Both Local Authority and Healthboard names required tidying to ensure consistency of naming across the SCHC data (e.g. use of and versus & plus spelling errors). A small number of care homes were assigned to two different Healthboard areas (NHS Greater Glasgow and Clyde and NHS Lanarkshire) over the study period, this reflects overlapping provision of services based on historic boundaries. For consistency, all affected care home services (n = 8) were assigned to NHS Lanarkshire, for consistency with their Local Authority area. For 28 care homes no Healthboard information was provided in the SCHC dataset, and this was imputed manually based on the Local Authority information provided.

The SCHC includes two variables about dementia: those care home staff report have a medical diagnosis of dementia (*medically diagnosed*) and those with no formal diagnosis, but whom care home staff believe are living with the clinical symptoms of dementia (*non-medically diagnosed*) [21, 22].

Recognising dementia as an umbrella term, a code list of International Classification of Disease (ICD) 9 and ICD-10 dementia code groupings was collated from published academic research [18, 19] and the codes used in the Royal College of Psychiatrists National Audit of Dementia [31] (Supplementary table 2). This enabled grouping of the following subtypes: Alcohol associated dementia; Alzheimer's disease; Fronto-temporal dementia; Lewy body dementia/Parkinson's disease dementia; Unspecified dementia; Vascular dementia and Other dementias. This classification was applied across the SMR01, SMR04 and NRS data to identify dementia and dementia subtypes.

Location of death was classified using location codes within NRS data [32]. Where locations were assigned as follows: Hospitals (H); Care homes including hospices (J, K, R, S, T, U, V), Non-institutional settings (N) and other locations (all other codes).

Dementia medication records were defined as cases where there were prior dispensed community prescriptions for one or more of the cholinesterase inhibitors (Donepezil, Galantamine or Rivastigmine) and/or the NMDA receptor antagonist Memantine.

Analysis and Software

Anonymised data without personal identifiers were provided to the research team in the secure Public Health Scotland National Safe Haven, with remote researcher access. With research aims focused on the availability and overlaps of information across different sources, we did not perform any inferential statistical comparisons. Descriptive statistics were generated reproducibly using *R* (v. 4.5.1) [33] in *RStudio* (v 2025.05.0+496 "Mariposa Orchid" Release) [34] with packages *openxlsx* [35] and *tidyverse* [36].

We described the dementia information in the SCHC across census years, in terms of consistency and change. We linked our cohort of care home residents with the other national data sources, to check if they had a dementia record there and to compare the overlap between care home, healthcare and administrative data sources. We were not anticipating that residents would be represented in all sources as they were indicative of need for specialist hospital care (e.g. inpatient psychiatry) or an indication for a dementia medication (whose use is restricted to specific dementia subtypes). In addition, not all comorbid conditions are included on an individual's death certificate. We compared the overlap between individual data sources, and we also combined the three healthcare data source results (SMR01, SMR04 and PIS) together to create

Table 1: Summary of Variables Included in Data

Variable	Data source
Care home characteristics	
– Registered subtype of service – Healthboard area – Local Authority area – Area-based deprivation status (SIMD quintile) – Urban-rural classification (8-fold)	Scottish Care Home Census Scottish Index of Multiple Deprivation lookup Urban-rural classification lookup
Individual characteristics	
Age – At census year – At care home admission – At death	CHI population spine Scottish Care Home Census National Records of Scotland Death Records
Sex	CHI population spine
Survival to end of follow-up	National Records of Scotland Death Records
Location of death	National Records of Scotland Death Records
Dementia information	
– Medically diagnosed – Non medically diagnosed	Scottish Care Home Census
Acute/general hospital discharge with dementia code – Main condition – Other condition [1 to 5]	Scottish Morbidity Record (01) – Daycase and Inpatient Hospital
Psychiatric hospital discharge with dementia code – Main condition – Other condition [1 to 5] OR if still an inpatient at end of study, admission with dementia code – Admission main condition – Other main condition [1 to 4]	Scottish Morbidity Record (04) – Daycase and Inpatient Psychiatry
Dispensed community prescription for: – Donepezil hydrochloride – Galantamine – Memantine – Rivastigmine	Prescribing Information System
Death certificate includes dementia code – Underlying cause – Other cause [0 to 9]	National Records of Scotland Death Records

a binary classification of dementia in healthcare data or not, compared to a combined care home dementia (including both types of SCHC dementia) or not.

Using the SCHC data as an illustrative reference standard for dementia diagnosis information, we calculated the sensitivity, specificity, positive and negative predictive values for each of the health and administrative data sources. We used Euler diagrams to visualise the overlap in people identified as having dementia between each data source (SCHC, SMR01, SMR04, PIS and NRS mortality data). We visualised pairwise comparisons between each of the respective data sources, where the area of each circle was proportionate to the number of people identified as having dementia from that data source, and the area of the overlap proportional to the number of people identified by both data sources. Plots were generated using the *eulerr* package in R [37]. Finally, analysis of dementia data by age group and data source was undertaken. The SCHC data include adults from 18 years and over, although the vast majority of the care home population are older adults.

Governance and Approvals

Ethical approval was obtained from South Central—Hampshire B Research Ethics Committee (16/SC/0242).

Permission for linking national data was granted by the Public Benefit and Privacy Panel for Scotland (1516–0438) and Scottish Government Social Care Analysis Division.

Data were provided by the electronic data research and innovation service (eDRIS). All outputs were subject to statistical disclosure control by eDRIS. This necessitated combining categories to avoid numbers less than five but greater than zero (e.g. Orkney and Shetland grouped together at care home level).

Results

Study Analysis Cohort

There were 146,152 records available from the SCHC between 2012/13-2015/16. After removing records which could not be indexed (N=14,947; 10.2%) and those under 18 years at each census year (n=78, 0.05%), the final analysis cohort included 131,113 records for n=63,308 people (Supplementary Figure 1). Only 14% of people had records in all four financial years, 42.9% had records in a single financial year (Supplementary Table 3).

Care Home and Resident Characteristics

Descriptive statistics summarising key characteristics for the 1,231 care home services are presented in Supplementary Table 4. In summary, the majority are registered as care homes for older people (922, 74.9% of homes), with care homes for adults with learning disabilities accounting for 14.6% of homes. They are in areas of population density in NHS Greater Glasgow and Clyde (N = 242, 19.6%), NHS Grampian (N = 174, 14.1%) and NHS Lothian (N = 159, 12.9%), with smaller numbers of care homes in island areas. Less than a quarter of care homes were concentrated in the most and least deprived areas of the country (N = 221, 17.9% in most and N = 164, 13.3% in least deprived). Most care homes were in urban or accessible areas (N = 910, 73.9%).

Our total cohort includes 63,308 care home residents; their characteristics are summarised in Table 2. The mean age at census year was 82.7 years [Standard Deviation (SD) 12.65]. Most of our sample were female (n=42,332, 66.8%). Most did not survive to the end of follow-up in May 2020 (n = 54,860 86.6% died). The average age at death was 86.5 years [SD 9.04]. Most deaths occurred in the care home (n = 43,159, 78.7%) with a fifth of residents dying in hospital (n = 10,542, 19.2%).

Scottish Care Home Census Dementia Status

Dementia status in the SCHC is recorded by care home staff as medically diagnosed, non-medically diagnosed or no dementia. In our cohort 31,676 people (50.0%) were medically diagnosed, 4,599 people (7.3%) were non-medically diagnosed and 27,033 people (42.7%) had no dementia. Combining the dementia variables results in 36,275 (57.3%) with dementia and 27,033 (42.7%) without dementia.

To understand the utility of the SCHC in describing individuals' dementia diagnosis over time, we looked at the 36,572 instances where individuals had more than one SCHC record. For 34,048 individuals (93.1%), their dementia diagnosis in SCHC did not change between census years, in 2,489 individuals (6.8%), their dementia diagnosis in SCHC changed once between census years and for 35 individuals (0.1%), their dementia diagnosis in SCHC changed twice between census years. We compared earliest SCHC dementia diagnosis and compared this to the latest SCHC dementia diagnosis to summarise change over time for thosewith a change in status. It was not possible to report the intermediate data for those whose status changed twice, as the numbers are too small to report due to statistical disclosure control.

Across census years 267 people changed from no dementia to medically diagnosed, 1,091 changed from no dementia to non-medically diagnosed, 383 from non-medically diagnosed to medically diagnosed and 178 from medically to non-medically diagnosed. Whereas across census years, 138 people changed from non-medically diagnosed to no dementia and 467 from medically diagnosed to no dementia. Of those who changed status, 69.0% moved into the group classified as having dementia, 7.1% remained in the dementia group but with a different category of SCHC dementia and 23.9% moved into the no dementia group.

Dementia Records in National Healthcare and Administrative Data Sources

Across our cohort of 63,308 care home residents, 36,275 (57.3%) had dementia in care home data, 14,269 (22.5%) had a prior prescription for a dementia medication, 13,920 (22.0%) had a discharge diagnosis of dementia from SMR01, 2,750 (4.3%) had a discharge diagnosis of dementia from SMR04 and 31,825 (50.3%) had dementia recorded on their death certificate. Combining the three healthcare data sources (community prescribing, SMR01 & SMR04) together, 24,461 (38.6%) people had dementia recorded at least once.

Table 3 provides complete reporting of the distribution of dementia medication use, acute general and psychiatric hospital discharge and death certificate dementia information, including recording of dementia subtypes, where available.

Using data from 2010 onwards, the median number of dispensed prescriptions per person was 20 for Donepezil, 21 for Galantamine and 14 for both Rivastigmine and Memantine.

Where dementia was included on the death certificate in around two-thirds of cases this was recorded as the underlying cause of death (n = 20,694, 65.0%). More than one dementia subtype was included on an individuals' death certificate in 1,025 (4.9%) of cases where dementia was the underlying cause of death and 175 (1.6%) of cases where dementia was included as one of the other causes of death.

Vascular dementia was the most common subtype in SMR01 (52.6%) and SMR04 (71.2%). Unspecified dementia accounts for <1% of records in SMR01 and SMR04 but is used in 37.2% and 42.9% of death certificates (underlying and other causes respectively).

Overlap Between National Data Sources

Comparing care home dementia with other data sources identifies n = 10,916 (78.4%) people with dementia in SMR01 had care home dementia, n = 2,127 (78.4%) people with dementia in SMR04 had care home dementia, n = 12,146

Table 2: Resident-level characteristics included in study cohort (N = 63,308)

Variable	Count (% of total number of residents)
Last SCHC financial year used	
2012/13	10302 (16.27)
2013/14	9460 (14.94)
2014/15	10573 (16.7)
2015/16	32973 (52.08)
Age at census year (years)	
Mean [SD]	82.7 [12.7]
Median [IQR]	85.5 [78.8 to 90.6]
Range	91.2
Age at care home admission (years)	
Mean [SD]	80.5 [13.6]
Median [IQR]	83.7 [76.7 to 88.9]
Range	90.1
Missing	127
Sex	
Male	20976 (33.13)
Female	42332 (66.87)
Survival to end follow-up	
Died	54860 (86.66)
Survived	8448 (13.34)
Age at death (n=54,860) (years)	
Mean [SD]	86.5 [9.0]
Median [IQR]	87.9 [82.3 to 92.5]
Range	91.1
Location of death (n=54,860)	*% based on those who died*
Care home including hospice	43159 (78.67)
Hospital	10542 (19.22)
Non-institutional setting	669 (1.22)
Other location	490 (0.89)

(85.2%) people with dementia medications had care home dementia. Of those dying from dementia, as underlying cause of death, n = 17,034 (82.3%) had care home dementia and where dementia was an 'other cause' of death, n = 8,603 (77.2%) had care home dementia (Supplementary Table 5).

If the SCHC care home data are used as an illustrative reference standard for dementia information, the relative performance of health and administrative data sources can be calculated (Table 4).

The Euler Diagram in Figure 1 shows the visual overlap of dementia data across datasets. This demonstrates the small contribution of SMR04, inpatient psychiatry, as a source of dementia data for people living in care homes. There is also significant overlap between dementia medication use and care home dementia and between care home dementia and dementia included on death certificate data. SMR01, inpatient hospital admissions and dementia medications each contribute additional healthcare dementia data, with comparatively small overlap.

Figure 2 presents the age distribution of dementia data across the included datasets, illustrating the lower proportion of the cohort in younger (18-60) age groups and the lower prevalence of dementia in these groups. Dementia in the oldest old (90 years and over) primarily comes from care home and death data.

Comparing care home dementia data with healthcare data we found 19,742 (54.4%) of those with care home dementia had healthcare data indicative of dementia and 22,314 (82.5%) without dementia in care home data also had no evidence of dementia in healthcare data. There were 4,719 (19.3%) people who had healthcare dementia data but were classed as not having dementia in the care home data.

Comparing all healthcare data and care home data with death certificate data, where dementia was the underlying cause of death, n = 18,790 (91.0%) people had prior evidence of dementia in other datasets. Similarly, where dementia was listed as an 'other cause' of death, n = 9833 (88.3%) people had prior evidence of dementia in other datasets. For those where dementia was not included on the death certificate, n = 14,138 (61.4%) people had no prior evidence of dementia in other datasets (Figure 3).

The overall overlap between national data sources found 30.2% of the cohort had no dementia data in any national data sources and 25.3% had dementia data in all three sources and 15.2% had dementia in care home data and death data

Table 3: Distribution of Dementia Information Across National Data Sets

	Variable	Number (% of cohort)
Prescribing Information System – Dementia Medication	Donepezil prescription ever	8030 (56.3)
	Galantamine prescription ever	2820 (19.8)
	Memantine prescription ever	4262 (29.9)
	Rivastigmine prescription ever	1874 (13.1)
	Medication combinations	*% based on n receiving dementia medication*
	Donepezil alone	6261 (43.9)
	Donepezil and Galantamine	205 (1.4)
	Donepezil and Memantine	1129 (7.9)
	Donepezil and Rivastigmine	225 (1.6)
	Galantamine alone	2007 (14.1)
	Galantamine and Memantine	390 (2.7)
	Galantamine and Rivastigmine	67 (0.6)
	Memantine alone	2319 (16.2)
	Memantine and Rivastigmine	211 (1.5)
	Rivastigmine alone	1219 (8.5)
	Donepezil, Galantamine and Memantine	84 (0.6)
	Donepezil, Galantamine and Rivastigmine	23 (0.2)
	Donepezil, Memantine and Rivastigmine	85 (0.6)
	Galantamine, Memantine and Rivastigmine	26 (0.2)
	All four drugs	18 (0.1)
SMR01 dementia code	Single dementia diagnosis coded	12744 (20.1)
	Combination dementia diagnoses coded*	1176 (1.9)
	Single dementia diagnosis	*% based on those with a single diagnosis*
	Alcohol associated dementia	718 (5.6)
	Alzheimer's disease	4051 (31.8)
	Frontotemporal dementia	48 (0.4)
	Lewy Body dementia/ Parkinson's disease dementia	271 (2.1)
	Other dementia	920 (7.2)
	Unspecified dementia	27 (0.2)
	Vascular dementia	6709 (52.6)
SMR04 dementia code	Single dementia diagnosis coded	2651 (4.2)
	Combination dementia diagnoses coded	99 (0.2)
	Single dementia diagnosis	*% based on those with a single diagnosis*
	Alcohol associated dementia	375 (14.1)
	Alzheimer's disease	257 (9.7)
	Frontotemporal dementia	7 (0.3)
	Lewy Body dementia/ Parkinson's disease dementia	37 (1.4)
	Other dementia	77 (2.9)
	Unspecified dementia	11 (0.4)
	Vascular dementia	1887 (71.2)
NRS deaths dementia code	**Distribution of dementia codes**	*%s based on underlying and other*
	Underlying cause – one dementia subtype	19939 (96.4)
	Underlying cause – more than one dementia subtype	1025 (4.9)
	Other cause – one dementia subtype	10956 (98.4)
	Other cause – more than one dementia subtype	175 (1.6)
	Underlying cause of death dementia subtype	*% based on those dying with dementia as underlying cause*
	Alcohol associated dementia	61 (0.3)
	Alzheimer's disease	6698 (32.4)
	Frontotemporal dementia	77 (0.4)
	Lewy Body dementia/ Parkinson's disease dementia	309 (1.5)
	Other dementia	14 (0.1)

Continued

Table 3: Continued

Variable	Number (% of cohort)
Unspecified dementia	7691 (37.2)
Vascular dementia	5844 (28.2)
Other cause of death dementia subtype – single dementia subtype	*% based on those with one dementia subtype other cause*
Alcohol associated dementia	211 (1.9)
Alzheimer's disease	2539 (23.2)
Frontotemporal dementia and Other dementia**	35 (0.4)
Lewy Body dementia/ Parkinson's disease dementia	204 (1.9)
Unspecified dementia	4704 (42.9)
Vascular dementia	3260 (29.8)

IQR – inter-quartile range.
*Combination dementia diagnosis coded includes where more than one dementia code used in the same hospital stay and across different hospital stays.
**Other dementia combined with Frontotemporal dementia for statistical disclosure control.

Table 4: Comparing the utility of health and administrative data for identifying care home dementia using SCHC care home dementia as illustrative reference standard

Dataset	Sensitivity	Specificity	Positive predictive value	Negative predictive value
Dispensed medication (PIS)	0.33	0.92	0.85	0.51
General hospital discharge (SMR01)	0.30	0.89	0.78	0.49
Psychiatric hospital discharge (SMR04)	0.06	0.98	0.77	0.44
Death record including dementia (NRS)	0.77	0.71	0.81	0.66

(Supplementary Table 6). We identified 3202 (5.1%) people who only had dementia included in death data.

Discussion

Findings in Context

We were able to create a large cohort of adult care home residents using routinely collected care home data and link this to healthcare and administrative dementia data to explore the usefulness of available data sources in understanding dementia prevalence in this population. Analysing the cohort finds that care home data on dementia are largely consistent over time and can identify those developing dementia while resident, with just 1% changing from dementia to no dementia across years. Inpatient psychiatry data, despite likely high accuracy, contribute a small number of cases of dementia among care home residents, compared to general hospital data and community prescribing of dementia medications. Not all prior dementia diagnosis information (e.g. from hospital records or prescribing) were reflected in care home records. Death certificate data is also an important source in identifying dementia among people who lived in care homes, who were not identified in hospital or prescribing datasets. However, death certificate data more commonly classified residents as having had Unspecified Dementia than other dementia data sources, where more specific diagnostic codes were used.

A key challenge with this work is the lack of a gold standard dataset or means to identify which source of information is 'correct' in dementia status. For example, just because someone has dementia, does not mean it will be included on their death certificate unless it is relevant to their cause of death. This lack of gold standard is common across routinely collected data, used for research, but often not formally explored. Similar challenges were faced when combining care home digital care record data with national health datasets in England, including the need to formalise hierarchical decision-making for defining dementia [38]. Dementia medication prescribing has been previously shown to identify less than half of those with a dementia diagnosis and hospital admission, in data from two Scottish Healthboards [39]. The importance of documenting dementia diagnoses in social care data has been demonstrated in Medicare home health data, where those with undocumented dementia had poorer experiences and outcomes of care than those where the dementia diagnosis was recorded [40].

Death certificate data for dementia is one area which has seen prior epidemiological research, including triangulation with other routine data [41]. Globally, dementia is under-reported on death certificates [42–44]. It has been noted that living in a care home was more strongly associated with the inclusion of dementia on death certificates than living elsewhere in the community [44, 45]. However, a lack of representation of specific dementia subtypes has been reported as common in England and the USA [43, 44]. Population studies making use of death certificate data must be mindful of these recording biases.

Strengths and Limitations

It is important to understand the strengths and limitations of routine data sources when using them for research and thus

Figure 1: Euler Diagram Demonstrating the Overlap Between Care Home, Healthcare and Administrative Data Sources

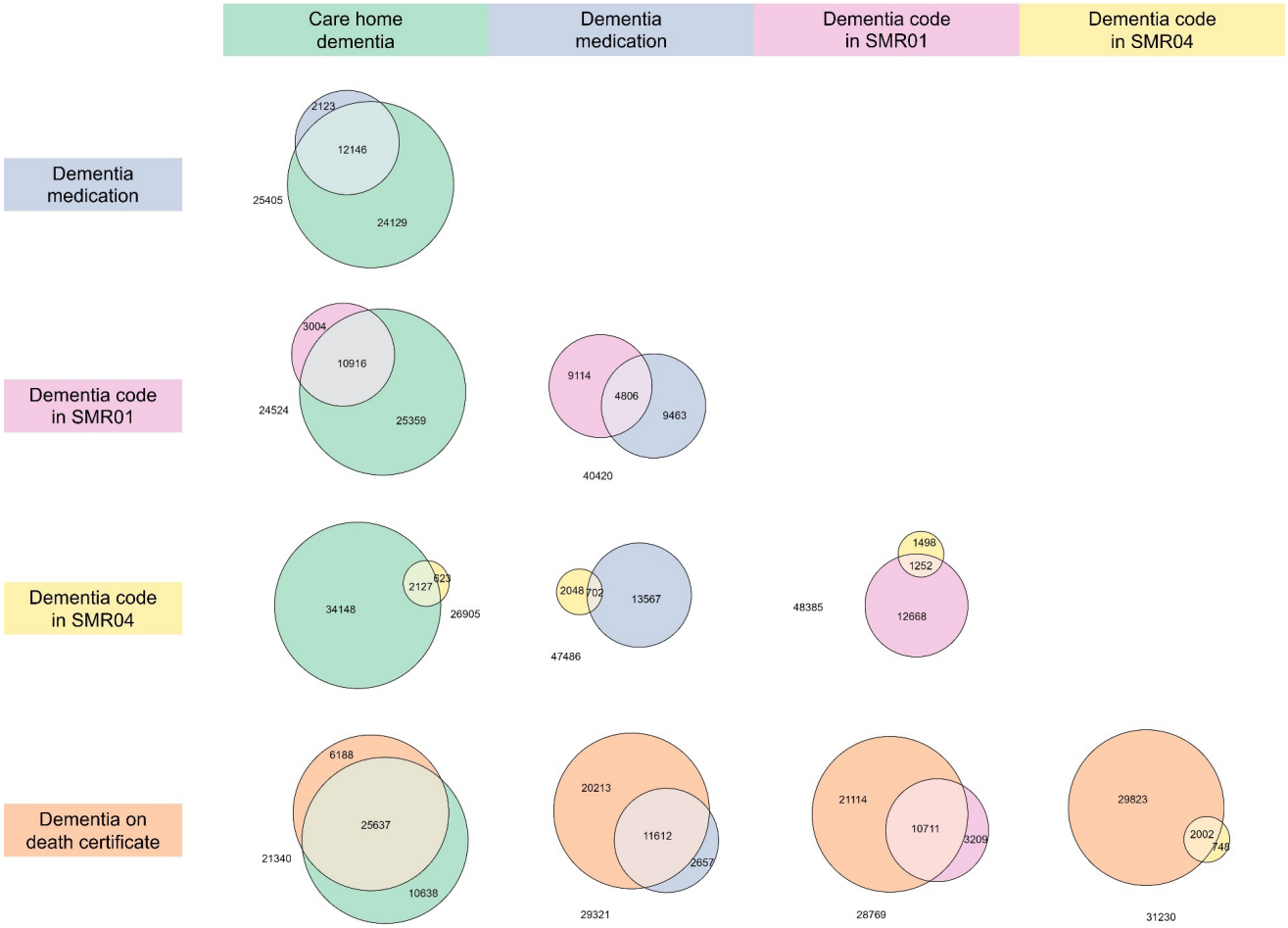


Figure 2: Age distribution of dementia data between datasets: A – dementia in any datasets versus none; B – dementia in healthcare datasets versus none; C – dementia in care home data versus none; D – dementia in death data versus death from another cause versus survived

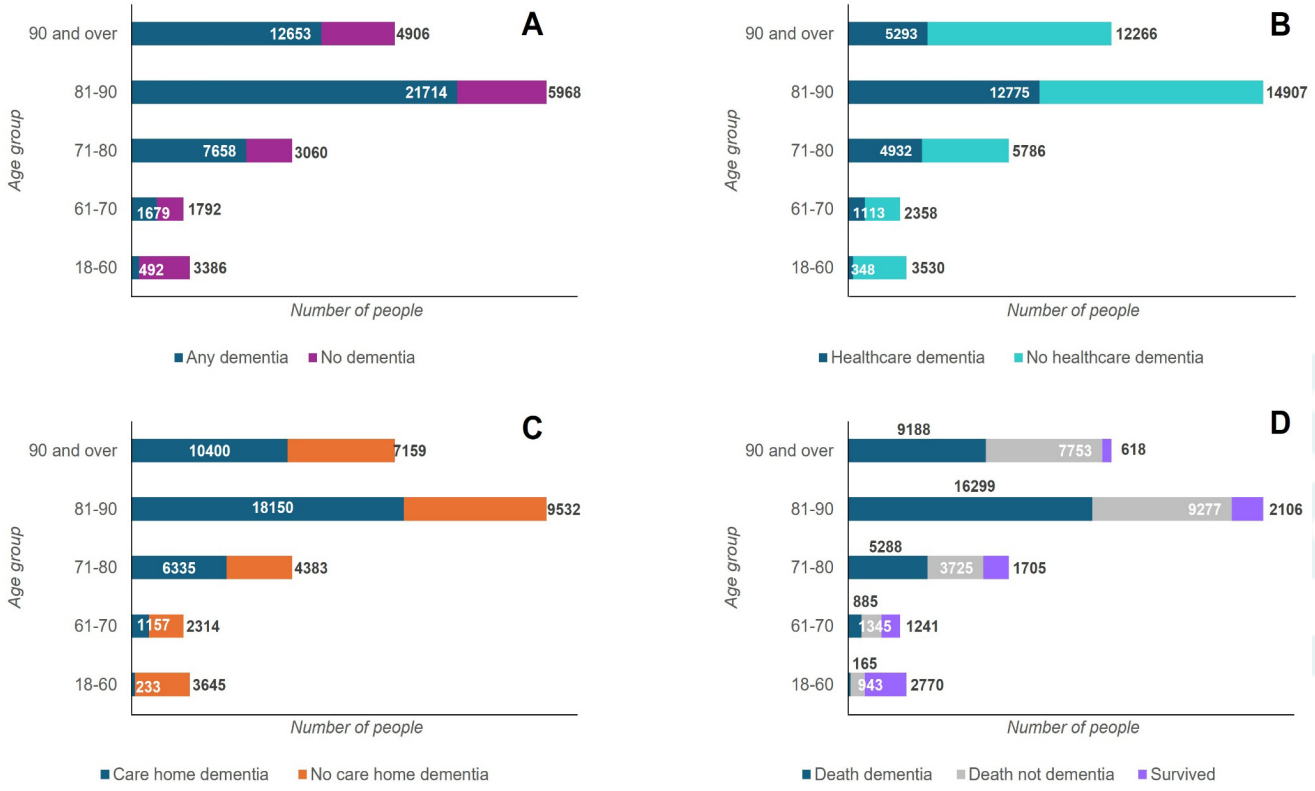
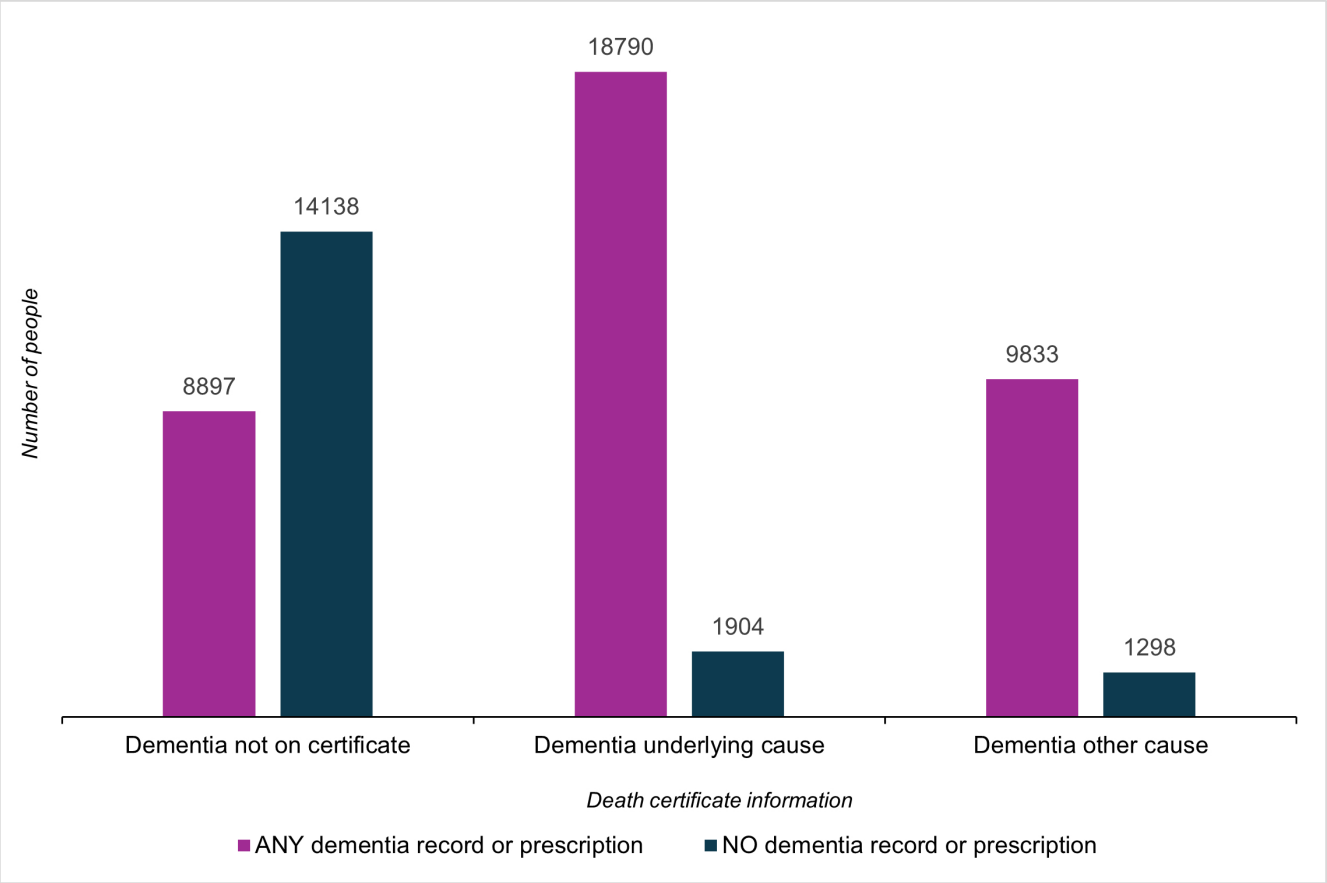


Figure 3: Bar Chart Demonstrating the Number of Participants With and Without Dementia on Death Certificate, Based on the Presence or Absence of Dementia in Other Data Sources



this work represents an innovative approach to exploring data in an under-researched population. It has made efficient re-use of existing data resources, collected by care home and healthcare staff in their routine care. It has not placed any additional burden on individuals (residents or staff) and is inclusive of those with and without capacity to participate in research, a significant challenge when undertaking care home research [46].

The aim was to provide transparency in the usefulness of different data sources, to explore the challenges working with overlapping data sources and make recommendations to assist other researchers working in the public sector and academia. Such methodological work is helpful in advancing use of new data sources.

It had been hoped to use care home data back to 2010. However, the lower rates of indexing (<80%) would have introduced bias, compared to those from 2012 onwards where a higher indexing rate was achieved as more personal identifiers were collected in the dataset. Even with this improved recording, we had to remove 10.2% of available records which could not be indexed to CHI as they could not be linked to other data sources. These are important findings for other research teams who may wish to use the SCHC for research.

It is also important to acknowledge the age of the data and potential impact this has on the utility of the results. The SCHC data from 2010-16 were indexed and made available for their first research use as part of a wider programme of national social care data research partnered between Scottish Government and academic researchers [26]. SCHC data collection and analysis were then disrupted due to data governance challenges and the COVID-19 pandemic [47]. The process to index, link and provide the data originally was time-consuming due to governance for new research use and technical challenges, highlighted in the National Care Home Data Review [48]. There has been greater interest in recent years in the use of routinely collected data for research, however, work exploring comparisons between data sources is critically important to inform the application and use of research findings. Individuals living with dementia continue to face barriers to accessing a diagnosis, particularly problematic among those living in care homes [49, 50]. Having demonstrated the utility of the SCHC data compared to other sources, it would be interesting to look at more recent, post-pandemic data, for any changes in recording and overlaps.

Our linkages to healthcare datasets were contingent on the presence of a dementia record in PIS, SMR01 or SMR04 and so we cannot calculate the proportions of the study cohort who had records in these datasets for conditions other than dementia. The project did not have access to national data on dementia post-diagnostic support referrals/provision as this national dataset is not available for research use. National primary care data were also not accessible to link for the study period. Finally, despite activity data for outpatient clinic attendances being collected in SMR00, this national dataset does not have diagnostic information collected which could provide another source of dementia information.

Implications

There is significant value in linking data across sectors to generate novel insights and make more effective use of existing resources [51, 52]. This paper demonstrates the contribution to better understand the needs of the population living in care homes with dementia and how this can be approached. It is consistent with wider emerging research interest in the methodological work to identify how best to use routine data as a research tool, for example in defining look-back periods to capture diagnoses adequately [53] and how to define and characterise multiple long-term conditions in secondary care data [54].

The national work creating a Dementia Index cohort to enable consistent identification of those living with dementia from routine data sources [23] underpins the value in methodological work such as this to describe the contributions of different data sources to increase the sensitivity of case finding.

Care homes do not routinely have access to residents' health data and rely on information shared by others. Exploring how this information can be shared between professionals to enable a complete picture of residents' medical history is essential in an integrated health and care system.

Future research would benefit from evaluating the impact of including post-diagnostic support data and primary care data as alternative national sources, which are likely to contribute additional information.

Conclusion

This study demonstrates the feasibility and utility of linking care home data with healthcare and administrative data to identify the population living with dementia in care homes and identifies the contributions of different data sources. It demonstrates the positive contribution routinely collected social care data resources, such as the SCHC, can make to understanding the needs of populations under-represented in research. This work can provide insights into the development of wider population cohorts to study dementia epidemiology and improve inclusive population-wide insights to inform practice and policymaking.

Data Availability Statement

The national datasets used in this study are collected/held/controlled by Public Health Scotland, National Records of Scotland, the Care Inspectorate and Scottish Government. Data access can be enabled by application to the Public Benefit and Privacy Panel for Health and Social Care and eDRIS team (Public Health Scotland).

Ethics Statement

Ethical approval was obtained from South Central—Hampshire B Research Ethics Committee (16/SC/0242).

Acknowledgements

The authors would like to acknowledge the support of the eDRIS Team (Public Health Scotland) for their involvement in obtaining approvals, provisioning and linking data and the use of the secure analytical platform within the National Safe Haven.

This work uses routinely collected data provided by care home staff about their residents and service, which is collected nationally by the Care Inspectorate on behalf of Scottish Government.

Funding

This research was supported by the Scottish Informatics and Linkage Collaboration. JKB is supported by a NES/CSO Postdoctoral Clinical Lectureship (PCL/21/01). SH was supported to undertake a vacation studentship by the Vivensa Foundation Academy Excellence Award (EA2402\51). The funders had no role in the design, conduct or interpretation of the study.

Declarations of Interest

KM, MD, SH, PH, TJQ & JKB have no conflicts of interest to declare

EL is employed as a Statistician in the Scottish Government Health and Social Care Analysis Division, which is joint data controller for the Scottish Care Home Census. The views in the paper represent the views of the authors and not the Scottish Government.

AI Disclosure Statement

The authors declare that no generative AI tools were used in the preparation of this manuscript.

References

1. Livingston G, Huntley J, Liu KY, Costafreda SG, Selbæk G, Alladi S, et al. Dementia prevention, intervention, and care: 2024 report of the Lancet Standing Commission. The Lancet. 2024;404(10452):572–628. https://doi.org/10.1016/S0140-6736(24)01296-0

2. Nichols E, Steinmetz JD, Vollset SE, Fukutaki K, Chalek J, Abd-Allah F, et al. Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019. The Lancet Public Health. 2022;7(2):e105–e25. https://doi.org/10.1016/S2468-2667(21)00249-8

3. Shepherd H, Livingston G, Chan J, Sommerlad A. Hospitalisation rates and predictors in people with dementia: a systematic review and meta-analysis. BMC Medicine. 2019;17(1):130. https://doi.org/10.1186/s12916-019-1369-7

4. Tonelli M, Wiebe N, Straus S, Fortin M, Guthrie B, James MT, et al. Multimorbidity, dementia and health care in older people:a population-based cohort study. CMAJ Open. 2017;5(3):e623–e31. https://doi.org/10.9778/cmajo.20170052

5. Stirland LE, Choate R, Zanwar PP, Zhang P, Watermeyer TJ, Valletta M, et al. Multimorbidity in dementia: Current perspectives and future challenges. Alzheimers Dement. 2025;21(8):e70546. https://doi.org/10.1002/alz.70546

6. Nelis SM, Wu Y-T, Matthews FE, Martyr A, Quinn C, Rippon I, et al. The impact of co-morbidity on the quality of life of people with dementia: findings from the IDEAL study. Age and Ageing. 2019;48(3):361–7. https://doi.org/10.1093/ageing/afy155

7. Shiells K, Pivodic L, Holmerová I, Van den Block L. Self-reported needs and experiences of people with dementia living in nursing homes: a scoping review. Aging & Mental Health. 2020;24(10):1553–68. https://doi.org/10.1080/13607863.2019.1625303

8. Haunch K, Thompson C, Arthur A, Edwards P, Goodman C, Hanratty B, et al. Understanding the staff behaviours that promote quality for older people living in long term care facilities: A realist review. Int J Nurs Stud. 2021;117:103905. https://doi.org/10.1016/j.ijnurstu.2021.103905

9. Mittal A, Arora I, Jayaram R, Yashwanth G, Rangarajan SK. Prevalence of Dementia in the Geriatric Population Residing in a Long-term Care Facility: An Updated Systematic Review and Meta-Analysis. Journal of Psychiatry Spectrum. 2025;4(1). https://doi.org/10.4103/jopsys.jopsys_35_24

10. Siegel EO, Backman A, Cai Y, Goodman C, Ocho ON, Wei S, et al. Understanding Contextual Differences in Residential LTC Provision for Cross-National Research: Identifying Internationally Relevant CDEs. Gerontology and Geriatric Medicine. 2019;5:2333721419840591. https://doi.org/10.1177/2333721419840591

11. Burton JK, Quinn TJ, Gordon AL, MacLullich AMJ, Reynish EL, Shenkin SD. Identifying published studies of care home research: an international survey of researchers. Journal of Nursing Home Research. 2017;3:99–102. https://doi.org/10.14283/jnhrs.2017.15

12. Stewart R, Hotopf M, Dewey M, Ballard C, Bisla J, Calem M, et al. Current prevalence of dementia, depression and behavioural problems in the older adult care home sector: the South East London Care Home Survey. Age and Ageing. 2014;43(4):562–7. https://doi.org/10.1093/ageing/afu062

13. Lithgow S, Jackson GA, Browne D. Estimating the prevalence of dementia: cognitive screening in Glasgow nursing homes. International Journal of Geriatric Psychiatry. 2012;27(8):785–91. https://doi.org/10.1002/gps.2784

14. Goodman C, Baron N, Machen I, Stevenson E, Evans C, Davies S, et al. Culture, consent, costs and care homes: Enabling older people with dementia to participate in research. Aging & Mental Health. 2011;15(4):475–81. https://doi.org/10.1080/13607863.2010.543659

15. Nocivelli B, Wood F, Hood K, Wallace C, Shepherd V. "Research happens a lot in other settings—so why not here?" A qualitative interview study of stakeholders' views about advance planning for care home residents' research participation. Age and Ageing. 2024;53(10):afae235. https://doi.org/10.1093/ageing/afae235

16. Moore DC, Hanratty B. Out of sight, out of mind? a review of data available on the health of care home residents in longitudinal and nationally representative cross-sectional studies in the UK and Ireland. Age and Ageing. 2013;42(6):798–803. https://doi.org/10.1093/ageing/aft125

17. Burton JK, Goodman C, Guthrie B, Gordon A, Hanratty B, Quinn T. Closing the UK care home data gap - methodological challenges and solutions. International Journal of Population Data Science. 2020;5(4):1391. https://doi.org/10.23889/ijpds.v5i4.1391

18. Wilkinson T, Schnier C, Bush K, Rannikmäe K, Henshall DE, Lerpiniere C, et al. Identifying dementia outcomes in UK Biobank: a validation study of primary care, hospital admissions and mortality data. Eur J Epidemiol. 2019;34(6):557–65. https://doi.org/10.1007/s10654-019-00499-1

19. Wilkinson T, Ly A, Schnier C, Rannikmäe K, Bush K, Brayne C, et al. Identifying dementia cases with routinely collected health data: A systematic review. Alzheimers Dement. 2018;14(8):1038–51. https://doi.org/10.1016/j.jalz.2018.02.016

20. Sibbett RA, Russ TC, Deary IJ, Starr JM. Dementia ascertainment using existing data in UK longitudinal and cohort studies: a systematic review of methodology. BMC Psychiatry. 2017;17(1):239. https://doi.org/10.1186/s12888-017-1401-4

21. Public Health Scotland. Care Home Census for Adults in Scotland: Statistics for 2014-20242024 [cited 2024 11th October]: Available from: https://publichealthscotland.scot/media/29259/2024-10-01-care-home-census-report_final.pdf.

22. Burton JK, Lynch E, Love S, Rintoul J, Starr JM, Shenkin SD. Who lives in Scotland's care homes? Descriptive analysis using routinely collected social care data 2012-16. The journal of the Royal College of Physicians of Edinburgh. 2019;49(1):12–22. https://doi.org/10.4997/JRCPE.2019.103

23. NHS Scotland, Scottish Government. Realistic Medicine: Critical Connections2025 [cited 2025 21st August]: Available from: https://www.gov.scot/publications/chief-medical-officers-annual-report-2024-2025-realistic-medicine-critical-connections/.

24. Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. PLoS Med. 2015;12(10):e1001885. https://doi.org/10.1371/journal.pmed.1001885

25. Alvarez-Madrazo S, McTaggart S, Nangle C, Nicholson E, Bennie M. Data Resource Profile: The Scottish National Prescribing Information System (PIS). International Journal of Epidemiology. 2016 Jun;45(3):714–5f. https://doi.org/10.1093/ije/dyw060

26. Henderson D, Burton J, Lynch E, Rintoul J, Clark D, Bailey N. Data Resource Profile The Scottish Social Care Survey (SCS) and the Scottish Care Home Census (SCHC). International Journal of Population Data Science. 2019;4:24. https://doi.org/10.23889/ijpds.v4i1.1108

27. Womersley J. The public health uses of the Scottish Community Health Index (CHI). Journal of public health medicine. 1996;18(4):465–72. https://doi.org/10.1093/oxfordjournals.pubmed.a024546

28. Care Inspectorate. Care Inspectorate: About us. 2025; Available from: https://www.careinspectorate.com/index.php/about-us.

29. NHS National Services Scotland. Scottish Index of Multiple Deprivation (SIMD). [cited 2018 13th February]; Available from: https://nhsnss.org/services/practitioner/dental/scottish-index-of-multiple-deprivation-simd/.

30. Scottish Government. Scottish Government Urban Rural Classification. 2016; Available from: https://www2.gov.scot/Topics/Statistics/About/Methodology/UrbanRuralClassification.

31. Royal College of Psychiatrists. National Audit of Dementia Round 6: List of eligible ICD-10 codes2023 [cited 2025 12th March]: Available from: https://www.rcpsych.ac.uk/docs/default-source/improving-care/ccqi/national-clinical-audits/national-audit-of-dementia/nad-round-6-(2023-2024)/mas-r6/list-of-icd-10-codes.pdf?sfvrsn=6d0be072_4.

32. National Records of Scotland. Code-lists Used in Vital Event Statistics: Institutions. 2022 [cited 2021 18th October]; Available from: https://www.nrscotland.gov.uk/files//statistics/vital-events/institution-codes-october-2021.xlsx.

33. R Core Team. A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2025.

34. Posit team. RStudio: Integrated Development Environment for R. Boston: Posit Software; 2025.

35. Schauberger P, A W. Read, Write and Edit xlsx Files R package version 4.2.6.1. 2024.

36. Wickham H, Averick M, Bryan J, Chang W, D'Agostino McGowan L, Francois R, et al. Welcome to the Tidyverse. The Journal of Open Source Software. 2019;4(43):1686. https://doi.org/10.21105/joss.01686

37. Larsson J, Godfrey A, Gustafsson P, Eberly D, Huber E, Prive F. eulerr: Area-Proportional Euler and Venn Diagrams with Ellipses. 7.0.2 ed2024.

38. Gordon AL, Rand S, Crellin E, Allan S, Tracey F, De Corte K, et al. Piloting a minimum data set for older people living in care homes in England: a developmental study. Age and Ageing. 2025;54(1):afaf001. https://doi.org/10.1093/ageing/afaf001

39. Hapca S, Burton JK, Cvoro V, Reynish E, Donnan PT. Are antidementia drugs associated with reduced mortality after a hospital emergency admission in the population with dementia aged 65 years and older? Alzheimer's & Dementia: Translational Research & Clinical Interventions. 2019;5:431–40. https://doi.org/10.1016/j.trci.2019.07.011

40. Burgdorf JG, Amjad H, Barrón Y, Ryvicker M. Undocumented Dementia Diagnosis During Skilled Home Health Care: Prevalence and Associated Factors. Journal of the American Geriatrics Society. 2025;73(7):2117–26. https://doi.org/10.1111/jgs.19491

41. Ahmad S, Carey IM, Harris T, Cook DG, DeWilde S, Strachan DP. The rising tide of dementia deaths: triangulation of data from three routine data sources using the Clinical Practice Research Datalink. BMC Geriatrics. 2021;21(1):375. https://doi.org/10.1186/s12877-021-02306-7

42. Garcia-Ptacek S, Kåreholt I, Cermakova P, Rizzuto D, Religa D, Eriksdotter M. Causes of Death According to Death Certificates in Individuals with Dementia: A Cohort from the Swedish Dementia Registry. Journal of the American Geriatrics Society. 2016;64(11):e137–e42. https://doi.org/10.1111/jgs.14421

43. Wachterman M, Kiely DK, Mitchell SL. Reporting dementia on the death certificates of nursing home residents dying with end-stage dementia. JAMA. 2008;300(22):2608–10. https://doi.org/10.1001/jama.2008.768

44. Perera G, Stewart R, Higginson IJ, Sleeman KE. Reporting of clinically diagnosed dementia on death certificates: retrospective cohort study. Age and Ageing. 2016;45(5):668–73. https://doi.org/10.1093/ageing/afw077

45. Gao L, Calloway R, Zhao E, Brayne C, Matthews FE. Accuracy of death certification of dementia in population-based samples of older people: analysis over time. Age Ageing. 2018;47(4):589–94. https://doi.org/10.1093/ageing/afy068

46. Drummond M, Cartin K, Shenkin SD, Burton JK. Facilitating equitable research access for people living in

care homes. Age Ageing. 2024;53(10). https://doi.org/10.1093/ageing/afae220

47. Office for Statistics Regulation, UK Statistics Authority. Care Home Census for Adults in Scotland Statistics. 2020; Available from: https://osr.statisticsauthority.gov.uk/wp-content/uploads/2020/02/Care_Home_Census_Dedesignation.pdf.

48. Scottish Government, Care Inspectorate, Public Health Scotland. Care Home Data Review - Full Report2024 [cited 2026 3rd January]: Available from: https://www.gov.scot/publications/care-home-data-review-full-report/documents/.

49. Burton J. Exploring advanced dementia in Scotland's care homes: prevalence and understanding2025 [cited 2025 14th September]: Available from: https://www.alzscot.org/wp-content/uploads/2025/09/Exploring-advanced-dementia-in-Scotlands-care-homes_-prevalence-and-understanding_Summary-Report.pdf.

50. International Longevity Centre UK. Living better with dementia through care and support: it's not rocket science2025 [cited 2025 7th September]: Available from: https://ilcuk.org.uk/wp-content/uploads/2025/06/ILC-living-better-with-dementia-factpack.pdf.

51. Todd OM, Burton JK, Dodds RM, Hollinghurst J, Lyons RA, Quinn TJ, et al. New Horizons in the use of routine data for ageing research. Age and Ageing. 2020;49:716–22. https://doi.org/10.1093/ageing/afaa018

52. Lugg-Widger F, Sydenham M, Oatley R, Scourfield J. Use of Linked Administrative Adult Social Care Data for Research: A Scoping Review of Existing UK Studies. The British Journal of Social Work. 2024:bcae151. https://doi.org/10.1093/bjsw/bcae151

53. Lewis J, Evison F, Doal R, Field J, Gallier S, Harris S, et al. How far back do we need to look to capture diagnoses in electronic health records? A retrospective observational study of hospital electronic health record data. BMJ Open. 2024;14(2):e080678. https://doi.org/10.1136/bmjopen-2023-080678

54. Cooper R, Bunn JG, Richardson SJ, Hillman SJ, Sayer AA, Witham MD. Rising to the challenge of defining and operationalising multimorbidity in a UK hospital setting: the ADMISSION research collaborative. Eur Geriatr Med. 2024;15(3):853–60. https://doi.org/10.1007/s41999-024-00953-8

Abbreviations

CHI:	Community Health Index
eDRIS:	electronic Data Research and Innovation Service
ICD:	International Classification of Disease
NHS:	National Health Service
NRS:	National Records of Scotland
PIS:	Prescribing Information System
SCHC:	Scottish Care Home Census
SIMD:	Scottish Index of Multiple Deprivation
SMR:	Scottish Morbidity Record
UK:	United Kingdom