Piloting a minimum data set for older people living in care homes in England: a developmental study

Adam L Gordon[1,2]
Stacey Rand[3]
Elizabeth Crellin[4]
Stephen Allan[3]
Freya Tracey[4]
Kaat De Corte[4]
Therese Lloyd[4]
Richard Brine[5]
Rachael E Carroll[1,2]
Ann-Marie Towers[6]
Jennifer Kirsty Burton[7]
Gizdem Akdur[8]
Barbara Hanratty[9]
Lucy Webster[6]
Sinead Palmer[3]
Liz Jones[10]
Julienne Meyer[10,11]
Karen Spilsbury[12,13]
Anne Killett[14]
Arne T Wolters[15]
Guy Peryer[14]
Claire Goodman[7,16]

[1]Wolfson Institute of Population Health, Queen Mary University of London, London. UK.
[2]Academic Centre for Healthy Ageing, Barts Health NHS Trust, London. UK
[3]Personal Social Services Research Unit (PSSRU), University of Kent, Canterbury, Kent. CT2 7NF.
[4]Improvement Analytics Unit, the Health Foundation, London, UK.
[5]Our Future Health, London, UK.
[6]Health and Social Care Workforce Research Unit, Kings Collee London, London. UK.
[7]School of Cardiovascular and Metabolic Health, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK.
[8]Centre for Research in Public Health and Community Care (CRIPACC), University of Hertfordshire, College Lane, Hatfield, AL10 9AB, UK
[9]Population Health Sciences Institute, Newcastle University, Newcastle upon Tyne, UK
[10]National Care Forum, London, UK
[11]City, University of London
[12]School of Healthcare, Faculty of Medicine and Health, University of Leeds, Leeds, UK
[13]NIHR Applied Research Collaboration Yorkshire and Humber (YHARC), Bradford, UK
[14]School of Health Sciences, University of East Anglia, Norwich, Norfolk, NR4 7TJ
[15]Cumbria, Northumberland, Tyne and Wear NHS Foundation Trust, Newcastle upon Tyne, UK.
[16]NIHR Applied Research Collaboration-East of England (ARC-EoE), Cambridge, UK

Corresponding author: Adam Gordon
Email: adam.gordon@qmul.ac.uk
Tel: +44 20 7882 5555
Address: Academic Centre for Healthy Ageing, Redevelopment Offices, Junction 2, Whipps Cross Hospital, London. E11 1NR.

All authors have seen and approved the manuscript. The authors have no competing interests to declare.

Abstract

Background
We developed a prototype minimum data set (MDS) for English care homes, assessing feasibility of extracting data directly from digital care records (DCRs) with linkage to health and social care data.

Methods
Through stakeholder development workshops, literature reviews, surveys and public consultation we developed an aspirational MDS. We identified ways to extract this from existing sources including DCRs and routine health and social care datasets. To address gaps we added validated measures of delirium, cognitive impairment, functional independence and Quality of Life to DCR software. Following routine health and social care data linkage to DCRs, we compared variables recorded across multiple data sources, using a hierarchical approach to reduce missingness where appropriate. We reported proportions of missingness, mean and standard deviation (SD) or frequencies (%) for all variables.

Results
We recruited 996 residents from 45 care homes in three English Integrated Care Systems. 727 residents had data included in the MDS. Additional data were well completed (<35% missingness at wave 1). Competition for staff time, staff attrition, and software-related implementation issues contributed to missing DCR data. Following data linkage and combining variables where appropriate, missingness was reduced (<=4% where applicable).

Discussion
Integration of health and social care is predicated on access to data and interoperability. Despite governance challenges we safely linked care home DCRs to statutory health and social care datasets to create a viable prototype MDS for English care homes. We identified issues around data quality, governance, data plurality and data completion essential to MDS implementation going forward.

Key points

- DCRs, health and social care datasets contain a range of information which can help provide a more complete picture of residents.
- We developed and implemented a Minimum Dataset linking care home DCRs to statutory health and social care records.
- Information governance for linking data across multiple data owners and data processors is complex and time consuming.
- Standardisation across Digital Care Records Systems would enable data to be used more effectively across the care home sector.
- Establishing shared priorities across key stakeholders interested in care home data is essential to effective MDS implementation.

Key words

Care homes; minimum dataset; data linkage; quality of life; digital care record

Background

Care homes provide around-the-clock residential care for people whose needs cannot be met by visiting care. Older people living in care homes often have needs defined by one or more of frailty, multiple long-term conditions, disability or cognitive impairment [1]. Homes can be registered as with or without nursing depending on whether they employ registered nurses to oversee and provide complex healthcare. In England, there are around 372,000 care home places [2].

Day-to-day care for residents generates abundant data spread across records held by care homes, statutory social care organisations, the National Health Service (NHS), residents and their families [3,4]. As records become increasingly digitised, there is an opportunity to collate data to inform decisions about commissioning, care planning and delivery, review and funding at the micro (individual resident), meso (care home and regional system) and macro (national system) levels [4].

Care home residents were amongst those most adversely affected by COVID-19 and the sector was devastated by outbreaks [5]. At pandemic outset, England lacked even rudimentary data on how many people lived in care homes to track COVID-19 incidence [6]. Emergency legislation, now repealed, enabled collated datasets and recognition of their potential to inform and transform care.

In other countries, Minimum Data Sets (MDSs) for care homes already exist. The most widely recognised of these are the US Medicare Minimum Data Set (MDS 3.0) [7] and InterRAI, deployed in multiple jurisdictions [8]. Implementation of MDSs is influenced by mandates and financial incentives supported by: ongoing training to motivate staff to engage with MDS completion; the extent to which completion is built into the working practices, monitoring, and record systems of all staff (including visiting professionals); and digital recording systems that care home staff use to document and discuss care [9]. At the time of writing, there is no national mandate or incentive framework for implementation of an MDS in any of the four UK nations, although plans are underway to standardise some aspects of social care data collection in England [10].

Against this background, we set out to pilot a prototype MDS for English care homes for older people, focusing on homes currently using digital care records (DCRs) [11]. Our objectives were to: (1) assess feasibility of extracting data directly from DCRs and linking these to routinely collected health and social care data to populate a pilot care home MDS; (2) to assess quality and completeness of MDS data; and (3) describe barriers and facilitators to implementation and use. In this article, we address the first two of these objectives. Implementation and use by care home staff and external stakeholders are addressed in a second paper [12].

Methods

This was a mixed-methods pilot of a prototype MDS. A full protocol is published elsewhere [11].

Sampling and resident recruitment

We aimed to recruit 20 care homes for older people in each of three Integrated Care Systems (ICSs), totalling 60 homes. ICSs are regional partnerships between NHS organisations, local government and others including third sector and social enterprises, which are responsible for co-ordinating and commissioning care in England. From the 42 English ICSs we chose three – in the South East, East Midlands, and North East – to sample different geographies, socio-economic deprivation indices, and care configurations. Assuming an occupancy rate of 90%, the sample size required for a true representation of the finite older care home population in each of the ICSs, with 90% confidence and 5% margin of error, was 262-268 residents per ICS [11].

Care homes were eligible for inclusion if using DCRs from one of two participating DCR software companies. Initial approaches were made by email, telephone and in-person, with homes recruited from those responding positively.

All permanent residents of participating care homes were eligible. We excluded: residents receiving respite or temporary/short stay care to minimise burden for people undergoing acute transitions; and

residents identified as in the last few days of life by care home staff to protect residents and families at a difficult time. Consent was obtained from residents to access and extract pseudonymised data from their care home, health and social care records and, separately, to link these. Capacity to provide consent to participate was assessed by a researcher at first meeting. For those without capacity, we asked care home staff to send a letter to a family member or friend who could act as a personal consultee as defined by the Mental Capacity Act. Consultee discussions were conducted either face-to-face or by telephone.

Selecting items for inclusion in the prototype MDS

MDS development was based upon: a review of international research literature summarising outcome measures used in care home studies [13]; a review of measures used in UK care home randomised controlled trials [14]; a systematic review on how contextual factors influence research processes, including data collation in care homes [15]; a series of consultation activities with stakeholders comprising care home managers and staff, and clinical specialists in healthcare of older people and primary care [16–18]; public involvement activity with care home residents, staff and family carers[19]; a survey of data currently collected and collated by English care homes [3]; and a scoping review of published MDSs. From these, we developed nine core principles to govern development and implementation of a care home MDS, previously published [20] (reproduced in Appendix 1).

A corollary of these findings was that a major barrier to implementing an existing MDS already deployed in other jurisdictions, such as interRAI or MDS 3.0, was the need for care homes to stop using existing DCR software and data approaches to start using these products. None of our care home stakeholders were motivated to do so. Additionally to form a complete dataset, interRAI or MDS 3.0 would either have to duplicate or replace data held in NHS records. The ability to draw from and connect with data already held was seen as important based upon our stakeholders. Therefore, based upon the nine core principles, we compiled an aspirational prototype MDS, containing agreed information and a plan for which routine datasets we hoped to collect these from [11] (summarised in Appendix 2). The systematic review on howcontextual factors influence research processes[15] informed our approach to MDS implementation. Having established what an aspirational MDS should contain, we then met with DCR providers to explore the variables contained in their datasets.

Digital care records (DCRs)

19 DCR software providers are at the time of writing accredited by NHS England for use in English care homes [21]. We worked with the independent Care Software Providers Association (https://caspa.care) to identify two leading care management software providers, who between them provide care software to 9500 of the circa 17000 care homes in the UK. Through an initial mapping exercise, based on demonstration of a 'standard' user interface by the software providers, we identified variables from the aspirational MDS likely to be included in DCRs.

A dummy data extract from both software providers, completed in summer 2022, identified several variables collected in free text or non-standardised formats. To address gaps in the MDS left by these, that could not be addressed through routine NHS and social care data, additional measures were added to each software system. These included seven validated measures of: *delirium* (I-AGeD) [22]; *cognitive impairment* (MDS Cognitive Performance Scale (MDSCPS)) [23]; *functional independence* (Barthel index)[24]; and *Quality of Life* (QoL) from the Adult Social Care Outcomes Tool Proxy (ASCOT-Proxy-Resident) [25,26], EuroQol 5 domain 5 level proxy version (EQ-5D-5L Proxy 2) (EuroQol)[27], ICECAP-O [28], and QUALIDEM [29].

The QoL measures were selected based on evidence of use in care homes, psychometric properties [30], relevance to different QoL constructs (health, social care, dementia, and older people), and advice from stakeholder consultations and public involvement activityClick or tap here to enter text.. Our decision to include QUALIDEM, rather than DEMQoL, as a dementia-specific Quality of Life measure was based upon rankings by stakeholders, published in full elsewhere [18]. Taking into account the high prevalence of cognitive impairment in care home residents[1], proxy versions were used. We further included the ASCOT pain item and low mood/anxiety subscale[31], as well as a question to rate overall QoL on a 7-point scale. This overall question was for resident completion where possible, or otherwise by staff proxy. The type of help needed by the resident, if any, was recorded.

Researchers provided specifications for the user interface format, data extract and outputs for these measures, which were then implemented by software providers and tested by researchers using a pilot interface, with revision as needed. In this process, it became evident that some specifications were not possible in both systems due to differences, for example, in how they dealt with missing data and/or because requirements were incompatible with a system's usual function or output.

Researchers met with care home staff to describe and explain the additional variables, and to highlight the need for these to be inputted manually in addition to usual care records. For routinely collected variables, data were extracted from existing records without additional input from care home staff, in the format(s) used by care homes and in an output format feasible for each software provider. This minimised burden on care homes and software providers but meant researchers had to clean raw data and derive variables.

All DCR variables were collected twice, six months apart, in March-June and September-November 2023. We collected a small amount of data directly from care homes through a short online survey at baseline to better understand context of care, including number of beds, residents, self-funding residents and staff employed by the care home.

Routinely collected health and social care datasets

We aimed to access the following data sources: general practice electronic medical records and prescribing data, hospital administrative data, operational datasets from emergency services, urgent care and community health, data from local authorities on social care funding, and data from CQC. We expected to access some of these sources at national (e.g. administrative hospital data) and others at local (e.g. community health) level (Appendix 2).

We developed a data flow diagram (Appendix 3) and legal bases for data sharing (Appendix 4).

Data management and linkage

As the Improvement Analytics Unit based at The Health Foundation (THF) led data management and linkage, data were hosted on THF's secure ISO27001/DSPT accredited Data Analysis Platform (DAP). Data were stored in AWS S3 buckets which only Data Managers and approved project data analysts could access. Access to data was controlled by Data Managers.

For extracts of health and social care information held by different data controllers to be created, pseudonymised and shared with THF, we securely transferred to software providers a unique NHS number salt key to enable pseudonymisation of subjects in the study. A separate salt key was used to pseudonymise the Care Quality Commission (CQC) location identifier (unique for each home). Both salt keys used the SHA256 hashing algorithm. Care home pseudonymisation minimised risk of re-identification of individuals based on location. Care home software providers securely transferred extracted DCRs and pseudonymised NHS numbers and care home identifiers for included residents. Data managers isolated pseudonymized NHS numbers and used a pre-computed rainbow table (password cracking tool) of hashed NHS number and salt combinations to determine actual NHS numbers of subjects. These were securely transferred to data processors of health and social care data to enable extraction of relevant records of consented residents. Salt keys were separately transferred so data processors could pseudonymise NHS numbers and care home identifiers in extracted health and social care information. Pseudonymised records were securely transferred to THF once all other identifiers were removed.

Non-personal, aggregated care home-level online survey data from care homes in the study were securely transferred to THF by University of Kent and pseudonymised by THF.

The salt keys, rainbow table of hashed NHS number and salt combinations, and data from the survey of care providers with clear CQC location identifiers from University of Kent, were stored in a location accessible only by Data Managers, separated from the extracted pseudonymised DCRs and health and social care records, and deleted after the datasets were linked.

Once data were received from data processors, the data was checked and cleaned and variables were derived (derivation described in Appendix 5a). Data cleaning and variable derivation code is published on Github: https://github.com/HFAnalyticsLab/DACHA. Datasets were linked via pseudonymised NHS numbers and pseudonymised CQC location identifiers.

Stakeholder engagement

We engaged technical experts within NHS England (NHSE) and the ICSs on information governance, data access and availability. We also engaged with wider stakeholders within each ICS to gain support for the project, to facilitate data sharing, and to inform analyses to be conducted on the MDS. Stakeholders included care home managers, staff, residents and family members, GPs, and local decision makers within the NHS and local authority. We also engaged with DHSC and NHSE programme teams (Enhanced Health in Care Homes and Ageing Well) at a national level to understand how an MDS could inform national policy priorities.

Importantly, initial buy-in from the three ICSs at the start of the study, three years before resident consent and data collation began, dissipated by the time discussions around data access started. This was due both to key stakeholders leaving and competing priorities for limited analytical and IG resource. Stakeholders who were able to influence data access and had clinical contact with care homes to inform discussions about data analysis differed between ICSs.

Deriving MDS variables

We designed a person-level, one row per resident MDS. The date on which additional care home measures were first completed by care home staff, or 1 June if missing, was the index date for all other MDS variables. The Elixhauser list of comorbidities[32,33] and a validated list of frailty syndromes [34] were identified from hospital admission data using ICD-10 codes for 3 years prior to each resident's index date. Potentially avoidable admissions were those due to a list of conditions originally developed by the Care Quality Commission [35]

Healthcare utilisation was collated for the year before the index date. By exception, ambulance activity was only calculated for the period between the first and second MDS measurements. "Out of hours" was defined as 18:00-08:00 and "long attendance" as being at Emergency Departments for 12 hours or longer. All variable derivations are detailed in the final MDS data specification (Appendix 5).

Data Analysis

Where variables were available from multiple data sources, we compared levels of completeness and agreement. To determine which data source(s) would populate the final MDS, we constructed a hierarchy, based on data quality and expert opinion. We distinguished between variables with a universal definition across datasets, such as date of birth or sex, and those which could be defined in multiple ways or vary over time, such as cognitive impairment or delirium. For the first category, we created a hierarchy collapsing all sources into one final variable. For the second, we presented a comparison but retained all variables in the final MDS. By exception, we took an additive approach for dementia. We used Personal Demographic Service (PDS)[36] as the master index based on NHSE guidance, and Secondary Uses Service (SUS) [37] where data were unavailable in PDS. The exception was ethnicity, where we used the care home record in the first instance, as self-reported ethnicity is more accurate than observational data commonly found in secondary care records [38,39].

Date of death can often generate disagreement between systems, mainly because dates of death notification and certification by the Office of National Statistics may differ [40]. However, they rarely vary more than 30 days, with negligible effect on analysis.

To understand the information contained within the MDS, we reported proportions of missingness, mean and standard deviation (SD) or frequency (%) as appropriate. We also derived two-way tables to provide worked examples of opportunities for more detailed descriptive statistics from the MDS, focusing on

emergency attendances and ambulance activity based upon discussions with stakeholders described above.

Evaluation of psychometric properties of the QoL measures (ASCOT-Proxy-Resident, ICECAP-O, EQ-5D-5L Proxy 2, QUALIDEM) are reported elsewhere [39,40]. These analyses identified limitations around using QUALIDEM in an older adult care home MDS, so we do not report QUALIDEM results here.

The analysis code is published on Github: https://github.com/HFAnalyticsLab/DACHA. We used R version 4.0.2, SAS Enterprise Guide version 8.3 (NHS and social care routine data), and Stata version 18 (DCR data).

Results

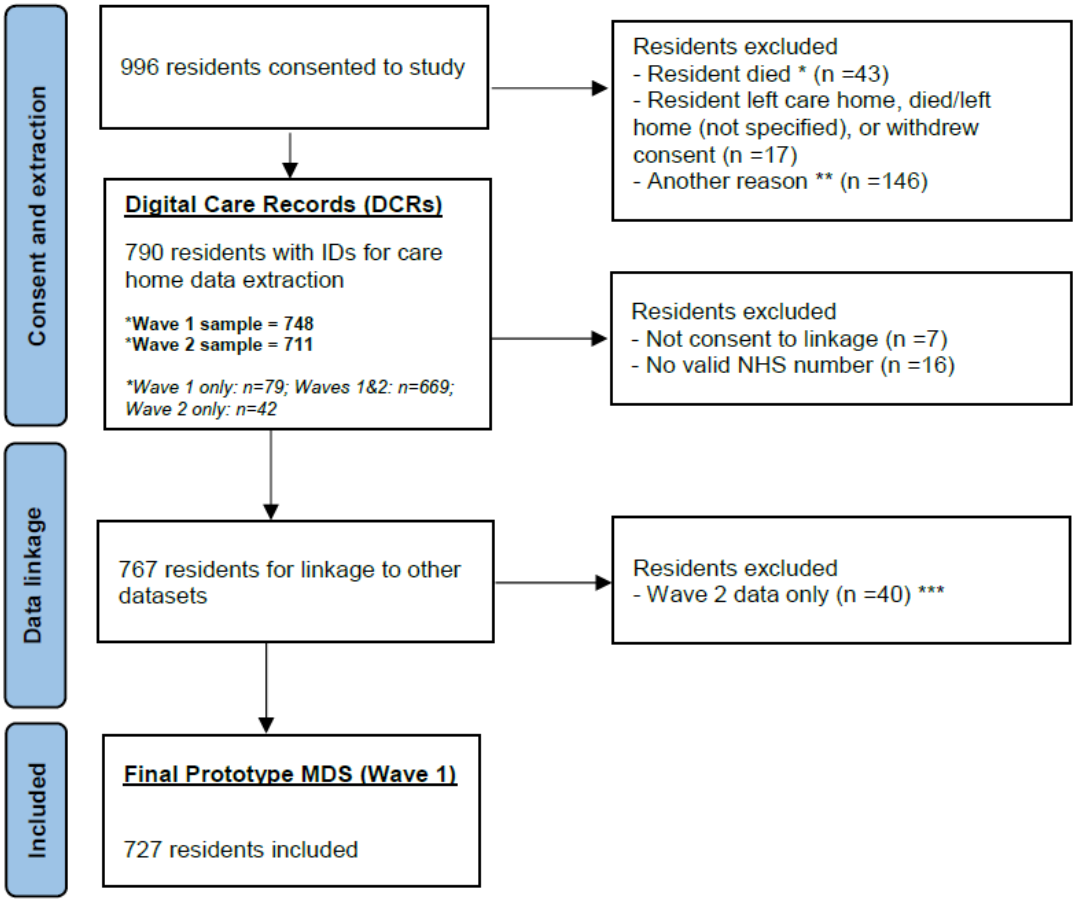
We recruited 996 residents from 45 care homes (Table **Table**). Working from lists of care home providers using particular DCR software meant brokering relationships with care homes often new to research. Success was greatest in ICS Area 1 because of long-established relationships between the researchers and their local care home community.

Table 1- Actual versus target recruitment by ICS area

ICS Area	Target recruitment	Actual recruitment
1	20 care homes	19 care homes
	320 residents	537 residents
2	20 care homes	15 care homes
	358 residents	286 residents
3	20 care homes	11 care homes
	292 residents	173 residents
Total	60 care homes	45 care homes
	970 residents	996 residents

From 996 eligible residents, 767 had data extracted which could be linked. Of these, 727 residents had complete data for baseline DCR data collection and were included in the final prototype MDS (Figure 1). Of these, 696 had a DCR with a valid CQC identifier enabling linkage to care home level data from CQC records and the online survey.

Figure 1 – Flow diagram of residents from recruitment into final prototype MDS (Wave 1)



* This only includes those residents who died before the data extraction from digital care records for Wave 1.

**These include (1) where care homes did not communicate to the research team that the resident had died or moved out of the care home (i.e., in addition to the n= 43+15 notified cases); and (2) where residents were not flagged by care home staff within the system, as was required by Provider 2, to enable them to identify participating residents and extract data from the digital care records. Of the n=146 cases, it is likely that n=97 are due to reason (2), which affected six care homes using Provider 2 software.

*** In these cases, data extracted from digital care records was missing at Wave 1, but not Wave 2. For the pilot MDS Wave 1 dataset, we excluded those residents without data from digital care records for Wave 1.

Digital Care Records (DCRs) from Care Homes

First, we describe DCR data extracted from care homes (Table 2) before we consider accessed datasets and subsequent linkage into the final prototype MDS. Table 2 includes data for the 790 residents (see Figure 1, under consent and extraction) who provided consent and had a valid ID for data extraction (n=748 at Wave 1, n=711 at Wave 2). For residents with complete data at Wave 1, but not Wave 2, most were attributable to resident death or care home drop-out from the study between waves (see Appendix 6).

Where data were already included in routine DCRs, some variables were more complete than others. CPR status was 99.6% complete. Care homes using Software Provider 2's system did not routinely complete fields including marital status, first language, power of attorney and malnutrition universal screening tool (MUST), which contributed to high levels of missing data. For National Emergency Warning Score 2 (NEWS2) variables, no data were entered by care homes using either software.

The measures added to DCRs for the pilot were more consistently completed compared to those routinely recorded (Wave 1: <35% missing data). This is perhaps to be expected, since we required software

providers to include these measures across participating homes, whereas homes could choose what routine data to record. We also devoted researcher time to explain the new variables and the rationale for their inclusion to care home staff.

In comparing Wave 1 and Wave 2, missing data increased by >8% for deprivation of liberty (7.9% to 21.2% for waves 1 and 2 respectively). For variables added to DCRs, missing data increased between 10% and 18% from Wave 1 to 2, except for the Barthel Index (increased by 36%) and ICECAP-O (increased by 37%). For Barthel, this was likely due to Provider 2 using a similar, but slightly different, version as their system default, which care homes reverted to using rather than the standardised version added for the study. Provider 2 did not return ICECAP-O data for five care homes at Wave 2.

Even with relatively high completion for QoL measures, there were issues with data quality in Wave 2. Provider 2 'carried over' Wave 1 scores; therefore, care homes had to manually overwrite prepopulated scores. By contrast, Provider 1 required data entry of new scores for Wave 2. As a result, all but one care home using Provider 2 software had a maximum of two residents with any change in ASCOT-Proxy-Resident score between Wave 1 and 2, whereas only three residents had the same ASCOT Proxy-Resident score across waves for homes using Provider 1's software.

Table 2 – Overview of data extracted from digital care records

			Wave 1			Wave 2			
MDS	**Variable**	**Categories (if applicable)**	**n**	**Mean, SD. (Range) or Freq. %**	**% Missing data [1]**	**n**	**Mean, SD. (Range) or Freq. %**	**% Missing data [2]**	**Other comments**
1	Ethnicity	Asian or Asian British White or White British	163	≤4% ≥96%	78.2%	122	≤4% ≥96%	82.8%	Derived variable. Coded differently by the two software providers.
1	Religion	No religion Christian Buddhist Other	119	11.8% ≥76.5% ≤5% 6.7%	84.1%	89	9.0% ≥79% ≤6% ≤6%	87.5%	Derived variable. Coded differently by the two software providers.
1	Marital Status	Married/cohabit Widowed Divorced/single/separated	114	35.1% 50.9% 14.0%	84.8%	120	34.2% 49.2% 16.6%	83.1%	Not available for Provider 2 (100% missing).
1	First Language	English Other	170	≥96% ≤4%	77.3%	134	≥96% ≤4%	81.2%	Not available for Provider 2 (100% missing).
1	Deprivation of Liberty	No Yes	689	81.3% 18.7%	7.9%	560	76.8% 23.2%	21.2%	
1	Weight	20-35kg 36-50kg 51-65kg 66-80kg 81-95kg 96-110kg 111-125kg 126-140kg	586	1.2% 19.8% ≥39.4% 26.3% 7.7% 3.6% ≤1% ≤1%	21.7%	573	1.9% 19.2% ≥37.6% 26.9% 9.1% 3.3% ≤1% ≤1%	19.4%	Derived variable. Provider 1: numerical, Provider 2: categorical. Majority of missing data (≥95%) are from Provider 2. §
1	Height	111-125cm 126-150cm 151-170cm 171-190cm 191-210cm	738	1.1% 9.8% ≥70.1% 18.0% ≤1%	1.3%	674	1.2% 11.0% ≥68.2% 18.6% ≤1%	5.2%	Derived variable. Provider 1: numerical, Provider 2: categorical.
1	DNACPR Status	No Yes	745	21.1% 78.9%	0.4%	707	18.7% 81.3%	0.6%	
1	Power of Attorney	No Yes	170	62.4% 37.6%	77.3%	168	60.7% 39.3%	76.4%	Not available for Provider 2 (100% missing).
3	Length of stay, days		747	873.7, 807.5 (<50 to >8,000)	0.1%	710	1,011.4, 815.6 (<50 to >8,000)	0.1%	Derived from date of entry to home. For Provider 2, only have month/year, so set to the 1st of the month. Data quality was improved across waves, since able to verify and correct

									anomalous dates (e.g. all set to the same date within a care home). Nevertheless, one case (n=1) omitted due to likely data entry error.
4	Textured food/diet	IDDSI 7 – regular IDDSI 7 – easy to chew IDDSI 6 – soft & bite sized IDDSI 5 – minced & moist IDDSI 4 – pureed IDDSI 3 - liquidized	687	≥71.3% 5.1% 9.5% 7.1% 6.0% ≤1%	8.2%	611	≥68.3 6.7% 8.8% 8.2% 7.0% ≤1%	14.1%	
4	Textured drink/fluid	IDDSI 0 – thin IDDSI 1 – slightly thickened IDDSI 2 – mildly IDDSI 3 – moderately IDDSI 4 – extremely	672	≥89.9 4.5% 3.6% ≤1% ≤1%	10.2%	605	89.2% 4.8% 5.0% ≤1% None	14.9%	
4	Cognitive impairment	Very severe Severe Moderately severe Moderate Mild Borderline intact Intact	630	12.5% 14.4% 12.9% 17.9% 14.0% 9.4% 18.9%	15.8%	479	14.6% 17.3% 12.9% 18.6% 12.1% 7.7% 16.7%	32.6%	Added to software for the MDS pilot. The MDS CPS is calculated from five items: comatose, problem with short-term memory, cognitive skills for daily decision making, being understood by others, and eating ADL. Scored per Morris et al[42].
4	Waterlow Score		542	17.8, 7.3 (5 to 44)	27.5%	473	18.3, 7.3 (5 to 43)	33.4%	Score of ≥10 indicates risk for pressure ulcer, with high risk ≥15 and very high risk ≥20. Full score range from 0 to 64. §§
4	Braden Score		345	16.6, 4.0 (7 to 23)	53.9%	277	16.4, 4.0 (8 to 23)	61.0%	Full score range of 6 to 23, with higher scores indicating lower risk of pressure ulcers. Scores ≤10-12 indicate high risk and ≤9 very high risk.
4	Barthel Index		582	41.5, 30.2 (0 to 100)	22.2%	288	34.9, 28.7 (0 to 100)	59.5%	Added to software for the pilot. Score from lowest (0) to highest (2) level of functional independence for each item. These are summed, x5, to create a score from lowest (0) to highest (100) independence.
4	Delirium / I-AGED		601	1.1, 1.8 (0 to 10)	19.7%	466	1.4, 2.1 (0 to 10)	34.5%	Added to software for the pilot. Each of the ten items is scored no (0) or yes (1) and summed to create a scale from 0 to 10. Score of ≥4 indicates delirium[43].

5	ASCOT Proxy-Resident		503	.83, .19 (-.17 to 1)	32.8%	384	.81, .19 (-.17 to 1)	46.0%	Added to software for the pilot. Required some recoding to combine. Applied preference weights for ASCOT SCT4 to generate index score from -.17 to 1.0 [44].
5	ASCOT: anxiety and low mood		503	4.0, 1.5 (0 to 6)	32.8%	403	3.9, 1.5 (0 to 6)	43.3%	Added to software for the pilot. Required some recoding to combine.
5	ASCOT: Pain		568	2.2, .8 (0 to 3)	24.1%	423	2.1, .9 (0 to 3)	40.5%	Added to software for the pilot. Required some recoding to combine.
5	ICECAP-O		583	.73, .21 (0 to 1)	22.1%	300	.71, .22 (0 to 1)	57.8%	Added to software for the pilot. Required some recoding to combine. Provider 2: Data not provided for some residents. Score (0 to 1) calculated using UK index values [45].
5	EQ-5D-5L Proxy 2		650	.33, .35 (-.59 to 1)	13.1%	494	.29, .34 (-.35 to 1)	30.5%	Added to software for the pilot. Required some recoding to combine. Score calculated using the mapping function to convert to EQ-5D-3L and applied UK index values. The UK value set for the EQ-5D-5L is still being developed [46,47]
5	ASCS QoL	So good Very good Good Alright Bad Very bad So bad	613	3.1% 26.3% 33.9% 28.9% 5.2% 1.5% 1.1%	18.1%	461	3.9% 25.8% ≥36.1% 27.3% 4.6% 1.3% ≤1%	35.2%	Added to software for the pilot. Required some recoding to combine.
6	NEWS2/ RESTORE2		0	N/A	100%	0	N/A	100%	Provider 1. Included in the software for data capture but no data entered by care homes (100% missing). Provider 2. Not available (100% missing).
6	MUST		169	.8, 1.2 (0 to 5)	77.4%	161	1.0, 1.3 (0 to 4)	77.4%	MUST scored in last 6 months. A MUST score of 1 indicates medium risk and ≥2 indicated high risk of malnutrition. Not available for Provider 2 (100% missing).

[1] **Wave 1:** n=748 (Provider 1 n=170; Provider 2 n=578)
[2] **Wave 2:** n=711 (Provider 1 n=168; Provider 2 n=543).
§ Included a flag to indicate cases with multiple entries (n=8) for Provider 2. As date of entry was unavailable, one entry was randomly selected.
§§ Included a flag to indicate cases with multiple entries for an individual (n=69) for Provider 2. As date of entry was not available, one entry was randomly selected.

Accessed routinely collected health and social care datasets

We were able to retrieve and link data from PDS, SUS Admitted Patient Care, Outpatient and Emergency Care datasets, CQC care home data and supplement this with data from our online survey of care homes as planned. We were additionally able to collect data from the newly available national ambulance [48], adult social care client level [49], and community services (CSDS) [50] datasets. A care home residency table created by Arden & GEM Commissioning Support Unit [51] based on PDS data and estimated care home residency dates, and ONS Index of Multiple Deprivation data were also accessed.

Due to information governance constraints, a new data sharing agreement with NHSE was required, which was signed in October 2023. This delayed access to NHSE datasets and restricted the analysis possible in the remaining time. This also adversely impacted set up of data sharing with ICSs.

These datasets were accessed only for consented residents and not for all care home residents in the ICSs as originally planned [11]. In addition to IG challenges, this was primarily because the underlying flow of data previously used to identify care home residents had been replaced, resulting in the complex algorithm [46] for care home identification needing to be redeveloped and validated by NHSE.

We were unable to access GP records because we couldn't establish data sharing agreements for two of the ICSs in time for the study. In the remaining ICS we were able to secure some data sharing agreements with GP practices by working through a Commissioning Support Unit (CSU), a regional body providing data support to NHS organisations. However patient data are held by individual GP practices, and we had to liaise with multiple Data Protection Officers within the same ICS. Ultimately, the number of resident records available from GP practices that signed agreements in time was too low to ensure residents could not be re-identified, and therefore it was not possible to proceed to extraction under General Data Protection Regulations (GDPR). A list of data items we would have accessed from one ICS where we established data sharing agreements, had we been able, is available in Appendix 7.

The inability to collect GP data was a major contributor to the differences between the aspirational and final prototype MDS, summarised in Appendix 8. Other contributors were poor feasibility of extraction from DCRs and high levels of missing data for some items in routine datasets, rendering reliable counts of activity linked to particular conditions or events impossible.

Creating derived variables in prototype MDS

Due to the absence of GP data, comorbidities were derived from SUS data, using a 3-year lookback period from the index date. We couldn't derive these for 144 residents (20%) who didn't have a hospital admission in that period. Activity summaries were reported for the year leading up to the index date, independent of whether residents joined their current care home within this time period. On average, residents in Wave 1 had been living in the current care home for 28.7 months, with 29% having moved in within the year leading up to their index date.

Hierarchy process

Table 3 presents the variables included in the hierarchy. For universally defined variables, there were high levels of consistency where recorded. Levels of completeness varied widely – from 1% missing for sex in CSDS to 80% missing for ethnicity in the care home record. Overall, the process of using information from several sources to populate the

final variable included in the MDS greatly reduced the level of missing data (missingness <=4% across variables).

Table 3 – Comparison of variables across data sources to determine hierarchy

Variable	Category	PDS		SUS		CSDS		Care home residency		Care home record		n with non-missing values in both/ all relevant data sets and % agreement														Final [4]	
												SUS vs CSDS		SUS vs PDS		SUS vs CHR		SUS vs care home record		CSDS vs PDS		CSDS vs care home record		PDS vs CHR		PDS vs care home record		All			
		n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%
Ethnicity	White	NA		586	81%	570	78%	NA		>=151	>=20%	473	98%	NA		NA		125	99%	NA		92	98%	NA		NA		78	99%	692	95%
	Black or Black British			0	0%	<=5	<=1%			0	NA																			<=5	NA
	Asian or Asian British			0	0%	0	0%			<=5	<=1%																			<=5	NA
	Mixed			<=5	NA	0	0%			0	NA																			<=5	NA
	Other			<=5	NA	<=5	<=1%			0	NA																			<=5	NA
	Missing			135	19%	149	20%			573	79%																			25	4%
Sex[1]	Female	453	62%	487	67%	504	69%	NA		178	23%	684	100%	NA		NA		240	99%	NA		245	99%	NA		NA		198	99%	542	71%
	Male	175	24%	205	28%	214	29%			73	10%																			225	29%
	Missing	99	14%	35	5%	9	1%			476	65%																			0	0%
Date of birth	Available	628	86%	692	95%	NA		NA		NA		NA		598	100%	NA		NA		NA		NA		NA		NA		NA		>=760	>=99%
	Missing	99	14%	35	5%																									<=5	<=1%
Record of death[2]	Present	<=5	NA	9	1%	NA		57	7%	NA		NA		<=5	100%	8	100%	NA		NA		NA		<=5	100%	NA		<=5	100%	58	8%
	Not present	>=720	>=99%	718	99%			670	92%																					709	92%

Dementia[3]	Yes	NA	394	54%	NA	NA	376	52%	NA	NA	NA	342	75%	NA	NA	NA	NA	NA	514	71%
	No		189	26%			199	27%											191	26%
	Missing		144	20%			152	21%											22	3%

[1] Source data sets refer to gender but data are recorded as 0/1 and labelled as male/ female so we understand this to be sex

[2] Date of death was determined as agreed where the two dates were within 30 days of each other

[3] Two code lists were used to identify dementia in SUS diagnosis codes (Charlson and frailty; Appendix 9). Additive approach taken where a record with either was identified as dementia present in SUS record.

[4] Collapsed variables were formed using the hierarchy PDS > SUS > CSDS > care home residency > DCR, with exception of ethnicity. For dementia, a record in either SUS or DCR resulted in a record in the final variable.

Final prototype MDS

Key variables from the final prototype MDS are summarised in Table 4. Appendix 10 shows the full version, which includes two approaches to healthcare utilisation – mean activity across all residents, and proportion of residents with at least one event. Appendix 11 contains worked examples, based upon our work with stakeholders, of how data from the MDS could be used to help understand Emergency Department and Ambulance contacts.

Table 4 – Selected variables from final prototype MDS. Numbers are reported for 727 residents unless otherwise specified

Domain	Variable	Categories (if categorical)	n	Mean (SD) or %
Demographics/ characteristics	Ethnicity (final)[1]	White	692	95%
		Black or Black British	<=5	NA
		Asian or Asian British	<=5	NA
		Mixed	<=5	NA
		Other	<=5	NA
		Missing	25	3%
	Sex (final)[1]	Female	513	71%
		Male	214	29%
	Date of birth record (final)[1]	Available	>=720	99%
		Missing	<=5	NA
	Date of death present in record (final)[1]	Present	58	8%
		Not present	669	92%
Palliative care needs	Discussed preferred death location indicator	Yes	18	3%
		No	383	53%
		Missing	326	45%
	Preferred death location	Care home	7	1%
		Care home services with nursing	27	4%
		Care home services without nursing	51	7%
		Hospice	<=5	NA
		Hospital	<=5	NA
		Patient's own home	16	2%
		Other (not listed)	<=5	NA
		Missing	623	86%
Care home stay	Client funding status	Health funded	7	1%
		Social care funded	18	2%
		Client funded	19	3%
		Joint client and social care funded	96	13%
		Other	<=5	NA
		Unknown in record	77	11%
		Missing	>=505	70%
Residents needs	Cognitive impairment	Borderline intact	56	8%

		Intact	116	16%
		Mild impairment	85	12%
		Moderate impairment	111	15%
		Moderately severe impairment	80	11%
		Severe impairment	88	12%
		Very severe impairment	76	10%
		Missing	115	16%
	Functional independence (Barthel index) (reported for 566 residents/ 22% missing)			41.40 (30.26)
Quality of life	Ascot Proxy-Resident (reported for 488 residents/ 33% missing)			0.83 (0.19)
	ICECAP-O (reported for 569 residents/ 22% missing)			0.73 (0.21)
	EQ-5D-5L Proxy 2			0.33 (0.35)
Diagnoses (based on previous 3 years hospital admission diagnosis codes) (reported for 583 residents/ 20% missing) apart from 'dementia (final)'	Dementia (final)[1]		514	71%
	Elixhauser conditions[2]			
	Number of Elixhauser conditions			3.59 (2.34)
	2 or more Elixhauser conditions		470	81%
	Anaemia		83	14%
	Congestive heart failure		86	15%
	Chronic pulmonary disease		110	19%
	Depression		129	22%
	Diabetes (complicated and uncomplicated)		127	22%
	Fluid and electrolyte disorders		226	39%
	Hypertension (complicated and uncomplicated)		353	61%
	Hypothyroidism		75	13%
	Liver disease		30	5%
	Obesity		39	7%
	Other neurological disorders		154	26%
	Peripheral vascular disease		47	8%
	Rheumatoid arthritis / collagen vascular diseases		179	31%
	Renal failure		39	7%
	Valvular disease		67	11%
	Weight loss		19	3%
	Frailty syndromes[3]			
	Number of frailty syndromes			2.17 (1.81)
	Cognitive impairment (delirium, dementia, senility)		457	78%
	Anxiety/Depression		168	29%
	Functional dependence		102	17%
	Falls/Fractures		291	50%
	Incontinence		105	18%
	Mobility problems		217	37%
	Pressure ulcers		62	11%
Healthcare utilisation			n (people with at least one event)	% who had at least one event

	Elective admissions (1 year history)		65	9%
	Emergency admissions (1 year history)		284	39%
	Potentially avoidable emergency admissions (1 year history) [4]		119	16%
	Emergency department attendances (1 year history)		370	51%
	Community services appointments (1 year history)		608	84%
	Face to face community services appointments (1 year history)		444	61%
	District nursing appointments (1 year history)		398	55%
	Ambulance call outs (1 June - 31 October 2023)		197	27%
	Ambulance attendances (1 June - 31 October 2023)		195	27%
	Ambulance conveyances (1 June - 31 October 2023)		147	20%
Care home characteristic and workforce characteristics	Service type	Nursing	403	55%
		Nursing and Residential	49	7%
		Residential	262	36%
		Missing	13	2%
	Registered bed capacity	Less than 50	211	29%
		50 or more	485	67%
		Missing	31	4%
	CQC rating	Outstanding	72	10%
		Good	511	70%
		Requires improvement	113	16%
		Missing	31	4%
	Years of service registration	Less than 10 years	238	33%
		More than 10 years	458	63%
		Missing	31	4%

Footnotes

1 - reporting variable as created in the hierarchy process – see Table 3
2 – Elixhauser list of comorbidities [32,33]
3 – Frailty Syndromes [34]
4 – Potentially avoidable emergency admissions [35]

Discussion

In the face of substantial challenges, many of which were not unique to this study[52,53], we accessed information from care home DCRs and safely linked data from multiple sources and data owners to create a viable prototype MDS for English care homes. Our prototype MDS was cross-sectional. Real-world deployment would be longitudinal, with data extracted at regular intervals, balancing the requirements of those funding, planning and delivering services against burden of data completion and collation.

We set out to collate routine administrative health and social care data for all care home residents in participating ICSs, with linkage to DCRs taking place only for those giving consent. This should have been technically feasible using methods outlined in this paper alongside a published algorithm to identify care home residents in routine data [54]. However, the algorithm was under redevelopment at the time of our pilot and couldn't be validated in time to be incorporated in our data flow. Our final prototype MDS was therefore limited only to residents providing consent to linkage. This may have introduced

systematic bias and data presented here should not be seen as representative of the wider UK care home population. For example, our data on healthcare resource use should be interpreted with caution – we do not know how health status influenced ability to provide consent.

Our data on health status, meanwhile, are limited by lack of access to GP records. This is reflected in lower reported prevalence of common conditions, such as dementia, than in previously published studies, although the prevalence based upon MDS CPS corresponds better to the prevalence cited elsewhere [1,55]. Long-term conditions such as incontinence and hearing loss, central to understanding healthcare needs in care home residents, are under-recorded in secondary care records [55]. If the MDS presented here is to be of use in practice, incorporating GP data is essential. The challenges encountered accessing GP data related to information governance and our role as researchers external to the ICS, coupled with time constraints. It was not due to resistance to the principle of data linkage. GP practices work as independent contractors commissioned by the NHS, each practice acts as data controller for their own patients' data and there is as of yet no national GP dataset.

Our design repurposed routinely collected care home data to minimise care staff burden and focus on capturing what was important to staff and residents. Where data were central to routine care delivery – such as CPR or Deprivation of Liberty status – they were largely complete. Variables that were incomplete were either: regarded as superfluous because care staff know these for their residents (e.g. ethnicity or marital status); captured in free text and difficult to analyse; or difficult to record in a dependent population, (e.g. weight). Variables added via external mandate (e.g. NEWS2, included at the request of healthcare providers) [56], were not completed. For variables added to DCRs by our research team for the pilot, we saw initial high completion rates fall during the second wave of data collection. This was multifactorial, with competition for staff time, staff attrition, and implementation issues including a duplicate Barthel index in some care homes' software, all contributing. We did not collect data on the amount of staff time spent completing additional variables – this limits our understanding of the quantitative impact of doing so upon their workload. These findings align with previous research on the importance of understanding the context of data collection when working with and interpreting data from social care [4,57].

An alternative approach to the one used here would be to implement an "off the peg" internationally validated MDS, such as interRAI or MDS 3.0. These would have a number of advantages including deploying well established and validated variables, deployed in a consistent way through licensed software, and which are regularly updated through reference to the evolving gerontological literature [8]. It is important, though, to note that this approach would not necessarily be a viable alternative for the UK. It superimposes a new system of data capture onto care homes, not linked to health and social care data held elsewhere and favours health data over quality of life and social care data. The issues we addressed around GDPR, the labour intensive and manual nature of linkage between care home and NHS data, the complex hierarchy of statutory databases into which an MDS has to interdigitate, and the need to train and invest in care home staff over time, would be the same. The interRAI is able to be deployed across multiple care settings, including acute hospitals and domiciliary care [8] – approaches to care records in these sectors in the UK are at least as fragmented as in the care home sector and could benefit from harmonisation, but the complexity of deploying a universal data solution increases with the number of care sectors involved. Previous attempts to use such MDSs in UK research studies found low completion rates and crucially, a higher burden associated with staff completing them on top of existing data requirements [58,59]. The work of implementation for uptake and sustained use is as significant and arguably more resource intensive than we found for our prototype dataset[60]. Dwelling excessively upon such approaches also misses the substantial progress made across health and social care data integration in multiple parts of the UK [61,62]. The challenge is to connect a care home

MDS into such approaches – a top-down reorganisation to implement a dataset developed elsewhere, and in other contexts, is at odds with these approaches focussed around making the most of what is already collected, and empowerment and enfranchisement of localities and the people that live and work within them.

We faced issues with standardising approaches to data collection across two software providers and forty-five care homes. This variation in approach across different providers would be multiplied if the approaches described were rolled out to all 19 DCR providers accredited by NHS England. Plans underway by the Department of Health and Social Care to develop a Minimum Operating Data standard (MODS) [10] might facilitate some standardisation going forward. However, this MODS has been designed without the comprehensive evidence review and stakeholder consultation conducted for our pilot, and it contains a fraction of the variables included in our prototype MDS. It is likely to be at best an adjunct to a more comprehensive solution and will likely require iteration as implementation challenges, of the sort described here, unfold.

Our prototype MDS focussed on healthcare variables. This reflects, in part, the prominence given to these by all contributors, including care home staff and public representatives, during stakeholder work [16–18]. It also reflects the fact that routine healthcare data are often collected in a way that enables systematic collation and linkage. Healthcare data, by its nature, is aligned to standardised international approaches to coding care data, such as SNOMED and ICD-10. We found some data in DCRs stored as free text – an approach that provides nuanced and personalised records but hampers collation and analysis at meso- and macro-levels using standardised coding approaches. For now, there is a trade-off between data collatable in an MDS and data held in free-text. This may, though, be addressed by advances in machine-based analysis of free-text in the future. Regardless, the integration of datasets across multiple sources represents an additional layer of challenge, as each source dataset may make its own changes over time. Keeping on top of these, and the data manipulation and derivation required for a linked MDS, has ongoing staffing resource requirements.

The incorporation of social care related QoL and wellbeing, in the form the ASCOT-Proxy-Resident and ICECAP-O measures, goes some way towards standardisation of data held in the social care record by providing person-centred data focussed around what matters to residents and relatives, collected in a standardised way. QoL data have been highlighted as essential for understanding quality in the sector[63].

We presented in an appendix how the MDS could facilitate understanding of care home residents' use of ambulance services and hospital emergency departments. Our stakeholder work revealed other areas where an MDS could generate insights, including reasons for hospital admissions to inform local service provision or training needs, and understanding pathways and access to services for residents with, for example, diabetes or mental health needs. Whilst this stakeholder wish list demonstrates the potential of an MDS to better understand resident needs, it also raises the challenge frequently reported in the care home literature, of care home staff and providers feeling that they are at the mercy of external forces beyond their control [4,12,59,64]. The evidence on what enables NHS services working with care homes to achieve improved outcomes consistently points to systems and practices that initiate and sustain quality of working relationships between health and social care staff and their organisations [59,65–69]. The powerful insights deliverable through an MDS come with attendant responsibilities. Ensuring that data are used in a way that foster trust between different stakeholder groups is an implementation imperative.

In conclusion, we have developed and demonstrated an MDS based on data-linkage for English care homes. We have identified issues around data quality, information governance, plurality of data and the need for implementation approaches that facilitate data completion, that are essential to implementation of any MDS in English care homes. We have also demonstrated the value of combining data sources to provide richer data and crucially reduce external requests for information from care homes. It is essential that this work moves forward to ensure that we can take data-informed approaches to care delivery, service design, commissioning and policy for the care home sector.

Ethics approval
The study has received ethical approval from the London Queen's Square Research Ethics Committee (22/LO/0250).

Conflicts
The authors have no conflicts of interest to declare.

Data availability
Anonymised data (digital care records and some associated variables) will be available on request from the corresponding author following a 24 month embargo from the date of publication

Funding
This project is funded by the NIHR Health Service Research and Delivery programme (HS&DR NIHR127234) and supported by the NIHR Applied Research Collaboration East of England. Several authors are supported by the NIHR Applied Research Collaborations in Kent, Surrey and Sussex (AMT); East Midlands (ALG); North East and North Cumbria (BH, AW); Yorkshire and Humber (KS); and East of England (GP, AK, and CG). AG, KS and CG are NIHR Senior Investigators. The views expressed in this publication are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

Acknowledgements

With thanks to Melissa Co, Sarah Opie-Martin and Tom Prendergast for their help with data cleaning, analysis and quality assurance of the analysis code, to James Lockyer for helpful discussions about routinely collected health and social care datasets, and to Sarah Hardy for her help with final write-up and proof-reading.

References

1 Gordon AL, Franklin M, Bradshaw L, *et al.* Health status of UK care home residents: a cohort study. *Age Ageing*. 2014;43:97–103.
2 Berg V. Care home stats: number of settings, population & workforce, available online at: https://www.carehome.co.uk/advice/care-home-stats-number-of-settings-population-workforce (last accessed 28th Jan, 2024).
3 Hanratty B, Wolters AT, Towers A-M, *et al.* Data Collection in Care Homes for Older Adults: A National Survey in England. *Journal of Long-Term Care*. 2023;288–96.
4 Burton JK, Goodman C, Guthrie B, *et al.* Closing the UK care home data gap – methodological challenges and solutions. *Int J Popul Data Sci*. 2020;5.
5 Aalto UL, Pitkälä KH, Andersen-Ranberg K, *et al.* COVID-19 pandemic and mortality in nursing homes across USA and Europe up to October 2021. *Eur Geriatr Med*. 2022;1:1–5.
6 Hanratty B, Burton JK, Goodman C, *et al.* Covid-19 and lack of linked datasets for care homes. *BMJ*. 2020;369:m2463.
7 Saliba D, Buchanan J. Making the Investment Count: Revision of the Minimum Data Set for Nursing Homes, MDS 3.0. *J Am Med Dir Assoc*. 2012;13:602–10.

8 Hirdes JP, Ljunggren G, Morris JN, *et al.* Reliability of the interRAI suite of assessment instruments: a 12-country study of an integrated health information system. *BMC Health Serv Res*. 2008;8:277.
9 Musa MK, Akdur G, Brand S, *et al.* The uptake and use of a minimum data set (MDS) for older people living and dying in care homes: a realist review. *BMC Geriatr*. 2022;22:1–14.
10 Department of Health and Social Care. An Overview of the Minimum Operational Data Standard (MODS), available online at: https://beta.digitisingsocialcare.co.uk/minimum-operational-data-standard-mods (last accessed 30th April 2024).
11 Towers AM, Gordon A, Wolters AT, *et al.* Piloting of a minimum data set for older people living in care homes in England: protocol for a longitudinal, mixed-methods study. *BMJ Open*. 2023;13:e071686.
12 Carroll RE, Smith N, Palmer SE, *et al* (2024). Piloting a Minimum Data Set (MDS) in English care homes: a qualitative study of professional perspectives on implementation and data use. Research Square. 2024. https://doi.org/10.21203/RS.3.RS-4564243/V1
13 Kelly S, Cowan A, Akdur G, *et al.* Outcome measures from international older adult care home intervention research: a scoping review. *Age Ageing*. 2023;52:1–10.
14 Irvine L, Burton J, Ali M, *et al.* Data Resource Profile: the Virtual International Care Homes Trials Archive (VICHTA). *Int J Popul Data Sci*. 2023;8.
15 Peryer G, Kelly S, Blake J, *et al.* Contextual factors influencing complex intervention research processes in care homes: a systematic review and framework synthesis. *Age Ageing*. 2022;51:1–16.
16 Feedback on DACHA study 2021 consultation events, available online at http://dachastudy.com/wp-content/uploads/2021/10/Report-DACHA-consultation-2021.pdf (last accessed 10th October, 2024)
17 DACHA 2022 Quality of Life Consultation Feedback Report, available online at: http://dachastudy.com/wp-content/uploads/2021/10/Report-DACHA-consultation-2021.pdf (last accessed 10th October, 2024)
18 Akdur G, Irvine L, Rand S, *et al.* . National stakeholder consultation on how to measure care home residents' quality of life. Under review. *Health Soc Care Deliv Res*.
19 Micklewright K, Killett A, Akdur G, *et al.* Activity provider-facilitated patient and public involvement with care home residents. *Res Involv Engagem*. 2024;10:1–12.
20 Burton JK, Wolters AT, Towers AM, *et al.* Developing a minimum data set for older adult care homes in the UK: exploring the concept and defining early core principles. *Lancet Healthy Longev*. 2022;3:e186–93.
21 NHS England. Assured Solutions for Digital Social Care, available online at: https://beta.digitisingsocialcare.co.uk/assured-solutions (last accessed 10th October, 2024)
22 Rhodius-Meester HFM, Van Campen JPCM, Fung W, *et al.* Development and validation of the Informant Assessment of Geriatric Delirium Scale (I-AGeD). Recognition of delirium in geriatric patients. *Eur Geriatr Med*. 2013;4:73–7.
23 Hartmaier SL, Sloane PD, Guess HA, *et al.* Validation of the Minimum Data Set Cognitive Performance Scale: Agreement with the Mini-Mental State Examination. *The Journals of Gerontology: Series A*. 1995;50A:M128–33.
24 Wade DT, Collin C. The Barthel ADL Index: a standard measure of physical disability? *Int Disabil Stud*. 1988;10:64–7.
25 Rand S, Caiels J, Collins G, *et al.* Developing a proxy version of the Adult social care outcome toolkit (ASCOT). *Health Qual Life Outcomes*. 2017;15:1–14.
26 Caiels J, Rand S, Crowther T, *et al.* Exploring the views of being a proxy from the perspective of unpaid carers and paid carers: Developing a proxy version of the Adult Social Care Outcomes Toolkit (ASCOT). *BMC Health Serv Res*. 2019;19:1–11.

27 EuroQOL. EQ-52-5L Proxy 2 version, available online at: https://euroqol.org/faq/proxy-version-2/ (last accessed April 19th, 2024).

28 Coast J, Peters TJ, Natarajan L, *et al.* An assessment of the construct validity of the descriptive system for the ICECAP capability measure for older people. *Qual Life Res*. 2008;17:967–76.

29 Dichter MN, Meyer G. Quality of life of people with dementia in nursing homes. *Dementia in Nursing Homes*. 2017;139–57.

30 Siette J, Knaggs GT, Zurynski Y, *et al.* Systematic review of 29 self-report instruments for assessing quality of life in older adults receiving aged care services. *BMJ Open*. 2021;11:e050892.

31 Towers A-M, Smith N, Allan S, *et al.* Care home residents' quality of life and its association with CQC ratings and workforce issues: the MiCareHQ mixed-methods study. *Health Services and Delivery Research*. 2021;9:1–188.

32 Elixhauser A, Steiner C, Harris DR, *et al.* Comorbidity measures for use with administrative data. *Med Care*. 1998;36:8–27.

33 Quan H, Sundararajan V, Halfon P, *et al.* Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care*. 2005;43:1130–9.

34 Soong J, Poots AJ, Scott S, *et al.* Developing and validating a risk prediction model for acute care based on frailty syndromes. *BMJ Open*. 2015;5.

35 Care Quality Commission., Great Britain. Parliament. House of Commons. The state of health care and adult social care in England in 2012/13. 2013;86.

36 NHS England/ Personal Demographics Service, available online at: https://digital.nhs.uk/services/personal-demographics-service (last accessed 30th April, 2024)

37. NHS England. Secondary Uses Service (SUS), available online at: https://digital.nhs.uk/services/secondary-uses-service-sus (last accessed 30th April 2024).

38 Saunders CL, Abel GA, El Turabi A, *et al.* Accuracy of routinely recorded ethnic group information compared with self-reported ethnicity: evidence from the English Cancer Patient Experience survey. *BMJ Open*. 2013;3:e002882.

39 Gallagher AM, Dedman D, Padmanabhan S, *et al.* The accuracy of date of death recording in the Clinical Practice Research Datalink GOLD database in England compared with the Office for National Statistics death registrations. *Pharmacoepidemiol Drug Saf*. 2019;28:563–9.

40 Rand S, Towers AM, Allan S, *et al.* Exploratory factor analysis and Rasch analysis to assess the structural validity of the Adult Social Care Outcomes Toolkit Proxy version (ASCOT-Proxy) completed by care home staff. *Quality of Life Research*. 2024;1–13.

41 Towers A-M, Rand S, Allan S, *et al.* Cross-sectional study assessing the feasibility of measuring residents' Quality of Life in English care homes and assessing the construct validity and internal consistency of measures completed by staff-proxy. *medRxiv*. 2024;2024.05.20.24307612.

42 Morris JN, Fries BE, Mehr DR, *et al.* MDS Cognitive Performance Scale. *J Gerontol*. 1994;49.

43 Urfer Dettwiler P, Zúñiga F, Bachnick S, *et al.* Detecting delirium in nursing home residents using the Informant Assessment of Geriatric Delirium (I-AGeD): a validation pilot study. *Eur Geriatr Med*. 2022;13:917–31.

44 Netten A, Burge P, Malley J, *et al.* Outcomes of social care for adults: developing a preference-weighted measure. *Health Technol Assess*. 2012;16:1–165.

45 Coast J, Flynn TN, Natarajan L, *et al.* Valuing the ICECAP capability index for older people. *Soc Sci Med*. 2008;67:874–82. doi: 10.1016/J.SOCSCIMED.2008.05.015

46 Rowen D, Mukuria C, Bray N, *et al.* UK Valuation of EQ-5D-5L, a Generic Measure of Health-Related Quality of Life: A Study Protocol. *Value in Health*. 2023;26:1625–35.
47 Van Hout B, Janssen MF, Feng YS, *et al.* Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. *Value Health*. 2012;15:708–15.
48 NHS England. Ambulance Data Set, avaiable online at: https://www.england.nhs.uk/urgent-emergency-care/improving-ambulance-services/ambulance-data-set/ (last accessed 30th April 2024)

49 NHS England. Adult Social Cere Client Level Data, available online at: https://digital.nhs.uk/data-and-information/data-collections-and-data-sets/data-sets/adult-social-care-client-level-data (last accessed 30th April 2024)

50 NHS England. Community Services Data Set (CSDS), available online at:https://digital.nhs.uk/data-and-information/data-collections-and-data-sets/data-sets/community-services-data-set (last accessed 30th April 2024).

51 Arden and GEM Commissioning Support Unit. The National Commissioning Data Repository – NCDR, available online at: https://www.ardengemcsu.nhs.uk/services/business-intelligence/ncdr/ (last accessed 30th April 2024.

52 Taylor JA, Crowe S, Pujol FE, *et al.* The road to hell is paved with good intentions: the experience of applying for national data for linkage and suggestions for improvement. *BMJ Open*. 2021;11:47575.
53 Williamson K, Nimegeer A, Lean M. Navigating data governance approvals to use routine health and social care data to evidence the hidden population with severe obesity: a case study from a clinical academic's perspective. *Journal of Research in Nursing*. 2022;27:623–36.
54 Santos F, Conti S, Wolters A. A novel method for identifying care home residents in England: a validation study. *Int J Popul Data Sci*. 2020;5.
55 Burton JK, Ciminata G, Lynch E, *et al.* Understanding Pathways into Care homes using Data (UnPiCD study): a retrospective cohort study using national linked health and social care data. *Age Ageing*. 2022;51:1–11.
56 Barker R, Stocker R, Russell S, *et al.* Distribution of the National Early Warning Score (NEWS) in care home residents. *Age Ageing*. 2019. 49(1):141-145.
57 Witham MD, Frost H, McMurdo M, *et al.* Construction of a linked health and social care database resource--lessons on process, content and culture. *Inform Health Soc Care*. 2015;40:229–39.
58 Onder G, Carpenter I, Finne-Soveri H, *et al.* Assessment of nursing home residents in Europe: the Services and Health for Elderly in Long TERm care (SHELTER) study. *BMC Health Serv Res*. 2012;12:5.
59 Goodman C, Davies SL, Gordon AL, *et al. Optimal NHS service delivery to care homes: a realist evaluation of the features and mechanisms that support effective working for the continuing care of older people in residential settings*. NIHR Journals Library 2017.
60 Kuha S, Vähäkangas P, Kanste O. Implementation leadership of the Resident Assessment Instrument in healthcare services for older people: An interview study of specialists. *J Adv Nurs*. 2024;80:2439–51.
61 Liverpool City Region Civic Data Cooperative, available online at: https://civicdatacooperative.com/ (last accessed 30th November, 2024)
62. Great North Care Record – a way of sharing patient information, available online at: https://www.greatnorthcarerecord.org.uk/ (last accessed 30th Novemer, 2024)

63 Spilsbury K, Charlwood A, Thompson C, *et al.* Relationship between staff and quality of care in care homes: StaRQ mixed methods study. *Health and Social Care Delivery Research*. 2024;12:1–139.
64 Devi R, Chadborn NH, Meyer J, *et al.* How quality improvement collaboratives work to improve healthcare in care homes: a realist evaluation. *Age Ageing*. 2021;50:1371–81.
65 Goodman C, Davies SL, Gordon AL, *et al.* Relationships, expertise, incentives, and governance: Supporting care home residents' access to health care: An interview study from England. *J Am Med Dir Assoc*. 2015;16.
66 Goodman C, Sharpe R, Russell C, et al. Care home readiness: a rapid review and consensus workshop on how organisational context affects care home engagement with health care innovation, available online at: http://researchprofiles.herts.ac.uk/portal/files/11749027/Vanguard_Care_home_readiness (last accessed 10th October, 2024)

67 Bunn, Goodman, Corazzini, *et al.* Setting Priorities to Inform Assessment of Care Homes' Readiness to Participate in Healthcare Innovation: A Systematic Mapping Review and Consensus Process. *Int J Environ Res Public Health*. 2020;17:987.
68 Devi R, Martin GP, Banerjee J, *et al.* Sustaining interventions in care homes initiated by quality improvement projects: a qualitative study. *BMJ Qual Saf*. 2022;0:1–11.
69 Hanratty B, Stocker R, Sinclair D, *et al.* Organising general practice for care homes, a multi method study. In press https://doi.org/10.3310/YNDV6358

APPENDIX 1 – Core Tenets of an MDS for long-term care homes, reproduced from Burton et al[20]

1. The MDS must primarily focus on measuring what matters most to support those living in care homes through systematic data collection and sharing.

2. The MDS must be evidence-based in design and contents, requiring co-production with key stakeholders.

3. The MDS must reduce data burden and duplication of effort for the care home. This will be achieved through piloting, collaboration, and ongoing engagement with homes.

4. The MDS will be most effective when underpinned by digital care planning and care records systems, within the care home, serving the day-to-day needs of residents, staff, families, and friends. This requires digital infrastructure and investment to deliver at scale.

5. The MDS will include information on the care home service, individual-level data on residents, and information on the model of staffing that supports them, but will not include individual-level data identifying the workforce in each home.

6. The MDS should bring together data from within the care home, coupled with data held externally about residents and care services.

7. Data sharing with external users of the MDS must have an agreed purpose. Data sharing pathways must be defined and formalised in data sharing agreements, using secure environments for access where appropriate. Care home residents' privacy rights must be protected.

8. Care homes should be supported to access and use the data they collect and share using electronic dashboards.

9. The MDS requires national infrastructure and integration with existing data systems.

APPENDIX 2 – the Aspirational Minimum Data Set with Proposed Dataset as published in the initial study protocol [11]

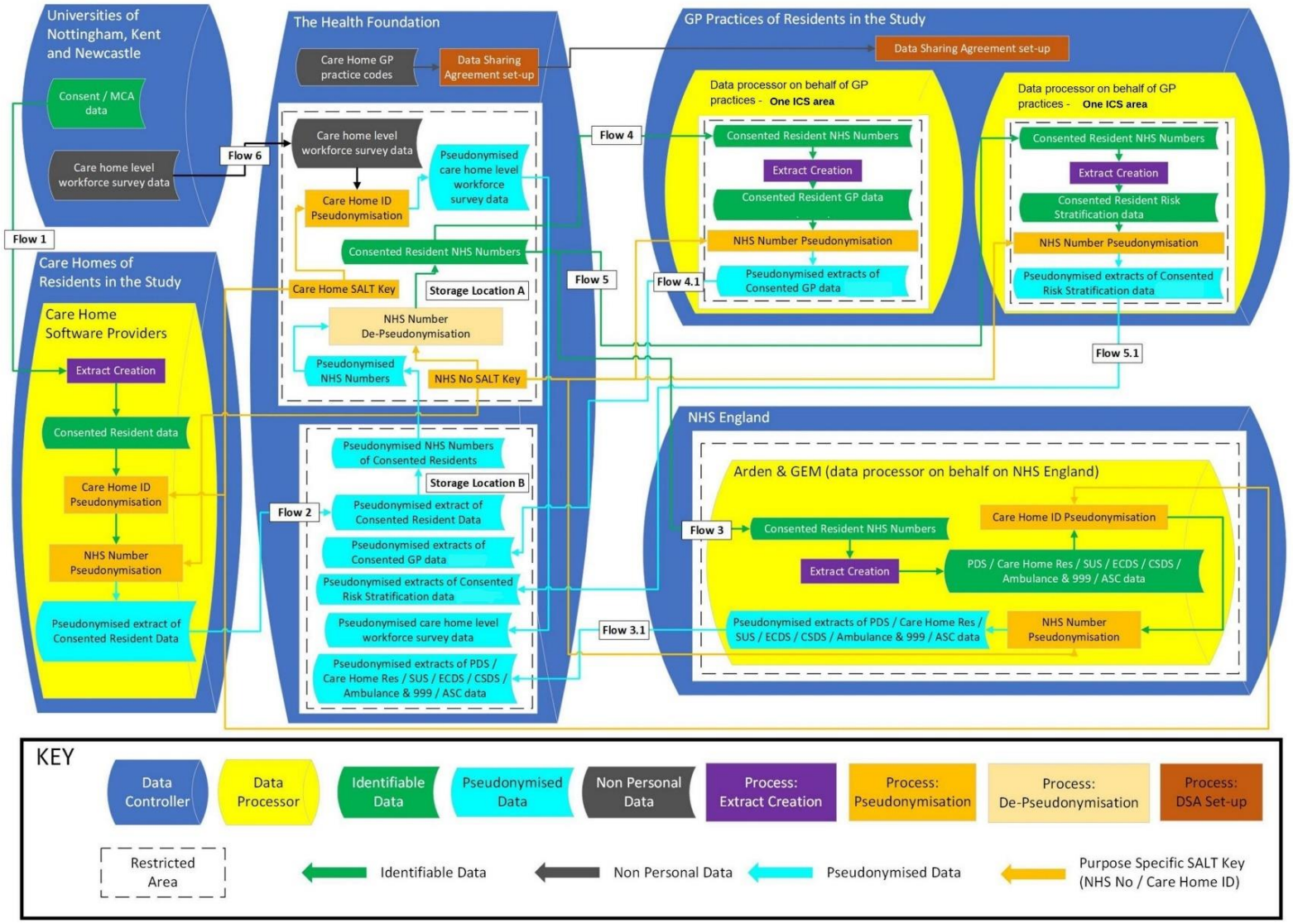
	Sections	Example variables	Digital care records	Health and social care datasets
1	Demographics/characteristics	Date of birth; sex; NHS no; area-based deprivation	No	Personal demographics service
		Religion, languages, marital or partnership status, deprivation of liberty	Yes	No
		Ethnicity; weight; height	Yes	GP data; secondary user services data
2	Palliative care needs	End of life pathway register	No	GP data
3	Care home stay	Date of entry to care home; date of death	Yes	No
4	Resident needs	Skin condition	Yes	No
		Cognitive impairment and impact on perception, understanding and need for support	Yes*	GP data; secondary user services data
		Oral/nutritional status	No	Secondary user services data
		Continence	No	Community datasets (where available)
		Ability to perform activities of daily living; cognitive performance; delirium	Yes*	No
5	Quality of life	Outcomes; mood; dementia quality of life	Yes*	No

	Sections	Example variables	Digital care records	Health and social care datasets
6	Complications/ adverse events	Infections	Yes	GP data; secondary user services data
		Falls (leading to hospital admission or GP visit)	Yes	Secondary user services data; 999 data; ambulance data
		Falls (only captured at care home level); early warning score; unintended weight loss	Yes	No
7	Diagnoses	Medical history	No	Secondary user services data; GP data
		Frailty	No	GP data
		Adverse reactions and allergies	No	GP data
8	Medication and vaccination	Prescribed medication and administered vaccines	No	GP data
9	Healthcare utilisation	Primary care use	No	GP data; NHS 111 data; 999 data
		Community nursing; community allied health professionals	No	Community services data set
		Out-of-hours contacts	No	GP out-of-hours data
		Ambulance call-outs	No	Ambulance data
		Accident and Emergency (A&E) attendance; emergency admissions; secondary care usage (outpatient	No	Secondary user services data

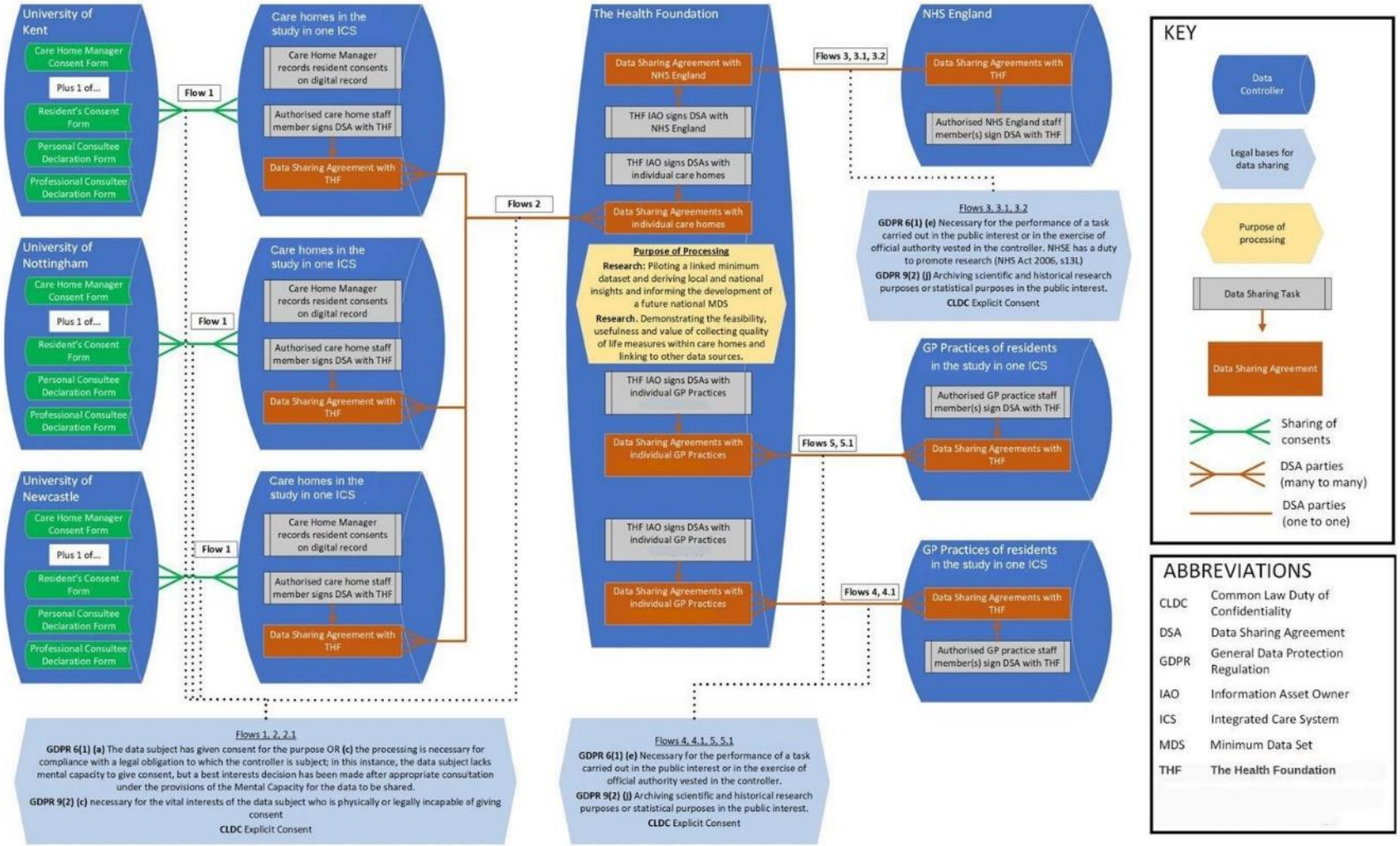
	Sections	Example variables	Digital care records	Health and social care datasets
		appointments and elective admission)		
10	Care home characteristics and workforce characteristics	Type of home; care home characteristics, specialities and client groups; location of care home; area-based deprivation; registered bed capacity; sector of provider; provider ownership type; CQC rating	No	CQC data
		Staffing model; staffing ratios; numbers and types of staff; no of agency staff; no and type of vacancies	No†	Skills for care data

***Added to the software for the purposes of the pilot study.**
†As the Skills for Care workforce survey is voluntary, participating homes were asked to provide some information on workforce as part of a short online survey for the pilot.
CQC: Care Quality Commission; GP: general practitioner; MDS: minimum data set; NHS: National Health Service.

APPENDIX 3 - Data Flow Diagram



APPENDIX 4 - Data Sharing Summary Diagram



APPENDIX 5 - Data specifications
Table a: Data specification for final prototype MDS

Domain	Variable name	Description	Dataset[1]	Raw or derived	Derivation method
Demographics/ characteristics	ethnicity_final	Ethnicity (final)	SUS, CSDS, DCR	Derived	From SUS APC, SUS OP and ECDS, mode was taken across all events for the patient (in patient episode, ED attendance or OP appointment). Where there was no modus, counts were generated by patient and ethnicity category. The following code is applied if there is any record in a given category: If there is any record as white and other records are not stated or missing, recorded as white. If there is any record as black and other records are not stated or missing, recorded as black. If there is any record as asian and other records are not stated or missing, recorded as asian. If there is any record as mixed and other records are not stated or missing, recorded as mixed. If there is any record as other and other records are not stated or missing, recorded as other. If there is any record as not stated then recorded as not stated. If none of the above, recorded as missing. The ethnicity record was taken from

					DCR, then SUS if missing in DCR, then CSDS if missing in SUS.
Demographics/ characteristics	sex_final	Sex (final)	SUS, PDS, CSDS, DCR,	Derived	From SUS APC, SUS OP and ECDS, mode was taken across all events for the patient (in patient episode, ED attendance or OP appointment). Sex was taken from PDS, then SUS if missing in PDS, then CSDS if missing in SUS, then DCR if missing in PDS.
Demographics/ characteristics	dob_final	Date of birth record	SUS, PDS	Derived	From SUS APC, SUS OP and SUS ECDS, mode was taken across all events for the patient (in patient episode, ED attendance or OP appointment). Date of birth was generated as the 1st of the month/year. The date of birth was taken from PDS, then SUS if missing in PDS.
Demographics/ characteristics	dod_final	Date of death record	SUS, PDS, Care home residency	Derived	From hospital records, death date was taken from the latest of: SUS APC (Episode End date where Discharge Method = 4) and ECDS (Departure Date where Discharge status = "75004002", or Treatment Date or Arrival Date is used if later) Date of death was taken from PDS, then SUS if missing in PDS, then Care home residency if missing in SUS.
Demographics/ characteristics	religion_new	Religion	DCR	Derived	Derived variable. Coded differently by the two software providers.
Demographics/ characteristics	maritalstatus_new	Marital status	DCR	Raw	NA
Demographics/ characteristics	firstlanguage	First language spoken	DCR	Raw	NA
Demographics/ characteristics	powerattorney	Power of attorney	DCR	Raw	NA

Demographics/ characteristics	dols	Deprivation of Liberty status	DCR	Raw	NA
Demographics/ characteristics	dnacpr	DNACPR status	DCR	Raw	NA
Demographics/ characteristics	weight_band	Weight	DCR	Derived	Provider 1 numerical data converted to align with provider 2 categorical data
Demographics/ characteristics	height_band	Height	DCR	Derived	Provider 1 numerical data converted to align with provider 2 categorical data
Demographics/ characteristics	IMD_quintile	Indices of Deprivation 2019 quintile	PDS; ONS	Derived	LSOA sourced from PDS and combined with ONS Indices of Deprivation 2019 to identify deprivation quintile
Palliative care needs	DiscussedPreferredDeathLocation_Indicator	Discussed preferred death location indicator	CSDS	Raw	NA
Palliative care needs	DeathLocationPreferred_Type	Preferred death location	CSDS	Raw	NA
Care home stay	Client_Funding_Status_ASC	Client funding status	ASC	Raw	N/A
Care home stay	discharge_ch_12h	Discharge from an in-patient spell to a care home (1 year history)	SUS-APC	Derived	Discharge to care home was derived based on Discharge Destination (codes '54', '65') from continuous in-patient spells.[2] Coded as 1 (yes) if there was any record of this in the year before the index date.
Care home stay	death_hosp_postindex	Death in hospital in the	SUS-APC	Derived	Death in hospital was derived based on Discharge Method (code '4') from continuous in-patient spells.[2]

		period between the index date and end of study.			Coded as 1 (yes) if there was any record of this between the index date and 31 October 2023.
Care home stay	los	Length of stay	DCR	Derived	Derived from date of entry to home. For Provider 2, only have month/year, so set to the 1st of the month.
Resident needs	mdscps	Cognitive impairment	DCR	Derived	The MDS CPS is calculated from five items: comatose, problem with short-term memory, cognitive skills for daily decision making, being understood by others, and eating ADL. Scored per Morris *et al*, 1994.
Resident needs	barthel_bowel	Bowel continence	DCR	Derived	Score from lowest (0) to highest (2) level of functional independence for each item. These are summed, x5, to create a score from lowest (0) to highest (100) independence.
Resident needs	barthel_bladder	Bladder continence	DCR	Derived	Score from lowest (0) to highest (2) level of functional independence for each item. These are summed, x5, to create a score from lowest (0) to highest (100) independence.
Resident needs	ascot_q11p	ASCOT: pain	DCR	Derived	Required some recoding to combine
Resident needs	ascot_mascore_p	ASCOT: anxiety and low mood	DCR	Derived	Required some recoding to combine
Resident needs	iddsi_food	Food texture requirements	DCR	Raw	NA

Resident needs	iddsi_drink	Drink thickness requirements	DCR	Raw	NA
Resident needs	allergy_food	Food allergy	DCR	Raw	NA
Resident needs	allergy_contact	Contact allergy	DCR	Raw	NA
Resident needs	allergy_med	Medication allergy	DCR	Raw	NA
Resident needs	allergy_penicillin	Penicillin allergy	DCR	Raw	NA
Resident needs	waterlow recent	Pressure ulcers (Waterlow score)	DCR	Raw	NA
Resident needs	braden	Pressure ulcers (Braden score)	DCR	Raw	NA
Resident needs	iaged score	Delirium (I-AGED score)	DCR	Derived	Each of the ten items is scored no (0) or yes (1) and summed to create a scale from 0 to 10.
Resident needs	barthel	Functional independence (Barthel score)	DCR	Derived	Score from lowest (0) to highest (2) level of functional independence for each item. These are summed, x5, to create a score from lowest (0) to highest (100) independence.
Resident needs	n_ED_attendances_ngproc	Emergency department attendances with nasogastr	SUS-ECDS	Derived	Coded as 1 (yes) if there was a SNOMED code for nasogastric procedure from Der_EC_Treatment_All ("87750000", "6125005", "112861000") in the year before index date, otherwise coded as 0 (no)

		ic procedure (1 year history)			
Resident needs	n_ngproc_op	Outpatient appointments with nasogastric procedure (1 year history)	SUS-OP	Derived	Coded as 1 (yes) if there was a procedure code for nasogastric procedure from Der_Procedure_All (OPCS codes) recorded for an outpatient appointment in the year before index date, otherwise coded as 0 (no)
Resident needs	mdscps	Cognitive impairment	DCR	Derived	MDS CPS is calculated from five items: comatose, problem with short-term memory, cognitive skills for daily decision making, being understood by others, and eating ADL. Scored per Morris *et al*, 1994.
Quality of life	ascs_qol_score	Quality of life overall	DCR	Derived	Required some recoding to combine
Quality of life	ascot_scrqol	Ascot Proxy-Resident	DCR	Derived	Required some recoding to combine. Applied preference weights for ASCOT SCT4 to generate index score from -.17 to 1.0.
Quality of life	icecap_qol	ICECAP-O	DCR	Derived	Required some recoding to combine. Score (0 to 1) calculated using UK index values [.
Quality of life	UK_crosswalk	EQ-5D-5L Proxy	DCR	Derived	Required some recoding to combine. Score calculated using the mapping function to convert to EQ-5D-3L and applied UK index values. The UK value set for the EQ-5D-5L is still being developed [45,46]

Complications/ adverse events	must_score_recent	MUST (malnutrition universal scoring tool) score	DCR	Raw	NA
Complications/ adverse events	adm_lrti_12h	Frequency of in-patient admissions with lower respiratory tract infection recorded (1 year history)	SUS-APC	Derived	Admissions with lrti recorded were derived based on ICD-10 codes in the primary diagnosis field . Count of admissions with lrti recorded from continuous in-patient spells in the year before the index date.[2]
Complications/ adverse events	adm_urti_12h	Frequency of in-patient admissions with upper respiratory tract infection recorded (1 year history)	SUS-APC	Derived	Admissions with urti recorded were derived based on ICD-10 codes in the primary diagnosis field. Count of admissions with urti recorded from continuous in-patient spells in the year before the index date.[2]
Diagnoses	cogimp_sus	Cognitive impairment (delirium,	SUS-APC	Derived	Diagnosis flags were derived based on ICD-10 codes in the diagnosis fields for each episode. Coded as 1 (yes) if there was a record of any of delirium, dementia or senility

		dementia, senility)			(using Soong et al 2015 code list) in the patient's hospital admissions in 3 years before the index date, otherwise coded as 0 (no)
Diagnoses	dementia_final	Dementia (final)	SUS-APC and DCR	Derived	Diagnosis flags were derived from SUS-APC based on ICD-10 codes in the diagnosis fields for each episode. Dementia status is selected first from SUS (a resident has dementia if identified through either the code list used for Charlson index or Soong et al Frailty), coded as 1 if there was any record of the condition in the patient's hospital admissions in the previous 3 years, and if missing, from DCR. Note if a person does not have dementia recorded in SUS but does in care home record then they will be recorded as having dementia in dementia_final.
Diagnoses	nr_elix_h36	Elixhauser no conditions based on 3 years history	SUS-APC	Derived	Diagnosis flags were derived based on ICD-10 codes in the diagnosis fields for each episode. Count of conditions in the Elixhauser list of comorbidities[3] in the patient's hospital admissions in 3 years before the index date.
Diagnoses	nr_elix_2_h36	Elixhauser conditions >=2 based on 3 years history	SUS-APC	Derived	Diagnosis flags were derived based on ICD-10 codes in the diagnosis fields for each episode. Coded as 1 (yes) if there were >= conditions in the Elixhauser list of comorbidities[3] in the patient's hospital admissions in 3 years before the index date.

Diagnoses	e_alco_abuse_h36	Alcohol abuse[2]	SUS-APC	Derived	Diagnosis flags were derived based on ICD-10 codes in the diagnosis fields for each episode. Coded as 1 (yes) if there was any record of the condition in the patient's hospital admissions in 3 years before the index date.
Diagnoses	e_anaemia_bloodloss_h36	Blood loss anaemia[2]	SUS-APC	Derived	Diagnosis flags were derived based on ICD-10 codes in the diagnosis fields for each episode. Coded as 1 (yes) if there was any record of the condition in the patient's hospital admissions in 3 years before the index date.
Diagnoses	e_anaemia_deficiency_h36	Deficiency anaemia[2]	SUS-APC	Derived	Diagnosis flags were derived based on ICD-10 codes in the diagnosis fields for each episode. Coded as 1 (yes) if there was any record of the condition in the patient's hospital admissions in 3 years before the index date.
Diagnoses	e_arrhythmias_h36	Cardiac arrhythmias[2]	SUS-APC	Derived	Diagnosis flags were derived based on ICD-10 codes in the diagnosis fields for each episode. Coded as 1 (yes) if there was any record of the condition in the patient's hospital admissions in 3 years before the index date.
Diagnoses	e_coagulopathy_h36	Coagulopathy[2]	SUS-APC	Derived	Diagnosis flags were derived based on ICD-10 codes in the diagnosis fields for each episode. Coded as 1 (yes) if there was any record of the condition in the patient's hospital admissions in 3 years before the index date.

Diagnoses	e_depression_h36	Depression[2]	SUS-APC	Derived	Diagnosis flags were derived based on ICD-10 codes in the diagnosis fields for each episode. Coded as 1 (yes) if there was any record of the condition in the patient's hospital admissions in 3 years before the index date.
Diagnoses	e_diab_comp_h36	Diabetes, complicated[2]	SUS-APC	Derived	Diagnosis flags were derived based on ICD-10 codes in the diagnosis fields for each episode. Coded as 1 (yes) if there was any record of the condition in the patient's hospital admissions in 3 years before the index date.
Diagnoses	e_diab_uncomp_h36	Diabetes, uncomplicated[2]	SUS-APC	Derived	Diagnosis flags were derived based on ICD-10 codes in the diagnosis fields for each episode. Coded as 1 (yes) if there was any record of the condition in the patient's hospital admissions in 3 years before the index date.
Diagnoses	e_drug_abuse_h36	Drug abuse[2]	SUS-APC	Derived	Diagnosis flags were derived based on ICD-10 codes in the diagnosis fields for each episode. Coded as 1 (yes) if there was any record of the condition in the patient's hospital admissions in 3 years before the index date.
Diagnoses	e_fluid_h36	Fluid and electrolyte disorders[2]	SUS-APC	Derived	Diagnosis flags were derived based on ICD-10 codes in the diagnosis fields for each episode. Coded as 1 (yes) if there was any record of the condition in the patient's hospital admissions in 3 years before the index date.

Diagnoses	e_ht_comp_h36	Hypertension, complicated[2]	SUS-APC	Derived	Diagnosis flags were derived based on ICD-10 codes in the diagnosis fields for each episode. Coded as 1 (yes) if there was any record of the condition in the patient's hospital admissions in 3 years before the index date.
Diagnoses	e_ht_uncomp_h36	Hypertension, uncomplicated[2]	SUS-APC	Derived	Diagnosis flags were derived based on ICD-10 codes in the diagnosis fields for each episode. Coded as 1 (yes) if there was any record of the condition in the patient's hospital admissions in 3 years before the index date.
Diagnoses	e_hypothyroid_h36	Hypothyroidism[2]	SUS-APC	Derived	Diagnosis flags were derived based on ICD-10 codes in the diagnosis fields for each episode. Coded as 1 (yes) if there was any record of the condition in the patient's hospital admissions in 3 years before the index date.
Diagnoses	e_liver_h36	Liver disease[2]	SUS-APC	Derived	Diagnosis flags were derived based on ICD-10 codes in the diagnosis fields for each episode. Coded as 1 (yes) if there was any record of the condition in the patient's hospital admissions in 3 years before the index date.
Diagnoses	e_lymphoma_h36	Lymphoma[2]	SUS-APC	Derived	Diagnosis flags were derived based on ICD-10 codes in the diagnosis fields for each episode. Coded as 1 (yes) if there was any record of the condition in the patient's hospital admissions in 3 years before the index date.

Diagnoses	e_obesity_h36	Obesity[2]	SUS-APC	Derived	Diagnosis flags were derived based on ICD-10 codes in the diagnosis fields for each episode. Coded as 1 (yes) if there was any record of the condition in the patient's hospital admissions in 3 years before the index date.
Diagnoses	e_other_neuro_h36	Other neurological disorders[2]	SUS-APC	Derived	Diagnosis flags were derived based on ICD-10 codes in the diagnosis fields for each episode. Coded as 1 (yes) if there was any record of the condition in the patient's hospital admissions in 3 years before the index date.
Diagnoses	e_peptic_nobld_h36	Peptic ulcer disease excl bleeding[2]	SUS-APC	Derived	Diagnosis flags were derived based on ICD-10 codes in the diagnosis fields for each episode. Coded as 1 (yes) if there was any record of the condition in the patient's hospital admissions in 3 years before the index date.
Diagnoses	e_psychoses_h36	Psychoses[2]	SUS-APC	Derived	Diagnosis flags were derived based on ICD-10 codes in the diagnosis fields for each episode. Coded as 1 (yes) if there was any record of the condition in the patient's hospital admissions in 3 years before the index date.
Diagnoses	e_pulmcirc_h36	Pulmonary circulation disorders[2]	SUS-APC	Derived	Diagnosis flags were derived based on ICD-10 codes in the diagnosis fields for each episode. Coded as 1 (yes) if there was any record of the condition in the patient's hospital admissions in 3 years before the index date.

Diagnoses	e_pvd_h36	Peripheral vascular disease[2]	SUS-APC	Derived	Diagnosis flags were derived based on ICD-10 codes in the diagnosis fields for each episode. Coded as 1 (yes) if there was any record of the condition in the patient's hospital admissions in 3 years before the index date.
Diagnoses	e_renalfail_h36	Rheumatoid arthritis / collagen vascular diseases[2]	SUS-APC	Derived	Diagnosis flags were derived based on ICD-10 codes in the diagnosis fields for each episode. Coded as 1 (yes) if there was any record of the condition in the patient's hospital admissions in 3 years before the index date.
Diagnoses	e_rheum_arth_h36	Renal failure[2]	SUS-APC	Derived	Diagnosis flags were derived based on ICD-10 codes in the diagnosis fields for each episode. Coded as 1 (yes) if there was any record of the condition in the patient's hospital admissions in 3 years before the index date.
Diagnoses	e_stumour_nomets_h36	Solid tumour without metastasis[2]	SUS-APC	Derived	Diagnosis flags were derived based on ICD-10 codes in the diagnosis fields for each episode. Coded as 1 (yes) if there was any record of the condition in the patient's hospital admissions in 3 years before the index date.
Diagnoses	e_valvular_h36	Valvular disease[2]	SUS-APC	Derived	Diagnosis flags were derived based on ICD-10 codes in the diagnosis fields for each episode. Coded as 1 (yes) if there was any record of the condition in the patient's hospital admissions in 3 years before the index date.

Diagnoses	e_weight_loss_h36	Weight loss[2]	SUS-APC	Derived	Diagnosis flags were derived based on ICD-10 codes in the diagnosis fields for each episode. Coded as 1 (yes) if there was any record of the condition in the patient's hospital admissions in 3 years before the index date.
Diagnoses	ec_chf_h36	Congestive heart failure[2]	SUS-APC	Derived	Diagnosis flags were derived based on ICD-10 codes in the diagnosis fields for each episode. Coded as 1 (yes) if there was any record of the condition in the patient's hospital admissions in 3 years before the index date.
Diagnoses	ec_cpd_h36	Chronic pulmonary disease[2]	SUS-APC	Derived	Diagnosis flags were derived based on ICD-10 codes in the diagnosis fields for each episode. Coded as 1 (yes) if there was any record of the condition in the patient's hospital admissions in 3 years before the index date.
Diagnoses	ec_plegia_h36	Hemiplegia / paraplegia[2]	SUS-APC	Derived	Diagnosis flags were derived based on ICD-10 codes in the diagnosis fields for each episode. Coded as 1 (yes) if there was any record of the condition in the patient's hospital admissions in 3 years before the index date.
Diagnoses	ec_stumour_mets_h36	Metastatic solid tumour / metastatic cancer[2]	SUS-APC	Derived	Diagnosis flags were derived based on ICD-10 codes in the diagnosis fields for each episode. Coded as 1 (yes) if there was any record of the condition in the patient's hospital admissions in 3 years before the index date.

Diagnoses	N frailty conditions	Number of frailty conditions	SUS-APC	Derived	Diagnosis flags were derived based on ICD-10 codes in the diagnosis fields for each episode. Frailty syndromes were cognitive impairment, anxiety or depression, functional dependence, falls or fractures, incontinence, mobility problems and pressure ulcers (Soong et al 2015). Count of frailty syndromes[4] recorded in the patient's hospital admission in 3 years before the index date.
Diagnoses	f_delirium_h36	Delirium[3]	SUS-APC	Derived	Diagnosis flags were derived based on ICD-10 codes in the diagnosis fields for each episode. Coded as 1 (yes) if there was any record of the condition in the patient's hospital admissions in 3 years before the index date.
Diagnoses	f_dementia_h36	Dementia[3]	SUS-APC	Derived	Diagnosis flags were derived based on ICD-10 codes in the diagnosis fields for each episode. Coded as 1 (yes) if there was any record of the condition in the patient's hospital admissions in 3 years before the index date.
Diagnoses	f_senility_h36	Senility[3]	SUS-APC	Derived	Diagnosis flags were derived based on ICD-10 codes in the diagnosis fields for each episode. Coded as 1 (yes) if there was any record of the condition in the patient's hospital admissions in 3 years before the index date.
Diagnoses	f_anxdep_h36	Anxiety/Depression[3]	SUS-APC	Derived	Diagnosis flags were derived based on ICD-10 codes in the diagnosis fields for each episode.

					Coded as 1 (yes) if there was any record of the condition in the patient's hospital admissions in 3 years before the index date.
Diagnoses	f_dependence_h36	Functional dependence[3]	SUS-APC	Derived	Diagnosis flags were derived based on ICD-10 codes in the diagnosis fields for each episode. Coded as 1 (yes) if there was any record of the condition in the patient's hospital admissions in 3 years before the index date.
Diagnoses	f_fallsfract_h36	Falls/Fractures[3]	SUS-APC	Derived	Diagnosis flags were derived based on ICD-10 codes in the diagnosis fields for each episode. Coded as 1 (yes) if there was any record of the condition in the patient's hospital admissions in 3 years before the index date.
Diagnoses	f_incont_h36	Incontinence[3]	SUS-APC	Derived	Diagnosis flags were derived based on ICD-10 codes in the diagnosis fields for each episode. Coded as 1 (yes) if there was any record of the condition in the patient's hospital admissions in 3 years before the index date.
Diagnoses	f_mobprob_h36	Mobility problems[3]	SUS-APC	Derived	Diagnosis flags were derived based on ICD-10 codes in the diagnosis fields for each episode. Coded as 1 (yes) if there was any record of the condition in the patient's hospital admissions in 3 years before the index date.

Diagnoses	f_pulcers_h36	Pressure ulcers[3]	SUS-APC	Derived	Diagnosis flags were derived based on ICD-10 codes in the diagnosis fields for each episode. Coded as 1 (yes) if there was any record of the condition in the patient's hospital admissions in 3 years before the index date.
Healthcare utilisation	adm_el_12h	Frequency of elective admissions (1 year history)	SUS-APC	Derived	Elective admissions were derived based on Admission Method indicating elective admission (codes '11','12','13'). Count of elective admissions from continuous in-patient spells in the year before the index date.[2]
Healthcare utilisation	adm_em_12h	Frequency of emergency admissions (1 year history)	SUS-APC	Derived	Emergency admissions were derived based on Admission Method indicating Emergency Admission (codes '21','22','23','24','25','28','2A','2B','2C','2D'). Count of emergency admissions from continuous in-patient spells in the year before the index date.[2]
Healthcare utilisation	adm_avoid_12h	Frequency of potentially avoidable emergency admissions (1 year history)[5]	SUS-APC	Derived	Potentially avoidable emergency admissions were defined as emergency admissions (Admission Method codes '21','22','23','24','25','28','2A','2B','2C','2D') with one of a list of conditions as the primary diagnoses for the first episode of the hospital spell. Count of potentially avoidable emergency admissions from continuous in-patient spells in the year before the index date.[2,5]
Healthcare utilisation	n_ED_attendances	Frequency of emergenc	SUS-ECDS	Derived	Count of non-duplicate ED attendances in 1 year before the index date

		y department attendances (1 year history)			
Healthcare utilisation	n_ED_attendances_via_ambulance	Frequency of emergency department attendances via ambulance (1 year history)	SUS-ECDS	Derived	Count of non-duplicate ED attendances that were via ambulance (as indicated by presence of codes 104803100000010, 1048021000000102, 1048041000000109, 1048051000000107 in field EC_Arrival_Mode_SNOMED_CT) in 1 year before the index date
Healthcare utilisation	n_out_of_hours_ED_attendances	Frequency of emergency department attendances after 6pm/ before 8am (1 year history)	SUS-ECDS	Derived	Count of non-duplicate ED attendances that occurred after 1800 or before 0800 hours in 1 year before the index date
Healthcare utilisation	n_long_ED_attendances	Frequency of emergency department	SUS-ECDS	Derived	Count of non-duplicate ED attendances that lasted more than 12 hours in duration in 1 year before the index date

		attendances lasting more than 12 hours (1 year history)			
Healthcare utilisation	n_ED_attendances_for_falls	Frequency of emergency department attendances due to a fall (1 year history)	SUS-ECDS	Derived	Count of non-duplicate ED attendances that were for falls (as indicated by presence of codes 161898004, 430576002, 54670004, 75941004, 240871000000104, 429482004 in field EC_Chief_Complaint_SNOMED_CT) in 1 year before the index date
Healthcare utilisation	n_OP_appts	Frequency of outpatient appointments (1 year history)	SUS-OP	Derived	Count of non-duplicate outpatient appointments in 1 year before index date
Healthcare utilisation	n_cs_appts	Frequency of community services appointments (1 year history)	CSDS	Derived	Count of community service appointments in 1 year before the index date
Healthcare utilisation	n_missed_OP_appts	Frequency of	SUS-OP	Derived	Count of non-duplicate appointments that were missed (as indicated by

		missed outpatient appointments (1 year history)			presence of codes 0, 1, 3 in field Attendance_Status) in 1 year before the index date
Healthcare utilisation	n_cs_appts_continence	Frequency of continence appointments (1 year history)	CSDS	Derived	Count of continence community service appointments (as indicated by Referral_TeamType =07) in 1 year before the index date
Healthcare utilisation	n_cs_appts_dn	Frequency of district nursing appointments (1 year history)	CSDS	Derived	Count of district nursing community service appointments (as indicated by Referral_TeamType =12) in 1 year before the index date
Healthcare utilisation	n_cs_appts_podiatry	Frequency of podiatry appointments (1 year history)	CSDS	Derived	Count of podiatry community service appointments (as indicated by Referral_TeamType =27) in 1 year before the index date

Healthcare utilisation	n_cs_appts_rehab	Frequency of community rehabilitation appointments (1 year history)	CSDS	Derived	Count of community rehabilitation community service appointments (as indicated by Referral_TeamType =29) in 1 year before the index date
Healthcare utilisation	n_cs_appts_SALT	Frequency of speech and language therapy appointments (1 year history)	CSDS	Derived	Count of SALT community service appointments (as indicated by Referral_TeamType =33) in 1 year before the index date
Healthcare utilisation	n_cs_f2f_appts	Frequency of face to face community services appointments (1 year history)	CSDS	Derived	Count of face to face community service appointments (as indicated by presence of code 1 in field Consultation_MethodUsed) in 1 year before the index date
Healthcare utilisation	n_cs_missed_appts	Frequency of missed community services appointments (1	CSDS	Derived	Count of missed community service appointments (as indicated by presence of codes 2, 3, 7 in field AttendanceStatus) in 1 year before the index date

		year history)			
Healthcare utilisation	n_cs_services	Number of community services seen (1 year history)	CSDS	Derived	Count of community service teams seen in 1 year before the index date
Healthcare utilisation	n_ambulance_call_outs_5m	Frequency of ambulance call outs (1 June-31 October 2023)	Ambulance	Derived	Count of ambulance call outs between 1 June 2023 and 31 October 2023
Healthcare utilisation	n_ambulance_attendances_5m	Frequency of ambulance attendances (1 June-31 October 2023)	Ambulance	Derived	Count of ambulance attendances between 1 June 2023 and 31 October 2023
Healthcare utilisation	n_ambulance_attendances_OOH_5m	Frequency of ambulance attendances after 6pm/ before 8am (1	Ambulance	Derived	Count of ambulance attendances that occurred after 1800 or before 0800 hours between 1 June 2023 and 31 October 2023

		June-31 October 2023)			
Healthcare utilisation	n_ambulance_conveyances_5m	Frequency of ambulance conveyances (1 June-31 October 2023)	Ambulance	Derived	Count of ambulance conveyances between 1 June 2023 and 31 October 2023
Care home characteristics and workforce characteristics	Der_CQC_Service_Type_CHR	Service type	Care home residency	Raw	NA
Care home characteristics and workforce characteristics	registered_bed_capacity	Registered bed capacity	CQC	Derived	Numeric converted into categorical variable
Care home characteristics and workforce characteristics	Rating_overall	CQC rating (overall)	CQC	Raw	NA
Care home characteristics and workforce characteristics	service_reg_yrs	Years since registration with CQC	CQC	Derived	Numeric converted into categorical variable
Care home characteristics and workforce characteristics	Q3_1_4	Number of beds	Care home survey	Raw	NA
Care home characteristics and workforce characteristics	Q3_2_4	Number of beds currently occupied	Care home survey	Raw	NA
Care home characteristics and workforce characteristics	Q3_3_4	Number of resident fully self-funding	Care home survey	Raw	NA

Care home characteristics and workforce characteristics	Q3_4_4	Number of staff	Care home survey	Raw	NA
Care home characteristics and workforce characteristics	Q3_5_4	Number of full-time staff	Care home survey	Raw	NA
Care home characteristics and workforce characteristics	Q3_6_4	Number of part-time staff	Care home survey	Raw	NA
Care home characteristics and workforce characteristics	Q3_7_4	Number of staff on permanent contracts	Care home survey	Raw	NA
Care home characteristics and workforce characteristics	Q3_8_4	Number of staff vacancies	Care home survey	Raw	NA
Care home characteristics and workforce characteristics	Q3_9_4	Number of agency staff	Care home survey	Raw	NA
Care home characteristics and workforce characteristics	Q3_10_4	Number of care workers	Care home survey	Raw	NA
Care home characteristics and workforce characteristics	Q3_11_4	Number of senior care workers	Care home survey	Raw	NA
Care home characteristics and workforce characteristics	Q3_13_4	Number of registered nurses	Care home survey	Raw	NA
Care home characteristics and workforce characteristics	Q3_12_4	Number of nursing associates	Care home survey	Raw	NA

Care home characteristics and workforce characteristics	Q3_14_4	Number of nursing assistants	Care home survey	Raw	NA
Care home characteristics and workforce characteristics	Q3_15_4	Number of allied health professionals	Care home survey	Raw	NA
Care home characteristics and workforce characteristics	Q3_16_4	Number of activities coordinators	Care home survey	Raw	NA
Care home characteristics and workforce characteristics	Q3_17_4	Number of staff in roles above on permanent contracts	Care home survey	Raw	NA
Care home characteristics and workforce characteristics	Q3_18_4	Number of staff vacancies in roles above	Care home survey	Raw	NA
Care home characteristics and workforce characteristics	Q3_19_4	Number of agency staff in roles above	Care home survey	Raw	NA

[1] Dataset abbreviations: ASC-CLD = Adult Social Care Client Level Dataset; CQC = Care Quality Commision; CSDS = Community Services Data Set; DCR = Digital Care Record; ONS = Office for National Statistics; PDS = Personal Demographics Service; SUS = Secondary Uses Service; SUS-APC = SUS Admitted Patient Care; SUS-ECDS = SUS Emergency Care Dataset; SUS-OP = SUS Outpatient.

[2] Continuous in-patient spells (CIPS) (sequence of spells from patient's first admission to hospital to patient final discharge home, including transfers to other hospitals as part of patient's care) were derived by grouping episodes with the same patient, provider and admission date (or previous episode end date the same as the following episode start date).
[3] From the list of Elixhauser conditions (M Elixhauser A, Steiner C, Harris DR, et al. Comorbidity measures for use with administrative data. Med Care. 1998;36:8–27. doi: 10.1097/00005650-199801000-00004; Quan H, Sundararajan V, Halfon P, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. Med Care. 2005;43:1130–9. doi: 10.1097/01.MLR.0000182534.19832.83.)
[4] From a validated list of frailty syndromes (Soong J, Poots AJ, Scott S, et al. Developing and validating a risk prediction model for acute care based on frailty syndromes. BMJ Open. 2015;5. doi: 10.1136/BMJOPEN-2015-008457)
[5] Potentially avoidable emergency admissions (Care Quality Commission, Great Britain. Parliament. House of Commons. The state of health care and adult social care in England in 2012/13. 2013;86.).

Table b: data dictionary for linked routinely collected source datasets

Dataset[1]	**Domain**	**Variable name**	**Description**
ALL	Linkage	pseudonhsno	Pseudonymised NHS number
SUS-APCE	Healthcare utilization	Generated_Record_ID	Generated Record ID
SUS-APCE	Healthcare utilization	Der_Postcode_LSOA_Code	LSOA code
SUS-APCE	Healthcare utilization	GP_Practice_Code	GP practice code
SUS-APCE	Healthcare utilization	Der_Postcode_CCG_Code	CCG code
SUS-APCE	Healthcare utilization	Der_Financial_Year	Financial year
SUS-APCE	Healthcare utilization	Der_Activity_Month	Activity month
SUS-APCE	Healthcare utilization	Admission_Date	Admission date
SUS-APCE	Healthcare utilization	Admission_Time	Admission time
SUS-APCE	Healthcare utilization	Admission_Method	Admission method

SUS-APCE	Healthcare utilization	Source_of_Admission	Source of admission
SUS-APCE	Healthcare utilization	Administrative_Category	Administrative category
SUS-APCE	Healthcare utilization	Discharge_Date	Discharge date
SUS-APCE	Healthcare utilization	Discharge_Time	Discharge time
SUS-APCE	Healthcare utilization	Discharge_Method	Discharge method
SUS-APCE	Healthcare utilization	Discharge_Destination	Discharge destination
SUS-APCE	Healthcare utilization	Der_Provider_Code	Provider code
SUS-APCE	Diagnoses	Der_Diagnosis_Count	Count of diagnosis codes
SUS-APCE	Diagnoses	Der_Primary_Diagnosis_Code	Primary diagnosis code
SUS-APCE	Diagnoses	Der_Secondary_Diagnosis_Code_1	Secondary diagnosis code 1
SUS-APCE	Diagnoses	Der_Secondary_Diagnosis_Code_2	Secondary Diagnosis Code 2
SUS-APCE	Diagnoses	Der_Secondary_Diagnosis_Code_3	Secondary Diagnosis Code 3
SUS-APCE	Diagnoses	Der_Secondary_Diagnosis_Code_4	Secondary Diagnosis Code 4
SUS-APCE	Diagnoses	Der_Secondary_Diagnosis_Code_5	Secondary Diagnosis Code 5
SUS-APCE	Diagnoses	Der_Secondary_Diagnosis_Code_6	Secondary Diagnosis Code 6
SUS-APCE	Diagnoses	Der_Secondary_Diagnosis_Code_7	Secondary Diagnosis Code 7
SUS-APCE	Diagnoses	Der_Secondary_Diagnosis_Code_8	Secondary Diagnosis Code 8
SUS-APCE	Diagnoses	Der_Secondary_Diagnosis_Code_9	Secondary Diagnosis Code 9
SUS-APCE	Diagnoses	Der_Secondary_Diagnosis_Code_10	Secondary Diagnosis Code 10
SUS-APCE	Diagnoses	Der_Secondary_Diagnosis_Code_11	Secondary Diagnosis Code 11
SUS-APCE	Diagnoses	Der_Secondary_Diagnosis_Code_12	Secondary Diagnosis Code 12

SUS-APCE	Diagnoses	Der_Secondary_Diagnosis_Code_13	Secondary Diagnosis Code 13
SUS-APCE	Diagnoses	Der_Secondary_Diagnosis_Code_14	Secondary Diagnosis Code 14
SUS-APCE	Diagnoses	Der_Secondary_Diagnosis_Code_15	Secondary Diagnosis Code 15
SUS-APCE	Diagnoses	Der_Secondary_Diagnosis_Code_16	Secondary Diagnosis Code 16
SUS-APCE	Diagnoses	Der_Secondary_Diagnosis_Code_17	Secondary Diagnosis Code 17
SUS-APCE	Diagnoses	Der_Secondary_Diagnosis_Code_18	Secondary Diagnosis Code 18
SUS-APCE	Diagnoses	Der_Secondary_Diagnosis_Code_19	Secondary Diagnosis Code 19
SUS-APCE	Diagnoses	Der_Secondary_Diagnosis_Code_20	Secondary Diagnosis Code 20
SUS-APCE	Diagnoses	Der_Secondary_Diagnosis_Code_21	Secondary Diagnosis Code 21
SUS-APCE	Diagnoses	Der_Secondary_Diagnosis_Code_22	Secondary Diagnosis Code 22
SUS-APCE	Diagnoses	Der_Secondary_Diagnosis_Code_23	Secondary Diagnosis Code 23
SUS-APCE	Diagnoses	Der_Diagnosis_All	All diagnosis codes
SUS-APCE	Healthcare utilization	Treatment_Function_Code	Treatment function code
SUS-APCE	Healthcare utilization	Main_Speciality_Code	Main speciality code
SUS-APCE	Healthcare utilization	Patient_Classification	Patient classification
SUS-APCE	Healthcare utilization	Episode_Start_Date	Episode Start Date
SUS-APCE	Healthcare utilization	Episode_End_Date	Episode End Date
SUS-APCE	Healthcare utilization	Der_Episode_Number	Episode Number
SUS-APCE	Healthcare utilization	Der_Spell_ID	Spell ID
SUS-APCE	Demographics/characteristics	Month_of_Birth_SUS	Month of birth

SUS-APCE	Demographics/characteristics	Year_of_Birth_SUS	Year of birth
SUS-APCE	Demographics/characteristics	Sex	Sex
SUS-APCE	Demographics/characteristics	Ethnic_Group	Ethnicity
SUS-ECDS	Healthcare utilization	Arrival_Date	Date of arrival
SUS-ECDS	Healthcare utilization	Arrival_Time	Time of arrival
SUS-ECDS	Healthcare utilization	EC_Treatment_Date_01	Date of treatment
SUS-ECDS	Healthcare utilization	EC_Seen_For_Treatment_Time_Since_Arrival	Waiting time from arrival to being seen for treatment
SUS-ECDS	Healthcare utilization	EC_Departure_Date	Date of departure
SUS-ECDS	Healthcare utilization	EC_Departure_Time	Time of departure
SUS-ECDS	Healthcare utilization	Der_EC_Duration	Duration of the A&E attendance
SUS-ECDS	Healthcare utilization	Der_Activity_Month	Activity month
SUS-ECDS	Healthcare utilization	Generated_Record_ID	Generated Record ID
SUS-ECDS	Healthcare utilization	Provider_Code	Provider code
SUS-ECDS	Healthcare utilization	EC_Department_Type	Department type
SUS-ECDS	Healthcare utilization	Der_Postcode_LSOA_2011_Code	LSOA code
SUS-ECDS	Healthcare utilization	Der_Postcode_CCG_Code	CCG code
SUS-ECDS	Healthcare utilization	GP_Practice_Code	GP practice code
SUS-ECDS	Healthcare utilization	Accommodation_Status_SNOMED_CT	Accommodation status (SNOMED CT)
SUS-ECDS	Healthcare utilization	EC_AttendanceCategory	Attendance category (SNOMED CT)
SUS-ECDS	Healthcare utilization	EC_Arrival_Mode_SNOMED_CT	Arrival mode (SNOMED CT)
SUS-ECDS	Healthcare utilization	EC_Attendance_Number	Attendance number (SNOMED CT)
SUS-ECDS	Healthcare utilization	Discharge_Destination_SNOMED_CT	Discharge destination (SNOMED CT)
SUS-ECDS	Healthcare utilization	Discharge_Follow_Up_SNOMED_CT	Discharge follow up (SNOMED CT)

SUS-ECDS	Healthcare utilization	EC_Acuity_SNOMED_CT	Acuity (SNOMED CT)
SUS-ECDS	Healthcare utilization	EC_Attendance_Source_SNOMED_CT	Attendance source (SNOMED CT)
SUS-ECDS	Healthcare utilization	EC_Discharge_Status_SNOMED_CT	Discharge status (SNOMED CT)
SUS-ECDS	Complications/adverse events	EC_Chief_Complaint_SNOMED_CT	Chief Complaint (SNOMED CT)
SUS-ECDS	Complications/adverse events	EC_Diagnosis_01	Primary diagnosis (SNOMED CT)
SUS-ECDS	Healthcare utilization	EC_Treatment_01	Primary treatment code (SNOMED CT)
SUS-ECDS	Complications/adverse events	Der_EC_Diagnosis_All	All diagnosis codes (SNOMED CT)
SUS-ECDS	Resident needs	Der_EC_Treatment_All	All treatment codes (SNOMED CT)
SUS-ECDS	Complications/adverse events	EC_Injury_Date	Injury date
SUS-ECDS	Complications/adverse events	EC_Injury_Time	Injury time
SUS-ECDS	Complications/adverse events	EC_Injury_Activity_Status_SNOMED	Injury activity status (SNOMED CT)
SUS-ECDS	Complications/adverse events	EC_Injury_Activity_Type_SNOMED_C	Injury activity type (SNOMED CT)
SUS-ECDS	Complications/adverse events	EC_Injury_Intent_SNOMED_CT	Injury intent (SNOMED CT)
SUS-ECDS	Complications/adverse events	EC_Injury_Mechanism_SNOMED_CT	Injury mechanism (SNOMED CT)
SUS-ECDS	Complications/adverse events	EC_Place_Of_Injury_SNOMED_CT	Place of injury (SNOMED CT)
SUS-ECDS	Complications/adverse events	AEA_Diagnosis_01	Primary diagnosis code (AEA code)
SUS-ECDS	Healthcare utilization	AEA_Treatment_01	Primary treatment code (AEA code)
SUS-ECDS	Complications/adverse events	Der_AEA_Diagnosis_All	All diagnosis codes (AEA codes)
SUS-ECDS	Resident needs	Der_AEA_Treatment_All	All treatment codes (AEA codes)

SUS-ECDS	Healthcare utilization	Org_Code_Patient_Pathway_ID_Issuer	Organisation code of the organisation that assigned the patient pathway identifier
SUS-ECDS	Demographics/characteristics	Month_of_Birth	Month of birth
SUS-ECDS	Demographics/characteristics	Year_of_Birth	Year of birth
SUS-ECDS	Demographics/characteristics	Sex	Sex
SUS-ECDS	Demographics/characteristics	Ethnic_Category	Ethnicity
SUS-OP	Healthcare utilization	Appointment_Date	Appointment date
SUS-OP	Healthcare utilization	Appointment_Time	Appointment time
SUS-OP	Healthcare utilization	Der_Activity_Month	Activity month
SUS-OP	Healthcare utilization	Der_Financial_Year	Financial year
SUS-OP	Healthcare utilization	Der_Provider_Code	Provider code
SUS-OP	Healthcare utilization	Der_Postcode_LSOA_Code	LSOA code
SUS-OP	Healthcare utilization	Der_Postcode_CCG_Code	CCG code
SUS-OP	Healthcare utilization	GP_Practice_Code	GP practice code
SUS-OP	Healthcare utilization	Der_Appointment_Type	Appointment type
SUS-OP	Healthcare utilization	Der_Attendance_Type	Attendance type
SUS-OP	Healthcare utilization	Attendance_Status	Attendance status
SUS-OP	Healthcare utilization	OPA_Referral_Source	Source of referral for the outpatient episode
SUS-OP	Healthcare utilization	Treatment_Function_Code	Treatment function code
SUS-OP	Healthcare utilization	Main_Speciality_Code	Main speciality code
SUS-OP	Resident needs	Der_Procedure_All	All procedure codes
SUS-OP	Demographics/characteristics	Month_of_Birth_SUS	Month of birth
SUS-OP	Demographics/characteristics	Year_of_Birth_SUS	Year of birth
SUS-OP	Demographics/characteristics	Sex	Sex

SUS-OP	Demographics/characteristics	Ethnic_Category	Ethnicity
PDS	Demographics/characteristics	Der_Practice_Code	GP practice code
PDS	Demographics/characteristics	Der_CCGofResidence	CCG code for residence location
PDS	Demographics/characteristics	Gender	Sex
PDS	Demographics/characteristics	DateofDeath	Date of death
PDS	Demographics/characteristics	Der_Postcode_LSOA_code	LSOA of residence
PDS	Demographics/characteristics	Der_DOBYearMnth	Year and month of birth
PDS	Demographics/characteristics	Change_Time_Stamp	Date and time to indicate when record was updated (for any column)
Care home residency[1] and CQC dataset	Linkage	Der_CQC_Location	Pseudonymised CQC location id
Care home residency	Care home stay	Der_Start_Date	Start of care home stay
Care home residency	Care home stay	Der_End_Date	End of care home stay
Care home residency	Care home characteristics and workforce characteristics	Der_CQC_Service_Type	Type of care home
Care home residency	Demographics/characteristics	CCG_Of_Residence	CCG of care home
Care home residency	Care home characteristics and workforce characteristics	CareHomeIndicator	? Unclear
Care home residency	Care home stay	DeathFlag	Flag to indicate resident died
Care home residency	Care home stay	YearMonthDeath_YYYYMM	Year and month of death
Care home residency	Care home stay	DeathInHospFlag	Flag to indicate death in hospital
Care home residency	Care home stay	LastResidencyFlag	Flag to indicate last residence of the person
Ambulance	Healthcare utilization	call_date	Date of call

Ambulance	Healthcare utilization	call_origin	Codes to identify the origin of the call into the ambulance control room, whether 999 was directly dialled or it was transferred from another agency.
Ambulance	Healthcare utilization	stop_codes	Codes to identify the reason the call was closed if no face to face response was received from the ambulance service.
Ambulance	Healthcare utilization	time_call_connected	Time call connected
Ambulance	Healthcare utilization	time_call_answered	Time call answered
Ambulance	Healthcare utilization	treatment_type	Codes to identify the overall outcome of the call.
Ambulance	Healthcare utilization	chief_complaint_call_triage_code	Codes to identify the initial chief complaint of the patient based on the information provided during telephone call to the ambulance control room.
Ambulance	Healthcare utilization	orgid_prov3	Provider code 3 characters
Ambulance	Healthcare utilization	time_resource_arrived_on_scene	Time resource arrived on scene
Ambulance	Healthcare utilization	receiving_location_type_cad	? Unclear
ASC CLD	Care home stay	Accommodation_Status_ASC	Accommodation status
ASC CLD	Care home stay	Client_Funding_Status_ASC	Client funding status
ASC CLD	Care home stay	Service_Component_ASC	Service component
ASC CLD	Care home stay	Service_Type_ASC	Service type
CSDS (CYP001MPI)	Demographics/characteristics	Gender	Person stated gender code or Sex
CSDS (CYP001MPI)	Demographics/characteristics	EthnicCategory	Ethnic category
CSDS (all files)	Healthcare utilization	EFFECTIVE_FROM	Date and time to indicate when record was updated (for any column)
CSDS (all files)	Healthcare utilization	RecordNumber	Record number
CSDS (CYP001MPI)	Healthcare utilization	LSOA	LSOA code
CSDS (CYP001MPI)	Healthcare utilization	Der_Postcode_yr2011_LSOA	LSOA code

CSDS (CYP001MPI)	Demographics/characteristics	AgeYr_RP_StartDate	Age of patient at reporting period start (days)
CSDS (CYP001MPI)	Care home stay	AgeYr_Death	Age at death (years)
CSDS (CYP001MPI)	Care home stay	Age_Death	Age at death (days)
CSDS (CYP001MPI)	Healthcare utilization	OrgIDICBRes	Organisation identier (ICB of residence)
CSDS (CYP001MPI)	Palliative care needs	DiscussPreferredDeathLocation_Indicator	Preferred death location discussed indicator
CSDS (CYP001MPI)	Palliative care needs	DeathLocationPreferred_Type	Death location type code (preferred)
CSDS (CYP001MPI)	Palliative care needs	DeathLocationActual_Type	Death location type code (actual)
CSDS (CYP001MPI)	Palliative care needs	NotAtPreferredLocation_Reason	Death not at preferred location reason
CSDS (CYP201CareContact)	Healthcare utilization	CareContactID	Care contact identifier
CSDS (CYP201CareContact)	Healthcare utilization	ServiceRequestID	Service request identifier
CSDS (CYP201CareContact)	Healthcare utilization	Contact_Date	Care contact date
CSDS (CYP201CareContact)	Healthcare utilization	Consultation_Type	Consultation type
CSDS (CYP201CareContact)	Healthcare utilization	Consultation_MediumUsed	Consultation mechanism e.g. face to face
CSDS (CYP201CareContact)	Healthcare utilization	Activity_LocationType	Activity location type code
CSDS (CYP201CareContact)	Healthcare utilization	OrgID_Provider	Organisation identifier (code of provider)
CSDS (CYP201CareContact)	Healthcare utilization	AgeYr_Contact_Date	Age at care contact date
CSDS (CYP201CareContact)	Healthcare utilization	AttendanceStatus	Attendance status
CSDS (CYP101Referral)	Healthcare utilization	SourceOfReferral	Source of referral
CSDS (CYP101Referral)	Healthcare utilization	Referring_StaffGroup	Referring care professional staff group (community care)
CSDS (CYP101Referral)	Healthcare utilization	PrimaryReferralReason	Primary reason for referral (community care)

CSDS (CYP102ServiceTypeReferredTo)	Healthcare utilization	ServiceRequestID	Service request identifier
CSDS (CYP102ServiceTypeReferredTo)	Healthcare utilization	TeamID_Local	Care professional team local identifier
CSDS (CYP102ServiceTypeReferredTo)	Healthcare utilization	TeamType	Service or team type referred to (community care)
CQC	Care home characteristics and workforce characteristics	Location_HSCA_start_date	Start of registration at location
CQC	Care home characteristics and workforce characteristics	Care_home?	Care home or not
CQC	Care home characteristics and workforce characteristics	Care_homes_beds	Number of beds
CQC	Care home characteristics and workforce characteristics	Location_Inspection_Directorate	Inspection Directorate
CQC	Care home characteristics and workforce characteristics	Location_Primary_Inspection_Cate	Primary inspection category
CQC	Care home characteristics and workforce characteristics	Publication_Date	Publication date
CQC	Care home characteristics and workforce characteristics	Location_Region	Region
CQC	Care home characteristics and workforce characteristics	Location_NHS_Region	NHS Region
CQC	Care home characteristics and workforce characteristics	Location_Local_Authority	Local Authority

CQC	Care home characteristics and workforce characteristics	Location_ONSPD_CCG	CCG
CQC	Care home characteristics and workforce characteristics	service_type	Service type e.g. nursing home
CQC	Care home characteristics and workforce characteristics	Location_Type	Location type
CQC	Care home characteristics and workforce characteristics	Rating_overall	Rating: overall
CQC	Care home characteristics and workforce characteristics	Rating_caring	Rating: caring
CQC	Care home characteristics and workforce characteristics	Rating_well_led	Rating: well led
CQC	Care home characteristics and workforce characteristics	Rating_effective	Rating: effective
CQC	Care home characteristics and workforce characteristics	Rating_reponsive	Rating: responsive
CQC	Care home characteristics and workforce characteristics	Rating_safe	Rating: safe
CQC	Care home characteristics and workforce characteristics	serv_user_dementia	Service users with dementia
CQC	Care home characteristics and workforce characteristics	years_since_registration	Years since registration

[1]Dataset abbreviations: ASC-CLD = Adult Social Care Client Level Dataset; CQC = Care Quality Commission; CSDS = Community Services Data Set; PDS = Personal Demographics Service; SUS-APC = Secondary Uses Service Admitted Patient Care; SUS-ECDS = Secondary Uses Service Emergency Care Dataset; SUS-OP = Secondary Uses Service Outpatient.

APPENDIX 6 – Completion of measures added into care home digital care record software, by wave, for linked DCR data*

Measure	N	i. Complete both	ii. Wave 1 only	iii. Wave 2 only	iv. Missing both [1]	Of ii… Resident died before Wave 2	Of ii… care home drop out or non-complete[1]
MDS CPS	767	58.7%	21.1%	2.4%	17.8%	27.8%	45.7%
Barthel	767	35.2%	38.6%	1.6%	24.6%	16.2%	§ 78.7%
IAGeD	767	55.3%	20.6%	4.0%	20.1%	26.6%	46.8%
ASCOT-Proxy-Resident	767	44.7%	18.9%	4.0%	32.3%	27.6%	46.2%
ASCOT: Anxiety/low mood	767	44.9%	18.9%	6.5%	29.7%	31.7%	51.0%
ASCOT: Pain	767	51.4%	20.7%	2.6%	25.3%	29.6%	49.1%
ICECAP-O	767	36.6%	37.6%	1.3%	24.5%	18.8%	§§ 76.0%
EQ-5D-5L Proxy	767	60.5%	21.8%	2.4%	15.4%	28.1%	44.9%
ASCS QoL item	767	56.6%	21.1%	2.2%	20.1%	29.0%	48.8%

*Since data on resident death between waves or grouping by care home were only available once the DCR data had been linked to other data, we only consider those residents who were eligible for data linkage, but do not omit residents with data at Wave 2 only (see Figure 1,).

[1] Due to drop out of n=5 care homes from Wave 1 to Wave 2, except when noted below (under § and §§).

§ Eleven additional care homes (**Provider 2 only**) returned no Barthel data at Wave 2 (n=157 (53.0%)). No additional information given as to why.

§§ Five additional care homes (**Provider 2 only**) returned no ICECAP-O data at Wave 2, although care staff completed other measures (n=96 (33.3%)). Despite a request for additional information, no reason was given for this omission.

APPENDIX 7: GP data items possible to access from one ICS

Type	Variable
Demographics	Ethnicity
Demographics	BMI
Demographics	Age in Years
Demographics	Has died Y/N /month and year of death
Demographics	Care home flag in GP record and/ or ICS system
Demographics	Registration with GP practice aligned to care home declined
Risk stratification	End-of-life pathway register
Risk stratification	Place of death discussed
Risk stratification	Preferred place of death
Risk stratification	Infection - urinary tract
Risk stratification	Infection - chest - lower respiratory tract infection
Risk stratification	Infection - skin
Risk stratification	Injury resulting from fall
Risk stratification	Haematological malignancies
Risk stratification	Depression
Risk stratification	Dementia
Risk stratification	Incontinence - fecal
Risk stratification	Delirium
Risk stratification	Difficulty swallowing
Risk stratification	Frailty index (eFi)
Risk stratification	eFI: activity limitation
Risk stratification	eFI: anaemia and haematinic deficiency
Risk stratification	eFI: arthritis
Risk stratification	eFI: atrial fibrillation
Risk stratification	eFI: cerebrovascular disease
Risk stratification	eFI: chronic kidney disease (CKD)
Risk stratification	eFI: diabetes
Risk stratification	eFI: dizziness
Risk stratification	eFI: dyspnoea
Risk stratification	eFI: falls
Risk stratification	eFI: foot problems
Risk stratification	eFI: fragility fracture
Risk stratification	eFI: hearing impairment
Risk stratification	eFI: heart failure
Risk stratification	eFI: heart valve disease
Risk stratification	eFI: housebound
Risk stratification	eFI: hypertension
Risk stratification	eFI: hypotension/syncope
Risk stratification	eFI Memory & cognitive problems
Risk stratification	eFI Mobility and transfer problems

Risk stratification	eFI: osteoporosis
Risk stratification	eFI: Parkinsonism and tremor
Risk stratification	eFI: peptic ulcer
Risk stratification	eFI Peripheral vascular disease
Risk stratification	eFI: polypharmacy
Risk stratification	eFI: requirement for care
Risk stratification	eFI: respiratory disease
Risk stratification	eFI: skin ulcer
Risk stratification	eFI: sleep disturbance
Risk stratification	eFI: social vulnerability
Risk stratification	eFI: thyroid disease
Risk stratification	eFI: urinary incontinence
Risk stratification	eFI: urinary system disease
Risk stratification	eFI: visual impairment
Risk stratification	eFI: weight loss and anorexia
Risk stratification	eFI: ischaemic heart disease
Medications	Numbers of current prescriptions
Medications	Most recent medication review
Medications	Penicillins (BFN 5.1.1)
Medications	Cephalosporins, carbapenems & other beta-lactams (BNF 5.1.2)
Medications	Tetracyclines (BNF 5.1.3)
Medications	Aminoglycosides (BNF 5.1.4)
Medications	Macrolides (BNF 5.1.5)
Medications	Sulfonamides and trimethoprim (BNF 5.1.8)
Medications	Quinolones (BNF 5.1.12)
Medications	Non-opioid analgesics (BNF 4.7.1)
Medications	Compound analgesic preparations (BNF 4.7.1)
Medications	Opioid analgesics (BNF 4.7.2)
Medications	Non-steroidal anti-inflammatory drugs (BNF 10.1.1)
Medications	Other drugs used as analgesics
Vaccinations	Flu
Vaccinations	Covid 19
Vaccinations	Pneumonia
Appointments/utilisation	Number of appointments by staff type
Appointments/utilisation	EHCH MDT referral
Appointments/utilisation	Hospital referral recorded?
Appointments/utilisation	A&E referral recorded?
Appointments/utilisation	Continence service referral recorded?
Appointments/utilisation	Community services referral recorded?
Appointments/utilisation	Community nurse referral recorded
Appointments/utilisation	Referrals to falls service
Appointments/utilisation	Face to Face

Appointments/utilisation	Telephone appointments

APPENDIX 8 – Reason for non-inclusion of planned variables in final MDS

Domain	Variable	Expected source[1]	Reason for not included/ reason for different source
Demographics/ characteristics	Date of birth	PDS	Included, but sourced from PDS and SUS
	Sex registered at birth	PDS	Included, but sourced from PDS, SUS, CSDS
	Religion	DCRs	Included, but high % missing data
	First language	DCRs	Included, but high % missing data
	Marital status	DCRs	Included, but high % missing data
	Ethnicity	DCRs, GP data	Included, but high % missing data
	Weight	DCRs, GP data	Included, but sourced from DCRs only
	Height	PDS linked to public IMD data	Included, but sourced from DCRs only
	Area-based deprivation (last known residence)		The version of PDS we accessed did not allow access to location of previous address
Palliative care needs	End-of-life pathway register	GP data	No access to GP data, but sourced information from CSDS
Care home stay	Date of death	PDS	Included, but sourced from PDS and SUS
	Admitted from hospital or community	SUS	Included, but limited to discharge from hospital to care home in previous year
Resident needs	Visual impairment	GP data	No access to GP data
	Hearing impairment	GP data	No access to GP data
	Cognitive impairment	DCRs, GP data, SUS	Included, but sourced from DCRs and SUS only
	Oral and nutritional status	GP data, SUS	Included, but sourced from SUS only
	Continence	DSRs, GP data SUS	Included, but sourced from DCRs and SUS only
Quality of life			
Complications/ adverse events	Infections	DCRs, GP data, SUS	Included, but sourced from SUS only and limited to upper and lower respiratory tract infections. Not able to extract from DCRs in standardised format.
	Falls leading to GP or hospital visit	DCRs, GP data, SUS, 999 / ambulance data	Included, but sourced from SUS only. Field in ambulance dataset required to derive this was 100% missing. Not able to extract from DCRs in standardised format.
	Falls (recorded in care home only)	DCRs	Not able to extract in standardised format.
	NEWS2/RESTORE	DCRs	Included, but 100% missing
Diagnoses	Medical history	GP data, SUS	Included, but sourced from SUS only
	Frailty	GP data, SUS	Included, but sourced from SUS only
	Adverse reactions and allergies	GP data	No access to GP data

Medications/ vaccinations	Prescribed medications	GP data	No access to GP data
	Administered vaccinations		
Healthcare utilisation	Primary care use Out of hours contacts Ambulance attendances for falls	GP data, NHS111, ambulance GP data Ambulance	Included, but sourced from ambulance data only. No access to GP data. However, we derived out of hours ambulance callouts. Field in dataset required to derive this was 100% missing so could not use to get reason for attendances
	Ambulance conveyances to ED		Field in dataset required to derive this was populated differently to as expected from data specifications

[1] Dataset abbreviations: CSDS = Community Services Data Set; DCR = Digital Care Record; PDS = Personal Demographics Service; SUS = Secondary Uses Service

APPENDIX 9 - Comparison of variables with inconsistent definitions across data sources

A small number of comorbidities were recorded in both hospital and care home records. However, these variables were not defined in a consistent way. Even within a data source, there can be several established definitions and code lists, which will give slightly different results. Table a shows that 11% of residents in our sample are identified as having dementia from SUS data according to one definition but not another (53 + 13/583).

The comorbidities were also collected differently: variables in the care home data were recorded at one particular point in time - therefore reflecting the resident's health status at that moment. For the SUS data, we collected information relating to these conditions over a three-year look-back period. For acute conditions, there were high levels of discrepancy between SUS and care home DCRs. For example, delirium was recorded in hospital records for 145 patients in the previous 3 years, but only 20 (14%) of these were recorded as delirious by care home staff at the time of recording. However, there were also 30 (48%, 30/62) residents recorded with delirium in the care home DCRs that were not reflected in hospital records prior to that date. For cognitive impairment, which tends to not improve over time, there was more consistency, although still substantial disagreement. Agreement between care home DCR and SUS record of dementia is recorded in Table 3 in main results.

Table a) Comparison of residents identified to have dementia based on Charlson index code list and Frailty syndromes code list

Comparing dementia prevalence within SUS		Frailty syndromes code list			
		No	Yes	Missing	Total
Charlson index code list	No	189	13	0	202
	Yes	53	328	0	381
	Missing	0	0	144	144
	Total	242	341	144	727

Table b) Comparison of residents identified to have delirium based on SUS data using Soong et al. List of frailty syndromes and care home DCR using I-AGED

Delirium		Care home DCR			
		No	Yes	Missing	Total
SUS	No	270	30	73	373
	Yes	145	20	45	210
	Missing	105	12	27	144
	Total	520	62	145	727

Table c) Comparison of residents identified to have cognitive impairment based on Soong et al. list of frailty syndromes and care home DCR, assess using Morris et al.

Cognitive impairment		Care home DCR								
		Intact	Borderline intact	Mild impair-ment	Moderate impair-ment	Moderately severe impairment	Sever impair-ment	Very severe impair-ment	Missing	Total
SUS	No	13	52	13	10	8	7	6	17	126
	Yes	33	45	61	84	52	60	43	79	457
	Missing	10	19	11	17	20	21	27	19	144
	Total	56	116	85	111	80	88	76	115	727

Note: 144 people are missing for each of the SUS measures as no inpatient hospital record

APPENDIX 10 – Complete final prototype MDS.

Domain	Variable	Categories (if categorical)	n[1]	Mean (SD) or %
Demographics/ characteristics	Ethnicity (final)[2]	White	692	95%
		Black or Black British	<=5	NA
		Asian or Asian British	<=5	NA
		Mixed	<=5	NA
		Other	<=5	NA
		Missing	25	3%
	Sex (final)[2]	Female	513	71%
		Male	214	29%
	Date of birth record (final)[2]	Available	>=720	99%
		Missing	<=5	NA
	Date of death present in record (final)[2]	Present	58	8%
		Not present	669	92%
	Religion	Christianity	93	13%
		Buddhist	<=5	NA
		Other	8	1%
		No religion	13	2%
		Missing	>=610	84%
	Marital status	Divorced/separated/single	15	2%
		Married/cohabiting	40	6%
		Widowed	54	7%
		Missing	618	85%
	First language spoken	English	160	22%
		Other	<=5	NA
		Missing	>=561	77%
	Power of attorney	Yes	61	8%
		No	103	14%
		Missing	563	77%
	Deprivation of Liberty status	Yes	126	17%
		No	544	75%
		Missing	57	8%
	DNACPR status	Yes	572	79%
		No	>=150	21%
		Missing	<=5	NA
	Weight	20-35kg	7	1%
		36-50kg	109	15%
		51-65kg	232	32%
		66-80kg	152	21%
		81-95kg	43	6%
		96-110kg	21	3%
		111-125kg	<=5	NA
		126-140kg	<=5	NA

		Missing	158	22%
	Height	111-125cm	8	1%
		126-150cm	67	9%
		151-170cm	508	70%
		171-190cm	130	18%
		191-210cm	<=5	NA
		Missing	>=8	1%
Palliative care needs	Discussed preferred death location indicator	Yes	18	3%
		No	383	53%
		Missing	326	45%
	Preferred death location	Care home	7	1%
		Care home services with nursing	27	4%
		Care home services without nursing	51	7%
		Hospice	<=5	NA
		Hospital	<=5	NA
		Patient's own home	16	2%
		Other (not listed)	<=5	NA
		Missing	623	86%
Care home stay	Client funding status	Health funded	7	1%
		Social care funded	18	2%
		Client funded	19	3%
		Joint client and social care funded	96	13%
		Other	<=5	NA
		Unknown in record	77	11%
		Missing	>=505	70%
	Discharge from an in-patient spell to a care home (1 year history)	Yes	32	4.40%
		No	695	95.5%
	Death in hospital in the period between the index date and end of study	Yes	<=5	NA
		No	>=720	99%
	Length of stay in care home			876.94 (812.60)
Resident needs	Cognitive impairment	Borderline intact	56	8%
		Intact	116	16%
		Mild impairment	85	12%
		Moderate impairment	111	15%

		Moderately severe impairment	80	11%
		Severe impairment	88	12%
		Very severe impairment	76	10%
		Missing	115	16%
	Bowel continence	Continent	201	28%
		Incontinent	289	40%
		Occasional accident	109	15%
		Missing	128	18%
	Bladder continence	Continent	143	20%
		Incontinent/Catheter	329	45%
		Occasional accident	124	17%
		Missing	131	18%
	ASCOT: Pain	High-level needs	21	3%
		Some needs	59	8%
		No needs	263	36%
		Ideal state	210	29%
		Missing	174	24%
	ASCOT: Anxiety and low mood			3.97 (1.45)
	Food texture requirements	IDDSI 3 - Liquidised	<=5	NA
		IDDSI 4 - Pureed	41	6%
		IDDSI 5 - Minced & Moist	47	6%
		IDDSI 6 - Soft & Bite-sized	65	9%
		IDDSI 7 - Easy to Chew	34	5%
		IDDSI 7 - Regular	482	66%
		Missing	>=54	8%
	Drink thickness requirements	IDDSI 0 - Thin	598	82%
		IDDSI 1 - Slightly thick	30	4%
		IDDSI 2 - Mildly thick	23	3%
		IDDSI 3 - Moderately thick	<=5	NA
		IDDSI 4 - Extremely thick	<=5	NA
		Missing	71	10%
	Food allergy	Yes	26	4%
		No	633	87%
		Missing	68	9%
	Contact allergy	Yes	<=5	NA
		No	156	21%
		Missing	>=564	78%
	Medication allergy	Yes	73	10%
		No	88	12%
		Missing	566	78%
	Penicillin allergy	Yes	27	4%
		No	134	18%
		Missing	566	78%

	Pressure ulcers (Waterlow score) (reported for 527 residents/ 28% missing)			17.87 (7.30)
	Pressure ulcers (Braden score) (reported for 338 residents/ 54% missing)			16.61 (4.02)
	Delirium (I-AGED score) (reported for 582 residents/ 20% missing)			1.11 (1.78)
	Functional independence (Barthel score) (reported for 566 residents/ 22% missing)			41.40 (30.26)
	Number of ED attendance or outpatient appointments with nasogastric feeding procedure			NA
Quality of life	Quality of Life overall (reported for 596 residents/ 18% missing)	So good, it could not be better	19	3%
		Very good	157	22%
		Good	201	28%
		Alright	171	24%
		Bad	32	4%
		Very bad	9	1%
		So bad, it could not be worse	7	1%
		Missing	131	18%
	Ascot Proxy-Resident (reported for 488 residents/ 33% missing)			0.83 (0.19)
	ICECAP-O (reported for 569 residents/ 22% missing)			0.73 (0.21)
	UK Crosswalk (reported for 631 residents/ 13% missing)			0.33 (0.35)
Complications/ adverse events	MUST (malnutrition universal scoring tool) score	0	101	14%
		1	23	3%
		2	20	3%
		3	12	2%
		4	6	1%
		5	<=5	NA
		Missing	>=562	78%
	Frequency of in-patient admissions with upper respiratory tract infection recorded (1 year history)			0.01 (0.10)
	Frequency of in-patient admissions with lower respiratory tract infection recorded (1 year history)			NA
Diagnoses (based on previous 3 years hospital admission diagnosis codes) (reported for 583 residents/ 20% missing)	Dementia (final)[2]		514	71%
	Elixhauser conditions[3]			
	Number of Elixhauser conditions			3.59 (2.34)
	2 or more Elixhauser conditions		470	81%
	Alcohol abuse		22	4%
	Anaemia		83	14%
	Cardiac arrhythmias		189	32%
	Chronic pulmonary disease		110	19%
	Coagulopathy		15	3%

apart from 'dementia (final)'	Congestive heart failure	86	15%
	Depression	129	22%
	Diabetes (complicated and uncomplicated)	127	22%
	Drug abuse	<=5	NA
	Fluid and electrolyte disorders	226	39%
	Hemiplegia / paraplegia	20	3%
	Hypertension (complicated and uncomplicated)	353	61%
	Hypothyroidism	75	13%
	Liver disease	30	5%
	Lymphoma	6	1%
	Metastatic solid tumour / metastatic cancer	11	2%
	Obesity	39	7%
	Other neurological disorders	154	26%
	Peptic ulcer disease excl bleeding	6	1%
	Psychoses	13	2%
	Pulmonary circulation disorders	25	4%
	Peripheral vascular disease	47	8%
	Rheumatoid arthritis / collagen vascular diseases	179	31%
	Renal failure	39	7%
	Solid tumour without metastasis	22	4%
	Valvular disease	67	11%
	Weight loss	19	3%
	Frailty syndromes[4]		
	Number of frailty syndromes		2.17 (1.81)
	Cognitive impairment (delirium, dementia, senility)	457	78%
	Anxiety/Depression	168	29%
	Functional dependence	102	17%
	Falls/Fractures	291	50%
	Incontinence	105	18%
	Mobility problems	217	37%
	Pressure ulcers	62	11%
Healthcare utilisation		*n (people with at least one event)*	*% who had at least one event*
	Elective admissions (1 year history)	65	9%
	Emergency admissions (1 year history)	284	39%
	Potentially avoidable emergency admissions (1 year history) [5]	119	16%
	Emergency department attendances (1 year history)	370	51%
	Emergency department attendances via ambulance (1 year history)	331	46%

Emergency department attendances after 6pm/ before 8am (1 year history)	218	30%
Emergency department attendances lasting more than 12 hours (1 year history)	156	21%
Emergency department attendances for falls (1 year history)	18	2%
Outpatient appointments (1 year history)	236	32%
Missed outpatient appointments (1 year history)	28	4%
Community services appointments (1 year history)	608	84%
Speech and language therapy appointments (1 year history)	49	7%
Continence appointments (1 year history)	159	26%
District nursing appointments (1 year history)	398	55%
Podiatry appointments (1 year history)	31	4%
Community rehabilitation appointments (1 year history)	79	11%
Face to face community services appointments (1 year history)	444	61%
Missed community services appointments (1 year history)	22	3%
Ambulance call outs (1 June - 31 October 2023)	197	27%
Ambulance attendances (1 June - 31 October 2023)	195	27%
Ambulance attendances after 6pm/ before 8am (1 June - 31 October 2023)	118	16%
Ambulance conveyances (1 June - 31 October 2023)	147	20%
	Total activity	*Mean (SD)*
Average number of emergency admissions (1 year history)	451	0.62 (1.02)
Average number of elective admissions (1 year history)	97	0.13 (0.60)
Average number of potentially avoidable emergency admissions (1 year history) [5]	143	0.20 (0.48)
Average number of emergency department attendances (1 year history)	752	1.03 (1.49)
Average number of emergency department attendances after 6pm/ before 8am (1 year history)	331	0.46 (0.85)
Average number of emergency department attendances lasting more than 12 hours (1 year history)	206	0.28 (0.63)
Average number of emergency department attendances via ambulance (1 year history)	605	0.83 (1.27)
Average number of emergency department attendances due to a fall (1 year history)	21	0.03 (0.20)

	Average number of outpatient appointments (1 year history)		424	0.58 (1.47)
	Average number of missed outpatient appointments (1 year history)		29	0.04 (0.20)
	Average number of community services appointments (1 year history)		15266	21.00 (65.83)
	Average number of missed community services appointments (1 year history)		36	0.05 (0.33)
	Average number of face to face community services appointments (1 year history)		6720	9.24 (30.47)
	Average number of district nursing appointments (1 year history)		11347	15.61 (61.82)
	Average number of speech and language therapy appointments (1 year history)		155	0.21 (1.03)
	Average number of podiatry appointments (1 year history)		286	0.39 (3.98)
	Average number of continence appointments (1 year history)		373	0.51 (1.44)
	Average number of community rehabilitation appointments (1 year history)		470	0.65 (3.85)
	Average number of ambulance call outs (1 June-31 October 2023)		333	0.45 (0.97)
	Average number of ambulance attendances (1 June-31 October 2023)		325	1.65 (1.21)
	Average number of ambulance attendances after 6pm/ before 8am (1 June-31 October 2023)		156	0.21 (0.57)
	Average number of ambulance conveyances (1 June-31 October 2023)		210	0.29 (0.72)
Care home characteristics and workforce characteristics	Service type	Nursing	403	55%
		Nursing and Residential	49	7%
		Residential	262	36%
		Missing	13	2%
	Registered bed capacity	Less than 50	211	29%
		50 or more	485	67%
		Missing	31	4%
	CQC rating	Outstanding	72	10%
		Good	511	70%
		Requires improvement	113	16%
		Missing	31	4%
	Years of service registration	Less than 10 years	238	33%
		More than 10 years	458	63%
		Missing	31	4%
	Number of beds (reported for 696 residents/ 4% missing)			54.33(16.68)
	Number of beds currently occupied (reported for 696 residents/ 4% missing)			42.54(14.75)

	Number of resident fully self-funding (reported for 650 residents/ 11% missing)		22.41(16.26)
	Number of staff (reported for 696 residents/ 4% missing)		67.52(32.16)
	Number of full-time staff (reported for 696 residents/ 4% missing)		45.55(20.35)
	Number of part-time staff (reported for 659 residents/ 9% missing)		19.32(13.42)
	Number of staff on permanent contracts (reported for 659 residents/ 9% missing)		62.54(28.14)
	Number of staff vacancies (reported for 696 residents/ 31% missing)		3.77(10.36)
	Number of agency staff (reported for 600 residents/ 17% missing)		1.52(4.28)
	Number of care workers (reported for 659 residents/ 9% missing)		32.44(20.69)
	Number of senior care workers (reported for 659 residents/ 9% missing)		9.45(4.43)
	Number of registered nurses (reported for 644 residents/ 11% missing)		5.05(6.07)
	Number of nursing associates (reported for 632 residents/ 13% missing)		0.43(1.31)
	Number of nursing assistants (reported for 632 residents/ 13% missing)		0.31(0.89)
	Number of allied health professionals (reported for 632 residents/ 13% missing)		0(0)
	Number of activities coordinators (reported for 659 residents/ 9% missing)		1.75(1.38)
	Number of staff in roles above on permanent contracts (reported for 632 residents/ 13% missing)		43.98(25.68)
	Number of staff vacancies in roles above (reported for 632 residents/ 13% missing)		1.93(1.84)
	Number of agency staff in roles above (reported for 592 residents/ 19% missing)		0.71(3.67)

[1] Numbers are reported for 727 residents unless otherwise specified
[2]Reporting variable as created in the hierarchy process – see Table 3
[3]Elixhauser list of comorbidities (E Elixhauser A, Steiner C, Harris DR, et al. Comorbidity measures for use with administrative data. Med Care. 1998;36:8–27. doi: 10.1097/00005650-199801000-00004; Quan H, Sundararajan V, Halfon P, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. Med Care. 2005;43:1130–9. doi: 10.1097/01.MLR.0000182534.19832.83)
[4]Frailty Syndromes (Soong J, Poots AJ, Scott S, et al. Developing and validating a risk prediction model for acute care based on frailty syndromes. BMJ Open. 2015;5. doi: 10.1136/BMJOPEN-2015-008457)
[5]Potentially avoidable emergency admissions (Care Quality Commission, Great Britain. Parliament. House of Commons. The state of health care and adult social care in England in 2012/13. 2013;86.).

APPENDIX 11 – Using MDS data to understand ED attendance and ambulance activity for care home residents: a worked example

In the main report, we presented key variables from the MDS with a full description of the MDS in Appendix 8. However, a key benefit of an MDS is the ability to explore sub groups of residents. These tables are examples of analyses that could help understand whether there are differences in outcomes or activities in different subgroups. Tables below are examples of such subgroup analyses which were suggested by stakeholders as of interest. For example, in our sample population, activity was in general higher across both A&E and ambulance services for those in residential care homes, compared to nursing homes, which could be informative when commissioning local services.

Table a) ED attendances for year leading to index date 1. * where no mean reported as denominator <=5. Mean reports the mean of all those eligible e.g. can only have ED attendance via ambulance if you have had an ED attendance

Variable	Total	Sex		Age				Nursing vs residential			Dementia			Deprivation[1]					
		Female	Male	<65	65-79	>=80	Missing	Nursing	Residential	Missing	No	Yes	Missing	Most deprived fifth	Second most deprived quintile	Middle fifth	Second least deprived fifth	Least deprived fifth	Missing
N	**727**	**513**	**214**	**17**	**125**	**486**	**99**	**452**	**262**	**13**	**191**	**514**	**22**	**81**	**158**	**72**	**119**	**198**	**99**
Mean (SD)																			
Number of A&E attendances	1.03 (1.49)	0.94 (1.44)	1.25 (1.61)	1.06 (1.34)	1.05 (1.46)	0.93 (1.41)	1.51 (1.85)	0.92 (1.44)	1.21 (1.57)	1.38 (1.50)	0.92 (1.39)	1.11 (1.55)	*	0.95 (1.37)	1.02 (1.40)	1.04 (1.30)	1.13 (1.63)	0.79 (1.35)	1.51 (1.85)
Number of out of hours A&E attendances	0.46 (0.85)	0.41 (0.77)	0.57 (1.00)	0.76 (1.25)	0.50 (0.81)	0.41 (0.81)	0.56 (0.95)	0.39 (0.82)	0.56 (0.90)	*	0.38 (0.81)	0.50 (0.87)	*	0.52 (1.01)	0.47 (0.78)	0.47 (0.87)	0.48 (0.85)	0.35 (0.75)	0.56 (0.95)
Number of A&E attendances lasting >12 hours	0.28 (0.63)	0.25 (0.62)	0.36 (0.65)	*	0.35 (0.65)	0.25 (0.59)	0.36 (0.80)	0.26 (0.58)	0.32 (0.72)	*	0.29 (0.71)	0.29 (0.62)	*	0.07 (0.26)	0.34 (0.66)	0.32 (0.60)	0.21 (0.57)	0.32 (0.66)	0.36 (0.80)
Number of A&E attendances via ambulance	0.83 (1.27)	0.76 (1.22)	1.00 (1.37)	0.71 (1.16)	0.82 (1.23)	0.79 (1.27)	1.09 (1.33)	0.72 (1.22)	1.02 (1.34)	1.00 (1.15)	0.76 (1.23)	0.89 (1.30)	*	0.85 (1.29)	0.91 (1.26)	0.90 (1.13)	0.79 (1.40)	0.63 (1.18)	1.09 (1.33)
Number of A&E attendances for falls	0.03 (0.20)	0.02 (0.16)	0.04 (0.26)	*	*	0.02 (0.19)	0.08 (0.31)	*	0.06 (0.30)	*	*	0.04 (0.22)	*	*	*	*	*	*	0.08 (0.31)

[1]Deprivation is based on LSOA of the care home

Table b) Ambulance activity for 5 month period from 1 June to 31 October. * where no mean reported as denominator <=5. Mean reports the mean of all those eligible e.g. can only have ED attendance via ambulance if you have had an ED attendance

Variable	Total	Sex		Age				Nursing vs residential			Dementia			Deprivation[1]					
		Female	Male	<65	65-79	>=80	Missing	Nursing	Residential	Missing	No	Yes	Missing	Most deprived fifth	Second most deprived quintile	Middle fifth	Second least deprived ffith	Least deprived fifth	Missing
N	**727**	**513**	**214**	**17**	**125**	**486**	**99**	**452**	**262**	**13**	**191**	**514**	**22**	**81**	**158**	**72**	**119**	**198**	**99**
Mean (SD)																			
Number of ambulance call outs	0.46 (0.99)	0.42 (0.90)	0.54 (1.19)	*	0.47 (1.22)	0.41 (0.91)	0.74 (1.09)	0.41 (1.02)	0.52 (0.91)	*	0.54 (1.08)	0.44 (0.97)	*	0.85 (1.59)	0.35 (0.88)	0.38 (0.76)	0.42 (0.81)	0.30 (0.81)	0.74 (1.09)
Number of ambulance attendances	0.45 (0.97)	0.42 (0.89)	0.52 (1.14)	*	0.45 (1.20)	0.41 (0.89)	0.71 (1.04)	0.40 (0.99)	0.52 (0.91)	*	0.53 (1.08)	0.43 (0.94)	*	0.83 (1.52)	0.35 (0.88)	0.38 (0.76)	0.42 (0.81)	0.28 (0.79)	0.71 (1.04)
Number of out of hours ambulance attendances	0.21 (0.57)	0.20 (0.50)	0.25 (0.71)	*	0.22 (0.58)	0.20 (0.55)	0.33 (0.69)	0.18 (0.53)	0.26 (0.63)	*	0.23 (0.58)	0.22 (0.58)	*	0.38 (0.92)	0.13 (0.41)	0.24 (0.52)	0.23 (0.51)	0.14 (0.47)	0.33 (0.69)
Number of ambulance conveyances	0.29 (0.72)	0.27 (0.67)	0.35 (0.85)	*	0.35 (0.98)	0.26 (0.67)	0.40 (0.65)	0.27 (0.77)	0.32 (0.62)	*	0.38 (0.87)	0.26 (0.67)	*	0.60 (1.13)	0.16 (0.64)	0.28 (0.56)	0.31 (0.69)	0.20 (0.64)	0.40 (0.65)

[1]Deprivation is based on LSOA of the care home