Statistical Analysis Protocol for an evaluation of providing enhanced support for care home residents in Rushcliffe

Therese Lloyd and Adam Steventon, October 2016

Addendum added November 2018

The Health Foundation

The Improvement Analytics Unit

The Improvement Analytics Unit is an innovative partnership between NHS England and the Health Foundation that provides robust analysis to help the NHS improve care for patients. We use advanced statistical techniques to provide evidence of whether local programmes are having an impact on improving the quality and efficiency of care. We do this by assessing whether the care delivered to patients as part a local programme (such as a new clinical model or an integrated care system) is different in any significant way from the outcomes of patients who have not experienced a similar initiative.

Our aim is that our analysis helps the local NHS and its partners identify whether implementation of an initiative is having the desired effect, or needs to change to succeed. At a national level, we support decision-makers to identify what works well and assess the impact of national priorities.

For more information see: www.health.org.uk/IAU

Contents

Glossary

Abbreviation	Description
A&E	Accident and emergency
CCG	Clinical commissioning group
CGA	Comprehensive geriatric assessment
CQC	Care Quality Commission
Emergency admission	Unplanned admission to hospital
GP	General practitioner
IAU	Improvement Analytics Unit
IMD	Index of Multiple Deprivation
MCP	Multispecialty community provider
NHAIS	National Health Applications and Infrastructure Services
NHS	National Health Service
PARR	Patients at Risk for Re-hospitalisation
RECORD	REporting of studies Conducted using Observational Routinely-collected health Data
SAP	Statistical analysis protocol
STROBE	STrengthening the Reporting of OBservational studies in Epidemiology
SUS	Secondary Uses Service
tNR	temporary National Repository

1. Background

Overview

Principia Partners in Health is a local partnership of GPs, patients and community services that aims to develop a better way to improve the care and health of the population of Rushcliffe in Nottinghamshire. Established as a Community Interest Company in 2006 it serves a population of 126,000. It has a history of engaging with central transformation initiatives and was a Department of Health social enterprise pathfinder and a Department of Health Integrated Care Pilot (2009–2011). There its focus was on managing patients with long-term conditions at high risk of admission, by means of virtual community wards and an integrated clinical pathway for people with severe chronic obstructive pulmonary disease.[1]

In April 2015 Principia Partners in Health was chosen as a New Model of Care Vanguard site with a remit to establish a multispecialty community provider (MCP). Its aim in establishing an MCP is to create a reorganised system that combines health and social care partners in a culture of integrated working and mutual accountability for patient experience and outcomes. The focus is on early intervention, living well at home and avoiding unnecessary hospital use, with an intended impact of reduction in fragmentation, delays, duplication and inefficiencies experienced by patients and carers.

Part of this transformation includes enhancing support to the area's care homes – a service transformation cited as an essential component of all MCPs in the recently published MCP framework.[2] The improvements in care homes began in April 2014 and the Vanguard programme has provided additional funding and central support to further develop this initiative.

The enhanced support to care homes that was developed and implemented from April 2014 by Principia is the subject of this evaluation.

The enhanced support was offered in all care homes caring specifically for frail older residents in Rushcliffe and in two care homes for frail older residents in neighbouring areas that were under the care of a Rushcliffe Clinical Commissioning Group (CCG) general practice.

The enhanced support is based on four elements:

1.0 enhanced GP specification for frail older people living in care homes

2.0 enhanced community nurse support and falls therapist for frail older people living in care homes

3.0 advocacy and independent support, delivered by Age UK Nottinghamshire, a charitable volunteer organisation

4.0 engaged care home managers.

The enhanced GP specification involves fortnightly resident ward review rounds in the care home, by a designated named GP to all the residents of the care home; better care planning with residents and their families, including end-of-life planning; regular medication reviews;

standardised long-term conditions planning; non-elective hospital activity review within 48 hours of discharge; review of new residents within five days and a comprehensive geriatric assessment (CGA) within two weeks; increased identification of dementia; and meetings between the care home managers and the CCG throughout the year.

The community nursing and specialist support includes nursing support during the resident ward rounds; falls support and training; access to specialist community services such as geriatrician, dietetics, heart failure and respiratory specialists; ongoing staff training such as on continence management and pressure care; a named community matron/senior nurse for each care home; direct access to a district nurse on call at weekends; and peer support for care home nurses.

Age UK's work, carried out by volunteers from the third sector, consists of providing independent information on GP alignment to the residents and their families (ie on changing to the designated named GP after admission to the care home), advocacy support, supplying a safe and trusted point of contact, facilitating the care home managers' network, and creating a space for residents and their families to raise any issues or concerns ('worry catcher').

The care home managers are able to engage with Rushcliffe CCG, have a voice and a point of contact in the MCP, experience shared ownership, and support service developments.

The theory of change behind this list of interventions is to improve the residents' care, involvement in their care and quality of life, through a multidisciplinary team where all stakeholders are engaged and working together, taking joint responsibility for the care and wellbeing of the care home residents.

Intended impact on outcomes

It is anticipated that as a result of improved proactive care to residents and support to care home staff, benefits will include a reduction in community acquired pressure sores, ambulance conveyances, A&E attendances and emergency (urgent and unplanned) admissions among the care home population. Access to a multidisciplinary team, including specialist community services and post-discharge follow-up, may also impact on the number of bed days, outpatient attendances and elective (planned) admissions. Quality of life may be improved as a result of residents being more involved in decisions relating to their care (including end-of-life care) and having the opportunity to raise any worries or issues with a person not directly involved in their care through the worry catcher.

Objectives of the analysis

This study aims to estimate the effectiveness of the enhanced support offered to residents in Principia-run care homes for older residents.

We will evaluate the overall effect of the Principia enhanced support in care homes on secondary care use for older residents of local care homes. This will include numbers of 'potentially avoidable' emergency admissions from care homes, all emergency admissions, total bed days, A&E attendances, outpatient attendances and elective admissions per person over a period of up to 24 months, allowing for a 'bedding-in' period for the enhanced support to become fully established. We will also examine the proportion of deaths that did not occur in hospital, as a proxy for residents dying in their preferred place of death.

We will not evaluate other potential impacts of the enhanced support, such as quality of life or improvement in working relationships, due to the limitations of the data available. Costs will not be evaluated in this study.

2. Methods

Study design

We will compare the outcomes of residents of the participating care homes with those of a retrospectively matched control group of care home residents.

The evaluation will consist of two separate analyses on two cohorts of residents. The main analysis will be based on the cohort of residents who entered a care home after August 2014 (resident cohort 2). A second analysis will be based on residents who were already living in a care home in August 2014 (resident cohort 1). The rationale for treating these separately is explained in more detail in 'Variable definitions' in section 2.

Matching will be done in two stages and at three levels. First, we will identify a group of local authorities that are comparable to Rushcliffe in terms of their demography, socioeconomic characteristics, rates of limiting long-term illness and emergency hospital use. Second, we will match residents of care homes in these areas to the residents receiving the interventions in Principia. This will be done separately for each cohort. At this stage, we will match on the characteristics of both care homes and residents (ie levels two and three of the matching process). In other words, line-level routine, patient and care home information will be used to characterise the intervention care home residents and their care homes at baseline as well as possible, and then select, for every resident, matched controls that did not receive the Principia enhanced support but had similar observed characteristics to the enrolled residents and also lived in similar care homes. These matched residents for cohort 1 and 2 will form the control groups for the evaluation of the programme.

Once matched controls have been selected, we will estimate the effect of the Principia enhanced support compared with the control group by fitting hierarchical mixed-model regression models. This will be done separately for each of the two cohorts.

Study cohorts

Definition of target population

The target population will be defined as all people aged 65 or over who are residents of a care home catering to older residents in the area covered by Principia.

Definition of study cohort

The study cohorts will be defined in terms of both resident and care home-level characteristics.

At resident level, we will identify people aged 65 or over who are resident in a care home at any time during the period from 17 August 2014 to 14 August 2016, which is the post-intervention period for which we can identify care home residents and have reliable data (see 'Variable definitions' in section 2). Cohort 2 will comprise people who moved into a care home after 17 August 2014. Cohort 1 will comprise people already resident in a care home on 17 August 2014.

The following residents will be excluded:

- residents without full address recorded in the National Health Applications and Infrastructure Services (NHAIS) data (see below)
- residents without a recorded month and year of birth
- residents without a record of prior emergency or elective admissions in the two-year pre-study period (defined in 'Pre-period' in section 2). These will be excluded because prior hospital data are required to define baseline resident characteristics
- residents who could not be matched to a control resident.

For cohort 2, an additional exclusion will be:

- people who were known to previously reside in a Principia care home, that is, who were resident in a Principia care home between 17 August 2014 and the date of admission to the care home in question. This exclusion criterion will be applied because any 'baseline' covariates would otherwise encompass the effect of the Principia enhanced support in the first care home.

At care home level, we will include care homes that cater to older residents, that were open during any period between 17 August 2014 and 14 August 2015, and that either benefitted from the Principia enhanced support or are located in one of the local authorities in England identified in the first step of the matching process. The cut-off date is set to 14 August 2015 so that the potential follow-up period for all care homes is at least a year.

The following care homes will be excluded:

- Care homes that operate other care home interventions during the follow-up period that we are aware of before the data transfer from the temporary National Repository (tNR) and that are deemed sufficiently similar to the Vanguard interventions, or unrepresentative of usual care in other parts of the country. These include other Vanguard sites[3] and care homes enrolled in the PEACH study (implementation of CGA in care homes),[4] other than Principia care homes. CGA is one of the interventions that Principia implemented as part of the enhanced support and Rushcliffe is one of the four CCGs taking part in the PEACH study.
- Care homes that are not either wholly under the Principia enhanced service or standard care. There was one care home outside Rushcliffe where it had been agreed that 20 (50%) of the patients were to receive the Principia enhanced support under a Principia general practice, while the other 20 patients were to be cared for by a non-Principia general practice. We will exclude this care home from the analysis, as only half of the care home residents were eligible for the intervention.
- Care homes likely to be specialist care homes for groups other than the frail older population, such as those for learning disabilities, as these would have very different primary care needs and usage of acute services.

Data on care home specialties are available from the Care Quality Commission (CQC), the independent regulator of all health and social care services in England that carries out regular inspections of all care homes in England.

The CQC data were not designed for research purposes and have not been properly validated. The CQC register allows multiple specialisations to be registered for care homes. Anecdotal evidence suggests that, especially when the Care Quality Commission was first established in 2009, registering care homes would often add more specialisations than intended. A few of the Principia care homes, although described by the Principia team as care homes for older frail adults, do nonetheless according to the CQC database cater also to adults aged under 65 or those with mental health needs, for example. We will therefore endeavour to exclude care homes likely to be genuine specialist care homes but include care homes for older people that are able to accommodate residents with more complex needs or who are younger than 65.

We will exclude care homes that are likely to be specialist care homes for groups other than the frail older population by excluding care homes that meet both of the following two criteria:

- have at least one of the following recorded specialties: learning disabilities or autistic spectrum disorder, people who misuse drugs and alcohol, people with eating disorders, people detained under the Mental Health Act, or people with sensory impairment
- are recorded as catering to additional age groups, other than just people aged 65 and over.

We will not exclude care homes due to specialisms in dementia care, mental health care or physical disability, as these categories are not inconsistent with the needs of older frail patients. The category mental health was recorded in the CQC website for a large number of care homes.

Although we are not excluding care homes that cater to younger age groups as well as older, the resident-level inclusion criteria ensure that only older patients will form part of the evaluation. In addition, we will match care homes according to the additional age groups they cater to.

There are around 840 residents in a total of 23 care homes for older people in the Principia intervention group. This includes 22 care homes in Rushcliffe. In addition, there was one care home outside of Rushcliffe that received enhanced support under a Principia general practice and is therefore included in the intervention group. Nine of the care homes were nursing homes (with an average of 47 beds each), while 13 were residential (average of 28 beds); one home had both a nursing and residential section (with 20 and 35 beds, respectively).

It is estimated that care home residents have a 45% risk of dying within the first year of admission to a care home,[5] with a median length of stay of 17.9 months for permanent residents and 4 weeks for temporary residents.[6] It is therefore estimated that cohort 2 will include at least 400 Principia care home residents.

Sources of data

We will have access to pseudonymised (ie anonymised in line with the Information Commissioner's Office code of practice on anonymisation) Secondary Uses Services (SUS) national administrative data, held by the tNR. SUS is a comprehensive repository for secondary healthcare data in England that is paid for by the NHS. It is used to support the NHS in the delivery of healthcare services and to trigger reimbursement for secondary care activity. We will request SUS data for the period 1 April 2012 to 30 August 2016. This period will cover the intervention period up to the latest point for which we will have access to reliable data, as well as the two years before the introduction of the interventions (needed for measuring study covariates). We believe that data to August 2016 will be of sufficiently good quality, but we will check that the quality of the data is similar to earlier months and if necessary shorten the follow-up period.

In addition, the tNR holds monthly extracts from NHAIS from August 2014 to September 2016. These monthly extracts, created on the first Sunday after the 13th of the month, contain demographic information about all registrations with general practices in England, including date of birth and full residential address. Addresses are updated in NHAIS when patients move general practice or when patients inform their GP of a change of address.

A manual review of unique address fields from NHAIS for Rushcliffe and the pool of comparator areas (without linked patient data) will enable the identification of care home addresses. The identified care home address fields will be used by the tNR to create a pseudonymised care home indicator, as well as to link to care home characteristics from the CQC data, for all care home residents.

Furthermore, these monthly extracts will enable the tNR to identify the month in which a person entered a care home (for cohort 2 only) and the month in which they left the care home. This will be done by analysing the history of address information in NHAIS, and identifying when it changed to or from the address of a care home. In addition, the date of death is also available in NHAIS for patients who died between August 2014 and September 2016. The tNR will derive and provide limited data from the NHAIS database, relating to a resident's month and year of birth, death, and move in or out of a care home. These data will be linked to the SUS data via a pseudonymised patient identifier. As there may be a time lag in information being updated, only data until the August 2016 extraction will be requested. We believe that data to August 2016 will be of sufficiently good quality, but we will check that the quality of the data is similar to earlier months. In particular, we can assess the quality of death data from NHAIS by checking if hospital deaths recorded in SUS are also recorded in NHAIS. If necessary, we will shorten the follow-up period.

Through the described method, we will be able to reliably identify the whole care home population from August 2014 to June 2016, even those residents who do not have any hospital admissions during the follow-up period. However, only for cohort 2 can we establish the date of entry into a care home and therefore a relatively accurate start date for the intervention or the equivalent control period (estimated as the date of the extraction in which the patient is first recorded as living at the care home address). For cohort 1, the start date is assumed to be the date of the August 2014 extraction date, 17 August 2014. Outcomes occurring between 1 April 2014 (when the enhanced support was implemented) and 16 August 2014 cannot be reliably classified as part of the intervention period or not.

Study endpoints

Primary outcome

The primary endpoint for both cohorts is the number of potentially avoidable hospital emergency (ie unplanned) admissions (6)[7] per resident over a period of up to 24 months. The length of the follow-up period will differ between residents, depending on their length of stay in the care home within the confines of the study period. For the main cohort (cohort 2), residents will have a follow-up period that starts from when they entered the care home; cohort 1 residents' follow-up period will start in August 2014. For both cohorts, the follow-up period will end when the resident leaves the care home or dies or the study period ends.

Emergency admissions are defined as separate hospital spells that either occur through the emergency room or as a result of direct, urgent referrals from a general practitioner or other professional. Potentially avoidable hospital admissions (also referred to as 'conditions amenable to early identification and intervention')[7,8] from care homes builds on the same notions as ambulatory care sensitive conditions, but is broader in scope and more targeted to older frail patients in care homes or community settings. It combines two separate concepts: conditions that should not occur, and conditions that should be manageable in the care home. The conditions include anaemia, congestive heart failure, hypertension, hypotension, hyper and hypoglycaemia diabetes with ketoacidosis or hyperosmolar coma, dehydration, acute renal failure, hypokalemia or hyponatremia, constipation, faecal impaction or obstipation, diarrhoea, C. difficile (may be preventable in some cases), gastroenteritis with nausea and vomiting, cellulitis, skin ulcers including pressure ulcers, lower respiratory: pneumonia or bronchitis, urinary tract infection, falls and trauma, altered mental status/acute confusion/delirium, psychosis or severe agitation, organic brain syndrome, COPD, asthma, chronic bronchitis, weight loss, nutritional deficiencies, adults failure to thrive, and seizures.[7,8] However, only ICD-9 codes for these outcomes are available from existing definitions of potentially avoidable hospital admissions. We will investigate translating the ICD-9 codes to ICD-10 codes and having the codes reviewed by a clinician. Alternatively, we will use the list of avoidable admission conditions in older frail people, used by CQC.[9,10] This list of conditions is however not specific to care home residents.

Secondary outcomes

The secondary endpoint is the proportion of total number of hospital bed days to resident bed days per resident over a period of up to 24 months. This endpoint reflects changes to the length of stay in hospital as well as the number of admissions.

A bed day is defined as a night in hospital following emergency or elective admission, but excluding 'regular day/night attendances' (determined by a specific code in the raw SUS data). This is consistent with the definitions of bed day used in NHS England[11] and the NHS England New Models of Care dashboard, which displays outcome data for all Vanguard sites.[12] An admission and discharge within the same day will therefore not count towards the total number of bed days.

Other endpoints will be (all calculated over a period of up to 24 months):

- number of emergency admissions per resident
- number of community-acquired pressure sores that result in a hospital admission per resident
- number of A&E attendances per resident
- number of elective admissions per resident
- number of outpatient attendances per resident (excluding 'did not attends', and defined using code Attended=5 or 6 in SUS)
- proportion of deaths outside hospital (*proxy for end-of-life care in place of choice)*

All of these endpoints will be modelled allowing for varying lengths of stay in the care home, by including an offset for amount of time at risk in the statistical analysis. As we anticipate high levels of attrition in this population, we will in addition perform all these analyses, apart from proportion of deaths outside of hospital, for a 6 and 12-month period respectively, for the subset of patients who were residents for at least that amount of time. We will conduct descriptive analysis of the length of stay in care homes and the reasons for attrition in the intervention and control groups. If we find marked differences that undermine our confidence in the main analysis, that is allowing for varying lengths of stay over 24 months using an offset, then the 12-month analysis will be considered the main analysis.

The proportion of deaths outside hospital will be evaluated as proxy for end-of-life care in place of choice. A significant proportion of older people prefer to die at home.[13,14] When patients are living in care homes, the care home is often considered home, especially among patients aged 85 or over.[14]

The number of outpatient attendances and elective admissions per resident could either increase or decrease with good care. Although interpretation may be difficult, these outcomes help create a fuller picture of how care home residents use secondary care, and they are therefore relevant endpoints.

Costing of secondary care could be done in future analyses.

Variable definitions

Exposure variables

A person is considered having received the intervention if they are resident in a Principia care home at any time during the follow-up period. These people will be identified according to whether their recorded address in the NHAIS monthly extracts matches a list of addresses of Principia care homes at any point between 17 August 2014 and 14 August 2016, which is the follow-up period for which we have reliable data to identify care home residents and their outcomes (see 'Variable definitions' in section 2).

Index dates and follow-up period

Although the enhanced support was introduced in April 2014, we will allow for a 'bedding-in' period to establish the enhanced support. As data from NHAIS, which enables the identification of persons residing in care homes, is only available from the first available extraction 17 August 2014, we will consider the bedding-in period to be approximately four and a half months (1 April 2014 to 16 August 2014).

A resident's follow-up period will start from the index date, which is defined as:

- Cohort 2: The date of the extraction in which they are known to first have a new care home address, if the resident moved after August 2014
- Cohort 1: 17 August 2014 if they were already resident in the care home at the time of the first available extraction

A resident's follow-up period ends when the resident either dies, moves away from the care home, or the study period ends. The end date is defined as the first of the following dates:

- 14 August 2016 if the resident still resides in the care home at that time
- date of death, estimated as the date of the extraction in which they are first recorded as dead
- date of move out of the care home, estimated as the date of the extraction in which they are first recorded with an address different from the care home they were residing in.

We estimate the date of death as the extraction date of the month in which they died. However, as the true death date could be as early as the day after the previous extraction date, there is a risk that we are overestimating the follow-up period by up to a month. However, as it unlikely that the day of the month that a person dies is other than random, this is unlikely to introduce bias between the intervention and control groups.

Pre-period

The baseline variables will be assembled using data recorded during the 'pre-period', which will encompass data at resident level going back two years. The definition of pre-period will differ between the two cohorts.

For cohort 2 – residents who moved into a care home later than August 2014 – the pre-period will consist of the two years before their index date.

For cohort 1 – residents who already resided in a care home in August 2014 – the pre-period will consist of 1 April 2012 to 31 March 2014. Thus, data from the bedding-in period 1 April 2014 to 16 August 2014 will be omitted for both Principia residents and their matched controls in cohort 1, as data cannot reliably be ascribed to either the intervention or pre-intervention phase.

Baseline variables

Both resident and care home-level baseline variables will be included in the matching and the regression model. All resident-level baseline variables are calculated on pre-period data.

Potential resident-level covariates are:

- approximate age at index date
- gender
- ethnicity
- number of emergency admissions in last year of the pre-period
- number of emergency admissions in first year of the pre-period
- number of potentially avoidable emergency admissions in last year of the pre-period
- number of potentially avoidable emergency admissions in first year of the pre-period
- number of elective admissions in last year of the pre-period
- number of A&E attendances in last year of the pre-period
- number of bed days in last year of the pre-period
- specific comorbidities linked to frailty (two-year lookback, see below)[15]
- other comorbidities predictive of hospital emergency admission, as identified in the Patients at Risk for Re-hospitalisation (PARR) algorithm[16]
- Charlson index (two-year lookback).[17]

Comorbidities linked to frailty are: anxiety or depression, functional dependence, falls and significant fracture, incontinence, mobility problems, pressure ulcers and cognitive impairment (composite of delirium, dementia and senility).[15] These will be defined using data from any hospital admission during the two-year pre-period. Although some of the frailty comorbidities can be cured, we will assume that any issues in the previous two years will be relevant to the overall frailty level of the resident and therefore also their risk of emergency admissions.

Comorbidities included in the PARR algorithm are: alcohol related diagnoses, cerebrovascular disease, chronic obstructive pulmonary disease, connective tissue disease/rheumatoid arthritis, developmental disability, diabetes, ischaemic heart disease, peripheral vascular disease, renal failure and sickle cell disease.[16]

For patients in cohort 2, additional covariates are:

- index dates, which allows for differences in, for example, care over time
- number of emergency admissions in the 60 days before the index date
- number of potentially avoidable emergency admissions in the 60 days before the index date.

Due to the limitations of our data, we estimate the date of arrival in a care home as the extraction date. However, as the true date of arrival could be as early as the day after the previous extraction date, the actual period of pre-care home admission in this 60-day period could differ by up to a month. However, as it is unlikely that the day of the month that a person enters a care home is other than random, this is unlikely to introduce bias between the intervention and control groups.

At care home level, potential covariates are:

- care home type: nursing or residential (CQC data)
- number of beds available (CQC data)
- care home age categories: whether the care home caters exclusively to adults aged 65 or over, or also to another age group (adults under 65 or whole population) (CQC data)
- urban/rural classification at lower layer super output area (LSOA) level, based on the 2011 census
- average socioeconomic deprivation deciles, based on the Index of Multiple Deprivation (IMD) 2015, available at LSOA level.

We will use CQC data from 1 April 2014 – the date the intervention was first implemented – for care homes that were open at that time. However, for care homes that opened later, we will use the CQC data from the month following the opening (when the data on that care home are first available). We will use IMD scores from 2015, as previous IMD scores were based on 2010 data and may no longer reflect the level of deprivation of the area. We think this approach is valid because the enhanced support is unlikely to affect IMD scores, at least over the time periods covered in this study.

Although we considered matching on the rate of hospital use measured at the care home level, unfortunately these data will not be available, as the lack of NHAIS date from before August 2014 excludes the identification of care home residents before that date. We do not have data on occupancy levels, what proportion of residents are funded by the local authority, or baseline nurse-to-bed ratio.

We will not use data on the care home inspection ratings assigned by the Care Quality Commission, as there is uncertainty around inter-rater reliability[18] and changes in scoring framework over time. In addition, the latest scoring may not be very recent and as the purpose of the CQC audits is to engender change, older scores may not reasonably reflect the care in the care homes in March 2014.

3. Statistical methods

Identifying control areas

Before performing matching at care home and resident level, we will identify other local authorities in England that are similar in terms of demographics, socioeconomic characteristics and other outcomes of interest in the pre-period.[19] We will use the established methodology used by the Office for National Statistics[20,21] for determining areas of England of comparable health and area classification, which calculates the squared Euclidean distance as a measure of similarity between areas. Its methodology includes a wide range of variables (including age structure, ethnic mix, education, employment rates, overall rates of long-term illness)[21] to which we will add variables of importance to this study that are publicly available at aggregate (local authority) level. In particular, we will try to find areas with similar proportions of care home residents that are funded by the local authority rather than privately.

The calculated distances, as well as the transformed and standardised census variables, as described in the methodological note,[21] are publicly available, but as they are calculated based on UK data they will need to be recalculated using data for England only. For the other variables, we will use the same method of data transformation and standardisation, that is the inverse hyperbolic sine method and inter-decile range standardisation.

A list of 10 potential local authorities with similar characteristics and their comparative values will be reviewed by the Principia Vanguard team, thereby allowing use of local/specialised knowledge. Any local authorities with known area-wide interventions relating to care homes deemed to differ substantially from 'standard care' will be excluded from the list. Where similar values of the squared Euclidean distance are available for several local authorities, preference will be given to areas in close proximity to Rushcliffe, as controls from geographically nearby local authorities may benefit from, for example, same hospital services or joint hospital transformation projects. The final list of six areas which will form the pool of potential matched controls will be agreed between the Improvement Analytics Unit (IAU) team and Principia.

Identifying control population

The control population will be determined using matching methods, to optimise the similarity of Principia care home resident and matched control groups with respect to variables that are likely to be predictive of any of the outcomes.

Matching will be done separately for cohorts 1 and 2.

Matched control observations will be selected using the genetic matching algorithm, which is a computer-intensive search procedure that produces more closely balanced groups than traditional approaches such as nearest neighbour matching or the propensity score.[22] If possible, matching will be done without replacement. We will match 1:1, that is each intervention resident to one control resident.

The genetic matching algorithm will try various distance functions to determine the 'closeness' of the match. However, for some variables we require an exact match or 'calliper' match where the variables are required to be within a fixed distance of one another. Table 1 shows those variables where special matching methods are to be applied.

Table 1: Matching variables and method of matching

Variable	Method of matching
Care home type	Exact
Care home age categories	Exact

We hope to match all Principia care home residents to controls. As we are not matching on length of stay in the care homes (as this could be correlated with quality of care), the intervention and their control residents may differ in this respect. We will conduct descriptive

analysis on care home length of stay, and address differences using an offset in the model that will allow for differences in time at risk.

Choice of matching variables

We will match on both resident and care home characteristics, to find similar resident case mix and similar care home types, such as whether the homes offer nursing support and the size of the home. Although we will ultimately assess balance for all of the variables detailed in 'Baseline variables' in section 2, we will only include a subset of those variables in the matching algorithm. We will adapt the subset of variables included in the matching algorithm to optimise balance between the two groups on those variables considered most strongly predictive of the outcome, for example the prior numbers of emergency admissions, but also aiming to optimise balance across the wider set of variables.

Matching will be done on both care home and resident characteristics at the same time, rather than matching first on care home and then resident, as this will enable a larger pool of potential controls, as all matching residents from one intervention care home would not need to be found within the same control care home.

Matching parameters

Table 2 gives the matching algorithm parameters to be used for this analysis.

Table 2: Matching parameters

Parameter	Value
Estimand	ATT
Number of controls per intervention patient	1
Control sampling method	Without replacement
Population size	2,000
Maximum generation	1,000
Number of generations to wait	100

Diagnostics

Balance will be assessed separately for cohorts 1 and 2. It will be assessed across all of the baseline variables listed in 'Baseline variables' in section 2, even though not all variables will be included in the matching algorithm.

Balance will be assessed using the standardised difference, which is defined as the difference in means as a proportion of the pooled standard deviation.[23] Although the

standardised difference should ideally be minimised without limit, a standardised difference below 10% has been used to describe negligible imbalance.[24] The standardised difference is a better measure of balance than formal statistical tests, as the latter depend on the size of the groups, as well as on the level of similarity.[25]

Statistical analysis

We aim to estimate the average treatment effect for the treated. Thus, once matched controls have been selected, we will estimate the effect of the Principia enhanced support compared with the control group by fitting hierarchical mixed-model regression models, both unadjusted and adjusted for covariates. The adjusted model will contain all variables that were used in the matching process, to adjust for any remaining observed imbalance, as well as any other covariates predictive of outcome.

As previously described, cohorts 1 and 2 will be analysed separately.

For the primary endpoint (number of potentially avoidable emergency admissions per resident) a Poisson regression will be used to estimate the rate ratio that compares the endpoint between intervention residents and matched control residents. We will also estimate the effect of the intervention on the absolute (as opposed to relative) number of admissions. The Poisson assumption will be assessed by examining over dispersion statistics, namely, the ratio of the deviance statistic to the residual degrees of freedom. If over-dispersion is detected then an alternative count model, such as the zero-inflated Poisson or Negative Binomial, will be fit.

For the secondary endpoint, the proportion of total number of hospital bed days to resident bed days per resident will be estimated using linear regression. However, we shall examine if a negative binomial model is a more suitable fit as has been suggested in recent literature.

Each of the other endpoints will be similarly examined for regression models appropriate to their distributional properties. As a guide Table 3 details the typical regression models and alternatives for each outcome.

Table 3: Regression models for each outcome

Outcome	Initial model	Alternative	Diagnostics
Number of potentially avoidable hospital admissions/emergency admissions/elective admissions/A&E attendances	Poisson	Negative binomial or Zero-inflated Poisson	Over dispersion Model fit
Number of hospital bed days	Linear regression (ordinary least squares regression/OLS)	Negative binomial	Model fit Distribution of model residuals Transformation of dependent variable Heteroscedasticity diagnostics

To account for attrition arising from death or moving away from the care home, an offset of the number of resident bed days will be added to the model. However, the offset assumes that the number of days that are 'missing' is random and that the rate of outcomes, such as emergency admissions, is constant, when in fact this is unlikely to be the case, for example, as residents may use more hospital services just before dying. We will therefore also perform all analyses with a 6 and 12-month endpoint on the subset of residents who have a care home stay of at least these periods of time, respectively.

Although no care home-level analyses will be done, the model will control for clustering within care homes, by including a random effect at care home level.

The matching may induce a correlation between the outcomes of patients in a matched pair. We may account for this potential correlation by adjusting the standard errors using either a robust estimator of the variance or by modelling the correlation directly via a mixed model in the case of continuous, approximately Gaussian outcomes, if appropriate.

Subgroup analysis

No results will be presented at care home level, as this would potentially jeopardise patient confidentiality. As some of the care homes are small, patients may have been identifiable based on their characteristics. However, we will conduct subgroup analysis according to whether a care home is nursing or residential, and assess the balance within each type of care home separately. No other subgroup analysis is planned.

Sensitivity analyses

One of the main threats to the validity of this study is unobserved confounding. That is, although we anticipate that the intervention and matched control groups will be similar in terms of observed variables (such as age and prior number of hospital admissions), there may be differences between these groups that we do not observe, for example aspects of the quality of care people received before entering the care home. However, the prior number of avoidable admissions to some extent measures the quality of care before admission to the care home.

Although there is no definitive way to assess the effect of unobserved confounding, we can compare the rates of the Principia and matched control groups on an endpoint unrelated to the intervention.[26] On the assumption that the enhanced support is unlikely to have had a large positive or negative impact on overall mortality within the follow-up period, then differences in mortality rates would make us doubt the performance of the matching. For example, if enrolled patients died at a higher rate than matched control patients, this might suggest that they were in worse health than controls at the point of enrolment.[27] However, there is also a possibility that good care may result in prolonged life and therefore fewer deaths during our follow-up period. We will therefore compare the rates of all-cause mortality over a period of up to 24 months. Similar mortality rates will be indicative of balanced groups, while differences may need to be interpreted with caution.

We may also assess sensitivity by calculating the Rosenbaum bounds for a binary outcome for the primary endpoint (whether a resident had at least one potentially avoidable admission in a 12 month period).[28,29,30,31]

Sample size calculation

No sample size was calculated for this study. The analysis is considered informative and will be carried out regardless of whether our study population is of a sufficient size to detect a statistically significant difference.

4. Limitations and sources of bias

Threats to validity

One of the main threats to the validity of this study is unobserved confounding. In our case, this can occur at area, care home or resident level. Geographically local controls are normally preferred for matching, as this minimises the risk of differences in context that might impact on outcomes, but local controls in the CCG or local authority are not possible for this study, as the Principia intervention encompasses the whole CCG. However, if there was a choice between several local authorities with comparable levels of similarity to Rushcliffe, we considered giving preference to local authorities in the geographical vicinity of Rushcliffe. In this study we are matching on a range of observed variables (such as age and prior number of hospital admissions, whether in a nursing or residential home), but there

may be differences between these groups that we do not observe (such as level of need, either at an individual or care home level) and that might contribute to their outcomes. In particular, we are not able to match on care home trends in the pre-period, due to the unavailability of data to determine care home residents before August 2014. Our findings might be biased if there are differences that we cannot account for; the potential impact of this bias will be assessed within the sensitivity analysis described earlier.

Principia has been implementing improvements to health care since before 2014. Matching on patient characteristics, including comorbidities, will hopefully ensure that the comparison groups are similar. Yet there is a risk that in cohort 1 – where residents may have been in a Principia care home for a long time – baseline variables such as number of previous emergency admissions may have been lower than would have been expected due to superior care. This patient may therefore be matched with a 'healthier' resident in a control care home.

Comorbidities are calculated from (primary and secondary) diagnosis fields in the SUS data. Although the study benefits from a single, national database of secondary care activity, in practice there are differences in coding and coding depth between hospitals[32] that may bias the detection of comorbidities.[33] Furthermore, there may be other differences in the characteristics of acute trusts that we are not able to allow for, such as in mortality or admission thresholds across different hospitals.

A limitation specific to cohort 1 is that we cannot identify the date of entry into a care home. This means that we cannot determine the date that residents in the intervention group started receiving enhanced support. We will therefore use the earliest date at which we are certain they were receiving the enhanced support, August 2014, as the index date. As a result, we cannot match on the real date of entry; we cannot determine baseline covariates using data from the full two years before their index date, or in the 30 days before the index date, as the Principia group may have already been receiving the intervention. In addition, as Principia has a history of engaging with central transformation initiatives, there is also a risk that long-term residents in Principia who entered pre-April 2014 received care that was different to 'standard care' even before the enhanced support was introduced.

There is a risk in cohort 2 that the Principia enhanced support could lead to care home selection bias. As the enhanced support becomes better known and positive results are announced, self-paying residents may choose Principia care homes over other care homes. Already the positive results of a local evaluation are being disseminated (www.rushcliffeccg.nhs.uk/principia-mcp-vanguard). However, it is likely that people tend to choose a care home in the vicinity of their home or their family. As the enhanced support was implemented across the whole borough, this kind of selection bias is unlikely to be a large issue, especially at this early phase of the implementation of the enhanced support.

As the results of the Principia enhanced support are disseminated, other care homes may copy some of the Principia interventions, biasing the results and making future evaluations of either Principia or other care home Vanguard sites more difficult. Again, this is unlikely within the time period of this study.

There may be differences between areas in availability of palliative care outside of the care homes.

The use of control groups is intended to isolate the impact of a specific set of interventions, and we are not aware of any other changes that occurred within Principia or the pool of comparison areas during the follow-up period. However, the endpoints may be affected by other service changes happening during the follow-up period, such as if a step-down care centre was opened, or other changes that we were not aware of.

Care home residents are identified based on primary care administrative data, depending on its data quality, timeliness and variation across different general practices. As addresses are updated in NHAIS when patients move general practice or inform their GP of a change of address, and the intervention encourages residents to change GP on entry to care homes, the address data are more likely to be updated in Rushcliffe than the other areas.

Care home characteristics were taken from CQC data. These data are not designed for research purposes and are not validated accordingly. However, we did some limited in-house validation of the CQC data by comparing it with data on Principia care homes supplied directly from the Vanguard and found that the CQC data were of sufficiently good quality to use in the manner described in our analyses (see 'Validation of care home information' in section 6).

SUS data is an administrative database and has not been subjected to the cleaning rules that Hospital Episode Statistics are. However, the IAU Data Management Team will perform data checks and cleaning.

Since we will use data that are linked based on full residential address, care home resident identification will be precise in most cases. However, this assumes that the addresses are updated in a timely manner in NHAIS. In addition, if there are staff living on the premises of the care home who are aged 65 or over, then those staff will be wrongly categorised as part of the study cohort. We are not aware that this is the case for any of the intervention care homes.

Index and end dates are only approximated for two reasons: we only have access to the monthly table extracts from August 2014 onwards and there is no date of when between the monthly snapshots the change occurred. This has several implications: Firstly, the offset period for each resident – the number of resident bed days – will only be approximate. Secondly, residents before August 2014 will not be able to be matched based on when they entered the care home. This will be less of an issue should we repeat the analyses in the future and focus more on new entrants.

The number of resident bed days, which will be used to determine the offset, will just be approximate, as we only have information on the month they moved in or out of a care home (imputed as the date of data extraction).

Furthermore, due to limitations to the scope of address field cleaning that will be done for this pilot study, we are not able to determine if the index date marked the start of the patient's first care home stay or if they moved from another care home (unless it was from another care home in the study).

Other limitations

We will not be able to evaluate other potential impacts of the enhanced support, such as quality of life or improvement in working relationships, as we only had access to secondary care data. Costing secondary care data is out of scope for this study.

5. Reporting

General reporting considerations

Results will be reported as the relevant measure of effect, such as odds or rate ratios, plus 95% confidence intervals and p values. Both the post-matching unadjusted and adjusted analysis will be presented and the variables used in the adjustment noted. Results will be presented to two decimal places for effect size and confidence intervals. P-values will be shown to two significant digits.

Special reporting requirements for this study

At a minimum the following are requirements for this study:

- adherence to the STROBE[34] and RECORD statement guidelines[35]
- adherence to the NHS Digital (previously Health and Social Care Information Centre, HSCIC) small number rules[36]
- compliance with the statistical code of practice.

6. Appendices

Cleaning

Data checks

Standard descriptive programs will generate a summary report of the distributions of individual data elements. These will be examined for missing values, outliers and erroneous values.

Missing values

Although we expect most of the study variables will not have missing data, the extent of missing values will be examined. In most cases a complete case analysis will be performed unless the importance of the missing data requires the use of an imputation method. In any event sensitivity analyses will be performed to understand the impact of missing values.

Outliers

Data outside the range of generally plausible values will be examined for erroneous values. A sensitivity analysis will be performed if the impact of the outliers is suspected to affect the results of the analysis.

Validation

Validation of care home information

Care home characteristics, such as whether the care home is nursing or residential, number of beds, full address and postcode and care home specialties, such as catering for older people, are readily available on the CQC website and will be used to match at care home level.

We checked the validity of the CQC data by comparing it with data on Principia care homes supplied directly from the Vanguard. The results show that the CQC data on variables of interest are of sufficiently good quality to use in our analysis.

Data flow diagram

Tables and figures for reporting matching results

Tables

A baseline table showing descriptive statistics for the intervention group and the matched and unmatched control populations, with:

1.0 continuous variables summarised by mean (SD) or median (IQR) depending on the distribution

2.0 categorical variables summarised by number (%)

3.0 standardised differences calculated for the intervention group versus the unmatched and matched control groups, and variance ratio for continuous variables.

Figures

The following figures would be a minimal requirement:

1.0 dot plot showing the standardised differences from both the matched and unmatched sample

2.0 histograms illustrating the baseline characteristics of the intervention and matched control groups in more detail (potentially with banding to ensure minimum cell counts are above 10).

Tables and figures for reporting statistical results

Tables

The following tables would be a minimal requirement:

1.0 A table showing the unadjusted estimates of treatment effect for the intervention and matched control groups:

1.1 for binary outcomes the number and proportion in each group
1.2 for count data the number of events and person time of exposure
1.3 for continuous data the mean and standard error
1.4 the size of the measure effect (eg odds ratio, rate ratio, hazard ratio or mean difference) together with a 95% confidence interval
1.5 for a difference-in-difference type analysis the table should show summary results in each time period, their difference and the difference between groups over time.

2.0 A table showing the adjusted results:

2.1 the size of the adjusted measure together with a 95% confidence interval and p-value
2.2 all adjustment variable listed and in some cases included in the table with the relevant effect sizes and 95% confidence intervals.

Figures

The following figures would be a minimal requirement:

1.0 forest plot showing the crude and adjusted results for each outcome measure.

7. Addendum: Main analysis

This section has been added to the original statistical analysis protocol (SAP) document to provide further clarifications or information on modifications to the SAP for the evaluation published in March 2017. The original SAP was agreed in October 2016 with this section added in November 2018.

Study design in section 2, page 7

Only the main analysis, on residents who entered a care home after August 2014 (ie cohort 2), was performed. We decided to not carry out the secondary analysis on cohort 1 as it was

not possible to determine the date of moving to the care home and we judged that we could therefore not reliably find a similar matched control group for this cohort. Prior hospital activity is predictive of future activity, but hospital activity is likely to differ depending on whether a person lives at home or in a care home and may also change depending on length of time in a care home. As the date of moving to the care home was unknown for residents who moved before August 2014, it was not possible to differentiate between prior hospital activity occurring before entering the care home and while resident in a care home (see 'Threats to validity' in section 4). Also, if care was better in Principia care homes before the introduction of the enhanced support, then this would have introduced unobserved confounding.

Study design in section 2, page 7; Statistical analyses in section 3, page 18

The analyses were performed using multivariable regression models, with no random effect at care home level. No clustering at care home level was performed.

Study cohorts in section 2, page 7

There were 20 Principia and 23 matched control residents who did not have an estimated follow-up as they moved to and either moved out or died in the same extraction month. These residents were included in the matching process and are therefore included in the baseline tables but were omitted from the analysis.

Sources of data in section 2, page 10

Due to circumstances outside the control of this project, no monthly extract of NHAIS was created in December 2014. Due to the nature of the administrative data, this extract could not be recreated in retrospect. Therefore, the extract was not included in the creation of the dataset. Changes in status of care home residents (ie moving into or out of a care home) after 16 November 2014 and on or before 18 January 2015 were all included in the January extract. Therefore, these residents were considered as having moved on 18 January 2015. Deaths are recorded separately, therefore deaths occurring after 16 November 2014 and on or before 14 December were dated as 14 December.

Study endpoints in section 2, page 11

Here we provide further clarification on how variables were defined.

For potentially avoidable emergency admissions, the definition used by CQC[10,10] was used.

An elective admission was defined as a planned admission, excluding maternity admissions, admissions due to births and transfers from other hospital providers other than in an emergency (admission codes 31,32, 81–83). However, regular day/night attendances were not excluded.

A bed day was defined as a night in hospital following an emergency or elective admission. Consistent with the definition above, elective admissions did not exclude 'regular day/night attendances'.

Study endpoints in section 2, page 11; Statistical analysis in section 3, page 18

Due to a combination of too few residents in the subgroup of people who were in the study for 6 or 12 months and too few events during those periods, it was not possible to model the outcomes in the sensitivity analyses for fixed 6 and 12-month periods. As the intervention

and matched control groups in the main analysis were similar in length of stay and reasons for attrition, the main analysis using an offset was considered robust.

Variable definitions: Index dates and follow-up period in section 2, page 12

Following quality assurance, it transpired that 17 residents (11 Principia and 6 matched comparison residents) had an attributed end-of-study date of 18 September 2016 instead of 14 August 2016. For these residents, outcomes were recorded until 31 August 2016 but the study length was computed until 18 September 2016. As a result, the study length for these residents was overestimated by 2 weeks. Due to both the minor difference in study length and the small number of residents affected, no action was taken.

Variable definitions: Baseline variables in section 2, page 15

Comorbidities predictive of hospital admission were originally to be based on the 2006 PARR paper identifying patients at high risk of readmission to hospital within the next 12 months.[16] However, instead, the variables identified in the 2012 updated PARR-30 were used, which are most predictive of 30-day emergency admission.[37] This was done before any data analysis was performed.

The number of emergency admissions and potentially avoidable emergency admissions, respectively, in the 60 days before the index date were not included in the matching or regression, as we considered that this was too narrow a window. This is because the estimated index dates were based on the extract date following the move and could include several weeks of post-baseline data.

One of the conditions defined in potentially avoidable admissions, 'food and drink issues', was not picked up due to a coding issue. This affected a small number of records (<5) in the pre-period. This did not affect the outcome variable potentially avoidable admissions.

Ethnicity was dichotomised as white/non-white.

Matching parameters in section 2, page 17

Matching was done with replacement.

Diagnostics in section 2, page 17

The standardised difference for a given variable was defined as the difference in means between the intervention and control groups, expressed as a proportion of the variable's standard deviation in the intervention group.[38]

Statistical analyses in section 3, page 18

In addition to modelling all baseline variables as covariates, models also including interaction terms for age*gender and age squared were included in the modelling options for count variables.

Adjusted absolute differences were not presented in the briefing.

For hospital bed days, expressed as the proportion of total number of hospital bed days relative to days in follow-up per resident, the following models were explored: Poisson, binomial and quasi-binomial models.

8. Addendum: Care home type subgroup analysis

This section has been added to the original statistical analysis protocol (SAP) document to provide further clarifications and information on modifications to the SAP specific to the care home type subgroup analysis. Unless otherwise stated, the changes made to the main analysis (see 'Addendum: Main analysis' in section 7) also refer to the subgroup analysis. The original SAP was agreed in October 2016 with this section added in November 2018.

Study cohorts in section 2, page 7

There were 20 Principia and 23 matched control residents who did not have an estimated follow-up as they moved to and either moved out or died in the same extraction month. These patients were excluded from the study (an additional exclusion criteria).

Variable definitions: Index dates and follow-up period in section 2, page 13

17 residents (11 Principia and 6 matched comparison residents) originally had an attributed end-of-study date of 18 September 2016 instead of 14 August 2016. For these residents, outcomes were recorded until 31 August 2016. The end-of-study date for these residents was set as 31 August 2016 and the study length was computed accordingly.

Variable definitions: Baseline variables in section 2, page 14

Instead of Index of Multiple Deprivation (IMD) deciles, quintiles were used in the descriptive tables and in the matching.

Identifying control population in section 3, page 16

Matching was done separately for nursing and residential care homes.

Identifying control population, choice of matching variables in section 3, page 16

Matching was tried using all baseline variables and the subset of variables considered 'core', respectively. The pre-specified core variables included age, gender, 2015 IMD quintiles, Charlson Index, number of frailty conditions, all hospital activity in previous year, and care home type. The matched dataset with best overall balance was from matching on all baseline variables for the residential subgroup and matching on core variables in the nursing subgroup. Matching was exact on care home age categories in both subgroups, as specified in the original SAP.

Statistical analysis in section 3, page 18

As well as modelling all baseline variables as covariates, modelling options included adjusting for a reduced set of variables, as the low number of events for some of the outcomes may otherwise have led to over-parametrised models. Explored modelling options consisted of including those variables that were either pre-specified 'core' variables (see section just above) or where there was imbalance between the groups post matching; those variables most predictive of a given outcome (using LASSO regression);[39] and the combined set of variables in the last two groups. Due to data sparsity in lower IMD quintiles (Table 2), these were grouped into three IMD grouping categories: quintiles 1 to 3, 4 and 5 in the regression models.

The hospital bed days outcome was changed in the subgroup analysis from the proportion of the total number of hospital bed days relative to days in follow-up per resident in the original analysis, to the number of hospital bed days with an offset to account for the length of follow-

up. Explored models included Poisson, quasi-Poisson, negative binomial regression and hurdle models.

A.4 Tables and figures for reporting matching results, page 24

For the journal manuscript, some tables and figures were omitted, either due to smaller sample sizes and therefore a higher risk of disclosive data from triangulating data from tables, or because the same data are presented in table format.

References

[1] Lewis R, Rosen R, Goodwin N, Dixon J. Where next for integrated care organisations in the English NHS? The Nuffield Trust. 2010. Available from: http://www.nuffieldtrust.org.uk/sites/files/nuffield/publication/where_next_for_integrated_care_organisations_in_the_english_nhs_230310.pdf
[2] NHS. The Multi-Speciality Community Provider (MCP) emerging care model and contract framework. 2016.
[3] NHS England. Spreading new care models. 2017. Available from: https://www.england.nhs.uk/ourwork/futurenhs/new-care-models
[4] Gladman J, Masud T, Sahota O et al. East Midlands Research into Ageing Network (EMRAN) Health care for older people research in Nottingham and Derby. 2015.
[5] Forder J, Fernandez J. Length of stay in care homes. 2011.
[6] Steventon A, Roberts A. Estimating length of stay in publicly-funded residential and nursing care homes: a retrospective analysis using linked administrative data sets. BMC Health Services Research. BMC Health Services Research; 2012;12(1):377. Available from: BMC Health Services Research.
[7] Walsh EG, Freiman M, Haber S et al. Cost drivers for dually eligible beneficiaries: potentially avoidable hospitalizations from nursing facility, skilled nursing facility, and home and community-based services waiver programs. Centers for Medicare and Medicaid Services. 2010;(0209853). Available from: https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Reports/downloads/costdriverstask2.pdf
[8] Walsh EG, Wiener JM, Haber S et al. Potentially avoidable hospitalizations of dually eligible medicare and medicaid beneficiaries from nursing facility and home- and community-based services waiver programs. Journal of the American Geriatrics Society. 2012;60(5):821–9.
[9] Care Quality Commission. The state of health care and adult social care in England in 2012/13. 2013. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21727839
[10] Care Quality Commission. The state of health care and adult social care in England. Technical Annex 1: Avoidable admissions. 2013.
[11] Department of Health. KH02 Quarterly bed availability and occupancy – data definitions. 2010. Available from: https://www.england.nhs.uk/statistics/statistical-work-areas/bed-availability-and-occupancy
[12] Operational Research and Evaluation Unit, NHS England. Evaluation strategy for new care model vanguards. 2016.
[13] Ahearn DJ, Nidh N, Kallat A, Adenwala Y, Varman S. Offering older hospitalised patients the choice to die in their preferred place. Postgraduate medical journal. 2013;89(1047):20–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23144224
[14] Abarshi E, Echteld MA, van den Block L et al. The oldest old and GP end-of-life care in the dutch community: A nationwide study. Age and Ageing. 2010;39(6):716–22.
[15] Soong J, Poots AJ, Scott S et al. Quantifying the prevalence of frailty in English hospitals. BMJ Open. 2015;5(10). Available from: N&AN=2015485395
[16] Billings J, Dixon J, Mijanovich T, Wennberg D. Case finding for patients at risk of readmission to hospital: development of algorithm to identify high risk patients. BMJ. 2006;333(7563):327.
[17] Quan H, Li B, Couris CM et al. Updating and validating the charlson comorbidity index and score for risk adjustment in hospital discharge abstracts using data from 6 countries. American Journal of Epidemiology. 2011;173(6):676–82.
[18] Walshe K, Addicott R, Boyd A, Robertson R, Ross S. Evaluating the Care Quality Commission's acute hospital regulatory model: final report. The Kings Fund. 2014.
[19] Steventon A, Grieve R, Sekhon JS. A comparison of alternative strategies for choosing control populations in observational studies. Health Services and Outcomes Research Methodology. 2015.

[20] Office for National Statistics. Methods for National Statistics 2001 area classification for health areas. 2001.
[21] Office for National Statistics. Methodology note for the 2011 area classification for local authorities. 2015. Available from: https://www.ons.gov.uk/methodology/geography/geographicalproducts/areaclassifications/2011areaclassifications/methodologyandvariables
[22] Sekhon JS, Grieve RD. A matching method for improving covariate balance in cost-effectiveness analyses. Health Economics. 2012;21(6):695–714.
[23] Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. Statistics in Medicine. 2009;28(25):3083–107.
[24] Normand S-LT, Landrum MB, Guadagnoli E et al. Validating recommendations for coronary angiography following acute myocardial infarction in the elderly: a matched analysis using propensity scores. Journal of Clinical Epidemiology. 2001;(54):387–98.
[25] Imai K, King G, Stuart EA. Misunderstandings between experimentalists and observationalists about causal inference. Journal of the Royal Statistical Society Series A: Statistics in Society. 2008;171:481–502.
[26] West SG, Thoemmes F. Campbell's and Rubin's perspectives on causal inference. Psychological methods. 2010;15(1):18–37.
[27] Steventon A, Bardsley M, Billings J, Georghiou T, Lewis G. An evaluation of the impact of interventions on hospital use. 2011.
[28] Dorie V, Harada M, Carnegie B, Hill J. A flexible, interpretable framework for assessing sensitivity to unmeasured confounding. Statistics in Medicine. 2016;35:3453–70.
[29] Rosenbaum PR, Rubin DB. Assessing Sensitivtiy to an Unobserved Binary Covariate in an Observational Study with Binary Outcome. Journal of the Royal Statistical Society, Series B. 1983;45(2):212–8.
[30] Rosenbaum PR. Design sensitivity in observational studies. Biometrika. 2004;91(1):153–164. Available from: https://academic.oup.com/biomet/article/91/1/153/218841
[31] Keele MLJ. Package 'rbounds'. 2015.
[32] Bottle A, Aylin P. Comorbidity scores for administrative data benefited from adaptation to local coding and diagnostic practices. Journal of Clinical Epidemiology. 2011;64:1426–33. Available from: http://dx.doi.org/10.1016/j.jclinepi.2011.04.004
[33] Wennberg JE, Staiger DO, Sharp SM et al. Observational intensity bias associated with illness adjustment: cross sectional analysis of insurance claims. BMJ. 2013;346:f549. Available from: http://www.bmj.com/content/346/bmj.f549
[34] STROBE. STROBE Statement – checklist of items that should be included in reports of observational studies. 2014. Available from: http://www.strobe-statement.org/index.php?id=strobe-home
[35] Benchimol EI, Smeeth L, Guttmann A et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. PLOS Medicine. 2015;12(10):e1001885.
[36] The Health and Social Care Information Centre. Hosptial Episode Statistics (HES) analysis guide. 2015. Available from: http://digital.nhs.uk/media/1592/HES-analysis-guide/pdf/HES_Analysis_Guide_Jan_2014.pdf
[37] Billings J, Blunt I, Steventon A et al. Development of a predictive model to identify inpatients at risk of re-admission within 30 days of discharge (PARR-30). BMJ Open. 2012;2(4):e001667–e001667. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22885591\nhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3425907
[38] Stuart EA. Matching methods for causal inference: A review and a look forward. Stat Sci. 2010;25(1):1–21.

[39] Friedman J, Hastie T, Tibshirani R et al. Package 'glmnet '. 2018. Available from: https://cran.r-project.org/web/packages/glmnet/glmnet.pdf