Statistical analysis protocol for the evaluation of Integrated Care Programmes in the North East Hampshire and Farnham and Fylde Coast CCGs

Paris Pariza, Therese Lloyd, Stefano Conti

April 2020

The Health Foundation

Contents

Summary

Purpose of this document

This statistical analysis protocol (SAP) sets out the broad aims and study design of three proposed Improvement Analytics Unit (IAU) quantitative analyses of Integrated Care Teams (ICTs) and concurrent NHS New Care Models vanguard (henceforth vanguard) initiatives, using the statistical Generalised Synthetic Control (GSynth) method.

Purpose of this evaluation

The proposed quantitative analyses build on three ICT evaluations already undertaken by the IAU: the North East Hampshire & Farnham (NEHF) vanguard[1] and the Extensive Care Service (ECS)[2] and Enhanced Primary Care (EPC) programmes, both in Fylde Coast[2]. These looked at individuals with long-term conditions and complex care needs, primarily of 65 years of age or over, who were referred to community care delivered by ICTs. These ICTs were set up to support local service and clinical integration strategies, usually led by the CCG, and included shared whole-system budgets or other organisational and governance incentives for multidisciplinary team collaboration. The interventions took place in three CCGs – Blackpool and Fylde and Wyre CCGs (collectively covering the Fylde Coast area) and NEHF CCG – with diverse demographics, rurality, social and health deprivation profiles that are broadly representative for the whole of England.

For the new analyses we will adopt a population approach to provide an overall picture of the effect of the different ICTs on health care for the local population of 65 years of age or over of the three geographical areas (Blackpool, Fylde and Wyre and NEHF CCGs). For each of the three geographical areas (the 'treated' areas), we will examine hospital resource use across the population of 65 years of age or over and compare it with that of a similar population in a 'synthetic control' area. The control areas are anticipated to provide an estimate of the hospital use that would have been expected in the treated areas in the absence of the ICTs or other vanguard initiatives. This will allow us to examine whether the effect of initiatives aimed at the population of 65 years of age or over – of which the ICTs are the main intervention – change over time and test the hypothesis that initiatives such as ICTs reduce emergency hospital use over longer timelines.

Furthermore, given that many of the vanguard initiatives were not limited to the older population, as a secondary analysis we will also evaluate the effect of the combined vanguard initiatives on the population of 18 years of age or over, in order to provide an overall picture of the effect of the initiatives over time.

What the evaluation will look at

We will conduct a rigorous empirical evaluation using the GSynth method to identify whether ICTs and concurrent vanguard initiatives impact on hospital use over an extended period of time (approx 6.5 years). For each of the three CCGs, we will compare the change in outcomes for patients of 65 years of age or over registered with a GP practice with that of patients of 65 years of age or over registered with selected GP practices in comparable CCGs from other parts of England.

A common finding of both peer reviewed and grey literature is that programme evaluations have often been informed by insufficient follow-up times and/or patient sample sizes to draw useful conclusions about the effect of interventions such as ICTs.[3] The IAU will use its pre-existing secure data environment infrastructure and assets – notably monthly Secondary Use Service (SUS) extracts from NHS Digital – and expertise in rigorous comparative effectiveness methods to analyse routine administrative data in order to explore whether the effect of the examined initiatives on hospital activity changes over a longer time. Although the evaluation will not be able to disentangle the effect of the individual components of the vanguard initiatives, as the ICTs were the main initiative targeting older frail patients the overall effect on the population of 65 years of age or over can be plausibly expected to be indicative of the effect of ICTs.

Data sources

The evaluation will use de-identified patient-level national SUS administrative hospital data for England from 2 years (24 months) prior to the start of each of the ICTs until the end of September 2019, as well as publicly available reference data. De-identified means that all direct identifiers (including, but not limited to name, address, date of birth and NHS patient number) are removed from the data. This reduces the risk of patient re-identification from the data.

Strengths and weaknesses of the evaluation

The evaluation will only study the impact of each of the CCGs' vanguard initiatives on secondary care activity as measured from national administrative hospital (SUS) data. Use of other health and care services, cost elements, impacts on quality of life, staff satisfaction, the quality of working relationships and other outcome areas that the vanguard has evaluated fall outside the scope of this analysis.

We are aiming to explore the effect of ICTs by evaluating the effect of a range of initiatives on the entire population of 65 years of age or over. The ICTs – and other initiatives – will have only cared for a small proportion of the population of 65 years of age or over, therefore the analysis may not be sensitive enough to account for the true effect of the initiatives.

The analyses on the population of 65 years of age or over aim to capture the effect of the ICTs, as the ICTs were the main initiative targeted at older frail people implemented in the three CCGs. The previous IAU analyses on patients referred to ICTs found that the mean age of patients referred to ICTs in Fylde Coast and NEHF was between 76 and 81 years. However, the treated population of 65 years of age or over will have benefited from other vanguard initiatives as well, therefore the findings will not be reflective of the ICTs alone. We will also evaluate in a secondary analysis the effect of the ICTs and other vanguard initiatives on the broader adult population (of 18 years of age or over); for this cohort, the analysis will likely less reflect the effect of ICTs over other initiatives.

For a full picture of the impact of the ICTs in the three treated CCGs, the evaluation should be viewed in conjunction with any qualitative research carried out by the local sites as part of the NHS New Care Models programme, as well as the prior IAU analyses on referred patients.

In order to assess the impact of GP practices in the three treated CCGs, our analysis will construct a counterfactual using the GSynth method. This method attempts to create a synthetic control for each of the treated CCGs by comparing GP practices in a treated CCG with similar practices elsewhere in the country. We identify similar GP practices based on both CCG and GP practice level characteristics, while excluding practices in the Greater London area and those in CCGs granted NHS New Care Home Vanguard Model status, as the level of care they provide is unlikely to be comparable to that offered as 'baseline' in control CCGs. The evaluation will not exclude NHS Integrated Care Pioneer sites from the pool of control CCGs, due to variations in the models locally adopted to join up health and social care and in the scale of those initiatives. A sensitivity analysis will however be carried out to ascertain if the GSynth counterfactuals built for the main analysis are robust to the removal of Integrated Care Pioneers from the corresponding pool of control CCGs. Furthermore, the GSynth counterfactual will control for other initiatives that are not specific to the treated group in the post period, with model-based risk adjustment applied to account for time-varying differences in patient and/or GP characteristics, as appropriate (see Tables 3 and 4 in the Statistical analysis section). Furthermore, the GSynth approach looks for any unobserved time-varying changes and adjusts for these. However, there is still a risk of time-varying unobservable characteristics (for example, factors related to social isolation or receptiveness to new approaches to managing patients' conditions), whose effect on outcomes also varies over time, that are not adjusted for when applying the GSynth method, potentially leading to biased estimates. There is also a risk of bias due to area-level changes in access or exposure to additional or alternative interventions in the post-intervention period (ie the timeline following roll-out of the intervention) which are specific to the treated or control regions.

The evaluation will study the impact of the ICTs and concurrent initiatives over a period when it was continuously developing at the three CCGs, so that estimates may not capture the full long-term effects of the individual components of programmes implemented in the later stages of the vanguard period. The results are nonetheless expected to provide a reliable summative assessment of the overall causal impact of the initiatives on hospital use that will help national policymakers and local teams implementing ICTs or other similar initiatives to assess what is working and identify potential areas for further investigation or improvement.

Background

The evaluations will cover two Vanguards: the Happy, Healthy, at Home Vanguard in the NEHF CCG, and the Fylde Coast Vanguard (made up of the Fylde and Wyre CCG and Blackpool CCG).

- The Happy, Healthy, at Home Primary and Acute Care System (PACS) vanguard in the NEHF CCG provided a broad range of initiatives. One of the core services introduced in July 2015 by the vanguard was the deployment of ICTs in each of the five localities within NEHF; other initiatives included enhanced recovery at home, which was rolled out in full from April 2017 (an interim service had been launched prior to April 2015), and a rapid response team whose roll-out began in March 2017 and was formed of specially trained paramedics. At the time of the IAU evaluation there were 23 GP practices within NEHF CCG.
- The Fylde Coast Multispecialty Community Provider (MCP) vanguard was jointly led by the Blackpool and Fylde and Wyre CCGs. The Fylde Coast NHS vanguard launched two complementary integrated care programmes: Extensive Care Service (ECS) in June 2015 and Enhanced Primary Care (EPC) in October 2016. ECS focuses on serving older people with more comorbidities, replacing usual GP care with a coordinated and specialist multidisciplinary team, whereas EPC supports an individual's existing GP relationship, where they have been identified as potentially benefiting from wrap-around multidisciplinary care such as health coaching. The two services were initially designed to be complementary, targeting somewhat different patient populations and with EPC serving as a 'step-down' bridge between ECS and routine primary and community care. A third vanguard initiative made available to the whole community was the Episodic Care model, which comprised a broad range of projects aimed at releasing capacity within primary care by targeting individuals with a minor short-term illness or health concern, for which the first point of access to support didn't need to be their GP practice. The city area of Blackpool, which at the time of the IAU evaluation contained 21 GP practices, experiences significant levels of deprivation, health inequalities and low life expectancy ranking among the worst in the country. The suburban and rural towns and villages of Fylde and Wyre CCG have a socio-economic profile closer to the English national average but face a growing proportion of older people and greater numbers of individuals with multiple and long-term conditions. At the time of the IAU evaluation there were 19 GP practices in Fylde and Wyre CCG.

For further details on the background of the integrated care programmes that were implemented in the two vanguard areas, please see the NEHF policy briefing[1] and the individual SAP of the evaluations published by the IAU.[2,4]

Methods

Study design

We will use national SUS and publicly available reference data to evaluate the impact of the ICTs and concurrent initiatives in the three geographical areas (Fylde and Wyre CCG, Blackpool CCG and NEHF CCG) for people of 65 years of age or over on hospital activity. The Fylde and Wyre and Blackpool CCGs will be analysed separately, even though they together form the Fylde Coast vanguard, due to their markedly differing population characteristics. For each of the three CCGs we will carry out the analyses by contrasting the change in outcomes for patients of 65 years of age or over and registered with a GP practice with a comparator based on patients of 65 years of age or over and registered with selected GP practices in comparable CCGs from other parts of England. The study start date will be 1 July 2015 for the Fylde Coast CCGs and 1 August 2015 for the NEHF CCG; the end date for all studies will be February 2020.

After first selecting a pool of similar CCGs, individual outcome and other activity, socio-economic and demographic data (referred to as covariates) will be collected at patient level for each month; for activities (eg A&E attendances) these will be collected as individual counts. Covariates will then be aggregated to monthly activity counts at GP practice level for each GP practice within the CCGs. These will be converted to rates (per 10,000 over GP practice population size) for the analysis. The treated group for each study comprises all GP practices that fall under the remit of the treated CCGs; all other GP practices will form part of the pool of potential GP practices for the control group. From this, a set of comparable GP practices will be selected as the control group. The final analysis data set will consist of aggregated outcome and covariate data for all treated and control GP practices as monthly series for 24 months before and all months after each of the intervention start dates up to February 2020 (56 and 55 months, respectively, for the Fylde Coast CCGs and NEHF), with the follow-up window closing prior to the onset of the COVID-19 pandemic. We will compare aggregated outcomes between GP practices in the treated and control groups, with adjustment for observed time-varying differences in patient mix or practice characteristics as well as pre-intervention (ie prior to the introduction of ICTs) trends in outcomes to obtain a so-called synthetic control group (the 'counterfactual'). Risk adjusted outcomes from the derived counterfactual for each treated GP over time will then be compared to the actual outcome in each post-intervention period to obtain an estimate of the intervention effects over time. These will be aggregated into consecutive yearly estimates to detect patterns of variation over time, as well as over the study follow-up period, to obtain an overall intervention effect estimate.

To build the synthetic control group we will be using a GSynth[5] approach, which generalises Difference-in-Difference (DiD) methods. DiD is a traditional approach to estimating treatment effects in this setting, where repeated measurements are available on the same set of units at different time points before and after an intervention of interest is made available. DiD stipulates that the post-intervention experience of patients in the control group, combined with the pre-existing difference between the groups' outcomes, represents an appropriate counterfactual for patients in the untreated group under the assumption that, in the absence of treatment, the expected difference in outcomes between the comparator groups is constant over time (the 'parallel-trends' assumption). However, if there are unobserved time-varying effects, estimates from DiD may be biased.[6,7] The GSynth method relaxes the parallel trends assumption and allows for the effects of both observed and unobserved confounders to vary over time.

Study cohorts

Target population

We adopt a population approach to the evaluation to provide an overall picture of the effect of the ICTs on health care for the local population of three administrative areas (Fylde and Wyre CCG, Blackpool CCG and NEHF CCG). We examine hospital use for the whole population (of either 65 or 18 years of age or over) of each of the three areas (treated regions) and compare it with that for the corresponding population of the derived synthetic control areas. The synthetic controls provide an estimate of the hospital use that would have been expected in the treated areas in the absence of the ICTs. The treated regions are represented by the group of GP practices belonging to the three respective CCGs that make up the region where each ICT programme is implemented (ie Blackpool, Fylde and Wyre and NEHF). The synthetic control region – one for each of the treated regions – is represented by a group of GP practices from comparable CCGs in other parts of the country. The target population for each study consists therefore of all patients of 65 years of age or over (18 years of age or over in a secondary analysis) registered with a GP practice in England (the control group). Note that this approach excludes hospital activity for individuals not registered with a GP practice.

Study cohort

The study cohort is defined by those individuals in the target population that were registered with a GP falling under the remit of either an intervention or a potential control CCG in England during the study period. Although the analysis uses activity data for continuous inpatient spells (CIPS) and A&E attendances, the study cohort is the entire target population regardless of whether they actually attend hospital for treatment, although the outcomes of interest will relate to their actual use of secondary care. This allows us to capture reductions in use as well as changes in the composition of hospital resources used.

Study outcomes

Outcomes of primary interest, all measured per head of GP practice size per month (see the Appendix for definitions) are:

- rate of A&E attendances
- rate of emergency admissions
- rate of chronic ambulatory care sensitive (ACS) emergency admissions
- rate of acute ACS emergency admissions
- average length of stay (LoS) for emergency admissions
- rate of deaths among people under 65 years of age (as a placebo test).

Outcomes are originally collected at the patient level as monthly counts of activity data and then aggregated to monthly counts at GP practice level. Rates are then calculated according to GP practice size, number of corresponding admissions, or attendances at A&E, as indicated, in order to make figures comparable across differently sized practices or changes in patient population or case-mix over time.

Sources of data

CCG and GP practice reference data

Data relating to key characteristics of CCGs and GP practices, which are collected from publicly available sources and used to identify control CCGs comparable to treated ones, form a pool of comparable GP practices, create the synthetic control groups and/or for risk adjustment, are next listed.

- NHS Digital. *Number of patients registered at a GP practice.* 2013–2020
- NHS Digital. *General and personal medical services.* 2015
- NHS Digital. *Quality Outcome Framework.* 2013–2020
- Office for National Statistics. *Census.* 2011
- Office for National Statistics. *Mid-year population estimates.* 2015–2020
- Office for National Statistics. *Lower Layer Super Output Area population density.* 2015
- Ministry of Housing, Communities and Local Government. *English indices of deprivation.* 2015
- Care Quality Commission. *CQC Care Directory.* 2015

We are including data over the two years prior to the start of each of the study periods as these will be used to characterise CCGs and GP practices in the pre-intervention period. Variables are available at either GP practice or Lower Layer Super Output Area (LSOA) level. Variables available at LSOA level will be mapped to GP practice level according to the LSOA of each patient registered at the GP practice. Variables available at GP practice level will be aggregated to CCG level and weighted according to the registered population size of each GP practice. Missing data will be inferred using the closest available value in time. The use of annual data reflects data availability and a belief that the variables will not exhibit substantial variation within a year.

Hospital activity data

Hospital activity data will be obtained from de-identified (ie anonymised in line with the Information Commissioner's Office code of practice on anonymisation) SUS extracts. SUS is a national, person-level database that is closely related to the widely used Hospital Episode Statistics (HES) database. It is used to support the NHS in the delivery of health care services and to trigger reimbursement for secondary care activity. The IAU has access to these data for its work and processes them in The Health Foundation's secure environment. All data are de-identified, meaning that they have been stripped of fields that can directly identify a patient (such as name, full date of birth and address). The NHS number is replaced with a pseudonym, which is used to link records for the same individual over time. The overall approach to information governance has been scrutinised by information governance experts at NHS Digital.

Patient-level monthly activity records for each outcome, as well as data relating to the characteristics of patients seeking treatment, were collected from two years before the intervention start date and up to September 2019 for all patients registered at treated or potential control GP practices.

Only activity data for patients registered with a treated or a potential control GP practice are retained. A&E visits for a patient who left before being seen or refused treatment, or where the visit is a duplicate, are excluded. Inpatient data are structured into CIPS, which may consist of several consultant episodes (since patients may be under the care of multiple consultants during a hospital stay) and stays at several hospitals (if patients are transferred). Spells that are missing an admission date, or where the discharge date preceded the admission date due to data quality problems, are excluded. A&E visits and spells with gender given as other than male or female are also excluded: although these records were considered valid, they cause technical difficulties for statistical modelling purposes.

Statistical methods

Selecting the control group

We will use national SUS and publicly available reference data to evaluate the impact of the ICTs and concurrent vanguard initiatives in Fylde and Wyre, Blackpool and NEHF CCGs for individuals of 65 years of age or over. We do this by contrasting the change in outcomes for patients of 65 years of age or over registered with a GP practice in each of these CCGs with that of patients of 65 years of age or over registered with selected GP practices in comparable CCGs from other parts of England. These CCGs and GP practices were chosen to be similar to their treated counterparts before the introduction of the ICTs. The study period started from the implementation of their ICTs (set as 1 July 2015 for both CCGs in Fylde Coast and 1 August 2015 for NEHF) and lasted up to the end of February 2020.

Each intervention CCG will be analysed separately, with independent counterfactuals being built for each. For each analysis, we will use a three-step process to identify the synthetic control group:

1. to ensure similarity at a regional (CCG) level, we first select the most comparable CCGs

2. to ensure similarity at a local (GP practice) level, we will then select a subset of GP practices within each comparable CCG, so that the 10 most similar practices to the GP practices in the treated CCG are selected to form the donor control group

3. create a synthetic control group consisting of synthetic control GP practices to each treated practice.

Selecting comparable CCGs

To assess similarity between units (CCGs or GP practices), we will adapt the method used in NHS England and NHS Improvement's NHS RightCare 'Similar 10 CCG' Explorer Tool.* This method assesses similarity by computing squared Euclidean distances (SED) between each unit across a set of variables (eg age structure, ethnic mix and level of deprivation), with a lower SED indicating greater similarity.† We included variables of relevance that are publicly available and pre-intervention hospital use data that could be derived from SUS data, including history of emergency admissions in the year prior to the implementation of the ICTs and the year before (Table 1), for three different age bands: 0–17, 18–64 and 65+ years. This aims to ensure that we exclude from the subsequent analysis those untreated CCGs that are most dissimilar to each of the treated ones across the pre-intervention period. Data unavailable at CCG level (eg recorded at LSOA, local authority or GP level) will be aggregated to CCG level as described above. Since the variables are measured on different scales, and hence are not directly comparable, the data will first need to be standardised. This will be done by using inter-decile range standardisation, obtained by subtracting the median from the original data and then dividing by the difference between their 90th and 10th percentiles‡, in line with Office for National Statistics' practice and the Commissioning for Value tool, prior to calculating the SED.

The most similar CCGs will be selected after adopting the following exclusions criteria:

- vanguards
- London CCGs (for the NEHF and Fylde and Wyre CCGs analyses only)
- CCGs that, together with NEHF CCG, formed part of the Frimley Health and Care ICS from April 2016, since these are likely to induce spillover effects due to close working relationships. Excluded CCGs (for NEHF CCG analysis only) are East Berkshire CCG (previously Bracknell and Ascot CCG, Windsor and Maidenhead CCG and Slough CCG) and Surrey Heath CCG.

The number of most similar CCGs retained will depend on the number of GP practices available in the potential control CCGs (given the practice-level inclusion/exclusion criteria outlined below). This is likely to be between 20 and 100 CCGs, with the aim of providing a pool of around 1,000 potential GP practices from which the control group will be selected.

We will not exclude other neighbouring areas as we do not consider the risk of spillover to be substantial, given that the intervention and control groups are based on the registered patient populations within the CCG and only these were eligible for ICTs.

We are evaluating the effect of the ICTs and other initiatives over and above 'usual care'. We will therefore exclude other vanguard areas, as these are also providing new health care improvement initiatives. However, we will not exclude other areas making efforts to

---

* The 'Similar 10 CCG Explorer Tool' calculates the 10 CCGs in England most similar to a given CCG. See www.england.nhs.uk/publication/similar-10-ccg-explorer-tool/.

† Suppose there are I units and K baseline variables. Let $\tilde{x}_{ki}$ represent the standardised version of x_{kj}, where x_{kj} is the k-th baseline variable observed in unit i, i=1, …I, k=1,…, K. Then the SED between unit i and unit j is calculated across K baseline variables $SED_{ij} = \sqrt{\sum_{k=1}^{K}(\tilde{x}_{ki} - \tilde{x}_{ki})^2}$ for i, j<I, i≠j.

‡ The standardised value of baseline variable is calculated as $\tilde{x}_{kj} = \frac{x_{kj} - median(x_{1j}, x_{2j}, \ldots, x_{Kj})}{90th\ perc(x_{1j}, x_{2j}, \ldots, x_{Kj}) - 10th\ perc(x_{1j}, x_{2j}, \ldots, x_{Kj})}$.

integrate care, as these are happening across the country and could therefore be reasonably considered as 'standard care', while lacking vanguard funding. This includes for instance areas participating in the Integrated Care and Support Pioneer programme, which started in November 2013, and the Better Care Fund programme (from which NEHF received funding in 2014 as part of Hampshire County).

Table 1: List of variables used to identify most comparable CCGs. Rates of hospital activity indicate number of events per 10,000 people in the specified age band registered in the CCG, calculated across each of the 2 years in the pre-period.

Variable	Source	Target cohort	Fylde & Wyre CCG	Blackpool CCG	NEHF CCG
Percentage of females	ONS		Mid-2015	Mid-2015	Mid-2015
Percentage of persons aged 0 to 4 years	ONS		Mid-2015	Mid-2015	Mid-2015
Percentage of persons aged 5 to 14 years	ONS		Mid-2015	Mid-2015	Mid-2015
Percentage of persons aged 15 to 44 years	ONS		Mid-2015	Mid-2015	Mid-2015
Percentage of persons aged 45 to 64 years	ONS		Mid-2015	Mid-2015	Mid-2015
Percentage of persons aged 65 to 74 years	ONS		Mid-2015	Mid-2015	Mid-2015
Percentage of persons aged 75 to 89 years	ONS		Mid-2015	Mid-2015	Mid-2015
Rural/Urban Indicator 2011	Census		2011	2011	2011
Rate of care home beds available according to CQC (per 10,000 registered population)	CQC	Whole population	Jun-15	Jun-15	Jul-15
Rate of GPs (FTE) (per 10,000 registered population)	NHS Digital	Whole population	2014–15	2014–15	2014–15
Population density – number of people per square kilometre	ONS		Mid-2015	Mid-2015	Mid-2015
Percentage of ethnicity recorded as White	Census		2011	2011	2011
Percentage of ethnicity recorded as Black	Census		2011	2011	2011
Percentage of ethnicity recorded as Asian	Census		2011	2011	2011
Percentage of ethnicity recorded as Mixed	Census		2011	2011	2011
Percentage of ethnicity recorded as Other	Census		2011	2011	2011
Individuals' day-to-day activities limited a lot or a little (standardised illness ratio)	Census		2011	2011	2011

Index of multiple deprivation (IMD) quintile	IMD		2015	2015	2015
Health deprivation and disability score	IMD		2015	2015	2015
Income Deprivation Affecting Older People Index (IDAOPI) score	IMD		2015	2015	2015
Rate of yearly emergency admissions for chronic ambulatory care sensitive conditions	SUS	age 65+	01/07/2013 – 30/06/2015	01/07/2013 – 30/06/2015	01/08/2013 – 31/07/2015
Rate of yearly emergency admissions for acute ambulatory care sensitive conditions	SUS	age 65+	01/07/2013 – 30/06/2015	01/07/2013 – 30/06/2015	01/08/2013 – 31/07/2015
Rate of yearly elective admissions	SUS	age 65+	01/07/2013 – 30/06/2015	01/07/2013 – 30/06/2015	01/08/2013 – 31/07/2015
Rate of yearly emergency admissions	SUS	age 65+	01/07/2013 – 30/06/2015	01/07/2013 – 30/06/2015	01/08/2013 – 31/07/2015
Rate of yearly outpatient appointments attended	SUS	age 65+	01/07/2013 – 30/06/2015	01/07/2013 – 30/06/2015	01/08/2013 – 31/07/2015
Rate of yearly A&E attendances	SUS	age 65+	01/07/2013 – 30/06/2015	01/07/2013 – 30/06/2015	01/08/2013 – 31/07/2015
Rate of yearly emergency admissions for chronic ACS conditions	SUS	age 18–64	01/07/2013 – 30/06/2015	01/07/2013 – 30/06/2015	01/08/2013 – 31/07/2015
Rate of yearly emergency admissions for acute ACS conditions	SUS	age 18–64	01/07/2013 – 30/06/2015	01/07/2013 – 30/06/2015	01/08/2013 – 31/07/2015
Rate of yearly elective admissions	SUS	age 18–64	01/07/2013 – 30/06/2015	01/07/2013 – 30/06/2015	01/08/2013 – 31/07/2015
Rate of yearly emergency admissions	SUS	age 18–64	01/07/2013 – 30/06/2015	01/07/2013 – 30/06/2015	01/08/2013 – 31/07/2015
Rate of yearly outpatient appointments attended	SUS	age 18–64	01/07/2013 – 30/06/2015	01/07/2013 – 30/06/2015	01/08/2013 – 31/07/2015
Rate of yearly A&E attendances	SUS	age 18–64	01/07/2013 – 30/06/2015	01/07/2013 – 30/06/2015	01/08/2013 – 31/07/2015
Rate of yearly emergency admissions for chronic ACS conditions	SUS	age 0–17	01/07/2013 –30/06/2015	01/07/2013 – 30/06/2015	01/08/2013 – 31/07/2015
Rate of yearly emergency admissions for acute ACS conditions	SUS	age 0–17	01/07/2013 – 30/06/2015	01/07/2013 – 30/06/2015	01/08/2013 – 31/07/2015
Rate of yearly elective admissions	SUS	age 0–17	01/07/2013 – 30/06/2015	01/07/2013 – 30/06/2015	01/08/2013 – 31/07/2015
Rate of yearly emergency admissions	SUS	age 0–17	01/07/2013 – 30/06/2015	01/07/2013 – 30/06/2015	01/08/2013 – 31/07/2015
Rate of yearly outpatient appointments attended	SUS	age 0–17	01/07/2013 – 30/06/2015	01/07/2013 – 30/06/2015	01/08/2013 – 31/07/2015
Rate of yearly A&E attendances	SUS	age 0–17	01/07/2013 – 30/06/2015	01/07/2013 – 30/06/2015	01/08/2013 – 31/07/2015

Selecting comparable GP practices

Once the most comparable CCGs to the three intervention areas are identified, we will draw from this pool the 10 most similar GP practices to each treated one. First, we will exclude from potential control CCGs GP practices:

- which opened or closed during the study period
- with registered patient population sizes outside the range of registered sizes in the treated group
- with patterns of key outcome variables indicative of reporting errors
- without records in both A&E and emergency CIPS files used in the analysis, so we would have the same list of GP practices in A&E and emergency CIPS files
- where there was an annual or quarterly change of at least 20% in important variables. While these changes may be legitimate – a large change in the number of registered patients may for instance be indicative of GP practices merging – we will not be able to determine when the change occurred, leading to any imputed values in the intervening months being unreliable. If a variable is unreliable but not necessary to the analysis, we may instead choose to drop it.

We will then use similar methods to those used for identifying comparable CCGs, thus selecting similar GP practices based on variables such as population size, number of full-time equivalent GPs and disease prevalence in the whole practice population (see Table 2). Variables either relate to the key reference characteristics of GP practices, and thus refer to the whole registered population, or are derived from SUS activity data. Variables will be mapped to produce yearly series of data across the study period aggregated at GP level; those only available at CCG level will be mapped to GP level by weighting according to the number of registered patients within each CCG area. Lastly, variables available at levels other than GP practice are mapped to GP practice level. Due to the large number of variables included, some of which are closely related to each other, these will be weighted according to how predictive they are of the rate of emergency hospital admissions in the whole population in the last year of the pre-intervention period, after adjusting for the other variables in the year prior to that (see Table 2). The weight given to each variable is determined by the absolute value of the corresponding T test statistic, estimated from a regression of the rate of emergency hospital admissions in the year prior to the start of the intervention on the variables for the preceding year.[8] We will incorporate multiple periods by calculating distance measures using annual estimates of each variable across yearly increments during the 2 years prior to the start of the ICTs. This aims to ensure that the 'most similar' untreated practices are similar to practices in the intervention areas during the pre-intervention period. Individual outcome and past hospital activity, socio-economic and demographic data are collected at the patient level and then aggregated to monthly activity counts at the GP practice level.

The pool of control GP practices will be ordered in terms of their similarity to each of the GP practices in the treated group in the 2 years prior to the start of the ICT intervention, with the 10 GP practices found to be most similar to each of the treated practices being ultimately selected. If the same control GP practice appears among the 10 most similar GP practices to multiple treated GP practices, this will be included in the control pool once but may be weighted more heavily in the GSynth building process.

We will assess whether the characteristics of GP practices in the selected control group pool are similar to those in the treated group in the pre-intervention period by calculating a standardized mean difference (SMD) or SED to compare the average values of the variables between treated and control groups.

We will apply the methods described to compute the SMDs or distance between all pairs of treated and control units and use these values to order the control units in terms of decreasing overall similarity as follows. First, we will select the set of control units that are first nearest to each of the treated units. We then arbitrarily order these, excluding any duplicates. We will then include into this list those control units that are second nearest to each of the treated units, excluding any duplicates in this set and in the list already established. And so on, until all control units are included in the list. The list will then provide a single ordered list of control units in terms of their decreasing similarity with the treated units across these variables. This list is used to select the control group for all outcomes.

When running the analysis for a given outcome the raw values of the outcome in the treated GP practices and in the ordered list of control GP practices over the study period will be reviewed: any practices exhibiting an unusual pattern in the outcome will be removed from the analysis – see the Diagnostics section for more details. The control practices for that outcome will be then be selected as the first 100 practices in the thinned list. A sensitivity analysis will assess the robustness of results to selection of a different set of 100 control practices from the thinned list, and to the first 250 and 500 control GP practices in the thinned list.

Table 2: List of variables used to identify most comparable GP practices.

Variable	Description	Date of collection	Level of collection	Source
Population size[*]	Number of registered patients	Quarterly 2013–15	GP practice	NHS Digital
Age	Proportion of registered patients aged 0–17, 18–24, 25–64, 65–74, 75+ years	Annually	GP practice	NHS Digital
Gender	Proportion of registered male patients	Annually	GP practice	NHS Digital
Ethnicity	Proportion of registered patients with self-reported race White, Black, Asian and Mixed	29 March 2011 (census day)	GP practice	Office for National Statistics
Education	Proportion of registered patients with at least third-level education (two or more A levels or equivalent)		GP practice	
Population density	Rate of persons per hectare in the nearest electoral ward		GP practice	

* As well as being included as a covariate, population size is included as an offset in the regression model to calculate weights.

Socio-economic deprivation	Weighted average of LSOA-level IMD scores according to LSOA of GP registered patients	2015	LSOA	Ministry of Housing, Communities and Local Government
Health deprivation	Weighted average of LSOA-level IMD scores on health deprivation, according to LSOA of GP registered patients		LSOA	
Rate of full-time equivalent general practitioners	Rate of full-time equivalent general practitioners per 10,000 people in the registered GP population	2015	GP practice	NHS Digital
QOF achievement score proportions	The proportion of points achieved, out of total number of points possible, for each QOF indicator. These include cardiovascular, respiratory, high dependency and other long-term conditions, as well as musculoskeletal QOF indicator groups	Annually 2014 and 2015	GP practice	NHS Digital
QOF disease prevalence	Proportion of registered population with atrial fibrillation, coronary heart disease, cardiovascular disease, heart failure, hypertension, peripheral arterial disease, stroke and transient ischaemic attack, asthma, COPD, cancer, chronic kidney disease (18 years of age or over only), diabetes, palliative care, osteoporosis (50 years of age or over only) and rheumatoid arthritis (16 years of age or over)		GP practice	
History of Elixhauser comorbidity categories (rate)	Total number of emergency admission CIPS in a given year for patients with a history of each of the Elixhauser comorbidity categories[*] (one variable per comorbidity in the Elixhauser list) per 10,000 of the registered GP population. Comorbidity flags are computed according to primary and secondary diagnosis ICD-10 Version 2015 classification codes for all their inpatient admissions in the preceding 24 months[9] for all patients with emergency admission activity in the month	Annually 2014 and 2015	Activity	SUS

[*] There are 31 Elixhauser comorbidity categories which are used to compute the Elixhauser comorbidity index. Here we exclude those for AIDS/HIV and obesity, therefore only 29 categories are considered.

History of Elixhauser Index ≥ 2 (rate)	Total number of emergency admission CIPS in a given year for patients with an Elixhauser index ≥ 2 per 10,000 of the registered GP population. Comorbidity flags are computed according to primary and secondary diagnosis ICD-10 Version 2015 classification codes for all their inpatient admissions in the preceding 24 months[9] for all patients with emergency admission activity in the month	Annually 2014 and 2015	Activity	SUS
History of dementia (rate)	Total number of emergency admission CIPS for patients with dementia per 10,000 of the registered GP population. The dementia flag is computed according to primary and secondary diagnosis ICD-10 Version 2015 classification code in any of their inpatient admissions in the preceding 24 months for all patients with activity in the month	Annually 2014 and 2015	Activity	SUS
Proportion of inpatient admissions by primary diagnosis	Total number of emergency CIPS admissions for patients with positive flags for 22 categories of primary and secondary diagnosis ICD-10 Version 2015 classification code in the preceding 24 months per 10,000 of the registered GP population	Annually 2014 and 2015	Activity	SUS
History of emergency admissions (rate)	Total number of emergency CIPS admissions in the preceding 24 months per 10,000 people in the registered GP population	Annually 2014 and 2015	Activity	SUS
History of A&E attendances (rate)	Total number of A&E attendances in the preceding 24 months per 10,000 people in the registered GP population	Annually 2014 and 2015	Activity	SUS

Creating a synthetic control group

The treated group comprises all GP practices within the intervention CCGs; the control group comprises comparable GP practices from comparable CCGs in other parts of England. For the counterfactual analyses we will use the GSynth method, which will create a synthetic control for each of the treated GP practices. These will then be aggregated to their respective treated and control CCGs to obtain an overall intervention effect. GSynth depends on an assumption of weak serial dependence of error terms.

Generalised Synthetic Control (GSynth)

The GSynth method, introduced by Xu et al[5], of which DiD is a special case, also generalises the original Synthetic Control (Synth) method.[10] It allows for multiple treated units and variable treatment periods. Despite its name, GSynth is not a traditional synthetic control method. Rather, it combines the efficiency gains of Interactive Fixed Effect (IFE) models with insights from the Synth methodology, allowing for unobserved covariates to have time-varying effects and enabling separate estimates for each treated unit. The intervention effects are estimated for each outcome in each treated GP practice at each post-intervention time period. Estimates are then averaged across all treated GP practices to provide an estimate of the Average Treatment effect among those Treated (ATT) at each post-intervention time period. The GSynth method has the advantage over IFE that it is unbiased even in the presence of treatment effect heterogeneity and retains the unbiasedness of DiD, while being more precise if the underlying model is correctly specified[11].

We apply the GSynth estimator to each outcome separately, including baseline covariates and pre-intervention outcomes for risk adjustment (see the Risk adjustment section below). The bundled effects of the ICTs with concurrent initiatives on each outcome at each treated GP practice are then estimated for each post-intervention time period. Weighted averages can be obtained across all treated GP practices to provide an estimate at each post-intervention time period, as well as across all post-intervention time periods, to obtain an estimate of the ATT. Significance of estimates will be calculated using bootstrapping methods.[12] GSynth estimates are derived in practice using the ‘gsynth’ package in R.

Risk adjustment

Aggregated outcomes across GP practices in the treated and control groups will be compared, after adjusting for differences in time-varying characteristics (eg GP practice characteristics, patient mix and pre-intervention trends in outcomes) in the GSynth estimation. The GSynth method will then provide an estimate of the counterfactual (risk adjusted) outcome for the CCGs in receipt of the intervention over time. We will identify time-varying (observed) characteristics to include in the model by graphical assessments (see the Diagnostics section). We anticipate that the number of full-time equivalent GPs and QOF prevalence rates will be useful to include in the risk adjustment. For the examined outcomes, two possible groups of risk adjustment variables can be devised.

- **A&E attendance and emergency admissions (including subsets of emergency admissions) outcomes (Table 3).** When looking at A&E visits, emergency admissions or total deaths, all patients in the relevant patient population (ie of 65 or 18 years of age or over, or under 65 years of age as applicable) registered with a treated or control GP practice are relevant. Characteristics adjusted for could include age (ie proportion of emergency CIPS involving patients of 65–74 or 75 years of age or over), gender, ethnicity and education degree, which are all derived from publicly available information at GP practice level; or QOF prevalence rate, which is not available for specific age groups. The latter is aiming to categorise the severity of illness in the GP practice population. As QOF prevalence rates and scores are only available on a yearly basis, we will linearly interpolate monthly values to reflect the change from one month to the next. We do not expect QOF prevalence rates or scores to be affected by the ICTs.
- **Average length of stay following emergency admission (Table 4).** Here only the characteristics of those patients with an emergency admission on record are relevant. These could include age, gender and ethnicity, primary diagnosis code at emergency admission, history of Elixhauser scores greater than 1 and history of Elixhauser comorbidity categories derived from patients with an emergency admission in a given month, looking back at their hospital records in the previous 2 years. The primary diagnosis code and the Elixhauser categories are aiming to categorise the type and severity of illness in the admitted patients.

Variables will first be collected monthly from SUS data for patients' hospital activity. Data on age, gender and ethnicity will be collected from A&E, emergency and elective admission activity, with the most common value across records being adopted. Data will then be aggregated to produce monthly event counts used for risk adjustment and comparison.

In addition to adjusting for specified covariates, the GSynth method uses outcome data from the pre-intervention period for both treated and control units, as well as from the post-intervention period for the controls, to impute unobserved time-varying components and includes these in the model as statistical 'factors'. Factors are formally the time-varying coefficients in the IFE model* proposed by Xu et al[5], while unit-specific intercepts, reflecting time-invariant unobserved variables, are called factor loadings. The number of factors to include is estimated using a data-driven approach; see the Appendix for technical details.

* The IFE model represents values of an outcome Y_{it} observed on unit i at time t as $Y_{it} = \boldsymbol{\beta}\, \boldsymbol{X}_{it} + \boldsymbol{\lambda}_t \boldsymbol{\mu}_i + D_{it}\alpha_{it} + \varepsilon_{it}$, where factors and factor loadings are respectively denoted by $\boldsymbol{\lambda}_t$ and $\boldsymbol{\mu}_i$.

Table 3: Variables used for risk adjusting A&E attendances and emergency admissions. Variables describe key reference characteristics of GP practices and, where possible, relate to patients aged according to the target population.

Variable	Description	Date of collection	Level of collection	Source
Population size	Number of registered patients	Quarterly 2013–20	GP practice	NHS Digital
Age*	Proportion of registered patients of 18–24, 25–64, 65–74 years of age and of 75 years of age or over	Annually 2013–20	GP practice	NHS Digital
Gender*	Proportion of registered male patients	Annually 2013–20	GP practice	NHS Digital
Ethnicity*	Proportion of registered patients with self-reported White, Black, Asian and Mixed ethnicity	29 March 2011 (census day)	GP practice	Office for National Statistics
Education	Proportion of registered patients with at least third-level education (ie two or more A levels or equivalent)	29 March 2011 (census day)	GP practice	Office for National Statistics
Population density	Rate of persons per hectare in the nearest electoral ward	29 March 2011 (census day)	GP practice	Office for National Statistics
Socio-economic deprivation	Weighted average of LSOA IMD scores according to the LSOA of GP registered patients	2015	LSOA	Ministry of Housing, Communities and Local Government
Health deprivation	Weighted average of LSOA-level IMD scores on health deprivation according to LSOA of GP registered patients	2015	LSOA	Ministry of Housing, Communities and Local Government
Number of full-time equivalent general practitioners	Rate of full-time equivalent general practitioners per 10,000 people in the registered GP population	2015	GP practice	NHS Digital
QOF achievement score proportions	The proportion of points achieved, out of total number of points possible, for each QOF indicator. These include cardiovascular, respiratory, high dependency and other long-term conditions, and musculoskeletal QOF indicator groups	Annually 2013–19	GP practice	NHS Digital
QOF disease prevalence	Proportion of registered population with atrial fibrillation, coronary heart disease, cardiovascular disease, heart failure, hypertension, peripheral arterial disease, stroke and transient ischemic attack, asthma, COPD, cancer, chronic kidney disease (18 years of age or over only), diabetes, palliative care, osteoporosis (of 50 years of age or over only) and rheumatoid arthritis (of 16 years of age or over)	Annually 2013–19	GP practice	NHS Digital

Table 4: Variables used for risk adjusting average length of stay following emergency admission. These relate to patients with hospital activity in the month and to the target population.

Variable	Description	Date of collection	Level of collection	Source
Age	Total number of emergency admission CIPS for patients of 65–74 years of age, and 75 years of age or over per 10,000 patients admitted as an emergency to hospital	Monthly 2011–20	Activity	SUS
Gender	Total number of emergency admission CIPS for male patients per 10,000 patients admitted as an emergency to hospital	Monthly 2011–20	Activity	SUS
Ethnicity	Total number of emergency admission CIPS for patients with self-reported race White, Black, Asian and Mixed per 10,000 patients admitted as an emergency to hospital	Monthly 2011–20	Activity	SUS
Proportion of inpatient admissions by primary diagnosis	Total number of emergency admission CIPS for patients with positive flags for 22 categories of primary and secondary diagnosis ICD-10 Version 2015 classification code in the preceding 24 months per 10,000 patients admitted as an emergency to hospital	Monthly 2011–20	Activity	SUS

Inference

The difference between the observed outcomes in the treated group and those estimated from the counterfactual provides an estimate of the effect of the ICTs and concurrent initiatives on hospital use. These estimates will be produced and reported both as yearly and overall figures throughout the follow-up study period, to allow discerning any pattern of impact variation over time and to provide a summary indication of effectiveness respectively. However, the precision of this estimate needs to be assessed. We will do this using parametric bootstrapping procedures as described in Xu et al[4] to generate p-values and confidence intervals.

Diagnostics

Pre-analysis diagnostics

Extrapolation biases can occur when the covariates of the treated and control practices do not share a ‘common support’, ie when their sampling distributions lack overlap. To verify that this is not the case:

- We will plot all raw outcomes for all GP practices in the donor pool of controls over the whole study period. The plots will be visually inspected and any practices that exhibit unusual patterns in the outcome – such as sudden dips for months at a time, or clearly implausible values – will either be removed from the analysis for the particular outcome, or have their outcomes inferred from contiguous months. Unusual trends are typically indicative of a lapse in reporting and are often common to all practices in a CCG but may also result from a mismatch between counts of activity and GP practices’ size following a merger between GP practices. We would do the same with the size of GP practices over time and check seasonality in the trends. If we observe large changes in the trends from one month to the other – other than seasonality around admissions in winter or those changes that were already removed after the previously outlined exclusion criteria – we would apply mean imputation for those unusual months. Additionally, we will calculate summary statistics (minimum, maximum, mean, median) for all variables to check if there are any problematic values.
- We will also remove GP practices that are problematic according to any outcome by the sorting procedure described in the Selecting the control group section. To check that control GP practices are more similar to the treated GP practices than the remaining GP practices in the control donor pool, we will calculate the average Euclidean distance from each treated practice to included controls and to the excluded ones and check that the majority of treated practices will be (on average) closer to the included than excluded GPs.
- Trends in the unadjusted outcomes of the control GP practices over the pre-intervention period will be compared to those in the treated GP practices to assess whether interpolation of the outcomes of the control GP practices can reasonably be expected to approximate the average outcomes of the treated GP practices. This will be carried out via Principal Component Analysis with the aim of identifying underlying drivers to the structure of GP practices. Specifically, we will first compute the eigenvectors (principal components) and eigenvalues of the matrix of SED values between all pairs of treated and control GP practices; see the Selecting the control group section for details of calculating the SED between a pair of units. Secondly, we will examine scatterplots of pairs of principal components illustrating the similarities between the treated and the selected control GP practices at the top of the ordered list, in contrast to control GP practices further down the list. These plots will be used to refine the choice of GP practices in the treated and control groups.

We will identify time-varying covariates to include in the risk adjustment process by plotting all observed variables described in Table 2 over time.

Post-analysis diagnostics

The GSynth method provides predicted outcomes for the treated units in the pre- and post-intervention periods. Intuitively, if the predicted outcome is not a good estimate of the observed outcome in the pre-intervention period it will unlikely be reliable in the post-intervention period either. We will assess whether the synthetic control provides a good fit via the following:

- We will compare observed and predicted outcomes from the group of treated GP practices during the pre-intervention period. To do this we will use the same methods used for estimating the impact of the ICTs in the post-intervention period. We will judge the synthetic control to provide a good fit if the null hypothesis of no difference is not rejected (ie p-value convincingly exceeding 0.05). If there is evidence of a poor fit, we will consider reducing the number of constraints on the synthetic control method by removing or averaging covariates or outcomes over pre-intervention periods in order to improve the fit.
- We will check the plausibility of any additional factors determined by the GSynth method by plotting factors and factor loadings and verifying that these overlap across treated and control practices.
- We will check similarity between treated and control GP practices by checking sensitivity to variables included for risk adjustment. This would mitigate the risk that arises when the risk adjustment model does not accurately account for differences in treated and control GP practices.

Sensitivity analyses

A few sensitivity analyses will be conducted to assess whether results are sensitive to the assumptions made in the base-case analysis. Namely:

- Using control groups comprising half the number of comparable GP practices in each CCG's control group (selecting the most similar ones) as well as twice as many GP practices (including practices that were not originally selected into the control group).
- Excluding CCGs that were part of the Integrated Care Pioneers.
- Choosing an earlier study start date, eg 6 months earlier. However, this needs to be considered within the context of different initiatives that might have been implemented at that earlier time, either in the treated or control CCGs (eg Integrated Care Pioneers or Better Care Fund).
- Reporting effect estimates from crude analyses (no risk adjustment).
- Accounting for history of long-term conditions by including individual Elixhauser conditions and a flag for having more than one of them in the risk adjustment for emergency admissions, A&E attendances and average length of stay following an emergency admission. These variables are not included in the main analysis as the recording of these conditions may be affected by the intervention.

Limitations and sources of bias

Threats to validity

See Box 1 for a summary of threats to validity and associated limitations and potential biases.

Box 1: Limitations, implications and mitigation

Limitation	Implication	Mitigation
Threats to internal validity		
Control practices have fundamentally different outcome and/or covariate values to the treated practices.	Unable to identify a suitable synthetic control that provides a sufficiently good fit in the pre-intervention period. Unsuitable synthetic control may lead to extrapolation biases.	Ensuring similarity between treated and control practices. Checking sensitivity to number and mix of GP practices in control group.
The risk adjustment model does not accurately account for differences between treated and control GP practices.	Inability to control for observed and unobserved confounders will lead to biased estimates.	Ensuring similarity between treated and control practices. Checking sensitivity to variables included for risk adjustment.
Treatment assignment may be correlated with time-varying unobserved confounders.[11]	Inability to account for unobserved confounders will lead to biased estimates. For example, attendance at A&E may depend on which local support services are available at the time and in the area where a patient lives (time-varying confounder), which is in turn correlated with which GP practice the patient is likely to be registered with (treatment assignment).	The GSynth method can infer presence of unobserved time-varying confounders with time-varying effects; these are reported as additional factor loadings. If detected by plotting them, these will be examined for plausibility and sensitivity to their inclusion.
The synthetic control is unduly capturing noise in outcomes, rather than underlying trends.	Credence is removed from the key assumption that similarity in pre-intervention outcome trends will also persist post-intervention. It is documented that if the number of pre-intervention periods increases, the bias of synthetic control estimates shrinks towards zero.[10] However, a criterion to determine a sufficient length for the pre-intervention period is not available.	Using a long pre-intervention of 24 months. Checking sensitivity to changes in the study start date (see the Sensitivity analyses section). Checking sensitivity to changes in the number of pre-intervention periods.
Interventions that influence outcomes in GP practices in the control group but not the treated group may have occurred in the pre-intervention period.	Key assumption that similarity pre-intervention will persist post-intervention will be implausible.	Excluding GP practices which are in vanguard areas. This does not exclude other areas that are implementing other models of ICTs. As a sensitivity analysis, we may also exclude pioneer sites and rerun the analyses. Also, checking sensitivity to using a different set of controls.

Interventions that influence outcomes in control GP practices may have occurred in the post-intervention period.	Key assumption that control GP practices reflect patterns of 'usual care' and that GP treated practices would have implemented policies with the same average effects on outcomes may be unwarranted.	Excluding GP practices which are in vanguard areas. However, this does not exclude other areas that are implementing other models of ICTs if these took place. Also, checking sensitivity to using a different set of control GP practices.
Threats to external validity		
SUS data may not fully reflect the hospital use of the population for which a GP practice is responsible, eg if relevant information in SUS is missing or incomplete.	Estimates of the impact of ICTs may be biased.	Missing data in SUS are assumed to be missing at random, so all GP practices should be affected similarly.
SUS data may not reflect the population if GP registration data are not up to date, eg if patients moved in or out of the area but did not change GP.	Estimates of the impact of ICTs may be biased.	Missing patient registration data are assumed to be missing at random, so all GP practices should be affected similarly.
Impact of ICTs may not have been fully realised within the post-intervention period or may vary over time.	Too short a post-intervention period may lead to erroneous results. Extrapolating the estimated impacts beyond the post-intervention period may be inappropriate.	Have long post-intervention period; evaluate the effect over several time periods to understand if and how any intervention impact varies over time. Selecting ICTs' practices with full-service uptake and where ICTs have been embedded for at least a year. Check with sites that initiatives are still ongoing past the end of the vanguards and for the entire post-intervention period.
These evaluations measure the effect of the particular roll-out, rather than effect of ICT more generally.	Estimates of the impact of ICTs may be biased.	Assume that variation in service delivery does not affect intervention impact. GP practices associated with unusual patterns to be scrutinised more closely.
Threats to construct validity		
Estimates may include the impact of other changes or initiatives that occurred during the study period.	Estimates of the impact of ICTs and concurrent initiatives may be biased.	Quantitative findings will be interpreted in the light of qualitative analyses describing other initiatives occurring at ICTs' practices.

The outcomes analysed do not represent all facets of the potential impact of ICTs.	Due to constraints with national data sets, some of the potential impacts of ICTs (eg on financial efficiency, patient satisfaction, staff morale and improvement in quality of care) will not be captured in the set of outcomes included here.	This phase of the analysis is intended to provide a broad snapshot of the effectiveness impact of ICTs and concurrent initiatives on key secondary care outcomes.
If an effect manifests itself only at the early or late stages of the post-intervention period, an average across the full post-intervention period would give an inadequate representation of the impact on patients of the intervention.	A misleading effect estimate would misrepresent the impact exerted over time by integrated care as it would implicitly assume that intervention maturity is reached upon roll-out and maintained throughout the post-intervention period.	Monthly estimates of treatment effect will be averaged over specific time spans throughout the post-intervention period.

General reporting considerations

Yearly and overall estimates of the ATT of the ICTs and concurrent initiatives with associated 95% confidence intervals and p-values will be reported for each outcome. These are calculated as described in the Statistical methods section. Plots of average outcomes in treated and counterfactual groups over the study post-intervention period will also be provided. Analysis with and without risk adjustment will be presented and the variables used in the risk adjustment model will be noted.

Secondary care resource utilisation definitions

Accident and Emergency (A&E) attendance

An A&E attendance is a non-duplicate visit by an individual to a hospital A&E department for a particular incident. A duplicate visit is defined as a recorded attendance by an individual to the same provider either at the same date and time as a previously recorded attendance, and where the primary diagnosis and treatment codes are the same; or within 1 hour of a previously recorded attendance. Depending on the analysis being undertaken, an A&E attendance may be further defined as one of the following:

- a non-duplicate, planned or *unplanned* visit
- a non-duplicate visit where the patient was *seen*
- a non-duplicate planned or *unplanned* visit where the patient was *seen.*

Avoidable admissions

An avoidable admission is an emergency admission for a condition that could have been managed or treated by timely or effective care within the community, hence being potentially avoided. Sets of clinical conditions which may lead to an avoidable admission include:

- A set of conditions that focus on older people experiencing health and social care – these include acute lower respiratory tract infections (such as acute bronchitis); chronic lower respiratory tract infections (such as emphysema and other chronic lung diseases); pressure sores; diabetes; food and drink issues (such as abnormal weight loss and poor intake of food and water due to self-neglect); food and liquid pneumonitis (inhaling food or drink); fractures and sprains; intestinal infections; pneumonia; and urinary tract infections.[13] An avoidable admission resulting from a condition in this set is referred to as a potentially avoidable admission.

- Ambulatory care sensitive (ACS) conditions are a set of clinical conditions for which the risk of emergency admission can be reduced by timely and effective ambulatory care.[14] Ambulatory care consists of primary care, community services and outpatient care.[15] Multiple definitions exist in the literature for ACS conditions[15]: the definition used by the IAU in this study will be the same as defined in the CCG Improvement and Assessment Framework (CCGIAF).[16] This framework was introduced in 2016/17 and was developed with input from NHS Clinical Commissioners, Clinical Commissioning Groups (CCGs), patient groups and charities. It was designed to play a part in the delivery of the *Five year forward view* for the NHS in England. Similarly to the NHS Outcomes Framework,[17] the CCGIAF differentiates between chronic and acute conditions:

  - Chronic ACS conditions: the definition of chronic ACS is the same as that for the NHS Outcomes Framework 2.3.i and the CCG Outcomes Indicator Set 2.6.[15] Among these conditions are epilepsy, diabetes and angina.[18]

  - Acute ACS conditions (also called urgent care sensitive conditions): acute ACS conditions are defined as unnecessary emergency admissions to hospital for conditions that should be dealt with effectively by the urgent care system without the need for admission to hospital. Such conditions include chronic obstructive pulmonary disease (COPD), cellulitis, deep vein thrombosis and falls.[16]

Length of stay following an emergency admission

Length of stay following an emergency admission is defined as the total number of nights spent in hospital following an emergency admission, calculated as the difference in days between the date of discharge and the date of admission. This is equivalent to the total number of emergency bed days. An admission and discharge within the same day will result in a length of stay of zero days.[19]

References

1. Lloyd T, Brine R, Pearson R, Caunt M, Steventon A. *The impact of integrated care teams on hospital use in North East Hampshire and Farnham*. Health Foundation; 2018 (www.health.org.uk/publications/impact-integrated-care-teams-hospital-use-north-east-hampshire-and-farnham).
2. Clarke GM, Vestesson E, Pariza P. *Statistical Analysis Protocol for the evaluation of Extensive Care Service and Enhanced Primary Care for Fylde Coast NHS Vanguard.* Health Foundation; 2020 (www.health.org.uk/sites/default/files/2020-01/fylde_coast_nhs_Vanguard_statistical_analysis_protocol.pdf).
3. Kumpunen S, Edwards N, Georghiou T, Hughes G. *Evaluating integrated care: why are evaluations not producing the results we expect?* Nuffield Trust; 2019 (www.nuffieldtrust.org.uk/resource/evaluating-integrated-care-why-are-evaluations-not-producing-the-results-we-expect).
4. Lloyd T, Steventon A. Statistical Analysis Protocol for an evaluation of integrated care teams as part of a primary and acute care system (PACS) in North East Hampshire and Farnham. Health Foundation; 2017 (www.health.org.uk/sites/default/files/2018-10/iau_nehf_sap_addendum.pdf).
5. Xu Y. Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models. *Polit Anal.* 2017; 25(1): 57–76.
6. Ryan AM, Burgess JF, Dimick JB. Why we should not be indifferent to specification choices for difference-in-differences. *Health Serv Res.* 2015; 50(4): 1211–35.
7. Robbins MW, Saunders J, Kilmer B. A Framework for Synthetic Control Methods With High-Dimensional, Micro-Level Data: Evaluating a Neighborhood-Specific Crime Intervention. *J Am Stat Assoc.* 2017; 112(517): 109–26.
8. O'Neill S, Kreif N, Grieve R, Sutton M, Sekhon JS. Estimating causal effects: considering three alternatives to difference-in-differences estimation. *Health Serv Outcomes Res Methodol.* 2016; 16: 1–21.
9. Office for National Statistics. Methodology note for the 2011 area classification for local authorities [webpage]. *ONS*; 2015 (www.ons.gov.uk/methodology/geography/geographicalproducts/areaclassifications/2011areaclassifications/methodologyandvariables).
10. Abadie A, Diamond A, Hainmueller J. Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *J Am Stat Assoc.* 2010; 105(490): 493–505 (www.tandfonline.com/doi/abs/10.1198/jasa.2009.ap08746).
11. King G, Zeng L. The dangers of extreme counterfactuals. *Polit Anal.* 2006; 14(2): 131–59.
12. Ferman B, Pinto C. Revisiting the Synthetic Control Estimator. Munich Personal RePEc Archive; 2016.
13. Care Quality Commission. *The state of health care and adult social care in England 2012/13.* CQC; 2013.
14. Sanderson C, Dixon J. Conditions for which onset or hospital admission is potentially preventable by timely and effective ambulatory care. *J Health Serv Res Policy.* 2000; 5(4): 222–30.
15. Bardsley M, Blunt I, Davies S, Dixon J. Is secondary preventive care improving? Observational study of 10-year trends in emergency admissions for conditions amenable to ambulatory care. *BMJ Open.* 2013; 3(1): 1–12 (www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3549201&tool=pmcentrez&rendertype=abstract).
16. NHS England: Analytical Team. *CCG improvement and assessment framework 2016/17: Technical annex.* NHS England; 2017.
17. NHS Digital. NHS Outcomes Framework [webpage]. Vol.2.4. Health and Social Care Information Centre. 2017 (https://digital.nhs.uk/data-and-information/publications/statistical/nhs-outcomes-framework).

18. NHS Digital Clinical Indicators Team. CCG Outcomes Indicator Set – June 2016 [webpage] (https://digital.nhs.uk/data-and-information/publications/statistical/ccg-outcomes-indicator-set/june-2016/ccg-outcomes-indicator-set---june-2016).
19. Department of Health. KH03 Quarterly bed availability and occupancy – Data definitions [webpage]. DoH; 2010 (www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2013/04/dh_124212-5.doc).