Improvement Analytics Unit

Statistical Analysis Protocol for an evaluation of the Mid-Nottinghamshire Better Together vanguard

Geraldine M. Clarke and Paris Pariza
February 2019

Addendum added in November 2019

Contents

Summary

Purpose of this document

This statistical analysis protocol describes in detail the proposed Improvement Analytics Unit (IAU) evaluation of the Mid-Nottinghamshire Better Together Primary and Acute Care Systems (PACs) vanguard Integrated Care Transformation Programme (ICTP). This document is intended to guide the analysis: it includes a summary of the background and context of the ICTP, the proposed evaluation design, statistical methods and the limitations of the analysis, and how these should be considered when interpreting the results.

This document has been agreed with the Mid-Nottinghamshire NHS vanguard team and written before the analysis begins, to ensure that all design and methods choices are made objectively, and are not influenced by what is found in the data. In rare instances, it may be necessary to make changes to the design of the study at a later stage; if so, this document will be appended accordingly. The IAU welcomes comments and questions on this document.

Purpose of this evaluation

The purpose of the evaluation is to feed back on the impact of the ICTP on secondary care activity in the Mid-Nottinghamshire Better Together vanguard region. The evaluation is intended to enable understanding of the impact of the different components of the ICTP in order to provide learning and improvement as well as assurance about continued investment in the service in terms of finance and quality.

What the evaluation will look at

The evaluation will study the impact of the ICTP on secondary care activity in the Mid-Nottinghamshire Better Together vanguard region between April 2013 and April 2018 for individuals aged 18 years and over. This is achieved by contrasting the change in outcomes for patients aged over 18 years and registered with a GP practice in the Mansfield and Ashfield (M&A) Clinical Commissioning Group (CCG) and Newark and Sherwood (N&S) CCG (the treated group) with that of similar patients aged over 18 years and registered with comparable GP practices in other parts of England (the control group), before and after the introduction of key interventions in the ICTP. After adjusting to take into account any underlying differences between the individuals in the control and the treated groups, the adjusted outcomes in the control group are assumed to represent the counterfactual outcomes for the treated group. In other words, they represent outcomes that would have occurred in the treated group if the treated group had not received the intervention. If the control group is a true counterfactual, the only difference between the treated and the control group is the treatment (intervention) itself, and the difference between a given outcome in the two groups provides an estimate of the causal impact of the intervention on that outcome.

The evaluation's primary outcomes will examine how the rate of Accident and Emergency (A&E) attendances, hospital admissions and outpatient attendances changed relative to their anticipated (i.e. counterfactual) level following the introduction of the ICTP. The evaluation's

secondary outcomes will examine impacts on the 4-hour A&E waiting time, the number of bed days for elective and emergency admissions, and the rate of admissions for specific conditions consistent with those targeted by the ICTP (including diseases of the respiratory system, cardiovascular diseases, chronic obstructive pulmonary disease (COPD), diabetes, disease of the musculoskeletal system and diseases of the eye and adnexa).

Data sources

The evaluation will use de-identified patient-level national Secondary Uses Service (SUS) administrative hospital data for England from 2 years prior to the start of the ICTP in April 2013 until the end of April 2018, as well as publicly available reference data. De-identified means that all direct identifiers (e.g. name, address, date of birth, NHS number for patients) are removed from the data. This reduces the risk that individual patients can be identified from the data.

Strengths and weaknesses of the evaluation

The evaluation will only study the impact of the ICTP on secondary care activity, which can be measured from national administrative hospital data (SUS data). Utilisation of other health and care services, impacts to quality of life, staff satisfaction, the quality of working relationships and other outcome areas that the vanguard has evaluated are outside the scope of this analysis, but could be investigated at a later stage. For a full picture of the impact of the ICTP in Mid-Nottinghamshire, the evaluation should be viewed in conjunction with the qualitative research carried out by the local evaluation of the Mid-Nottinghamshire Better Together vanguard.

The analysis will attempt to control for differences between GP practices in the treated and control groups by ensuring that the pool of GP practices in the control group are from CCGs that are similar to M&A and N&S GP practices in terms of both CCG and GP practice level influences. Further risk adjustment will be applied to control for differences in patient activity characteristics in the statistical analysis. However, there is still a risk that they are different in ways that cannot be observed (for example, in terms of their response to social isolation or receptiveness to new approaches to managing their conditions), leading to bias of estimates. There is also a risk of bias due to area-level differences in access, or exposure, to additional or alternative interventions.

The evaluation will study the impact of the ICTP over a period when it was continuously developing and estimates may not capture the full long-term effects of the individual components of programmes implemented in the later stages of the ICTP. The results are nonetheless expected to provide a robust summative assessment of the overall causal impact of the ICTP on hospital use that, together with the local evaluation and other evidence, will help the Mid-Nottinghamshire Better Together vanguard understand what is happening on the ground, assess what is working and identify potential areas for further investigation or improvement.

Background

Mid-Nottinghamshire

Health services for Mid-Nottinghamshire are coordinated by NHS Mansfield and Ashfield (M&A) CCG and NHS Newark and Sherwood (N&S) CCG. The M&A and N&S CCGs are part of Nottingham and Nottinghamshire Health and Care Sustainability Transformation Partnership (STP) (Figure 1), which is developing into an integrated care system (ICS).

M&A CCG consists of 27 GP practices; N&S CCG consists of 14 GP practices.* Together they serve a population of 328,782† individuals with an age structure slightly older than the national average. The older population are more likely to experience disability and long-term illnesses and there are increasing numbers of individuals with multiple comorbidites as well as individuals who need complex care. There are high levels of deprivation across both urban and rural areas, giving rise to an increasing demand for health and social care. The CCGs estimated that people with long-term conditions (LTCs) account for 50% of GP consultations and 70% of hospital inpatient bed stays, which translated into an expenditure of £4.45 million on unplanned admissions in 2011/12[1]. Overall, the CCGs have estimated that the increased costs could lead to a funding gap of approximately £140 million in 10 years' time[2].

Specific challenges faced by Mid-Nottinghamshire are:

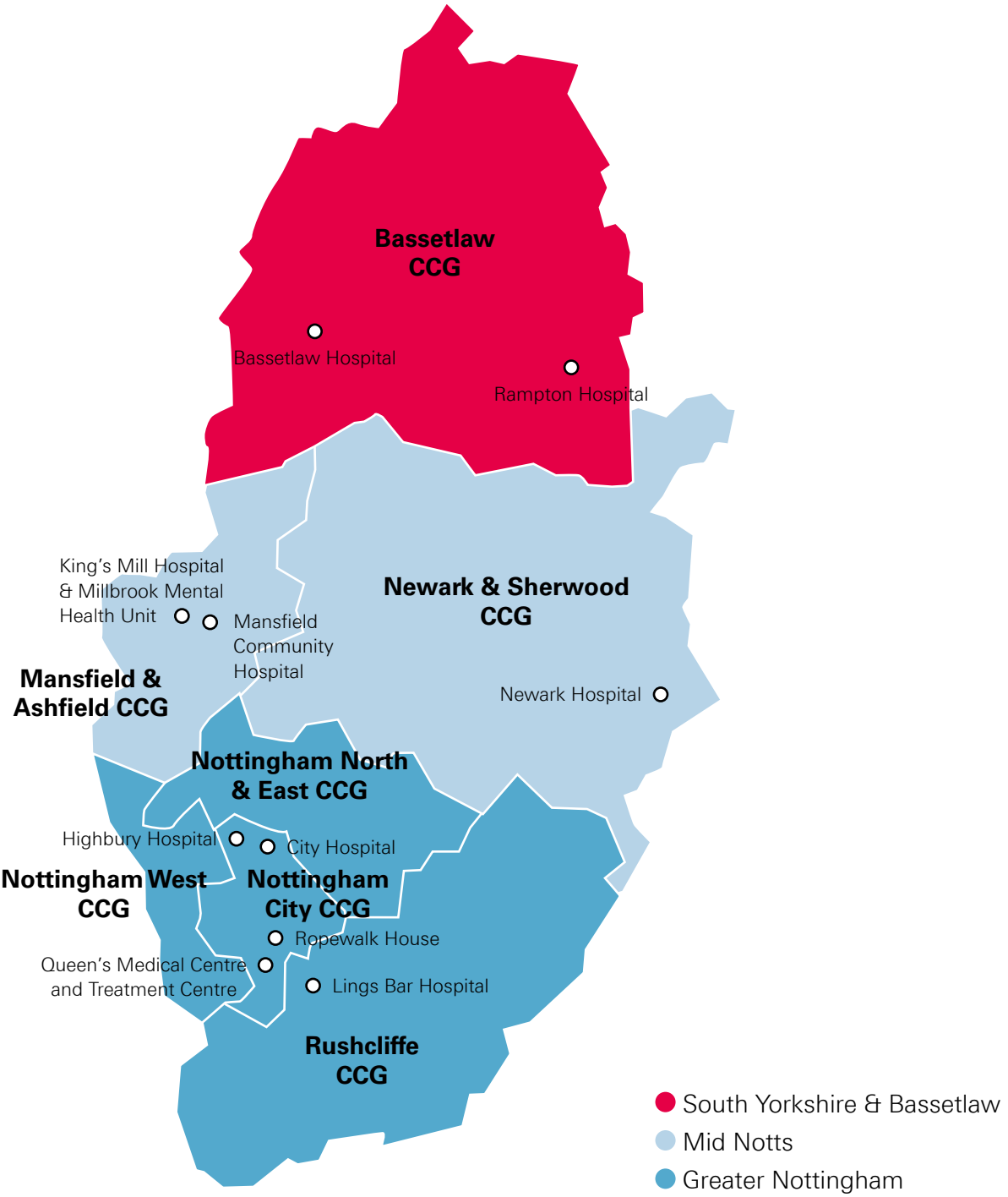
- increasing numbers of older people and people with complex long-term medical conditions
- communities with a diverse range of health care needs
- lack of coordination in the current health and care system, leading to a poor experience for patients and their families
- many patients have conditions which are not managed as well as they could be, and so often go to hospital when they could be better supported in a community setting or at home
- the cost of providing health care is rising faster than the funding received.

In January 2013, healthcare services in Mid-Nottinghamshire set out their vision and strategy for changes to health and care services over the subsequent five years. The need for better health and social care outcomes and improved patient experiences led to a radical change of care delivery. New models of care were proposed with the aim of integrating hospital, community, social and primary care services.

* https://www.nhs.uk/Services/Trusts/GPs/DefaultView.aspx?id=89801 and https://www.nhs.uk/services/trusts/gps/defaultview.aspx?id=89804

† ONS March 2018 mid-year estimate

Figure 1: Nottingham and Nottinghamshire Sustainability and Transformation Partnership



Mid Nottinghamshire Better Together Alliance

The Mid-Nottinghamshire Better Together Integrated Care Transformation Programme (ICTP) was established in 2013 as a partnership between M&A CCG and N&S CCG, seven NHS health providers and voluntary sector partners. In March 2015, Mid-Nottinghamshire Better Together was selected as an integrated Primary and Acute Care Systems (PACS) vanguard, one of the

first 29 New Care Model vanguards.* The vanguard funding was used to support the ongoing ICTP implementing new models of health and social care. An Alliance agreement contract was agreed from April 2016, entering the partners into a contractual joint venture. The Alliance-led ICTP is now a key component of the Nottingham and Nottinghamshire STP.

The ICTP aims to ensure that people across the Alliance region receive the best possible care in the community and in hospital, with high-quality, sustainable services delivered by a proactive, coordinated and properly resourced multidisciplinary community-based team. The original strategic objectives[2] were:

- 15% reduction in avoidable A&E attendances
- 19% reduction in non-elective admissions (generally urgent and emergency hospital admission)
- 0.5% reduction in non-elective bed days (generally patients admitted because of urgent and emergency referral)
- 25% reduction in admission to nursing and residential care homes
- 9.8% reduction in secondary care elective referrals (planned hospital day case visits or planned consultant appointments)
- 20% reduction in paediatric non-elective admissions (generally urgent or emergency hospital admissions).

In 2017, the Alliance refined the objectives. The 9.8% reduction in secondary care elective referrals was replaced with a 5% reduction in outpatient first appointments referred by GP, and 8% reduction in referrals from all sources. New objectives included a 6% reduction in follow-up outpatient appointments and an 8% reduction in elective admissions.

Integrated Care Transformation Programme

The Mid-Nottinghamshire Better Together ICTP[3] comprises four main work programmes: Proactive and Urgent Care, Elective (Early and Planned) Care, Mental Health and Community, and Maternity and Children. Key vanguard-funded interventions likely to have an impact on secondary care resource utilisation are concentrated in the Proactive and Urgent Care and Elective Care programmes (Figure 2: Chronology of key interventions in the Mid-Nottinghamshire ICTP).

Proactive and Urgent Care

Key vanguard-funded interventions in the Proactive and Urgent Care programme include:

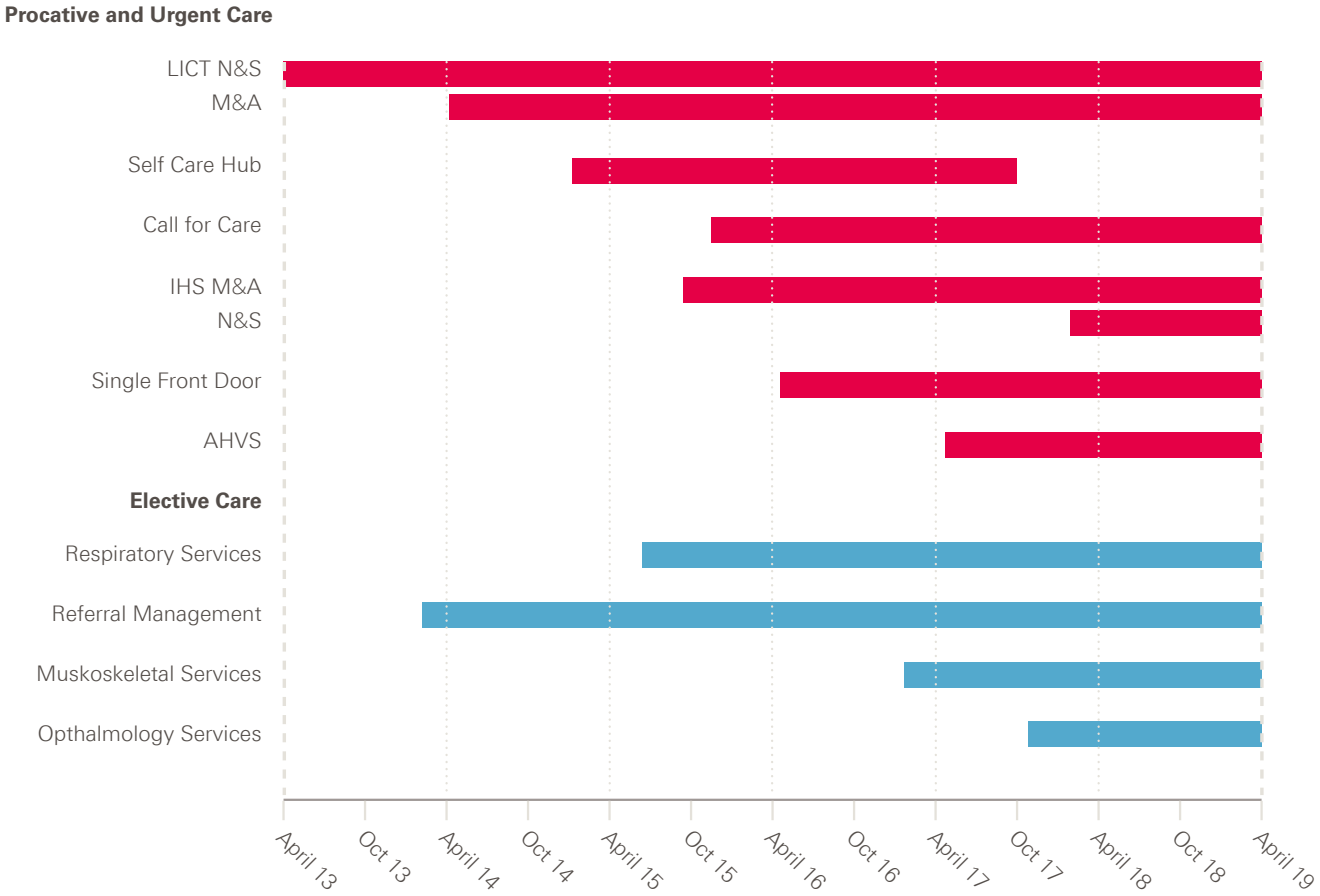
Local Integrated Care Teams (enhanced community services)

Local Integrated Care Teams (LICTs) have been implemented across the N&S CCG region since April 2013 and in the M&A CCG region since April 2014. The LICTs were introduced to provide enhanced community services in response to various challenges including large

* https://www.england.nhs.uk/new-care-models/about/primary-acute-sites/

numbers of patients receiving care from fragmented and disjointed teams, confusion between patients, carers and GPs about which services were available, pressures on primary care and out-of-hours services, and unplanned admissions for people with LTCs.

Figure 2: Chronology of key interventions in the Mid-Nottinghamshire ICTP



Note: Dashed grey lines indicate the start and end of the study period. Dotted lines indicate financial years for which we will report estimates of impact.
M&A - Mansfield and Ashfield CCG; N&S - Newark and Sherwood CCG.

The work of the LICTs is based on the Profiling Risk, Integrated Care and Self-Management (PRISM) model. The aim is to provide preventative care to patients aged 18 years and over who are deemed to be at high risk of future admission. The LICT facilitate monthly multidisciplinary team (MDT) team meetings at each GP practice to discuss these patients. Prior to each meeting, the LICT prepare a list of patients for discussion based on the GP Repository for Clinical Care (GPRCC)*, which uses a number of sources to identify key cohorts of patients with COPD, heart failure, stroke, dementia, who are on the end-of-life register, or who are in the top 2% of those most at risk of hospital admission. In practice, only those patients who have had any interaction with secondary care services have a risk score calculated, and an early summative report by Capita (available from the vanguard) found that many patients were also included for discussion as a result of local knowledge of the GPs and LICTs. Once an individual has been deemed suitable for the LICT, they are placed on a 'virtual ward' where MDTs manage the delivery of coordinated care in the

* The GPRCC is a local repository, which allows GP practices and local care teams to regularly review key cohorts of patients who have COPD, heart failure, stroke, dementia, are on the end-of-life register or on the 2% admission avoidance list.

patients' homes using the staffing, systems and daily routines of a hospital ward. Decisions to continue, review, enhance or discharge patients from the virtual ward are also taken at the monthly MDT meetings.

LICTs are co-located with other community nursing and therapy services wherever possible, allowing earlier mobilisation of health and social care, including intermediate care, so that hospital admissions are prevented and earlier discharge supported. A specialist LTC nurse can provide support to LICT teams either by supporting MDTs or providing a clinical intervention on a virtual ward patient. LICTs also work closely with community-based clinics (cardiovascular diseases, COPD, diabetes) that have consultant specialist support, community nursing teams and the voluntary sector. The LICT aims to provide a holistic assessment of needs and ensure patients are signposted to appropriate support services available in the community that historically the GP may not have been aware of.

There are three LICT teams in the N&S CCG region and four in the M&A CCG region, each based around populations of 30,000–50,000 individuals. The Mid-Nottinghamshire vanguard reported that 11,743 proactive actions, designed to reduce future risk of admission, were delivered between April 2014 and March 2016. A recent review concluded that the LICTs have the potential to reduce total bed days by 28,122 days, emergency department attendances by 4,687 and ambulance conveyances by 4,218 per year[4].

Self-Care Hub

The Self-Care Hub was introduced in September 2013 as a part of the LICT. It aims to increase the number of patients managing their LTCs in the community and has two main elements. Firstly, Self-Care Advisors work within the LICT as part of an MDT to support patients around self-care and support navigation to appropriate services. Secondly, the Self-Care Hub is a 'one stop shop' for service users, the public and professionals to provide advice, signposting and information and to facilitate access to relevant self-care support. In addition, the Self-Care Hub makes use of a virtual telephone platform to provide information on support available and to provide telephone coaching, telephone mentoring and Internet peer support groups. Services are mainly provided by Self Help UK.

The 2-year pilot period has seen the hub provider working collaboratively with specific disease specialists to enhance and encourage self-care in patients to improve outcomes. The service has helped promote self-care to both public and health care professionals across Mid-Nottinghamshire.

The Self-Care Hub was decommissioned in November 2017 after it was assessed as a cost ineffective model of delivery for self-care. The CCGs are now exploring the Patient Activation Measure (PAM) tool – a tool that enables health care professionals to understand a patient's activation level, or their level of knowledge, skills and confidence to manage their LTCs – as an alternative way of ensuring self-care is promoted.

Call for Care (care navigation service)

Call for Care is a care navigation service for health and social care professionals aiming to increase out of hospital urgent care and reduce unplanned acute activity. It is a dedicated call streaming service operating 7 days a week from 8am to 8pm for patients* with an urgent care need[5], providing clinical triage and a 2-hour response from community clinicians. Crisis response is provided by the Call for Care team for up to 48 hours. They can then refer individuals to the Intensive Home Support (IHS) service. GPs and care home staff are encouraged to use the service for all unplanned hospital admissions, except in case of clear life threatening conditions and for children.

An early implementation of Call for Care was introduced in November 2015 for use by East Midlands Ambulance Service staff and out-of-hours care professionals only. It was rolled out more widely in April 2016. It was originally set up to handle 11,600 calls per annum provided by a team of 24 full-time staff. For the year to February 2018, there were 6,046 calls leading to an estimated reduction of 2,297 A&E attendances and 608 hospital admissions.†

Intensive Home Support (Specialist Intermediate Care Service)

Intensive Home Support (IHS), previously called the Specialist Intermediate Care Service, is a community-based crisis response service operating 24 hours a day, 7 days a week, which provides time limited, goal focused rehabilitation, medical monitoring and nursing care for patients being discharged from hospital, or to prevent a hospital admission. IHS aims to enable a patient to remain living in his or her own home as independently as possible. Patients are referred via Call for Care. The LICTs can also refer to IHS to prevent a hospital admission.

IHS has only operated in the M&A CCG region since September 2015. It was extended to cover the N&S CCG region in January 2018. By the end of March 2017, there had been 719 referrals to the IHS team with an average length of hospital stay of 25 days[4].

Single Front Door

The Single Front Door intervention started in April 2016. The aim of this intervention was to simplify the access points for the public into urgent care and to integrate access for walk-in patients on the King's Mill Hospital site.

The co-located A&E and primary care service at King's Mill Hospital are accessed via a single entrance and a single reception[4]. Streaming is led by the emergency department staff as per national guidance and is based on jointly agreed streaming principles. Nurses supported by a GP undertake clinical triage to identify the level of patient need. A 'see and treat' model is in place, so that a patient can be discharged from the department if their needs can be met at the initial contact. In cases where this is not possible, any immediately necessary tests (e.g. X-ray, blood tests) are ordered.

---

* Patients aged over 18 years and registered with a GP in the Alliance regions. Patients with dementia are included, but people with other mental health conditions are excluded.

† Data supplied by Mid-Nottinghamshire Alliance team from Proactive and Urgent Care Dashboard for February 2018.

Currently up to 23% of patients are streamed to and discharged by the primary care service against a target of 20% nationally. This actively supports achievement of the 4-hour standard at Sherwood Forest Hospitals NHS Foundation Trust and NHS England has recommended the model as an alternative to the Luton and Dunstable model.*

Respiratory service

The respiratory service was introduced in the M&A CCG region in January 2017. It provides a home oxygen review and assessment service, pulmonary rehabilitation sessions and respiratory nurse led clinics.

Elective Care

The Elective Care programme comprises a collection of approaches designed to improve elective referrals and elective processes. Key vanguard funded initiatives include:

Referral management

Prior to February 2014, each CCG had its own processes for managing GP referrals. The CHEC Gateway Referral Process was managing GP referrals on behalf of N&S CCG from February 2012 until August 2017. To enable standardisation of processes the following changes, to be implemented over time, were proposed:

- peer to peer reviews and development of a clinical pathways website to support GPs to access collective information and experience to ensure the most appropriate referral
- standard referral template and guidelines to help reduce the time it takes to complete a referral and to ensure that common criteria are applied to all individuals
- referral management through an administrative team who will help patients to make their choice of time and place of appointment.

Musculoskeletal service

The musculoskeletal service is an integrated service providing a single point of access for patients to combine physiotherapy, rheumatology, chronic pain management and elective orthopaedic services into a single coordinated service. It enables early decision making through triage to identify those who can benefit from less invasive treatment and minimise waiting times to be seen by the most appropriate specialist as necessary. The service was first introduced in January 2017 and has been continuously developing since then, including self-referral to therapy.

Ophthalmology service

A community-based ophthalmology service has been developed to treat patients with eye conditions that require ongoing management. Previously these patients were referred to hospital for treatment. Community monitoring is intended to enable quicker appointments, reduce patient anxiety and convenience. Services were first rolled out in October 2017.

* See https://www.hospedia.com/luton-and-dunstables-a-and-e-target-success/

Other programmes

The Mental Health and Community programme includes workstreams such as proactive mental health, mental health core contract and community services core contract. Interventions of the programme for Maternity and Children include acute paediatrics, maternity service specification, community paediatrics and urgent paediatric surgery[2]. As well as vanguard funded schemes, the ICTP also includes other services, e.g. Proactive Care Homes Service, Ambulatory Emergency Care and Integrated Urgent Care[4].

Intended impact of interventions

The overarching aim of the ICTP was to create a strategically different model of care, with more care being provided outside of acute hospital settings. An additional aim was to create an environment where care professionals were encouraged to work together across organisational and professional boundaries to deliver improved outcomes and make efficiency savings. As outlined in the strategic objectives, the activity shift was predicted to result in an estimated 15%, 19% and 0.5% reduction in A&E attendances, A&E admissions and emergency bed days, respectively. Expectations are that delays in the provision of care will be reduced significantly and planned care will be delivered in a more effective and sustainable way as a result of the ICTP activities.

Methods

Study design

We use national SUS and publicly available reference data to evaluate the impact of the ICTP for individuals aged over 18 years on hospital activity during the study period of April 2013 to April 2018. We do this by contrasting the change in outcomes for patients aged over 18 years and registered with a GP practice in the M&A and N&S CCGs with that of patients aged over 18 years and registered with selected GP practices in comparable CCGs from other parts of England before and after the introduction of key interventions at three different time points.

Individual outcome and other activity and demographic (referred to as covariate) data are collected at the patient level as monthly activity counts and then aggregated to monthly activity counts at the GP practice level. The treated group comprises all 47 GP practices that are members of M&A or N&S CCGs; the control group comprises comparable GP practices from comparable CCGs in other parts of England. The final analysis data set consists of aggregated outcome and covariate activity data for all treated and control group GP practices as monthly series for 24 months before and 60 months after April 2013. Aggregated outcomes between GP practices in the treated group and the control group will be compared with adjustment for differences in patient mix, and pre-intervention trends in outcomes, across time to provide an estimate of the counterfactual (risk adjusted) outcome for M&A and N&S CCGs.

Difference-in-Difference (DiD) methods are a traditional approach to estimating treatment effects in this setting where we have repeated measurements on the same units at different time points pre- and post- an intervention of interest. DiD assumes that the post-intervention experience of patients in the control group represents an appropriate counterfactual for patients in the untreated group, commonly referred to as the parallel trends assumption. However, if there are unobserved time varying effects, estimates from DiD may be biased[6,7]. Synthetic control methods, which relax the parallel trends assumption and allow for the effects of both observed and unobserved confounders to vary over time, offer an alternative to DiD. For this analysis, we consider one of two synthetic control methods, Generalised Synthetic Control (GSC)[8]and Micro Synthetic Control (MSC)[9].

Study cohorts

Target population

Here we adopt a population approach to the evaluation to provide an overall picture of the effect the ICTP had on health care for the local Mid-Nottinghamshire population. This approach is consistent with the aim of M&A and N&S CCGs to create an accountable care organisation with other CCGs in Nottinghamshire. We examine hospital use for the whole population of Mid-Nottinghamshire (the treated region) and compare that to hospital use for the whole population of a control region. The control region is anticipated to provide an estimate of hospital use that would have been expected in Mid-Nottinghamshire in the absence of the ICTP. The treated region is represented by the group of GP practices belonging to either M&A CCG or N&S CCG, and the control region is represented by a group of GP practices from comparable CCGs in other parts of the country. The target population is therefore all patients registered with a GP practice that belongs to either M&A CCG or N&S CCG (the treated group), or to a GP practice in another part of England (the control group). Note that this approach excludes hospital activity for individuals who are unregistered, or who are registered with other GP practices not in the treated or control group.

Study cohort

The study cohort is defined as the patients in the target population during the study period. Although the analysis uses activity data for continuous inpatient spells (CIPS), outpatient visits and A&E visits, the study cohort is the entire target population regardless of whether they actually attend for treatment, although the outcomes of interest will relate to their actual utilisation of secondary care. This allows us to capture reductions in utilisation as well as changes in the composition of utilisation.

Study outcomes

The ICTP is anticipated to impact on a range of outcomes (see Background section). Primary and secondary outcomes* that will be examined here are listed below.

Primary outcomes:

- rate of A&E attendances per head of GP practice size per month
- rate of emergency admissions per head of GP practice size per month
- rate of elective admissions per head of GP practice size per month
- rate of avoidable emergency admissions per head of GP practice size per month
- rate of outpatient appointments (excluding 'did not attends', and defined using code Attended=5 or 6 in SUS data) per head of GP practice per month.

Secondary outcomes:

- average length of stay of emergency admissions (for all admissions initiated in the month, even if they last longer than the month)
- average length of stay of elective admissions (for all admissions initiated in the month, even if they last longer than the month)
- proportion of A&E attendances seen within 4 hours
- subject to sufficient numbers of individuals and data quality, we will also look separately at the rate of emergency admissions and rate of elective admissions per head of GP practice size per month for ICD-10 Version 2015 primary admission diagnosis code, for:

– diseases of the respiratory system
– cardiovascular diseases
– COPD
– diabetes
– musculoskeletal system
– the eye and adnexa.

Outcomes are 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. Under the assumption that the ICTP did not alter the size of the GP practice population (only GP practices open for the duration of the study are included and so are expected to be fairly stable during this time), expressing the outcome as a rate in this manner mitigates the potential bias that would arise if the intervention altered the case-mix of secondary care utilisation, while ensuring the outcome is directly comparable across GP practices.† Considering these outcomes together will provide an overview of the effects of the ICTP on the utilisation of secondary care.

---

* For definitions of items highlighted in bold, please see the SAP Appendix.

† We could look at activity counts instead of rates, but then the counts across different GP practices may be very dissimilar if the population is much larger/smaller. The synthetic control method by default requires that the outcomes of the treated units can be approximated by interpolation of the donor units' outcomes.

Sources of data

GP practice-level and CCG level reference data

CCG level data for all CCGs in England and GP practice level reference data for all GP practices in the target population will be obtained from data that is publicly available at the GP practice and CCG level over the study period. This is used to summarise key characteristics of CCGs and GP practices. We will map information that is only available at CCG level to GP practice level, and vice versa, by weighting according to the number of registered patients in the practice that live in each CCG area. CCG level reference data is obtained on a yearly basis; GP reference level data is obtained on a quarterly basis. Since data for some variables may not be available in all years or quarters, we will proxy this missing data using data from the closest available year or quarter. See Table 1: Variables relating to the characteristics of CCGs and GP practices for details and a list of sources.

Table 1: Variables relating to the characteristics of CCGs and GP practices
Variables are collected from indicated source at indicated date and mapped to produce monthly series of data across the study period aggregated at either GP or CCG level. Variables available at the CCG level only are mapped to GP level by weighting according to the number of registered patients within each CCG area, and vice versa.

Variable	Description	Date of collection	Level of collection	Source
Population size	Number of registered patients	Annually 2011–2017	GP practice	*NHS Digital. Number of patients registered at a GP practice.* 2011–2015
Age	Proportion of registered patients aged <5, 5–14, 15–24 and 75+ years	Annually 2011–2015	GP practice	
Gender	Proportion of registered male patients		GP practice	
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. *Census*. 2011
Education	Proportion of registered patients with at least third level education (two or more A-levels or equivalent)		GP practice	
Population density	Number of persons per hectare in the nearest electoral ward		GP practice	
Socioeconomic deprivation	Weighted average of Lower Super Output Area (LSOA) level index of multiple deprivation (IMD) scores according to LSOA of GP registered patients	2015	LSOA	Department for Communities and Local Government. *English indices of deprivation*. 2015
Health deprivation	Weighted average of LSOA level IMD scores on health deprivation according to LSOA of GP registered patients		LSOA	
Number of full time equivalent general practitioners	Number of full time equivalent general practitioners per 1,000 people in the registered GP population	2015	GP practice	NHS Digital. *General and personal medical services.* 2015

Number of care home beds	Number of care home beds (residential and nursing) per 1,000 registered GP patients	2015	GP practice	Care Quality Commission. *Register of Care Homes.* 2015
Quality and Outcomes Framework (QOF) achievement scores	Total of achievement scores on all QOF indicators across cardiovascular, respiratory, high dependency and other LTCs, and musculoskeletal QOF indicator groups	Annually 2011–2015	GP practice	NHS Digital. Quality Outcome Framework. 2013–2015
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+ only), diabetes, palliative care, osteoporosis (50+ only) and rheumatoid arthritis (16+)		GP practice	

Activity data

Hospital activity data is obtained from de-identified (i.e. anonymised in line with the Information Commissioner's Office code of practice on anonymisation) SUS data. SUS is a national, person-level database that is closely related to the widely-used Hospital Episode Statistics (HES). 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 a secure environment based at the Health Foundation. 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 the programme oversight group and by information governance experts at NHS Digital.

For this study, we will use data on A&E visits, inpatient and outpatient attendances from April 2011 to April 2018. The data will be pre-processed to ensure that variables that are not expected to change, e.g. gender and ethnicity are consistently recorded across time.

Only activity data for patients who are registered with a known GP practice are included. This is to ensure that activity can be ascribed to a GP practice. Hence, any records that are missing a de-identified NHS number are excluded. A&E visits for a patient who left before being seen or refused treatment, or where the visit is a duplicate, are excluded. Outpatient appointments where the patient did not attend, or where the outpatient appointment is a duplicate, are excluded. Inpatient data is 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, outpatient appointments 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 the statistical modelling.

Setting

Study periods

The study period includes a pre-period of 24 months from April 2011 to March 2013, and a post-intervention period from April 2013 to the end of April 2018.

Baseline variables

GP reference baseline variables

The analysis will first identify a set of GP practices that are considered to be comparable to the GP practices in the treated group in terms of aggregate characteristics in the pre-intervention period. Baseline variables listed in Table 1: Variables relating to the characteristics of CCGs and GP practices obtained at an annual level for each GP practice for the financial years 2011/12 and 2012/13 will be used to identify these GP practices, which will then comprise the control group, or 'donor' pool, of GP practices against which the treated group of GPs will be compared. Since data for some variables may not be available in both years, we will proxy missing data using data from the other year. The use of annual data reflects data availability and a belief that the variables do not exhibit a lot of variation within a year.

Activity level baseline variables

After identifying the donor GP practices, the A&E records, CIP spells and outpatient records for all patients registered to these GP practices will be identified. In addition to patient outcomes, the baseline variables listed in table 2 will also be collected to be used for risk adjustment to account for differences in the patient demographics, comorbidities and prior hospital use across GP practices over time, as described in the statistical analysis. Data is first collected at the patient level and then aggregated to GP practice level in monthly series.

Quan (2005) provides algorithms that classify patients' comorbidity into the 17 categories of the Charlson comorbidities and into the 30 categories of the Elixhauser comorbidities using their ICD-10 codes[10]. A number of studies suggest that the Elixhauser comorbidity categorisation leads to a superior comorbidity risk-adjustment model[11]. Bottle et al (2014) conducted a systematic review of multiple comparison studies on comorbidity measures/indices in use with administrative data and conclude that, for general purposes, comorbidity is currently best described by the set of 30 Elixhauser comorbidities plus dementia[12]. We will include the average Elixhauser comorbidity score for all patients from a GP practice with hospital activity in a given month in the risk adjustment model.

Table 2: Variables relating to the characteristics of patients at GP practices
Variables are first collected at patient level per month from SUS data and then aggregated to produce total activity counts per GP practice per month. The aggregate counts indicate key characteristics related to hospital utilisation of patients registered at each GP practice each month.

Variable	Description
Age	Total number of records for patients aged <5, 5–14, 15–24 and 75+ years
Gender	Total number of records for male patients
Ethnicity	Total number of records for patients with self-reported race white, black, Asian and mixed
Number of hospital admissions per primary diagnosis code	Total number of records for each ICD-10 Version 2015 classification code I – XXII
History of elective admissions	Total number of elective admissions in the preceding 24 months for all patients with activity in the month
History of emergency admissions	Total number of emergency admissions in the preceding 24 months for all patients with activity in the month
History of A&E visits	Total number of A&E visits in the preceding 24 months for all patients with activity in the month
Elixhauser index >=2	Total number of patients with an Elixhauser comorbidity score (number of comorbidities defined by Elixhauser et al 14]) greater or equal to 2 in the preceding 24 months for all patients with activity in the month
History of primary diagnoses	Total number of patients with an indicator for a primary diagnosis in the preceding 24 months for all patients with activity in the month (one variable for each primary diagnosis based on summary hospital level mortality indicator categories)

Statistical methods

Identifying selected GP practices in the control group

We aim to select GP practices in the control group that are comparable to those in the treated group across the two years prior to the start of the ICTP in April 2013. We aim to ensure similarity at both the CCG and GP practice level. We do this in two steps:

1) **Select a set of comparable CCGs.** The comparable CCGs comprises a subset of the 209 CCGs in England that are most similar to M&A and N&S CCGs according to key demographic and socio-economic characteristics in the 2 years prior to the start of the ICTP. From the set of 209 CCGs, we exclude:
   a. CCGs in London, because they are assumed to serve inherently different populations to the Mid-Nottinghamshire CCGs;
   b. CCGs in neighbouring Mid-Nottinghamshire, because they may have spillover effects;

c. 59 CCGs participating in New Care Model vanguards (see **Supplementary figure 1** in the SAP Appendix), because they may be exposed to similar interventions.

From the remaining CCGs, we select the 20 that are most similar to M&A CCG, and the 20 that are most similar to N&S CCG (see below). Some of the 20 selected for M&A, and the 20 for N&S, are expected to be the same, resulting in a total of up to 40 comparable CCGs. 20 is chosen arbitrarily to allow for a sufficient number of constituent GP practices, whilst ensuring manageability.

2) **Select a set of comparable GP practices from the comparable CCGs.** The comparable GP practices will comprise a subset of all those in the comparable CCGs. After identifying all the GP practices in the comparable CCGs, we exclude any that opened later than two years prior to the start of the ICTP in April 2013, or which closed before the end of the study period in April 2018. We also exclude any GP practices with a population size outside the range of population sizes of GP practices in the treated group. From the remaining, we select 1,000 with similar values of key demographic and socio-economic characteristics as those in the treated group in the 2 years prior to the start of the ICTP. 1,000 units was chosen to ensure sufficient data dimensionality for MSC and GSC methods. We will assess sensitivity to the number of GP practices included in the treated group.

To identify CCGs or GP practices with similar values of key characteristics as described, we apply the method used in NHS England's Commissioning for Value tool* to assess similarity in terms of key characteristics (Table 1: Variables relating to the characteristics of CCGs and GP practices). Where available, annual data for each CCG or GP practice for the financial years 2011/12 and 2012/13 will be included.

Similarity is assessed by calculating the squared Euclidean distance (SED) between each treated and each untreated unit (CCG or GP practice) across these variables. Lower SEDs between units indicate greater similarity.† Since the variables were measured on different scales and hence were not directly comparable, the data will first be standardised using inter-decile range standardisation‡ as used by the Office for National Statistics and in the Commissioning for Value tool, prior to calculating the SED[13,14]. Since many variables are included, some of which may be closely related to each other, the variables are also weighted according to how predictive they are of the rate of hospital admissions in 2012/13 (controlling for the other variables). The weight given to each variable is determined by its squared standardised coefficient in a regression of the rate of hospital admissions in 2012/13 on the variables for the preceding year[15]. Variables expected to receive greater weight will be the Quality and Outcomes Framework (QOF) achievement scores and the past rates of elective and emergency admissions.

In the Commissioning for Value tool, the similarity measure is based on a particular point in time. Here we are using longitudinal data and must decide whether to focus on (a) similarity at a particular point in time, e.g. immediately preceding the ICTP or (b) to incorporate multiple periods within the similarity measure. One potential drawback with the former

---

* The Commissioning for Value tool is used to determine the 10 most similar CCGs in England for a given CCG. See https://www.england.nhs.uk/rightcare/products/nhs-rightcare-intelligence-tools-and-support

† Let $\tilde{x}_{ki}$ represent the standardised version of x_{kj} where x_{kj} is the k^{th}, baseline variable in unit i, $i=1, \ldots I$, $k=1,\ldots,K$. Here a unit represents a GP practice. Then the SED between unit i and unit j is calculated across K baseline variables as $SED_{ij} = \sqrt{\sum_{k=1}^{K}(\tilde{x}_{ki} - \tilde{x}_{ki})^2}$ for $i,j<I$.

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

approach is that ensuring similarity at a single point in time would not correct for different time trends for the units, which is problematic here. Here we incorporate multiple periods by calculating the distance measures using annual estimates of each variable across financial years 2011/12 and 2012/13. The overall measure of similarity between two units over the pre-intervention years is then computed as the geometric mean of the SED in 2011/12 and 2012/13. This aims to ensure that the 'most similar' untreated units were similar to the unit across the pre-intervention period.

The nearest *U* untreated units to each of the treated units are then selected, where U is chosen at each step to ensure that there are a total of approximately 40 CCGs in the set of comparable CCGs, and 1,000 GP practices in the set of comparable GP practices. U is unknown until the analysis is run as some untreated units may be similar to more than one treated unit.

Counterfactual analysis

One of two analytic approaches will be used to estimate the effects of the intervention using a panel dataset of aggregated outcomes and covariate data from treated and control group GP practices as monthly series between April 2011 and March 2018: (a) MSC and (b) GSC. MSC and GSC build on the Original Synthetic Control (OSC) method[16,17], which offered an alternative to the commonly used Difference in Difference (DiD) estimator. MSC and GSC both depend on an assumption of weak serial dependence of error terms. Unpublished work by the IAU shows that in the presence of strongly serial correlation of error terms, the MSC estimator is a more efficient estimator; otherwise the GSC estimator has been shown to more efficient. Our analysis will check for the presence of serial correlation and use the GSC estimator unless there is evidence of high serial correlation.

Original Synthetic Control (OSC)

The central idea of a synthetic control approach is to construct a weighted combination of the units in the untreated group to represent a new group whose hospital use represents that which would have been expected in Mid-Nottinghamshire in the absence of the ICTP, the so-called 'counterfactual'. The weights are chosen so that the treated group and the synthetic control group have similar values of the outcome and covariates over a pre-intervention period. A comparison between the outcomes in the synthetic control group and the average of those in the treated group provides a risk adjusted estimate of the treatment at each post-intervention time point.

DiD is commonly used to evaluate an intervention in the panel data setting that we have here. DiD depends on the assumption that the outcomes follow parallel trends over time. This means that there are no undetected variables that vary in the effect they have on the outcomes in a treated or a control group unit over time. For example, if an intervention in another CCG has a varying impact on outcomes at one or more GP practices in the control group over the pre-intervention period, then the parallel trends assumption will be violated. Since our analyses will use a 2-year pre-intervention period, the parallel trends assumption is unlikely to hold. OSC relies on the alternative assumption that, given the pre-intervention outcomes, the post-intervention outcomes are independent of treatment assignment. In other words, the

gains or losses as a result of the ICTP would be the same in any region with the same pre-intervention history. This allows for the effects of unobserved predictors of the outcomes to vary over time.

However, the OSC method is not designed computationally for high dimensional data, which are increasingly commonplace in scientific fields and, when applied to low dimensional data, the uncertainty estimates are cumbersome and not easily interpretable[8].

See 'Analysis methods' in the Appendix for technical details for DiD and OSC estimators.

Micro Synthetic Control (MSC)

The recently developed MSC method[9] generalises the OSC method to allow for multiple outcomes. Like the OSC, it constructs a weighted combination of the treated units to represent the counterfactual and is designed to provide an estimate of the average treatment effect across all treated units,. Moreover, it is designed specifically for high-dimensional data, as we have here with multiple treated units, outcomes and covariates. To calculate the weights for the synthetic control efficiently, the MSC method exploits methods commonly used in the analysis of surveys for the reweighting of a survey sample according to known characteristics from a target population. When the outcomes and covariates of the treated units are fundamentally different from the untreated units, the MSC estimator will not necessarily return weights that have practical utility, i.e. they cannot be used to accurately represent the counterfactual, and resulting estimates may be biased.

For each study period, we apply the MSC estimator to each outcome separately including all baseline covariates for risk adjustment and a pre-intervention period of 2 years prior to the start of the study period. The weights will then be chosen so that the synthetic control has the same value of the outcome and covariates as the average of those in the treated GP practices at each time point in the two years prior to the start of the study period. The effect of the ICTP on the average outcome across all treated GPs is estimated at each post-intervention time point by comparing the observed averages to those of the synthetic control. Estimates are then averaged across all post-intervention time points to provide an estimate of effect called the average treatment effect on the treated (ATT). Significance of estimates will be estimated using placebo groups generated through permuation[9]. Estimates using the MSC estimator are made using package 'microsynth' in R.*

See 'Analysis methods' in the Appendix for technical details.

Generalised Synthetic Control (GSC)

The GSC method, introduced by Xu et al.[8], of which DiD is a special case, also generalises the OSC method. It allows for multiple treated units and variable treatment periods. Despite its name, GSC is not a traditional synthetic control method. Rather, it combines the efficiency gains of interactive fixed effect (IFE) models with insights from the synthetic control methodology, allowing for unobserved covariates to have time varying effects and enabling

* https://cran.r-project.org/web/packages/microsynth/index.html

separate estimates for each treated unit. It has the advantage over IFE that it is unbiased even in the case of treatment effect heterogeneity and maintains the unbiasedness of DiD but is more precise if the underlying model is correctly specified[8].

As for MSC, we apply the GSC estimator to each outcome separately including all baseline covariates for risk adjustment and a pre-intervention period of 2 years prior to the start of the study period. The effects of the ICTP on each outcome at each treated GP practice at each post-intervention time period are then estimated. Estimates can be averaged across all treated GP practices to provide an estimate at each post-intervention time period and further, across all post-intervention time periods, to provide an estimate of the ATT. Significance of estimates will be calculated using bootstrapping methods[8]. Estimates using the GSC estimator are made using the 'gsynth' package in R.*

See 'Analysis methods' in the Appendix for technical details.

Diagnostics

After identifying the selected GP practices in the control group

To assess whether the control group is more similar to the treated group than the full set of GP practices, the average values of each GP reference baseline over the pre-intervention period will be compared between the GP practices selected for the control group, the GP practices not selected for the control group and the GP practices in the treated group. The standardised differences between the average value of each variable in the control and treated groups over the pre-intervention period will also be calculated. Inevitably there is a trade-off between the imbalance we are willing to tolerate and the number of GP practices included in the control group. The impact of imbalances in the standardised variables will depend on the extent to which the variables correlate with outcomes, however we would envision a difference of less than 0.2 (0.1) being an indicator that the control group GP practices are (very) similar to the treated group GP practices. If a particular variable is responsible for a lack of similarity, its importance will be reconsidered. The GP practices will be identified for inclusion in the control group without reference to the outcome data to ensure this is carried out objectively.

Pre-analysis diagnostics

After aggregating the outcomes for each month to the GP practice level:

- The trend in the unadjusted outcomes of the control group GP practices over time will be compared to the trend of the treated group GP practices to assess whether the treated GP practices lie within the convex hull of the control group GP practices, i.e. whether interpolation of the outcomes of the control group GP practices can reasonably be expected to approximate the average outcomes of the treated group GP practices. If not, the MSC estimator may return impractical weights and resulting estimates may be biased.

* https://cran.r-project.org/web/packages/gsynth/index.html

- The average trend in unadjusted outcomes between control group and treated group GP practices over time will be compared to assess the plausibility of the parallel trends assumption.

A time series regression analysis will be used to check for serial auto-correlation of outcomes using the Durbin-Watson test. If significant autocorrelation is detected, we will use the MSC estimator; otherwise we will use the GSC estimator.

Diagnostics of synthetic control weights for the MSC estimator

The pre-intervention fit of the synthetic control for each outcome will be assessed by calculating the Root Mean Square Error (RMSE) between the value of the outcome in the synthetic control and the average value of the outcome across all the treated group GP practices. If there is evidence of a poor fit, or that the treated units lie outside the convex hull of the controls, reducing the number of constraints on the synthetic control, e.g. removing or averaging covariates or outcomes over pre-intervention periods, will be considered in order to improve the fit.

Subgroup analyses

The impact of the ICTP on the primary outcomes will be estimated separately for each of the two CCGs.

Sensitivity analyses

A number of sensitivity analyses will be conducted to assess whether results are sensitive to the assumptions made in the baseline analysis.

1. To assess the sensitivity of results to the units included in the donor pool, the analysis will be repeated after varying the number of GP practices included in the donor pool and assessing the impact on covariate balance and on estimates.
2. To assess the sensitivity of results to the models used to risk adjust outcomes, we will compare estimates without risk adjustment to those after risk adjustment.
3. The sensitivity of estimates to changes in the length of the pre-treatment period will be assessed by varying the length of the pre-intervention period and/or collapsing data into quarterly, rather than monthly, intervals.

Limitations and sources of bias

Threats to validity

Internal validity

Threats to internal validity include:

- **GP practices in the control group have fundamentally different outcome and/or covariate values to those in the treated group.** In this case, it may not be possible to find a suitable counterfactual and estimates may be biased. We mitigate this bias by making sure that GP practices in the control group have similar CCG level and GP level reference characteristics. In particular we exclude London CCGs, which are assumed to serve inherently different populations to the Mid-Nottinghamshire CCGs. We will explore differences in the activity level covariates between GP practices in the treated and control groups to identify any that differ significantly. We will also assess the sensitivity of results to changing the number of GP practices included in the control group.
- **The risk adjustment model does not accurately account for differences in the intervention and control groups.** GP practices are selected to have similar values of key reference variables at both CCG and GP level, which mitigates against large differences between treated and control groups. Sensitivity of estimates to variables included for risk adjustment will be explored.
- **GSC and MSC methods may be unable to find a feasible solution.** A drawback of methods is that a feasible solution may not exist. Feasibility is data dependent, therefore we are limited in terms of our ability to predict when a solution will, and will not, be feasible. Feasibility is more likely in setting with large numbers of treated and untreated units, hence by working with multiple GP practices in both treated and control groups, we are mitigating this risk. If no solution can be found for a given outcome, we will report results according to a DiD framework.
- **The methods select a counterfactual that unduly reflects noise in outcomes.** GP practices in the control group may only appear to have similar pre-intervention outcomes to those in the treated group due to random variation. In this case, the similarity may not persist into the post-intervention period, leading to biased estimates of intervention effects. Abadie et al.[17] report that as the number of pre-intervention periods increases, the bias of synthetic control estimates shrinks towards zero. However, they do not provide criteria by which to assess whether the number of periods available is sufficient. By using a long pre-treatment period, we reduce the potential for this bias. The sensitivity of estimates to changes in the length of the pre-treatment period will be assessed.
- **Interventions that influence outcomes in GP practices in the control group may have occurred in the pre-intervention period.** An identifying assumption of our analysis is that outcomes in the GP practices in the control group are similar to those in treated groups in the pre-treatment period and that this similarity would have persisted into the post-intervention period. If this likeness is due to other interventions in the control group GP practices, it makes this assumption less plausible. We mitigate this

bias by excluding GP practices from the control group if they belong to CCGs that are participating in the New Care Model vanguard programme, minimising the risk that major policy interventions influenced outcomes in these control groups during the pre-intervention period.

- **Interventions that influence outcomes in GP practices in the control group may have occurred in the post-intervention period.** As above, this would violate the identifying assumption unless we believe that these interventions formed part of 'usual care' and that the treated group GP practices would have implemented policies that would have had the same average effect on outcomes. GP practices in the control group will be examined to minimise the risk that they implemented major policy interventions influencing outcomes in the post-intervention period.
- **Effects of interventions in GP practices in the area surrounding the Mid-Nottinghamshire Better Together vanguard may spill over into nearby GP practices and vice versa, reducing estimates of effect.** To mitigate this bias, we exclude all GP practices belonging to neighbouring CCGs from the control pool.

External validity

Threats to external validity include:

- **SUS data may not fully reflect the population for which Mid-Nottinghamshire Better Together is the responsible vanguard area.** For instance, relevant information in SUS may be missing or incomplete. In the analysis, missing data is assumed to be missing at random; it if is not then estimates of effect for the full population may be biased.
- **The effects of the ICTP may not have been fully realised within the study period(s) or the effects may vary over time.** Caution is therefore advised in extrapolating the estimated effects beyond the study period(s).

There are also a number of threats to external validity in relation to extrapolating the effects estimated for the Mid-Nottinghamshire Better Together vanguard as an estimate of the effects of similar interventions in other regions. The ICTP is multifaceted and depends on the societal, economic, health system and environmental context in which it is delivered, making its impact unique and affecting how results can be generalised to other settings.

Statistical conclusion validity

The GSC and MSC estimators make a variety of assumptions. For any estimator there is a risk of drawing wrong conclusions if the assumption underlying the method does not hold. Standard diagnostics will be performed to check the underlying assumptions of the method used.

Construct validity

Threats to construct validity include:

- **We will not know to what extent effects are attributed to different aspects of the intervention.** Results will estimate the effect of the evolution of the ICTP over the study period.
- **Estimates may be biased if there are large numbers of patients who are unregistered, or who are registered at a GP practice outside of Mid-Nottinghamshire, and who seek hospital treatment at a hospital or trust in Mid-Nottinghamshire.** These individuals may have been un-impacted by the interventions taking place as part of the ICTP and their activity may dilute observed effects. To investigate the extent of this bias, we will report the number of patients with hospital activity in a Mid-Nottinghamshire hospital or trust who are unregistered, or who are registered at a GP practice outside of Mid-Nottinghamshire.
- **Estimates may be harder to detect, if there are large numbers of patients who are registered at a GP practice in Mid-Nottinghamshire but who seek hospital treatment at hospitals or trusts outside of Mid-Nottinghamshire.** Their hospital activity will not be detected in the treated group and any changes in hospital activity because of the intervention will not be captured in the treated group. Further, if these changes are captured in the control group instead, then estimates will also be biased. To mitigate any bias, we exclude all GP practices in neighbouring CCGs from the control pool.
- **Estimates may include the impact of other changes that occurred in the Mid-Nottinghamshire Better Together vanguard region in the post-intervention period.** We partially address this concern by considering three separate study periods to try to isolate the impact of key interventions.
- **Estimates may not be properly risk adjusted if the risk adjustment model is not correctly specified.** Methods make the implicit assumption that the true risk adjustment model does not vary over time, since they use pre-intervention observations to risk adjust the post-intervention outcomes. By considering three separate outcome periods, we are able to update the risk adjustment model each time to ensure that it reflects the timeliest pre-intervention observations.
- **The outcomes analysed do not represent all facets of the potential impact of the ICTP.** Some of the potential impacts of the ICTP (e.g. on utilisation of non secondary care health and care services, impacts on quality of life, staff satisfaction, the quality of working relationships and on children) will not be captured in the set of outcomes included. Analysis of these impacts could be considered in future work. For a full picture of the impact of the ICTP in Mid-Nottinghamshire, the evaluation should be viewed in conjunction with the other research carried out by the local evaluation of the Mid-Nottinghamshire Better Together vanguard.

Reporting

General reporting considerations

Estimates of the average treatment effect of the ICTP on the treated (ATT) with associated 95% confidence interval and p-value will be reported for each outcome. These are calculated as described in the statistical methods. Plots of average outcomes in treated and counterfactual groups over the study periods 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.

Special reporting requirements for this study

At a minimum, the following are requirements for this study:

- adherence to the Strengthening of The Reporting of Observational Studies in Epidemiology (STROBE)[18] which hampers the assessment of its strengths and weaknesses and of a study's generalizability. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE and The REporting of studies Conducted using Observational Routinely-Collected health Data (RECORD)[19] guidelines
- adherence to the NHS Digital (previously Health and Social Care Information Centre, HSCIC) small number rules[20]
- compliance with the statistical code of practice.

Appendix

Analysis methods

Framework

We begin by setting out the framework that is used throughout. Let Y_{it} denote the observed value of a selected outcome in unit i at time t, and let X_{it} denote an ($r \times 1$) vector of observed covariates for unit i at time t. We assume there are T separate time periods of which T_0 are prior to an intervention, so $t \in (1, \cdots, T_0, T_0 + 1, \cdots T)$; and I units (e.g. GP practices) of which the first I_o are untreated, so $i \in (1, \cdots, I_0, I_0 + 1, \cdots I)$. Further, let D_{it} be a treatment indicator, which equals 1 if unit i has been exposed to the treatment after time t and zero otherwise (i.e. $D_{it} = 1$ if $i > I_o$ and $t > T_o$, and $D_{it} = 0$ otherwise).

To formalise the notion of causality, we use the potential outcomes framework for causal inference[21,22,23]. Let Y_{it}^0 and Y_{it}^1 be the potential outcome in unit i at time t when $D_{it} = 0$ and $D_{it} = 1$ respectively. The observed outcome can be written as

$$Y_{it} = D_{it} Y_{it}^1 + (1 - D_{it}) Y_{it}^0 \tag{1}$$

Following Abadie et al[17], we assume that Y_{it}^0 is derived linearly via a factor model with mean zero shocks:

$$Y_{it}^0 = \boldsymbol{\beta} \boldsymbol{X_{it}} + \boldsymbol{\lambda_t \mu_i} + \varepsilon_{it}$$

where $\boldsymbol{\beta}$ is an ($1 \times r$) vector of unknown parameters, $\boldsymbol{\mu}_i$ is an ($F \times 1$) vector of unobserved time invariant variables (factor loadings) with $\boldsymbol{\lambda_t}$ a ($1 \times F$) vector of their effects (common factors) at time t, and ε_{it} are exogenous, unobserved idiosyncratic zero mean shocks.

Assuming an additive treatment effect, α_{it}, on the outcome in unit i at time t, we can write the potential outcome under treatment as

$$Y_{it}^1 = \boldsymbol{\beta} \boldsymbol{X_{it}} + \boldsymbol{\lambda_t \mu_i} + \alpha_{it} + \varepsilon_{it}$$

and, more generally, the observed outcome as

$$Y_{it} = \boldsymbol{\beta} \boldsymbol{X_{it}} + \boldsymbol{\lambda_t \mu_i} + D_{it} \alpha_{it} + \varepsilon_{it} \tag{2}$$

The individual treatment effect on the outcome in unit i at time t is then $\alpha_{it} = Y_{it}^1 - Y_{it}^0$ for $i > I_0$ and $t > T_0$. An estimand of interest is the average treatment effect for the treated (ATT_t) at time $t > T_0$ which is given by the average treatment effect on the outcome across all units in the treated group at time t:

$$\text{ATT}_t := \alpha_t = \frac{1}{I - I_o} \sum_{i=I_0+1}^{I} \alpha_{it} = \frac{1}{I - I_o} \sum_{i=I_0+1}^{I} (Y_{it}^1 - Y_{it}^0) = \bar{Y}_t^1 - \bar{Y}_t^0 \tag{3}$$

where $\bar{Y}_t^1 = \frac{1}{I - I_o} \sum_{i=I_0+1}^{I} Y_{it}^1$ and $\bar{Y}_t^0 = \frac{1}{I - I_o} \sum_{i=I_0+1}^{I} Y_{it}^0$ are the average potential outcomes in the treated units in the presence and absence of treatment, respectively. Averaging ATT_t across all treated time periods gives a final estimate of the average treatment effect for the treated, $ATT = \frac{1}{T - T_o} \sum_{t=T_0+1}^{T} \alpha_t$.

In the post-intervention period, while Y_{it}^1 is observed for treated units ($Y_{it}^1 = Y_{it}$ for $i > I_0$ and $t > T_0$), Y_{it}^0 is not. Therefore, approximation of α_t requires estimation of Y_{it}^0 for $i > I_0$ and $t > T_0$, or of its average value across the treated units: $\bar{Y}_t^0 = \frac{1}{I - I_o} \sum_{i=I_0+1}^{I} Y_{it}^0$ for $t > T_o$.

Difference in Differences (DiD)

Letting $\boldsymbol{\mu_i} = [1, \mu_i]'$ and $\boldsymbol{\lambda_t} = [\lambda_t, 1]$, equation (2) is then a two-way fixed effects model:

$$Y_{it} = \boldsymbol{\beta X_{it}} + \mu_i + \lambda_t + D_{it}\alpha_{it} + \varepsilon_{it} \quad (4)$$

Here, the effect of the unobserved variable, μi, is assumed not to vary over time, thus satisfying the parallel trends assumption[24,25]:

$$E\left(Y_{it}^0 - Y_{it'}^0 | D_{it} = 1, \boldsymbol{X_{it}}\right) = E\left(Y_{it}^0 - Y_{it'}^0 | D_{it} = 0, \boldsymbol{X_{it}}\right) \forall\, t > T_o \quad \text{(A1: "Parallel trends")}$$

The DiD estimate of the ATT is estimated by fitting this two-way fixed effect regression model[24,26,27,28,29]. Under assumption A1, DiD will provide consistent estimates of the ATT even when treatment effects are heterogenous.

Original Synthetic Control (OSC)

The central idea of the Original Synthetic Control (OSC) method proposed by Abadie et al.[30,31] is to construct a weighted combination of untreated units to represent a counterfactual treatment free outcome for a treated region. The OSC method is designed for the case where the treated region is comprised of a single treated unit and the untreated region of multiple untreated units. However, if there is more than one treated unit, the treated unit can be aggregated to create a single treated unit[32].

The similarity between the treated region and the synthetic control region for the outcome is then determined by a set of characteristics from the T_0 pre-intervention periods comprising two kinds of predictors. The first is given by M linear combinations of the outcome in the pre-intervention period; the second is given by r covariates that are predictive of the outcome in the pre-intervention period. All p predictors ($p = M + r$) are combined in a (p x 1) vector $\boldsymbol{Z_1}$ for the treated unit, and in a corresponding $(p\ x\ I_o)$ matrix $\boldsymbol{Z_0}$ for the untreated units. The synthetic control unit is then formed by finding a $(I_o x\ 1)$ vector of positive weights $\boldsymbol{W}^* = (w_1 \ldots w_{I_0})'$ that sums to one and which minimizes the distance metric $(\boldsymbol{Z_1} - \boldsymbol{Z_0W})'\ \boldsymbol{V}(\boldsymbol{Z_1} - \boldsymbol{Z_0W})$ where $\boldsymbol{V}$ is a $(p \times p)$ diagonal matrix that captures the relative importance of the p predictors. The optimisation process comprises two steps. The 'inner optimisation' step attempts to minimise the distance metric subject to the 'outer optimisation' step, which deals with finding the $\boldsymbol{V}$ matrix.

The optimal set of weights creates a synthetic control, which approximates the average of the treated unit outcomes ($\bar{Y}_t = \sum_{i=I_0+1}^{I} Y_{it}$) and covariates ($\bar{\boldsymbol{X}}_{\boldsymbol{t}} = \sum_{i=I_0+1}^{I} \boldsymbol{X}_{\boldsymbol{it}}$) in each pre-intervention period:

$$\sum_{i=1}^{I_0} w_i Y_{it} = \bar{Y}_t, \forall t \leq T_o \qquad (5)$$

$$\sum_{i=1}^{I_0} w_i \boldsymbol{X}_{\boldsymbol{it}} = \bar{\boldsymbol{X}}_{\boldsymbol{t}}, \forall t \leq T_o \qquad (6)$$

with $0 \leq w_i \leq 1$, and $\sum_{i=1}^{I_0} w_i = 1$.

Given ***W**** that satisfy equations (5) and (6), we can estimate $\bar{Y}_t^0$, the average treated counterfactual outcome, by the post-intervention outcome of the synthetic control, $\hat{\bar{Y}}_t^0 = \sum_{i=1}^{I_o} w_i Y_{it}$ for $t > T_o$. Hence an estimate of the ATT_t is

$$\hat{\alpha}_t = \bar{Y}_t^1 - \hat{\bar{Y}}_t^0$$

for t > T_0.

Abadie et al.[17] show that, assuming the data-generating model of the potential outcomes is linear, as in equation (2), and that the pre-intervention time period is sufficiently long, $\hat{a}_t$ is an approximately unbiased estimator for the ATT_t.

Micro Synthetic Control (MSC)

The Micro Synthetic Control (MSC) method proposed by Robbins et al.[9] makes a variation to the OSC method in the way that the weights for the synthetic control are selected. It is designed to be computationally efficient for high dimensional data comprising multiple treated and untreated units and multiple pre- and post-intervention time periods. As for the OSC method, the central idea is to construct a weighted combination of the untreated units to represent a counterfactual treatment free outcome for a treated region. Like the OSC method, the weights are chosen so that the treated region and the synthetic control region have similar values of a set of characteristics from the T_0 pre-intervention periods. Unlike the OSC method, these characteristics can include multiple outcomes enabling a synthetic control that can be simultaneously applied to multiple outcomes.

To extend notation to cover *K* different outcomes of interest, let $Y_{k,it}$ denote the observed value of outcome k in unit i at time *t*, $k \in (1, \cdots, K)$. Denote

$$\boldsymbol{X}_{\boldsymbol{i}} = (1, Y_{1,i1}, \cdots, Y_{1,iT_o}, Y_{2,i1}, \cdots, Y_{2,iT_o} Y_{K,i1}, \cdots, Y_{K,iT_o}, \boldsymbol{X}_{\boldsymbol{i1}}', \boldsymbol{X}_{\boldsymbol{i2}}', \cdots, \boldsymbol{X}_{\boldsymbol{iT_o}}')$$

as a vector of all outcomes and covariates at all pre-intervention time periods (with an intercept term) for unit *i*. The treated units are aggregated in to a single treated region and the target totals for the treated region are $\boldsymbol{t_x} = \sum_{i=I_0+1}^{I} \boldsymbol{X_i}\ (I_o x\ 1)$. The synthetic control unit is then formed by finding a $(I_o x\ 1)$ vector of positive weights $\boldsymbol{W}^* = (w_1 \cdots w_{1_0})'$ aiming to satisfy a set of calibration equations given by

$$\sum_{i=1}^{I_0} w_i \boldsymbol{X_i} = \boldsymbol{t_x} \qquad (7)$$

with $0 \leq w_i \leq 1$, and $\sum_{i=1}^{I_0} w_i = I - I_o$. Having the weights add up to the number of treated units, rather than 1, as for the OSC method is arbitrary and arguably more intuitive for count covariates. The synthetic control weights are derived using calibration techniques to minimise a given distance metric subject to reproducing the target totals.

Given $\boldsymbol{W^*}$ that satisfy equation (7), we can estimate the average treated counterfactual value for outcome k, $\bar{Y}^0_{k,t} = \frac{1}{I-I_o} Y^0_{k,it}$ by the post-intervention value for outcome *k* of the synthetic control, $\hat{\bar{Y}}^0_{k,t} = \frac{1}{I-I_0} \sum_{i=1}^{I_o} w_i Y_{k,it}$, and hence an estimate of the effect of outcome *k* at time *t* is

$$\hat{\alpha}_{k,t} = \bar{Y}^1_{k,t} - \hat{\bar{Y}}^0_{k,t}$$

for t > T_0. Averaging this estimate over the treated units gives the ATT_t.

Under certain conditions, algebraic expressions can be used to approximate the sampling distribution of estimated treatment effects, but more generally resampling techniques are required to estimate standard errors and confidence intervals.

Generalised Synthetic Control (GSC)

The Generalised Synthetic Control (GSC) method proposed by Xu[8] combines insight from the OSC method with interactive fixed effect (IFE) models. It is designed for multiple treated units and multiple time periods and allows for heterogenous estimates of effect.

GSC assumes that each outcome Y_{it} is given by a linear factor model with functional form as in equation (2) and, for identifiability, that all factors are normalised and orthogonal to each other. GSC also assumes strict exogeneity:

$$\varepsilon_{it} \perp D_{js}, \boldsymbol{X_{js}}, \boldsymbol{\lambda_s}, \boldsymbol{\mu_j} \text{ for } i, j \in (1, \cdots, I) \text{ and } t, s \in (1, \cdots, T)$$

This means that the error term of any unit at any time is independent of treatment assignment, observed covariates, factors and factor loadings of any other unit at any other time and further implies conditional mean independence:

$$E(\varepsilon_{it} | D_{it}, \boldsymbol{X}_{it,} \boldsymbol{\lambda}_t, \boldsymbol{\mu}_i) = E(\varepsilon_{it} | \boldsymbol{X}_{it,} \boldsymbol{\lambda}_t, \boldsymbol{\mu}_i) = 0$$

This holds true if the same data generating process underlies the outcomes for both the treated and untreated units. If not, then estimates will be biased. GSC assumes that this underlying model is an IFE model and estimates treatment effects for each treated units as follows.

IFE models rely on an alternative set of estimation approaches for the common factor structure $\boldsymbol{\lambda_t \mu_i}$ in equation (2)[33]. In the first step an IFE model is estimated for the control units only across the entire treatment period. This provides estimates $\widehat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\lambda}}_t, t = 1, \ldots, T$, which are assumed to be the same for treated and control units (as a result of the assumed conditional mean independence), and estimates $\widehat{\boldsymbol{\mu}}_i, i \leq I_0$, for the control units only. To determine estimates $\widehat{\boldsymbol{\mu}}_i, i > I_0$, for the treated units, the second step finds the values $\boldsymbol{\mu}_i, i > I_0$, that minimise the pre-treatment difference between the observed outcome Y_{it} and the imputed outcome Y_{it} for each treated unit based on:

$$\hat{Y}_{it} = \widehat{\boldsymbol{\beta}}\, \boldsymbol{X}_{it} + \hat{\boldsymbol{\lambda}}_{\mathbf{t}} \boldsymbol{\mu}_i, i > I_0, t \leq T_0$$

Finally, the treated counterfactual is based on $\widehat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\lambda}}_t$, and $\widehat{\boldsymbol{\mu}}_i$:

$$\hat{Y}_{it}^{\,0} = \widehat{\boldsymbol{\beta}}\, \boldsymbol{X_{it}} + \hat{\boldsymbol{\lambda}}_{\boldsymbol{t}} \widehat{\boldsymbol{\mu}}_{\boldsymbol{i}} \quad i > I_o, t > T_o$$

and an estimate of the ATT_t is

$$\hat{\alpha}_t = \frac{1}{I - I_o} \sum_{i=I_0+1}^{I} \left(Y_{it}^1 - \hat{Y}_{it}^0\right)$$

for $t > T_o$.

Standard errors and confidence intervals are estimated using parametric bootstrap methods. The GSC method is more dependent on modelling assumptions that the OSC method and requires more pre-intervention data than fixed effects estimators, such as DiD, otherwise incidental parameters can lead to biased estimates of treatment effects[33].

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 one 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 admission

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 and hence which could have been 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[34]. An avoidable admission resulting from a condition in this set is referred to as a potentially avoidable admission.
- Ambulatory care sensitive (ACS) conditions – ACS conditions are a set of clinical conditions for which the risk of emergency admission can be reduced by timely and effective ambulatory care[35]. Ambulatory care consists of primary care, community services and outpatient care[36]. There are a variety of definitions of ACS conditions[36]. The definition used by the IAU will be the same as defined in the CCG improvement and assessment framework (CCGIAF)[37]. 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[38], 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 CCG Outcomes Indicator Set 2.6[37]. Conditions include epilepsy, diabetes and angina[39].

– Acute ACS conditions, also called urgent care sensitive conditions: urgent care sensitive 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. Conditions include chronic obstructive pulmonary disease (COPD), cellulitis, deep vein thrombosis and falls[37].

Elective admission

An elective admission is defined as an admission that has been arranged in advance, either planned, booked in advance or from a waiting list. It does not include an emergency admission, a maternity admission or a transfer from a hospital bed in another health care provider. Depending on the analysis being undertaken, an elective admission may include one, some or all of the following patient classifications:

- ordinary admission
- day case admission
- regular day admission
- regular night admission
- mother and baby using delivery facilities only.

Elective bed days

An elective bed day is defined as a night in hospital following an elective admission. Some elective admissions may be excluded from bed days calculations depending on the patient classifications being included in the definition of an elective admission (see above).

Emergency (non-elective) admission

An emergency admission, also called a non-elective admission, is defined as a separate hospital spell that either occurs through an A&E department, or because of direct, urgent referrals from a GP or other professional.

Emergency bed days

An emergency bed day is defined as a night in hospital following an emergency admission. This is consistent with the definitions of bed day used within NHS England[40] and the NHS England New Models of Care dashboard, which displays outcome data for all vanguard sites[41].

Emergency readmissions within 30 days of discharge

An emergency readmission within 30 days of discharge is defined as an emergency admission occurring within 30 days of discharge following an earlier hospital admission (regardless of whether the earlier admission was emergency or elective). Admissions for cancer and obstetrics are excluded as they may be part of the patient's care plan[17].

Length of stay following an elective admission

Length of stay following an elective admission is defined as the number of nights spent in hospital following an elective admission, calculated as the difference in days between the date of discharge and the date of admission. Some elective admissions may be excluded from length of stay calculations depending on the patient classifications being included in the definition of an elective admission (see above). An admission and discharge within the same day will result in a length of stay of zero days[42].

Length of stay following an emergency (non-elective) 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[42].

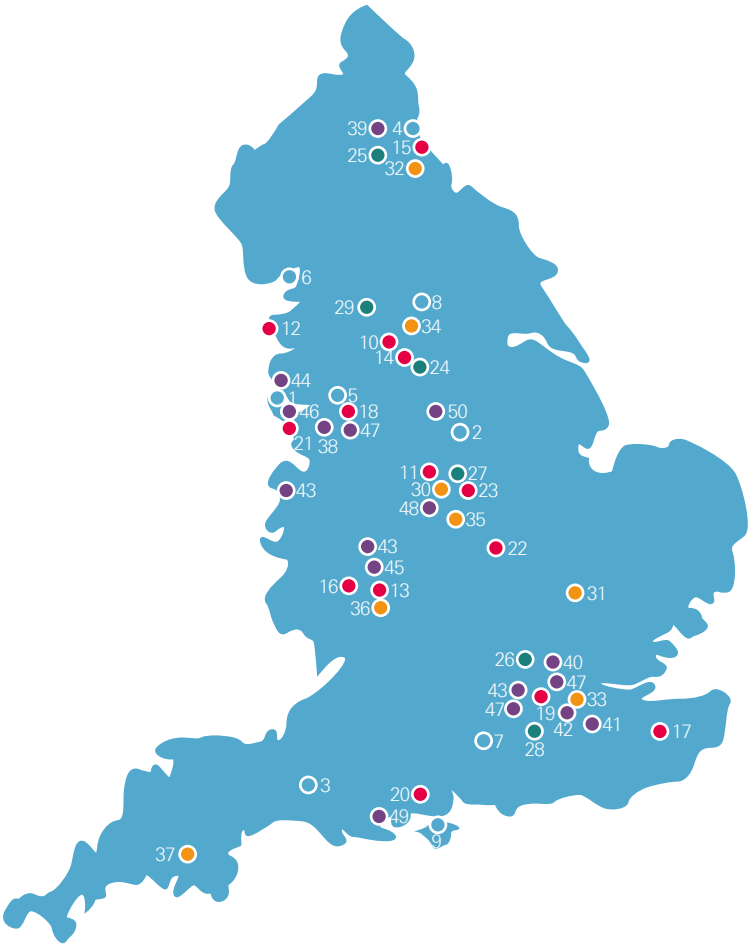
Outpatient appointment

An outpatient appointment is a non-duplicate appointment for a patient to see, or have contact with, a care professional at an outpatient clinic. A recorded appointment by an individual to the same provider at the same date and time as a previously recorded appointment, and where the main specialty and treatment function codes are the same, is defined as a duplicate appointment Depending on the analysis being undertaken, an outpatient appointment may be further defined as one of the following:

- a non-duplicate appointment where the patient was seen
- a non-duplicate appointment with an acute provider
- a non-duplicate appointment with an acute provider where the patient was seen.

Supplementary figures

Supplementary figure 1: Map of New Care Model vanguard sites



Integrated primary and acute care systems – joining up GP, hospital, community and mental health services

1 Wirral Partners
2 Mid Nottinghamshire Better Together
3 South Somerset Symphony Programme
4 Northumberland Accountable Care Organisation
5 Salford Together
6 Better Care Together (Morecambe Bay Health Community)
7 North East Hampshire and Farnham
8 Harrogate and Rural District Clinical Commissioning Group
9 My Life a Full Life (Isle of Wight)

Multispecialty community providers – moving specialist care out of hospitals into the community

10 Calderdale Health and Social Care Economy
11 Wellbeing Erewash
12 Fylde Coast Local Health Economy
13 Modality Birmingham and Sandwell
14 West Wakefield Health and Wellbeing Ltd
15 All Together Better Sunderland
16 Dudley Multispecialty Community Provider
17 Encompass (Whitstable, Faversham and Canterbury)
18 Stockport Together
19 Tower Hamlets Together
20 Better Local Care (Hampshire)
21 West Cheshire Way
22 Lakeside Healthcare (Northamptonshire)
23 Principia Partners in Health (Southern Nottinghamshire)

Enhanced health in care homes – offering older people better, joined up heath, care and rehabilitation services

24 Connecting Care - Wakefield District
25 Gateshead Care Home Project
26 East and North Hertfordshire Clinical Commissioning Group
27 Nottingham City Clinical Commissioning Group
28 Sutton Homes of Care
29 Airedale & Partners

Urgent and emergency care – new approaches to improve the coordination of services and reduce pressure on A&E department

30 Greater Nottingham System Resilience Group
31 Cambridgeshire and Peterborough Clinical Commissioning Group
32 North East Urgent Care Network
33 Barking and Dagenham, Havering and Redbridge System Resilience Group
34 West Yorkshire Urgent and Emergency Care Network
35 Leicester, Leicestershire & Rutland System Resilience Group
36 Solihull Together for Better Lives
37 South Devon and Torbay System Resilience Group

Acute care collaborations - linking hospitals together to improve their clinical and financial viability

38 Salford and Wigan Foundation Chain
39 Northumbria Foundation Group
40 Royal Free London
41 Foundation Healthcare Group (Dartford and Gravesham)
42 Moorfields
43 National Orthopaedic Alliance
44 The Neuro Network (The Walton Centre, Liverpool)
45 MERIT (The Mental Health Alliance for Excellence, Resilience, Innovation and Training (West Midlands)
46 Cheshire and Merseyside Women's and Children's Services
47 Accountable Cancer Network (ACN)
48 EMRAD - East Midlands Radiology Consortium
49 Developing One NHS in Dorset
50 Working Together Partnership (South Yorkshire, Mid Yorkshire, North Derbyshire

Addendum

Study outcomes

The following outcomes were added to the study:

- proportion of A&E attendances that resulted in an emergency admission
- rate of non-ambulatory care sensitive admissions per head of GP practice size per month
- proportion of emergency admissions with a non-zero length of stay
- proportion of elective admissions with a non-zero length of stay
- rate of emergency re-admissions, defined as the proportion of admissions (emergency and elective) that resulted in an emergency admission within 30 days of the original admission.

The following amendments were made to study outcomes originally selected:

- analysis of length of stay for emergency and elective admissions only took into account those that resulted in a non-zero length of stay
- analysis of outpatient appointments considered only the first outpatient appointment in any series of appointments.

Study periods

The end of the study period was extended from March 2018 to March 2019. Estimates of impact are provided for each financial year starting 2013, 2014, 2015, 2016, 2017 and 2018.

Baseline variables

This evaluation only looks at individuals aged 18 years and older. Age categorisation noted in Table 2 should be corrected to 18–24 years, 25–64 years, 65–75 years and >75 years.

Subgroup analyses

The impact of the ICTP on primary study outcomes was also estimated separately for individuals aged over 65 years.

Acknowledgements

We are grateful for the support of the Mid-Nottinghamshire Better Together vanguard in providing review and feedback on the statistical analysis protocol.

References

1. Better Together Mid-Nottinghamshire. Mid-Nottinghamshire NHS Better Care Fund Case Study [Internet]. 2014. Available from: https://webcache.googleusercontent.com/search?q=cache:DBiTic3LQ_IJ:https://www.england.nhs.uk/wp-content/uploads/2014/09/bcf-cs-mid-notts.pdf+&cd=1&hl=en&ct=clnk&gl=uk
2. Better Together Mid-Nottinghamshire. Welcome to Better Together [Internet]. [cited 2018 Jun 11]. Available from: http://www.bettertogethermidnotts.org.uk/
3. Better Together Mid-Nottinghamshire. Mid-Nottinghamshire Integrated Care Transformation Programme Overarching Programme Proposal [Internet]. 2013. Available from: http://webcache.googleusercontent.com/search?q=cache:wil5oVv0cUIJ:www.bettertogethermidnotts.org.uk/modules/downloads/download.php%3Ffile_name%3D124+&cd=1&hl=en&ct=clnk&gl=uk
4. Better Together Mid-Nottinghamshire. Better Together – Development Journey. June 2017 [Internet]. Available from: http://webcache.googleusercontent.com/search?q=cache:evFQDAmmoiMJ:www.bettertogethermidnotts.org.uk/modules/downloads/download.php%3Ffile_name%3D152+&cd=1&hl=en&ct=clnk&gl=uk
5. Better Together Mid-Nottinghamshire. Mid Notts pathways. Call for Care Evaluation [Internet]. Available from: https://midnottspathways.nhs.uk/guidance-pathways/out-of-hours-care-urgent-care/call-for-care/
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. 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(1–2).
8. Xu Y. Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models. *Polit Anal*. 2017;25(1):57–76.
9. 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.
10. Quan H, Sundararajan V, Halfon P, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. Medical Care. 2005;43(11):1,130–9. Available from: https://insights.ovid.com/crossref?an=00005650-200511000-00010
11. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. Med Care. 1998;36(1):8–27.
12. Bottle A, Aylin P. Comorbidity scores for administrative data benefited from adaptation to local coding and diagnostic practices. Journal of Clinical Epidemiology. 2011;64(12):1426–33. Available from doi: 10.1016/j. jclinepi.2011.04.004
13. Vickers D, Rees P. Creating the UK National Statistics 2001 output area classification. *J R Stat Soc Ser A Stat Soc.* 2007;170(2):379-403.
14. Office for National Statistics. Methodology note for the 2011 area classification for local authorities [Internet]. 2015. Available from: https://www.ons.gov.uk/methodology/geography/geographicalproducts/areaclassifications/2011areaclassifications/methodologyandvariables
15. Bring J. How to standardize regression coefficients. *Am Stat.* 1994;48(3).
16. Abadie A, Gardeazabal J. The economic costs of conflict: A case study of the Basque country. Am Econ Rev. 2003;93(1):113–32.
17. 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* [Internet]. 2010;105(490):493–505. Available from: https://www.tandfonline.com/doi/abs/10.1198/jasa.2009.ap08746

18. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol.* 2008; Apr;61(4):344-9.
19. Nicholls SG, Quach P, Von Elm E, Guttmann A, Moher D, Petersen I, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement: Methods for arriving at consensus and developing reporting guidelines. *PLoS One.* 2015;10(5): e0125620.
20. NHS Digital. Hosptial Episode Statistics (HES) analysis guide [Internet]. Available from: https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics
21. Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. *J Educ Psychol.* 1974;66(5):688–701.
22. Holland PW. Statistics and causal inference. *J Am Stat Assoc.* 1986;81(396):945–60.
23. Splawa-Neyman J, Dabrowska DM and Speed TP. On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9. *Stat Sci.* 1990;5(4):465–72.
24. Jones AM, Rice N. Econometric Evaluation of Health Policies. In: Glied S, Smith PC (eds) The Oxford Handbook of Health Economics. Oxford University Press. 2012.
25. Angrist JD, Pischke J-S. Parallel Worlds: Fixed Effects, Differences-in-Differences, and Panel Data. In: Angrist JD, Pischke J-S (eds) Mostly Harmless Econometrics. Princeton University Press. 2009. pp.221–46.
26. Bertrand M, Duflo E, Mullainathan S. How Much Should We Trust Differences-In-Differences Estimates? *Q J Econ* [Internet]. 2004;119(1):249–75. Available from: https://economics.mit.edu/files/750
27. Carpenter CS, Stehr M. The effects of mandatory seatbelt laws on seatbelt use, motor vehicle fatalities, and crash-related injuries among youths. *J Health Econ.* 2008;27(3):642–62.
28. Fletcher JM, Frisvold DE, Tefft N. Non-linear effects of soda taxes on consumption and weight outcomes. *Health Econ.* 2015;24(5):566–82.
29. Wen H, Hockenberry JM, Cummings JR. The effect of medical marijuana laws on adolescent and adult use of marijuana, alcohol, and other substances. *J Health Econ.* 2015;42:64–80.
30. Abadie A, Gardeazabal J. The economic costs of conflict: A case study of the Basque country. *Am Econ Rev.* 2003;93(1):113–32.
31. Abadie A, Diamond A, Hainmueller J. Comparative Politics and the Synthetic Control Method. *Am J Pol Sci.* 2015;59(2):495–510.
32. Kreif N, Grieve R, Hangartner D, Turner AJ, Nikolova S, Sutton M. Examination of the Synthetic Control Method for Evaluating Health Policies with Multiple Treated Units. *Health Econ* [Internet]. 2015;19(11):n/a-n/a. Available from: http://doi.wiley.com/10.1002/hec.3258
33. Bai J. Panel data models with interactive fixed effects. *Econometrica.* 2009;77(4):1229–79.
34. Care Quality Commission. The state of health care and adult social care in England 2012/13. CQC. 2013.
35. 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.
36. 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* [Internet]. 2013 Jan [cited 2014 Dec 12];3(1):1–12. Available from: https://bmjopen.bmj.com/content/3/1/e002007
37. NHS England: Analytical Team. CCG improvement and assessment framework 2016/17: Technical annex. NHS England. 2017.
38. NHS Digital. NHS Outcomes Framework [Internet]. Vol. 2.4. Health and Social Care Information Centre. 2017.

39. NHS Digital Clinical Indicators Team. CCG Outcomes Indicator Set – Indicator 2.6: Unplanned hospitalisation for chronic ambulatory sensitive conditions. 2016.
40. Department of Health. KH03 Quarterly bed availability and occupancy – Data definitions [Internet]. DoH. 2010. Available from: https://webcache.googleusercontent.com/search?q=cache:zHFTTqEg8lYJ:https://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2013/04/dh_124212-5.doc+&cd=2&hl=en&ct=clnk&gl=uk
41. Operational Research and Evaluation Unit. Evaluation strategy for new care model vanguards. NHS England. 2016.
42. Bottle A, Aylin P. Application of AHRQ patient safety indicators to English hospital data. *Qual Saf Health Care.* 2009;18(4):303–8.