

# Statistical analysis protocol for pooled analysis of three evaluations of community-based multidisciplinary teams in England

Julia Shen, Therese Lloyd

August 2020

# Contents

<b>Summary</b> .....	<b>4</b>
Purpose of this document .....	4
Purpose of this study .....	5
What the study will look at .....	5
Data the IAU will use .....	7
Period and residents covered by the study.....	7
Main strengths and weaknesses of the study .....	7
<b>Background</b> .....	<b>9</b>
Multidisciplinary teams and settings .....	10
Original study designs .....	11
Original study results .....	11
Potential causes of increased emergency hospital use for patients enrolled in MDTs .....	19
Objectives of the evaluation .....	24
<b>Methods</b> .....	<b>27</b>
Study cohorts .....	27
Sources of data .....	27
Study endpoints .....	28
Patient subgroups .....	28
Exploratory analyses .....	30
Reweighting of the pooled sample to build a new counterfactual using entropy balancing.....	30
Exploratory data analyses on heterogeneity, on stacked original analysis data (using matched controls) .....	32
Exploratory data analyses: improving on unobserved confounding by exploring social context factors and a proxy for end of life .....	36
Handling of differential information from risk scores.....	37
Comparison of matching alternatives in the pooled data set.....	39

<b>Main analysis .....</b>	<b>39</b>
Accounting for some unobserved confounding.....	39
Pooled and subgroup analyses .....	40
Conventional pooled analyses on the ‘stacked’ matched data set – in case no valid alternative to genetic matching is found within the constraints of the timelines .....	44
Patient and Public Involvement and Engagement (PPIE).....	44
Potential further work .....	45
<b>Limitations and sources of bias.....</b>	<b>45</b>
<b>Acknowledgements .....</b>	<b>47</b>
<b>References .....</b>	<b>48</b>

# Summary

## Purpose of this document

This protocol describes a detailed statistical analysis plan for the Improvement Analytics Unit (IAU) to carry out pooled meta- and subgroup analyses across three recently completed IAU matched control studies. These evaluations studied the impact of multidisciplinary teams (MDTs) offering integrated care in the community. We studied the impact of these teams on unplanned hospital use among service users, who were typically older people living with chronic conditions, in the Fylde Coast and North East Hampshire & Farnham (NEHF) clinical commissioning groups (CCGs) in England. This document describes detailed considerations on selecting and finalising study design for pooled analyses of the three previous studies, including sensitivity checks for robustness.

Overall, the objective of our research is to further explore the effectiveness of community MDTs, particularly within specific patient groups and across intervention settings. Under which general conditions, if any, do MDTs achieve their intended impact to reduce unplanned hospital use? A large body of health services research in the UK and elsewhere – including the original IAU evaluations of the three MDTs – has demonstrated that this aim is difficult to achieve.<sup>1,2,3,4,5</sup> English integrated care programmes involve substantial heterogeneity, with large variation across patients, settings and the clinical services that are offered within the MDT.<sup>6</sup> This heterogeneity may obscure preventive impacts on emergency hospital activity for certain patients, settings, and services, as well as pose challenges for generalising findings. This research addresses whether MDTs reduce or prevent unplanned hospital use as intended under specific circumstances, even though this effect may not occur in a general population. This may inform future targeting and refinement of these programmes, particularly as they feature prominently in the NHS Long Term Plan.

This pooled analysis will form part of a larger project on multidisciplinary teams. As part of the project, we are also undertaking a comparative area-level analysis using synthetic controls to assess whether these MDTs, combined with other concurrent integrated care initiatives implemented as part of the new care models programme, have reduced hospital use over a longer timeline.<sup>7</sup> We will contextualise these studies using a systematic ‘review of reviews’ of community MDTs, and narrate the history and policy rationale for these programmes in the UK.

This statistical analysis protocol (SAP) has been written before any analyses were undertaken, to ensure that all design and methods choices are made objectively and are not influenced by what is found in the data. Some exploratory work will be done before making some of the design decisions (see section Methods: Exploratory analyses). These exploratory analyses are specified in the SAP and the decisions taken based on these will be documented in an addendum to the SAP before further analyses are undertaken. It may be necessary to make changes to the design of the study at a later stage; if so, these changes will be documented and the rationale explained as a further addendum to this protocol.

This summary section provides a more accessible overview of the planned study, while the rest of the protocol is a technical document written to guide analytical processes. The IAU welcomes comments and questions on this protocol.

## Purpose of this study

This study is primarily intended to support the evidence base concerning the design and implementation of community-based MDTs to provide joined-up care for people with chronic conditions. These interventions, which bring together medical and non-medical care staff, form a core ambition of the NHS Long Term Plan to establish primary care networks (PCNs), including a £4.5bn fund for 'expanded community multidisciplinary teams' which will improve care for patients with, eg dementia, respiratory disease and severe mental ill health.<sup>8</sup> This research may also be of interest to others involved in health and care evaluation.

The proposed quantitative analysis builds on three evaluations of MDTs in three CCGs, already undertaken by the IAU between 2018 and 2019, which were commissioned and operated by two governing structures under the new care model<sup>6</sup> in the Fylde Coast and NEHF: integrated care teams (ICTs) in NEHF<sup>3,9</sup> as well as the Extensive Care Service (ECS) and the Enhanced Primary Care (EPC) programmes, both in Fylde Coast.<sup>5</sup> These MDTs targeted individuals at highest need or at increased risk of crisis, who are primarily aged over 65 and typically live with long-term conditions (LTCs) and complex care needs. These three CCGs cover areas of varied demographics, health and social care need, rurality and deprivation. All three studies found that patients enrolled in MDTs had higher emergency hospital use than the carefully selected control (comparison) groups (see Table 4).

## What the study will look at

There are several plausible hypotheses consistent with the increased emergency hospital use we observed in the original studies, as well as in previous research.<sup>10,11</sup>

- H1 Residual unobserved confounding:** although the control groups were similar to the intervention groups on observable characteristics, there may have been unobserved differences between the intervention and control groups which affected the outcomes.
- H2 Patient-specific effects:** MDTs may be best suited to support certain patient subgroups, for instance with particular long-term conditions.
- H3 Limited 'impactability' of riskiest patients:** there may be limited scope to reduce emergency hospital use for those identified as being at the highest risk, especially if they are nearing their end of life.
- H4 Delayed impact:** MDTs increase hospitalisation initially but sustain eventual reductions in hospital use after an initial delay caused by patient and/or provider factors (eg identifying unmet need in the short term and/or time needed to embed complex change).
- H5 Induced demand:** MDTs may have led to patients being more aware of their health needs or risks or health professions being risk averse, which in turn led to greater hospital and other service use.
- H6 Measuring hospital use does not adequately reflect quality:** the value of integrated care teams may primarily lie in measures other than reducing emergency hospital admissions.

We will explore some of these hypotheses.

Firstly, the IAU seeks to further explore the potential impact of **unobserved confounding** as an explanation for the increase in hospital use associated with MDTs in the original evaluations. We will try to improve the level of similarity between the referred and matched control groups by exploring characteristics related to 'social context', some of which may be proxies for social isolation. These are: living alone, recently started to live alone, living with somebody with certain LTCs, a recent bereavement within the household or frequent moves in the last year. Stokes et al found that, when identifying patients for MDTs, medical practitioners felt that the patients' needs were often primarily related to social factors such as isolation, poor housing or living arrangements and other socio-economic issues.<sup>2</sup> Factors such as living alone have been found to be associated with emergency hospital use.<sup>12</sup>

Secondly, the IAU will address **whether MDTs impact differently on certain patient subgroups**. These will focus on pre-specified patient groups who have multimorbidity, frailty, or specific LTCs that may be mostly likely to benefit from MDTs. The specified LTCs – dementia, chronic pulmonary disease and serious mental ill health – are based on conditions specified in the NHS Long Term Plan.<sup>8</sup> We will also assess whether patients at the end of life are impacted differently from those who were alive for more than 3 months (90 days) after MDT enrolment. Lastly, we will assess the effect of MDTs on patients with certain social context factors.

Similar impacts of MDTs were observed across the three programmes and CCGs, suggesting some general effects of such MDTs, even under varied conditions, that may hold across English communities that adopt such programmes. While we may identify impacts that may be generalised to England, the effects estimated here may also relate to important differences in how the respective MDTs have been implemented. These differences, or **heterogeneity**, occur between, and sometimes even within, MDT programmes in terms of the population, intervention, and standard of usual care that was available outside the MDT.

It may be impossible to quantify the impact of these differences. In particular, heterogeneity across regions and programmes will be identifiable, but the previous implementation and data do not allow us to infer clearly whether any differences arise from the patient population, intervention delivery or system setting. However, the pooled analysis may shed some light on directional trends, based on design and implementation choices in these programmes. Some differential effects may be inferred by trends arising from staged MDT implementation within each CCG, as well as contrasts between the two Fylde Coast CCGs for ECS or EPC as programmes with some overlap in eligible patient populations and time. For example, Fylde ECS replaces routine GP care for a temporary time whereas the Fylde EPC and NEHF ICT support standard primary care with wrap-around support. Consistent differences in the impact seen with Fylde ECS compared to the other MDTs could point to a different impact when replacing rather than supplementing standard GP-led care, particularly where other factors are consistent (ie patient population and other programme implementation factors are similar). Post hoc analyses and further evaluation may allow the IAU to validate these differences, which would further clarify how MDTs ought to be designed, commissioned, implemented and targeted.

## Data the IAU will use

The IAU will use the same data as used in the original studies<sup>13,14</sup> using pseudonymised patient-level national Secondary Uses Services (SUS) administrative hospital data for England, data on GP registrations and deaths, derived from pseudonymised resident-level National Health Applications and Infrastructure Services (NHAIS) data and the pseudonymised local data supplied by the CCGs. This included a list of patients referred to each ICT and their date of referral for NEHF and the equivalent information on enrolment in Fylde. In addition, we had access to the previously used list of risk scores supplied by the Fylde Coast CCGs for all of their intervention and the potential control patients.<sup>14</sup> These risk scores were based on the Combined Predictive Model (CPM) that quantifies the risk of future hospital (re)admission in the following 12 months.<sup>15</sup>

In addition, we will access pseudonymised addresses, also derived from pseudonymised resident-level NHAIS data, to identify factors relating to social context.

Pseudonymised means that all direct identifiers (eg name, address, date of birth, NHS number for patients) are removed from the data. Pseudonymisation reduces the risk that individual patients can be identified from the data.

## Period and residents covered by the study

The period, intervention patients and resident pools will be the same as in the original IAU studies.<sup>13,14</sup> There are a total of 5,411 individuals who received MDT care across NEHF and Fylde Coast, with a pool of potential control patients in excess of 130,000. The time periods vary across MDTs (NEHF July 2015 – June 2017, ECS August 2015 – April 2018, EPC November 2016 – April 2018).

## Main strengths and weaknesses of the study

We are exploring whether we can identify proxies for social isolation and other factors that may affect both the decision to refer individuals to an MDT and their health outcomes and emergency hospital use. The variables that we are investigating – living alone, recently started living alone, living with somebody with dementia or frailty, frequent moves and recent bereavement within the household – do not necessarily reflect an individual person's need or risk. However, we believe these factors signal important social context at population level that may be very important for assessing the impact of MDTs on hospital use and therefore warrant investigation.

Although this pooled analysis will aim to improve the similarity between the compared groups and refine the original findings from the original evaluations by including social context indicators, there is still a risk that the intervention and control groups remain different in important ways, for instance in comorbidities only recently diagnosed by GPs but not recorded in hospital or other risk factors that cannot be observed by the IAU. It is not currently possible to quantify this risk.

The IAU only had access to secondary care data and was not able to evaluate the impact of the MDTs on other outcomes, such as quality of life, experience of care, staff satisfaction, improvement in working relationships or cost. A consistent finding across the local mixed-method and descriptive evaluations from all three MDTs was that the MDTs often provided support that was more wide-ranging than just medical care, including eg benefits advice, help with lifestyle changes, and practical improvements in service users' homes.<sup>16,17</sup> These interventions may affect outcomes such as quality of life or experience of care, which we were not able to evaluate. Furthermore, although these interventions may also reduce hospital use in the long term, the effect may not be detectable within the time frames of these studies. There will, however, be a review of reviews of MDT interventions carried out in parallel with this work, which will aim to draw out the findings on the effect of MDTs on other outcome measures.

In this analysis, we explore the effect of MDTs on specific patient groups (eg individuals nearing end of life, with dementia, chronic pulmonary disease or serious mental ill health), by selecting patients with these conditions from the group of patients enrolled in the ICT, ECS and EPC multidisciplinary programmes. However, it may be that MDTs that are designed to specifically support a specific condition may have a different effect from that of MDTs that care for a broader range of conditions and needs. All three evaluated MDTs were designed to reach a broad population.

This study does not directly analyse the effect of other concurrent vanguard interventions such as a rapid home response by a specially trained community paramedic or social prescribing, nor will this study be able to quantify how other, non-MDT changes to the health and care systems in Fylde Coast and NEHF may have impacted outcomes.

The IAU used very similar statistical analysis methods, including the choice of covariates and outcomes, for the original analyses of all three programmes. We will standardise these procedures into a common data set for this pooled analysis, allowing us to measure the impact of an MDT programme on individuals enrolled in one of them. Nevertheless, there remain some natural differences (heterogeneity) in the patient populations, interventions, and comparative usual care between the three MDTs. This study will partially deal with this heterogeneity, as the continued use of robust statistical techniques will rely on substantial overlap or 'exchangeability' (ie similarity irrespective of whether enrolled in an MDT or not) of patients in terms of their observed characteristics. The three MDTs were similar in their design and implementation, perhaps to a greater extent than might be expected from the locally led programme design that was scoped under the new care models. The study populations appear to show substantial overlap or 'exchangeability', and we will be able to assess this formally when creating a counterfactual (comparison group) and applying regression in the pooled analysis. It is notable that these studies generated quite similar estimates of the impact (average treatment effect on the treated, or ATT) of MDTs on hospital use, suggesting there are generalisable effects regardless of the differences in settings, populations and programmes.



This study aims to understand why emergency hospital use is higher for patients enrolled to MDTs compared with other patients with similar observable characteristics, through robust quantitative analysis of administrative data. Although outside the scope of this project, it would be very useful to complement these analyses with further qualitative study, where MDT service users and professionals could be interviewed about patients' hospital use.

The results are nonetheless expected to enable learning that, together with other evidence, will help national and local commissioners to better understand MDTs, recognise the realistic impacts that can be expected from them, and identify potential areas for further investigation or improvement, which can, for example, help inform the implementation of MDTs within the new PCNs.

## Background

The proposed quantitative analysis builds on three evaluations of MDTs in three CCGs, already undertaken by the IAU between 2018 and 2019, which were commissioned and operated by two governing structures under the new care models vanguards<sup>6,18</sup> in the Fylde Coast and NEHF: the integrated care team (ICT) in NEHF<sup>3</sup>, as well as the Extensive Care Service (ECS) and the Enhanced Primary Care (EPC) programmes, both in Fylde Coast.<sup>5</sup> Previous research commissioned as a definitive evaluation by NHS England and NHS Improvement examined the impact of all English vanguard areas compared to areas that did not participate in the vanguards over the same period. This research found a statistically discernible but relatively modest reduction in unplanned hospital use from the policy, albeit with the greatest apparent impacts in care home vanguard areas and appearing after a sustained implementation period of 2–3 years.<sup>19</sup>

The MDTs evaluated by the IAU targeted individuals primarily living in their own homes in the community – rather than in care homes – with the highest needs or at increased risk of crisis, who were mostly aged over 65 and typically living with long-term conditions and complex care needs. These MDTs were set up to support local service and clinical integration strategies, usually led by the CCG. They included shared whole-system budgets or other organisational and governance incentives for multidisciplinary team collaboration, between traditional primary care providers (GPs) and other health and care professionals. All three MDTs were started as part of the Integrated Care Vanguards programme (primary and acute care system (PACS) or multispecialty community provider (MCP)).

The interventions were in three CCGs (Blackpool CCG and Fylde & Wyre CCG, collectively called the Fylde Coast vanguard, and NEHF CCG), covering areas of varied demographics, rurality and deprivation, relative to the English average. In total, the three CCGs covered a population of approximately 552,000 residents registered with 63 general practices. The three MDT intervention groups included 5,411 patients during the periods examined in our studies. Enrolment dates for the MDTs were specific to individual patients and recorded as such rather than applying at the same time across an area or programme.

The effect of being enrolled in an MDT was assessed by comparing the emergency hospital use of patients who enrolled in each of these interventions with the corresponding hospital use of carefully chosen matched control groups from the same local community, who were similar individuals on a range of known characteristics, eg age, gender, LTCs and prior hospital use.

Further details of the various study settings, populations, and intervention characteristics are detailed below and in Tables 1–4.

## Multidisciplinary teams and settings

The three MDTs in NEHF and Fylde Coast were similar in many respects (Table 1). All three programmes in Fylde Coast and NEHF targeted patients at higher risk of unplanned hospital use, and brought together community-based nurses or nurse practitioners with other medical professionals and non-medical care staff such as coordinators, social workers, etc. The MDTs universally applied care coordination and case management techniques to a patient group with multimorbidity, supported by regular meetings at least once per week to prioritise and stage care. All three programmes created or scaled multiple MDTs, with each one anchored in neighbourhoods or localities covering smaller geographic areas within each CCG. These MDTs served selected patients within communities of roughly the same size as the forthcoming PCN target list size of 30,000–50,000 people.<sup>8</sup> Notably, Fylde ECS completely replaced routine GP care on a temporary basis, whereas the other two MDTs in Fylde and NEHF provided wrap-around support to patients who remained registered with their GPs for routine primary care.

There are, however, some important differences (heterogeneity) between the populations and interventions (Table 2). Fylde Coast and NEHF represent diverse characteristics across patients, communities and health systems, with a broad range of social and individual-level factors. The MDT patients and matched controls span ages, genders and medical history of LTCs and hospital use. Their communities vary across rural vs urban, deprivation levels and other geographic factors. Fylde Coast is in western Lancashire in the North West of England, whereas NEHF covers parts of both Hampshire and Surrey in the South East, within commuting distance to London. Blackpool in Fylde Coast is an urban seaside resort town that currently faces significant social challenges and includes neighbourhoods among the most deprived in England, while the surrounding Fylde & Wyre boroughs are more suburban or rural with a similar profile of life expectancy and disease burden to English national averages.<sup>14,20</sup> NEHF is a largely suburban and affluent area that has better overall life expectancy, income, employment and housing conditions than the averages for England and the South East, but with pockets of deprivation.<sup>13,21</sup>

The local health and care systems surrounding each of the MDTs vary accordingly. For acute needs, Fylde Coast is served by one large NHS trust, Blackpool Teaching Hospitals NHS Foundation Trust, and the private hospital Spire Fylde Coast. NEHF is also served by two acute hospitals run by NHS Foundation Trusts: Frimley Park Hospital and the Royal Surrey County Hospital. Community medical and social services are structured differently according to local needs and policies. While English medical services are commissioned by local CCGs, social services are primarily commissioned and governed by local authorities. These bodies jointly commission and staff MDTs in local health and wellbeing strategies and integrated care systems.

As a result, there has been local tailoring of each MDT, including in exact staffing arrangements, back-end support such as electronic forms and patient records, and care packages that were offered. The Fylde Coast vanguard was jointly managed across two CCGs:

Blackpool CCG and Fylde & Wyre CCG rolled out ECS and EPC together as two complementary MDT models for patients in the highest and second highest risk groups, respectively, with some clinical judgement applied to final referrals after initial stratification using the CPM.<sup>16</sup> The NEHF vanguard covers the same area as the CCG, and the ICTs targeted the patients who were considered by referring professionals to have the highest need, to be at greatest risk of going into crisis and who would most benefit from multidisciplinary support, although there was some local variation in targeting between the five locality teams. In both Fylde Coast and NEHF, referrals typically came from GPs. In Fylde Coast, final enrolment decisions were determined by a clinical assessment carried out by the MDT and patient consent to starting the service. In NEHF, consent was obtained before referral, and therefore all patients in NEHF referred to an ICT are considered enrolled in the programme.

## Original study designs

The IAU used very similar study designs and nearly identical data sources (SUS and local enrolment data) to carry out the original evaluations (Table 3). All studies were individually matched control studies using generalised Mahalanobis distance matching (within the genetic matching algorithm<sup>22</sup>) with very similar matching and regression procedures, modelling for nearly identical outcomes from very similar covariates. However, predictive risk scores using past clinical data were available and deployed in Fylde Coast, whereas they were not applied in NEHF during the evaluation period and therefore were also not included in the statistical analysis. In the outcome modelling across all evaluations, data-driven lasso regression<sup>23</sup> was generally used for outcome estimation except in a small number of cases where the algorithm failed to converge with sparse data and a generalised linear model was selected from several pre-specified distributions and covariate lists based on the Akaike Information Criterion.<sup>24</sup> Because the IAU carried out all three studies using essentially the same pre-processing and statistical outcome modelling methods, we expect that measurement error or methodological differences have negligible impact on differences in observed impact between the three interventions. All MDTs were associated with higher unplanned hospital use than in the matched controls. However, this analysis will identify whether observed impacts are robust to the notable difference in covariates created by inclusion (or not) of the CPM risk score.

## Original study results

Across all three studies, we found that patients who used MDTs had higher unplanned hospital use than matched controls, even after regression adjustment (Table 4). Our evaluations estimated that A&E attendances were, on average, between 26% and 40% higher for each of the interventions, and similarly emergency admissions were between 27% and 43% higher. Consistent increases in hospital activity were also observed across type of emergency admission – for chronic ambulatory care sensitive (chronic ACS) or urgent care sensitive (UCS) conditions – and for emergency bed days in Fylde Coast and emergency length of stay in NEHF. In other words, we consistently found that MDTs appeared to increase unplanned hospital use in these three programmes, contrary to the strategic objectives motivating their design. These results are consistent with other independent evaluations.<sup>1,2,4,25,26,27</sup> The research questions in this pooled analysis arose from this finding.

**Table 1: MDT programmes reviewed in the pooled analysis**

<b>Multidisciplinary teams – intervention design</b>	<b>Fylde Coast Extensive Care Service (ECS)</b>	<b>Fylde Coast Enhanced Primary Care (EPC)</b>	<b>NEHF Integrated Care Teams (ICT)</b>
<b>Relationship to primary care</b>	<b>Replaces usual GP</b> ('carve-out' from usual primary care)	'Wrap-around' providing <b>additional</b> coordination and <b>services to usual GP care</b>	
<b>Implementation start date</b>	August 2015	November 2016	July 2015
<b>Implementation and team structure</b>	10 ECS teams were mobilised, across six hubs. 2 care hubs were launched with the service in June 2015, 2 more opened in May 2016, and the final 2 started in October 2016.	8 EPC teams were fully mobilised across the Fylde Coast vanguard area by February 2017, and largely overlapped with the ECS care hubs.	5 ICTs, 1 in each of NEHF's 5 localities, Farnborough, Farnham, Yateley, Fleet and Aldershot, were introduced in July 2015.
<b>Staffing</b>	Area divided into <b>10 'neighbourhood' hubs</b> , each covering ~20–30k population. Hub-based team mostly led by advanced nurse practitioners, also comprising GPs, clinical care coordinators, wellbeing support workers, etc. All teams medically supervised and receiving regular input from one consultant extensivist doctor.	<b>8 care teams</b> based at same neighbourhood hubs as ECS. Staffing across nurses, therapists, mental health specialists, social workers, care coordinators, and wellbeing support workers. The latter group were specifically hired into new roles created for EPC. Primary responsibility for medical supervision remained with regular GP.	<b>5 teams for each locality</b> in NEHF. Each ICT had a clinical lead, coordinator, community matron, social worker or care manager, mental health practitioner, ambulance service or community paramedic, social prescribing coordinator, dementia practitioner and pharmacist. Medical supervision remained with regular GP.
<b>Clinical services offered</b>	Primary care <b>replacement</b> of GP services on a temporary basis including <b>care plan</b> , with the aim of more proactive and patient-centred coordination and delivery, allowing for more <b>frequent or longer visits</b> . MDT <b>clinical huddles</b> improved care.	Primary care <b>support</b> including <b>care plan</b> with medical monitoring, <b>health coaching</b> , and support with social skills where applicable. Focus was enabling more self-care and self-management, but MDT could also refer to <b>other specialists</b> upon discretion.	Primary care <b>support</b> including <b>care plan</b> , with the aim of delivering more joined-up care and ability to draw from <b>other specialists</b> case by case as needed, eg palliative care nurses. MDT staff could organise prompt outpatient visits, eg dementia assessments.
<b>Organisation and delivery of MDT care</b>	<p>Across all three interventions:</p> <ul style="list-style-type: none"> <li>• A <b>single point of contact</b> was assigned to coordinate and/or oversee 'care navigation' for each patient.</li> <li>• In Fylde Coast, MDTs were mostly <b>co-located</b> working full time as a <b>dedicated</b> service alongside each other in their primary tasks.</li> <li>• In NEHF, the teams convened once a week for the MDTs as <b>designated</b> providers for their patients but continued to provide other non-MDT care within the community.</li> </ul>		

Multidisciplinary teams – intervention design	Fylde Coast Extensive Care Service (ECS)	Fylde Coast Enhanced Primary Care (EPC)	NEHF Integrated Care Teams (ICT)
<b>Patient information sharing</b>	An <b>MDT-specific database</b> was developed and used by staff. However, due to technology and IT constraints, this was not shared routinely with hospital staff, though specific MDT clinicians could follow up on this as needed. This shared care record was on the same EMIS platform and thus was integrated with GP records and information access.		NEHF aimed to create a shared care record, but this was delayed due to technology and information governance constraints. Patient information was instead shared verbally in the case management meetings.
<b>Overlap in treatment with other area-level interventions</b>	Approximately 260 ECS patients (16%) also used EPC at some point before the study end in April 2018. 8.7% of ECS patients and <10 EPC patients (<1%) were also enrolled in Lancashire & Cumbria Innovation Alliance Telehealth Test Bed. Unknown overlap with other vanguard interventions (Episodic Care, such as care signposting in pharmacies) but these programmes were relatively lower scale and were also generally available to all Fylde Coast residents.		Other vanguard interventions were generally available to all NEHF residents but introduced to different localities and over different time periods. The CCG estimates that at least one-third of ICT patients also received other vanguard interventions, eg Enhanced Recovery at Home.
<b>Care planning and coordination</b>	New patients were discussed following referral; an action plan was created and assigned to a primary contact within the MDT.		
	Each neighbourhood ECS team met <b>weekly</b> to discuss patients, as needed.	Each EPC team met <b>regularly</b> to discuss patients, as needed.	Each ICT <b>met weekly</b> to discuss patients, as needed. Farnham ICT held informal daily discussions.

Multidisciplinary teams – intervention design	Fylde Coast Extensive Care Service (ECS)	Fylde Coast Enhanced Primary Care (EPC)	NEHF Integrated Care Teams (ICT)
<p><b>Target population</b></p> <p><b>Procedure for determining eligibility</b></p> <p><b>Risk stratification</b></p>	<p>Originally aimed at the cohort with <b>highest risk</b> for hospital use:</p> <ul style="list-style-type: none"> <li>– 60+ years old</li> <li>– at least 2 long-term conditions (LTCs) from a pre-specified list*</li> <li>– predicted CPM<sup>†</sup> <b>risk score ≥20</b>.</li> </ul> <p>Eligibility <b>criteria amended</b> in Feb 2016 to allow individuals at least 2 prior A&amp;E admissions or 2 out-of-hospital contacts within last 2 months, instead of risk score ≥20.</p>	<p>Originally aimed at a moderate or '<b>second highest</b>' <b>risk cohort</b>, alongside ECS, with patients:</p> <ul style="list-style-type: none"> <li>– 16+ years old</li> <li>– predicted CPM<sup>†</sup> <b>risk score ≥20</b></li> <li>– attention to mental ill health and/or difficult personal or social circumstances.</li> </ul> <p>Expanded to include <b>professional clinical judgement</b> about anyone who could benefit from increased support.</p>	<p>Intended to reach patients with <b>highest need</b> and at <b>highest risk</b> of going into crisis, primarily assessed by <b>professional clinical judgement</b>.</p> <p>In practice, health professionals also selected patients based on who they considered would benefit most from a multidisciplinary approach.</p> <p><b>Risk scores</b> (using Johns Hopkins ACG system) initially planned but ultimately not used during the period of the study, except for a small minority of cases. Farnham used risk stratification and other patient data from March 2017.</p>
<b>Median age (IQR) of MDT patient</b>	<b>80 years</b> (73, 85)	<b>76 years</b> (63, 85)	<b>81 years</b> (72, 87)
<b>Source of referrals</b>	Originally only GPs. Later broadened to hospital and community professionals.	Most referrals were made by GPs, but also possible from others (community, mental health and social care staff).  Blackpool patients could self-refer but this was rare.	Most referrals were made by GPs, but also possible from others (community, mental health and social care staff).
<b>Assessment of eligibility</b>	<b>MDT assessed eligibility</b> ; patient or service could <b>refuse enrolment</b> for various reasons.  ECS patients could not simultaneously be on EPC. Typically, where a single individual did receive care from both MDTs, EPC was used as a step-down service from ECS.		Health professionals referring used their clinical assessment. Patients could refuse referral; all referred patients were seen by MDT.

\* Across coronary arterial disease (CAD), atrial fibrillation, congestive heart failure (CHF), chronic obstructive pulmonary disease (COPD), diabetes, dementia.

† Combined Predictive Model using inpatient, outpatient, A&E and GP data.

<b>Multidisciplinary teams – intervention design</b>	<b>Fylde Coast Extensive Care Service (ECS)</b>	<b>Fylde Coast Enhanced Primary Care (EPC)</b>	<b>NEHF Integrated Care Teams (ICT)</b>
<b>Discharges</b>	<b>All MDTs</b> were intended to provide temporary support to anticipate and prevent duplicative or some unplanned hospital use.		
	In the minority of cases where a patient received care from both MDTs, EPC was typically used as a ‘step down’ for ECS.		Patients referred to an ICT would stay on the ICT register in perpetuum, although differentiation was made between active and dormant patients.
<b>Relationship with end-of-life care</b>	End-of-life care pathways and their interaction with ECS/EPC had not yet been well defined during the follow-up period.		Palliative care nurses were not part of the core team but they could be called in for special end-of-life support.
<b>Unplanned changes or challenges during implementation</b>	<p>Fewer patients were recruited than targeted in the first year of implementation, ie time to reach ~1,000 patients was slower than anticipated for both programmes.</p> <p>Staff could not be fully recruited for all periods, and in particular staffing configuration for the ECS had to change. Whereas each of the teams was originally meant to be led by specialist GPs or medical doctors, recruitment difficulty meant these roles were ultimately led by advanced practice nurses.</p> <p>Estates issues led to some implementation delays, such that EPC was established 16 months after ECS (resulting in approximately half the maximum follow-up time we observed for the latter). In the first 3–6 months of planned implementation, patient recruitment or acceptance onto ECS from GP referrals was slower than initially anticipated.</p> <p>The initially envisaged programme underwent some refinements as not all of the initially proposed vanguard budget was confirmed for central funding.</p>		<p>Less systemic referral process due to lack of risk stratification tool at launch.</p> <p>Unplanned delay in implementing urgent care hubs in Farnham and Yateley due to estates issues.</p> <p>Fewer patients were recruited than targeted in first year of implementation, ie time to ~1,000 patients was slower than anticipated.</p> <p>There were some staff recruitment issues in some of the localities, resulting in a reduced core team during some periods. In particular, 3 of the 5 localities struggled to recruit mental health staff.</p>

**Table 2: Settings of the MDTs**

<p><b>NHS New Models of Care vanguard programme</b></p>	<p><b>Fylde Coast</b> MDTs funded as a 'multispecialty community provider (MCP)' model</p>	<p><b>NEHF</b> MDTs funded within 'Primary and Acute Care System (PACS)'</p>
<p><b>CCGs and population covered</b></p>	<p>Blackpool (~172,000 people at 21 GPs) Fylde &amp; Wyre (~155,000 people at 19 GPs)</p>	<p>NEHF (~225,000 people at 23 GPs)</p>
<p><b>Socio-economic characteristics</b></p>	<p>Urban Blackpool faces significant deprivation, health inequalities and low life expectancy, among the worst in England.  Suburban and rural Fylde &amp; Wyre has a similar profile to national averages but has a growing proportion of older people with multiple long-term conditions.</p>	<p>A mostly suburban area spanning Hampshire and Surrey with a relatively affluent population, mostly living within 1 hour's commute of London.</p>
<p><b>Historic integrated care interventions</b></p>	<p>Neighbourhoods in Blackpool had historically some multidisciplinary teams working in the community, whereas this was a new model and way of working introduced in Fylde &amp; Wyre.</p>	<p>Some work on integrating delivery of services and establishing a joint integration team had already started before NEHF received vanguard status, following funding from the Better Care Fund in 2014 to Hampshire County.</p>



**Table 3: Study design and characteristics of the original IAU evaluations**

<b>Study design aspects</b>	<b>Fylde Coast Extensive Care Service (ECS)</b>	<b>Fylde Coast Enhanced Primary Care (EPC)</b>	<b>NEHF Integrated Care Teams (ICT)</b>
<b>Start and end dates</b>	<p>Intervention patients were entered from first confirmed start of first spell on the MDT service to death, de-registration from CCG for primary care or end of the study period.</p> <p>The IAU only included MDT patients who were enrolled for at least 1 month (28 days), to allow for a minimal effective service 'dose'.</p> <p>Matched control patients were selected from the local area, with data for all potential controls re-entered for each month (to align to varying follow-up times of the intervention patients and allow for different start dates from the same control individual to potentially match multiple MDT patients). Their records were similarly censored upon death, de-registration or end of the study period.</p>		
<b>Maximum study follow-up</b>	<b>33 months</b> (mid-August 2015 to mid-April 2018)	<b>18 months</b> (mid-November 2016 to mid-April 2018)	<b>21.5 months</b> (end of July 2015 to mid-June 2017)
<b>Average study follow-up</b>	<b>393 days</b> (13 months)	<b>220 days</b> (7 months)	<b>202 days</b> (7 months)
<b>Matching</b>	<p>Local matched controls, ie the control patients were registered with a GP in the same area (neighbourhood or locality).</p> <p>Matching done within MDT-specific catchment areas (neighbourhoods/localities rather than across CCG/vanguard).</p> <p>Matching with replacement, ie the same control record could be matched to different treated individuals if most similar across baseline covariates to them. Control records could include multiple time periods of observation for the same matched control individual.</p>		
	<b>Matching based on patient characteristics and risk score (including GP data in Combined Predictive Model)</b>		Matching based on patient characteristics
	1:1 matching	2:1 matching (pre-specified for statistical efficiency with large N)	1:1 matching
<b>Sample size in original IAU evaluation</b>	<b>n=1,626 patients</b> in each group with matched records from 1,438 unique control individuals	<b>n=3,011 patients</b> in each group with matched records from 3,772 unique control individuals	<b>n=774 patients</b> in each group with matched records from 731 unique control individuals

**Table 4. Original evaluation findings: relative differences or average treatment effect on the treated for MDT users (95% confidence interval)**

<b>Study findings</b>	<b>Fylde Coast Extensive Care Service (ECS)</b>	<b>Fylde Coast Enhanced Primary Care (EPC)</b>	<b>NEHF Integrated Care Teams (ICT)</b>
<b>A&amp;E attendances*</b>	26% higher (15% to 38% higher)	40% higher (28% to 52% higher)	33% higher (16% to 54% higher)
<b>Emergency admissions*</b>	27% higher (15% to 41% higher)	42% higher (29% to 56% higher)	43% higher (23% to 67% higher)
<b>Chronic ACS emergency admissions*</b>	62% higher (33% to 98% higher)	32% higher (4% to 66% higher)	105% higher (32% to 222% higher)
<b>UCS emergency admissions*</b>	24% higher (3% to 48% higher)	46% higher (21% to 75% higher)	76% higher (35% to 129% higher)
<b>Average length of emergency stay, nights*</b>	N/A	N/A	33% higher (8% to 63% higher)
<b>Emergency hospital bed days</b>	8% higher (9% lower to 28% higher)	57% higher (31% to 88% higher)	N/A
<b>Emergency readmissions within 30 days of discharge</b>	N/A	N/A	4% higher (11% lower to 21% higher)
<b>Elective admissions (ordinary or non-regular day admissions)</b>	8% higher (5% lower to 24% higher)	2% lower (14% lower to 10% higher)	24% lower (2% to 41% lower)
<b>Elective bed days</b>	1% higher (34% lower to 54% higher)	18% lower (25% to 10% lower)	N/A
<b>Outpatient attendances</b>	10% higher (4% to 17% higher)	11% higher (5% to 16% higher)	3% higher (8% lower to 17% higher)
<b>Deaths in hospital</b>	43% higher (2% to 100% higher)	21% higher (12% lower to 67% higher)	27% lower (42% lower to 2% higher)
<b>Deaths* (placebo test)</b>	4% higher (16% lower to 28% higher)	58% higher (33% to 89% higher)	22% higher (12% lower to 68% higher)

Note: all relative differences are rate ratios, apart from deaths and deaths in hospital, which are odds ratios.

\* Outcomes proposed for the pooled analysis.

## Potential causes of increased emergency hospital use for patients enrolled in MDTs

There are several plausible hypotheses consistent with the results we observed and previous research<sup>1,10</sup>:

- H1** Residual unobserved confounding (differences in eg social context, end of life)
- H2** Patient-specific effects (subgroup heterogeneity)
- H3** Limited 'impactability' of riskiest patients
- H4** Delayed impact caused by patient and/or provider factors
- H5** Induced demand from clinical risk aversion
- H6** Hospital use does not adequately measure quality.

### **H1. Residual unobserved confounding accounts for impacts, as MDT patients were still different from their matched controls in unseen but important ways**

Unseen differences between the respective intervention and matched control groups could potentially account for some or even all the difference in hospital use between the intervention and control groups (unobserved confounding). This is the case when the matched control patients appear very similar to the intervention patients but have important unseen differences that may impact on their eligibility for MDTs as well as their hospital use. The risk of unobserved confounding is always a limitation of observational studies but is a particular concern where the intervention was available for the control individuals but they were not selected, even though they had similar observable characteristics. This is the case in these studies: controls were from the same areas as the patients enrolled in MDTs and appear very similar, but were not referred to an MDT (in the case of NEHF) or not referred for assessment and potential enrolment in an MDT (in the case of ECS and EPC).

For the original evaluations, although we adjusted for factors such as age, level of deprivation, LTCs and prior hospital use, we may not have captured other important factors that contributed to the decision to refer to an MDT. We found that there was strong statistical evidence that patients enrolled on EPC had a higher rate of death during the follow-up period, as compared to the matched control group. That suggests that there was greater baseline medical risk in this group prior to receiving EPC that could not be accounted for in the accessible data. However, the observed crude risk difference in deaths (3.6 percentage points higher, for an overall crude mortality rate of 10.6%) is highly unlikely to account for the observed differences in hospital activity (approximately 40% greater A&E attendances and emergency admissions among all EPC users as compared to matched controls). EPC patients who died during the follow-up period did not have hospitalisation rates that were over 10 times higher (40% divided by 3.6%) than the matched control group. We can investigate whether nearing end of life may be an unobserved confounder in the original analyses by subsetting patients depending on if they died within 90 days of being enrolled in an MDT/entering the study.

Unobserved confounding was a particular concern in the NEHF study, where health professionals referred patients based predominately on clinical judgement (rather than on objective criteria) on which patients would most benefit from support from a multidisciplinary

team, because they were at highest risk and need. This may have been based on information not available in our original data, such as not managing a long-term condition well or being socially isolated, where such factors also included an element of clinical judgement. In Fylde Coast, risk scores were used for selection to both types of MDTs but particularly EPC was aimed at individuals who a GP felt could benefit from increased support because of a lack of family support or other factors that meant they could not effectively manage their own health.

Stokes et al found that when identifying patients for MDTs, medical practitioners felt that the patients' needs were often primarily related to social factors such as isolation, poor housing or living arrangements, and other socio-economic issues.<sup>2</sup>

One such factor could be social isolation, which has been found to be associated with both increased morbidity and mortality.<sup>28,29</sup> Social isolation is a concept related to but distinct from loneliness; while loneliness is a subjective feeling associated with actual or perceived isolation, social isolation reflects a lack of social ties, social integration or sense of community.<sup>28,30</sup>

It is not possible to quantify social isolation in our data, but there is potential to identify proxies of it relating to social context. One of these is whether someone is living alone, and this has been found to be associated with a 50% higher likelihood of visiting A&E than those living with others.<sup>12</sup> However, while living alone could result in social isolation, it is not necessarily the case; approximately one-third of people aged 65 or over live on their own<sup>31</sup>, many of whom may have friends or family living nearby. Therefore, identifying whether somebody is living alone may not be sufficient to identify somebody who is socially isolated. Other, related, factors that may affect health outcomes and a health professional's decision to refer a patient to an MDT may be life events such as a recent bereavement<sup>32</sup> or losing one's job. Another factor could be housing instability. Social isolation is also likely to be associated with having LTCs (something we already account for in our analyses) or living and caring for somebody in the family with LTCs, as this can lead to the person being more housebound or having less time available for social interactions or leisure activities, especially if it may not be safe to leave the household member on their own. This may particularly be the case if the person has dementia, serious mental ill health or profound mobility problems. It may also be more difficult to look after one's own health while also caring for somebody else; therefore, this may be important social context to account for in its own right.

If in fact these social context factors were poorly matched in our previous studies, the original evaluations would be limited by biased information and the 'true' impact of these interventions ought to be reassessed in light of the newly observed data. However, it is also plausible that the conclusions from our original evaluations remain unchanged despite including further potential sources of unobserved confounding. Statistically, this issue cannot be eliminated or precisely quantified in an observational study.\* Yet careful study designs to evaluate impact can and do replicate the treatment effects of clinical and other health interventions seen from randomised controlled trials<sup>33,34</sup> and can match real-world experience more closely than the very controlled and sometimes artificial conditions of most trials.<sup>35</sup>

---

\* There are some statistical methods available, eg Rosenbaum bounds (<https://cran.r-project.org/web/packages/rbounds/rbounds.pdf>) which calculates how sensitive results are to unobserved confounding. However, these are not currently available for count outcomes.

There are several further hypotheses which relate to patients' 'impactability', ie the notion that some people may be more amenable to preventive care and that individuals at the highest overall risk for emergency hospital use do not necessarily correspond with the group of people for whom such hospitalisation can actually be prevented or avoided.<sup>36</sup> This is related to but distinct from the identification of unmet need and a non-linear impact on hospital use over time, as discussed in the fourth and fifth hypotheses below.

## **H2. MDTs may be best suited to support specific patient subgroups**

MDTs may be better suited to support certain patient subgroups, eg those with serious mental illness, diabetes or heart failure, rather than the general population of individuals who are identified as having high risk of hospitalisation in the following 12 months. It may be more difficult to detect a positive impact overall in evaluations of MDTs serving a broader general population if there is heterogeneity in characteristics and in the MDTs' effect on outcomes.

Although MDTs have been demonstrated to reduce emergency hospitalisation in several randomised trials, this data has been sparse and research protocols were of varying quality and high heterogeneity, eg evidence from the US where system incentives and clinical risk thresholds are quite different from the UK, and studies pre-dating the landmark 2012 Health and Social Care Act in the UK. In both cases, MDTs may be less effective or even counterproductive in England today, due to unignorable differences in how NHS and social care are now organised.

Recent evidence is equivocal about the impact of MDTs on hospital activity. Integrated care interventions in primary and community care settings, including MDTs, have a mixed record of success when applied to patients with general multimorbidity<sup>37</sup>, including no change in hospital use for 'primary intensive care' case management models in the US.<sup>38</sup> Multidimensional assessment and management of older people in primary care does not appear to impact patient outcomes in the UK.<sup>39</sup> However, other research, for example a multicentre randomised trial in the north of England, has found greater independence for older people needing rehabilitation when providing community-based MDTs.<sup>40</sup>

Historically, some but not all measures of reduced hospital use – for example shorter length of stay but not lower emergency admissions – have been demonstrated across systematic reviews of international trials in congestive heart failure<sup>41,42</sup> and stroke<sup>43</sup> and with relatively modest impacts in specific UK trials for severe mental illness.<sup>44,45</sup> Conversely, integrated care managers in the US have previously believed that serious mental ill health makes it less likely that case or disease management will reduce hospital use among patients who have these conditions.<sup>36</sup> Integrated care may reduce length of stay and long-term dependency among stroke patients.<sup>43</sup>

However, even if there are certain conditions that are more amenable to multidisciplinary support, these conditions may not be equally well supported by an MDT that cares for a broader range of conditions and needs as by one specifically targeting patients with a specific diagnosis, particularly if condition-specific staffing and skills are required for intervention effectiveness.

### **H3. There may be limited scope to reduce or impact emergency hospital use for the highest risk patients, especially those nearing end of life**

Linked to the concept of 'impactability', it may be that there is limited scope to reduce emergency hospital use for those identified as being at the highest risk, especially if they are nearing their end of life.<sup>10,36,46</sup> In qualitative feedback, Fylde Coast MDT staff noted for instance that some patients had elevated 'acuity' and therefore may not have represented a group for whom future hospital use could actually be avoided.<sup>16</sup>

If the aim is to reduce emergency admissions, it may be that MDTs should not target the patients identified as at highest risk. There may instead be more scope, especially at a population level, to reduce unplanned hospital use in a lower or more moderate risk group of patients whose long-term conditions have not yet progressed to the most serious or unmanageable stages. MDTs may be able to delay and compress future periods of frailty for a high-impact group of patients, but these may not be the same population that the vanguards were able to identify and target in their initial phase.

The currently available tools for risk stratification are somewhat useful but not definitive for targeting MDTs. While they use extensive data to predict hospital admission, there is still substantial likelihood of false negatives and positives in their use, and most of them do not address impactability directly.<sup>10,47,48</sup> These tools also cannot capture or report information that is not routinely collected in aggregated NHS data systems but which might be important for targeting MDTs, for instance on social isolation, homelessness, or similar personal circumstances. In some cases, continuous development and improvement of these tools is no longer supported (eg NHS support for continuous development on CPM was halted in 2011), which limits their continued applicability and validity. Such tools may be most useful when deployed alongside clinical judgement to target population health interventions.<sup>49</sup>

### **H4. MDTs increase hospitalisation initially but can sustain reduced use over a longer time period**

Greater hospital use by MDT patients may reflect a real impact that occurs in the year immediately following the intervention, but this may be offset by reduced hospital use over the longer run. In other words, it is possible that the short-run increases we observed are a typical impact of MDTs, but overall these are offset by reductions when considering a longer time period following MDT enrolment.

Such a pattern could be due to MDTs identifying urgent unmet medical needs that might otherwise have only been identified later at a more problematic stage or gone untreated otherwise. It is possible that for many of the people targeted by these MDTs – including those living alone or with other social factors that we have not previously been able to account for – additional care is needed in the short term to stabilise their condition and promote self-management after acute crises (where such crises signal eligibility for potential MDT benefit in the first place), but these interventions show preventive benefits only later over a longer follow-up period of years rather than months. In other words, it may be that reduced hospital use is only observable beyond the median 7–13 months of follow-up in the original IAU evaluations.

Alternatively – or additionally – a change in MDT impact over time may result from programme implementation lag. Consistent with change management theory<sup>50</sup> complex integrated care programmes like the MDTs may require significantly more time to ‘bed in’ than allowed for under vanguard funding arrangements. Operational setup and management of complex interventions needs time to take effect, as teams introducing these changes must set up new routines of work, typically within complex institutional bureaucracies, with staff learning from experience and monitoring over time about how to optimise their delivery to achieve strategic aims. For example, in Fylde Coast ECS, the first recruited patients were typically enrolled for longer than current patients, roughly 12 as opposed to 6 months, according to local programme managers<sup>16</sup> and consistent with observed data.<sup>5</sup> It is thus possible that vanguard MDTs can reduce (and potentially sustain lower) hospital activity over a longer follow-up because the programme can only make these impacts after an initial period of operational start-up and improvement. Alternatively, it may be the case that such a lag is inherent to the effect of MDTs, as provider teams may require a certain number of visits or time with each individual patient in order to optimise the support interventions required to prevent future hospitalisation.

The IAU has previously evaluated the Mid-Nottinghamshire integrated care vanguard, which included, among others, an MDT like the Fylde and NEHF teams.\* Evidence from Mid-Nottinghamshire suggests that there was an initial area-level increase in emergency hospital use, which was followed by a decrease in the longer term, apparent only after about 2 years after approval of the vanguard and 4–5 years after initiation of the integrated care transformation programme.<sup>51</sup> This is consistent with at least 2–3 years before a modest reduction or prevention of comparative rates of hospitalisation, as found in other recent peer reviewed research on all the English new care models vanguards<sup>19</sup> and pioneers<sup>52</sup>. In an accompanying study, we plan to replicate the Mid-Nottinghamshire evaluation methods for Fylde Coast CCGs (Blackpool and Fylde & Wyre) and NEHF. If this trend over time holds across the three areas, this would be compelling evidence that integrated care models overall – and potentially also MDTs – take a longer time to generate positive impacts. Further qualitative and process evaluation would be required to quantify the reasons for this delayed impact, but there are at least two plausible explanations.

##### **H5. MDTs may have led to induced demand, from health awareness among patients or heightened risk aversion among professionals**

MDTs may have led to patients being more aware of their health needs or risks, or health and care professionals being more risk averse, which in turn led to greater hospital and other service utilisation. This usage may address previously unmet medical need as described above and may therefore be justified or appropriate. However, quality of life impacts – as opposed to service utilisation and mortality trends – can only be measured reliably by further data collection with service users. This phenomenon, whether caused by increased care-seeking by patients or more frequent referral and admission by medical staff, is known as ‘supplier induced demand’ in health economics. We do not have directly observed or consistent data

---

\* In Mid-Nottinghamshire, the MDT and other vanguard initiatives were evaluated at an area or whole-population level using synthetic controls rather than as an individually matched study, as no individual-level MDT enrolment data were available. Therefore, a matched-control study was not possible there and it cannot be included in the pooled analysis.

on the clinical risk thresholds or self-reported health awareness across the MDT and matched control sample. As a result, we will not test this hypothesis in the quantitative analyses, but the review of reviews may identify relevant findings in other empirical studies.

#### **H6. Multidisciplinary integrated care teams may improve quality in ways that are not adequately measured by emergency hospital use**

The value of integrated care teams may primarily lie in improving other outcomes, such as coordination of care, patients' wellbeing or experience of care or better management of LTCs, rather than reducing emergency hospital admissions. However, these evaluations did not examine, for example, the impact of the MDTs in improving patient-recorded outcome measures of quality of life as they are not routinely collected or recorded in NHS administrative systems.

Unplanned hospital use is an imperfect proxy for quality of care. In many cases, particularly at the end of life, hospital use may not be preventable and may be medically appropriate, eg to check that a chemotherapy has not caused dangerous side effects. In such cases, hospitalisation is essentially inevitable and would not represent a bad outcome for patients, but may represent essential care that improves quality of life by alleviating pain and other symptoms. As such, increases in unplanned (ie unscheduled) hospital use as measured by A&E attendances and emergency admissions could be an appropriate outcome signalling greater quality of care where MDTs address otherwise unmet medical need, and where further complications or worsening of LTCs is not truly preventable. Separate quality improvement literature on end-of-life care has noted the costliness of hospitalisation that motivates many interventions. However, optimal clinical management is nuanced and typically considered in terms of compressing the acute use of medical resources at end of life, ie in reducing length of stay, rather than preventing hospital attendance or admission in the first place for patients with terminal LTCs.<sup>53</sup>

## **Objectives of the evaluation**

This analysis aims to understand why patients enrolled in MDTs have in general higher emergency hospital use than their comparison groups. As such, we will explore a number of questions relating to the hypotheses above.

The study will examine and contextualise impacts on unplanned hospital use for the 5,411 'treated' patients who used MDT services in Fylde Coast and NEHF during the period of the respective IAU evaluations. Although these impacts have already been studied as compared to local matched control groups for each study area, the goal of this pooled analysis is to further explore the impacts between and across all three evaluations.

The research questions we aim to answer are:

1. Can we improve on the comparability between the intervention and control groups by including variables relating to 'social context', which may act as proxies for social isolation or other factors that could affect both treatment assignment and outcomes? What is the effect of MDTs on emergency hospital use after these factors have been taken into account? **(H1)**



We can also explore whether a difference in deaths between the intervention and control groups explains the observed impacts on hospital use, in particular in the EPC study. EPC contributes most of the pooled study sample (3,011 out of 5,411 MDT intervention patients) but the difference in observed death rates between treated patients and matched controls suggests that unobserved confounders may be differentially affecting the outcomes. We will carry out a sensitivity analysis across the pooled sample and within the EPC study (see below) to check how robust the hospital use findings are, after adjustment for end of life. This sensitivity analysis will compare hospitalisation outcomes between a patient subgroup who were at the end of life against all other patients, and statistically assess whether the intervention effect is different between these subgroups after risk adjustment for known factors.

2. From these three evaluations, what is the generalisable impact of MDTs? Where differences in impact have been observed, are these explained by patient- and system-level differences (heterogeneity)? **(H2)**

The communities within the three CCGs under study vary across rural vs urban, deprivation levels and other geographic factors and are therefore quite broadly representative of the English population outside London, and our initial analyses independently found fairly similar treatment effects for the primary outcomes of A&E attendances and emergency admissions (Table 4). As there was relatively high consistency in treatment effects (ie increased emergency hospital use) observed across studies and outcomes, the pooled data will inherently confirm this qualitative finding. In other words, as all three previous studies agreed on the direction of impact (an increase), the pooled analysis will find an increase. However, pooling and increasing the sample size will allow us to narrow the anticipated quantitative range of this impact (as given by confidence intervals).

There are two levels of heterogeneity that may be addressed by contrasting effects observed in the pooled analysis: the first is geographic population or study setting differences. By pooling the studies and quantifying the impact of differences in CCG area (NEHF, Blackpool and Fylde & Wyre), we will be able to estimate the impact of geographic heterogeneity, which could reflect population differences, system or setting nuances, or differences between the MDT interventions themselves. This will particularly hold in comparing the ECS and EPC, respectively, between Blackpool and Fylde & Wyre CCGs. As we account for many population-level differences in the matching and regression covariates, such residual heterogeneity will point to system or setting differences driven by local care providers and/or latent population variables.

The second level of heterogeneity – differences in effect arising from variation in the MDT interventions – is more difficult to observe, but directional findings may be possible from the unique implementation of ECS and EPC in the same geographies. As both Fylde Coast MDTs were implemented as complementary interventions in two CCG populations, it is possible to more formally quantify contrasting effects arising from intervention as opposed to CCG differences (ie considering the impact of ECS vs EPC in Fylde & Wyre

CCG, where substantial patient overlap exists in observed risk factors such as age, pre-existing conditions, past hospitalisation and potentially social context). Furthermore, because Fylde EPC and NEHF ICT are more similar interventions (as they do not replace GP care) observed in somewhat differing patient populations, some inference is possible by comparing and contrasting the impact of these interventions among more vs less similar enrolled patient groups. The large sample of 3,011 EPC service users is widely distributed across risk factors, with substantial overlap in patient profile among specific subsets of intervention users to patients in both ECS and NEHF.

3. Among pre-specified patient subgroups, are there any who are more likely to see either a reduction or an increase in unplanned hospital use after receiving MDT services? (**H2, H3**). These subgroups will focus on pre-specified patients who have multimorbidity, frailty, or specific LTCs that may be mostly likely to benefit from MDTs (see section Methods: Patient subgroups). The specified LTCs – dementia, chronic pulmonary disease and serious mental ill health – are based on conditions specified in the NHS Long Term Plan.<sup>8</sup> We will also assess whether patients at the end of life are impacted differently to those who were alive for more than 3 months (90 days) after MDT enrolment. Using new data linkages, we will also assess the effect of MDTs on patients with certain social context factors, such as living alone.
4. Is there any evidence that differences in impact between the Fylde Coast MDTs and the NEHF ICT can be explained by the use of formal risk stratification tools in Fylde Coast? (**H2, H3**)

The sensitivity analysis for inclusion of CPM risk score information as a covariate in the Fylde Coast studies will be required to validate how to pool these data with NEHF, where risk scores were absent, and the interpretation of ATT estimates across all three programmes and CCGs. This analysis would also serve to quantify the impact of using risk stratification tools: for both ECS and EPC, we will compare the estimated treatment impacts when matching or weighting control with risk score information included vs not.

Although previous post hoc analysis of the Fylde ECS results did not suggest substantial differences in MDT impact by risk score, sample size and follow-up scores were relatively small. Thus, pooled analysis across both ECS and EPC may help identify clearer impacts.

Further research questions that we will attempt to answer through other strands of work are:

5. Is there evidence for a delayed exposure effect of MDTs, in which initially increased hospital use changes to a reduction over the long run once the intervention has become more embedded? This separate analysis will evaluate the effect of the whole population aged 65+ in each of the vanguard areas (NEHF, Blackpool and Fylde & Wyre CCGs) to investigate whether the vanguard interventions, of which the MDTs were the most important, had an effect on emergency hospital use over a longer time period (**H4** indirectly, via an accompanying synthetic controls study and review of reviews).
6. Is there evidence that MDTs have an effect on other types of outcomes, eg quality of life? As part of this project, a review of reviews of the literature will aim to provide wider evidence of the effect of community MDTs.

# Methods

## Study cohorts

The pooled analysis study cohort will consist of the original evaluations' study cohorts.<sup>13,14</sup> The same inclusion/exclusion criteria will be applied, apart from the criteria for 'slimming' potential donor controls by age, which has been adapted for consistency across the three studies: control pools are pre-specified to be 2 years around the minimum and maximum ages of all MDT intervention patients, by gender, in the respective MDT studies. There are a total of 5,411 individuals who received MDT care across NEHF and Fylde Coast, with a pool of potential control patients in excess of 130,000. The time periods vary across MDTs (NEHF July 2015 – June 2017, ECS August 2015 – April 2018, EPC November 2016 – April 2018).

## Sources of data

The IAU will use the same data as used in the original studies and for the same time periods<sup>13,14</sup>, including pseudonymised patient-level national SUS administrative hospital data for England, data on registrations and deaths derived from pseudonymised resident-level NHAIS data, as well as the pseudonymised local data supplied by the CCGs. This included a list of patients referred to each MDT, and their date of referral for NEHF and the equivalent information on patients enrolled in Fylde. In addition, we previously used a list of risk scores supplied by the Fylde Coast CCGs for all of their intervention and matched control patients.<sup>14</sup> These risk scores are based on the Combined Predictive Model and are calculated based on the risk of hospital admission in the following 12 months. No additional data was requested from local CCGs for this analysis beyond the scope covered by data sharing agreements supporting the original evaluations.

We will also access pseudonymised address information for all individuals registered with a GP, in the form of a pseudonymized ONS Unique Property Reference Number (UPRN), matched to each person's full address. These data are derived from NHAIS patient registration data, which is available to the National Commissioning Data Repository (NCDR)\*. By comparing the pseudonymised UPRN of each person, we can determine whether a person is living alone or with others, and link this to the individuals in our previous data sets.

Pseudonymised means that all direct identifiers (eg name, address, date of birth, NHS number for patients) are removed from the data. Pseudonymisation reduces the risk that individual patients can be identified from the data.

---

\* The NCDR holds patient-level hospital data on behalf of NHS England and NHS Improvement.

## Study endpoints

The overall outcomes underpinning this study relate to emergency hospital use, as this analysis aims to understand why patients enrolled in MDTs have in general higher emergency hospital use than their comparison groups. While these measures are not the only informative outcomes for assessing the value and impact of integrated care, national and local health and care commissioners in the UK set out to reduce unplanned hospital use by establishing these MDTs in the community, to complement and in some cases replace routine primary care. Measures of unplanned hospital use are conventionally used to summarise the health services impact of interventions, particularly in England.<sup>54,55</sup> Using common outcome measures allows for more straightforward comparison between different studies.

The **primary outcomes** to be examined are:

- rate of A&E attendances
- rate of emergency (non-elective) hospital admissions
- average emergency admission associated length of stay (LoS)
- rate of chronic ambulatory care sensitive (chronic ACS) emergency admissions
- rate of urgent care sensitive (UCS, also known as acute ACS or AACS) emergency admissions.

We will also examine:

- rate of deaths (independent of whether they occurred in and outside of hospital), as a diagnostic check. We use mortality rate as a 'placebo test', 'falsification', or 'negative control' for potential unobserved confounding, ie to check that the treatment and counterfactual groups are similar at baseline, as mortality risk may be a proxy for severity of disease. We typically do not expect relatively short exposures to MDTs to materially impact life expectancy. Prior evidence about MDTs and mortality is weak; based on previous programme theory and evaluation, it may be the case that improved quality of care through MDTs can delay death. Given the safety profile of MDT health coaching and interventions to coordinate communication and care referrals, we would not expect that MDTs plausibly increase deaths.

## Patient subgroups

These analyses will be carried out for the overall population of MDT-treated patients and the counterfactual, alongside up to seven subgroup analyses to assess patient-level heterogeneity. These subgroups are pre-specified below on the basis of past research on MDTs and English policies that have identified these groups as suitable for community case management.<sup>8</sup>

It is possible to identify further patient subgroups, but we have prioritised a set of seven key contrasts. Interpretation becomes more problematic whenever more subgroups are specified, and our statistical procedures are at greater risk of finding spurious 'significant' differences in impact purely by random chance due to multiple testing across many hypotheses using the same data set. This risk of multiple testing will be addressed by Dunn-Sidak corrections to hypothesis tests (as a more precise alternative to Bonferroni adjustment). We will report both the Dunn-Sidak corrected and uncorrected p-values arising from hypotheses tests of outcome modelling across the pooled sample and the pre-specified subgroups.

All below analyses will be undertaken as stratified subgroups, but will only be reported where there is a minimum sample size across the three evaluations of 50 MDT-treated patients and 250 potential control individuals, ie where statistical estimates are expected to be stable and reliable.

We will construct the counterfactual for the MDT-treated patients in each subgroup separately (unless otherwise specified below) using entropy balancing or inverse probability of treatment weighting (IPTW), used as appropriate.

1. **By social context:** based on the exploratory analyses for incorporating address-level data, we will determine the most useful subgroup based on the social context variables identified (eg living alone).
2. **By end-of-life status:** contrasting the effect of MDTs between those individuals who died within 90 days of enrolment compared to those who did not.<sup>56</sup> This will explore whether the effect of MDTs on emergency hospital use differs between patients nearing the end of life and those that are not, as well as removing a potential source of unobserved confounding. Controls have previously been matched to these individuals in each of the three studies without accounting for information on death outcomes. Using contrasts from these pre-selected matched control groups within each study thus avoids endogeneity in sensitivity analysis for the possibility that death occurring within the 3 months after MDT enrolment (ie 'end-of-life' status) confounds MDT effects because it is a proxy for other unobserved medical or social risks among intervention users that were not present among matched controls. For a pooled sample, a similar procedure will be applied in which re-estimation of treatment effects will separately be conducted for the intervention patients who were at end of life and a reweighted counterfactual among the control pool. In other words, we construct the counterfactual for MDT-treated patients by weighting and matching only on pre-intervention characteristics, but we will compare the observed ATT to what is observed when splitting the treatment group into end-of-life and non-end-of-life populations.
3. **By multimorbidity (LTCs):**
  - 2+ comorbidities, as defined by Elixhauser records flags in hospital<sup>57,58</sup> in the 3 years prior to study enrolment
4. **By frailty:**
  - 1+ marker of frailty, as defined by the frailty index flags<sup>59</sup> from the same data

**By specific comorbidity groups** aligned with target groups identified as priority patients for integrated care and primary care quality improvement in the NHS Long Term Plan:<sup>8</sup>

5. **Dementia** as determined by the Charlson index<sup>60</sup>
6. **Chronic pulmonary disease** as determined by the Elixhauser flag, as a proxy for 'respiratory diseases'<sup>57</sup>
7. **Serious mental ill health**, based on the definition used in the original NEHF evaluation<sup>\*</sup>, which was broadly based on the QualityWatch definition<sup>61</sup>

\* A subset of the mental ill health cohort who had at least one inpatient admission or outpatient appointment with a primary diagnosis of schizophrenia, bipolar disorder or psychosis (ICD-10 codes F20–29 and F30–31) (White et al, 2014) in a given year.

## Exploratory analyses

We will do some initial exploratory analyses to inform some of the final design decisions, as detailed in this SAP. These will be documented in an addendum to the SAP before starting the main analyses.

### Reweighting of the pooled sample to build a new counterfactual using entropy balancing

Directly combining the three original analysis data sets to specify subgroups is expected to result in some imbalance in the MDT and matched control groups, due to differences in characteristics used to select the counterfactuals (in particular, the use of risk scores in these ECS and EPC programmes and their evaluations, but not in NEHF). Original matching procedures were also not optimised for subgroup balance, but rather balance across the whole population sample. This would therefore introduce bias to the statistical estimates in crude post hoc analysis.

The previous method used for robustly generating the counterfactual in the IAU evaluations of MDTs – genetic matching (based on the generalised Mahalanobis distance<sup>22</sup>) – is attractive for reducing bias. It applies relatively loose parametric assumptions extending from the principles of propensity score matching to offer a principled, replicable approach to identifying the most similar control units from high-dimension, large samples of data. However, genetic matching is very computationally intensive and would require unwieldy calculation time for the pooled analysis.

An alternative algorithm to create a counterfactual, entropy balancing<sup>62</sup> is much quicker and more efficient than genetic matching, computing in a matter of minutes rather than weeks or months. This is because entropy balancing reweights all available control units to be similar to the treated sample, rather than selecting individually matched units to map to each intervention unit, which excludes many potential control units that fail to minimise the distance metric of interest. As such, it is less straightforward to interpret, but it generates an equally robust counterfactual in order to estimate treatment impacts. The Health Foundation data analytics team has previously used entropy balancing and found consistent results with genetic matching in evaluation of integrated care pathways in Tower Hamlets.<sup>63</sup> Other empirical literature using entropy balancing in health policy and economics has been widely demonstrated and accepted.<sup>64,65</sup>

As a check before overall application of this method, we will first validate that effect estimates for the impact of the three MDTs derived from applying the entropy balancing approach are consistent with the previously estimated impacts from genetic matching (see Table 4). Entropy balancing models are also subject to failure or unstable weights where intervention groups are very small, eg  $n < 50$ . In such cases, the more straightforward approach of IPTW yields more robust counterfactuals, with even less calculation time than entropy balancing.

In addition to standard diagnostics to confirm the convergence and general statistical validity of estimates generated by entropy balancing, we will validate the consistency of the entropy balancing effect estimates with the previous effect estimates generated from Mahalanobis distance matching (as a variant of genetic matching) from the original IAU evaluations. We aim for final effect estimates to be within 15% of the relative (not absolute)

magnitude of previously identified median ATTs, with overlap in 95% confidence intervals checked visually. Although we will aim to replicate our previous modelling procedures for genetic matching as much as possible, we will proceed with the estimates that provide the most statistical efficiency (ie lowest root mean square error (RMSE) and closeness to zero across covariates on standardised mean difference (SMD) plots) while remaining consistent under the above threshold.

Early exploratory analysis has found that further granularity to sub-CCG-level areas (ie 'localities' in NEHF or 'neighbourhoods' in Fylde Coast) leads to some small sample sizes (less than approximately  $n=50$ ). In these cases, the entropy balancing algorithm fails to appropriately converge or compute at all, or results demonstrate some extreme weights and SMD/RMSE that provides inferior counterfactual balance to IPTW, including after re-aggregation across the study sample. This result from 'exact matching' on sub-CCG area was shown despite relaxation of the covariate sets during reweighting to allow for as low as 70% predictiveness of the propensity to treatment and outcomes of interest. As a result of the lack of power and precision found from the small sample sizes when pooling from sub-CCG areas, we do not plan to use these results or pursue further estimates of heterogeneity at this level, though we may include a sensitivity check during outcome modelling of sub-CCG area level as an interaction term.

For consistency with previous IAU procedures, entropy balancing will use control pools as pre-specified in each of the original studies, apart from the criteria for 'slimming' potential donor controls by age. The original groups of potential donor controls were deterministically 'slimmed' by age, ie had matching calipers applied. However, the exact procedure varied somewhat between the three studies, as the primary purpose of this slimming was to reduce the computational load from control pool iterations in the genetic matching algorithm:

- In Fylde ECS, potential controls were subclassified by neighbourhood and gender, and only potential controls who were within 2 years of the oldest or youngest MDT intervention patient were retained. As a purely illustrative example, if the youngest female and male patients in neighbourhood 1 were respectively 68 and 57 years old, the control pool for neighbourhood 1 only included females who were at least 66 years old and males who were at least 55 years old, with similar logic applying to the maximum age.
- In Fylde EPC, these calipers were similarly applied within neighbourhood and gender, and set more narrowly to within 1 year of the minimum and maximum age. Additionally, potential controls who were older than the minimum age or younger than the maximum age were further manually 'slimmed' to reflect the extreme nature of minimum and maximum outliers in the age distribution, by removing 'in-between' controls who were at least 2 years older than the youngest EPC patient but also at least 2 years younger than the second youngest EPC patient. For illustration, if the youngest male EPC user in neighbourhood 1 was aged 21 and the second youngest male EPC user was aged 47, the potential control pool for males in neighbourhood 1 only included males aged 19–23 years old (as potential matches for the youngest patient) and otherwise males aged at least 45 years old (as no EPC patients were in the intervening age range to require a match).

This variation on the ECS procedure was carried out to pragmatically reduce matching computation time, as a much larger group of intervention patients and thus potential controls comprising essentially the entire CCG population were available. Unlike ECS where guidance was for patients to be at least 60 years old, in EPC all adults were eligible and some MDT patients were even minors.

- In NEHF, potential controls (consisting of individuals registered with a GP in NEHF) were only defined by age across the study sample, irrespective of locality (the equivalent to Fylde neighbourhoods). Thus, the youngest control individual retained for analysis was at most 2 years younger than the youngest overall NEHF patient (irrespective of gender), with this rule applied symmetrically for maximum ages.

For the planned analyses, the criteria for ‘slimming’ potential donor controls by age has been adapted for consistency across the three studies: control pools are pre-specified to be 2 years around the minimum and maximum ages of all MDT intervention patients, by gender, in the respective MDT studies.

### **Exploratory data analyses on heterogeneity, on stacked original analysis data (using matched controls)**

First, we will ‘stack’ or directly combine the three original matched data sets, according to a commonly mapped variable list that has been pre-defined (Table 5). Variable names and definitions are largely but not entirely consistent between the three IAU evaluations. Some will be adjusted so that both our stacked matched data set and our pooled data set have consistent definitions, variables and outcomes across all three studies. For example, for Fylde we will calculate emergency admission-associated length of stay, while the original analysis looked at emergency hospital bed days. Similarly, some of the flags used for risk adjustment will change, including removing some historic hospital activity and updated definitions for mental ill health flags in Fylde (see Table 5). In this initial phase, the individual intervention and matched control patients, as well as follow-up period, will be exactly the same units as those included in the original evaluations.

We will also add a set of new variables that capture social context characteristics, which aim to improve the similarity between the intervention group and their counterfactual. These are discussed in the next section.



**Table 5. Potential baseline variables for regression on stacked data sets and for entropy balancing on pooled data**

Variables at patient level	Changes to the original individual study analyses	
	Fylde Coast ECS and EPC	NEHF ICTs
<p><b>Demographics and socio-demographics</b></p> <ul style="list-style-type: none"> <li>• Approximate age at index date</li> <li>• Gender</li> <li>• Ethnicity across five high-level categories (White, Black, Asian, Mixed, Other, Unknown)</li> <li>• Average socio-economic deprivation deciles, based on the Index of Multiple Deprivation (IMD) 2015, available at LSOA level</li> <li>• Urban/rural classification at Lower Layer Super Output Area (LSOA) level, based on the 2011 census</li> </ul>	<p>Ethnicity in Fylde was White yes/no</p>	<p>Ethnicity in NEHF had three categories: White, Other, and Unknown</p>
<p><b>Prior hospital use</b></p> <ul style="list-style-type: none"> <li>• Number of emergency admissions in the last 60 days of the pre-period</li> <li>• Number of emergency admissions in year -1, the last year of the pre-period (ie 365 to 1 day before index date)</li> <li>• Number of emergency admissions in year -2 of the pre-period (ie 730 to 366 days before index date)</li> <li>• Number of chronic ACS emergency admissions in the last 60 days of the pre-period</li> <li>• Number of chronic ACS emergency admissions in year -1 of the pre-period</li> <li>• Number of chronic ACS admissions in year -2 of the pre-period</li> <li>• Number of UCS emergency admissions in the last 60 days of the pre-period</li> <li>• Number of UCS emergency admissions in year -1 of the pre-period</li> <li>• Number of UCS admissions in year -2 of the pre-period</li> <li>• Number of elective admissions in year -1 of the pre-period*</li> <li>• Number of A&amp;E attendances in year -1 of the pre-period</li> <li>• Number of outpatient attendances in year -1 of the pre-period*</li> <li>• Number of missed outpatient visits in year -1 of the pre-period*</li> <li>• Average length of stay following emergency admission in year -1 of the pre-period</li> </ul>	<p><b>Removed:</b></p> <ul style="list-style-type: none"> <li>– Number of A&amp;E attendances in the last 60 days of the pre-period</li> <li>– Number of A&amp;E attendances in year -2 of the pre-period (ie 730 to 366 days before index date)</li> <li>– Number of emergency bed days in year -1</li> </ul> <p><b>Added:</b></p> <ul style="list-style-type: none"> <li>– Average length of stay following emergency admission in year -1 of the pre-period</li> </ul>	<p><b>Removed:</b></p> <ul style="list-style-type: none"> <li>– Number of emergency bed days in year -1</li> <li>– Average length of stay following elective admission in year -1 of the pre-period</li> <li>– Number of emergency readmissions in year -1 of the pre-period</li> </ul>

\* Elective and outpatient use in the pre-period will not be included in the entropy balancing weighting but may be included at the regression stage.

<p><b>Health variables*</b></p> <ul style="list-style-type: none"> <li>Elixhauser list of comorbidities, identified in the pre-period</li> <li>Specific comorbidities linked to frailty, identified in the pre-period</li> <li>Selected comorbidities predictive of hospital emergency admission, as identified by the IPAEOPGP model, identified from SUS data in the pre-period: Charlson index, myocardial infarction, cerebrovascular disease, dementia and cognitive dysfunction</li> <li>Number of Elixhauser comorbidities in the pre-period</li> <li>Number of frailty comorbidities in the pre-period</li> <li>Mental ill health and severe mental ill health (as defined in NEHF SAP)</li> </ul>	<p>Number of Elixhauser comorbidities in the pre-period did not count peripheral vascular disease in error</p> <p>Different definition of mental ill health in Fylde</p> <p><b>Removed:</b></p> <ul style="list-style-type: none"> <li>Frailty variables delirium, senility and dementia were redundantly included when these were also captured using the composite variable cognitive impairment</li> <li>Number of comorbidities (frailty, Elixhauser and variables from IPOPAEGP) due to double counting</li> <li>The risk stratification score and the associated derived GP data for the risk scores (frailty score and list of long-term conditions)</li> <li>Some overlapping variables, eg for mental ill health and rheumatoid arthritis</li> </ul>	<p>History of peripheral vascular disease (an Elixhauser comorbidity) was previously omitted in error</p> <p>Number of Elixhauser comorbidities in the pre-period did not count peripheral vascular disease in error</p> <p>For Elixhauser comorbidities where conditions were differentiated by level of complexity, a patient could be recorded as having both levels. For the total number of Elixhauser comorbidities, patients recorded as having both cancer with and without metastases were double counted (in both NEHF and Fylde)</p>
<p><b>Support</b></p> <ul style="list-style-type: none"> <li>Living in a care home</li> </ul>	<p>Not included in original study</p>	<p>Included in the original study but using an older address matching algorithm</p>
<p><b>Time period</b></p> <ul style="list-style-type: none"> <li>Index date/period (by quarter)</li> </ul>	<p>Unchanged</p>	
<p><b>'Counterfactual' level of care available outside the MDT</b></p> <ul style="list-style-type: none"> <li>CCG</li> <li>Neighbourhood/locality in which the patient is registered with a GP</li> </ul>	<p>Unchanged</p>	
<p><b>Social context</b></p> <ul style="list-style-type: none"> <li>Living alone</li> <li>Living with somebody with dementia or frailty<sup>†</sup></li> <li>Recently living alone (change in last X months)<sup>†</sup></li> <li>Bereavement in last X months<sup>†</sup></li> <li>Number of moves in last 12 months</li> </ul>	<p><b>New variables to be refined and added as a key output of the pooled analysis</b> – ATT estimated effects from the previous evaluations may or may not be robust to inclusion of such social context factors when identifying and matching/weighting for potential controls. If ATTs are not robust, the analysis will demonstrate clear evidence of previously unobserved confounding</p>	

Note: these are potential covariates and may not all have been included in both the matching and regression stages in the original studies (see technical reports for further details).

\* Health variables are identified from hospital records in the 3-year pre-period.

† Exploratory analysis will establish exact definition (see section on social context).

The Elixhauser and Charlson lists of comorbidities are routinely used for risk adjustment.<sup>57,58,60,66</sup> The Elixhauser list is broader than the Charlson list and consists of the following 30 comorbidities: congestive heart failure; chronic pulmonary disease; hemiplegia or paraplegia; metastatic solid tumour or metastatic cancer; acquired immune deficiency syndrome or human immunodeficiency virus; peripheral vascular disease; cardiac arrhythmias; valvular disease; pulmonary circulation disorders; hypertension (uncomplicated); hypertension (complicated); other neurological disorders; diabetes (uncomplicated); diabetes (complicated); hypothyroidism; renal failure; liver disease; peptic ulcer disease (excluding bleeding); lymphoma; solid tumour without metastasis, rheumatoid arthritis or collagen vascular diseases; coagulopathy; obesity; weight loss; fluid and electrolyte disorders; blood loss anaemia; deficiency anaemia; alcohol abuse; drug abuse; psychoses; depression.<sup>57,58</sup> Most of the comorbidities on the Charlson list are covered by the Elixhauser list.<sup>58,60,66</sup> One notable exception is dementia; therefore, dementia is included separately (see Table 5).

The following list of comorbidities is considered linked to frailty: anxiety or depression; functional dependence; falls and significant fracture; incontinence; mobility problems; pressure ulcers; and cognitive impairment (composite of delirium, dementia and senility), consistent with other IAU analyses.<sup>59,67</sup>

The IPAEOPGP is a risk prediction model of hospital admissions that builds on and improves on earlier models such as the PARR algorithm and identifies a number of comorbidities predictive of hospital admission.<sup>68,69</sup> Selected variables from IPAEOPGP identifiable in inpatient hospital records and not captured in the Elixhauser list were included in the list of potential baseline variables.

After this **stacked** data set is generated across MDT-treated units and control units originally selected within each study by genetic matching, we will run standard quality checks for completeness and plausibility across the sample, in order to avoid the inclusion of erroneous data. In certain cases, we may address missing data by appropriate multiple imputation or complete case analysis.

We will **divide the data set into the pre-specified subgroups detailed above** and check the sample sizes. Where **the pre-specified subgroups are too small** (ie fewer than 50 intervention patients across the stacked data), no further analyses on that subgroup will be done. This is because small sample sizes will compromise our ability to generate reliable and stable estimates of the impact of MDTs and there will be lower power to detect effects.

If, however, the pre-specified subgroups are large enough (at least n=50 interventions, and at least n=250 potential controls in the donor pool for weighting), we will address the subgroup analysis through creating new counterfactuals for each subgroup by reweighting using entropy balancing methodology as detailed above.

## Exploratory data analyses: improving on unobserved confounding by exploring social context factors and a proxy for end of life

By using pseudonymised address information for all individuals registered with a GP from the NHAIS data and linking this to SUS, we will investigate the relevance of a number of variables relating to social context, which may be proxies for social isolation, instability or other factors that may affect an individual's risk of emergency hospital use and may be a referral criterion for MDTs.

We will identify and explore a number of indicators:

1. **Living alone flag.** This variable shows whether a person was living alone at baseline (start date, or index date). For intervention patients, this will be their status in the month prior to being enrolled in an MDT. For potential controls, this will be calculated monthly, for each of the monthly index dates that were assigned, between the start of the study period and 1 month before the end of the study period (which was the last date a patient could join the study, as per the original studies, to ensure at least 1 month of follow-up).
2. **Living with somebody who has pre-specified long-term conditions (dementia or frailty, to be determined).** This variable captures whether a person, at baseline, is living with somebody who has at least one pre-specified condition. We will investigate dementia and a broader group of indicators of frailty (including dementia) as two potential conditions or characteristics that may lead to social isolation for a carer if they do not feel able to leave the person at home on their own or have difficulty in managing their own conditions or accessing the care they need. Other potential variables which were discarded due to likely small sample sizes in our data were serious mental ill health and paralysis. As in the original analyses, conditions are identified from the patient's hospital SUS records in the 3 years prior to the index date, using the same definitions as for patients' baseline characteristics. For dementia, we will use the definition used in the Charlson index,<sup>58,60,66</sup> for frailty, we will use the Soong definition.<sup>59</sup> The 'living with somebody with dementia/frailty' flag will not apply to individuals living in care homes or at addresses where seven or more people are registered. This is because larger numbers could be indicative of other living establishments, eg prison. According to the Office for National Statistics, approximately 2% of households in the UK contain six or more people<sup>31</sup> therefore setting the threshold at seven will exclude only a very small percentage of households (<2%).
3. **Recent change to living alone.** This variable will capture the household shrinking from two or more people to one person, within a specific period (lookback period to be determined). This shrinkage could be as a result of bereavement, divorce, partner moving to a care home or a single parent's child moving out.
4. **Bereavement.** Any bereavement within a household, independent of number of people left in the household within a specific period (lookback period to be determined). This variable will not apply to individuals living in care homes or at addresses where seven or more people are registered.
5. **Change: Multiple moves in last 12 months.** This is where a person changed address at least twice, as identified as having had three or more different pseudo-UPRNs, within 12 months.

For bereavement and recent change to living alone, before determining the lookback period we will first explore the time period that best characterises the patients that were enrolled in MDTs, ie affected treatment assignment. We will explore 3 and 12 months' lookback. Based on baseline characteristics, we will also determine whether to look at dementia or frailty for other household members.

We will start by examining whether there are differences in the social context variables at baseline between the original intervention and matched control patients, in each of the three studies individually. For those variables that look promising, we will quantify their effect on the outcomes for the original intervention and matched control groups, by including these social context variables in a subsequent outcome regression.

We may also explore five or fewer pre-specified two-way interaction effects, eg:

- recent bereavement with living alone
- mental ill health with bereavement
- mental ill health with recently living alone.

If these variables are shown to be important to adjust for we will include them in the entropy balancing and subsequent analyses, as appropriate.

We will also create a flag for whether the person died within 90 days of being enrolled in an MDT or being selected as a control, as a proxy for nearing their end of life. We will check for any imbalances between intervention and matched controls in this variable, which may suggest previously unobserved differences in disease severity or frailty that may confound emergency hospital admission rates. We will also check for any imbalances between intervention patients in the three studies, which would indicate heterogeneity.

We will create subgroups based on this EOL flag and check the balance between the intervention and matched control groups on patient characteristics.

### **Handling of differential information from risk scores**

As previously noted, there were differences in what characteristics were matched on (eg risk scores in ECS and EPC studies but not in NEHF). It is likely that SMDs and other balance diagnostics across covariates will reveal imbalances between the studies upon stacking the sample; this is a particular issue because risk stratification used for determining treatment assignment in Fylde Coast (and therefore included in matching and regression there) was not used in NEHF.

ECS and EPC used the CPM,<sup>15</sup> which derives risk scores from inpatient, outpatient, A&E, and GP data sources to inform eligibility and selection of patients to receive MDT services in the Fylde Coast.<sup>14</sup> In NEHF, the intention was to similarly deploy the Johns Hopkins Adjusted Clinical Groupings (ACG) system.<sup>70,71</sup> The ACG does not directly map onto the CPM in any case; however, due to technical and information governance constraints, the ACG risk stratification tool was not deployed in NEHF for the ICT during our study follow-up period, except for a small minority of patients in one of the five localities.<sup>13</sup> As it had not been generated for the vast majority of individuals and did not inform their enrolment or case management of the ICT intervention, it was not included in the NEHF analysis, either in matching or regression stages.

Because the CPM is not straightforward to generate, it is not feasible for the IAU to carry this out for the NEHF sub-sample in order to make it generalisable with the Fylde Coast MDTs. Although CPM relies on conventional logistic regression for statistical modelling, it maps onto a complex set of predictive covariates from over 1,000 diagnostic codes, not only across Hospital Episode Statistics but also from general practice Read codes, which are not routinely available to the IAU. Notably, CPM has not been funded for continued development within the NHS since 2011, which is likely to impair the predictive value of its algorithms given major health system changes that have changed patterns of hospitalisation in England over the past decade.

Therefore, **we will instead generate comparative effect estimates from the ECS and EPC samples using the entropy balancing algorithm as described below, but remove the risk scores** that were used for matching and regression in the original Fylde Coast evaluations. Alongside other procedures as described above to map all variables across the three data sets, this will allow for greater comparability and would in principle standardise the risks of omitted variable bias when estimating results across Fylde ECS, Fylde EPC and the NEHF ICT.

Furthermore, as the IAU has already included past hospital activity from inpatient, outpatient and A&E data sources across the evaluations of these MDTs, much of the information used in the CPM risk score is already captured in our analysis (ie risk score is largely collinear with other SUS covariates already included in our analysis). We believe that the statistical value of including the CPM risk scores in Fylde Coast was primarily from inclusion of GP-level data, for instance on newly diagnosed LTCs recorded in primary care that had not yet been included in hospital-level records. In validation work on the successor model for the CPM, Billings et al had found that inclusion of GP data improved the predictive accuracy (as measured by the c-statistic) by an incremental amount, from 75.2% to 78.0% across the English patient sample of 1.8 million patients from five primary care trusts.<sup>68</sup>

As a check on the robustness of our proposed approach to dealing with risk scores, we will pursue one of the following two options:

1. **Repeat outcome modelling** within the Fylde MDT sample, using a newly generated counterfactual from entropy balancing that excludes risk scores.
  - a. If impact effect estimates in the ATTs are largely unchanged (ie no greater than 10% difference in estimated median effects), then we will be satisfied that our estimand is robust for inclusion vs exclusion of risk score information. Although the previous studies found strong statistical evidence that CPM risk scores have a meaningful interaction with the treatment effect, the magnitude of this interaction may be relatively small. If it is the case that omitting the risk scores yields <10% relative rather than absolute difference in the effect estimates (ie for a previously estimated 40% increase, the new estimate is between 36% and 44%), then we will conclude the additional information provided by risk scores is not clinically significant for MDT policy. In this case, we will conclude that outcome modelling with and without the risk scores is sufficiently similar, ie the pooled analysis and the individual study estimates for Fylde Coast are directly comparable ATT estimates.

- b. If effect estimates are changed by more than 10% from the inclusion of risk scores, we will conclude that the NEHF modelling is subject to omitted variable bias. In such a case, we will treat the lack of risk scores from NEHF as a missing data problem, in which the missing risk scores can be imputed from the known correlation with observed variables of historic hospital activity in Fylde Coast. Although we lack primary care data that contributes to the risk score calculation, we believe that a robust missing data multiple imputation approach such as predictive mean matching (PMM) will account for random variation in these unobserved variables. Extrapolating the relationship between CPM risk scores and historic hospital use from Fylde Coast to NEHF is theoretically valid, as CPM algorithms were optimised for predictiveness on the whole English population and thus would follow the same calculation method between different CCGs.
- c. Even if the condition (a) is satisfied, we may nevertheless carry out (b) as a sensitivity analysis for robustness of the pooled effect estimate overall, if time allows.

## Comparison of matching alternatives in the pooled data set

If the validation process of entropy balancing does not produce consistent results across the studies, we will also compare effect estimates as above with two older matching/weighting methods that are more canonically used in the medical statistics and health policy literature. These are inverse probability of treatment weighting (IPTW), and propensity score matching (PSM).<sup>72</sup> However, these two methods rely on stronger inferential assumptions and can thus have higher risk of model misspecification and thus biased impact estimates than genetic matching and entropy balancing. Please see scoping document for IAU Matching Alternatives project for more detail (available on request).

If entropy balancing does not produce consistent results to genetic matching and time does not allow for exploring other alternative methods, we will consider the reliability of the 'stacked' analyses based on the original matching and if this is considered good enough, we may run analysis on this instead.

## Main analysis

### Accounting for some unobserved confounding

If the exploratory work on social context showed that there was poor balance between the intervention and matched control groups on these variables and that they may affect outcomes, we will rebuild the counterfactual by reweighting the pool of unmatched controls within each of the three intervention data sets, including these new indicators using the entropy balancing algorithm. We will then be able to compare the estimated MDT effects with and without the social context flags.

The difference in the estimated impacts of MDTs with and without the social context indicators will constitute an informal sensitivity analysis for previous unobserved confounding. These results will be particularly notable if the effect estimates are substantively different, for instance if one or more indicators interacts with other covariates to have a qualitative impact

on our previous findings, or if any given interaction term contributes to the final outcome model based on statistical evidence ( $p < 0.05$  for inclusion in the model, from standard joint probability F-tests). If this is the case, this may provide a quantifiable and compelling argument for the value of including this flag in routinely collected data sets, risk stratification algorithms and health service evaluations of all integrated care interventions in England.

We will also carry out sensitivity analyses comparing the MDT impact among patients who died within 90 days vs those who did not, across the pooled sample. The sensitivity analysis of 'end of life' is not sufficient on its own to resolve the issue of unobserved confounding, in particular because it is possible though unlikely that increased hospital use is on a causal pathway between MDT usage and death. However, prior evidence about MDTs and mortality is weak, with previous research not finding that service MDTs focused on delivering intangible services such as health coaching have any impact, much less increased risk of death. This sensitivity analysis is particularly relevant to the results seen in Fylde EPC, but may also be important for clarifying findings from the NEHF ICT and Fylde ECS. We will report the findings of the sensitivity analysis for the overall pooled sample after inclusion of the social context flags, and may further disaggregate this by MDT if the effects of the sensitivity analysis show substantive heterogeneity between study settings.

### Pooled and subgroup analyses

The value of the pooled analysis is primarily in surfacing potential heterogeneity in the impact of MDTs for **patient subgroups** where the individual evaluations did not previously contribute large sample sizes, eg end-of-life patients. Pooling allows us to increase sample sizes, eg for congestive heart failure, that may be more viable in modelling across all three studies where effects could not be statistically determined in the individual samples.

As in the original studies, we will **estimate outcomes** based on four sets of covariates: all 'base' covariates (as described in Table 5); a smaller set of 'core' covariates that exclude historic hospital activity; covariates selected from data-driven lasso regression; and unadjusted or crude results. The lasso is a variation of ridge regression that applies a machine learning algorithm to principled variable selection – additional terms are penalised such that included variables achieve an optimal and replicable signal-to-noise ratio and improves upon the ad hoc nature of previously conventional procedures such as stepwise selection.<sup>23</sup> The final model will always be adjusted for covariates unless the algorithm entirely failed to converge due to sparse data. The most appropriate generalised linear model, selected from several pre-specified distributions, and covariate list will be chosen based on the Akaike Information Criterion<sup>24</sup> and other diagnostics, such as comparing observed vs predicted zeros (for count variables), and root mean squared errors (RMSE). Depending on the outcome being modelled, the likelihood distributions used are typically Poisson or inverse negative binomial (for count outcomes) – potentially with some corrections for zero-values inflation or over-dispersion – or logistic (for binary outcomes).

For the overall pooled data set, we will generate generic **forest plots** per conventional meta-analytic summaries that depict the weighted ATT estimates from across the three evaluations, and also compare the results from fixed-effect vs random-effect models (where the latter will be more valid if the study samples contribute 'true' heterogeneity). The results of



the model fitting may also be reported as tabular effect estimates and 95% confidence intervals. We will also calculate and report commonly accepted measures for study heterogeneity in the meta-analysis, namely the  $H$  and  $I^2$  statistics<sup>73</sup> as well as  $\tau^2$ <sup>74</sup>. We will also report the difference in effect estimates by fixed and random effects for the MDT treatment.

In terms of **pre-processing or counterfactual construction** prior to outcome modelling, we plan to use entropy balancing. If entropy balancing generates consistent estimates with the previous IAU evaluations of these MDTs when specifying the same set of variables, we will then apply it to the overall as well as subgroup pooled analyses as specified above, using a set of covariates as described above, taking into account the findings from the exploratory stage. Where entropy balancing fails to converge due to data sparsity, or where we ascertain that its underlying assumptions are invalid given extreme weights identified in diagnostic plots (pseudo-propensity scores) and statistical tests (variance ratio, Kolmogorov-Smirnov), we may use inverse probability of treatment weighting (IPTW) as an alternative where the hypothesis is viewed as particularly relevant by policymakers. However, it is worth noting that IPTW and related propensity score matching or weighting methods are subject to greater relative risk of model misspecification and thus bias in the outcome results.<sup>75</sup>

For the overall analysis, entropy balancing will be applied within each MDT or evaluation (of ECS, EPC, and ICT respectively), and the set of Fylde MDT-specific CCG contrasts. We will also carry out entropy balancing across the entire pooled sample, ie allowing for conditional exchangeability of control individuals between different CCGs and intervention teams. We will use standard diagnostics to confirm the convergence and general statistical validity of estimates generated by entropy balancing. We will proceed with the estimates that provide the most statistical efficiency (ie lowest RMSE and closeness to zero across covariates on SMD plots) while remaining consistent under the above threshold. This implies up to eight causal contrasts:

1. Entropy balancing **across the entire pooled sample**.

This assumes conditional exchangeability of all potential control units to all MDTs and study areas (CCGs), ie that treatment is homogeneous regardless of exact MDT configuration and local study setting factors, so long as observed demographics, comorbidities, and historic hospital use are adjusted. In other words, potential control individuals from all three CCGs and all three MDTs are included for the single effect estimate of MDT impact across all. This method is least similar to the previous IAU procedures, and features the largest sample size (n=5,411 interventions with ~2 million potential control individuals).

2. Entropy balancing **by MDT or intervention study**, ie three contrasts, across **Fylde Coast ECS, Fylde Coast EPC and NEHF ICT**, then combined via a weighted estimate from across the three samples.

This assumes conditional exchangeability of all potential control units within a given MDT intervention, eg all potential control individuals for ECS from either Blackpool or Fylde & Wyre CCGs may be included in the effect estimate for that MDT, and similarly for EPC.

3. Entropy balancing **by MDT plus CCG**, ie four contrasts, reweighted across respective study samples, across **ECS-Blackpool, ECS-Fylde & Wyre, EPC-Blackpool and EPC-Fylde & Wyre**.

This comparison is of particular interest to give a directional signal of area or setting vs intervention effects, and assumes no conditional exchangeability of out-of-area individuals or those who could have received a different MDT from the one in question. The NEHF ICT is not included as it was a single programme implemented in a single area, and is thus identical to the contrast in #2 above. This method is the most similar to previous IAU genetic matching procedures, and features the smallest sample sizes for algorithm convergence as the intervention and control pools are subdivided by both MDT and CCG.

Exploratory analysis has found that further granularity in the above contrasts to sub-CCG-level areas (ie 'localities' in NEHF or 'neighbourhoods' in Fylde Coast) leads to small sample sizes less than approximately  $n=50$  that fail to appropriately converge, or which demonstrate some extreme weights and SMD/RMSE that provides inferior counterfactual balance to IPTW, including after re-aggregation across the study sample. This result from 'exact matching' on sub-CCG area was shown despite relaxation of the covariate sets during reweighting to allow for as low as 70% predictiveness of the propensity to treatment and outcomes of interest. We will, however, include a sensitivity check during outcome modelling after the entropy balancing of sub-CCG area, by including this as an interaction term in the main pooled analysis. Entropy balancing will use control pools that are pre-specified to be 2 years around the minimum and maximum ages of all MDT intervention patients, by gender, in the respective MDT studies. This is identical to specifying a fixed caliper or exact match criteria for the age covariate, by gender and study. This procedure is most similar to the potential control pool identified for the NEHF study, and ensures that all control individuals who were previously selected by genetic matching in the previous IAU evaluations can be included in the pooled analysis. Potential control units who had been excluded for falling outside of the neighbourhood-specific age calipers in Fylde ECS and EPC are now included in the pooled estimate, as computation time is less of a concern and in order to allow for the greatest chance of finding an optimal counterfactual across all control individuals when applying different assumptions of exchangeability across MDT and study area.

We will initially identify subgroup effects based on the statistical significance and absolute magnitude of their coefficients during outcome regression. Where these terms are significant, we may carry out further post hoc stratified analysis by repeating the weighting algorithm for MDT patients in this subgroup only, in order to generate conditional effect estimates that may be of particular interest to policymakers in planning future targeted MDT implementation. In other words, we may run the entropy balancing algorithm up to seven times to obtain an appropriate counterfactual for each of the patient subgroups of interest. Combined with the reweighting of the overall MDT pool and each of the MDTs, this implies up to 15 sets of entropy balancing and/or IPTW results.

However, we will drop subgroup analyses where there appears to be poor 'overlap' of key characteristics such as age or gender distribution between the MDT patients and the reweighted controls. This is because such exchangeability between intervention and counterfactual populations is a prerequisite for the validity of treatment effect estimates.<sup>62,72</sup>

For example, it will not be valid or useful to compare hospital activity among a population of MDT patients with dementia who are primarily female and over 70 with a population of control patients who are primarily male and under 50. Where exchangeability is poor, entropy balancing typically fails to converge due to data sparsity. PS or IPTW may converge but results should be interpreted cautiously due to the risk of model misspecification and biased estimates. As poor overlap signals missing data observations that would otherwise have to be entirely extrapolated, no statistical method can address this flaw if it arises.

In the event of non-significant regression coefficients, small sample sizes, poor intervention-counterfactual overlap, or other issues we may uncover during the exploratory analyses that constrain statistical feasibility, we will only carry out and report a smaller set of regression-adjusted subgroup analyses – where we are confident the results are meaningful and reliable (ie overlap appears to be satisfied and statistical models including many covariates are successfully fitted). We anticipate that some of the subgroups will most likely fail one or more of these conditions.

To **check for outliers** in the pooled sample and subgroup samples that may skew or overly bias our estimates, we will use **standard visual plots** such as covariate plots and quantile-quantile ('Q-Q' plots) to inspect for unexpected or problematic patterns in regression results. Although we do not anticipate this to be an issue as we do not apply simple linear regression in any cases, we may resolve problematic covariate-outcome patterns using standard transformations (eg higher-order terms, log variables). For entropy balancing/PSM/IPTW, we will **apply further matching diagnostics of balance** (eg standardised mean difference plots or SMDs, Kolmogorov-Smirnov test (KS), pseudo-propensity score plots (PPS), variance ratio tests (VRT)). Residual imbalances above a 10% SMD will be noted, and where they show systematic and large bias, we may drop the corresponding subgroup analysis. With the entropy balancing algorithm, SMDs are typically optimised to zero across (nearly) all covariates as an inherent property of the algorithm, and merely signal that the mathematical model has successfully calculated. Thus, other diagnostics such as PPS plots are important to check that the entropy balancing algorithm is not invalid, eg due to extreme weights applied to a very small number of control units relative to the intervention group. For the exact statistical tests (KS, VRT), we pre-specify a failure threshold of  $p < 0.05$ . In all cases, we will report our diagnostics in technical appendices to the main outcome report, particularly where judgement involves visual inspection.

## Conventional pooled analyses on the ‘stacked’ matched data set – in case no valid alternative to genetic matching is found within the constraints of the timelines

If we do not find a valid alternative to genetic matching within the timelines set, we may use the stacked matched data set to generate crude outcome comparisons between MDT users and their matched controls, overall, for each study and for all subgroups.

For all analyses and finalised list of subgroup analyses, we will apply conventional meta-analysis techniques based on assumed homogeneity of effects. We would then generate **estimated treatment effect estimates across all three MDTs**. The effect estimates from the three studies should be more numerically **precise** due to the larger sample size available from pooling, and based on the previous ATTs we can already anticipate that for the overall study pool these will show an approximately 30% to 40% increase in the primary study outcomes (A&E attendances and emergency admissions).

We will generate forest plots depicting the outcomes from the ‘stacked’ data set, and also compare the results from fixed-effect vs random-effect models (where the latter will be more valid if the study samples contribute ‘true’ heterogeneity) as described above, but anticipate that this crude estimate will be less precise than the reweighted estimates described above.

Relatively poor balance from SMD plots or other exploratory diagnostics may suggest that a conventional approach to pooled analysis using the original combined data set is biased or invalid, for one or many of the results sets. Even if SMD plots and other diagnostics do not suggest imbalance, homogeneity of effects is a quite strong causal assumption that has been critiqued elsewhere.<sup>76</sup> The effect estimates from conventional meta-analysis across a small sample of n=3 studies would therefore be relatively weak evidence regardless of ultimate effect estimates, and suggest hypotheses for further validation.

For this reason, we strongly prefer effect estimates generated from entropy balancing and appropriate reconstruction of counterfactuals across the sample of interest, as they provide much stronger evidence for causal inference about the impact of MDTs.

## Patient and Public Involvement and Engagement (PPIE)

In 2019, the data analytics directorate at The Health Foundation, including the IAU, began to embed purpose-driven Patient and Public Involvement and Engagement (PPIE). PPIE is fundamental in our work to ensure that data is used in a legitimate and transparent way. We recognise that our assumptions around the needs of communities may not be accurate and that it is through working in partnership with the population that we improve health outcomes, make better use of resources, reduce inequalities and build public trust. The data analytics directorate’s PPIE vision of ‘involvement for better data’ recognises patients, carers and the public as stakeholders with rights and responsibilities for publicly-funded services, research and education. The vision is based on the principles of partnership, respect, inclusivity and transparency. We are working towards longer-term outcomes of improving research; supporting directorate strategy and planning to take account of what matters to the public; working to shape policy that improves health and care for communities; and improving how we involve and engage patients and the public in the longer term.

Before finalising this SAP, we engaged with a patient representative. Although several topics were touched upon, the discussion was predominately around the effect of social isolation and other factors relating to social context on health and health outcomes. This conversation helped inform the social context indicators described in the SAP.

Once we have preliminary results, we will be seeking input from stakeholders and a small group of (two to five) patient representatives, in order to understand the findings and explore whether there are any follow-on analyses required.

### Potential further work

Over the course of the pooled analysis, we may identify that specific methodological elements or interim analyses may be worthy of further investigation and reporting in their own right, including submission to peer reviewed journals. These analyses would be further developed after this primary work is carried out through March 2021, and may include:

- **National descriptive analysis** of social context factors and any association between them and hospital use
- **Validity and interpretation of missing data multiple imputation for CPM risk scores**, based on Fylde Coast correlation with observed SUS variables applied to NEHF as an exemplar for similar cases
- **Comparison of matching and weighting alternatives** in theoretical and empirical terms when applied to these MDT data sets, potentially with further simulated data and tutorial discussion
- **Further subgroup analysis using emergent machine learning tools to detect heterogeneous treatment effects**, based on replicable algorithms applied to patterns in the data rather than our pre-specified subgroups. This is particularly relevant in this case as the underlying theory and evidence for heterogeneity of MDT service impacts in different patient groups is relatively weak, with few quantifiable as opposed to qualitative trends. Methods derived from random forests or regression trees are particularly applicable here and offer principled approaches, such as generalised random (causal) forests<sup>77</sup> and Bayesian additive regression trees.<sup>78</sup>

## Limitations and sources of bias

We are exploring whether we can identify proxies for social isolation and other factors that may affect both the decision to refer individuals to an MDT and their emergency hospital use. The variables that we are investigating – living alone, recently living alone, living with somebody with dementia or frailty, frequent moves and recent bereavement – do not necessarily reflect an individual person's need or risk. However, we believe these factors signal meaningful social context at population level that may be very important for assessing the impact of MDTs on hospital use and therefore warrant investigation.

Although this pooled analysis will aim to improve the similarity between the compared groups and refine the original findings from the original evaluations by including social context indicators, there is still a risk that the intervention and control groups remain different

in important ways, for instance in comorbidities only recently diagnosed by GPs but not recorded in hospital or other risk factors, that cannot be observed by the IAU. It is not currently possible to quantify this risk.

During the conversation with the patient representative, some other potentially important risk factors that could affect a person's feeling of social isolation, health and mental health were raised, for example losing one's job and domestic abuse. Employment status is not recorded in hospital data and may not be well recorded in GP data. Although there are codes in hospital records for physical, sexual or psychological abuse, these are likely to be under-reported. Further work – outside of the scope of this project – could be to explore linking pseudonymised employment data and GP data to hospital data in order to further understand these factors' effect on MDT referral, health and health care service use.

The IAU only had access to secondary care data and was not able to evaluate the impact of the MDTs on other outcomes, such as quality of life, experience of care, staff satisfaction, improvement in working relationships or cost. The local evaluations in both NEHF and Fylde Coast included interviews or case studies with patients enrolled in the MDTs. A consistent finding across the qualitative work from all three MDTs was that the MDTs often provided support that was more wide-ranging than just medical care, including eg benefits advice, help with lifestyle changes and practical improvements within their homes.<sup>16,17</sup> These interventions may affect outcomes such as quality of life or experience of care, which we were not able to evaluate. Although these interventions may also affect hospital use in the long term, the effect may not be detectable within the time frames of these studies. There will, however, be a review of reviews of MDT interventions carried out in parallel with this work, which will aim to draw out the findings on the effect of MDTs on other outcome measures.

In this analysis, we explore the effect of MDTs on specific patient groups (eg individuals with dementia, chronic pulmonary disease or serious mental ill health, or nearing end of life), by selecting patients with these conditions from the group of patients enrolled in the ICT, ECS and EPC MDTs. However, it may be that MDTs that are designed to specifically support only specific conditions may have a different effect from that of MDTs that care for a broader range of conditions and needs, as is the case in the three MDTs evaluated here.

This study does not directly analyse the effect of other concurrent vanguard interventions such as rapid home response by a specially trained community paramedic or social prescribing, nor will this study be able to quantify how other, non-vanguard differences in the health and care systems in Fylde Coast and NEHF may have impacted outcomes.

The risk of study or measurement heterogeneity is relatively low, as the IAU previously used standardised data sets and nearly equivalent statistical analysis methods, including the choice of covariates and outcomes for all three programmes. We will standardise these procedures into a common data set for this pooled analysis, allowing for consistent estimation of the average treatment effect on the treated (ATT), which would measure the impact of an MDT programme on individuals enrolled in one of them. Nevertheless, there remains some natural heterogeneity in the patient populations, interventions, and comparative usual care between the three MDTs (see detail in Background section, Table 1). This study will partially deal with this heterogeneity, as the continued use of robust statistical techniques will rely on substantial overlap or 'exchangeability' (ie similarity irrespective of MDT participation status) of patients

in terms of their observed characteristics. The three MDTs were ultimately very similar in their design and implementation (see Background section, Table 2), perhaps to a greater extent than might be expected from the locally led programme design that was scoped under the new care models. The study samples appear to show substantial overlap or exchangeability (Table 3), and we will be able to assess this formally when applying reweighting and regression in the pooled analysis. It is notable that these studies generated quite similar estimates of the impact (ATT) of MDTs on hospital use (Table 4), suggesting there are generalisable effects regardless of the differences in settings, populations and programmes. Further post hoc analysis may be possible, including in other research, to explore and quantify the impact of heterogeneity in the intervention and comparative standard of care in the MDTs.

This project aims to understand why emergency hospital use was higher for patients enrolled in MDTs. As highlighted by the patient representative, it would be useful to ask enrolled patients this question. Both Fylde Coast and NEHF did qualitative interviews or case studies with referred patients as part of local vanguard evaluations, however these were done prior to the IAU evaluation results and do not specifically explore their use of and reasons for using emergency hospital services. A further qualitative study interviewing patients referred to MDTs about their care needs and their hospital use, although outside of the scope of this project, would therefore complement our work.

The results are nonetheless expected to enable learning that, together with other evidence, will help national and local commissioners to better understand MDTs, recognise the realistic impacts that can be expected from them, and identify potential areas for further investigation or improvement, which can for example help inform the implementation of MDTs within the new PCNs.

## Acknowledgements

We thank Lynn Laidlaw who provided a patient perspective, and Richard Grieve, Stephen O'Neill and Stefano Conti for reviewing and commenting on an earlier draft of this SAP.

# References

1. Roland M, Steventon A, et al. Case management for at-risk elderly patients in the English integrated care pilots: observational study of staff and patient experience and secondary care utilisation. *Int J Integr Care*. 2012; 12 (July).
2. Stokes J, Panagioti M, Alam R, Checkland K, Cheraghi-Sohi S and Bower P. Effectiveness of Case Management for 'At Risk' Patients in Primary Care: A Systematic Review and Meta-Analysis. *PLoS One*. 2015; 10(7).
3. Lloyd T, Brine R, Pearson R, Caunt M and Steventon A. *The impact of integrated care teams on hospital use in North East Hampshire and Farnham*. Health Foundation; 2018.
4. Baxter S, Johnson M, Chambers D, Sutton A, Goyder E and Booth A. The effects of integrated care: A systematic review of UK and international evidence. *BMC Health Serv Res*. 2018; 18(1).
5. Vestesson E, Shen J and Pariza P. *Technical Report: The impact of Extensive Care Service and Enhanced Primary Care in Fylde Coast: Findings from the Improvement Analytics Unit*. Health Foundation; 2020.
6. NHS England. *New Care Model: Vanguard - developing a blueprint for the future of NHS and care services*. NHS England; 2016.
7. Paris P, Lloyd T, Conti S. *Statistical analysis protocol for the evaluation of integrated care programmes in the North East Hampshire and Farnham and Fylde Coast CCGs*. Health Foundation; 2020 (available on request).
8. NHS England. *The NHS Long Term Plan*. NHS England; 2019.
9. Lloyd T, Brine R, Pearson R, Caunt M and Steventon A. *Technical appendix: The impact of integrated care teams on hospital use in North East Hampshire and Farnham*. Health Foundation; 2018.
10. Roland M and Abel G. Reducing emergency admissions: are we on the right track? *BMJ*. 2012; Sep 18; 345: e6017.
11. Wallace E, Smith SM, Fahey T and Roland M. Reducing emergency admissions through community based interventions. *BMJ*. 2016; 352(4).
12. Dreyer K, Steventon A, Fisher R and Deeny SR. The association between living alone and health care utilisation in older adults: a retrospective cohort study of electronic health records from a London general practice. *BMC Geriatrics*. 2018; 18: 15–17.
13. Lloyd T and Steventon A. *Statistical Analysis Protocol for an evaluation of integrated care teams as part of a primary and acute care system (PACS) in North East Hampshire and Farnham*. Health Foundation; 2017.
14. Clarke GM, Vestesson E and Pariza P. *Statistical Analysis Protocol for an evaluation of Extensive Care Service and Enhanced Primary Care for Fylde Coast NHS vanguard*. Health Foundation; 2018.
15. Wennberg D, Siegel M, Darin B, Billings J, Dixon J and Curry N et al. *Combined predictive model: final report and technical documentation*. Health Dialog, King's Fund and New York University; 2006.
16. Fylde Coast. *Fylde Coast Vanguard Local Evaluation Report*. 2018.
17. Wessex Academic Health Science Network. *Independent evaluation of the North East Hampshire and Farnham Vanguard: Farnham Integrated Care Team*. 2017.
18. NHS England. *Five year forward view*. NHS England; 2014.
19. Morciano M, Checkland K, Billings J, Coleman A, Stokes J, Tallack C, et al. New integrated care models in England associated with small reduction in hospital admissions in longer-term: A difference-in-differences analysis. *Health Policy*. 2020; 124(8) 826–833.
20. Lancashire County Council. *Lancashire Joint Strategic Needs Assessment (JSNA)*. 2014.



21. North East Hampshire & Farnham Clinical Commissioning Group. *Joint Strategic Needs Assessment (JSNA)*. 2013.
22. Diamond A and Sekhon JS. Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies. *Rev Econ Stat*. 2013; 95(3): 932–945.
23. Tibshirani R. Regression Shrinkage and Selection Via the Lasso. *J R Stat Soc Ser B*. 1996; 58(1) 267–288.
24. Posada D and Buckley TR. Model selection and model averaging in phylogenetics: Advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. *Syst Biol*. 2004; 53(5): 793–808.
25. Wang KY, Seed P, Schofield P, Ibrahim S, and Ashworth M. Which practices are high antibiotic prescribers? A cross-sectional analysis. *Br J Gen Pract*. 2009; 59(567).
26. Georghiou T and Keeble E. *Age UK's Personalised Integrated Care Programme: Evaluation of impact on hospital activity*. Nuffield Trust; 2019.
27. Kumpunen S, Edwards N, Georghiou T and Hughes G. Why do evaluations of integrated care not produce the results we expect? *Int J Care Coord*. 2020; 23(1).
28. Aiden H. *Isolation and loneliness: An overview of the literature*. British Red Cross; 2016.
29. Holt-Lunstad J, Smith TB, Baker M, Harris T and Stephenson D. Loneliness and Social Isolation as Risk Factors for Mortality: A Meta-Analytic Review. *Perspect Psychol Sci*. 2015; 10(2): 227–237.
30. Hawkley LC and Cacioppo JT. Loneliness matters: a theoretical and empirical review of consequences and mechanisms. *Ann Behav Med*. 2010; 40(2) 218–27.
31. Sanders S. *Families and households in the UK: 2019*. Office for National Statistics; 2019.
32. Stroebe M, Schut H and Stroebe W. Health outcomes of bereavement. *Lancet*. 2007; 370(9603): 1960–73.
33. Hernán MA and Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *Am J Epidemiol*. 2016; 183(8): 758–764.
34. Keele L, O'Neill S and Grieve R. Comparing the Performance of Statistical Adjustment Methods By Recovering the Experimental Benchmark from the REFLUX Trial. 2020 (<https://arxiv.org/pdf/2001.08170v1.pdf>)
35. Van Belle S, Wong G, Westhorp G, Pearson M, Emmel N, Manzano A, et al. Can 'realist' randomised controlled trials be genuinely realist? *Trials*. 2016; 17(1).
36. Lewis GH. 'Impactibility models': Identifying the subgroup of high-risk patients most amenable to hospital-avoidance programs. *Milbank Q*. 2010; 88(2): 240–255.
37. Smith SM, Soubhi H, Fortin M, Hudon C and O'Dowd T. Managing patients with multimorbidity: Systematic review of interventions in primary care and community settings. *BMJ*. 2012; 345(7874).
38. Sledge WH, Brown KE, Levine JM, Fiellin DA, Chawarski M, White WD, et al. A Randomized Trial of Primary Intensive Care to Reduce Hospital Admissions in Patients with High Utilization of Inpatient Services. *Dis Manag*. 2006; 9(6): 328–338.
39. Fletcher AE, Price GM, Ng ES, Stirling SL, Bulpitt CJ, Breeze E, et al. Population-based multidimensional assessment of older people in UK general practice: A cluster-randomised factorial trial. *Lancet*. 2004; 364(9446): 1667–1677.
40. Young J, Green J, Forster A, Small N, Lowson K, Bogle S, et al. Postacute care for older people in community hospitals: A multicenter randomized, controlled trial. *J Am Geriatr Soc*. 2007; 55(12): 1995–2002.
41. Holland R, Battersby J, Harvey I, Lenaghan E, Smith J and Hay L. Systematic review of multidisciplinary interventions in heart failure. *Heart*. 2005; 91(7): 899–906.

42. McAlister FA, Stewart S, Ferrua S and McMurray JJJV. Multidisciplinary strategies for the management of heart failure patients at high risk for admission: A systematic review of randomized trials. *J Am Coll Cardiol.* 2004; 44(4): 810–819.
43. Langhorne P, Baylan S and Early Supported Discharge Trialists. Early supported discharge services for people with acute stroke (Cochrane review) [with consumer summary]. *Cochrane Database Syst Rev.* 2017; no.7.
44. Tyrer P, Evans K, Gandhi N, Lamont A, Harrison-Read P and Johnson T. Randomised controlled trial of two models of care for discharged psychiatric patients. *Br Med J.* 1998; 316(7125): 106–109.
45. Malone D, Marriott S, Newton-Howes G, Simmonds S and Tyrer P. Community mental health teams for people with severe mental illnesses and disordered personality. *Cochrane Database Syst Rev.* 2007; no.3.
46. Steventon A and Billings J. Preventing hospital readmissions: the importance of considering 'impactibility' not just predicted risk. *BMJ Qual Saf.* 2017; 12(June).
47. Roland M. Continuity of care: betrayed values or misplaced nostalgia. *Int J Integr Care.* 2012; 12: e200.
48. Wallace E, Stuart E, Vaughan N, Bennett K, Fahey T and Smith SM. Risk Prediction models to predict emergency hospital admission in community-dwelling adults: A systematic review. *Med Care.* 2014; 52(8): 751–765.
49. NHS England. *New care models Risk stratification: Learning and Impact Study.* NHS England Operational Research and Evaluation Unit; 2017
50. Phillip L. Change management in integrated care: what helps and hinders middle managers – a case study. *J Integr Care.* 2012; 20(4): 246–256.
51. Clarke GM, Pariza P and Wolters A. *The long-term impacts of new care models on hospital use: An evaluation of the Integrated Care Transformation Programme in Mid-Nottinghamshire.* Health Foundation; 2020.
52. Keeble E, Bardsley M, Durand MA, Hoomans T and Mays N. Area level impacts on emergency hospital admissions of the integrated care and support pioneer programme in England: Difference-in-differences analysis. *BMJ Open.* 2019; 9(8).
53. Fries J. The Compression of Morbidity The Finite Life Span. *Milbank Q.* 2005; 83(3): 801–23.
54. NHS England et al. *Delivering the Forward View: NHS planning guidance 2016/17 – 2020/21.* NHS England; 2015.
55. Steventon A, Deeny S, Friebel R, Gardner T and Thorlby R. *Briefing: Emergency hospital admissions in England: which may be avoidable and how?* Health Foundation; 2018.
56. Georghiou T and Bardsley M. *Exploring the cost of care at the end of life.* Nuffield Trust; 2014.
57. Elixhauser A, Steiner C, Harris DR and Coffey RM. Comorbidity measures for use with administrative data. *Med Care.* 1998; 36(1): 8–27.
58. Quan H, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care.* 2005; 43(11): 1130–1139.
59. Soong J, et al. Quantifying the prevalence of frailty in English hospitals. *BMJ Open.* 2015; 5(10).
60. Charlson ME, Pompei P, Ales KL and MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *J Chronic Dis.* 1987; 40(5): 373–383.
61. Dorning H, Davies A and Blunt I. *Quality Watch - Focus on: People with mental ill health and hospital use.* Nuffield Trust; 2015.
62. Hainmueller J. Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Polit Anal.* 2012; 20(1): 25–46.

63. Parry W, Wolters AT, Brine RJ and Steventon A. Effect of an integrated care pathway on use of primary and secondary healthcare by patients at high risk of emergency inpatient admission: a matched control cohort study in Tower Hamlets. *BMJ Open*. 2019; 9(6).
64. Parish WJ, Keyes V, Beadles C and Kandilov A. Using entropy balancing to strengthen an observational cohort study design: lessons learned from an evaluation of a complex multi-state federal demonstration. *Heal Serv Outcomes Res Methodol*. 2018; 18(1): 17–46.
65. Everding J and Marcus J. The effect of unemployment on the smoking behavior of couples. *Heal Econ*. 2020; 29,(2): 154–170.
66. Quan H, Li B, Couris C, Fushimi K, Graham P, Hider P, et al. Updating and validating the Charlson comorbidity index and score for risk adjustment in hospital discharge abstracts using data from 6 countries. *Am J Epidemiol*. 2011; 173(6): 676–682.
67. Soong J, Poots AJ, Scott S, Donald K and Bell D. Developing and validating a risk prediction model for acute care based on frailty syndromes. *BMJ Open*. 2015; 5: e008457.
68. Billings J, Georghiou T, Blunt I and Bardsley M. Choosing a model to predict hospital admission: an observational study of new variants of predictive models for case finding. *BMJ Open*. 2013; 3(8).
69. Billings J, Blunt I, Steventon A, Georghiou T, Lewis G and Bardsley M. Development of a predictive model to identify inpatients at risk of re-admission within 30 days of discharge (PARR-30). *BMJ Open*. 2012; 2(4).
70. Lemke KW, Weiner JP and Clark JM. Development and validation of a model for predicting hand prehensile movements. *Med Care*. 2012; 50(2): 131–139.
71. The Johns Hopkins ACG® System.
72. Austin PC and Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Stat Med*. 2015; April.
73. Higgins JPT and Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med*. 2002; 21(11): 1539–1558.
74. Rücker G, Schwarzer G, Carpenter JR and Schumacher M. Undue reliance on I<sup>2</sup> in assessing heterogeneity may mislead. *BMC Med Res Methodol*. 2008; 8: 2–9.
75. Austin PC and Stuart EA. The performance of inverse probability of treatment weighting and full matching on the propensity score in the presence of model misspecification when estimating the effect of treatment on survival outcomes. *Stat Methods Med Res*. 2017; 26(4): 1654–1670.
76. Dahabreh IJ, Petito LC, Robertson SE, Hernán MA and Steingrimsson JA. Toward Causally Interpretable Meta-analysis: Transporting Inferences from Multiple Randomized Trials to a New Target Population. *Epidemiology*. 2020; 31(3): 334–344.
77. Wager S and Athey S. Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. *J Am Stat Assoc*. 2018; 113(523): 1228–1242.
78. Hill J, Linero A and Murray J. Bayesian additive regression trees: A review and look forward. *Annu Rev Stat Its Appl*. 2020; 7: 251–278.