



STUDY PROTOCOL

Unlocking data to inform public health policy and practice: Exploring barriers and creating solutions for public health intelligence using integrated datasets across Kent, Sussex and Surrey (KSS)

Short Title	Unlocking Data
Acronym	None
Version number	Version 1.0, 8 th July 2021
Funder	NIHR Public Health Research Programme
Funders Reference	NIHR133761
Sponsors Number	ER/BSMS2730/7
Chief Investigators	Dr Elizabeth Ford; Dr Melanie Rees-Roberts
Host Institution	Brighton and Sussex Medical School, University of Sussex
Start date	1st May 2021
Project duration	9 months

VERSION CONTROL TABLE

Version Number	Date	Changes made
0.1	19 th April 2021	First Draft
0.2	29 th April 2021	Changes made to proposed workshop agenda and removal of Project Oversight Committee.
0.3	30 th April 2021	
0.4	10 th June 2021	Update and changes to public focus group methods following Ethics Committee review.
1.0	8 th July 2021	Changes above accepted by NIHR. Signed by CI.

FUNDING AND SUPPORT IN KIND

FUNDER(S)	FINANCIAL AND NON-FINANCIAL SUPPORT GIVEN
National Institute of Health Research University of Southampton, Alpha House, Enterprise Road, Southampton, SO16 7NS Telephone: 0238 059 7501 E-Mail: netspostawardsetup@nihr.ac.uk	Financial support

SIGNATURE PAGE

The undersigned confirm that the following protocol has been agreed and accepted and that the Chief Investigator agrees to conduct the study in compliance with the approved protocol and will adhere to the principles outlined in the Declaration of Helsinki, the Sponsor's SOPs, and other regulatory requirement.

I agree to ensure that the confidential information contained in this document will not be used for any other purpose other than the evaluation or conduct of the investigation without the prior written consent of the Sponsor

I also confirm that I will make the findings of the study publicly available through publication or other dissemination tools without any unnecessary delay and that an honest accurate and transparent account of the study will be given; and that any discrepancies from the study as planned in this protocol will be explained.

For and on behalf of the Study Sponsor:			
			
Chief Investigator:			
Dr Elizabeth Ford	Chief Investigator	Date:	8 th July 2021

TABLE OF CONTENTS

Contents

STUDY PROTOCOL.....	1
VERSION CONTROL TABLE	2
FUNDING AND SUPPORT IN KIND	2
SIGNATURE PAGE	3
TABLE OF CONTENTS	4
KEY STUDY CONTACTS	5
SCIENTIFIC ABSTRACT	7
PLAIN ENGLISH SUMMARY	8
ROLE OF STUDY SPONSOR AND FUNDER	10
ROLES AND RESPONSIBILITIES OF THE STUDY MANAGEMENT GROUPS	10
ETHICAL APPROVALS	10
DATA PROTECTION AND SHARING	11
GLOSSARY OF ABBREVIATIONS.....	12
STUDY PROTOCOL.....	13
Background and Rationale	13
RESEARCH QUESTIONS	15
RESEARCH PLAN AND METHODS.....	15
Pre-grant and preparatory activity:.....	15
Strand 1: Foster networks and partnership activities between academic research and public health practice with a focus on supporting training and capacity.....	16
Strand 2: Exploring how to maximise the use of data to gain a greater cross-sector understanding of health needs and health inequalities and facilitate cross-sector working to improve population health and reduce health inequalities.....	17
Strand 3: Understanding public priorities and perceptions to inform publicly supported governance structures for ICS datasets.....	19
PATIENT AND PUBLIC INVOLVEMENT	22
DISSEMINATION AND OUTPUTS	24
PROJECT TIMETABLE	25
REFERENCES	27

KEY STUDY CONTACTS

Chief Investigator	Dr Elizabeth Ford	Department of Primary Care and Public Health Brighton and Sussex Medical School Watson Building, Village Way, Falmer Brighton BN1 9PH e.m.ford@bsms.ac.uk
Joint lead applicant	Dr Melanie Rees-Roberts	Centre for Health Services Studies, University of Kent, Canterbury, Kent, CT2 7NF M.Rees-Roberts@kent.ac.uk
Research Assistant	Ms Kathryn Stanley	Department of Primary Care and Public Health Brighton and Sussex Medical School Watson Building, Village Way, Falmer Brighton BN1 9PH k.stanley@bsms.ac.uk
Research Administrator	Ms Katie Goddard	Department of Primary Care and Public Health Brighton and Sussex Medical School Watson Building, Village Way, Falmer Brighton BN1 9PH k.goddard@bsms.ac.uk
PPI Coordinator	Ms Kathryn Stanley	As above
Sponsor's Representative	Ms Tanya Telling, University of Sussex	Head of Joint Clinical Research Office JCRO@sussex.ac.uk . Tt34@sussex.ac.uk
Funder	NIHR	University of Southampton, Alpha House, Enterprise Road, Southampton, SO16 7NS Telephone: 0238 059 7501 E-Mail: netspostawardsetup@nihr.ac.uk
Co-applicants	Ms Victoria Spencer-Hughes Dr Abraham George Dr Jo Armes Prof Jackie Cassell Prof Anotida Madzvamuse Prof Chris Farmer Ms Sarah Giles	East Sussex County Council Kent County Council University of Surrey Brighton and Sussex Medical School University of Sussex University of Kent Lay co-applicant
Collaborators	Dr Gurprit Pannu Katherine Church Kate Gilchrist Melanie Dumelow Jane Lodge	Sussex Health and Care Partnership Surrey Heartlands Health and Care Partnership Brighton and Hove City Council East Sussex County Council Sussex NHS Commissioners

	Pippa Shaw	ARC KSS
--	------------	---------

SCIENTIFIC ABSTRACT

The triangulation of population level health needs, data, and health service use is done by public health teams in local authorities (LAs). In the region of Kent, Surrey and Sussex (KSS), LAs, together with NHS partners, have launched programmes to create integrated datasets of citizens' NHS, social care and administrative records. While the infrastructure for hosting and linking data is proceeding, to unlock the potential for these datasets to improve health and quality of life for the local population, several barriers must be tackled. Currently, analytics capacity and capability in KSS public health teams is not at a level to deal with the complexity of the integrated data, and due to the pandemic, LAs have not had the capacity to develop those skills through training or recruitment. Second, legal and technical solutions for linkage of data to enable cross sector working (e.g. between health and housing or education) are needed. Third, public involvement and engagement in various aspects of the use of routinely collected data to support decision making in local authorities and across health and care systems has been largely neglected.

In this project we will tackle each of these three barriers. We will create and foster a partnership network between academic researchers, data-intensive scientists, and public health practitioners with a focus on sharing expertise in the technical linkage and analysis of multi-dimensional integrated datasets, kickstarted by two data-sprint events. Through this network, and by funding two 6-month analyst fellowships, we will support knowledge exchange, training and capacity in the public health teams, and tackle important public health case studies identified by our local authority partners ('learning by doing').

In Sussex, the fellow will explore how to use integrated data to identify patterns of service use or clusters of factors which are robust proxies for being at high risk of decline in health status, multimorbidity, or increase in care needs. This would inform better targeted early intervention, and improvement in provision of more holistic and less duplicated care across services.

In Kent, the fellow will explore technical solutions within current legal frameworks for linking non-NHS, non-social care data, such as from housing, education and employment, with health and care datasets, to explore of socio-economic and wider determinants of health. These approaches will maximise the use of different sources of data to facilitate cross-sector working across service providers, to gain a greater understanding of health needs and health inequalities, particularly focussing on the local needs of promoting early & preventive service use. In addition, they will reduce inefficient provision or duplication for people using multiple services, both of which are key public health and LA focus.

Finally, we will conduct a series of deliberative focus groups to understand public perception of the use of routinely collected data to support decision making in local authorities and ensure public involvement is embedded in dataset governance structures.

We will disseminate results from projects through bespoke stakeholder and public-facing events, reports, journal articles, and regional and national conferences. This work will enable a step-change in productivity in terms of analysing citizens' data for local and regional public health intelligence to improve health for the local population.

PLAIN ENGLISH SUMMARY

Aim of the research:

This project aims to support and develop skills in using data for planning public health services by:

- Improving data analysis skills in Kent, Sussex and Surrey (KSS) public health teams
- Solving current problems around linking existing data
- Understanding how we can involve the public in deciding how different health and administrative data can be linked and stored safely, and who has access to them for improving services and research.

Background:

Local authorities (LAs) in England are responsible for improving the health of their local population. This work is led by public health (PH) teams.

In KSS, the LAs want to improve services for communities. They also want to find ways to improve health. They want to do this by joining up (or linking) data from health, social care, housing and education services. Linked datasets have been started, but LAs have some barriers which stop them using the data fully.

Firstly, PH teams do not have analysts with advanced data analysis skills to answer complex data questions. Secondly, there are difficulties in linking different types of non-health data (housing, education) with health and social care data (social services, GP records). Thirdly, local Integrated Care Systems (ICS), who plan and fund local health and care services and manage datasets, do not currently know how the public feel about their data being linked together and used to improve services and public health.

Better ways of involving the public, that are informed by public views, need to be put in place.

Design and methods used:

- We will bring together data scientists at universities with members of PH teams, database teams, and public representatives. This new group will gather to discuss challenges, barriers and create solutions to using available data in the best way to improve the public's health.
- We will train two PH analysts to solve problems of local public needs through the use of existing data. They will be trained by university data scientists. The analyst in Kent will focus on the use of different sources of data to understand what things affect the public's health e.g. the types of environment people live in. The analyst in Sussex will look at using available data to spot earlier opportunities to keep people well and independent for longer.
- We will hold public discussion groups to find out what people think about the use of health and other data to improve services. This will advise the ICS on how to be clear and trustworthy about data use. It will also suggest ways the public can be included in decision-making around using data.

Patient and public involvement

The project was designed with and reviewed by multiple members of the public. A member of the public will be a study team member and will contribute to decision-making and support the running of the public focused work. We will consult with the public throughout the project and we will include the public in our working community. The public will be asked to review documents and help us interpret results from our discussion groups.

Dissemination

The results of this project will be made available online:

- Project report
- Publication in a scientific journal
- Lay summary.

We will share our findings through events for the public, local researchers, service providers and analysts. The working community will continue beyond the project and will continue the development of analytical and research skills to help the public stay well and prevent ill health.

ROLE OF STUDY SPONSOR AND FUNDER

University of Sussex is the sponsor for this research and will assume overall responsibility for the initiation and management of the study. University of Sussex as research Sponsor indemnifies its staff, research participants and research protocols with public liability insurance. These policies include provision for indemnity in the event of a successful litigious claim for proven non-negligent harm.

The programme is funded by the NIHR Public Health Research Programme (PHR). The NIHR PHR will monitor progress and be informed of all changes to the protocol. The NIHR PHR will be sent all outputs at least 28 days before publication/dissemination. All published outputs will acknowledge funding and include the following disclaimer:

‘This project is funded by the National Institute for Health Research (NIHR) Public Health Research programme (NIHR133761). The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.’

All other decisions about the study design, conduct, data analysis and interpretation, manuscript writing and dissemination of results will be made by the Chief Investigator and study management group (see below) and will not be within the responsibility of the sponsor or funder.

ROLES AND RESPONSIBILITIES OF THE STUDY MANAGEMENT GROUPS

The trial is funded by the National Institute for Health Research (NIHR) Public Health Research Programme (NIHR133761). University of Sussex will be the project sponsor and the host organisation, with Dr Ford the Chief Investigator. Subcontracts will be put in place between University of Sussex and other partner organisations, detailing the budget resources allocated, the responsibilities and the expected contributions of each party.

The core Project Management Group will comprise the project leads (Ford and Rees-Roberts), research assistant (Stanley), lay co-applicant (Giles) co-applicants (Madzvamuse, Spender-Hughes) and the analyst fellows (TBC). The core project management group will meet twice a month to ensure smooth running of the project. Specialist input from other members of the research team will be available as and when needed.

The full research team including senior members (Cassell, Armes, Farmer, George) will meet every 6 weeks to ensure milestones are achieved, oversee progress, trouble shoot if problems arise, plan the next stage and agree timelines. This steering group will include all co-applicants and partners will be invited.

The Chief Investigator has overall responsibility for the study and will oversee all study management. The data custodian will be the Chief Investigator.

ETHICAL APPROVALS

Ethical approval will be sought for the Strand 3 Public Discussion Group work from Brighton and Sussex Medical School Research Governance and Ethics Committee. Documents were submitted for assessment on 21st April 2021 and a response will be received after 12th May 2021 when the committee are meeting.

No participant recruitment will occur until ethical approvals have been granted.

DATA PROTECTION AND SHARING

Strand 1 & 2 Analysis using Sussex and Kent Integrated Datasets

Kent: All contributors to the Kent Integrated Dataset (KERNEL) are joint data controllers, and the University of Kent has also signed up as a joint data controller, meaning access to the Kent dataset is assured for all Kent parties in the collaboration.

Sussex: A data-sharing agreement will be drawn up between the University of Sussex and the joint data controllers of the Sussex Integrated Dataset (SID) as part of the collaboration agreement of this award. This will allow access to parts of the SID for the University of Sussex-hosted data-sprints.

Data will be protected according to guidance from the dataset information governance leads and all parties using data will be trained and informed about secure access and protection of data according to agreed protocols.

Strand 3: Public Discussion Groups

Identifiable participant information will be collected on the consent form and the participant demographics survey. There may also be identifiers in the recorded audio files from the discussion groups. These data will all be in electronic format. All of these data will be stored securely on password-protected university storage (not on personal laptops). They will be deleted at the end of the study when the demographic data has been transferred to an anonymised database and patients have been sent results summaries.

Audio files will be transcribed verbatim but identifiers such as names, addresses, hospitals or other healthcare provider names will be redacted and replaced with placeholders such as [GP surgery]. The transcriptions will not contain any identifiers. Transcripts will be labelled only with focus group numbers and stored on password-protected network storage. These will be uploaded into NVivo software for analysis. NVivo files will also not contain any identifiers and will be stored on password protected networked storage. Recordings will be deleted after transcription is completed and transcripts checked for quality.

Participants will not be identified in the analysis, write up or dissemination of the study findings. When demographics of the participants are described in project write up, characteristics will be aggregated so that no individual can be identified from a unique combination of characteristics.

Data from the discussion groups which needs to be shared between team members in different organisations will be uploaded to the approved, secure, collaboration platform "Glasscubes".

GLOSSARY OF ABBREVIATIONS

A&E – Accident and Emergency Department

ARC-KSS – NIHR Applied Research Collaboration in Kent Surrey and Sussex

BHCC – Brighton and Hove City Council

BSMS – Brighton and Sussex Medical School

CCGs – Clinical Commissioning Groups

DISCUS – Data Intensive Science Centre at the University of Sussex

KSS – Kent Surrey and Sussex

ESCC – East Sussex County Council

GP – General Practice or General Practitioner.

HR – Human Resources

ICS – Integrated Care Systems

KCC – Kent County Council

KERNEL - Kent Research Network for Education and Learning

KID – Kent Integrated Dataset

LA – Local Authority

ML – Machine Learning

NHS – National Health Service

NIHR – National Institute of Health Research

PH – Public Health

PHR – Public Health Research Programme

PPI – Patient and Public Involvement

QOF – Quality and Outcomes Framework

RA – Research Assistant

RQ – Research Question

SCC – Surrey County Council

SID – Sussex Integrated Dataset

SSC – Study Steering Committee (Project Oversight Committee)

UoK – University of Kent

WSCC – West Sussex County Council

STUDY PROTOCOL

Background and Rationale

The Kent, Sussex and Surrey (KSS) Context

National funding for regional datasets has enabled the development of infrastructure to take delivery of, and link data from local authorities (LAs), general practitioners (GPs), acute, community and mental health NHS Trusts, adult social care, and other health and care settings. In the Kent, Sussex and Surrey region, this funding has enabled data to be curated in this way for the first time, as previously no such data infrastructure existed. Public health teams in KSS LAs plan to make use of linked data to inform care and better understand the health and care needs of the local population, in particular identifying individuals with multiple needs. Furthermore, to understand existing and emerging health inequalities and inform public health interventions that benefit diverse local populations, the use of linked data across services needs to be contextualised with data on the wider determinants and socio-economic factors of health that impact outcomes at a population level [1].

LAs recognize that information governance is the critical enabler for better data and analytics. Substantial progress has been made in each county, onboarding data from primary care networks, acute and mental health trusts and local authorities, creating pseudonymisation mechanisms, and linking and storing data in local data warehouses. However, barriers continue to exist in using these data to full potential. This is due to a lack of analytical and intelligence capacity plus under-developed research skills to tackle datasets' full complexity for the improvement of health and care services and population health.

Two LAs in the region, East Sussex County Council (ESCC) and Kent County Council (KCC) will work with academics at the Universities of Sussex, Kent, and Surrey to achieve the following:

- 1) Create and foster a partnership network between academic research, local authorities and public health practice across the Kent, Surrey and Sussex region with a focus on sharing existing expertise in the creation and use of extremely large integrated data.
- 2) Through this network, support knowledge exchange, training and capacity to build public health data science research using existing datasets to inform strategic planning and detailed needs assessments in the health and social care system.
- 3) Explore how to maximise the use of different sources of data to facilitate cross-sector working across service providers to gain a greater understanding of health needs and health inequalities through use of integrated data, particularly focussing on the local needs of targeted prevention and service improvements for public health in the context of multi-morbidities and health inequalities, and understanding risk factors.
- 4) Understand public priorities and perceptions of the use of routinely collected data to support decision making in local authorities and evaluate how public involvement can be embedded in dataset governance structures in local Integrated Care Systems (ICS).

Further collaborative partners include: Brighton and Hove City Council (BHCC), West Sussex County Council (WSCC), Medway Council, Surrey County Council (SCC), and Surrey Heartlands Health and Care Partnership (see letters of support).

Current barriers this project will address

- 1) Analytics capability and capacity in the LAs is falling behind.

ESCC states that the biggest gap in their public health team portfolio presently is a lack of analysts with programming skills and experience. Team members are adept at utilising

routine tools such as Excel, SPSS and Tableau, but are behind similar teams in terms of coding/programming, to make use of multi-dimensional, linked, longitudinal data. Their experience is mainly with single condition or aggregated local statistics. ESCC contribute to the Sussex Health and Care Partnership in which the Sussex Integrated Dataset (SID) is being developed. The public health team at ESCC want to use the linked health and social care data in the SID to track individuals' patterns of service use, understand the distribution of long terms conditions within the population, build a complex description of co-morbidities clustering and accumulation over time, and learn methods to present multidimensional longitudinal data, available in the SID for the first time, in a way which is easily understood by decision makers.

A recent NIHR-funded study in Kent (PHR NIHR131931) highlighted that despite limited current research activity in the councils, there was a strong appetite to do more. LAs recognized the potential value of data intensive research, but did not have funding or mechanisms for using data for full benefit or to inform practice [2].

2) Linkage of data and access models to facilitate cross-sector working across local authorities and other providers are currently unclear

KCC has identified technical, linkage and governance needs which are barriers to making full use of local integrated datasets. Although Kent and Medway was an early implementer site for linked datasets for population health analytics since 2013, having developed the Kent Integrated Dataset (KID) pilot [3], changes in governance, funding priorities, and organisation restructures have proved challenging. Kent's local health system has not yet focused resources on governance and data infrastructure to enable wider use of the data for service planning and research.

KCC are currently working within the Kent & Medway Shared Health and Care Analytics Board, with whom work commenced in early 2020 to develop a new linked dataset known as the Kent Research Network for Education and Learning (KERNEL). KCC have a strategic interest in developing methods for linking LA data into the KERNEL which are not health or social care related, such as housing, education and employment. Local joint strategic needs assessments have identified disparities in outcomes across populations, linked to deprivation. Use of wider cross-sector data would allow better understanding of the social determinants of health to address these existing inequalities.

In Kent, all NHS organisations including the CCGs have signed up to Joint Data Controller Arrangements, moving towards a consistent set of processes and procedures for linked data access and data linkage system requests. However, it is not clear how to set up processes for linkage and regulation of non-NHS, non-social care data, especially as there is no consistent approach between the NHS and wider public sector for a unique patient/citizen identifier, and national legislation impedes the regular use of NHS identifiers outside the NHS.

3) Public perception of the use of routinely collected data to support decision making across integrated care systems are unknown

There is currently very little local evidence of public awareness as well as little evidence testing "reasonable expectations" of patients and public around population data linkage (health and non-health) for analytics. For example, for the Sussex Integrated Dataset (SID), shared data controller responsibilities are still being developed and the route to ask questions or propose analysis of the data is not clear. A publicly-supported and informed governance and data access structure would ensure the questions asked of the data are appropriate, ethical, and can be prioritized against public benefit and need. In Kent and Medway, local research suggests that members of the public are enthusiastic about public health research and being involved in this work, to ensure research reflects community priorities. However, further work is needed, in a capacity challenged environment, to facilitate better public contributions in public health delivery and research. Across the region,

the current lack of public awareness of datasets is an obstacle to developing publicly supported governance models for data access and use, which also enable confidence across service providers contributing data. This becomes increasingly important as data may become available to a wider range of stakeholders such as academic partners.

While use of health and social care data for service evaluation and improvement is a statutory obligation for the LAs, public engagement around this task is still valuable and important [4]. According to social licence theory, the public expect that organisations who, for example, hold and use the public's data, will go beyond the requirements of formal regulation, and adhere to voluntary codes of trustworthy behaviour and transparency [5]. Where the public are satisfied that the motivations of the organisation are trustworthy, they may confer a "social licence" to operate. Previous work by our group has indicated that the public consider core research ethics principles [6] when weighing up their approval of data-sharing and linkage schemes [7-9]:

- 1) Do the methods of data collection, linkage, storage and usage **respect individual patient autonomy**? (Respect for Autonomy)
- 2) Are the objectives and the intended outputs primarily concerned with contributing to the **public good**? (Beneficence)
- 3) Is any agreement between the local authorities, NHS, and organisations offering or requesting analytics (private or public) **fair and just**? (Justice)
- 4) Could granting access to the data, or granting a particular use of the data, lead to **individual or collective harm**? (Non-maleficence)

Beyond this, we note two key issues: The public must trust service provider and commissioning or research organisations' **competence** with data handling, and must trust organisations' **motivations** for data analysis. The public are often willing to share their data for the benefit of a **community** they belong to. Focusing public engagement activities on community building with a diverse range of stakeholders is a way of creating a greater sense of ownership of the research by the public, and also enables data holders to better meet the needs of citizens [9].

Work is urgently needed to engage with the citizens of Kent, Sussex and Surrey, to understand public views around integrated dataset use, inform ways of including the public in governance models, explore routes to public involvement in prioritisation and decision-making for data use, and advise best practice for transparency and engagement with the public for data oversight and use.

RESEARCH QUESTIONS

- 1) What capacity and analytic training is needed within LA PH teams to make use of linked datasets to inform strategic planning?
- 2) How can we expand local authority capacity and analytics training through partnerships with academic researchers?
- 3) How can we harness this new analytics capacity to facilitate Integrated Care Systems improvement of population health at system, place and locality?
- 4) How can public engagement and involvement be integrated in the Kent, Sussex and Surrey integrated dataset governance models?

RESEARCH PLAN AND METHODS

Pre-grant and preparatory activity:

- 1) We have conducted an evidence review and synthesis of governance structures of datasets across the UK, with a focus on how public involvement and engagement has been embedded throughout, to inform local set up of datasets in the region.
- 2) We will write a protocol, co-design study materials, and apply for ethical approval for the public engagement aspects of this work, prior to grant start.

Strand 1: Foster networks and partnership activities between academic research and public health practice with a focus on supporting training and capacity

Background and Aims

We will build a **strategic partnership** between KCC, ESCC, BHCC, WSCC, SCC, Medway Council, Surrey Heartlands, Sussex Health and Care Partnership, Kent & Medway Shared Health and Care Analytics Board, Brighton and Sussex Medical School (BSMS), the Data Intensive Science Centre at the University of Sussex (DISCUS) and the Universities of Kent and Surrey. The partnership will be initiated within this grant and supported in the longer term by the NIHR Applied Research Collaboration Kent, Surrey and Sussex (ARC KSS) through hosting of partnership meetings every 6 months. The aim of the partnership will be to share research projects, exemplars of data use for service delivery improvement, ongoing barriers, skills, expertise and create collaborative opportunities to inform policy and practice. These events will be open to all leaders, strategic decision-makers, providers and researchers and public members to involve all relevant expertise across the region.

Strategy

1) Data Sprints

DISCUS will provide in kind support by funding and hosting **two “data-sprints”**. These are days which bring together a range of DISCUS analysts from students through to professors, alongside LA public health staff, NHS analysts, decision-makers and public representatives. Using non-identifiable citizens’ data, teams will work together to kickstart solution-finding to technical or analytics problems brought in by the two LAs.

Data Sprint Agenda:

These events will start with a presentation on the data source, with some examples, so that attendees can become familiar with the data involved. Real anonymized or synthetic data extracts will be used to aid understanding.

Following this, there will be a presentation from the PH team about what they want to achieve – outlining the problem to be solved, and a presentation from a commissioner or service lead so that analysts understand how services are organized before interpreting the data. The data-sprint will then involve discussion and brainstorming involving the following:

- the variables, and way these variables are derived, that we think are important to enable the analyses planned.
- planning an analysis pipeline (defining and creating variables, using linked data, analysis approaches).

Where possible a data extract, with a simple research question, can be used to plan out an initial approach for analysis, which could be generalized to wider or more complex questions. This means we can create coding and an analysis pipeline in the data-sprint which can be shared with the PH team and fellow for kick-starting the fellowship. This process will also enable documentation of the skills mix and timescales needed to make any approach part of routine LA use of linked datasets.

2) 6-month analytics fellowships in public health teams

One fellowship will be hosted in ESCC and one in KCC. These fellows will be secondments within UoK and ESCC who already have access agreements in place for the datasets, and will be closely supervised by Ford, Madzvamuse, Rees-Roberts and Farmer, with access to all DISCUS resources, in order to support them in developing additional analytics and programming skills. The focus of these fellowships and are described below.

Strand 2: Exploring how to maximise the use of data to gain a greater cross-sector understanding of health needs and health inequalities and facilitate cross-sector working to improve population health and reduce health inequalities

For each fellowship, a period of exploration will be required at the beginning. This will be kick-started by the 2 data sprints designed to explore and brainstorm potential solutions to the data problem/questions as a starting point. The fellowships and data-sprints therefore feed into each other. Activities at the data-sprint which feed directly into the fellowship include:

- a. Evaluating what data are available to answer the data problem.
- b. Descriptive work exploring the data and other considerations needed for analysis e.g. what data sources are available for each patient, how many datasets they appear in, the quality of data from each source (GP, mental health, hospital visits/stays, other outpatient appointments e.g. physiotherapy, and adult social care).
- c. Identifying how we can make the project as *efficient* as possible. This will depend on the programming approach which will be explored in the data-sprints and supported for the duration of the fellowships by existing regional and research team expertise.
- d. Understanding *a priori* definitions of the concepts we wish to find and quantify in the data. Definitions will differ depending on the source of data, coding structure, etc.
- e. Defining the clinical concepts at the core of the research questions and drawing up lists of how they might be represented in the different sources of data. E.g. Some outcomes/exposures, such as multiple service use, high/low number of appointments will be defined by combinations of pieces of information over set time periods.

The analyst fellows will then proceed with working on the project supported by the research team. This will involve the following steps:

- a. Cleaning and working with the data to construct a meaningful dataset for analysis e.g. write logic-based code to create concept definitions, categorising patients and creating variables which indicate patterns of population features or service use such as number of appointments, number of services attended. Some variables may be created using other methods e.g. unsupervised clustering methods.
- b. Interrogating data completeness and data quality, prevalence and incidence estimates, comparing against other sources of data, which will inform us what we should expect from our datasets.

- c. Converting the data into an analysable data-frame potentially with one row per patient and categorical data about all of that patient's conditions, or possibly keeping it in long form with multiple rows per patient.
- d. Use this data-frame in statistical or machine learning models to answer the data problem. E.g. for Sussex fellow - look for associations between hypothesised exposures/risk factors and outcomes (see below).

Fellowship 1: East Sussex – Led by Spencer-Hughes, Ford and Madzvamuse

Public health aim: ESCC want to reduce the gap in healthy life expectancy between the least and most deprived populations by delaying age of onset of multiple long-term conditions. To do this they need to better understand the pattern of onset of morbidities, the relationship with service use, service provision, and socio-demographic data in order to improve targeted interventions.

Barrier: The data that is currently accessible and used by ESCC public health team derives from primary care Quality and Outcomes Framework (QOF) reports of prevalence of single conditions, and/or is aggregated data, where it is impossible to find numbers and characteristics of individuals with 2 or more long term conditions. ESCC has contributed to the development of the SID to enable linkage of multiple sources of patient level health and care data, which would enable data analysis. Currently, Covid-19 pandemic response means the PH team at ESCC does not have capacity to develop suitable analytics skills or data presentation skills to produce intelligence from the data, and the SID access/governance model is currently immature.

Our solution: An analyst from ESCC, BHCC or WSCC will be seconded to the ESCC PH team for 6 months, closely supervised through a partnership between DISCUS analysts and public health to build analytics skills and supply novel solutions. Access to the SID data is assured, as ESCC is a data controller. The project will specifically:

- 1) Identify efficient methods for evaluating numbers of data sources for each individual within an extremely large and complex dataset.
- 2) Evaluate data quality and completeness of diagnostic information for long term conditions.
- 3) Develop early methods for describing multiple service use, and making predictions about it (e.g. number of appointments, number of services used). Methods will comprise statistical and machine learning (ML) data analytical techniques. These will include supervised and ML techniques such as classification, clustering, regression and neural networks
- 4) Explore application of longitudinal modelling to identify individuals developing 2 or more long term conditions at different ages and their socio-demographic risk factors, and their service use.
- 5) Develop advanced presentation methods to communicate intelligence from these models to decision makers in an easy to understand way.
- 6) With the results of (4) and (5) draw up recommendations for areas where engagement with services could be improved.
- 7) Identify the infrastructure, competencies and skills mix required across an intelligence / analyst team to maximise use of linked datasets within local authorities

Impact of work: Identifying the number of local people with multiple long-term conditions and the intersection of this with service use, locality, deprivation and other socio-demographic risk factors would enable the local authority to identify key target groups and to take appropriate actions to improve health and wellbeing. Development of a competencies framework for utilising complex large datasets would aid embedding of work within local authorities as well as collaboration.

Fellowship 2: Kent – Led by Rees-Roberts, Farmer and George.

Public health aim: KCC want to explore the use of non-NHS provider data to inform public health interventions by better understanding the impact of socio-economic factors and wider determinants of health. In the longer term, this will help to inform public health services and needs assessment in order to address some of Kent and Medway's significant health problems linked to some of the lowest life expectancy and significant health inequalities [10].

Barriers: The data that will inform these wider determinants of health are available for integration, however, there are technical barriers to linkage. Firstly, there is a lack of clarity around how to link non-NHS, non-social care data, due to the lack of unique patient / citizen identifier and no consistent approach between NHS and wider public sector. Second, there is not enough local evidence of public awareness; 'reasonable expectations' of patients and public around population data linkage (health and non-health) for analytics have not been explored.

Our solution: An analyst-fellow will work at the University of Kent, closely supervised with partners from East Kent Hospitals University NHS Foundation Trust (utilising data linkage expertise in place), DISCUS and KCC to explore solutions for linking NHS and non-NHS data that might enable elucidation of socio-economic and wider determinants of health [11]. The project will specifically:

- 1) examine data available across the current NHS and non-NHS providers within the KERNEL dataset.
- 2) identify data fields and methods that might enable linkage of data at different levels e.g. individual, ecological (group), geographical level or other methods.
- 3) evaluate data quality and completeness of information that can be linked between NHS and non-NHS data,
- 4) identify socio-demographic and wider determinant data points available for interrogation with NHS data
- 5) explore potential methods for using linked data to inform health or other public outcomes e.g. links between police data and A&E attendance of victims to identify domestic abuse.
- 6) With the results of (5) draw up recommendations for areas where socio-economic and wider determinants of health can be explored in relations to public health services and outcomes.

Impact of work: This work will identify if local linkage of NHS and non-NHS data is feasible and explore potential solutions that enable the data to be used to interrogate socio-economic or wider determinants of health even partially. This would then pave the way for future work focussing on addressing health inequalities and wider determinants of health in Kent and Medway.

Strand 3: Understanding public priorities and perceptions to inform publicly supported governance structures for ICS datasets

We plan to hold public and stakeholder deliberative discussion focus groups to gather views on issues of governance, to tackle the barriers outlined above:

- 1) To understand public views around dataset use,
- 2) To inform ways of including the public in governance models regionally,
- 3) To explore public involvement in prioritisation and decision making for the use of data,
- 4) To advise on best practice for transparency and engagement with the public around data oversight and use.

All Health and Care Partnerships and Councils will be involved in this work-strand. A series of four deliberative discussion focus groups will be conducted with citizens of KSS,

according to the methodology outlined by Rothwell et al [12]. One further focus group will be conducted with data stakeholders and system leaders across the region.

Justification of methodology

Because so little is publicly known about the regional use of routinely collected health and administrative data for service planning and evaluation, participants will need to be informed and supported to understand the different issues before deliberating on their opinions and views. Deliberative research is an approach for gathering wider views for health issues when there are many complex issues to weigh against each other [8]. Deliberative discussion focus groups include a range of informative presentations, interspersed with facilitated plenary and breakout discussions. Information presented should be balanced, comprehensive and incorporate the pros and cons of the topic of interest. The resulting discussion must be inclusive of all group members.

Recruitment and Procedure

Regional leads for public and patient involvement and engagement (From ARC KSS, ESCC, Sussex CCG, Surrey Heartlands) will use their networks to recruit members of the public from KSS. A purposefully diverse sample will be recruited using a range of community organisations, with whom we already have links, aiming for inclusive representation. We will create a strategy with our PPI advisors to recruit seldom heard populations. Regional leads for public and patient involvement from ARC KSS, local authorities, and integrated care systems will advertise the study and invite their contacts to express an interest in participation through emails and social media, and through posters and in person meetings only where possible with current restrictions. The ARC KSS covers all three counties (Kent, Surrey and Sussex). It also has links with Healthwatch which has a diverse membership. We will also use our contacts with local community groups in Sussex through the SPICES project who will circulate study information in the same way. Furthermore, we will work with Dr Priya Paudyal who through ARC funded research mapped all the Ethnic Minority Community Groups in KSS. She will offer introductions to community leaders who we will approach.

A flyer or advertisement will be circulated via the above networks giving a link to a URL which will give brief information about the study, and ask for an expression of interest and some brief demographic, availability and contact details. A phone number and email address will be listed for questions and informal enquiries. This step is to manage any over-subscription to the focus groups. If we can offer a place to the participant, we will send a link which will show the full participant information sheet and the opportunity to consent to taking part in the study and give additional demographic information. When participants have completed these online forms, they will be contacted to be invited to a discussion group. If we have too many expressions of interest we will screen participants on age, sex, ethnicity, location (Kent, Surrey or Sussex) and views on data-sharing, and aim for maximum diversity within the groups. Other participants will be added to a waiting list.

Focus groups will be conducted remotely using video-conferencing software and will be recorded within Zoom. We will use a version of Zoom which allows subtitles and chat for those who are hard of hearing or those who prefer to send written comments than speak aloud. The 3 hour agenda has been split into two sessions (1h10 and 1h50) which will be held a week apart at the same time and day. Each focus group will have up to 15 participants with a total sample of up to 60. Groups will be facilitated by a research assistant trained in online focus group facilitation, with up to 4 breakout room facilitators, and will feature a number of guest speakers from the KSS region. Some written information and links to videos will be sent in advance of the focus group as preparatory material, to allow more time for discussion.

Participants will be offered a £50 voucher for their participation.

Addressing Digital Exclusion

We recognize the risk of digital exclusion. We will tackle this in four ways:

- 1) We will hold discussion groups at different times of the day and in the evening. The maximum attendance time will be 3.00 hours, I split into 2 shorter sessions per group (1h10 and 1h50) so that participants do not need to find such a large chunk of free time to participate.
- 2) If an individual is not proficient in English they will be invited to attend with another member of household to help with English. Due to Covid restrictions and funding constraints, we cannot offer any other translation services.
- 3) On the flyer and invitation email we will provide a university landline number that can be called instead of signing up online. This landline will be automatically diverted to the research assistant's project phone (as she is working from home). Her personal number will not be given out. Individuals with no or limited internet access will be sent study documentation by post. They will be able to call into the Zoom group from a phone at the cost of landline call.
- 4) For those who wish to use Zoom but who are inexperienced in the software, we will send them a simple step-by-step Zoom user guide which we have previously prepared for a research project providing remote health coaching over Zoom in low income communities. We will offer a pre-group Zoom check in session (15 minutes) for anyone who wants to check their use of the software before the focus group.

Workshops planned content

The outline agenda for the discussion groups is given in Table 1, this has been finalised after consultation with public advisors and speakers. For the dataset stakeholder focus group, a similar outline will be used but we will remove presentation elements that the audience will already know (such as what an ICS is) and condense the Q&A sessions, so the overall session will be shorter.

Analysis

Audio files from the discussion group will be transcribed by an external company following signed non-disclosure agreement. De-identified transcripts will be analysed by thematic analysis, with an inductive approach informed by the evidence synthesis and the teams' prior research and the needs and aims of the dataset teams. We will follow the 6 steps in Braun and Clark's 2006 Thematic Analysis paper for the transcripts of data.

Demographic data and quantitative data collected in polls will be analysed manually in MS Excel or SPSS and reported descriptively (number of responses in each category and percentages).

Table 1: Proposed Discussion Group Agenda.

Time	Activity Description	Person (s) responsible
Pre-discussion group	Complete PIS, consent form, and demographics sheet on line	All participants. To be emailed out by KS
Pre-discussion group	Videos and written information on: 1. What are integrated care systems? 2. Why integrated care systems need data – what they use it for 3. What integrated datasets (IDs) are and what they could be used for.	. Content contributed by LA and ICS partners. Videos taken from open sources. Content and design feedback by public advisors.

	4. How is patient privacy protected when NHS data is reused for service planning and research? 5. What types of public or population health problems analysing data in integrated datasets could be addressed? 6. What is data linkage and how and why does it happen? Including with non-health datasets. Resource pack of technical terms and visuals to be provided online and on paper.	
	On Day 1:	
	1. Welcome, housekeeping, introductions (10 mins) 2. Discussion of Pre-Group Information (10 mins) 3. Presentation 1: What integrated datasets could be used for in public health (10 mins) 4. Any questions (5 mins) 5. Break Out room discussions (25 mins) 6. Summary discussion session (10 mins) Total Time: 1 hour 10 mins	
	On Day 2:	
	1. Welcome, housekeeping, introductions (10 mins) 2. Presentation 2: Governance issues for patient data (10 mins) 3. Any questions (5 mins) 4. Break Out room discussions (25 mins) 5. Coffee Break (10 mins) 6. Presentation 3: How the public are involved in datasets (10 mins) 7. Any questions (5 mins) 8. Break Out room discussions (25 mins) 9. Summary session (10 mins) Total time 1h50m	

PATIENT AND PUBLIC INVOLVEMENT

Firstly, we will consult with a public group of advisors on the protocol and study materials for the focus groups. This will be conducted using the Public and Community Involvement and Engagement (PCIE) group through ARC KSS, and through the Sussex NHS Commissioners public involvement lead. This will happen prior to project start (i.e. in January – March 2021) to inform the study ethics application which will be made in this period. This consultation will happen remotely using Zoom software.

Secondly, we have a lay member of the study team (GILES) who will:

- Contribute to fortnightly core group meetings as an equal member.

- Review the protocol and participant information documents during preparation for ethical approval and when any changes are made
- Input into the focus group structure, discussion questions and information presentations for participants and stakeholders
- Will be invited to facilitate a breakout room
- Offer ongoing review of the project data, findings and progress throughout.
- Contribute to dissemination and ideas of how to disseminate project findings
- Help to ensure a variety of Patient and Public Involvement is included in the project.

The project team will provide appropriate training to ensure the lay co-applicant can be fully involved in the project. Training needs will be discussed with each lay co-applicant at the start of the study. A training budget of £220 has been included.

The focus of Strand 3 is to consult with the public about their views on a variety of issues with the integrated datasets. Their views will be key to producing recommendations to make sure public involvement is embedded in governance structures for these datasets and that local authorities achieve transparency and accountability for the use of these data. All participant views will be transcribed and analysed carefully to ensure balance is achieved within the recommendations at the end of the project.

We will have a dissemination event specifically for members of the public to discuss our findings and recommendations. We will advertise the event through all available networks to ensure a large attendance. We will create a public facing information leaflet as a key output of the study.

DISSEMINATION AND OUTPUTS

Our dissemination and output plans are:

- 1) 5000-word project report and 20 slide presentation for NIHR (due Dec 2021).
- 2) Regional dissemination events for (a) cross-region data stakeholders and system leaders, and (b) members of the public (online or face-to-face depending on university policy) to showcase results from fellowships and public workshops and generate discussion [to report on strands 1-3].
- 3) Three academic papers to be submitted to International Journal of Population Data Science or similar; one reporting the public workshop outcomes combined with pre-grant governance review. One paper reporting each fellowship activity and results [one for each fellowship (strand 2) and one for strand 3].
- 4) Conference presentations at regional ARC events and other national dataset infrastructure events such as the OpenDataSavesLives initiative, and at national conferences such as Administrative Data Network conference or similar [on fellowships and strand 3 work]

Our research will produce the following:

- A sustainable community of practice partnership between experts in the field across system leaders, service providers and academics in the region to support capacity development of digital and analytics skills for public health [from strand 1].
- A skills framework for PH analysts to inform recruitment and personal development plans [from strands 1&2].
- Solutions to technical barriers to enable use of data by public health teams to inform improvement of population health and wellbeing, which will be shared for regional and national learning [from strands 1&2].
- A thematic analysis of KSS regional public views on the most acceptable and trustworthy governance structures for local integrated datasets, which may be generalisable to other regions [from strand 3].
- A set of implementable recommendations for embedding public involvement in the governance models of the local datasets, to present for discussion to KSS dataset programme boards [from strand 3].
- Public facing materials to communicate the existence and purpose of the KSS integrated datasets and set out how the public can be involved in prioritisation and decision-making [from strand 3].

PROJECT TIMETABLE

	Public Engagement	LA-based fellowships	Networking, dissemination, and project management
Pre-project	Protocol development w/ ethics application for focus groups Recruitment and contracting of RA	HR/contracts for fellows	Governance review. Prepare for first partnership event, identifying speakers and setting programme.
May	RA (with ARC KSS) recruitment and set up for focus groups	HR/contracts for fellows	Networking, partnership and agenda setting meeting for all partners. Planning of data-sprints
June	RA conducts stakeholder focus group	Fellows and LA attend data-sprints; handover period.	1 st data-sprint run by DISCUS with ESCC
July	RA conducts 3 public focus groups (transcription ongoing; initial analysis)	1 st July Fellows start. Evaluation of data quality for key issues in case study and linkage issues/service use parameters	2 nd data-sprint run by DISCUS with KCC.
August	RA conducts 3 public focus groups (transcription ongoing; initial analysis)	Descriptive analysis of service use and multi-morbidity/data sources and points for linkage	3-month progress report to NIHR and SSC; SSC meeting.
September	Transcription and data analysis	Modelling, linkage or prediction techniques tried and evaluated	
October	Data analysis	Data presentation techniques tried and evaluated	
November	Report written; full team discussion event; recommendations made; dissemination events organised.	Results finalised, report writing	6-month progress report to NIHR and SSC; SSC meeting. Second partnership meeting
December	Design and production of public facing materials. Stakeholder and public dissemination event.	30 th Dec fellows finish. Reports delivered to project team. Results presented at dissemination event.	Stakeholder and public dissemination event

January	Reports / papers finalised and submitted.		Reports / papers finalised and submitted.
---------	---	--	---

REFERENCES

1. Marmot M. Health equity in England: the Marmot review 10 years on: Institute of Health Equity. ; 2020 [Available from: <http://www.instituteofhealthequity.org/the-marmot-review-10-years-on>].
2. Forbes L, Bates A, Rees-Roberts M, Hotham S, Kendall S. Developing a Public Health Research System to Support Local Government in Kent. Kent and Medway NIHR; 2020.
3. Lewer D, Bourne T, George A, Abi-Aad G, Taylor C, George J. Data Resource: the Kent Integrated Dataset (KID). *International Journal of Population Data Science*. 2018;3(1).
4. Aitken M, Tully MP, Porteous C, Denegri S, Cunningham-Burley S, Banner N, et al. Consensus statement on public involvement and engagement with data-intensive health research. *International Journal of Population Data Science*. 2019;4(1).
5. Gunningham N, Kagan RA, Thornton D. Social License and Environmental Protection: Why Businesses Go Beyond Compliance. *Law & Social Inquiry*. 2004;29(2):307-41.
6. Beauchamp TL, Childress JF. *Principles of Biomedical Ethics*. 6th ed. New York: Oxford University Press; 2009.
7. Stockdale J, Cassell J, Ford E. "Giving something back": A systematic review and ethical enquiry into public views on the use of patient data for research in the United Kingdom and the Republic of Ireland. *Wellcome Open Res*. 2019;3:6-.
8. Ford E, Oswald M, Hassan L, Bozentko K, Nenadic G, Cassell J. Should free-text data in electronic medical records be shared for research? A citizens' jury study in the UK. *Journal of Medical Ethics*. 2020;46(6):367-77.
9. Ford E, Boyd A, Bowles JKF, Havard A, Aldridge RW, Curcin V, et al. Our data, our society, our health: A vision for inclusive and transparent health data science in the United Kingdom and beyond. *Learning Health Systems*. 2019;3(3):e10191.
10. Kent Public Health Observatory. Joint strategic needs assessment 2021 [Available from: <https://www.kpho.org.uk/joint-strategic-needs-assessment>].
11. Moscrop A, Ziebland S, Bloch G, Iraola JR. If social determinants of health are so important, shouldn't we ask patients about them? *BMJ*. 2020;371:m4150.
12. Rothwell E, Anderson R, Botkin JR. Deliberative Discussion Focus Groups. *Qual Health Res*. 2016;26(6):734-40.
13. Ford E, Nicholson A, Koeling R, Tate AR, Carroll JA, Axelrod L, et al. Optimising the use of electronic health records to estimate the incidence of rheumatoid arthritis in primary care: What information is hidden in free text? *BMC Med Res Methodol*. 2013;13:105.
14. Ford E, Carroll J, Smith H, Davies K, Koeling R, Petersen I, et al. What evidence is there for a delay in diagnostic coding of RA in UK general practice records? An observational study of free text. *BMJ Open*. 2016;6(6):e010393.
15. Ford E, Rooney P, Oliver S, Hoile R, Hurley P, Banerjee S, et al. Identifying undetected dementia in UK primary care patients: a retrospective case-control study comparing machine-learning and standard epidemiological approaches. *BMC Med Inform Decis Mak*. 2019;19(1):248.
16. Ford E, Starlinger J, Rooney P, Oliver S, Banerjee S, van Marwijk H, et al. Could dementia be detected from UK primary care patients' records by simple automated methods earlier than by the treating physician? A retrospective case-control study. *Wellcome Open Res*. 2020;5(120):120.
17. Jones KH, Ford EM, Lea N, Griffiths LJ, Hassan L, Heys S, et al. Toward the Development of Data Governance Standards for Using Clinical Free-Text Data in Health Research: Position Paper. *J Med Internet Res*. 2020;22(6):e16760.
18. Campillo-Funollet E, Van Yperen J, Allman P, Bell M, Beresford W, Clay J, et al. Forecasting COVID-19: Using SEIR-D quantitative modelling for healthcare demand and capacity. *International Journal of Epidemiology*. 2021;(under revision).