Data Resource: the Kent Integrated Dataset (KID)

Lewer, D[1], Bourne, T[2], George, A[2*], Abi-Aad, G[2], Taylor, C[2], and George, J[1]

Abstract

Introduction
Electronic healthcare records from the UK are accessible to researchers via several platforms, but these platforms typically include data from a limited subset of health and care services. The Kent Integrated Dataset (KID) provides insight into system-wide health and care utilisation for the whole population of Kent and Medway.

Processes
The KID uses pseudonymisation-at-source to link patient-level records from services including general practices, hospitals, community health services and social care. Data is refreshed monthly and processes to monitor data quality have been developed.

Data contents
For each episode of care, the KID includes date of the episode, the type of service accessed, the cost of the episode and clinical information such as the health condition being treated and results of diagnostic tests. The dataset also includes contextual information such as the neighbourhood deprivation.

Conclusions
The KID is a unique and rich dataset available to researchers who are investigating a broad range of public health questions. It provides system-level insight into patient journeys and care utilisation and supports commissioning based on patient needs.

Background

Research using electronic health records in the United Kingdom

The United Kingdom has a long history of research based on electronic healthcare records (EHRs). Large samples of primary care data are available from several databases. The Clinical Practice Research Datalink, formerly the General Practice Research Database, now includes records of over eleven million patients, with linkage to hospital records, cancer registries, social deprivation information and cause-specific mortality (1). Over 1000 studies have been published based on this resource, including classical risk factor epidemiology, health services research and randomised controlled trials that use EHRs to measure outcomes. Similarly, The Health Improvement Network (THIN) (2) and QResearch (3,4) databases capture records from subsets of general practices.

Following the examples of Welsh and Scottish whole-population EHR research centres (5,6), some local areas in England are seeking to create data platforms that cover their whole population and incorporate data from a broad set of public services. This approach to EHR platforms offers new opportunities to (a) research social patterns in health and healthcare use, (b) map patients' pathways across multiple organisations, (c) evaluate the impact of interventions and changes to services across multiple organisations, and (d) understand the efficiency and effectiveness of a whole care system. While they are used to support healthcare planning, such data platforms provide a largely untapped resource for research. The Kent Integrated Dataset (KID) is a relatively mature exemplar of such a population-level linked dataset, covering almost two million residents in South East England. For researchers, the KID makes two important additions to existing EHR resources in England: it includes data from a wider range of health and care services and it covers the entire local population.

Initiation of the project

The purpose of the KID is to provide planners in Kent and Medway with insight into population health and system-level use of services. It aims to integrate data from health and social care providers and therefore allow analysis of the 'patient

---

*Corresponding Author:
 Email Address: Abraham.George@kent.gov.uk (A George)

journey' or 'citizen journey'. This is a shift from the traditional approach of collecting and analysing data at the level of organisations. In addition to providing information about utilisation of health and social care services, the KID includes information about individuals' socioeconomic and environmental contexts, allowing insight into the wider determinants of service use and health.

Start-up funding was provided through NHS England's 'National Long Term Conditions Year of Care' programme from 2013-2016. Two objectives of this funding were (a) to carry out detailed analyses of the prevalence of multi-morbidity and associated costs in health and social care, and (b) to provide a dataset for planning and evaluating an integrated care service in Kent. After this programme ended, Kent County Council and local National Health Service (NHS) commissioners agreed to continue developing the KID to support the long-term approach of 'place-based commissioning' (7).

About Kent and Medway

The KID covers Kent and Medway, an area of South-East England with a population of approximately 1.8 million in mid-2015 (8). It includes urban and rural areas (figure 1, map). The population is varied, with areas in the west generally more affluent than the east. Deprivation is concentrated in urban areas and particularly in coastal towns. Life expectancy ranges from 73 years in the most deprived area to 90 in the least deprived area (9). Like most areas of the UK, the average age of the population is increasing.

The KID draws on data from public services in Kent and Medway (see 'data contents', below). Most healthcare in England is provided through the taxpayer-funded NHS, which is free at the point of use. Individuals register with local family doctors, who are the gatekeepers to secondary and tertiary care. The majority of the English population use publicly funded healthcare; estimates suggest that only 3% of primary care (10) and 9% of acute hospital care (11) is paid for by patients (for example through insurance or direct payments). Local government is responsible for social care, which includes providing funding for personal care in residential care homes and clients' own homes.

Processes

Source and linkage of datasets

The KID comprises individual-level linked EHRs from the following services located in Kent and Medway: primary care providers (including general practices, out-of-hours providers and walk-in centres), community health providers, mental health services, acute hospitals (including accident and emergency, inpatient and outpatient episodes), public health services, adult social care and palliative care hospices. The dataset includes records of interactions between residents of Kent and Medway and these services.

Each service provider/data owner has securely uploaded data monthly since April 2014. The data is processed by the KID team and is available for use within three months of being originally recorded.

Across the NHS and many social care providers, individuals are given a unique identifier in the form of a 10 digit 'NHS number'. An encrypted version of this identifier is used to link individuals across the constituent datasets. Names are excluded and other potentially identifiable information is coarsened to prevent re-identification of individuals. For example, dates of birth are replaced by single-year-of-age and postcodes are replaced by Lower Super Output Areas (a geographical area covering approximately 1500 residents).

Data validation and quality checks

Data owners are responsible for validating and checking the quality of data before it is fed into the KID. These processes have been developed for purposes such as invoicing commissioners, and are carried out by the data owners' analysts. After each monthly upload to the KID, the data owners check that the correct total number of records is registered in the KID. The KID team then runs five checks on each 'service function' (primary care, social care, hospitals, etc.) to monitor data quality; these checks are summarised in Table 1.

Governance

The principles in the design of the KID's governance have been that (a) the organisations that contribute data should participate in the development of the dataset and (b) the uses of the KID should benefit service planning and improve the health and wellbeing of residents and patients in Kent and Medway.

The KID is overseen by a steering group that includes representatives of Kent County Council and local health commissioners. Sub-groups consider issues such as information governance, development of the dataset and considering applications for use of the data. Kent County Council public health team provide day-to-day administration and project management. Patients can opt-out of contributing to data to the KID by informing their GP surgery that they do not want their data to be shared with external organisations.

Data contents

Variables

All organisations submitting data include the following information about each episode of care: the date of the episode, the type of service accessed, the cost of the episode/interaction and clinical information such as the health condition being treated. Each dataset also includes further fields, specific to the type of care delivered. The dataset includes many variables and a full set is available on request. Table 2 shows selected datasets that feed into the KID and example variables included.
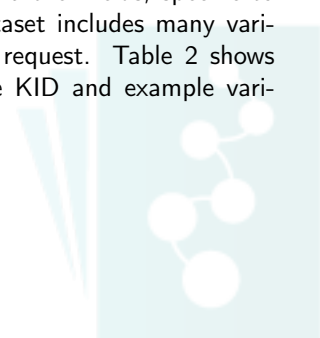
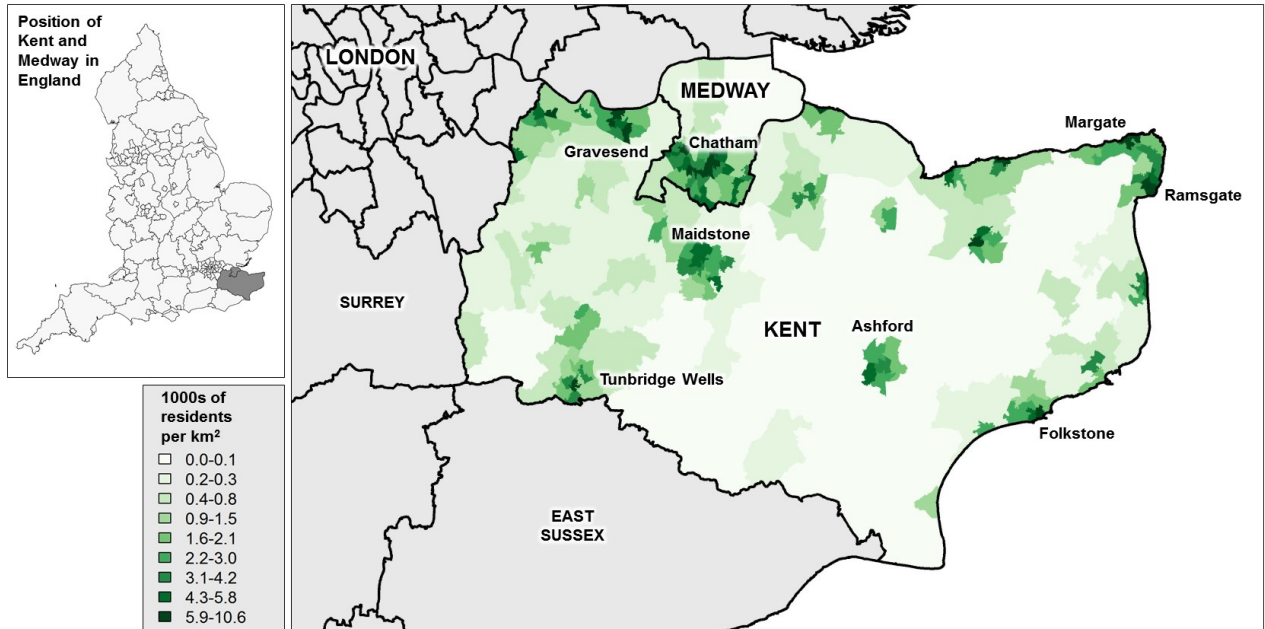Figure 1: Geographical location and population density of Kent and Medway

Table 1: Monthly data quality checks for the Kent Integrated Dataset

Data receipt	Data quantity	Data accuracy and completeness
1. All participating organisations have provided data for each of their service areas*	2. Data received is stable over time (i.e. within expected stochastic tolerances for month-to-month changes) 3. Data volumes are comparable with external reference data, such as published numbers of hospital admissions, accident & emergency attendances and GP consultations.	4. All data items include key variables (such as those listed in the 'data contents' section) 5. Coding of events (such as Read Codes) appears consistent across data providers

* 'Service areas' refer to services provided by participating organisations. For example, hospitals contribute data for a wide range of inpatient, outpatient and emergency services.

Table 2: Selected datasets included in the Kent Integrated Dataset

Dataset	Source system	Linkage	Example variables
Primary care (general practices)	EMIS and Vision clinical computing systems	NHS number	- The site of consultation (surgery, home or telephone) - Clinical information; including diagnoses, diagnostic tests and results, vaccinations, prescriptions, referrals, records of risk factors (such as alcohol, tobacco and blood pressure) - Staff type ('doctor' or 'other') - GP practice code (can be linked to practice characteristics)
Social care (Kent County Council)	SWIFT	NHS number (available for 94% of cases)	- Care package type - Client category type (e.g. 'Adult Learning Disability')
Secondary Care (Acute NHS Trusts located in Kent)	Secondary Uses Service	NHS number	- Type of episode (emergency, admitted care, outpatient care) - Length of admission - Discharge location - Causes of admission (using ICD-10 diagnoses) - Type of procedure (such as operation) received, if relevant

Dataset	Source system	Linkage	Example variables
Community health services (Kent Community Health NHS Foundation Trust)	CIS	NHS number	• Type of service provided • Description of treatment received • Referral source
Mental health (Kent and Medway NHS and Social Care Partnership Trust)	Servelec RiO	NHS number	• Type of service provided • A classification of mental health need, based on severity and type of mental health problems(12)
Out of hours (IC24)	CLEO	NHS number	• Location of consultation (phone, walk-in centre, home visit) • Outcome (e.g. GP referral, hospital episode)
Population register	Derived from other datasets	NHS number	• Age • Sex • LSOA of residence • Index of Multiple Deprivation decile(13) of LSOA of residence • Commercial geodemographic segmentation at household level: Mosaic and ACORN • Pseudonymised Unique Property Reference Number, allowing members of the same household to be linked

The primary care data is one of the richest sources of clinical information in the KID. It includes a wide range of events such as diagnoses, referral letters, prescriptions, and requests and results for diagnostic tests. The data is encoded using the Read Code System (which will be replaced by SNOMED codes in 2018), which is a taxonomy of clinical terms used to record patient findings and procedures in primary care IT systems.(14) Consultations in primary care often have multiple Read Codes, which can be linked via a unique consultation identifier. The data do not include free text consultation notes.

All episodes of care include an estimate of the cost of the episode. The methodology for estimating costs differs between datasets and typically relates to the type of service provided. The costs of primary care interactions are taken from Personal Social Services Research Unit 'Unit Costs' (15), a compendium of estimated unit costs in health and social care based on data such as salary scales, consultation length and typical overhead costs. The appropriate unit cost is selected using the location (telephone, surgery or home visit) and the type of healthcare professional delivering the service. Costs of secondary care services are taken from national NHS tariffs (which dictate the amount paid to NHS hospitals by NHS commissioners for each episode of care). Methodologies behind the costs in each dataset are available on request.

Quantity and completeness of data

The KID is a dynamic dataset and the steering group regularly considers new sources of linked data. As of December 2017, 221/238 (93%) primary care providers in Kent and Medway have agreed to submit data. Table 3 shows the rate of service use recorded in the KID.

There are known and quantified data gaps in the KID, including data for individuals who have declined to share their information outside of their GP surgery (2.3% of patients at the time of publication) and some 'sensitive' data in primary and secondary care datasets, including data relating to sexual health, suicides and children's social care. In addition, the KID excludes hospital care that is not funded through national NHS tariffs, such as privately funded care (though care provided in independent providers and funded by the NHS is included), and records of Kent and Medway residents' interactions with health and care providers that are located outside of Kent and Medway.

Example uses

This section includes three examples of analyses that have been undertaken using the KID.

Economic analysis of frailty

International guidelines recommend routine identification of frailty to provide evidence-based treatment (16), but many available tools require primary data collection from patients. An electronic frailty index (eFI) has been developed by Clegg et al. (17), based on linked electronic health records, allowing healthcare professionals to draw on routinely available data to generate a frailty score. The process uses primary care data

to count 'deficits', including symptoms, conditions and disabilities, in people aged 65 and over. This generates a frailty score, which can be subdivided by severity. This method has been used in the KID and extended to include costs of care, allowing an economic comparison between frail patients and patients of the same age who are not frail.

Estimating prevalence of rare conditions

Disease registers for the 20 most common long-term conditions are routinely maintained in primary care, incentivised through a national scheme known as the Quality Outcomes Framework. The KID is being used to estimate prevalence of less common conditions such as acute macular degeneration and autism spectrum disorders, supporting decisions about funding of specialist treatment.

Comparing risk of non-elective hospitalisation by general practice

The KID has been used to measure the risk of non-elective (i.e. unplanned) hospitalisation among patients of general practices in Kent and Medway. Practices were grouped according to their age structure and deprivation to allow for valid comparisons. A relatively high risk of hospitalisation compared to peer practices may suggest a need to review community care for people with long-term conditions.

Strengths and limitations

The strengths of KID as a research resource lie in its population coverage, service coverage, variety of variables, timeliness of data availability and the use of a unique reference number for linkage. First, it includes a complete list of patients registered with GP surgeries in Kent and Medway, providing whole population coverage. While some groups are less likely to be registered with a GP, such as young adult men and migrants, previous research has indicated that 99% of the UK population is registered with a GP (18). Second, the KID covers more services than many available EHR research platforms, with community health, mental health and social care providers typically not included in existing EHR research platforms. Third, it includes many variables that allow for new studies of aetiology and health care services. In particular, all datasets include the cost of the episode, allowing for economic modelling. Fourth, the data platform is updated regularly. Data is updated monthly and is available for research within three months, providing planners and researchers with opportunities for rapid evaluation of service changes. Finally, the unique reference number used across all datasets allows individual patients to be tracked across services and primary care practices, providing insight into the paths that patients take across the health and social care system. This also leads to high-quality linkage with low risk of errors.

The limitations relate to data quality, the exclusion of mortality data and generalisability of the data. First, data quality is variable and differs across participating organisations. An understanding of the sources is required to design research appropriately. For example, GP Read Codes provide a large amount of information about consultations, but should be used

Table 3: Rate of service use recorded in the Kent Integrated Dataset, by age group, 2015-2016

| | Patient age group | | | | |
	0-15	16-34	35-59	60+	All ages
Population (mid-2015) (8)	347,950	417,076	594,434	441,751	1,801,221
Rate of service use per 1000 persons					
Primary care consultations*	2,482	2,846	3,704	6,618	3,973
Adult social care interactions	n/a	98	130	538	197
Hospital admissions	152	149	193	514	254
A&E attendances	342	339	251	363	316
Community health contacts	564	263	345	2,310	850
Adult mental health services contacts	3	399	392	365	312
Out-of-hours consultations	121	80	54	114	88

*The rate of primary care consultations is based on a subset of primary care practices that supplied data in from April 2015-March 2016.

with care because they are not always recorded consistently and the way they are used may change over time (19). Similarly, in the UK, only around half of attendances at a hospital emergency department have a valid diagnostic code (20), partly because the service is self-referral and many patients are not considered unwell. Second, the KID is not linked to the UK's official mortality records. A significant proportion of deaths can be identified from the constituent datasets, including patients who die in hospital (42% of deaths in Kent and 48% of deaths in Medway in 2015(21)) and those whose death is recorded on general practice clinical systems. However, the timeliness of these data is not currently known and the data may not include the date or cause of death. Finally, researchers should bear in mind that the service utilisation recorded in the KID may differ from populations in other regions and countries.

Data access

Licensed access to the KID for research purposes is available on condition that the research is likely to provide some benefit to the Kent and Medway health and care economy. Researchers should contact Dr Abraham George, Consultant in Public Health and lead for the KID at Kent County Council, who can advise on whether the research objectives fit the allowed purposes of the KID and how to make an application (see corresponding author for contact details). Currently, individual-level data can only be viewed and analysed on Kent County Council's computer systems, with access provided physically at Kent County Council or via a secure remote desktop.

Conclusion

The KID is extremely rich in terms of the services that contribute data and the variables that are available. It provides opportunities for new analyses of patient journeys across different health and care providers and new epidemiological insight into the wider determinants of health. To date, the data has been mainly used to support healthcare planning and is rela-

tively untapped for research purposes. The quality and depth of the data varies and an understanding of the data sources and structured terminologies (such as ICD-10 codes in hospital data and Read codes in primary care) is required to design research appropriately. With support from local partners, linked datasets such as the KID can provide powerful support for joined-up planning of services and new research opportunities.

Statement on conflicts of interest

The authors declare that they do not have any conflicts of interest.

Funding

References

1. Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, Staa T van, et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). International Journal of Epidemiology. 2015;44(3):827–36. https://doi.org/10.1093/ije/dyv098

2. THIN Database Research Team. THIN Database Research. [cited 2018 Jan 17]. Available from: http://www.ucl.ac.uk/pcph/research-groups-themes/thin-pub/

3. Hippisley-Cox J. A Description of the 4th Version of the QRESEARCH Database: An analysis using QRESEARCH for the Department of Health. University of Nottingham; 2007.

4. Hippisley-Cox J, Coupland C. Predicting risk of emergency admission to hospital using primary care data: derivation and validation of QAdmissions score. BMJ Open. 2013;3(8):e003482. https://doi.org/10.1136/bmjopen-2013-003482

5. Scottish Health Informatics Study. SHIP: A Blueprint for Health Records Research in Scotland. 2012 [cited 2018 Jan 17]. Available from: http://www.scot-ship.ac.uk/sites/default/files/Reports/SHIP_BLUEPRINT_DOCUMENT_final_100712.pdf

6. Ford DV, Jones KH, Verplancke JP, Lyons RA, John G, Brown G, et al. The SAIL Databank: Building a national architecture for e-health research and evaluation. BMC Health Services Research. 2009;9:1–12. https://doi.org/10.1186/1472-6963-9-157

7. Kent County Council, Medway Council, Ashford CCG, Canterbury Coastal CCG, Dartford Gravesham and Swanley CCG, Medway CCG, et al. Transforming health and social care in Kent and Medway: Sustainability and Transformation Plan. 2016 [cited 2018 Jan 17]. Available from: http://kentandmedway.nhs.uk/wp-content/uploads/2017/03/20161021-Kent-and-Medway-STP-draft--as-submitted-ii.pdf

8. Office for National Statistics. Population estimates analysis tool. 2016 [cited 2018 Jan 17]. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesanalysistool

9. Kent County Council. Kent Annual Public Health Report 2015: Health Inequalities. 2015 [cited 2018 Jan 17]. Available from: http://www.kpho.org.uk/__data/assets/pdf_file/0005/57407/Final-Public-Health-Annual-Report-2015.pdf

10. Commission on the Future of Health and Social Care in England. The UK private health market. The King's Fund; 2014. p. 6. Available from: https://www.kingsfund.org.uk/sites/default/files/media/commission-appendix-uk-private-health-market.pdf

11. Office for National Statistics. UK Health Accounts: 2014. 2014 [cited 2018 Jan 17]. p. 1. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthcaresystem/bulletins/ukhealthaccounts/2014

12. NHS England. Mental Health Clustering Booklet 2016/17 (V5.0). 2016 [cited 2018 Jan 17]. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/523135/Annex_D_Mental_health_clustering_booklet.pdf

13. Smith T, Noble M, Noble S, Wright G, McLennan D, Plunkett E. The English Indices of Deprivation 2015: Technical Report. 2015 [cited 2018 Jan 17]. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/464485/English_Indices_of_Deprivation_2015_-_Technical-Report.pdf

14. Benson T. The history of the Read Codes: The inaugural James Read memorial lecture 2011. Informatics in Primary Care. 2012;19(3):173–82. https://doi.org/10.14236/jhi.v19i3.811

15. Curtis L, Burns A. Unit Costs of Health & Social Care 2016. 2016 [cited 2018 Jan 17]. Available from: http://www.pssru.ac.uk/project-pages/unit-costs/2016/

16. Morley JE, Vellas B, Abellan van Kan G, Anker SD, Bauer JM, Bernabei R, et al. Frailty consensus: A call to action. Journal of the American Medical Directors Association. 2013;14(6):392–7. https://doi.org/10.1016/j.jamda.2013.03.022

17. Clegg A, Bates C, Young J, Ryan R, Nichols L, Ann Teale E, et al. Development and validation of an electronic frailty index using routine primary care electronic health record data. Age and Ageing. 2016;45(3):353–60. https://doi.org/10.1093/ageing/afw039

18. Social Exclusion Task Force. Social Exclusion Task Force. Inclusinon Health: improving the way we meet the primary health care needs of the socially excluded. 2010.

19. Kendrick T, Stuart B, Newell C, Geraghty AWA, Moore M. Changes in rates of recorded depression in English primary care 2003-2013: Time trend analyses of effects of the economic recession, and the GP contract quality outcomes framework (QOF). Journal of Affective Disorders. 2015;180:68–78. https://doi.org/10.1016/j.jad.2015.03.040

20. HSCIC. Hospital Episode Statistics: Accident and Emergency Attendances in England 2014-15. 2016 [cited 2018 Jan 17]. Available from: http://digital.nhs.uk/catalogue/PUB19883

21. Public Health England. End of Life Care Profiles. 2015 [cited 2018 Jan 17]. Available from: https://fingertips.phe.org.uk/profile/end-of-life/data#page/0/gid/1938132883/pat/6/par/E12000008/ati/102/are/E06000036

Abbreviations

A&E	Accident & Emergency
CPRD	Clinical Practice Research Datalink
eFI	Electronic Frailty Index
EHRs	Electronic Health Records
GP	General Practitioner
ICD	International Classification of Diseases
KID	Kent Integrated Dataset
NHS	National Health Service
THIN	The Health Improvement Network