Institute of Education
EPPI-Centre

UCL

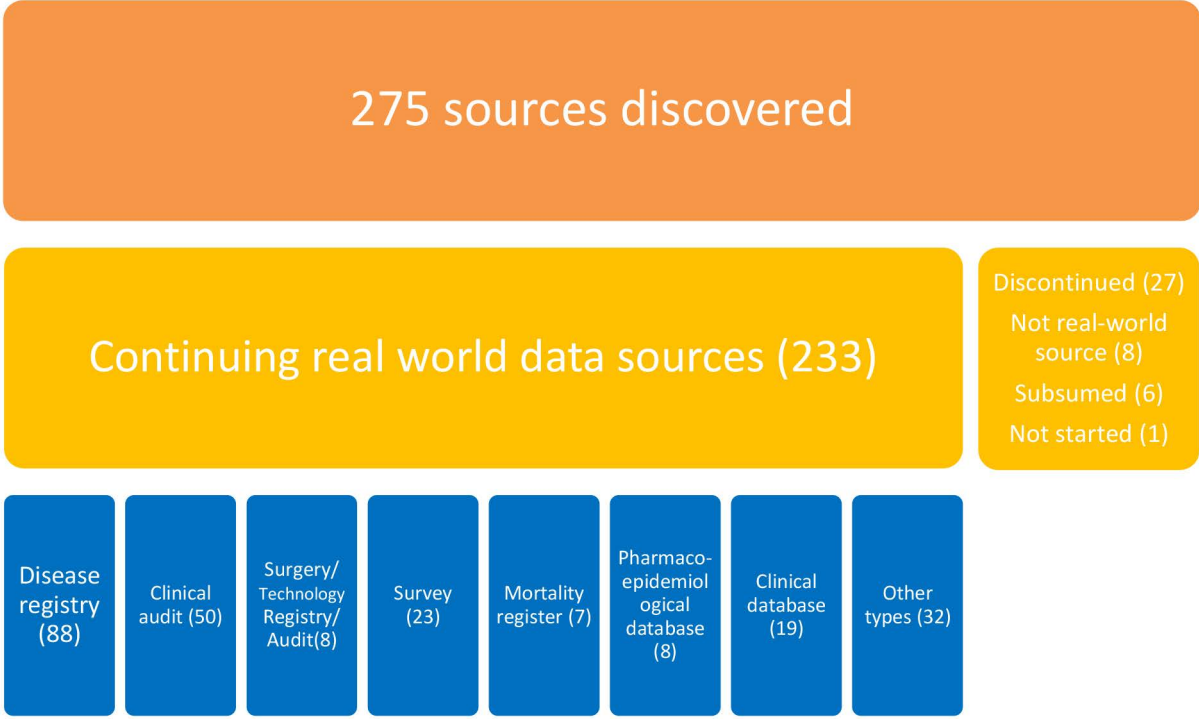# Identifying and appraising promising sources of UK clinical, health and social care data for use by NICE

Executive Summary

275 sources discovered

Continuing real world data sources (233)

Discontinued (27)

Not real-world source (8)

Subsumed (6)

Not started (1)

Disease registry (88)

Clinical audit (50)

Surgery/ Technology Registry/ Audit(8)

Survey (23)

Mortality register (7)

Pharmaco-epidemiological database (8)

Clinical database (19)

Other types (32)

Dylan Kneale, Meena Khatwa, James Thomas

ePPI CENTRE

SSRU Social Science Research Unit

Identifying and appraising promising sources of UK clinical, health and social care data for use by NICE

This is a summary of key findings from a report prepared by the Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre), UCL Institute of Education for the NICE Science Policy and Research Programme. The full report contains additional details of the EPPI-Centre's findings. The views expressed here are those of the authors and not necessarily those of NICE, the EPPI-Centre, or any of the stakeholders involved in this research (see acknowledgements section for a list). **All data were correct as of August 2015 and any errors and inaccuracies are the sole responsibility of the authors.**

Executive Summary

Introduction

Understanding the topography of this real-world data landscape is of prime interest to NICE (National Institute for Health and Care Excellence) in this study, as well as gaining a snapshot of the key areas of debate in the field. Specifically NICE seeks to understand the way in which different real-world data sources could help to support NICE to realise its strategic objectives and to this end, NICE has identified five key uses of real-world data that can help the organisation meet its overall strategic objectives. These are to:

(a) **Research the effectiveness of interventions or practice** in real-world (UK) settings (e.g. through monitoring outcomes or proxy outcomes). Data could be used to inform the modelling of clinical and/or cost effectiveness outcomes as part of guidance production. Real-world data can also help to resolve uncertainties that have been identified in existing NICE guidance.

(b) **Audit the implementation of guidance.** For example, to assess the equity of implementation across different groups (including socioeconomic, geographic, demographic and groups differentiated by different diseases/health conditions); this may also form part of performance monitoring systems

(c) **Provide information on resource use** and evaluate the potential impact of guidance.

(d) **Provide epidemiological information.** For example prevalence/incidence of diseases, natural history, co-morbidities and information on current practice.

(e) **Provide information on current practice to inform the development of NICE quality standards**

EPPI-Centre Key Findings and Recommendations

Key Findings

- The real-world data landscape remains complex and heterogeneous and composed of sources with different purposes, structures and collection methods. This heterogeneity may increase with opportunities stemming from the incorporation of new technologies in data collection (current quality assured sources are limited in number)
- Some real-world data sources are purposefully either set-up or re-developed to enhance their data linkages and to examine the presence/absence/effectiveness of integrated patient care; however, such sources are in the minority. Furthermore, the small number that are designed to enable the monitoring of care across providers, or at least have the capability to do so at a national level, have been utilised infrequently for this purpose in the literature.
- Data that offer the capacity to monitor transitions between health and social care do not currently exist at a national level, despite the increasing recognition of the interdependency between these sectors.
- Among the data sources we included, it was clear that no one data source represented a panacea for NICE's real world data needs. This does highlight the merits and importance of data linkage projects and is suggestive of a need to triangulate evidence across different data, particularly in order to understand the feasibility and impact of guidance.

Key Overall Recommendation

- There exists no overall catalogue or repository of real-world data sources for health, public health and social care, and previous initiatives aimed at creating such a resource have not been maintained. As much as there is a need for enhanced usage of the data, there is also a need for taking stock, integration, standardisation, and quality assurance of different sources. This research highlights a pressing need for a systematic approach to creating an inventory of sources with detailed meta-data and the funding to maintain this resource. This would represent an essential first step to support future initiatives aimed at enhancing the use of real-world data.

Key Recommendations for NICE

Increased utilisation of existing sources beyond clinical databases:

- Making recommendations is difficult around the use of specific data sources. However, NICE's current use of real-world data differs substantially from the landscape with respect to its low utilisation of clinical audit, disease registry and survey data. Several of the datasets profiled in-depth highlight the potential of different sources of survey, clinical database and audit data.
- We also recommend that NICE further review its use of disease registry and audit data and engage in dialogue with collectors and depositors of these data to explore the utility of these types of data. Many sources of data available from disease registries and clinical audits are currently underutilised.

Investment in capacity and partnership building

- Use of real-world data requires substantial investment of resource that allows for the organisation to develop an in-depth understanding and experience of using different real-world sources. The extent of this undertaking should not be underestimated; any commitments and real-world data usage strategies should be matched by resources that allow for developing expertise in-house and in developing partnerships with data depositors and academic experts.
- Many of the data sources profiled either have active user groups or hold regular consultative exercises. NICE should further investigate these opportunities and capitalise on these.

Strategy and influence

- NICE has the potential to influence the availability of real-world data sources and good practice around the collection and utilisation of real-world data. This influence could be used to develop good practice around aspects such as obtaining informed consent from patients or obtaining investment around the creation of data linkages. NICE should develop and publish an outward-facing policy around its use of real-world data which includes transparent means of influencing the state of the landscape, in order to ensure that sources continue to meet its organisational needs and to ensure alignment with national strategy.. Exerting such influence could not only lead to benefits to NICE, but will have broader positive impacts across other stakeholders more widely, and could lead to improved patient and service user outcomes. This influence could also extend to developing quality standards around the way in which data are collected that can be shared across the sector.
- Care.data represents an initiative that could potentially meet many of NICE's real-world data needs. NICE should engage in discussions with the Health and Social Care Information Centre (HSCIC) to better understand and prepare for potentially using these data, while continuing to monitor whether and how the initiative overcomes challenges identified in earlier stages.

Understanding implementation

- Finally, while NICE is potentially able to monitor the implementation of guidelines using several sources, it may still lack information on the underlying mechanisms as to how or why guidelines succeed or fail in implementation. Starting its own programme of real-world data collection in the form of surveys of practitioners may be a way of understanding the mechanisms of un/successful implementation. Such an approach has been adopted elsewhere, for example by the Swedish Council on Healthcare Technology Assessment (SBU).

Green shoots

There are three key factors as to why the state of the real world data landscape should be regarded with some optimism for NICE and more generally.

1. Firstly, while data linkage and the capacity to research patient journeys is not at the point where many would desire, there are several examples where these efforts have

been met with success and some of these have been met with a high degree of public acceptance. On a national level, the care.data initiative has restarted after a pause, and if these efforts succeed, they could meet many of NICE's real-world data requirements

2. Secondly, while we have been critical in the study about the representation of sources of patient reported outcomes, there are examples featured in the main report where patients have become more involved and have become gatekeepers to their own data (e.g. Salford Integrated Record), providing a possible model for the future. In addition, the ubiquity of smartphone technology and apps mean that ways of patients providing and managing their own information are increasing at pace.

3. Thirdly, methodological advances in the design and analysis of studies continue to ensure that real-world data becomes of greater utility for organisations, such as NICE, who wish to understand the implications of their decisions in real-world settings. These advances include the development of pragmatic trials using electronic health data which offer a balance between the methodological rigour of RCTs and the generalisability of observational studies. Several UK based organisations and teams – some of which are represented among the expert stakeholders involved in the present study – are involved in driving these advances and it is likely that future studies will feature the results of these undertakings extensively in their findings.
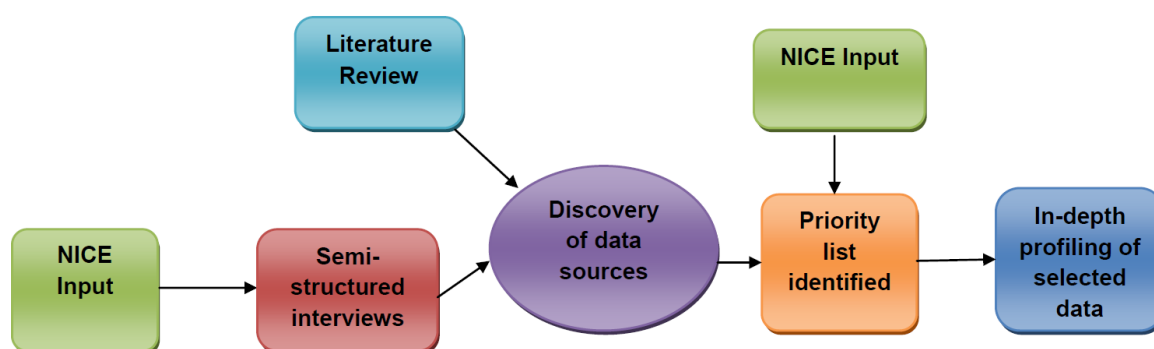
What is real world data?

The definition of real-world data can be contentious and different stakeholders have different views as to what constitutes 'real-world' data. Real world data is defined in this report through two key tenets:

a. The collection of real world data reflects the usual care or treatment provided to populations of patients, service users or the public. This therefore excludes conventional Randomised Controlled Trials (RCT(s)) but could include other forms of RCT design, namely pragmatic RCTs[1].

b. Real world data provides enough depth to assess trends around everyday practice, service usage, or to assess the effectiveness of interventions and their outcomes.

To meet the needs of NICE, we do not pay close attention to sources of data that have limited geographic representation, and prioritise those sources with national or regional representativeness.

---

[1] Pragmatic randomised controlled trials aim to mimic real life conditions and test the effectiveness of a range of interventions that are known to be safe. They can be instrumental in understanding the relative effectiveness where there is no apparent clinical advantage/disadvantage among currently accepted treatment (1.van Staa T-P, Goldacre B, Gulliford M, Cassell J, Pirmohamed M, Taweel A, et al. Pragmatic randomised trials using routine electronic health records: putting them to the test. Bmj. 2012;344:e55.)

Study Approach

This study is focussed on identifying some of the available opportunities to NICE in terms of real-world data sources. To reflect the remit of NICE, in this report we consider data that spans clinical, public health and social care fields. To create a topographical map of real-world data sources we use:

- Data from interviews with expert stakeholders
- Data from studies discovered in the literature

After drawing a long-list of sources, with the aid of NICE, we then identified those data sources that were not in current use, or were under-utilised, but were of interest to the organisation, and created an in-depth profile of eleven of these data sources.

Results - mapping of real-world datasets

In creating the map based on the literature and on interviewees' responses we discovered a total of 275 different sources of real-world data (figure 1), of which 233 are analysed further, being of most relevance to NICE. The remaining data sources were found to either have been discontinued (27) or subsumed into other studies (6 sources), were not actually real-world data sources (e.g. they were procedures or standards for application in the real world (8 sources)), or were at the protocol stage (one source).
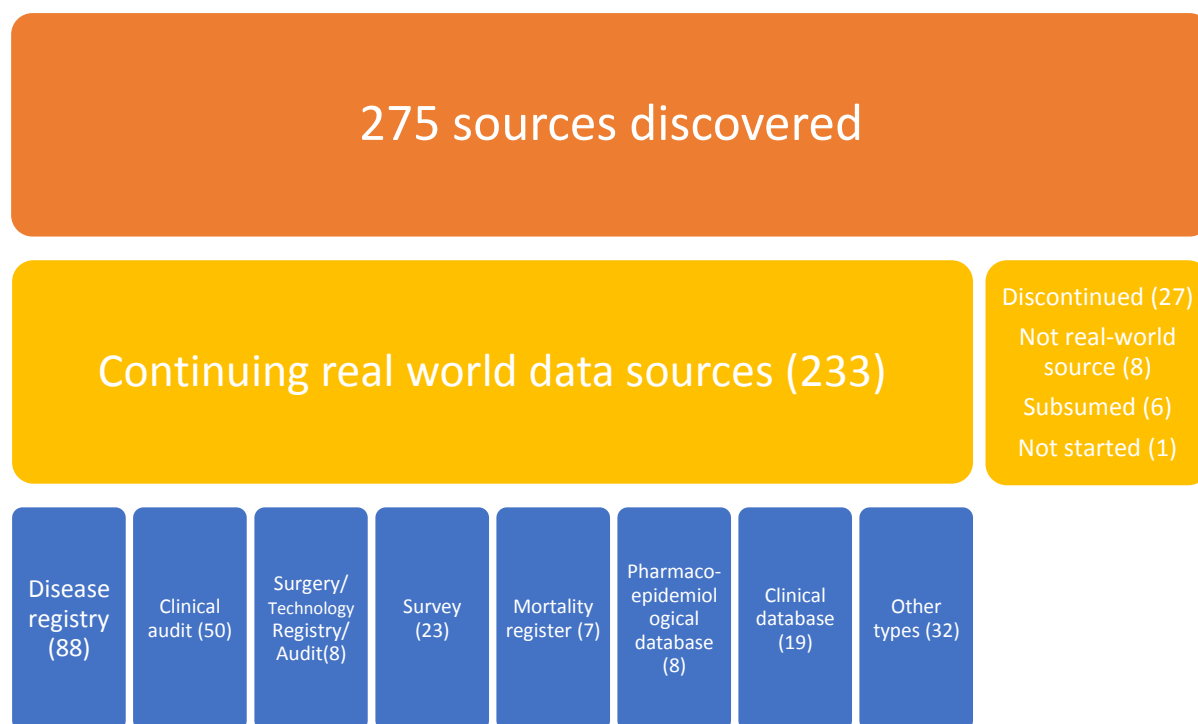
How does NICE currently use real-world data?

Internally, NICE conducted a review of its use of real-world data across different teams. This review asked teams to name which data source was currently being used, how these data were accessed, processes employed for accessing data, associated costs; and a brief description of how the data were used:

The majority of data were found to be used, internally at least, to either (i) inform on the uptake of NICE guidance and/or explore use of medication (nine reports of usage could be described in this way (one represented future plans)); or (ii) for health economic modelling (twenty reports could be described in this way). Some of the data appear to support the development of quality standards particularly around safe levels of staffing (three reports could be described in this way) and there was one reported use of data for monitoring epidemiological and demographic trends. One dataset was described as being used to establish the effectiveness of interventions.

Evidence suggested NICE's internal use of real-world data differs from the real-world data landscape in the following ways:

- The use of clinical audit data by NICE does not match the widespread availability of these data
- The use of disease registry data by NICE does not match the widespread availability of these data
- Several sources of survey data are currently not being utilised internally

---

[2] It was also acknowledged that the HSCIC website held a great number of sources that could also be potentially profiled – future exercises could include a more detailed inventory of the HSCIC datasets.

- Most of the data sources/functions currently in use appear to allow for cross-sectional analyses or repeated cross-sectional analyses only; patient-level longitudinal analyses appear to be conducted rarely
- Few datasets currently in use capture Patient Reported Outcomes
- NICE use an extensive array of different datasets to support understanding trends and the economic modelling of changes in prescribing trends
- There were no reports of NICE requesting additional data to be collected alongside standard data in any of the real-world sources
- Primary care data is based on The Health Improvement Network (THIN) data.
- Few of the existing sources allow for linkage across different services which a patient or service user may experience; however, these types of data were also underrepresented in the results of the mapping exercise
- For social care, none of the datasets described are directly sourced and explicitly focus on monitoring trends in private provision despite private provision being hugely important for the sector; these data were also underrepresented in the results of the mapping exercise

It should be noted that this review did not capture the multitude of data that are being used in work that NICE commissions from partner organisations.

What are some of the other broad debates and themes occurring in real-world data that NICE should be aware of?

There is no standard definition of real-world data and the term can be problematic

Often, real world data collected in primary care is viewed as a by-product of administrative or performance management activities. In social care, the definition of real world data tends to be broader from the outset and survey data was much more likely to be included in definitions. The breadth in the definition of real-world data was viewed as problematic by some and there is a need to clarify the distinction between 'real world data' and 'just all data'. Furthermore, the term 'real-world data' is not a familiar one across all disciplines.

Data for tracking patient and service user journeys are rare

Obtaining data that enables the tracking of patient and service user journeys and trajectories is fraught with difficulty. Much of this challenge is attributed to difficulties in being able to link data between sources using a common identifier, although considerable efforts are underway to link across some datasets. Data that enable monitoring of transitions between health and social care are especially underrepresented.

Using real world data may involve using multiple datasets in addressing a single research question

Real world data's particular strength is the potential to provide the most complete picture available of the health and care status of the nation, and the services and interventions that are received in maintaining or improving health and care status [2]. These data are derived from a representative subset of the population and can provide a population-based snapshot of (i) illness or care needs; (ii) contacts with providers that take place; (iii)

information on treatments or care packages, and (iv) ideally provides enough information on the outcomes of individuals. However, the real-world data landscape remains fragmented and the different elements that can provide a holistic understanding of patient and service user trajectories are stored in different sources with no means of linking.

Real-world data analysts will often encounter a trade-off between data that provides a depth of information on patient characteristics and data that provides a breadth of information on the services or interventions they receive. Real-world data projects often incorporate data from a number of different sources in order to overcome limitations within any given data source. In other real-world data projects, data from different sources are used in order to triangulate the findings and overcome potential concerns around representativeness or bias.

> *"There is a trade-off between having more information in terms of numbers and information in terms of breadth and depth of indicators. So survey data such as ELSA [English Longitudinal Study of Ageing] will give you a lot more in terms of quality of characteristics – income, wealth, needs, households' composition, service users etc. Certain outcomes will be much more limited on the other hand data from services; there will be thousands of cases in other sources - but much more limited – and the data and may not be of the same quality. We try to combine the data, look at patterns from both".*

Real world data complements the findings from randomised controlled trials

The main defining advantage of real-world data, besides apparent advantages in terms of cost, sample size and representativeness, is its (ostensibly) high external validity [1]. The external validity reflects both the delivery of an intervention to a group that is representative of the general population, but more crucially in the delivery of the control, which usually involves an alternative treatment regimen (best available alternative) as opposed to a placebo. While there is an expanding literature citing studies and study protocols that have been conducted using real-world data, interviewees (especially those from clinical backgrounds) emphasised that real-world data was not a replacement for/superseded the findings from RCT studies. Real-world data is prone to forms of epidemiological bias unlikely to be replicated in findings from well designed and executed RCT studies [1, 3]; however, as several interviewees pointed out, RCT study data can also be subject to bias, and some identified that observational data was subject to greater scrutiny despite its superior properties in terms of transparency, than RCT data are.

Future directions

Two themes emerged around future potential of real-world data. The first of these is around the expanding potential of pragmatic clinical trials (PCTs). Unlike traditional RCTs,

PCTs are trials that take place within real-world environments and among representative samples of patients, thereby placing the focus on establishing the effectiveness of interventions, as opposed to their efficacy. Within a PCT, patients are randomised to receive an intervention or control treatment but the focus on mimicking real-world conditions means that, among other factors: (i) the control treatment provided often represents the best viable alternative already in place (as opposed to a placebo as can be the case in some RCTs), (ii) the patients randomised reflect the normal range of patients in terms of disease severity, comorbidity and demographic characteristics; and (iii) the measures of effectiveness collected as outcomes are valid and easily understood by a range of stakeholders, including clinicians, patients, policy-makers, and health commissioners. Real-world data collected through electronic health records was viewed as the basis for designing and undertaking a greater number of pragmatic trials (PCTs) and a number of real-world sources theoretically provide the means of implementing studies and monitoring outcomes in real-time. Evidence from PCTs is likely to be of substantial interest to NICE in establishing the effectiveness of interventions in real world settings while maintaining randomisation, thereby eliminating or at least substantially reducing the occurrence of channelling bias; the proliferation of real world data sources may facilitate this form of evidence to become increasingly frequent in the future.

A second theme that emerged was around new technologies stimulating new forms of real world data to be collected. Methods of collecting patient reported outcomes are shifting from paper to digital devices (smartphones and tablets): *"we have a lot of interest in technology where people get messages on their mobile phone to fill out symptoms, whether these are severe and so on. Uptake is very good and this type of model can be utilised for trials quite easily… where you have mobile phone technology sending information you don't have lots of paperwork… modern technology can help a lot with that. Also with ipads there is a strong movement to increases use in that."*

Recommended specific sources to consider from interviews included:

1) Opportunities are available for assessing individual level service user outcomes through the Adult Social Care Survey (ASCS)

2) Understanding patient journeys and experiences using the broad scope of data contained within the National Cancer Data Repository

3) Assessing resource usage using National Minimum Data Set for Social Care (Skills for Care)

4) Exploring primary care practice using three of the large GP datasets
   a) QResearch
   b) Clinical Practice Research Datalink (CPRD)
   c) The Health Improvement Network (THIN)

5) Understanding the contribution of risk factors to disease outcomes using the Whitehall II study

6) Exploiting the longevity and near-universality of the Myocardial Ischaemia National Audit Project (MINAP)

7) Examining epidemiological trends using Health Survey for England

8) Examining life course experiences on patterns of ageing using the National Child Development Study

9) Gaining a snapshot of social care and health service usage and needs of older people using the English Longitudinal Study of Ageing

10) Understanding epidemiological and care trends among households, including ethnic minorities, using Understanding Society

11) Monitoring Social and Health care trends, experiences and monitoring the implementation of standards using Care Quality Commission data and reports

12) Gaining an insight into patient experiences using Patients Like Me

13) Tracking data on patient journeys in integrated delivery networks: the potential of Scottish Health Informatics Programme (SHIP)

14) Exploiting Hospital Episodes Statistics Data as a multipurpose dataset

15) Exploring Epidemiological Trends using the Avon Longitudinal Study of Parents and Children (ALSPAC)

16) Using information from an integrated learning system through the Salford Integrated Record

17) Investigating the application of GP records

18) Understanding the effectiveness of interventions using National Joint Registry as a registry that is collecting longitudinal outcomes and patient reported outcomes

19) Understanding the effectiveness of interventions and monitoring the impact of guidance using the Sentinel Stroke National Audit Programme (SSNAP)

20) Understanding the effectiveness of interventions and monitoring the impact of guidance using the Renal Registry

21) Understanding the effectiveness of interventions and monitoring epidemiological trends using Adult critical care case mix programme (managed by ICNARC)

22) Harnessing the potential of cardiovascular audit and register data to address NICE's real world data needs

23) Data from the National Diabetes Audit; "the most advanced for long-term conditions"

24) Capturing genetic information on biomarkers in the UK Biobank

25) Calculating cost effectiveness based on data from the Personal and Social Services Research Unit

26) Mental Health and Learning Disabilities Data Set

27) Understanding trends in screening rates, healthcare and epidemiology using Quality and Outcomes Framework (QoF) data

28) Understanding epidemiological trends and measuring the effectiveness of interventions using the UK Inflammatory Bowel Disease Audit

29) Harnessing the potential of audit data to address NICE's real world data needs through clinical audits conducted by the Royal College of Surgeons Clinical Effectiveness Unit

30) Data used to populate NICE's Return on Investment Tools

31) Data from private health providers and insurers

Further sources that were shortlisted for consideration based on the literature/input from NICE were: (i) Care.data (ii) Prescribing observatory for Mental Health. It was also acknowledged that the HSCIC website held a great number of sources that could also be potentially profiled – future exercises could include a more detailed inventory of the HSCIC datasets.

Specific data profiled

From the earlier long-list of 30+ datasets, a selection of eleven data sources was chosen for in-depth profiling based on input from the NICE steering group. Datasets were prioritised if they were not in current use by NICE and where they appeared to meet some of the broader gaps in usage or addressed any of the themes emerging from the interviews. A template was developed to capture the properties of different sources according to their suitability for NICE's intended usage.

Focus on the potential utility of different datasets for NICE

All the profiled data sources are likely to have some utility to NICE dependent on the research question and making a specific recommendation around use is challenging as this is very much dependent on the context and the focus of the research question. The following section summarises the utility of the different sources for NICE. A full description of each dataset is provided in the main report.

English Longitudinal Study of Ageing (ELSA)

- ELSA has been used to establish the effectiveness of interventions at a population level using observational methods, for example in a cost-benefit analysis of cataract surgery among ELSA respondents [6]. ELSA may be less suitable for establishing the effectiveness of more specialist interventions/practice, or establishing how interventions/practice vary among minority groups.
- ELSA can be used to determine the implementation of guidance through examining broad population-level temporal changes in the receipt of common interventions or practice. For example ELSA data were used to examine shortfalls in care for chronic conditions using set quality indicators [7]. Without further linkages, ELSA data may be less suitable for explaining the underlying mechanisms around the implementation of guidance, beyond patient/service-user characteristics.
- ELSA data can be used to provide information on some aspects of resource use, for example how many people receive common interventions, and can be used to establish how access may vary by individual patient characteristics.
- ELSA data can be used to establish self-reported levels and determinants of many age related conditions and non-communicable diseases and more broadly information on lifestyle behaviours and attitudes among older people.
- ELSA data may be less suitable for establishing the incidence/prevalence/outcomes of very uncommon diseases/conditions/interventions.

Community Mental Health Survey (CMHS)

- The CMHS data have been used to monitor the implementation of guidance, for example in monitoring the implementation of guidance aiming to strengthen support for service users during times of turnover in staffing [8]. The data have also been used to draw together guidance around expected standards of care [9]. There may also be potential to use the data to monitor different aspects of resource usage.
- The focus of the survey is on service user experiences and there is less information on outcomes following receipt of different forms of care, limiting the utility of the data with respect to establishing the effectiveness of interventions. The data are less suitable as a tool for monitoring epidemiological patterns in mental health.

Clinical Practice Research Datalink (CPRD)

- CPRD data have utility for NICE through the flexibility in being able to collect additional fields. CPRD data are also available to medical researchers based outside UK universities potentially expanding the pool of potential partners with which NICE could

work in using the dataset. The long established nature of CPRD (based on the General Practice Research Database) means that several retrospective studies could also be potentially conducted using these data.

- There are numerous examples where CPRD (and GPRD) data have been used in studies that cover all of NICE's intended uses of real-world data. For example, CPRD have been used to evaluate changes in cancer diagnostic intervals following the introduction of NICE guidance [11]. Given the potential to draw large samples, studies could be implemented that examine the epidemiology/outcomes/implementation of rare or less common conditions and procedures. Unlike survey-based sources, for example ELSA and HSE, and in the absence of further data collection, there is potential to examine only a limited range of patient-level intrinsic factors, although these may be sufficient for many studies.

- Data linkages will expand the utility of CPRD data for NICE; current linkages include those with MINAP data, National Cancer Intelligence Network data and HES data. Area level data are also available including Index of Multiple Deprivation data and Townsend deprivation scores [10]. Further data linkages are planned.

QResearch

- QResearch is of interest to NICE for many of the real world data uses identified by NICE, but access appears to be restricted to research consortiums led by academic institutions. Nevertheless, given the substantial potential of these data, NICE could consider ways of developing research projects based on QResearch data led by universities.

- There is potential for QResearch data to be used in studies that cover all of NICE's intended uses of real-world data. The use of QResearch data in developing risk prediction scores may also be of interest to NICE, potentially around forecasting and modelling future disease burden.

- Given the potential to draw large samples, studies can be implemented that examine the epidemiology/outcomes/implementation of rare or less common conditions and procedures. One example is a study of peanut allergy, where a prevalence rate of 0.51 per 1000 patients in the UK was estimated [14].

- The study depositors state that QResearch data are suitable for case control studies designed to examine risk factors for onset of disease, cross sectional surveys, cohort studies and sample size calculations (for non-observational studies) [15].

- As is the case for all three large primary care databases, there is potential to examine only a limited range of patient-level intrinsic factors, although these may be sufficient for many studies.

The Health Improvement Network (THIN)

- There are numerous examples where THIN data have been used in studies that cover all of NICE's intended uses of real-world data. For example, THIN data have been used to examine equity in access to cancer screening among people with Intellectual Disabilities compared to those without across different types of cancer [19].

- Data linkages expand the utility of THIN, and THIN data have been linked with Hospital Episodes Statistics (HES) data, providing potential for studying continuity in care between primary and secondary care. A number of patient postcode-based socioeconomic, ethnicity and environmental indicators are available to researchers including Townsend deprivation quintile scores.

- Overall there is a wide scope for analysing data reflecting outcomes and experiences of morbidity and mortality at primary care level, as well as trends in the care and treatment provided. These data can also be linked to HES data allowing for potential tracking of patient journeys between primary and secondary care. As is the case for all three large primary care databases, there is potential to examine only a limited range of patient-level intrinsic factors, although these may be sufficient for many studies.

- THIN data have utility for NICE through the flexibility in being able to collect additional fields and the potential to conduct research based on free-text fields. THIN data are also available to medical researchers based outside UK universities potentially expanding the pool of potential partners with which NICE could work with in utilising real world data.

National Minimum Data Set for Social Care (NMDS-SC)

- NMDS-SC is a specialist dataset suitable for monitoring trends in the social care workforce. This data can potentially help NICE to understand workforce capabilities and undertake preliminary work to understand the feasibility of implementing new standards and guidance in social care settings.
- The data may be suitable to examine changes following the implementation of NICE guidance at a workforce level in terms of indicators such as pay, training or necessary skills. They may also be useful in helping to set benchmarks and develop quality standards around workforce capacity and skills. The data have also been incorporated into calculations of resource use in the literature [24]. While the data do not provide insight into epidemiological trends per se, they do provide insight into the workforce preparedness for responding to epidemiological challenges, such as dementia [25].
- As social care outcomes are not collected in NMDS-SC, it is unlikely that these data are suitable for researching the effectiveness of interventions and practice.

Health Survey for England (HSE)

- HSE was suggested in the context of monitoring epidemiological trends although the potential usage extends beyond this purpose alone and potentially HSE data can be used to gain an understanding of trends over time in terms of resource utilisation, trends in social care needs and usage, trends in lifestyles and social determinants of health, and some trends in prescribing, service usage and attitudes to health. With regards to researching the effectiveness of interventions, in the absence of data linkages, there may be more limited potential to measure the effectiveness of interventions or changes in practice. Examples where data have been linked to explore later outcomes include an examination of fruit and vegetable intake and mortality [30].
- The survey data may be of great utility for NICE in gathering contextual information critical in the assessing feasibility of different forms of guidance aimed at public health and social care challenges. The data also have the added advantage of being relatively easy to obtain for further secondary data analysis and are free to use.
- There is scope for auditing the implementation of guidance through examining change in practice at a population level; one of the strengths of HSE data in doing so is the ability to examine social or medical inequalities in the implementation of guidance. Some HSE information may be suitable in providing information for the development of NICE quality standards and these data may be particularly useful where the standard is based on meeting a certain level of patient satisfaction or experience.

Adult Social Care Survey

- ASCS is a survey of users' satisfaction with the care that they receive. Such data can be used in forming guidance that is based on user experience and patient reported outcomes. There may be limited scope for undertaking secondary analysis of the individual service user data without further permissions being sought. Nevertheless, the detailed reports and tables produced may allow for gaining a good level of understanding of aspects of service user satisfaction with their care and broader aspects of wellbeing.
- With regards to measuring the effectiveness of practice, while it may be possible to undertake repeated cross-sectional studies and examine the impact of changing practice on user experiences, fully assessing the effectiveness of interventions through measuring longitudinal changes at a service-user level will be challenging with these data. However, it may be possible to assess whether guidance is being implemented,

particularly around service user satisfaction or service user reported experiences, through analysing change (for example at a Local Authority level).

- With regards to using the data as an epidemiological tool, the study provides a snapshot of general health trends and social care needs but among a population who are receiving LA assistance for these health needs (the sample design represents a caveat around the applicability of the data). There may be scope for the data to be used to form quality standards around social care experiences and trajectories – for example around information advice and guidance received by older people in accessing care.

Salford Integrated Record

- SIR was suggested as a source of data that may have the potential to overcome the limitations of other data source and examine patients' integrated care pathways. The potential of the data for research purposes are likely to be in the process of being realised and there are comparatively few publications using these data in the literature; the data may have been used initially to mainly facilitate clinical decision-making and performance management. Perhaps one of the most appealing characteristics of the data, given the current climate around the use and ethics of electronic health records in medical research, is the high degree of patient involvement and the ability of patients to access their own records.
- The data hold substantial potential for improving patient care. The integration of primary and secondary care data allows for research tracking patient outcomes across care providers (through examining Integrated Care Pathways (ICP)). One initiative using the data in this way is the Collaborative Online Care Pathway Investigation Tool that is being used to examine missed opportunities in patient care – that is where primary prevention opportunities were missed which could lead to adverse health outcomes. This initiative is focussed on modelling the circumstances and frequency of variance between idealised ICP and the actual care provided [37].

Prescribing Observatory for Mental Health

- One of the key criteria for choosing a topic focus of the POMH-UK is that the topics are relevant for monitoring the implementation of NICE guidelines. This has direct relevance to one of the intended uses of real-world data by NICE. An example of study directly assessing the implementation of NICE guidance can be found in a study of renal and thyroid functioning among patients who are prescribed lithium [39].
- The utility of the data for other more research-focused or evaluative activities, for example in assessing the effectiveness of interventions or monitoring epidemiological trends, may be more limited. The data are not widely used in the literature and it is unclear the extent to which these data are made available for re-analysis, reflecting their primary function as a quality improvement tool. Nevertheless, there are several important questions that could be addressed for NICE as there may be potential to understand whether practice/outputs have changed over time. In addition, this source represents one of the few specialist sources of real-world data on mental health encountered.

Care.data

- If successfully implemented, care.data would make a substantial contribution to the real-world data needs of NICE and other organisations. The data could allow for establishing the long-term effectiveness of interventions through the capacity to track patient journeys through primary and into secondary care as standard, something that rarely occurs as standard in real-world data projects and sources. Uniquely, it could also potentially, allow for insight into patterns of social care and their relationship clinical and public health data.
- At the time of writing it is too early to tell the extent to which care.data has been able to overcome the challenges encountered, particularly around consent and

conditions around data usage. The results of the pathfinder exercise will offer further insight into the viability of the whole project; the majority of testing in pathfinder areas is due to begin later this year.

References

1.   van Staa T-P, Goldacre B, Gulliford M, Cassell J, Pirmohamed M, Taweel A, et al. Pragmatic randomised trials using routine electronic health records: putting them to the test. Bmj. 2012;344:e55.
2.   Gnani S, Majeed A. A user's guide to data collected in primary care in England. Cambridge: Eastern Region Public Health Observatory; 2006.
3.   Newton J, Garner S. Disease registers in England. Oxford: Institute of Health Sciences, University of Oxford, 2002.
4.   Steptoe A, Breeze E, Banks J, Nazroo J. Cohort profile: the English longitudinal study of ageing. International journal of epidemiology. 2013;42:1640–8.
5.   Gray M. A review of Data linkage procedures at NatCen London: NatCen, 2009.
6.   Weale M. A cost-benefit analysis of cataract surgery based on the English Longitudinal Survey of Ageing. Journal of health economics. 2011;30(4):730-9.
7.   Steel N, Bachmann M, Maisey S, Shekelle P, Breeze E, Marmot M, et al. Self reported receipt of care consistent with 32 quality indicators: national population survey of adults aged 50 or more in England. Bmj. 2008;337.
8.   CQC. National Summary of the Results for the 2014 Community Mental Health Survey. London: Care Quality Commission, 2014.
9.   Kendall T, Crawford MJ, Taylor C, Whittington C, Rose D. Improving the experience of care for adults using NHS mental health services: summary of NICE guidance. Bmj. 2012;344.
10.   Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, van Staa T, et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). International journal of epidemiology. 2015:dyv098.
11.   Neal RD, Din NU, Hamilton W, Ukoumunne OC, Carter B, Stapley S, et al. Comparison of cancer diagnostic intervals before and after implementation of NICE guidelines: analysis of data from the UK General Practice Research Database. British journal of cancer. 2014;110(3):584-92.
12.   Hippisley-Cox J, Stables D, Pringle M. QRESEARCH: a new general practice database for research. Journal of Innovation in Health Informatics. 2004;12(1):49-50.
13.   QResearch. A summary of public health indicators using electronic data from primary care Nottingham: QResearch and the Health and Social Care Information Centre, 2008.
14.   Kotz D, Simpson CR, Sheikh A. Incidence, prevalence, and trends of general practitioner–recorded diagnosis of peanut allergy in England, 2001 to 2005. Journal of Allergy and Clinical Immunology. 2011;127(3):623-30. e1.
15.   QResearch. What is QResearch? Nottingham: University of Nottingham; 2012. Available from: http://www.qresearch.org/SitePages/What%20Is%20QResearch.aspx.
16.   Lewis JD, Schinnar R, Bilker WB, Wang X, Strom BL. Validation studies of the health improvement network (THIN) database for pharmacoepidemiology research. Pharmacoepidemiology and drug safety. 2007;16(4):393-401.
17.   Cai B, Xu W, Bortnichak E, Watson DJ. An algorithm to identify medical practices common to both the General Practice Research Database and The Health Improvement Network database. Pharmacoepidemiology and drug safety. 2012;21(7):770-4.
18.   UCL THIN Research. The THIN database: UCL; 2015 [cited 2015 31st May]. Available from: https://www.ucl.ac.uk/pcph/research-groups-themes/thin-pub/database.
19.   Osborn DPJ, Horsfall L, Hassiotis A, Petersen I, Walters K, Nazareth I. Access to cancer screening in people with learning disabilities in the UK: cohort study in the health improvement network, a primary care research database. 2012.
20.   Skills for Care. NMDS-SC key information and statistics reports: May 2015. Leeds: Skills for Care, 2015.
21.   Skills for Care. NMDS-SC Raw Data User Guide - Worker File. Leeds: Skills for Care, 2013.
22.   Hussein S, Manthorpe J. Structural marginalisation among the long-term care workforce in England: evidence from mixed-effect models of national pay data. Ageing and Society. 2014;34(01):21-41.
23.   Skills for Care. The state of the adult social care sector and workforce in England 2015. Leeds: Skills for Care, 2015.
24.   Curtis L. Unit Costs of Health and Social Care. Canterbury, Kent: PSSRU, University of Kent, 2009.
25.   Hussein S, Manthorpe J. The dementia social care workforce in England: secondary analysis of a national workforce dataset. Aging & mental health. 2012;16(1):110-8.
26.   Thompson J, Wittenberg R, Henderson C, Darton R. Social care: need for and receipt of help. In: Craig R, Mindell J, editors. (Vol 1) Health, social care and lifestyles. Leeds: Health and Social Care Information Centre; 2014.
27.   Mindell J, Biddulph JP, Hirani V, Stamatakis E, Craig R, Nunn S, et al. Cohort profile: the health survey for England. International journal of epidemiology. 2012;41(6):1585-93.
28.   Nazroo JY, Falaschetti E, Pierce M, Primatesta P. Ethnic inequalities in access to and outcomes of healthcare: analysis of the Health Survey for England. Journal of epidemiology and community health. 2009;63(12):1022-7.
29.   Andrew MK. Social capital, health, and care home residence among older adults: A secondary analysis of the Health Survey for England 2000. European Journal of Ageing. 2005;2(2):137-48.

30.   Oyebode O, Gordon-Dseagu V, Walker A, Mindell JS. Fruit and vegetable consumption and all-cause, cancer and CVD mortality: analysis of Health Survey for England data. Journal of epidemiology and community health. 2014:jech-2013-203500.
31.   Smith N, Malley J. Understanding and addressing underrepresentation in a postal survey of social care users. London: London School of Economics, 2012.
32.   HSCIC. Personal Social Services Adult Social Care Survey, England 2013-14, Final release. Leeds: Health and Social Care Information Centre, 2014.
33.   Diggle P, editor Statistics: a Data Science for the Twenty-First Century. CHICAS; 2015 June; Lancaster University Medical School.
34.   New JP, Bakerly ND, Leather D, Woodcock A. Obtaining real-world evidence: the Salford Lung Study. Thorax. 2014:thoraxjnl-2014-205259.
35.   Davis KJ, New JP, Delderfield MR, Stein ND, Leather D, Vestbo J, et al. Evaluating Feasibility of an EMR-enabled Randomized Clinical Trial in the UK [cited 2015 July 6th]. Available from: http://www.idrn.org/documents/events/presentations/primarycaredatabases/Posters/Davis%20K.pdf.
36.   Sperrin M, Thew S, Weatherall J, Dixon W, Buchan I, editors. Quantifying the longitudinal value of healthcare record collections for pharmacoepidemiology. AMIA Annual Symposium Proceedings; 2011: American Medical Informatics Association.
37.   Ainsworth J, Buchan I. COCPIT: a tool for integrated care pathway variance analysis. Studies in health technology and informatics. 2011;180:995-9.
38.   Paton C, Barnes TRE, Shingleton-Smith A, McAllister-Williams RH, Kirkbride J, Jones PB, et al. Lithium in bipolar and other affective disorders: prescribing practice in the UK. Journal of Psychopharmacology. 2010;24(12):1739-46.
39.   Collins N, Barnes TRE, Shingleton-Smith A, Gerrett D, Paton C. Standards of lithium monitoring in mental health trusts in the UK. BMC psychiatry. 2010;10(1):80.
40.   HSCIC. General Practice Extraction Service (GPES): Customer Requirement Summary. Leeds: HSCIC, 2013.
41.   Cabinet Office. Major Projects Authority Annual Report 2014-15. London: Major Projects Authority, 2015.

Acknowledgements

The Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre) is part of the Social Science Research Unit (SSRU), UCL Institute of Education, University College London.

The EPPI-Centre was established in 1993 to address the need for a systematic approach to the organisation and review of evidence-based work on social interventions. The work and publications of the Centre engage health and education policy makers, practitioners and service users in discussions about how researchers can make their work more relevant and how to use research findings.

Founded in 1990, the Social Science Research Unit (SSRU) is based at the UCL Institute of Education, University College London. Our mission is to engage in and otherwise promote rigorous, ethical and participative social research as well as to support evidence-informed public policy and practice across a range of domains including education, health and welfare, guided by a concern for human rights, social justice and the development of human potential.

The views expressed in this work are those of the authors and do not necessarily reflect the views of the EPPI-Centre or the funder. All errors and omissions remain those of the authors.

This document is available in a range of accessible formats including large print.

Please contact the UCL Institute of Education for assistance:
telephone: +44 (0)20 7947 9556   email: info@ioe.ac.uk