Check for updates

Research Article

Linkage of routinely collected NHS data to evaluate liaison mental health services: challenges and lessons learned

Elspeth Guthrie,[1*] Allan House,[1] Chris Smith,[2] Sam Relton,[1] Daniel Romeu,[1] Sonia Saraiva,[1] Peter Trigwell,[3] Robert West,[1] Farag Shuweihdi,[1] Mike Crawford,[4] Matt Fossey,[5] Jenny Hewison,[1] Claire Hulme[6] and Sandy Tubeuf[7]

[1]Leeds Institute of Health Sciences, University of Leeds, Leeds, UK
[2]Department of Computer Science, University of York, York, UK
[3]National Inpatient Centre for Psychological Medicine, Leeds and York Partnership NHS Foundation Trust, Leeds, UK
[4]Department of Brain Sciences, Faculty of Medicine, Imperial College, London, UK
[5]Veterans and Families Institute for Military Social Research, Faculty of Health, Social Care and Education, Anglia Ruskin University, Chelmsford, UK
[6]Faculty of Health and Life Sciences, University of Exeter, Exeter, UK
[7]Institute of Health and Society (IRSS) and Institute of Economic and Social Research (IRES), Université catholique de Louvain, Louvain, Belgium

*Corresponding author e.a.guthrie@leeds.ac.uk

Abstract

Background: Liaison mental health services provide mental health care to patients in acute hospital settings. Evaluation of liaison services is challenging due to their heterogeneous organisation and delivery, high case throughput and varied patient case mix. We aimed to link routinely collected National Health Service data from secondary care settings, chosen for their service characteristics, to data from primary care to evaluate hospital-based liaison mental health services in England.

Methods: We planned to compare patients referred to hospital-based liaison services with comparable patients in the same hospital not referred to liaison services and comparable patients in hospitals without any liaison services. We designed and enacted a methodology to link data from: (1) Hospital Episode Statistics, a database controlled by the National Health Service Digital and (2) ResearchOne, a primary care database controlled by The Phoenix Partnership.

Results: Obtaining approvals for the steps prespecified in the methodological protocol took 907 days. Enactment following approvals took 385 days. Data supplied from Hospital Episode Statistics contained 181,063 patients from 6 hospitals (mean = 30,177, standard deviation = 28,875.86) who matched the inclusion and exclusion criteria. Data supplied from ResearchOne contained 33,666 (18.6%) of these patients from the 6 hospitals (mean = 5611, standard deviation = 5206.59).

Discussion: Time required for approvals and enactment was attributable to slowness of data handling processes within each data holder and to resolution of technical and organisational queries between them. Variation in number of patients for which data was supplied between databases and between hospitals was attributable to coding inconsistencies and to the limited intersection of patient populations between databases and variation in recording practices between hospitals.

Conclusion: Although it is technically feasible to link primary and secondary care data, the current system is challenging, complicated, unnecessarily bureaucratic, time consuming and costly. This limits the number of studies that could be conducted with these rich data sources.

Funding: This article presents independent research funded by the National Institute for Health and Care Research (NIHR) Health and Social Care Delivery Research programme as award number 13/58/08.

A plain language summary of this research article is available on the NIHR Journals Library Website https://doi.org/10.3310/WCPA5283.

Background and introduction

Liaison mental health services (LMHS) provide assessment and treatment for people with co-existing physical and mental health problems.[1–4] In the UK, such services are generally based in an acute hospital setting. Different terms are used for these services including liaison psychiatry, consultation-liaison psychiatry and psychological medicine services, but we have chosen to use LMHS, as this is the preferred term currently in the UK. There has been a recent expansion of LMHS in England such that all acute hospitals with an emergency department now have a liaison mental health team.[5] Liaison services have the potential to improve both the quality of care and overall outcomes for people with mental and physical health problems. However, research evidence relating to the cost-effectiveness of liaison mental health teams, as opposed to the individual interventions used within those services, is currently limited. Several small-scale local evaluations have shown that liaison services may result in large cost savings, primarily by reducing length of stay of acute hospital inpatients,[6–9] but the research methodology in these studies is weak, as most did not employ a comparator to control for any routine changes to service delivery. A recent re-evaluation of one of these studies[6] compared changes in the hospital with the liaison service to a similar acute hospital that did not have a liaison service during the study period.[10] An interrupted time series analysis showed a reduction in length of stay in both hospitals during the same periods of time with no evidence for any specific effect of the liaison service.[10]

A recent systematic review of the effectiveness of consultation-liaison psychiatry service models identified only eight trials, all of which were at least 10 years old, and none of which reflected usual clinical practice.[11] All trials recruited patients who had been identified by some sort of screening process which is not routine practice in the UK, four excluded patients who had severe cognitive impairment or were unable to provide informed consent, six excluded patients with severe physical illness and three excluded patients who had expressed suicidal thoughts. The authors of the systematic review concluded that all the trials had methodological limitations and there was no evidence of effectiveness.

Although there are other more modern models of delivering care than the consultation model (in which the patient is referred directly to the liaison team from acute hospital staff), the same difficulties in evaluating heterogeneous services on a national basis still apply.

There is one very large trial (*n* > 3500 participants) of proactive integrated consultation-liaison psychiatry (patients are identified proactively on admission and reviewed on a daily basis) which is currently being evaluated for its impact on length of stay of older patients on acute medical wards.[12] The trial has yet to report and will provide valuable information regarding the impact of this model on a subset of patients seen by liaison services, but it will not be able to determine the overall effectiveness of liaison services for all the patients they currently assess and treat. Most medium- to large-sized liaison services see over 100 patients per week; just over half of these patients (55%) are seen on acute medical wards and one-third are assessed in the emergency department.[13] Of those assessed on acute wards, a significant proportion are working-age adults.[13]

Liaison teams vary in their staff composition, working practices, relations with other mental health teams, policies such as responses times, hours of coverage and location (ward or Emergency Department based),[14] making a randomised controlled trial impracticable due to the number of hospitals and patients required to account for the heterogeneity in service delivery. In addition, a randomised controlled trial may not be the best vehicle for the evaluation of routine clinical services: the internal validity of a trial becomes increasingly compromised by the heterogeneity of the intervention under study, and external validity is also compromised as even cluster designs necessarily have to exclude many patients who would routinely be referred to and assessed by clinical services. The analysis of routinely collected clinical NHS data, which is not predicated on individual patient consent or ability to speak English and can accommodate heterogeneity of service delivery, provides an alternative to evaluate NHS services, using all relevant patient data. High-quality evaluations using this approach are now included in its evidence reviews by the UK's National Institute for Health and Care Excellence (https://www.nice.org.uk/about/who-we-are/corporate-publications/the-nice-strategy-2021-to-2026).

A robust evaluation of the impact of LMHS requires the ability to: (1) identify patients attending specific hospitals, (2) characterise these patients with respect to their physical and mental health, (3) determine the interaction of these patients with primary and secondary care services and (4) identify outcomes for these patients beyond the immediate spell in hospital.

We determined to conduct a large-scale study based on the analysis of routinely collected NHS data. The

intended study was part of the liaison psychiatry-measurement and evaluation of service types, referral patterns and outcomes (LP-MAESTRO)[15] programme of research funded by the National Institute of Health and Care Research to evaluate the cost-effectiveness and efficiency of LMHS in acute hospital settings. Other parts of the programme which were successfully completed included mapping and characterising liaison services,[14] staff and patient experience,[16,17] processes involved in commissioning of liaison services,[18] developing programme theory to understand how services achieve their intended outcomes[19] and a robust re-evaluation of the effectiveness of a liaison service previously reported to demonstrate cost-effectiveness.[10] The aim of the present study was to evaluate the impact of LMHS on the outcomes of patients in acute hospital settings.

Data that are relevant to the evaluation of health services are routinely collected by healthcare professionals. However, the collection of such data is not centralised and data are collected independently by the relevant different organisations and only those variables required by each specific service are collected. Consequently, data collected by different NHS organisations vary in their coverage of patient populations and their view of the health and health care of these populations. Moreover, organisations are subject to variation in the format and quality of data collected, and the processes by which data may be provided for purposes beyond care, including research. Several organisations in the UK collate standardised subsets of the data collected by individual organisations and provide a single point of access to these data for research purposes. However, such databases are themselves subject to limitations in their coverage of patient populations and their view of health and health care. We determined that no single organisation could provide the data required to evaluate the impact of LMHS and that linkage of data from different organisations was required.

This paper describes the challenges we faced in being able to link primary and secondary care data. Since we began this project other researchers have reported severe difficulties in data linkage and have described the system as 'incredibly complex, arduous and slow, stifling innovation and delaying scientific progress'.[20] Our aim is to describe the barriers we experienced in being able to access the data we required to conduct our analyses in a timely fashion. There are important lessons to be learnt from our experience which we will detail below.

Aims and objectives

Aim
To use routinely collected NHS data to construct care pathways for the main target populations of LMHS, and to estimate outcomes and costs associated with care.

Objectives
To compare the outcomes and costs for patients admitted to hospital and seen by a LMHS compared with:

1. a patient group that was admitted to the same hospital in the same study period but not seen by a LMHS
2. a patient group that was admitted to a different hospital without a LMHS in the same study period.

It was planned to match all three patient groups on patients' characteristics determined from primary and secondary care.

Methods

Study design
The study was a retrospective cohort study. For each hospital with a liaison service, patient groups admitted to hospital and seen by the hospital LMHS would be compared with a patient group at the same hospital not seen by the liaison service and another patient group at a separate hospital with no LMHS during the study period. We intended to match all patient groups on characteristics determined from primary and secondary care data and compare outcome for certain marker conditions which are commonly assessed by liaison services (such as physical–mental comorbidity, acute behavioural disturbance, cognitive impairment/dementia) in different LMHS and configurations. Cost of care pathways for patients referred to LMHS and the matched comparison patient groups and the main determinants of those costs over 12 months after an index hospital episode would then be estimated. All analyses of the linked data would be examined by the research team at the University of Leeds. The published protocol describes the rationale for the design and the planned analyses in more detail.[21]

Data sources
Routinely collected NHS data from primary and secondary care were required to construct care pathways. Secondary care data were required to determine those patients attending specific hospitals and to determine whether the patient was seen by a LMHS (if applicable). Primary and secondary care data were required to characterise these

patients with respect to their physical and mental health, determine their interaction with primary and secondary care services and identify their outcomes beyond the immediate spell in hospital.

Hospital Episode Statistics (HES)[22] is a database controlled by NHS Digital[23] that contains data that are routinely collected in secondary care and relate to emergency department, inpatient and outpatient episodes for patients at hospitals in England. Data within HES can be used to determine those patients attending specific hospitals, to characterise those patients with respect to certain aspects of their physical and mental health, to determine their interactions with secondary care services and to determine certain outcomes. Episodes represent discrete periods of care under a particular consultant and a spell is composed of one or more episodes – for example, when a patient is transferred from a medical to a surgical team for treatment or from an intensive care unit to a lower-intensity medical ward.

ResearchOne (RO)[24] is a database controlled by The Phoenix Partnership (TPP) that contains data that are routinely collected in primary care provided by organisations using the SystmOne clinical information system.[25] SystmOne (34% coverage) and Egton Medical Information Systems (EMIS)[26] (56% coverage) are the most prevalent clinical information systems used by organisations in general practice.[27] No specific information was available a priori regarding the intersection of patient populations between HES and RO. Based on the prevalence of SystmOne use by organisations in primary care, we expected that a sufficient number of patients with episodes in the HES data for the specific hospitals to be studied would have data relating to their primary care collected by organisations that use SystmOne and contained within the RO database. We chose not to use the clinical practice research datalink (CPRD)[28] which is a database controlled by the Medicines and Healthcare Regulatory Authority[29] that contains data that are routinely collected in primary care due to the limited number of organisations whose primary care data were included in CPRD at the time of research design, and the geographical distribution of these organisations.

We divided the study into two phases. In the first phase, we planned to use the data from HES and RO only, to construct care pathways for patients attending hospitals without a LMHS (i.e. to provide data for one of the control groups). This paper focuses on this process, which became so challenging that we were unable to progress the study further.

Population

Patients were included in the study population if they had an accident and emergency (A&E) attendance or inpatient admission to 1 of 11 hospitals without a mental health liaison service between 1 April 2013 and 31 March 2014 (Index Period) for which they were aged 18 or over. At the time we started the study, none of these 11 hospitals had a LMHS, although all now have a service due to the recent expansion in liaison services in England. Patients were excluded from the study population if they had registered a Type 2 objection[30] with NHS Digital to prevent their identifiable data from any health and social care settings being released. NHS Digital selected the study population by applying this inclusion/exclusion criteria to emergency department attendances and inpatient admissions recorded as episodes within HES.

Data items

Data items to be included for each patient in the study population were those determined by the research team to be necessary and sufficient for analysis. No patient-identifiable data were included in the data required by the research team.

For patients included in the study population, the following data items were selected for inclusion from HES for the period from 1 April 2012 to 31 March 2015 (Index Period ± 1 year):

- Episodes from HES (A&E). For each episode, data items relating to the following were included: (1) attendance; (2) diagnoses; (3) investigations; (4) treatments; (5) socioeconomic status; (6) provider; and (7) demographics.
- Episodes from HES (Admitted Patient Care). For each episode, data items relating to the following were included: (1) admission; (2) critical care; (3) cause; (4) diagnoses; (5) procedures; (6) specialty; (7) discharge; (8) spell; (9) provider; (10) demographics; (11) legal status of patient; (12) psychiatric care and status; and (13) socioeconomic status.
- Episodes from HES (Outpatient). For each episode, data items relating to the following were included: (1) appointment; (2) referral; (3) diagnoses; (4) procedures; (5) specialty; (6) provider; (7) demographics; and (8) socioeconomic status.

For patients included in the study population whose data relating to primary care are included in RO specific data items relating to the following were selected for inclusion from RO for the period prior to 31 March 2015:

- Demographics and socioeconomic status.
- Selected primary care events including: (1) coded diagnoses/observations that relate to selected

comorbidities; (2) prescriptions and repeated prescriptions for selected drugs; (3) non-coded referrals; (4) appointments; and (5) practice registrations.

Linkage

Linkage was required to enable data items relating to the same patient in HES and RO to be determined. No persistent link exists between HES and RO and, to our knowledge, no linkage methodology had been previously enacted between the two databases. We worked with NHS Digital and TPP to design a methodology that would: (1) be able to link data from HES and RO for the study, (2) be acceptable to the data controllers from a technical and organisational perspective and (3) be consistent with legal and ethical frameworks applicable to the use of routinely collected NHS data for research purposes (in the absence of individual patient consent).

Patients are referenced by source-specific identifiers in HES and RO. Different identifiers are used to reference the same patient in HES and RO. No persistent mapping exists or indeed is feasible ex post between these source-specific identifiers. Data from HES and RO are therefore completely disjoint based on these source-specific identifiers alone. Linkage requires a mapping to be established between these source-specific identifiers. NHS Digital and TPP require access to at least one common data item for each patient that can be used to establish this mapping. Moreover, any data item used for this purpose must demonstrate sufficient quality (e.g. completeness) to ensure the integrity of the resultant linkage. To protect the privacy of patients, pseudonyms are generated from NHS numbers by the data controllers and these pseudonyms are used to generate a mapping between source-specific identifiers.

Figure 1 shows a summary of the key data flows and processing activities of the methodology. Linkage is undertaken by NHS Digital, in accordance with recommendations provided in the Caldicott Review: information governance in the health and acre system document published by the National Data Guardian.[31] NHS Digital and TPP generate two unique references for each patient in their databases: (1) a pseudonym, generated by applying a one-way cryptographic hash function (SHA-512) to an input that comprises a cryptographic salt and the NHS number and (2) a source-specific identifier. For a patient with a given NHS number, NHS Digital and TPP will generate the same pseudonym but a different source-specific identifier. Both the pseudonym and source-specific identifier generated for each patient are specific to the study. Pseudonyms are then used by NHS Digital to: (1) communicate to TPP those patients for whom data

are required from RO and (2) generate mappings between different source-specific identifiers for each patient. NHS Digital and TPP provide the required data items to the research team at the University of Leeds, including only the source-specific identifier as the unique reference for each patient. Mappings generated by NHS Digital are then provided to the research team and used to enable data items relating to the same patient across databases to be determined.

Infrastructure

Infrastructure provided by the Leeds Institute of Clinical Trials Research (LICTR) at the University of Leeds was used for the study. Information security standards defined in the Data Security and Protection Toolkit[32] were met by an infrastructure and the Data-Sharing Framework Contract between the University of Leeds and NHS Digital (REF: CON-315426-K3W7R) including the use of this infrastructure.

Agreed project-specific terms for processing activities involving NHS Digital were defined in a Data-Sharing Agreement established between the University of Leeds and NHS Digital. Agreed project-specific terms for data flows and processing activities involving TPP were defined in the RO Data Request Form provided to TPP by the University of Leeds and approved by the RO Project Committee, and in a Data Processing Agreement established between the University of Leeds and TPP.

Data flows between organisations were performed using one of the following secure data transfer services:

- Secure file transfer (SFT): Provided by LICTR and used for data flows between University of Leeds and TPP.
- Secure electronic file transfer (SEFT) service:[33] Provided by NHS Digital and used for data flows between University of Leeds and NHS Digital, and between TPP and NHS Digital.

Approvals

Approval was required from Research Governance at the Sponsor (University of Leeds). A favourable ethical opinion was sought from an NHS Research Ethics Committee (REC) and the Health Research Authority (HRA)[34] to ensure consistency with applicable ethical frameworks for medical research, including the Declaration of Helsinki.[35] Additionally, an application was made to the Confidentiality Advisory Group (CAG)[36] at the HRA to obtain (if applicable) support under Section 251 of the NHS Act 2006.[37] Figure 2 shows a summary of the envisaged progression of approval processes.
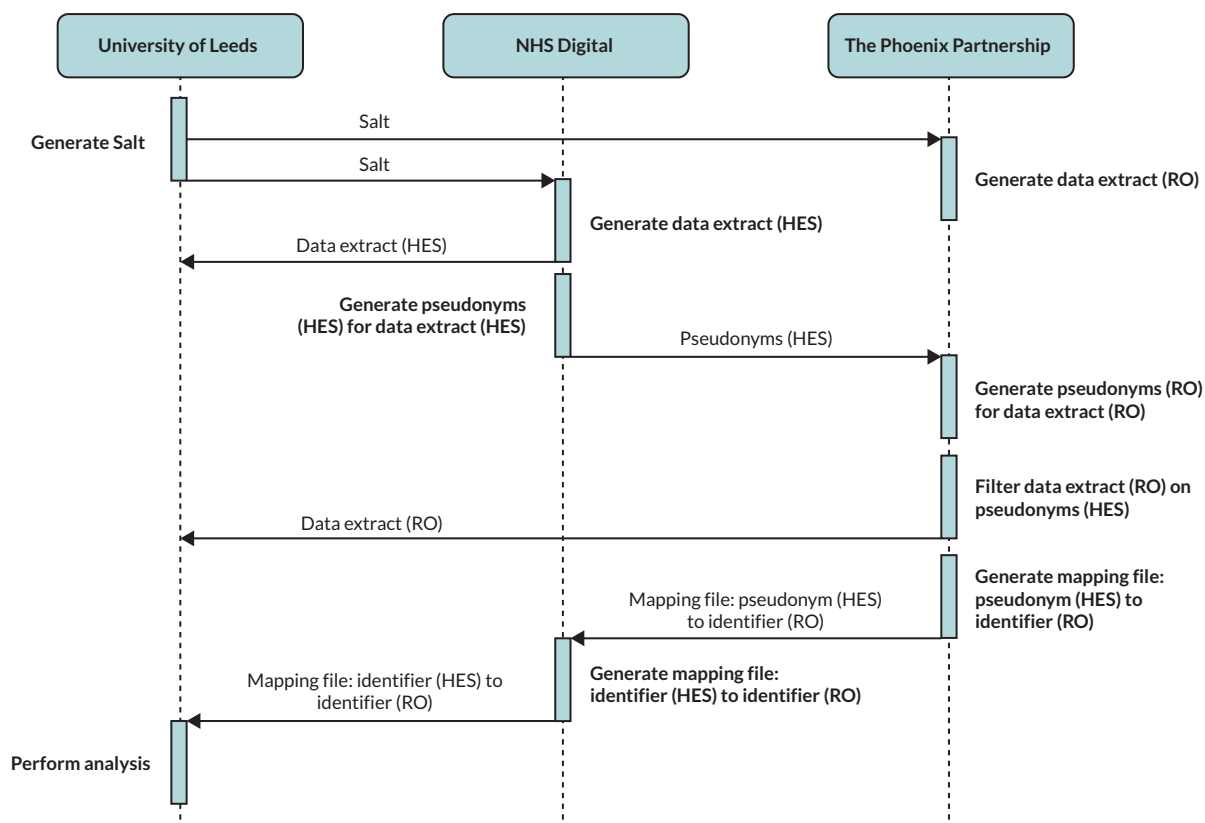
FIGURE 1 A summary of the key data flows and processing activities of the methodology.
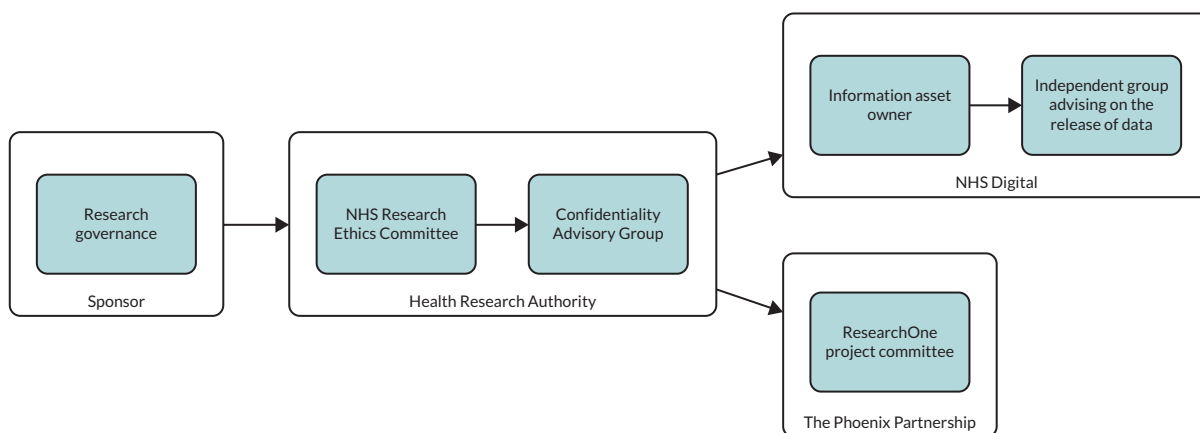
**FIGURE 2** A summary of the envisaged progression of approval processes.

Results

Approvals

The study received: (1) approval from Research Governance at the Sponsor, (2) a favourable ethical opinion from NHS REC (North of Scotland) (REF: 16/NS/0025), (3) management permission from a subset of NHS Trusts which manage the hospitals whose patients were to be included in the study population, (4) a decision from the CAG at the HRA that Section 251 support was not required 'on the basis that there is no disclosure of patient identifiable data without consent' (REF: 16/CAG/0037), (5) approval from the Independent Group Advising on the Release of Data (IGARD)[38] at NHS Digital, (6) approval from the Information Asset Owner for HES at NHS Digital and (7) approval from RO Project Committee at TPP. Approvals took 907 days in total. *Figure 3* summarises the progression of the approval process.

Favourable opinion from NHS REC (North of Scotland) was contingent on management permission from the NHS Trusts (*n* = 8) which manage the hospitals (*n* = 11) whose
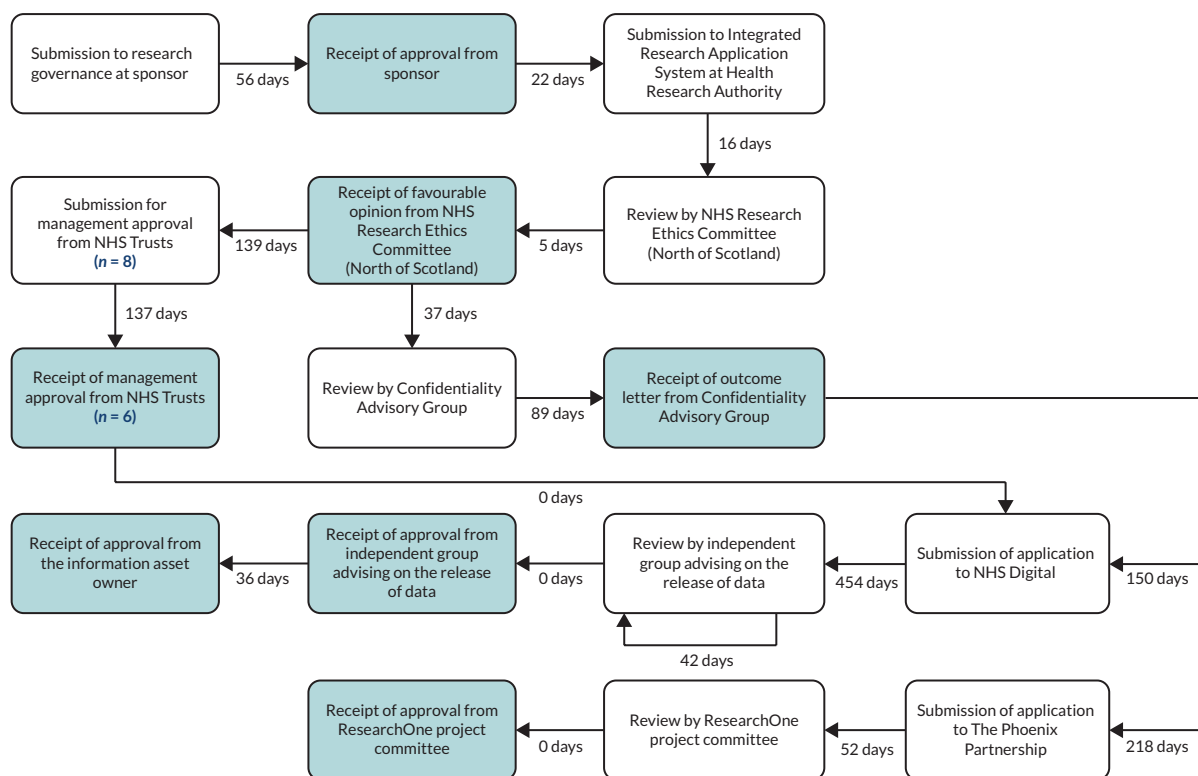
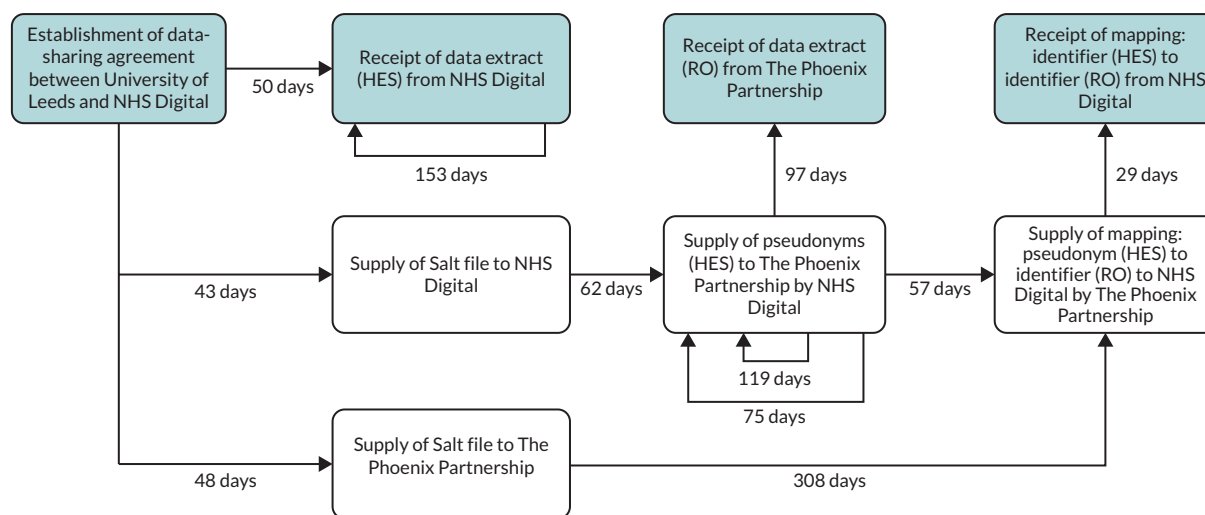**FIGURE 3** A summary of the progression of the approval processes and the associated time periods.

patients were to be included in the study population. Requests for management permission were submitted to the NHS Trusts following receipt of the outcome from the CAG. We started by trying to obtain linked data from NHS Trusts with no identifiable LMHS during the study period, with the intention of then widening the data collection to NHS Trusts with liaison services. This part of the study, therefore, proceeded based on a subset of the hospitals only (*n* = 8) – that is, hospitals from which one group of control subjects could potentially be identified.

Following application to TPP, 52 days elapsed before approval by the RO Project Committee. TPP supplied an approval date of April 2017. An indicative date of 15 April 2017 was used to calculate the period of 52 days. Review and approval by the ResearchOne Project Committee were assumed to take place on the same date. Following submission of application to NHS Digital for data from HES [REF: NIC-77953], 454 days elapsed before the application was reviewed by IGARD. Delays were attributable to the resolution of technical and organisational queries raised by NHS Digital in their consideration of the application, including: (1) the process of pseudonym generation, (2) the data flows involved in the linkage methodology, (3) the organisation with legal responsibility for RO, (4) the data controller and data processor responsibilities for organisations involved in the linkage methodology and (5) the data storage locations for University of Leeds and

TPP. NHS Digital noted that the linkage methodology was a 'first of type' and took the application to IGARD for advice prior to submission for review by IGARD. Following review by IGARD, further information was requested from the project team, which was subsequently supplied and a recommendation to approve the application was provided in a subsequent IGARD meeting. A Data-Sharing Agreement was established between the University of Leeds and NHS Digital, 118 days after approvals – attributable to the University of Leeds awaiting required information from TPP.

Enactment

Figure 4 shows a summary of the progression enactment processes and associated time periods. Following establishment of a Data-Sharing Agreement between the University of Leeds and NHS Digital, the methodology progressed to enactment in which the defined processing activities were undertaken by the University of Leeds, NHS Digital and TPP. NHS Digital generated pseudonyms for patients in the cohort using a Salt file generated and supplied by the University of Leeds. Incorrect pseudonyms were generated by NHS Digital and supplied to TPP on two occasions, which resulted in no matches being found within RO. On the first occasion, NHS Digital applied an incorrect hashing algorithm (SHA-256). On the second occasion, NHS Digital used an incorrect format for the input to the hashing algorithm. NHS Digital determined that the HES data supplied to the University of Leeds

FIGURE 4 A summary of the progression of the enactment processes and the associated time periods.

had to be resupplied following the first occasion, but not following the second occasion. Following the second occasion, NHS Digital regenerated the pseudonyms using the correct hashing algorithm and values in the correct input format and resupplied to TPP. Receipt of a valid set of pseudonyms from NHS Digital by TPP required 313 days.

Following supply of a valid set of pseudonyms, TPP were able to match pseudonyms with those generated for patients in RO. TPP generated and supplied a file to NHS Digital which mapped these pseudonyms to corresponding source-specific identifiers from RO. NHS Digital used the file to generate a file which mapped source-specific identifier in HES to source-specific identifier in RO. This mapping file was supplied by NHS Digital to the University of Leeds. TPP then supplied the data extract from RO to University of Leeds. Enactment took 385 days in total.

Table 1 summarises the data supplied from HES and RO. Data supplied from HES contained 181,063 patients from six hospitals (mean = 30,177, standard deviation = 28,875.86) who matched the required criteria (see Population). Data supplied from RO contained 33,666 (18.6%) of these patients from the 6 hospitals (mean = 5611, standard deviation = 5206.59). For two hospitals (G and H), no patients who matched the required criteria could be determined.

In response to a follow-up query about low rates of case identification, NHS Digital determined the probable cause for this result to be the use of the Provider Code (5 character) (PROCODE5)[39,40] to determine the hospital associated with an episode. NHS Digital advised that the use of both the PROCODE5 field and the Site Code

of Treatment (SITETRET) field would likely yield a higher number of patients for each hospital, including those for which no patients with an index event were determined. From September 2021 NHS Digital has made changes to the way they process the HES data set, which means the data referenced are no longer available online.[41]

Due to the impact on project timescales, we determined that this change to the data specification would not be feasible. After considerable discussion with National Institute for Health and Care Research and a further delay because of the COVID-19 pandemic, it was decided not to proceed with the analysis of the data we held from the six hospital sites. Such an analysis would have been proof that the data linkage we were attempting could be achieved, but its value in terms of the overall outcomes we aimed to achieve was limited. We therefore proceeded to destroy the data we had obtained, and a certificate of destruction was signed by the University of Leeds investigators and data management staff on 15 July 2021.

Discussion
Our objective was to design and enact a methodology to evaluate hospital-based liaison psychiatry services in England, by linking routinely collected NHS data from secondary care settings, chosen for their service characteristics, to data from primary care. To achieve this objective, we designed and enacted a methodology to link routinely collected NHS data from HES and RO. No persistent link exists between these two databases, and, to our knowledge, no linkage methodology had been previously enacted between the two databases. Linkage of these databases using the described methodology represented at the time a 'first of type' linkage of routinely collected NHS data from primary and secondary care in England.

TABLE 1 A summary of the number of patients who matched the required criteria from HES and RO

	HES	RO	
	n	*n*	%
Number of patients	181,063	33,666	18.6
Hospital of index event			
Hospital A	89,755	13,202	14.7
Hospital B	31,312	7057	22.5
Hospital C	30,154	11,069	36.5
Hospital D	21,181	2219	10.5
Hospital E	8517	101	1.2
Hospital F	144	18	12.5
Hospital G	0	0	0
Hospital H	0	0	0

Linkage of routinely collected NHS data has the potential to extend the patient populations, time periods and dimensions of health and health care that can be considered within research studies. Linkage of data from HES and RO in this study provides a view across primary and secondary care for a subset of patients defined by their attendance at, or admission to, a specific set of hospitals in England within a defined time period. Linkage of data from HES and RO and appropriate use of similar linkage methodologies between other databases has the potential to move research studies beyond those research questions that can be answered using data from the small number of databases between which persistent linkages already exist, for example, between CPRD and HES.

Linkage of routinely collected NHS data using the methodology described maintains the privacy of patients, the protection of their personal data and the confidentiality of their information. No patient-identifiable data are communicated between organisations or received by the research team and linkage is undertaken by NHS Digital (in accordance with national guidance). Linkage is also transient and purpose-specific,[42] enabling the benefits and risks of linkage and subsequent analysis, along with the ethical and legal implications, to be appropriately scrutinised by the relevant bodies within the context of a specific project. The recent Goldacre Review[43] argued that the maintenance of patient trust was essential if existing data were going to be used to improve health care, and our approach is fully in keeping with both the spirit and the practical implications of the Review's arguments.

Significant difficulties were however encountered in the approvals and enactment of the methodology. Co-ordination between multiple organisations at a technical and a governance level was a challenge. Different organisations and organisational units required information to be presented at different levels of granularity, with different emphases, and in adherence to different presentational formats. Timely provision of this information was also contingent on successful traversal of complex organisational structures and internal processes over which we had no control. Moreover, the governance processes that drive decision-making within organisations are subject to change in both definition and interpretation over time. Delays in provision of information have the potential to cause further delays. Moreover, technical issues can arise when processing activities must be performed by different organisations in a specific manner, as demonstrated by the issues with pseudonym generation at NHS Digital. Absence of published information relating to the number of patients that can be expected from each database at the unit of analysis (hospital) and relating to the quality of fields on which cohort selection is determined means that the feasibility of proposed analyses may only be determinable once data have been received. Given the resources required by the approvals and enactment, such uncertainty currently represents a significant risk for research projects.

Lessons learnt

We are not the only group of researchers to encounter significant delays and difficulties with data access and linkage. Taylor and colleagues (2021)[20] reported it taking them 2.5 years to be able to link a series of databases

to study outcomes for congenital heart disease. To link 5 relevant data sets, they were required to produce 47 documents at the application stage for 11 data controllers or departments which were submitted 162 times in total. Although similar study information was requested for each application, each required different wording, structure and detail.

The biggest barriers to realising the full potential of data linkage as a powerful research tool are gaining and maintaining public trust while simultaneously reducing the cost, delays and inefficiencies in how privacy-preserving data linkage is carried out.[44]The necessities of linking primary and secondary data during the COVID-19 pandemic to facilitate research and rapid transfer of knowledge led to the development of such initiatives as OpenSAFELY[45] and the British Heart Foundation's Cardio Vascular Disease-Coronovirus Disease-united Kingdom (CVD-COVID-UK) consortium.[46] These initiatives show what is possible in principle, and in the distinctively permissive legal context of the pandemic, but they are not easily generalised to enable wider adoption by researchers.

On reflection, we dramatically underestimated the degree of complexity in obtaining permissions to link databases and for the linkage work to be done. At the time of the study, linking primary and secondary care data in the manner we envisioned was a 'first of type' linkage and required the various organisations involved to undertake work they had previously not carried out. OpenSAFELY[41] was started in response to the need to obtain rapid data at the time of COVID-19 and required a change in law (effective 18 March 2020); so it was not an option for the LP-MAESTRO study. The approach of OpenSAFELY, where data linkage is carried out in-house, would have undoubtedly reduced the complexity of the approval processes for our programme, but the legal basis for accessing and linking data in OpenSAFELY, despite researcher pressure, has not changed, and it is only permitted for COVID-19 purposes [https://digital.nhs.uk/coronavirus/coronavirus-covid-19-response-information-governance-hub/the-nhs-england-opensafely-covid-19-service-privacy-notice#top (20 July 2023)]. Although the advent of OpenSAFELY has illustrated that in the right legal as well as technical environment a basic primary/secondary data linkage is readily achievable, studying more complex care pathways, such as those involving both emergency care and mental health secondary care services, could still have been a problem, as our original plans required access to specific hospital databases which would have required bespoke linkages. At the time of our study, mental health data recorded via HES were also limited in their scope, and much of the work of liaison services were not captured nationally.

Implications for research

Accessible and functional pathways are required to enable projects to adopt research methodologies based on the linkage of routinely collected NHS data that aim to protect the privacy of patients, their personal data and the confidentiality of their information. In the absence of such pathways, there is a risk that such methodologies will not be adopted. This may lead to a detrimental impact on the rights of patients, the quality of research designs and processes, and the ability of researchers to meet policy requirements for research that better reflects populations served. Improved knowledge and skills are required to develop more effective and efficient processes and structures to support research governance across all stakeholder organisations, including academic institutions, data providers and approval/advisory bodies.

Researchers who are planning to apply for funding for a data linkage study need to consider the following when developing their application with reference to: (a) the research team; (b) research governance (includes sponsor, ethics, HRA); (c) the funder; and (d) the data controllers. Several of these issues, including complexity, process delays, bureaucracy, time taken to acquire data post approval, were highlighted by Taylor and colleagues[20] but were also experienced by us.

a.   Research team

Staff morale – what seem to be insurmountable delays results in low staff morale, as there is inevitable pressure to meet funding deadlines and also a desire to conduct the research which has been planned.

Career progression and institutional reputational damage – failure to complete or even start a planned programme of research due to delays beyond the control of the research team impacts upon the career progression of middle-grade researchers who are employed on the project who are not able to publish as widely or influentially as they would have planned. The Sponsor may also be at risk of reputation damage.

b.   Research governance

Complexity – data request forms vary across different data sets, some of which are more suitable for requests for data in relation to randomised controlled trials, rather than large data projects. Streamlining and standardising data request forms across different organisations, while maintaining robust data governance procedures commensurate with the Data Protection Act 2018, would reduce time delays.

Process delays – we encountered numerous complex process delays which have major cost implications for any research study.

Delays lead to further delays – a delay in one part of the acquisition process leads to delays in other parts of the system.

Bureaucracy – each time research staff working on the project are recruited, some data-sharing agreements have to be updated. Minor changes can take weeks to be processed.

Fear of legal misstep – failure to comply with the Data Protection Act 2018 and potentially incur a costly fine mean that some data controllers require excessive safeguards before agreeing to release data when sufficient safeguards are in place.

c.   Funder

Budget to access the data sets – it is difficult to anticipate the final costs of access to a database, as limited information is usually available online prior to completion of a full application which cannot be started until funding has been received for the research. Although applications are for the full length of a project, annual data-sharing extensions are required which involve additional costs. Taylor and colleagues[20] reported that in the case of NHS Digital, even one additional field request could result in a charge equal to that of the original data extraction.

Most grant applications are for periods of 3 years, during which it is not feasible at present to conduct complete data linkage processes and carry out planned data analyses. Underestimation of the time involved in acquiring linked data also results in costs for the research programme, as researchers employed to conduct data analyses cannot carry out this work until the data are securely acquired.

d.   Data controllers

Acquiring data – once all approvals are in place, it can still take at least 6 months to acquire the data.

Difficult-to-spot small mistakes in the data request process that have major implications for the data which are supplied – two out of the eight hospitals for which we requested data were supplied with empty data fields, due to a misunderstanding about the different ways and data fields which HES uses to record hospital site.

Patient and public involvement

Patient and public involvement (PPI) representatives were involved in the design of the project and in the delivery of all other components of the overall programme, all of which were completed within the original programme timeline

are available in a series of peer-reviewed publications (see *Publications*). The delays we encountered in delivering this strand of work meant that maintaining meaningful PPI was not possible.

Conclusions

We conclude that routinely collected NHS data from primary care and secondary care available in RO and HES can be linked using the methodology described. However, the time required for approvals and enactment, the variation exhibited in the supplied data and the unpredictability of these factors during research design currently limit the viability of such linkage for timely and robust evaluation of health services. Current processes need to be streamlined and standardised with designated clear response times for the different organisations.

Additional information

Contributions of authors
Elspeth Guthrie (https://orcid.org/0000-0002-5834-6616) (Current Principal Investigator) revised the manuscript.

Allan House (https://orcid.org/0000-0001-8721-8026) (Former Principal Investigator) designed this part of the research programme and gave input on the manuscript.

Chris Smith (https://orcid.org/0000-0002-0184-3590) (Co-applicant) designed this part of the research programme, carried out the attempted linkage work, drafted the manuscript and developed and prepared the figures. He left the University of Leeds in August 2021 and did not make any further contributions to the manuscript.

Sam Relton (https://orcid.org/0000-0003-0634-4587) (Associate Professor) discussed the implications of the findings and provided comments and critical revision of the final manuscript.

Daniel Romeu (https://orcid.org/0000-0002-2417-0202) (Academic Clinical Fellow) discussed the implications of the findings and provided comments and critical revision of the final manuscript.

Sonia Saraiva (https://orcid.org/0000-0002-2305-9246) (Senior Research Fellow) discussed the implications of the findings and provided comments and critical revision of the final manuscript.

Peter Trigwell (Co-applicant) discussed the implications of the findings and provided comments and critical revision of the final manuscript.

Robert West (https//orcid.org/0000-0001-7305-3654) (Co-applicant) designed this part of the research programme.

Farag Shuweihdi (https//orcid.org/0000-0003-1199-2992) (Data Analyst) discussed the implications of the findings and provided comments and critical revision of the final manuscript.

Mike Crawford (https://orcid.org/0000-0003-3137-5772) (Co-applicant) discussed the implications of the findings and provided comments and critical revision of the final manuscript.

Matt Fossey (https://orcid.org/0000-0002-9725-7464) (Co-applicant) discussed the implications of the findings and provided comments and critical revision of the final manuscript.

Jenny Hewison (https://orcid.org/0000-0003-3026-3250) (Co-applicant) designed this part of the research programme and gave input on the manuscript.

Claire Hulme (https://orcid.org/0000-0003-2077-0419) (Co-applicant) designed this part of the research programme.

Sandy Tubeuf (https://orcid.org/0000-0001-9001-1157) (Health Economist) discussed the implications of the findings and provided comments and critical revision of the final manuscript.

Disclosure of interests

Data-sharing statement

No data are accessible for this project. Any queries should be addressed to the corresponding author.

Ethics statement

Individual patient consent was not feasible for this study. Favourable ethical opinion has been obtained from the NHS Research Ethics Committee (North of Scotland) (REF: 16/NS/0025) for Work Stream 2 (phase 1) of the LP-MAESTRO.

Information governance statement

The Confidentiality Advisory Group at the Health Research Authority determined that Section 251 approval under Regulation 5 of the Health Service (Control of Patient Information) Regulations 2002 was not required for the study 'on the basis that there is no disclosure of patient identifiable data without consent' (REF: 16/CAG/0037).

Department of Health and Social Care disclaimer

Publications

Walker A, Barrett JR, Lee W, West RM, Guthrie E, Trigwell P, *et al.* Organisation and delivery of liaison psychiatry services in general hospitals in England: results of a national survey. *BMJ Open* 2018;**8**:e023091. https://doi.org/10.1136/bmjopen-2018-023091

House A, Guthrie E, Walker A, Hewison J, Trigwell P, Brennan C, *et al.* A programme theory for liaison mental health services in England. *BMC Health Serv Res* 2018;**18**:742. https://doi.org/10.1186/s12913-018-3539-2

Smith C, Hewison J, West RM, Guthrie E, Trigwell P, Crawford MJ, *et al.* Liaison psychiatry – measurement and evaluation of service types, referral patterns and outcomes (LP-MAESTRO): a protocol. *BMJ Open* 2019;**9**:e032179. https://doi.org/10.1136/bmjopen-2019-032179

Jasmin K, Walker A, Guthrie E, Trigwell P, Quirk A, Hewison J, *et al.* Integrated liaison psychiatry services in England: a qualitative study of the views of liaison practitioners and acute hospital staffs from four distinctly different kinds of liaison service. *BMC Health Serv Res* 2019;**19**:522. https://doi.org/10.1186/s12913-019-4356-y

Saraiva S, Guthrie E, Walker A, Trigwell P, West R, Shuweidi F, *et al.* The nature and activity of liaison mental services in acute hospital settings: a multi-site cross sectional study. *BMC Health Serv Res* 2020;**20**:308. https://doi.org/10.1186/s12913-020-05165-x

Fossey M, Godier-McBard L, Guthrie EA, Hewison J, Trigwell J, Smith CJ, House A. Understanding liaison psychiatry commissioning: an observational study. *Ment Health Rev J* 2020;**25**:301–16. https://doi.org/10.1108/MHRJ-03-2020-0016

House A, West R, Smith C, Tubeuf S, Guthrie E, Trigwell P. The effect of a hospital liaison psychiatry service on inpatient lengths of stay: interrupted time series analysis using routinely collected NHS Hospital Episode Statistics. *BMC Psychiatry* 2020;**20**:27. https://doi.org/10.1186/s12888-020-2441-8

Guthrie E, Romeu D, Czoski-Murray C, Relton S, Walker A, Trigwell P, *et al.* Experiences of people seen in an acute hospital setting by a liaison mental health service: responses from an online survey. *BMC Health Serv Res* 2021;**21**:1050. https://doi.org/10.1186/s12913-021-06974-4

Submitted for publication

Romeu D, Guthrie E, House A. Staff experiences of liaison mental health services following implementation of CORE 24 informed by service user preferences.

Funding

This article reports on one component of the research award Linkage of routinely collected NHS data to evaluate liaison mental health services: challenges and lessons learnt. For more information about this research please view the award page (https://www.fundingawards.nihr.ac.uk/award/13/58/08).

About this article

The contractual start date for this research was in January 2015. This article began editorial review in February 2023 and was accepted for publication in January 2024. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The Health and Social Care Delivery Research editors and publisher have tried to ensure the accuracy of the authors' article and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this article.

Copyright

List of abbreviations

A&E	accident and emergency
CAG	Confidentiality Advisory Group
CPRD	clinical practice research datalink
HES	Hospital Episode Statistics
HRA	Health Research Authority
IGARD	Independent Group Advising on the Release of Data
LICTR	Leeds Institute of Clinical Trials Research
LMHS	liaison mental health services
LP-MAESTRO	liaison psychiatry-measurement and evaluation of service types, referral patterns and outcomes
PPI	patient and public involvement
PROCODE5	Provider Code (5 character)
REC	Research Ethics Committee

| RO | ResearchOne |
| TPP | The Phoenix Partnership |

References

1. Aitken P, Lloyd G, Mayou R, Bass C, Sharpe M. A history of liaison psychiatry in the UK. *BJPsych Bull* 2016;**40**:199–203. https://doi.org/10.1192/pb.bp.116.053728

2. Lipowski ZJ. Consultation-liaison psychiatry: the first half century. *Gen Hosp Psychiatry* 1986;**8**:305–15. https://doi.org/10.1016/0163-8343(86)90046-0

3. Mayou R. The history of General Hospital psychiatry. *Br J Psychiatry* 1989;**155**:764–76. https://doi.org/10.1192/bjp.155.6.764

4. Schwab JJ. Consultation-liaison psychiatry: a historical overview. *Psychosomatics* 1989;**30**:245–54. https://doi.org/10.1016/S0033-3182(89)72268-4

5. NHS England. *Report of the Fourth Survey of Liaison Psychiatry in England*. 2019. URL: www.england.nhs.uk/publication/report-of-the-4th-survey-of-liaison-psychiatry-in-england/ (accessed 19 October 2023).

6. Tadros G, Kingston P, Mustafa N, Johnson E, Pannell R, Hashmi M. Impact of an integrated rapid response psychiatric liaison team on quality improvement and cost savings: the Birmingham RAID model. *The Psychiatrist* 2013;**37**:4–10. https://doi.org/10.1192/pb.bp.111.037366

7. Breckon JKS, McClimnes A, Mubarak I, Burley K. *Adult Mental Health Hospital Liaison Service Evaluation*. Sheffield, UK: Sheffield Hallam University; 2016. URL: http://shura.shu.ac.uk/16729/3/Kelly%20Adult%20Mental%20Health%20Hospital%20Liaison%20Service%20Evaluation.pdf (accessed 19 October 2023).

8. Opmeer BC, Hollingworth W, Marques EMR, Margelyte R, Gunnell D. Extending the liaison psychiatry service in a large hospital in the UK: a before and after evaluation of the economic impact and patient care following ED attendances for self-harm. *BMJ Open* 2017;**7**:e016906. https://doi.org/10.1136/bmjopen-2017-016906

9. Becker L, Saunders R, Hardy R, Pilling S. *The RAID Model of Liaison Psychiatry: Report on the Evaluation of Four Pilot Services in East London*. London: UCL Partners Academic Health Science Partnership; 2016. URL: www.ucl.ac.uk/pals/sites/pals/files/uclp_raid_evaluation_report.pdf (accessed 19 October 2023).

10. House A, West R, Smith C, Tubeuf S, Guthrie E, Trigwell P. The effect of a hospital liaison psychiatry service on inpatient lengths of stay: interrupted time series analysis using routinely collected NHS Hospital Episode Statistics. *BMC Psychiatry* 2020;**20**:27. https://doi.org/10.1186/s12888-020-2441-8

11. Toynbee M, Walker J, Clay F, Hollands L, van Niekerk M, Harriss E, Sharpe M. The effectiveness of inpatient consultation-liaison psychiatry service models: a systematic review of randomized trials. *Gen Hosp Psychiatry* 2021;**71**:11–9. https://doi.org/10.1016/j.genhosppsych.2021.04.003

12. Walker J, Burke K, Toynbee M, van Niekerk M, Frost C, Magill N, *et al*. The HOME Study: study protocol for a randomised controlled trial comparing the addition of Proactive Psychological Medicine to usual care, with usual care alone, on the time spent in hospital by older acute hospital inpatients. *Trials* 2019;**20**:483. https://doi.org/10.1186/s13063-019-3502-5

13. Saraiva S, Guthrie E, Walker A, Trigwell P, West R, Shuweihdi F, *et al*. The nature and activity of liaison mental services in acute hospital settings: a multi-site cross sectional study. *BMC Health Serv Res* 2020;**20**:308. https://doi.org/10.1186/s12913-020-05165-x

14. Walker A, Barrett JR, Lee W, West RM, Guthrie E, Trigwell P, *et al*. Organisation and delivery of liaison psychiatry services in general hospitals in England: results of a national survey. *BMJ Open* 2018;**8**:e023091. https://doi.org/10.1136/bmjopen-2018-023091

15. National Institute for Health and Care Research. *Liaison Psychiatry: Measurement and Evaluation of Service Types, Referral Patterns and Outcomes (LP-MAESTRO)*. 2020. URL: https://fundingawards.nihr.ac.uk/award/13/58/08 (accessed 19 October 2023).

16. Guthrie E, Romeu D, Czoski-Murray C, Relton S, Walker A, Trigwell P, *et al*. Experiences of people seen in an acute hospital setting by a liaison mental health service: responses from an online survey. *BMC Health Serv Res* 2021;**21**:1050. https://doi.org/10.1186/s12913-021-06974-4

17. Jasmin K, Walker A, Guthrie E, Trigwell P, Quirk A, Hewison J, *et al*. Integrated liaison psychiatry services in England: a qualitative study of the views of liaison practitioners and acute hospital staffs from four distinctly different kinds of liaison service. *BMC Health Serv Res* 2019;**19**:522. https://doi.org/10.1186/s12913-019-4356-y

18. Fossey M, Godier-McBard L, Guthrie EA, Hewison J, Trigwell J, Smith CJ, House A. Understanding liaison psychiatry commissioning: an observational study. *Ment Health Rev J* 2020;**20**:301–16. https://doi.org/10.1108/MHRJ-03-2020-0016

19. House A, Guthrie E, Walker A, Hewsion J, Trigwell P, Brennan C, *et al*. A programme theory for liaison mental health services in England. *BMC Health Serv Res* 2018;**18**:742. https://doi.org/10.1186/s12913-018-3539-2

20. Taylor JA, Crowe S, Espuny Pujol F, Franklin RC, Feltbower RG, Norman LJ, *et al*. The road to hell is paved with good intentions: the experience of applying for national data for linkage and suggestions for improvement. *BMJ Open* 2021;**11**:e047575. https://doi.org/10.1136/bmjopen-2020-047575

21. Smith C, Hewison J, West RM, Guthrie E, Trigwell P, Crawford MJ, *et al*. Liaison psychiatry – measurement and

evaluation of service types, referral patterns and outcomes (LP-MAESTRO): a protocol. *BMJ Open* 2019;**9**:e032179. https://doi.org/10.1136/bmjopen-2019-032179

22. NHS Digital. *Hospital Episode Statistics (HES)*. 2020. URL: https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics (accessed 19 October 2023).

23. NHS Digital. 2020. URL: https://digital.nhs.uk (accessed 19 October 2023).

24. The Phoenix Partnership. *ResearchOne*. 2023. URL: https://tpp-uk.com/products/data-and-research/ (accessed 19 October 2023).

25. The Phoenix Partnership. *SystmOne*. 2020. URL: https://tpp-uk.com/products/ (accessed 19 October 2023).

26. Egton Medical Information Systems. *EMIS Web*. 2020. URL: www.emishealth.com/products/emis-web (accessed 19 October 2023).

27. Kontopantelis E, Stevens RJ, Helms PJ, Edwards D, Doran T, Ashcroft DM. Spatial distribution of clinical computer systems in primary care in England in 2016 and implications for primary care electronic medical record databases: a cross-sectional population study. *BMJ Open* 2018;**8**:e020738. https://doi.org/10.1136/bmjopen-2017-020738

28. Medicines and Healthcare products Regulatory Agency. *Clinical Practice Research Datalink*. URL: https://cprd.com (accessed 19 October 2023).

29. Medicines and Healthcare products Regulatory Agency. 2021. URL: www.gov.uk/government/organisations/medicines-and-healthcare-products-regulatory-agency (accessed 19 October 2023).

30. Health and Social Care Information Centre. *Patient Objections Management for General Practices in England*. 2020. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/469290/Data-Provision-Notice_Patient_Objections_Management_19.10.15.pdf (accessed 19 October 2023).

31. National Data Guardian. *Caldicott Review: Information Governance in the Health and Social Care System*. 2020. URL: www.gov.uk/government/publications/the-information-governance-review (accessed 19 October 2023).

32. NHS Digital. *Data Security and Protection Toolkit*. 2020. URL: www.dsptoolkit.nhs.uk (accessed 19 October 2023).

33. NHS Digital. *Secure Electronic File Transfer Service*. URL: https://digital.nhs.uk/services/transfer-data-securely (accessed 19 October 2023).

34. Health Research Authority. 2020. URL: www.hra.nhs.uk (accessed 19 October 2023).

35. World Medical Association. *WMA Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Subject*. 2020. URL: www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/ (accessed 19 October 2023).

36. Health Research Authority. *Confidentiality Advisory Group*. 2020. URL: www.hra.nhs.uk/about-us/committees-and-services/confidentiality-advisory-group (accessed 19 October 2023).

37. HM Government. *National Health Service Act*. 2006. URL: www.legislation.gov.uk/ukpga/2006/41/contents (accessed 19 October 2023).

38. NHS Digital. *Independent Group Advising on the Release of Data*. 2023. URL: https://digital.nhs.uk/about-nhs-digital/corporate-information-and-documents/independent-group-advising-on-the-release-of-data (accessed 19 October 2023).

39. NHS Digital. *HES Data Dictionary – Accident and Emergency*. 2021. URL: https://digital.nhs.uk/binaries/content/assets/website-assets/data-and-information/data-toolsand-services/data-services/hospital-episode-statistics/hes-data-dictionary/dd-ae_v12.pdf (accessed 27 January 2021).

40. NHS Digital. *HES Data Dictionary – Admitted Patient Care*. 2021. URL: https://digital.nhs.uk/binaries/content/assets/web-site-assets/data-and-information/data-toolsand-services/data-services/hospital-episode-statistics/hes-data-dictionary/dd-apc_v12.pdf (accessed 27 January 2021).

41. NHS Digital. *Hospital Episode Statistics. Changes to HES Data Dictionary*. 2021. URL: https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics/hospital-episode-statistics-data-dictionary#changes-to-hes-data-dictionary-2021 (accessed 19 October 2023).

42. Smith C, Hewison J, House A. Designing a privacy-preserving protocol to support transient and purpose-specific data linkages. *Int J Popul Data Sci* 2017;**1**:353. https://doi.org/10.23889/ijpds.v1i1.374

43. Goldacre B, Morley J. *Better, Broader, Safer: Using Health Data for Research and Analysis. A Review Commissioned by the Secretary of State for Health and Social Care.* Department of Health and Social Care; 2022. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1067053/goldacre-review-using-health-data-for-research-and-analysis.pdf (accessed 19 October 2023).

44. Harron K. Data linkage in medical research. *BMJ Med* 2022;**1**:e000087. https://doi.org/10.1136/bmjmed-2021-000087

45. Secure analytics platform for NHS electronic health records. *OpenSAFELY*. URL: www.opensafely.org (accessed 19 October 2023).

46. Pujol FE, Pagel C, Brown KL, Taylor JA, Doidge J, Feltbower RG, *et al*. Linkage of National Congenital Heart Disease Audit data to hospital, critical care and mortality national data sets to enable research focused on quality improvement. *BMJ Open* 2022;**12**:e057343. https://doi.org/10.1136/bmjopen-2021-057343