

# *The road to hell winds on: the high administrative burden of maintaining linked national health data*

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access

Taylor, J., Pagel, C., Espuny-Pujol, F. ORCID: <https://orcid.org/0000-0001-9085-7400> and Crowe, S. (2026) The road to hell winds on: the high administrative burden of maintaining linked national health data. *International Journal of Population Data Science*, 11 (1). ISSN 2399-4908 doi: 10.23889/ijpds.v11i1.3000 Available at <https://centaur.reading.ac.uk/129739/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

Identification Number/DOI: 10.23889/ijpds.v11i1.3000  
<<https://doi.org/10.23889/ijpds.v11i1.3000>>

Publisher: Swansea University

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

[www.reading.ac.uk/centaur](http://www.reading.ac.uk/centaur)

**CentAUR**

Central Archive at the University of Reading

Reading's research outputs online

## The Road to Hell Winds on: The High Administrative Burden of Maintaining Linked National Health Data

 Julie A. Taylor<sup>1,\*</sup>, Christina Pagel<sup>1</sup>, Ferran Espuny Pujol<sup>1,2</sup>, and Sonya Crowe<sup>1</sup>

### Submission History

Submitted:	02/06/2025
Accepted:	13/02/2026
Published:	30/04/2026

<sup>1</sup> Clinical Operational Research Unit, Department of Mathematics, University College London, 4 Taviton Street, London, WC1H 0BT, UK

<sup>2</sup> Department of Computer Science, University of Reading, Pepper Lane, Reading RG6 6DH, UK

### Abstract

#### Introduction

Data linkage studies generate opportunities to answer important health questions, but are also notoriously time consuming, with delays accessing data lasting years. Previously we described our experience of accessing and linking five national datasets for our study 'LAUNCHES QI'. However, the delays and challenges we encountered did not end once data were received - maintaining access and research amendments also required significant time and resources.

#### Objective

To outline the costs, time, and workload associated with LAUNCHES, including set-up, annual requirements, and amendments. We also describe the experience of re-using the linked dataset for a second study, CHAMPION, which we wrongly assumed would be simpler. We highlight new delays and challenges, offer advice to researchers, and suggest improvements to the system.

#### Results

LAUNCHES set-up involved 39 documents and data access took 2.5 years. Subsequent amendments/renewals added 75 more documents. Repurposing the data for CHAMPION required restarting the entire process from scratch. It took nearly two years to gain initial approvals and another 18 months to receive the datasets, with 44 documents submitted for set-up and 92 more for amendments and annual processes.

New challenges and delays included changes of data controller, data extract errors, and the repetitive amendment process.

#### Conclusions

With growing recognition of the challenges in using health data for research and our own increasing experience navigating the system, we had hoped for positive change. Instead, there are more forms, and processes have become harder, not easier. We believe simple changes could streamline the system, yet with each new study, the document set grows, and the forms get longer. Research with linked datasets is incredibly painful, and the system discourages amendments, stifling innovation. Whilst we've realised the value of linked data, delays to research are significant, and researchers may give up without simpler processes to link and update data.

#### Keywords

data linkage; audit; public health; health services research

\*Corresponding Author:

Email Address: [julie.a.taylor@ucl.ac.uk](mailto:julie.a.taylor@ucl.ac.uk) (Julie A. Taylor)



## The Unfulfilled Potential

The opportunity generated by data linkage studies using health data, the time and workload involved, and the challenges encountered are being increasingly reported [1–28]. In some cases, publicly funded studies that have invested substantial time, funding, and resources in navigating approvals and data application processes are ultimately unable to obtain all of the requested data [19, 29]. There are ongoing efforts to improve the system, including a government commissioned review on using health data for research and analysis [30] and an independent review of the UK health data landscape [31].

As researchers attempt to answer important health questions by setting up larger data assets with linked datasets [32], the barriers and amount of time it takes is delaying research progress. Solutions such as using simulations [33], self-report for health outcomes [34, 35], synthetic data [36, 37] or specific trusted research environments (TREs) have been explored [38–40]. A TRE is a highly secure computing environment that provides remote access to health data for approved researchers to use in research. These ‘data safe havens’ are being established by custodians to hold several de-identified datasets with a process to allow approved external researchers to analyse the data. However, they may not always hold the right datasets to address specific research questions (e.g. a specific audit). When they do have the required datasets, the data can only be processed for pre-specified research questions. Therefore, many studies still require the use of bespoke linked datasets [41, 42].

In a seemingly positive move, the government has recently announced plans to establish a national health data service, the Health Data Research Service (HDRS) [43]. However, given the anticipated multi-year timeline for its implementation, it remains essential to address limitations within the current data access and governance system to enable high-quality, timely research. Doing so is critical for improving health services and for generating the evidence needed to inform the updating of policies and procedures.

## The Road to Hell Winds On

In previous work [8] we reported the challenging and sometimes demoralising burden of setting up our data linkage study LAUNCHES QI: Linking AUDit and National datasets in Congenital HEArt Services for Quality Improvement (Box 1). LAUNCHES QI has now ended, 3.5 years later than intended (partly, but definitely not wholly, due to the pandemic). During the project (and after) we found that the administrative burden and associated costs do not end once you have your data.

In this paper we take you further along the road to hell, highlighting the ongoing requirements and challenges from our experience of applying for pseudonymised patient level health data from audits and national datasets in England to ultimately link together to analyse. We uniquely describe the total costs, time, and workload associated with the entire project, including study set-up, annual requirements, and amendments. We also describe our experience of re-using the dataset for another study CHAMPION, Congenital Heart Audit: Measuring Progress In Outcomes Nationally (box 1) (which we wrongly assumed would be more straightforward), and

report on new challenges encountered. Finally, we summarise the research made possible with the final linked datasets.

## Total Document Burden

### Set-Up

The full list of documents required for the set-up of both studies is given in Supplementary Table 1 (minor amendments not included in this table). The study set-up for LAUNCHES required 39 documents spanning 345 pages, shared with 11 regulators, with multiple submissions of many of the documents (133 submissions in total). Although re-using the LAUNCHES dataset (plus some additional data), CHAMPION was considered a new study and required the same set-up process as LAUNCHES (described in box 1) with additional documents required for the extra datasets, the addition of online forums, and an increase in document requirements and document size due to changing processes. In total, CHAMPION set-up involved 44 documents spanning 445 pages, shared with 12 regulators, with 143 document submissions in total. For one data controller, CHAMPION was added as an amendment to the LAUNCHES Data Sharing Agreement, which added complexity when amendments were made to one study and not the other.

### Once Studies Are Underway

The statistics above just relate to study set-up. Table 1 shows the documents required throughout the rest of each study for LAUNCHES and CHAMPION.

Annual requirements involved the maintenance of our approvals: a review for the research ethics committee (REC), a report for the Confidentiality Advisory Group (CAG), information governance updates (university specific), and reviews submitted to most of the data controllers to renew the Data Sharing Agreement (DSA) annually. We note that the projected end dates (several years in the future) were known and shared with all regulators from the start.

As each study progressed and we began working with the datasets, changes and improvements such as the addition of objectives, changing project staff, and extensions to the study end date, needed to be made which required the completion of amendment forms and processes. Combined with the required annual documentation to maintain compliance, this resulted in a considerable ongoing burden throughout each study.

During LAUNCHES we submitted five minor amendments. When combined with annual requirements, 75 documents were added to the study document set over the course of the study, increasing the overall document total for LAUNCHES to 104.

During CHAMPION we submitted three minor and two substantial amendments. We completed an extra 26 amendment/annual renewal forms and submitted 66 study documents to accompany the forms (these included new versions of study documents, and approvals received for the amendments), meaning in total an additional 92 documents for amendments and annual requirements, with a study total of 136. These totals are correct as of March 2025 but will increase as annual requirements will continue for as long as we retain the data.

## Box 1: Description of LAUNCHES and CHAMPION

LAUNCHES and CHAMPION are data linkage studies using national audit and routinely collected health data to explore service use and outcomes for patients with congenital heart disease.

The aim of **LAUNCHES** is to support improvement in services for congenital heart disease and provide a template for other lifelong conditions.

Objectives:

- Establish patient trajectories through the healthcare system.
- Identify variations, including over time and by demographics.
- Identify what is meaningful for quality improvement.

The study involved the linkage of five datasets: The National Congenital Heart Disease Audit (NCHDA), the Paediatric Intensive Care Network (PICANet), the Intensive Care National Audit and Research Centre (ICNARC), Hospital Episode Statistics (HES), and civil registrations mortality data from the Office of National Statistics (referred to as mortality or ONS data throughout).

Required were university approvals for: data protection, insurance, information governance and approval of the completed Integrated Research Application System application, protocol and other study documents. This information and the approvals made up the document set (supplementary table 1) to submit to the Research Ethics Committee (which was required as LAUNCHES involved patients identified in the context of their past use of services in the National Health Service), and the Confidentiality Advisory Group to apply for section 251 approval to link and process the data (as it is not feasible to contact patients from the audits for retrospective individual consent). While these applications were under review, we began four data applications for the five data sets, to submit for approval to the three data controllers.

**CHAMPION** was set-up to reuse this already linked dataset for a new purpose, with additional linkage of the National Adult Congenital Surgery Audit (NACSA), and the analysis of National Congenital Anomaly and Rare Disease Registration Service (NCARDRS) and a hospital clinical audit from an NHS Trust.

The aim of CHAMPION is to improve how quality of services for congenital heart disease (CHD) is measured and reported in England.

Objectives:

- Define which cases of adults undergoing surgery or interventional cardiology should be included in national audit reporting of adult CHD outcomes.
- Identify key diagnoses for which long term outcomes should be reported.
- Capture the views of patients and parents/carers about which procedure- based outcomes for adults and which longer-term outcomes by diagnosis would be most meaningful.
- To build prototype software tools that specify the automated mechanisms required to extract, generate, and visualize the outcomes.

Setting up the CHAMPION study required a new ethics and CAG application, new university processes and a new application to each data controller, just for the reuse of the datasets already held for LAUNCHES. Additionally, a new application was required for each of the extra datasets.

Table 1: Additional Documents Required Throughout LAUNCHES and CHAMPION

Renewal and amendment documents	LAUNCHES (study length 6 years)		CHAMPION (study length 5 years)	
	Forms completed	Accompanying documents <sup>a</sup> submitted	Forms completed	Accompanying documents <sup>a</sup> submitted
<b>Annual requirements</b>				
Ethics annual review	5	0	3	0
CAG annual report	5	0	3	0
UCL Information governance	5	8	3	4
DSA extension - PICANet	4	8	1	2
DSA extension - NICOR (NCHDA and NACSA <sup>c</sup> data)	3	6	0	0
DSA extension - NHS Digital	1	2	n/a <sup>b</sup>	n/a
<b>Minor amendments</b>				
IRAS amendment tool	5	0	1	2
CAG amendment form	5	0	1	2
DSA amendment - PICANet	3	6	1	2
DSA amendment - NICOR (NCHDA and NACSA <sup>c</sup> data)	3	6	1	2
DSA amendment - NHS Digital	0	0	0	0
<b>Substantial amendments</b>				
IRAS amendment tool			2	9
CAG amendment form			2	9
DSA amendment form - PICANet			2	8
DSA amendment form - NICOR (NCHDA and NACSA <sup>c</sup> data)			2	12
DSA amendment form - NHS Digital			3	9
DSA amendment form - NHSE (NCARDRS)			1	5
<b>Extra documents</b>	<b>39</b>	<b>36</b>	<b>26</b>	<b>66</b>
<b>Total extra documents</b>	<b>75</b>		<b>92</b>	

<sup>a</sup>New versions of protocol, data linkage diagram, privacy notice, ethics amendment approval, CAG amendment approval, and in one case the re-sharing of all documents from the original data application.

<sup>b</sup>The CHAMPION NHSD application was submitted as an amendment to the LAUNCHES application, therefore the annual renewal was covered in the LAUNCHES renewal.

<sup>c</sup>NACSA data was for CHAMPION only.

CAG = Confidentiality Advisory Group; DSA = Data Sharing Agreement; UCL = University College London; IRAS = Integrated Research Application System; PICANet = Paediatric Intensive Care Network; NICOR = National Institute for Cardiovascular Outcomes Research; NCHDA = National Congenital Heart Disease Audit; NACSA = National Adult Congenital Surgery Audit; NHSE = NHS England; NCARDRS = National Congenital Anomaly and Rare Disease Registration Service.

## Total Data Costs

Data linkage studies are expensive, with charges possible from each data processor and controller for application review, data processing, linkage, and transfer costs.

Table 2 shows the total data costs for each dataset for both LAUNCHES and CHAMPION. At the time of reporting, the total data costs for LAUNCHES are £55,650. Following the study end in October 2023, data is being retained to complete analyses and publications. This retention will require annual renewals for some of the data sharing agreements, with associated costs. Therefore, the costs will continue for as long as we hold the data.

The cost to re-use the data for CHAMPION, to receive two additional datasets, and for the annual requirements came to £15,756. The substantial amendment that included an update of the NCHDA and ONS data extracts from 2017 to 2022 came to £8,700. The total data costs for CHAMPION to date are £24,456.

## Total Time Taken

Figure 1 shows the timeline for CHAMPION from study set-up and for each substantial amendment. Supplementary Figure 1 (SF1) shows the timeline for LAUNCHES updated from the one previously reported [8].

For LAUNCHES, the time taken to receive approval from each of the regulators ranged from 4 to 9 months. From study start to receiving the final dataset was 2.5 years. Minor amendments (not all shown) took between 1 and 4 months for each of the necessary approvals.

For CHAMPION, permissions to re-use the data we already held at UCL took 1 year and 10 months with individual approval times ranging from 1 to 22 months. It was a further 1 year and 6 months before we had received all additional datasets requested.

During CHAMPION, two substantial amendments were submitted. The first was to add an objective, to request an update of the NCHDA data and the mortality data, and

Table 2: Data Controller and Processor Costs for LAUNCHES and CHAMPION

Dataset	Type of cost	LAUNCHES	CHAMPION	
		Cost (£)	Cost (£)	Amendment for extra data (cost (£))
NCHDA/NACSA <sup>c</sup>	NICOR processing costs	9,250	4,250	1,200
	NICOR annual costs	2,200	0	0
	HQIP costs: original application fee	0	1,800	
	HQIP <sup>a</sup> costs: annual renewals/amendments	600	1,200	0
PICANet	Processing costs <sup>b</sup>	0	0	0
	HQIP costs: original application fee	0	1,800	0
	HQIP costs: annual renewals/amendments	3,000	2,460	600
ICNARC	All costs	10,000	0	0
HES and mortality data	NHS Digital controller and processing costs	27,900	984	5,400
	Annual renewals/amendments	2,700	2,264	
NCARDS data	Processing and data controller costs		998	1,500
<b>TOTAL (£)</b>		<b>55,650</b>	<b>15,756</b>	<b>8,700</b>

<sup>a</sup>Data controller changed to NHSE for the amendment for extra data.

<sup>b</sup>PICANet did not charge processing costs at the time of these two studies, but have subsequently introduced charges that will apply to future studies.

<sup>c</sup>NACSA data was for CHAMPION only.

PICANet = Paediatric Intensive Care Network; NICOR = National Institute for Cardiovascular Outcomes Research; NCHDA = National Congenital Heart Disease Audit; NACSA = National Adult Congenital Surgery Audit; HES=Hospital Episode Statistics; HQIP = Health Quality Improvement Partnership; NHSE = NHS England; NCARDS = National Congenital Anomaly and Rare Disease Registration Service.

to perform an additional link of the NCARDS dataset to NCHDA and mortality data. The second amendment was to add an objective that had been flagged as important by stakeholders, to extend the study end date to have time to complete the new objective, and to make changes to data items that we would be receiving for the NCARDS dataset (we were early users of the dataset and so despite liaising with the data controllers had requested variables that were not fit for purpose in the initial application).

It took ten months to receive all approvals for substantial amendment 1, and a further seven months to receive the NCHDA and mortality data requested. The dataset for NCARDS data was only received two years later, and with only one month left to go until study funding ended.

## Process Delays

Throughout the studies we encountered major delays specific to both studies and some delays through broader system challenges. With the benefit of our experience from LAUNCHES, we were prepared for the challenges for setting up CHAMPION, or so we thought. Challenges from LAUNCHES, such as uncertainty over the processes and intricacies required of individual data controllers, were not the issue, as we had the experience from LAUNCHES. However, there were new challenges, as set out below.

## Study Specific Delays

### The Pandemic

Delays due to the pandemic were impossible to anticipate, but it is important to note that this contributed in part to the delays in the approval process, where ethics, CAG, and the data controllers rightly prioritised applications for COVID research, and the funders allowed us to pause the study with only the study set-up continuing. Early communication with the funders has been key for all delays, and thankfully, they were understanding.

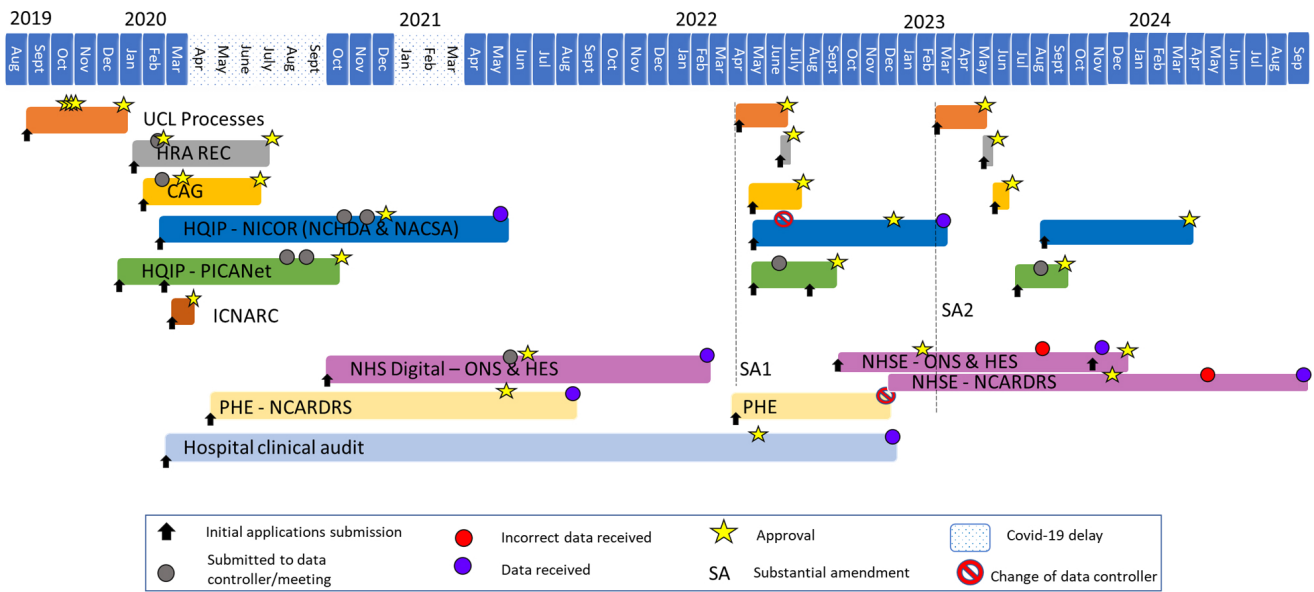
### Data Errors

Three out of nine of our data extracts were missing requested fields or detail. Waiting for the corrected extracts caused a 1 year and 5 months delay, a 4-month delay, and a 5-month delay.

### Change in Data Controller

Two of the data controllers with whom we had data sharing agreements changed during the study. This occurred whilst we were submitting an amendment and there were no processes in place to deal with our request. This resulted in approvals taking six months for one application and a one year wait for the second application, just for approvals. As one new data controller was not familiar with the details of the study, we had to complete the full application form again, and re-share

Figure 1: Timeline of the Permission Processes and Time Taken in CHAMPION: From First Submission to Approval and Data Received if Applicable, for Each Regulator



The timeline runs from the first submission at the beginning of CHAMPION to the last approval of the second substantial amendment (minor (non-substantial) amendments are not shown).

CAG = Confidentiality Advisory Group, HQIP = Health Quality Improvement Partnership, PICANET = Paediatric Intensive Care Audit Network, ICNARC = Intensive Care National Audit & Research Centre, NICOR = National Institute For Cardiovascular Outcomes Research, HRA REC = Health Research Authority Research Ethics Committee, UCL = University College London, NACSA = National Adult Congenital Surgery Audit, NCARDS = National Congenital Anomaly and Rare Disease Registration Service, NCHDA = National Congenital Heart Disease Audit, NHSE = NHS England, ONS = Office for National Statistics, HES Hospital Episode Statistics, PHE = Public Health England.

original documents from the initial approvals, as well as provide additional clarifications. The data teams from the original data controllers tried their best to help us push our applications through, but there was just no process in place.

Receiving the data then took an additional 7 months for one data controller, and the second dataset was only received 2 years later, with just one month left on the study. As we would not have been able to complete the analysis planned for this data, we had to request a variation to contract and minor amendments to extend the study end date.

## System Challenges

### The Preparation

Although not a new challenge, the preparation time is relevant to both studies, and an important consideration for anyone thinking of starting a data linkage study. Our timelines (Figure 1 and Supplementary Figure 1) start from submission to the regulators and do not show the huge task of preparing and submitting the set-up documents. It took two months to prepare the document set and complete other aspects of study set-up prior to submission for LAUNCHES, and as CHAMPION was a mixed methods study the preparation time was a little longer.

### The Increasing Burden

In the year from the study set-up of LAUNCHES in 2018 to the study set-up of CHAMPION in 2019, the number of documents required, the number of questions on application

forms, and the number of clarifications required all increased. In 2023 we started a third data linkage study, and the requirements have increased again.

Encouragingly, following the change of data controller from NHS Digital to NHS England in 2022, there were attempts to streamline the DARS system. As these were our lengthiest applications, and the ones requesting the most clarifications, this is a welcome change. As hopeful as we were for that development, a current application for HES data for a third data linkage study has been submitted, there are significant delays due to staff shortages, and we were waiting to be assigned a case officer for 6 months. The reason for this delay was the complexity of our study, which although not greater in complexity than LAUNCHES and CHAMPION, did not fall directly into the new automated system. Now with the DSA approved and signed after working with DARS for 5 months, we received the NHSE data just under one year from submission, with only the bridge files to allow linkage left to follow.

There are other signs of systems growing in complexity. One of our funders is now asking for an additional form from award holders of research programmes to detail data access management plans (DAMPs). They state that in standardising the process they are aiming to reduce the duplication. But this new form *is* the duplication. We can assure funders that the system already has data management covered, in great detail. Perhaps these are useful for studies not requiring ethics and CAG but, in our experience, this was another deadline to meet, received at the worst possible time, during the set-up of the study. The content was mainly covered in our CAG application,

and any other questions (e.g. format of data to be received) would be covered when applying for the data.

### **Patient and Public Involvement and Engagement (PPIE)**

PPIE is incredibly important and detailed PPIE plans are required in funding applications. For both LAUNCHES and CHAMPION we had a PPIE group to provide feedback and for CHAMPION a PPI co-applicant and three partner charities. In addition to this, CAG requested at review of our application that additional PPIE activities were completed to ascertain that the use of the data for the specific research questions has the support of patients. Although important, PPIE activity is not effective as a tick box exercise. It can be time consuming, requires proper and experienced facilitation, and support from relevant communities. It is difficult to achieve as additional unfunded work, and if done incorrectly could be damaging to a study or to public support.

### **Amendments**

During a study, if anything changes from the original application then the amendment process must be completed. Amendments, classified as minor or substantial, increase paperwork and can take several months to be approved. Each amendment requires not only the completion of the amendment form to describe the changes, but a careful re-evaluation of what was written in the original application sometimes written years previously to ensure that the study remains compliant.

Minor amendments included common project events such as adding researchers or extending the study end date and were required for each regulator. Each resulted in a delay of a few months. The irony that delays in data access are often the reason for study extension in the first place is not lost on us. Substantial amendments such as adding objectives, requesting an update of data, or additional linkage caused a significant delay that resulted in the extension of our study end date for CHAMPION.

### **Insanity is Doing the Same Thing Over and Over Again**

The repetition of information across multiple forms and documents required for study set-up is repeated for amendments, on several forms with different formats and slightly different wording for ethics, CAG, and each data application. By the second substantial amendment, the absurdity of this really kicks in.

Some data controllers have made amendments a little easier by having a separate amendment form where only the changes need to be completed, but you still need to repeat it for every dataset. However, despite this, on two occasions we were asked to complete a full application form rather than just the amendment form as there was confusion about the background of the study. Where they do not have an amendment form (e.g. NHS Digital now NHSE), we were often re-questioned on aspects of the original study that had already been clarified and approved.

Table 3 shows the extent of the repetition in the application forms. Most of the information provided in the CAG form is then required again for every data application. Unfortunately,

it is not just a case of copy and paste, as each question will be slightly different.

### **It's Urgent, When it's Our Move**

Renewals and clarifications are chased with urgency when it is our turn to respond but can then sit within the system for months at a time, where the urgency apparently disappears. For example, we waited several months to receive the new process for amending and renewing DSAs following a change of data controller and were told that we would be informed once it was in place. We were not informed, but instead contacted 18 months later to say we were now outside of our current DSA and the renewal was urgent. We completed our side urgently but received the final signed off version 8 Months later.

Some data controllers requested that renewals were submitted up to 3 months before the DSA expiry, but when it's taken several months to get approval, and then several more months to receive the data, it is frustrating to have to look at it again so soon.

### **The Lonely Burden of the Research Coordinator**

Keeping up with changing processes, several data controllers, annual requirements, and the annual expiry of DSAs can become stressful. Research coordinators are doing their best to keep the study compliant and navigate the system but must worry about the smallest deviation from wording written several years earlier being considered a breach of the data sharing agreement. The research coordinator role is crucial and yet often not adequately funded and valued (or remunerated), is needed for the full length of these projects, and ideally full time (at least for the set-up of the project).

### **The Fate of the Dataset**

As LAUNCHES has ended and CHAMPION is in its last few months, the data controller's concern becomes the destruction of the dataset. We have a data retention period for the completion of analyses and publications but after that would be expected to destroy all the data. During data retention, we will need to keep renewing the DSAs which means more documents and costs. Because these fall outside of the funding period, this may not be feasible for the full time planned.

The ultimate destruction of the dataset will be huge waste, given the time, money, and resources used to access, link, clean and prepare the data, as many research questions that could be addressed with these linked datasets, but outside the scope of our studies, remain unanswered. However, the data governance requirements to set up a secure repository where the data could be used by others and updated have already discouraged two potential custodians for the data. Instead, our funder for LAUNCHES provided additional funds to curate the computer code we developed to process the datasets into a public shared repository, which will make it easier for future researchers to recreate and then use linked datasets.

### **Not Sure? Add a Form**

The system's answer to new regulations or oversight is often to add in a form, a question, a clarification, or a process without considering where else this may already be answered

Table 3: The Repetition of Information Across Application Forms and Documents

	Protocol	IRAS	CAG	HQIP PICANet	NHSE (Arden & Gem - NICOR)	NHSE - HES & Mortality	NHSE - NCARDRS	PHE - NCARDRS	ICNARC <sup>a</sup>
<b>Information requested in forms</b>									
Objectives/Rationale	✓	✓	✓	✓	✓	✓	✓	✓	✓
Methodology	✓	✓	✓	✓	✓	✓	✓	✓	✓
Inclusion criteria	✓	✓	✓	✓	✓	✓	✓	✓	✓
Expected measurable benefits		b	b	✓	✓	✓	✓	✓	
Lay title and summary		✓	✓	✓	✓	✓	✓	✓	
Publications & other outputs	✓	✓	✓	✓	✓	✓			✓
Project funding	✓	✓	✓	✓	✓	✓	✓	✓	✓
Processing location and data flows (extra to methodology)		✓	✓	✓	✓	✓	✓	✓	✓
Data access (information on employment contracts, training etc)						✓	✓	✓	✓
Data access including named project team		✓	✓	✓	✓				
Data protection	✓	✓	✓	✓	✓	✓	✓	✓	
GDPR legal basis	✓		✓	✓	✓	✓	✓	✓	✓
Common law duty of confidentiality	✓		✓	✓	✓	✓	✓	✓	
Fair processing	✓		✓	✓	✓	✓	✓	✓	
Data security	✓	✓	✓	✓	✓	✓	✓	✓	✓
Retention & destruction	✓	✓	✓	✓	✓	✓			✓
Intention to link data (extra to data flows)		✓	✓	✓	✓	✓	✓	✓	

Each tick indicates information required in the corresponding form

<sup>a</sup>ICNARC updated their form in March 2024 to include benefits. <sup>b</sup>Ethics and CAG review the research as a whole to establish that public benefit justifies accessing data.

CAG = Confidentiality Advisory Group; IRAS = Integrated Research Application System; PICANet = Paediatric Intensive Care Network; HQIP = Health Quality Improvement Partnership; NICOR = National Institute for Cardiovascular Outcomes Research; NHSE = NHS England; PHE = Public Health England; NCARDRS = National Congenital Anomaly and Rare Disease Registration Service; HES = Hospital Episode Statistics; ICNARC = Intensive Care National Audit and Research Centre; GDPR = General Data Protection Regulation.

or covered. Additions are seen as small administrative tasks, so it'll be quick. However, whilst it may seem like a small change to an individual data processor or controller, it all adds to the burden. We have found university and NHS R&D processes can be particularly prone to responding in this way.

### “You Will be in Breach of the DSA if . . .”

There is a pressure on all parties to get the governance exactly right, but there are grey areas. When we asked questions to ensure we got it right (e.g. whether adding a researcher was allowed without an amendment given that had *optionally* specified the number of researchers accessing the data initially), we often received generic text back, which typically states that any deviation from what is written in the Data Sharing Agreement (DSA) is a breach of that contract, which certainly contributes to *our* fear of legal misstep. Studies change, objectives are added, team members leave and are replaced. It is important that this can happen to allow the best possible research to be carried out. However, any change means that the precise information provided at the beginning of the study might no longer be correct (and so require an amendment). Therefore, from the research team point of view, to navigate the fear of legal misstep the initial application process becomes a trade-off of trying to provide all information and being completely transparent but not providing detail that may later change and make us in breach of the DSA. It is exhausting.

More broadly, it is our belief that the fear of legal misstep within the system is contributing to the burden and delays in accessing data, with each actor behaving as if it is the only regulator and therefore must assess all aspects of the project

for itself, despite the study already having been scrutinised by a number of other regulators.

## Advice for Researchers (and Their Employers) Planning a Data Linkage Study

This advice is additional to that offered in Taylor et al [8], and the list is available in Supplementary Table 2. Whilst this advice may help prevent some delays, our experience and learning did not result in a noticeably easier process.

### Keep Submitted Versions of All Applications

Keep all previous versions of submitted applications. Depending on the system, changes may be made to online applications by the data controllers themselves. If these changes are not correct and are missed by the research team, then this could be a breach. For example, one data sharing agreement included the information for both LAUNCHES and CHAMPION. When edits were made by the data controller team, the information for one study had been removed entirely. If we had not picked this up before signing, we would have had one study proceed without the DSA in place. Keeping track of the original request will help if any errors are made.

### Prepare for Extra Costs

Allow for amendments and renewals in data costings. The amendments will normally extend the date of the DSA,

however costing in both these and the renewals will provide a buffer should costs change or an amendment be required just after a renewal.

## Study set Preparation Time

Factoring in suitable research coordinator and PI time at the beginning of the study will aid the preparation of complete and good quality applications. Many decisions are still being made at the start and assuring the wording is correct in the ethics and CAG applications (the first to be submitted) is important to allow consistency across the applications. Any changes made in a later application may cause a cycle of the amendment process. The more complex the study, and the more study methods used, the more preparation time needed.

## Look After the Research Coordinators

They've probably just been told they're in breach of a DSA because of a typo from 3 years ago, whilst carrying on with the general coordination of the study. They can keep on top of the documents, expiry dates of DSAs, and will be able to organise annual requirements and amendments. Having a supportive research team makes all the difference.

## Suggestions for System Improvements – It's Not Us, It's You

There will be times when researchers do not know the full requirements or miss new requirements that have just been added. That said, we've often received very positive feedback on the quality of our applications (most recently for the extensive ethics application for our latest project). Sadly, we've not experienced the process becoming easier or quicker despite the amount of experience we've amassed. The delays are so widespread, with many independent examples [1–23], that we are confident in saying that the system is the problem. Ultimately, we believe that the only way to make things significantly less burdensome is to change the system.

Figure 2 is an updated diagram of our suggested system improvements. Additions from Taylor et al [8] include:

### Sample Data Extracts Before Application

Sample data extracts of requested fields and levels of completeness for these fields, would enable researchers to understand which fields are needed, as well as make sure that the data requested is suitable for the analyses planned. It may also help to trouble shoot any misunderstanding about derived fields (which are particularly important for data minimisation) before the extract has been finalised and so reduce additional work for the data processor in arranging a corrected extract and avoid delays to the data recipient.

### New Data Controllers to Have Processes in Place Before Taking on New Datasets

New data controllers should have processes in place before they take over. Studies have been left in limbo without a process, wasting time and funding.

## PPIE Assistance

If the integration of the PPIE into the study design that is currently in place is not seen as adequate for CAG and data law requirements, then centralised support from CAG in accessing PPIE participants, groups, and facilitation would be needed.

Such support would also be useful from funders, as finding and securing required numbers of PPIE participants that are able to commit to the full study time, feel properly included and heard and incorporating their opinions and experiences is a large task.

## Partnership Working Between Funders and Data Providers

Greater partnership between research funders and data providers at a strategic and operational level remains important to ensure aligned support of health data linkage studies.

## Single Application Form (and a Simpler Amendment Process)

We previously suggested a single application form and process for data applications, and after several amendments this seems even more pertinent. This is not only a time burden at the beginning of the study, but minor amendments that need to be completed on 5+ forms become a substantial piece of work, and a substantial amendment becomes an epic one. Even if a single application form is not feasible, there is so much overlap between forms (Table 2) that the same structured questions would surely be possible, with each controller being able to add their own agreement terms. The CAG form is the first to be completed, and the most comprehensive. If these questions could be used in data application forms, then it would allow the same answers (that would still be relevant and correct), to be used.

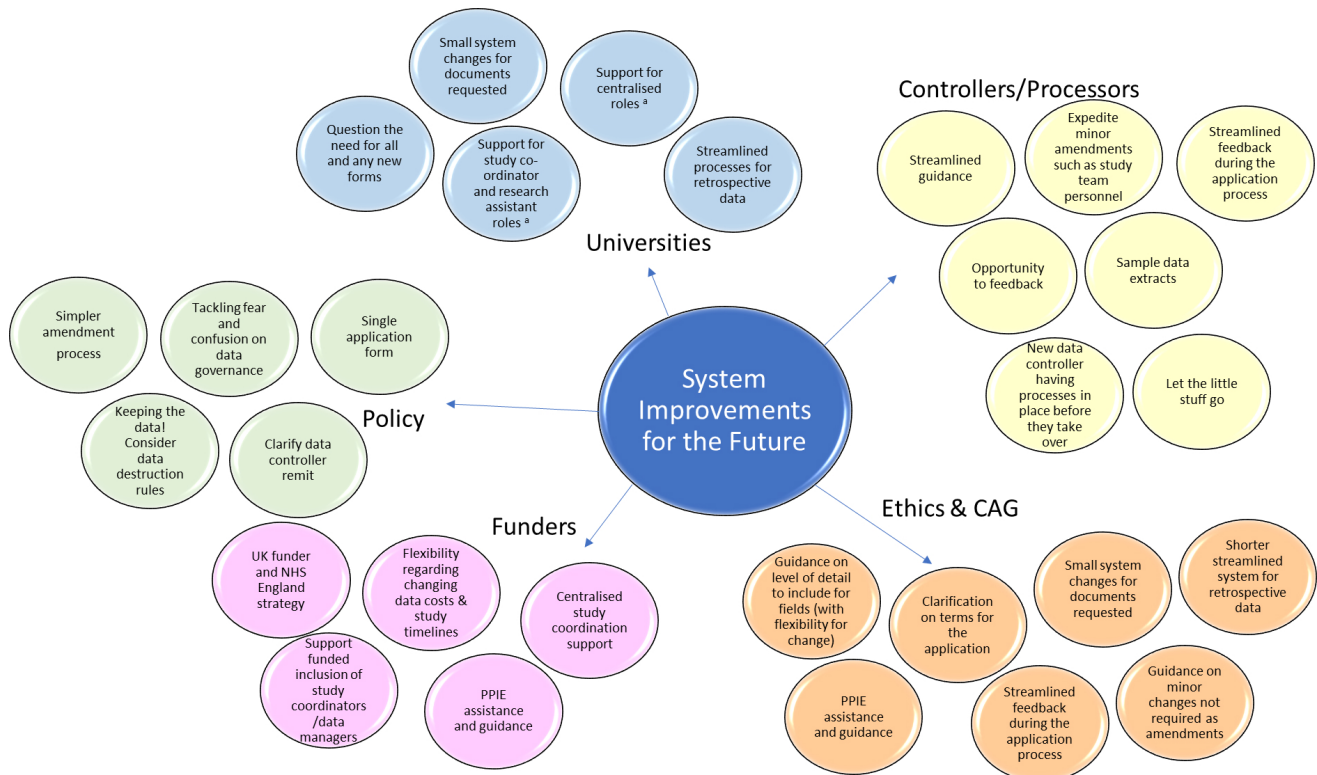
## Let the Little Stuff Go

Greater flexibility would reduce the need for amendments, saving time for both data controllers, processors, and study teams. For example, having to list the team members when a study can go on for years makes amendments inevitable. Not all data controllers request this information but may still want to know the number of people accessing the data. Even allowing an estimated maximum number of researchers, and so be flexible to small changes, would be useful. That way, a study team hasn't immediately set themselves up for a breach when they hire someone 2 years later.

Another example is requiring specific grammar. Some forms are required to be written in the first person, some in the third person. Surely this doesn't matter and often prevents the re-use of already prepared and relevant answers. Small system changes could make a big difference.

A few of the forms required for ethics (e.g. the Organisational Information Document) may not be relevant to the study, but you still must complete the first section, select that this form is not applicable to you, obtain authorisations from the university, and submit it. Getting your head round lengthy forms that are not necessary whilst trying to arrange

Figure 2: Illustrates Systems Improvements for the Future



The coloured bubbles contain the suggestions, with each colour representing an authority that should consider these changes. CAG, Confidentiality Advisory Group.

<sup>a</sup>Whilst one of our suggested system improvements is for centralised support for study coordinator and research assistant roles, we support the inclusion of these roles to remain within the study team. This is important for teamwork, maintaining enthusiasm for the study aims and study progress, and gives a sense of belonging to staff in a team. All qualities that are being lost as increasing numbers of university professional service roles are centralised, undervalued, and sidelined.

the other 30 or so forms you do need, is not helpful. Often the lengthy approval process will not begin until all forms, including those not applicable, are submitted.

Most data controllers provide feedback once, requesting any clarifications, which if made well, means the application can proceed. However, some will request clarifications over several months and your application may bounce repeatedly and quickly back to you for small reasons, which can be really disheartening.

## Keep Going - the Value of Linked Datasets

Conscious that we do not want to deter researchers from data linkage studies, we would like to say that although it was incredibly challenging and morale-sapping, we did get there, were able to complete the study objectives, and completed analyses that would not have been possible without linking the data.

For the LAUNCHES objectives we have published eight papers to date with five in preparation, and from CHAMPION we have published nine papers, with two in preparation. We have reported the resource use, outcomes, and difference in hospital burden for several congenital heart disease diagnoses [44–47] and reported for the first time the differences

between hospitals when children transfer to adult services [48], and will be reporting on how the adult CHD population use services. We have also looked at long term outcomes by CHD diagnosis [49], and developed risk models for use in routine monitoring of 30-day and 90-day survival in adults undergoing surgery for CHD and for 30-day complications following CHD surgery in children and in adults [50–52]. Using the delayed NCARDS dataset we evaluated outcomes for all detected structural CHD cases in England from fetal life to the age of 1 year, and reported on the under estimation of cases due to no intervention [53].

## Conclusion

With increasing recognition of the challenges of using health data for research and our own increasing experience at navigating the system, we were hoping for positive change in the 3.5 years since Road to Hell was published. Instead, the forms have proliferated, and processes have become harder, not easier. We believe that simple changes could greatly streamline the system, and yet with each new study we undertake (we are currently doing this process for a new project), the document set increases, and the application forms, the steps to approval and the delays get longer. Doing research with linked datasets is incredibly painful and there is a big disincentive to make amendments which stifles innovation.

We have been able to realise the research value of the linked dataset, but the current system delays research and many researchers might just give up without simpler processes to routinely link and update data.

## Acknowledgements

This report is independent research funded by the NIHR Policy Research Programme (grant number PR-R20-0318-23001) and the Health Foundation, an independent charity committed to bringing about better health and healthcare for people in the UK (Award number 685009). The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

We would like to thank the data application teams at PICANet, ICNARC, NICOR, HQIP and NHS England for their help and guidance as we negotiated the data application system.

## Statement on Conflicts of Interest

None declared.

## Ethics Statement

LAUNCHES was approved by the North of Scotland NHS Research Ethics Committee on 26<sup>th</sup> October 2018 (REC number 18/NS/0106), and the Health Research Authority Confidentiality Advisory Group (CAG) on 29<sup>th</sup> November 2019 (CAG number 18CAG0180), which permits the use of registry data for specific research purposes without consent.

CHAMPION was approved by the North of Scotland NHS Research Ethics Committee on 14<sup>th</sup> February 2020 (REC number 20/NS/0022), and the Health Research Authority Confidentiality Advisory Group (CAG) on 12<sup>th</sup> July 2020 (CAG number 20/CAG/0027), which permits the use of registry data for specific research purposes without consent.

## Data Availability Statement

The research dataset is not available for sharing under the current data sharing agreements.

## AI Disclosure Statement

AI-assisted tools were not used in any stage of the research or writing process.

## References

1. Lugg-Widger F AL, Cannings-John R, et al. . Challenges in accessing routinely collected data from multiple providers in the UK for primary studies: Managing the morass. *International Journal of Population Data Science*. 2018;3(3). <https://doi.org/10.23889/ijpds.v3i3.432>

2. Dattani N, Hardelid P, Davey J, Gilbert R. Accessing electronic administrative health data for research takes time. *Arch Dis Child*. 2013;98(5):391–2. <https://doi.org/10.1136/archdischild-2013-303730>
3. Jones KH, Laurie G, Stevens L, Dobbs C, Ford DV, Lea N. The other side of the coin: Harm due to the non-use of health-related data. *Int J Med Inform*. 2017;97:43–51. <https://doi.org/10.1016/j.ijmedinf.2016.09.010>
4. Jones KH, Ford DV. Population data science: advancing the safe use of population data for public benefit. *Epidemiol Health*. 2018;40:e2018061. <https://doi.org/10.4178/epih.e2018061>
5. Jones KH, Heys S, Tingay KS, Jackson P, Dibben C. The Good, the Bad, the Clunky: Improving the Use of Administrative Data for Research. *Int J Popul Data Sci*. 2018;4(1):587. <https://doi.org/10.23889/ijpds.v4i1.587>
6. Mourby MJ, Doidge J, Jones KH, Aidinlis S, Smith H, Bell J, et al. Health Data Linkage for UK Public Interest Research: Key Obstacles and Solutions. *Int J Popul Data Sci*. 2019;4(1):1093. <https://doi.org/10.23889/ijpds.v4i1.1093>
7. Cavallaro FL-W, F; Cannings-John, R; Harron, K. Reducing barriers to data access for research in the public interest—lessons from covid-19. *BMJ Opinion*. 2020.
8. 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(8):e047575. <https://doi.org/10.1136/bmjopen-2020-047575>
9. Ferrando-Vivas P, Shankar-Hari M, Thomas K, Doidge JC, Caskey FJ, Forni L, et al. Improving risk prediction model quality in the critically ill: data linkage study. *Health Services and Delivery Research*. 2022;10(39). <https://doi.org/10.3310/EQAB4594>
10. Harron K. Data linkage in medical research. *BMJ Med*. 2022;1(1):e000087. <https://doi.org/10.1136/bmjmed-2021-000087>
11. Jefferson E, Cole C, Mumtaz S, Cox S, Giles TC, Adejumo S, et al. A Hybrid Architecture (CO-CONNECT) to Facilitate Rapid Discovery and Access to Data Across the United Kingdom in Response to the COVID-19 Pandemic: Development Study. *J Med Internet Res*. 2022;24(12). ARTN e40035. <https://doi.org/10.2196/40035>
12. Mc Grath-Lone L, Jay MA, Blackburn R, Gordon E, Zylbersztejn A, Wiljaars L, et al. What makes administrative data "research-ready"? A systematic review and thematic analysis of published literature. *International Journal of Population Data Science (Ijpd)*. 2022;7(1). ARTN 6 <https://doi.org/10.23889/ijpds.v7i1.1718>

13. Moorthie S, Hayat S, Zhang Y, Parkin K, Philips V, Bale A, et al. Rapid systematic review to identify key barriers to access, linkage, and use of local authority administrative data for population health research, practice, and policy in the United Kingdom. *Bmc Public Health*. 2022;22(1):1263. <https://doi.org/10.1186/s12889-022-13187-9>
14. Nishiki MR, Cabecinha M, Knowles R, Peters C, Aitkenhead H, Ifederu A, et al. Establishing risk factors and outcomes for congenital hypothyroidism with gland in situ using population-based data linkage methods: study protocol. *Bmj Paediatr Open*. 2022;6(1). ARTN e001341. <https://doi.org/10.1136/bmjpo-2021-001341>
15. Williamson K, Nimegeer A, Lean M. Navigating data governance approvals to use routine health and social care data to evidence the hidden population with severe obesity: a case study from a clinical academic's perspective. *J Res Nurs*. 2022;27(7):623–36. <https://doi.org/10.1177/17449871221122040>
16. Cavallaro FL, Cannings-John R, Lugg-Widger F, Gilbert R, Kennedy E, Kendall S, et al. Lessons learned from using linked administrative data to evaluate the Family Nurse Partnership in England and Scotland. *Int J Popul Data Sci*. 2023;8(1):2113. <https://doi.org/10.23889/ijpds.v8i1.2113>
17. Claridge H, Tan J, Loane M, Garne E, Barisic I, Caverro-Carbonell C, et al. Ethics and legal requirements for data linkage in 14 European countries for children with congenital anomalies. *Bmj Open*. 2023;13(7):e071687. <https://doi.org/10.1136/bmjopen-2023-071687>
18. Cavallaro F, Clery A, Gilbert R, van der Meulen J, Kendall S, Kennedy E, et al. Evaluating the real-world implementation of the Family Nurse Partnership in England: a data linkage study. *Health Soc Care Deliv Res*. 2024;12(11):1–223. <https://doi.org/10.3310/BVDW6447>
19. Goodacre S, Sutton L, Ennis K, Thomas B, Hawksworth O, Iftikhar K, et al. Prehospital early warning scores for adults with suspected sepsis: the PHEWS observational cohort and decision-analytic modelling study. *Health Technol Assess*. 2024;28(16):1–93. <https://doi.org/10.3310/NDTY2403>
20. Guthrie E, House A, Smith C, Relton S, Romeu D, Saraiva S, et al. Linkage of routinely collected NHS data to evaluate liaison mental health services: challenges and lessons learned. *Health Soc Care Deliv Res*. 2024:1–15. <https://doi.org/10.3310/WCPA5283>
21. Gordon AL, Rand S, Crellin E, Allan S, Tracey F, De Corte K, et al. Piloting a minimum data set for older people living in care homes in England: a developmental study. *Age Ageing*. 2025;54(1). <https://doi.org/10.1093/ageing/afaf001>
22. Mumford L, Hogg R, Taylor A, Lanyon P, Bythell M, McPhail S, et al. Impact of SARS-CoV-2 spike antibody positivity on infection and hospitalisation rates in immunosuppressed populations during the omicron period: the MELODY study. *Lancet*. 2025;405(10475):314–28. [https://doi.org/10.1016/S0140-6736\(24\)02560-1](https://doi.org/10.1016/S0140-6736(24)02560-1)
23. Shahbaz M, Harding JE, Milne B, Walters A, Underwood L, von Randow M, et al. Time and cost of linking administrative datasets for outcomes assessment in a follow-up study of participants from two randomised trials. *BMC Med Res Methodol*. 2025;25(1):21. <https://doi.org/10.1186/s12874-025-02458-9>
24. Eisinger-Mathason TSK, Leshin J, Lahoti V, Fridsma DB, Mucaj V, Kho AN. Data linkage multiplies research insights across diverse healthcare sectors. *Commun Med (Lond)*. 2025;5(1):58. <https://doi.org/10.1038/s43856-025-00769-y>
25. Nelson E, Burns F. Impact through engagement: Co-production of administrative data research and the approach of the Administrative Data Research Centre Northern Ireland. *Int J Popul Data Sci*. 2020;5(3):1369. <https://doi.org/10.23889/ijpds.v5i3.1369>
26. O'Reilly D, Bateson O, McGreevy G, Snoddy C, Power T. Administrative Data Research Northern Ireland (ADR NI). *Int J Popul Data Sci*. 2020;4(2):1148. <https://doi.org/10.23889/ijpds.v4i2.1148>
27. Oderkirk J, Ronchi E, Klazinga N. International comparisons of health system performance among OECD countries: opportunities and data privacy protection challenges. *Health Policy*. 2013;112(1-2):9–18. <https://doi.org/10.1016/j.healthpol.2013.06.006>
28. Van Der Wel KA, Ostergren O, Lundberg O, Korhonen K, Martikainen P, Andersen AN, et al. A gold mine, but still no Klondike: Nordic register data in health inequalities research. *Scand J Public Health*. 2019;47(6):618–30. <https://doi.org/10.1177/1403494819858046>
29. Bhaskaran K, Rentsch CT, MacKenna B, Schultze A, Mehrkar A, Bates CJ, et al. HIV infection and COVID-19 death: a population-based cohort analysis of UK primary care data and linked national death registrations within the OpenSAFELY platform. *Lancet HIV*. 2021;8(1):e24–e32. [https://doi.org/10.1016/S2352-3018\(20\)30305-2](https://doi.org/10.1016/S2352-3018(20)30305-2)
30. Goldacre BM, J. Better, Broader, Safer: Using health data for research and analysis. A review commissioned by the Secretary of State for Health and Social Care. London: Department of Health and Social Care; 2022.
31. C S. Uniting the UK's Health Data: A Huge Opportunity for Society.; 2024.
32. Health Data Research UK Innovation Gateway 2025 [Available from: <https://www.hdruk.ac.uk/>].
33. Joseph Farrington SA, Martin Utley, Kezhi Li, Wai Keong Wong. Many happy returns: machine learning to support platelet issuing and waste reduction in hospital blood banks (Preprint). *arXiv - Preprint*. 2024.

34. Shahbaz M, Harding JE, Milne B, Walters A, Underwood L, von Randow M, et al. Comparison of outcomes of the 50-year follow-up of a randomized trial assessed by study questionnaire and by data linkage: The CONCUR study. *Clin Trials*. 2024. <https://doi.org/10.1177/17407745241259088>
35. Shahbaz M, Harding JE, Milne B, Walters A, von Randow M, Gamble GD, et al. Effect of utilizing either a self-reported questionnaire or administrative data alone or in combination on the findings of a randomized controlled trial of the long-term effects of antenatal corticosteroids. *Plos One*. 2024;19(8). ARTN e030841410.1371/journal.pone.0308414.
36. Kokosi T, De Stavola B, Mitra R, Frayling L, Doherty A, Dove I, et al. An overview of synthetic administrative data for research. *Int J Popul Data Sci*. 2022;7(1):1727. <https://doi.org/10.23889/ijpds.v7i1.1727>
37. Little C, Elliot M, Allmendinger R. Federated learning for generating synthetic data: a scoping review. *Int J Popul Data Sci*. 2023;8(1):2158. <https://doi.org/10.23889/ijpds.v8i1.2158>
38. Brophy R, Bellavia E, Bluemink MG, Evans K, Hashimi M, Macaulay Y, et al. Towards a standardised cross-sectoral data access agreement template for research: a core set of principles for data access within trusted research environments. *Int J Popul Data Sci*. 2023;8(4):2169. <https://doi.org/10.23889/ijpds.v8i4.2169>
39. Priou S, Lame G, Jankovic M, Kempf E. "In conferences, everyone goes 'health data is the future' ": an interview study on challenges in re-using EHR data for research in Clinical Data Warehouses. *AMIA Annual Symposium proceedings / AMIA Symposium AMIA Symposium*. 2023;2023:579–88.
40. London Secure Data Environment 2026 [Available from: <https://www.onelondon.online/london-secure-data-environment/>].
41. Buckel M, Proudfoot AG. Time for a rethink in cardiogenic shock: the shock to survival framework document. *Br J Hosp Med (Lond)*. 2023;84(7):1–8. <https://doi.org/10.12968/hmed.2023.0139>
42. Sonesson E, Das S, Burn AM, van Melle M, Anderson JK, Fazel M, et al. Leveraging Administrative Data to Better Understand and Address Child Maltreatment: A Scoping Review of Data Linkage Studies. *Child Maltreat*. 2023;28(1):176–95. <https://doi.org/10.1177/10775595221079308>
43. Health Data Research Service (HDRS): Department of Health and Social Care; 2025 [Available from: <https://www.gov.uk/government/news/prime-minister-turbocharges-medical-research>].
44. Dorobantu DM, Espuny Pujol F, Kostolny M, Brown KL, Franklin RC, Crowe S, et al. Arterial Switch for Transposition of the Great Arteries: Treatment Timing, Late Outcomes, and Risk Factors. *JACC Adv*. 2023;2(5):100407. <https://doi.org/10.1016/j.jacadv.2023.100407>
45. Dorobantu DM, Huang Q, Espuny Pujol F, Brown KL, Franklin RC, Pufulete M, et al. Hospital resource utilization in a national cohort of functionally single ventricle patients undergoing surgical treatment. *JTCVS Open*. 2023;14:441–61. <https://doi.org/10.1016/j.xjon.2023.03.014>
46. English KM, Espuny-Pujol F, Franklin RC, Crowe S, Pagel C. Secundum atrial septal defect closure in adults in the UK. *Eur Heart J Qual Care Clin Outcomes*. 2025;11(1):78–88. <https://doi.org/10.1093/ehjqcco/qcae019>
47. Dorobantu DM, Huang Q, Espuny-Pujol F, Brown KL, Franklin R, Crowe S, et al. Contemporary outcomes of childhood aortic coarctation interventions: a national registry analysis of mortality, reinterventions and hospital resource use. *Heart*. 2025. <https://doi.org/10.1136/heartjnl-2024-325346>
48. Espuny Pujol F, Franklin RC, Crowe S, Brown KL, Swan L, Pagel C, et al. Transfer of congenital heart patients from paediatric to adult services in England. *Heart*. 2022;108(24):1964–71. <https://doi.org/10.1136/heartjnl-2022-321085>
49. Brown KL, Huang Q, Espuny-Pujol F, Taylor JA, Wray J, van Doorn C, et al. Evaluating Long-Term Outcomes of Children Undergoing Surgical Treatment for Congenital Heart Disease for National Audit in England and Wales. *J Am Heart Assoc*. 2024;13(21):e035166. <https://doi.org/10.1161/JAHA.124.035166>
50. Espuny-Pujol F, Pagel C, Ambler G, Brown KL, Coats L, Franklin RC, et al. New Models for Risk-Adjusted Monitoring of Postsurgical Complications and Mortality in Adult Congenital Heart Disease in England and Wales. *The Annals of thoracic surgery*. 2025. <https://doi.org/10.1016/j.athoracsur.2025.07.030>
51. Mitchell HK, Abdelmoumene M, Espuny-Pujol F, Ambler G, Taylor J, Franklin RCG, et al. Postoperative complications after paediatric cardiac surgery: the role of ethnicity and deprivation - a national cohort study. *Arch Dis Child*. 2025. <https://doi.org/10.1136/archdischild-2025-329267>
52. Mitchell HK, Espuny-Pujol F, Franklin RC, Ambler G, Stickley J, Taylor JA, et al. Risk Models for Monitoring Postoperative Complication Rates After Paediatric Cardiac Surgery. *Eur J Cardiothorac Surg*. 2025;67(10). <https://doi.org/10.1093/ejcts/ezaf317>
53. Huang Q, Franklin RCG, Seale AN, Jowett V, Pagel C, Crowe S, et al. Congenital heart disease in England: a national cohort study from fetal diagnosis to end of infancy. *Heart*. 2025. <https://doi.org/10.1136/heartjnl-2025-326369>

## Abbreviations

CAG:	Confidentiality Advisory Group	JRO:	Joint Research Office
DSA:	Data Sharing Agreement	NACSA:	National Adult Congenital Surgery Audit
GDPR:	General Data Protection Regulation	NCARDRS:	National Congenital Anomaly and Rare Disease Registration Service
HES:	Hospital Episode Statistics	NCHDA:	National Congenital Heart Disease Audit
HRA:	Health Research Authority	NHSE:	NHS England
HQIP:	Health Quality Improvement Partnership	NICOR:	National Institute for Cardiovascular Outcomes Research
ICNARC:	Intensive Care National Audit and Research Centre	PHE:	Public Health England
IG:	Information Governance	PICANet:	Paediatric Intensive Care Network
IRAS:	Integrated Research Application System	UCL:	University College London



# Supplementary appendices

Supplementary Table 1: Document set for study set-up for LAUNCHES and CHAMPION

	Regulator													LAUNCHES		CHAMPION	
	Data protection	Information Governance	JRO office	HRA REC	CAG	Named NHS Site	NHS site of 1 collaborator	HQIP NICOR	HQIP PICANet	ICNARC	NHS Digital	Public Health England	Pages/ screens	Times submitted	Pages/ screens	Times submitted	
<b>Study forms filled out</b>																	
Data Protection form													7	3	8	3	
<b>Governance forms (online)</b>													1	1	5	1	
IRAS form													32	4	37	4	
<b>CAG form – IRAS</b>													40	5	42	5	
CAT advice form													5	1	0	0	
<b>HQIP form (NICOR)</b>													29	4	20	2	
HQIP form (PICANet)													24	2	20	2	
<b>ICNARC form</b>													5	2	5	2	
DARS form (online)													21	2	20	2	
PHE form for NCARDRS													0	0	15	2	
NHS DSA for Barts clinical audit													0	0	12	2	
<b>Data fields &amp; justification (NICOR)</b>													4	1	4	1	
Data fields & justification (PICANet)													4	1	4	1	
<b>Data fields &amp; justification (ICNARC)</b>													4	1	0	0	
<b>Data fields &amp; justification (NCARDRS)</b>													0	0	8	1	
Data Linkage Diagram													1	9	0	0	
<b>Detailed data linkage diagram</b>													1	1	1	9	
Risk Assessment													3	3	2	1	
<b>IG toolkit confirmation</b>													1	8	0	0	
Correspondence with NICOR													1	2	0	0	
<b>Correspondence with charities</b>													1	1	0	0	
Letter from SIRO													1	3	1	3	
<b>Privacy notice</b>													2	9	4	10	
<b>Privacy notice - short description</b>													0	0	1	10	
Protocol													25	6	14	6	
<b>Insurance registration form</b>													6	1	14	1	
Statement of Activities													22	4	0	0	
<b>Organisational Info Document</b>													7	4	10	4	
<b>Schedule of events</b>													3	4	3	4	
Cover Letters													3	3	2	3	
<b>Checklist</b>													2	2	2	2	
<b>DPIA</b>													0	0	3	4	
<b>Signature &amp; delegation log</b>													0	0	3	2	
<b>Patient information</b>													0	0	21	4	
<b>PPI opinions on data use</b>													0	0	2	1	
<b>Documents organised/shared</b>																	
<b>HQIP letter of support</b>													1	3	1	1	
Insurance Certificate													1	1	1	1	
<b>REC provisional opinion</b>													5	2	6	2	
REC favourable opinion													4	8	6	8	
<b>CAG provisional approval</b>													8	4	10	4	
CAG approval													6	8	9	8	
<b>Funder Contract</b>													38	2	52	2	
PI CV													4	4	2	4	
CV's from team													10	1	4	1	
Reviewers Comments													4	4	2	4	
HRA/HCRW approval letter													4	7	6	8	
Training certificates													5	2	6	2	
Service agreement													0	0	14	3	
Draft collaboration agreement													0	0	37	2	
Honorary contracts													0	0	6	1	
<b>Totals</b>													345	133	445	143	

**Key:**  
 Form required for both LAUNCHES and CHAMPION  
 Form required for LAUNCHES only  
 Form required for CHAMPION only

Block shading indicates documents required for each corresponding regulator. The 'times submitted' column sums the number of times each document was required by a different regulator. \*Data protection, Information Governance and JRO processes as at UCL.

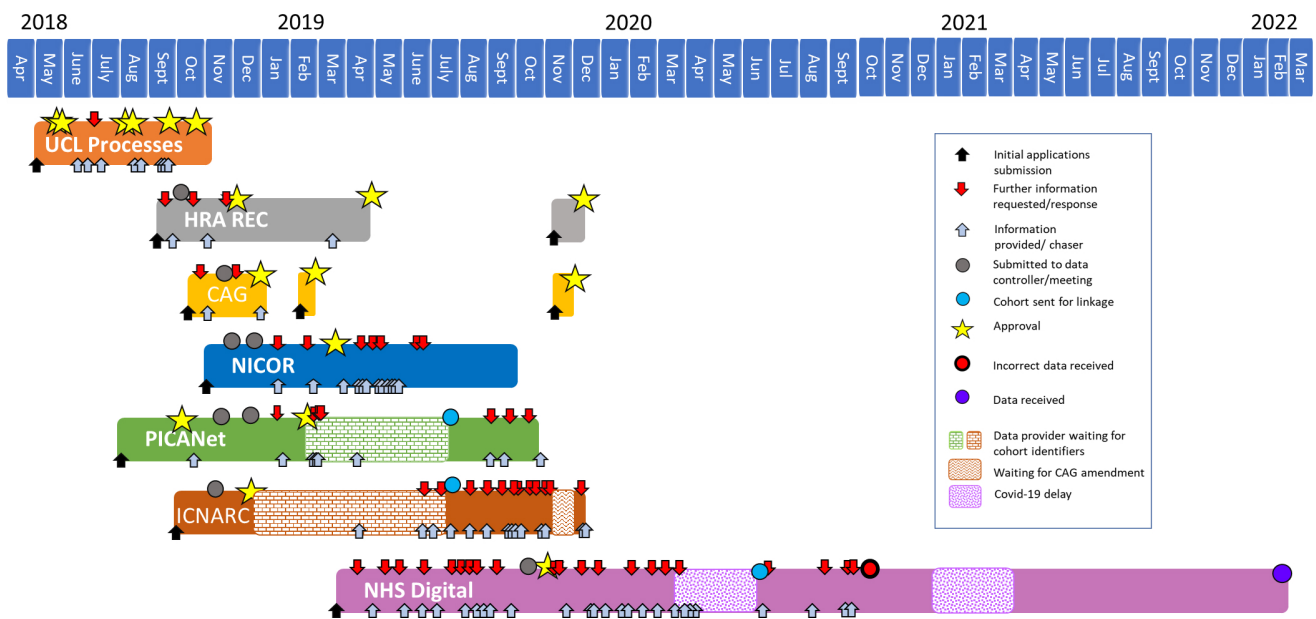
IRAS = Integrated Research Application System, CAG = Confidentiality Advisory Group, CAT = Confidentiality Advice Team, PICANET = Paediatric Intensive Care Audit Network, ICNARC = Intensive Care National Audit & Research Centre, DARS = Data Access Request Service, NICOR = National Institute For Cardiovascular Outcomes Research, IG = Information Governance, SIRO = Senior Information Risk Owner, HQIP = Health Quality Improvement Partnership, HRA REC = Health Research Authority Research Ethics Committee, PI = Principal Investigator, CV = Curriculum Vitae, HCRW = Health And Care Research Wales, UCL = University College London, JRO = Joint Research Office.

Supplementary Table 2: Previously reported advice for your data linkage studies [8]

Advice	Explanation
Find people who have been there before you	Researchers who were further ahead in their data application journey shared their learning and expertise with us. This helped speed up the initial process of preparing the applications.
Love thy research assistants/study coordinators (the legal amount)	These are roles that often get squeezed at application stage to reduce costs and can be challenging to resource if part-time. But these roles are absolutely crucial in accessing data.
Have a flexible start up period	Having an initial data application period in the project with just the PIs and a study coordinator funded allowed us to react early to delays and quickly adjust the timeline.
Add clinical and audit experts to the team	Input from clinicians on data fields helped with data minimisation versus research realisation (particularly a problem for exploratory work). We also found it useful to include, where possible, collaborators working for data controllers and audits.
Plan your data minimisation and beware of differences in how some terms are interpreted	You will need to describe how you have minimised the need for patient identifiers, and all fields requested will require justification. In doing so be sure to correctly state if data will be de-identified, pseudonymised or further anonymised. Some consider anonymised data to be aggregate data only, and amendments may be needed to clarify terms.

PI=Principal Investigator.

Supplementary Figure 1: Updated Timeline of the Permission Process Order and Time Taken in LAUNCHES QI: From First Submission to Data Acquisition



Each box represents the time from submission of the data application to transfer of the data to UCL, for each data set. Note that not all requests for further information and responses have been included! CAG, Confidentiality Advisory Group; HRA, Health Research Authority; ICNARC, Intensive Care National Audit and Research Centre; NHS, National Health Service; NICOR, National Institute for Cardiovascular Outcomes Research; PICANet, Paediatric Intensive Care Audit Network; REC, Research Ethics Committee; UCL, University College London.