BMJ Open

Linkage of National Congenital Heart Disease Audit data to hospital, critical care and mortality national data sets to enable research focused on quality improvement

Ferran Espuny Pujol ,[1] Christina Pagel ,[1] Katherine L Brown ,[2] James C Doidge ,[3] Richard G Feltbower ,[4] Rodney C Franklin,[5] Arturo Gonzalez-Izquierdo ,[6,7] Doug W Gould ,[3] Lee J Norman ,[4] John Stickley,[8] Julie A Taylor ,[1] Sonya Crowe [1]

To cite: Espuny Pujol F, Pagel C, Brown KL, *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. doi:10.1136/bmjopen-2021-057343

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2021-057343).

Received 25 September 2021
Accepted 22 April 2022

For numbered affiliations see end of article.

Correspondence to
Dr Ferran Espuny Pujol; f.pujol@ucl.ac.uk

ABSTRACT

Objectives To link five national data sets (three registries, two administrative) and create longitudinal healthcare trajectories for patients with congenital heart disease (CHD), describing the quality and the summary statistics of the linked data set.

Design Bespoke linkage of record-level patient identifiers across five national data sets. Generation of spells of care defined as periods of time-overlapping events across the data sets.

Setting National Congenital Heart Disease Audit (NCHDA) procedures in public (National Health Service; NHS) hospitals in England and Wales, paediatric and adult intensive care data sets (Paediatric Intensive Care Audit Network; PICANet and the Case Mix Programme from the Intensive Care National Audit & Research Centre; ICNARC-CMP), administrative hospital episodes (hospital episode statistics; HES inpatient, outpatient, accident and emergency; A&E) and mortality registry data.

Participants Patients with any CHD procedure recorded in NCHDA between April 2000 and March 2017 from public hospitals.

Primary and secondary outcome measures Primary: number of linked records, number of unique patients and number of generated spells of care. Secondary: quality and completeness of linkage.

Results There were 143 862 records in NCHDA relating to 96 041 unique patients. We identified 65 797 linked PICANet patient admissions, 4664 linked ICNARC-CMP admissions and over 6 million linked HES episodes of care (1.1M inpatient, 4.7M outpatient). The linked data set had 4 908 153 spells of care after quality checks, with a median (IQR) of 3.4 (1.8–6.3) spells per patient-year. Where linkage was feasible (in terms of year and centre), 95.6% surgical procedure records were linked to a corresponding HES record, 93.9% paediatric (cardiac) surgery procedure records to a corresponding PICANet admission and 76.8% adult surgery procedure records to a corresponding ICNARC-CMP record.

Conclusions We successfully linked four national data sets to the core data set of all CHD procedures performed between 2000 and 2017. This will enable a much richer analysis of longitudinal patient journeys and outcomes. We hope that our detailed description of the linkage process will be useful to others looking to link national data sets to address important research priorities.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ We linked five national established, high-quality, data sets using bespoke methods for the preprocessing of identifiers and establishing matches to maximise linkage.
⇒ In our final data set, data consistency has been checked at patient level using year and month of birth, postcodes and diagnosis codes, and also clinically sense checked at spell level for spells containing congenital heart procedures.
⇒ We created meaningful spells of care for each patient in the data set covering inpatient and outpatient interactions with secondary and tertiary care, covering up to 20 years of life of patients with congenital heart disease (CHD), representing an important step to understanding patient care for people with CHD.
⇒ Data completeness, quality and availability were worse in earlier years, meaning that linkage was poorer for earlier eras.
⇒ We do not yet have data on hospital care for patients outside England or on longer term adult follow-up for patients whose full CHD history is captured, since most cardiac procedures start in early live—the national CHD audit started on April 2000.

INTRODUCTION

Measuring, reporting and learning from patient outcomes should drive quality improvement (QI), but this is particularly challenging for lifelong conditions where outcomes need to be interpreted in the context of different phases of treatment,

changing treatment options, changing service provision and the natural evolution of disease.[1 2] Given the complex longitudinal care trajectories of such patients, rich data sets and careful multidisciplinary analysis are required to understand how patients interact with health services and to identify relevant outcomes and meaningful variations. These then provide opportunities for more targeted QI. Services for congenital heart disease (CHD) provide one such example. They span a patient's lifetime, but their quality in the UK is mainly measured by 30-day survival following children's heart surgery or catheter-based procedures. This is no longer a sufficient proxy and a more sophisticated approach is required.[3]

Information on patients with CHD, and their utilisation of specialised care services in England and Wales, is not available in a single data set. Since April 2000, the main source of information on the early outcomes of therapeutic paediatric and congenital cardiovascular procedures for patients with CHD in UK has been the *National Congenital Heart Disease Audit* (NCHDA).[4 5] Submission is mandatory for all centres and data quality is subjected to external validation. The key feature of this data set is the detailed recording of cardiac-related diagnosis and procedural information using the *European Paediatric and Congenital Cardiac Code* short list descriptors.[6]

By linking NCHDA with other national data sets, both validated registries and administrative, we aimed to build a unique combined data set for understanding patient journeys through the secondary and tertiary healthcare system. The four relevant national data sets are the Paediatric Intensive Care Audit Network (PICANet) for patient admissions to paediatric intensive care units (PICU)[7]; the case mix programme (CMP) from the Intensive Care National Audit & Research Centre (ICNARC-CMP) for patient admissions to adult intensive care units[8]; death registrations from the Office for National Statistics (ONS); hospital episode statistics (HES) routine administrative data on admitted patient care (APC), accident and emergency (A&E) attendances and outpatient (OP) appointments at National Health Service (NHS) hospitals in England.[9 10]

The research project 'LAUNCHES QI: Linking Audit and National data sets in Congenital Heart Services for Quality Improvement' aims to: describe patient trajectories through secondary and tertiary care; identify useful metrics for driving QI and informing commissioning and policy; explore variation across services to identify priorities for QI. In this paper, our objective is to describe the methods used to link the NCHDA data to HES, ONS, PICANet and ICNARC-CMP data sets and report the general characteristics, strengths and limitations of the resulting LAUNCHES data set. The process and challenges involved in the application for the approvals needed to link the LAUNCHES data sets have been described elsewhere.[11]

METHODS

Data

The core data set in LAUNCHES is NCHDA,[4 5] from which we obtained data for all records between 1 April 2000 and 31 March 2017 (figure 1). Each record relates to a single CHD procedure carried out in public hospitals in England and Wales. Most patients are resident in England and Wales, but patients from Northern Ireland and Scotland and overseas are also represented. NCHDA provides detailed demographic, diagnosis and procedural information for CHD procedures in children and adults

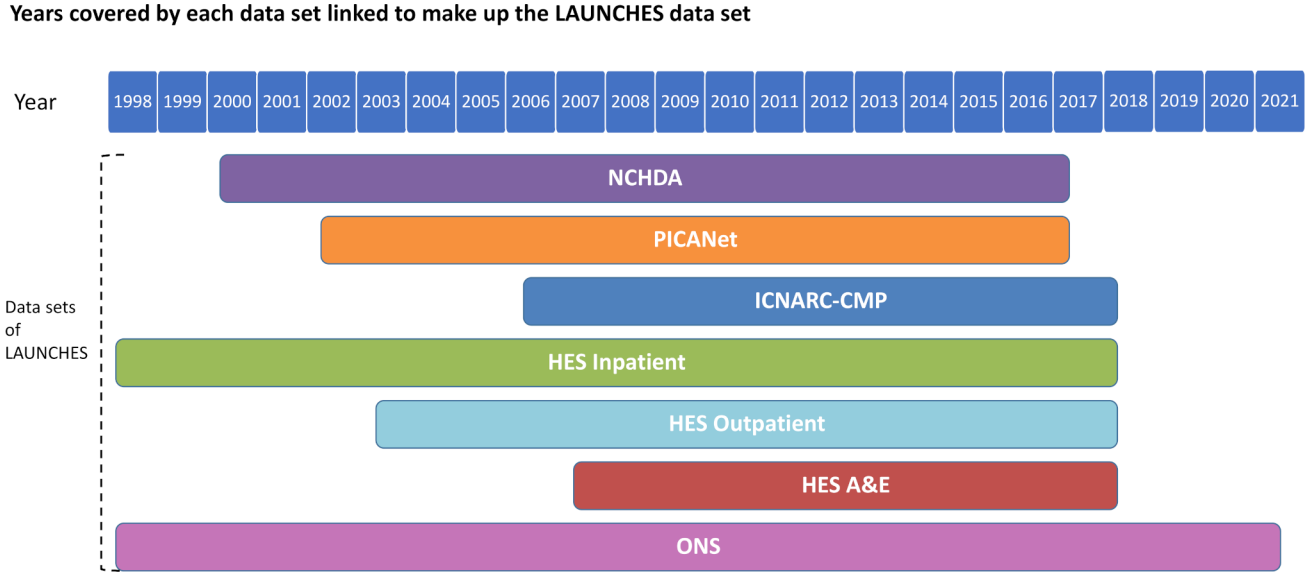

Figure 1 Data sets and years covered to make up the LAUNCHES data set. Calendar years are displayed at the top of this figure, while the data were obtained by financial years, which run from 1 April to 31 March. A&E, accident and emergency; HES, hospital episode statistics; ICNARC-CMP, Intensive Care National Audit & Research Centre Case Mix Programme; NCHDA, National Congenital Heart Disease Audit; ONS, Office for National Statistics (mortality); PICANet, Paediatric Intensive Care Audit Network.

as well as short-term survival outcomes (in-hospital and at 30 days).[12] Online supplemental table S1 contains all NCHDA fields that we obtained for LAUNCHES.

We applied to link to the following HES data sets (figure 1): APC inpatient (not limited to cardiac) admissions to hospitals in England between financial years 1998/1999 (starting 1 April 1998) and 2017/2018 (ending 31 March 2018); HES OP appointments between financial years 2003/2004 (first year available) and 2017/2018; HES A&E attendances between financial years 2007/2008 (first year available) and 2017/2018.[9 10 13] Online supplemental tables S2–S4 contain all HES fields that we obtained for LAUNCHES.

The ONS mortality data are the most complete source for the assessment of patient survival, recording all deaths registered in England and Wales.[14] Linked to HES data,[15] we obtained the ONS life status of patients of patients resident in England and Wales. See online supplemental table S5 for all ONS fields.

The PICANet contains records for all children admitted to PICU within UK and Ireland.[7] We requested all PICANet admissions in England up to March 2017 that could be linked to records in NCHDA (see online supplemental table S6 for all PICANet fields).

The CMP collects data from adult general critical care units in England, Wales and Northern Ireland.[8 16] We requested all ICNARC-CMP admissions up to August 2018 that could be linked to records in NCHDA (see online supplemental table S7 for all ICNARC-CMP fields).

The selected HES years correspond to all years of HES data with available HES identifiers (HES IDs) and NHS numbers (see HES Data Dictionary[17]) at the application time, where HES APC year 1997/1998 was not requested because we were informed that NHS numbers were largely missing (55.5%).

No dates of patient events were requested, other than year and month of birth (online supplemental tables S1–S6). Instead, ages (in years) to 4 decimal places at each event were requested from data providers to facilitate construction of detailed healthcare trajectories (enabling ordering of multiple events on the same day) while minimising identifiability of the linked data.

Table 1 Identifiers used for linkage

Identifier	Description and processing undertaken	Data set
NHS number	NHS numbers are 10-digit identifiers assigned to people registered for NHS care in England, Wales, or the Isle of Man. They are assigned to patients soon after birth (since year 2002) or the first time they receive NHS care or treatment.[30] *Processing*: removed non-numeric characters and blanks. *Invalid values:* 10-digit numbers that are all the same; dummy value '2333455667'; format 'n00000000n' (eg, '6000000006').[15] *Valid values*: Not invalid (above) and satisfying the checksum digit check.[31]	NCHDA, PICANet, ICNARC-CMP, HES/ ONS
Hospital patient ID	Hospitals use their own local patient identifiers, which in combination with the centre ID constitute a unique patient identifier that we refer to as 'hospital patient identifier'. A patient can have multiple hospital identifiers across their records for example, associated with care in different hospitals at different times. *Processing:* standardised the centre ID values, and removed blanks, leading zeroes and leading/ trailing special characters from the local patient identifiers.[15] *Valid values:* any value was considered valid.	NCHDA, PICANet
Date of birth (DoB)	Date of birth of the patient is available as recorded in the data sets *Processing:* standardised the format to day/month/year (eg, 17/11/2007). *Invalid values:* Any date after 01/04/2017 or before 01/01/1895. Equal to either 01/01/1901 or 31/12/1899.[15] *Valid values:* Not invalid (see above) and a feasible date.	NCHDA, PICANet, ICNARC-CMP, HES/ ONS
Name/surname	*Processing:* converted to upper case; removed prefixes and titles (eg, MISS, MSTR, MASTER, MRS, MS, MR, MAST, DR, SGT, SHEIKHA, SULTANA, SHEIKH, SULTAN), removed generic values (eg, BABY, INFANT, TWIN, TRIPLETS, BOY, GIRL, NAME1, NAME2). Removed special characters (apostrophes and accents). *Valid values:* non-empty values (after processing the fields).	NCHDA, PICANet
Postcode	*Processing:* converted to uppercase, removed blanks and special characters (only alphanumeric characters allowed). *Valid values:* postcodes included in the historical list of postcodes from the Organisation Data Service[32] and not corresponding to country postcodes (starting with 'ZZ') and not from an NHS trust site.[33]	NCHDA, PICANet, ICNARC-CMP, HES/ ONS

HES, hospital episode statistics; ICNARC-CMP, Intensive Care National Audit & Research Centre Case Mix Programme; NCHDA, National Congenital Heart Disease Audit; ONS, Office for National Statistics (mortality); PICANet, Paediatric Intensive Care Audit Network.

Data identifiers used for linkage

Table 1 lists the identifiers used for linkage, the data sets each were present in, and any prelinkage processing that was undertaken. NHS numbers have some limitations,[18 19] particularly that they are likely to be missing for overseas patients or those from Scotland and Northern Ireland. Hospital identifiers are unique to a patient, and records with the same hospital identifier will relate to the same patient. But hospital identifiers change between hospitals and so are not useful for linking patient records across different hospitals. In the absence of a matching NHS number or hospital patient identifier, we used date of birth, name and postcode to identify records as pertaining to the same patient but only if all three matched across records. We categorised the quality of each identifier for each record as: valid (for linkage), invalid or missing (table 1).

Linkage method

We developed an algorithm to link NCHDA data both internally (to identify records pertaining to the same person within NCHDA) and externally, to records in the other data sets. Our hierarchical method, shown in figure 2, treated NHS number and hospital patient ID as primary identifiers, while date of birth, patient name and postcode were treated as weaker identifiers. The possible linkage states when comparing a processed identifier across two records were:

- Exact agreement, if each identifier was valid and they were exactly the same.
- Partial agreement only used for valid dates of birth and names and defined in detail below.
- Any missing, if either or both identifiers were missing or invalid.

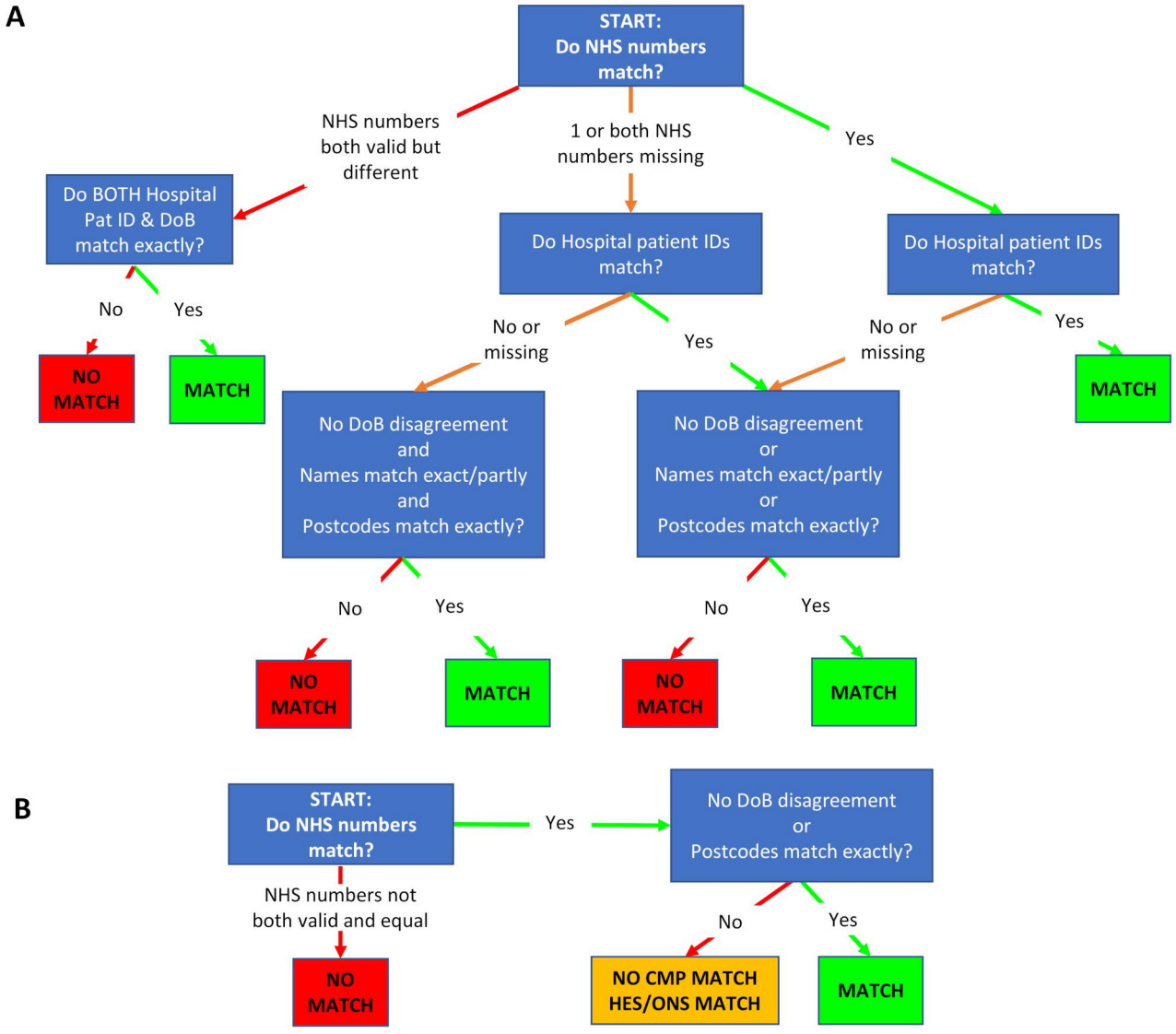

Figure 2 The linkage algorithm for deciding whether two records pertain to the same patient. A: linkage of NCHDA records internally and to PICANet records. B: linkage of NCHDA to ICNARC-CMP and to HES/ONS. 'No DoB disagreement' means that the dates of birth either match (exactly or partially) or one or both of those dates are missing. HES, hospital episode statistics; ICNARC-CMP, Intensive Care National Audit & Research Centre Case Mix Programme; NCHDA, National Congenital Heart Disease Audit; ONS, Office for National Statistics (mortality); PICANet, Paediatric Intensive Care Audit Network.

- Disagreement, if both values were valid and non-missing but did not match (exactly or partially).

Two valid dates of birth (DoB) were considered to be in partial agreement if either: the two DoB values were no more than 5 days apart; the two DoB values were not the same, but either two components (ie, YYYY, MM or DD) of the two DoB values matched or two components of the two DoB values matched when the MM and DD parts of one of them were swapped. Partial agreement of names occurred between two records if there were previous and current versions of names and at least one matched the other record.

An auxiliary lookup table (online supplemental table S8) between NCHDA organisations and PICUs was used by PICANet when comparing hospital patient identifiers as part of the NCHDA to PICANet linkage (figure 2A), given that the two data sets use different names for centres.

For NCHDA to ICNARC-CMP linkage, two records were matched by ICNARC only if there was exact agreement of NHS numbers and either the DoB did not disagree or postcodes matched exactly (figure 2B). NCHDA to HES/ONS linkage was performed by NHS Digital and required the exact match of NHS numbers (agreement in postcode was reported but not required). See online supplemental table S9 for the HES/ONS linkage method.

Finally, note that all linkages were done at record level. This resulted in many-to-many record matches that were resolved to identify records as pertaining to the same patient across all five data sets once pseudonymised data sets had been received at University College London (UCL).

Data flows

Record-level patient identifiers in the core data set (NCHDA) were sent for linkage via secure transfer to each of the three data controllers for the other four data sets, along with a study-specific pseudonymised record identifier. Each data controller then searched for records within their data sets with matching patient identifiers and returned the pseudonymised, clinical data (without patient identifiers) for all records that had at least one match to an NCHDA record to UCL Clinical Operational Research Unit. We used secure transfer and all data are stored in the UCL data safe haven, which complies with the NHS Information Governance Toolkit. Only pseudonymised study-specific record and patient IDs were shared with or stored at UCL. Linkage results were provided as lists of corresponding pairs of records with a code indicating the quality of linkage for each record-to-record match (concatenated agreement category for each identifier).

Patient-level consistency and quality assurance

The national audit body (National Institute for Cardiovascular Outcomes Research; NICOR) identified unique patients within the NHCDA using the linkage algorithm and then checked for inconsistencies on site as part of data quality assurance. Inconsistencies in DoB (missing values, procedures before birth, different DoB for a same patient) were identified and sent to submitting hospitals for correction and were then revised by NICOR. Cleaned record identifiers were then sent for linkage to the other data processors. An additional internal detailed clinical review was undertaken of pairs of records that were not linked but similar to some extent (eg, those pairs solely agreeing in NHS number) and pairs of records linked but with only moderate agreement in identifiers (eg, pairs with matched names, DoB and postcode but NHS numbers missing) and internal patient categorisation updated.

Both HES and PICANet have their own internal unique patient IDs across records. Pseudonymised versions of these were included in the returned records. We then assessed the level of agreement between the identified patients from the NCHDA and patient identifiers from the linked PICANet and HES data sets. PICANet and HES patients linked to more than one LAUNCHES patient were discussed with each processor and patient categorisation was revised on a case-by-case basis. Numbers of records and patients before, during and after quality assurance will be reported, together with available years of follow-up.

Spells of care and completeness of linkage

Once the linked data set was created, we combined overlapping events into ‘spells of care’. Gaps of less than 24 hours were considered to be overlapping, since times of events were not routinely collected and so records could have a 12-hour uncertainty in either direction. Figure 3 illustrates an example of event records that would be

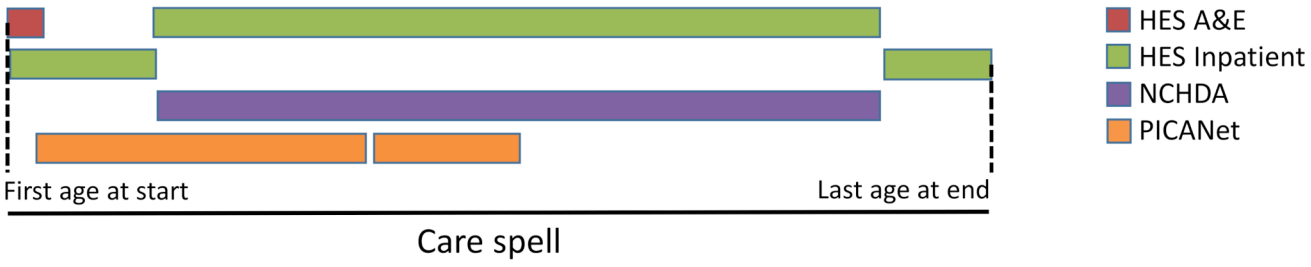

Figure 3 Example of Care spell consisting of several time-overlapping events involving different services. A&E, accident and emergency; HES, hospital episode statistics; ICNARC-CMP, Intensive Care National Audit & Research Centre Case Mix Programme; PICANet, Paediatric Intensive Care Audit Network.

combined into a single (paediatric) spell. Number of spells per year/patient/data set will be reported.

Cardiac surgeries typically require intensive care recovery. Catheter-based interventions and diagnostic procedures are far less likely to require ICU admission. Our first consistency check was to look at how many spells containing a cardiac surgery procedure also contained an accompanying ICU stay, enabling an assessment of the completeness of linkages from NCHDA to PICANet and NCHDA to ICNARC-CMP. While we would not expect 100% of NCHDA surgeries to have a linked record, we would expect a high proportion to. A second consistency check was for HES linkage completeness. We would expect a HES-linked record (either inpatient admission or OP attendance) to be part of the same spell as any NCHDA procedure, as long as the NHCDA record had a valid NHS number. In addition, at least one of the ICD-10 diagnostic codes used within HES for inpatient admissions should denote CHD for HES records linked to NCHDA surgical procedures (a list of valid congenital codes and other cardiac non-congenital codes that are sometimes used for patients with CHD is provided in online supplemental table S10). Summary statistics will be provided on the completeness of linkage per data set and the clinical sense checking of HES linked data.

Patient and public involvement statement

We have patient and public representatives on the independent study advisory group. The advisory group was consulted on linkage design and execution and approved the process.

RESULTS

Quality of identifiers in each data set

The NCHDA data set contained 143 862 CHD records of which 94.7% had valid NHS numbers. Unsurprisingly, the percentage of valid NHS numbers was higher for patients with residence in England (98.8%) or Wales (99.1%) as determined by their postcode at the time of procedure. The breakdown of NHS numbers by residence is given in online supplemental table S11. PICANet records for patients born before 14 October 2001 were available only if they had a PICANet event between 14 October 2014 and 13 October 2019, due to the terms of the PICANet Health Research Authority (HRA) Confidentiality Advisory Group (CAG) approval for processing identifiable information.[20] There were 179 791 PICANet records available for linkage, of which 90.5% had valid NHS numbers. Hospital patient identifiers were available for 100% of NCHDA and PICANet records, as were DoB; names/surnames were available for 99.6% and 98.9% of records, respectively, and postcodes were valid for 95.0% and 97.2% of records. ICNARC-CMP had 1 853 568 records of which 88.7% had valid NHS numbers. The total of records and percentage of valid NHS numbers for HES data were: 314 445 082 (93.8%) for HES inpatient, 1 288 711 692 (98.0%) for HES OP and 194 572 279 (93.3%) for HES A&E. We did not know the quality of identifiers in ONS mortality data, which we obtained linked to HES data. The quality of the identifiers improved over time (online supplemental table S12).

Linked data sets before quality assurance

There were 6 408 673 records across the final component data sets before any quality assurance was carried out (online supplemental table S13), with each non-NCHDA record linked to at least one NCHDA record.

Quality of the record-level linkage

The use of a bespoke method for linking NCHDA-NCHDA and NCHDA-PICANet records (figure 2A) allowed us to identify more linked records than had we relied solely on NHS numbers:

- 95.0% of the NCHDA-NCHDA matches and 92.3% of the NCHDA-PICANet matches were identified by an exact agreement of NHS numbers.
- 4.9% of the NCHDA-NCHDA and 7.0% of the NCHDA-PICANet matches were identified by exact agreement in hospital patient identifiers (allowing for missing NHS number).
- 0.1% of the NCHDA-NCHDA and 0.7% of the NCHDA-PICANet matches were identified by other options of our bespoke linkage algorithm.

Patient-level results

There were 47 753 internal NCHDA-linked records (out of a total of 143 862 NCHDA records), representing patients with more than one recorded procedure within the NCHDA data set.

Once patients had been defined across NCHDA records, 649 inconsistencies in DoB affecting 219 patients were detected and corrected. There was a very high level of agreement between the identified patients from the linked PICANet data and the LAUNCHES linkage definition of patients: only seven PICANet patients (0.0% of the 34 507 linked PICANet patients) were linked to two LAUNCHES patients each. Investigation of those cases by each audit resulted in a further minor revision. In a similar exercise, we excluded 88 HES IDs (0.1% of the total 89 098 linked HES IDs) that were linked to two LAUNCHES patient IDs each. It was not possible to determine which HES records corresponded to each patient (mainly because they pertained to twins). Inconsistencies between 42 HES and NCHDA patients linked with disagreement in year–month of birth and postcode were also resolved.

This detailed review of linked NCHDA records resulted in a final total of 96 041 unique patients with a total of 6 381 600 records (table 2). Of those, 66 453 patients (69.2%) had at least one NHCDA record as children (age at procedure under 16), whereas the remaining 29 588 patients (30.8%) had all their NHCDA records as adults.

A total of 90 678 patients (94.5%) were linked to at least one external data set: 91.5% of patients had some form of HES/ONS record, 35.9% had at least one linked PICANet

Table 2 Number of linked records in each data set after quality assurance, by estimated financial year

Financial year	NCHDA	PICANet	ICNARC-CMP	HES inpatient	HES outpatient	HES A&E	Total
1998	0	0	0	16431	0	0	16431
1999	0	0	0	19811	0	0	19811
2000	6421	15	2	29113	0	0	35551
2001	6161	11	1	33210	0	0	39383
2002	6137	952	0	36870	0	0	43959
2003	7402	3226	0	42805	132364	0	185797
2004	6968	3464	0	45314	149544	0	205290
2005	7684	3828	0	50097	176383	0	237992
2006	8152	4052	6	52001	195655	0	259866
2007	7984	4136	154	56577	223402	23268	315521
2008	8294	4275	215	59782	254476	27482	354524
2009	8719	4748	273	65190	292972	32732	404634
2010	8987	4891	388	69084	322196	35862	441408
2011	9102	5103	407	70564	347096	38854	471126
2012	9013	5176	411	70908	368160	41598	495266
2013	9593	5435	473	71781	406805	42830	536917
2014	9639	5435	447	72751	440554	44913	573739
2015	11492	5546	629	75959	468434	47219	609279
2016	12114	5504	686	72899	476727	46885	614815
2017	0	0	572	51814	424473	43432	520291
All years	143862	65797	4664	1062961	4679241	425075	6381600

Financial years (running from April to March) were estimated using the ages at events and the estimated date of birth (we took day 15th of the known month of birth as date of birth). The estimation is likely wrong for 27 records from PICANet and ICNARC-CMP with estimated year pre-2002, but we could not fix the needed ages or dates of birth at the time of submission (such inconsistencies are likely to be excluded in future analyses).
A&E, accident and emergency; HES, hospital episode statistics; ICNARC-CMP, Intensive Care National Audit & Research Centre Case Mix Programme; NCHDA, National Congenital Heart Disease Audit; PICANet, Paediatric Intensive Care Audit Network.

record and 3.6% had at least one linked ICNARC-CMP record. The main reasons for non-linkage of the remaining 5363 patients (5.6% of all NCHDA patients) were: missing NHS number; residence not recorded or outside England; and/or record from before 2003 when data quality was poorer. The final linked data set covers up to 20 years of life of patients, with a median (IQR) coverage of 12 (6, 16) years for 87735 patients with no known age of death and 4 (1, 13) years for 8306 patients with known age of death.

Spell-level results

We identified 4908153 spells of care for the 96041 patients in the LAUNCHES data set. Only 2.6% of the spells contained at least one NCHDA procedure compared to the 99.7% of spells that included at least one HES record (799890 inpatient spells in total). Only 1.0% of spells included at least one PICANet record, and 0.1% of spells included at least one ICNARC-CMP record. Patients had a median (IQR) of 3.4 (1.8, 6.3) spells per year, with a median (IQR) of 0.1 (0.1, 0.3) spells with NCHDA procedures per year. This high level of healthcare interaction was expected in this population, since patients with CHD require regular specialist follow-up.

Sense checking the completeness of the linkage

PICANet

Out of all paediatric cardiac surgeries, 93.9% (42 512/45 265) were linked to an associated PICANet record where linkage was in principle feasible. The corresponding percentage for paediatric catheter-based procedures was 11.2% (2047/18 268).

ICNARC-CMP

Out of all adult cardiac surgeries (resp catheters), 76.8% (906/1180) (resp 2.6%: 69/2610) were linked to ICNARC-CMP when the procedures were post-March 2009 at centres submitting regularly to ICNARC, and where a valid NHS number was recorded. Unfortunately, many hospitals carrying out congenital heart procedures submitted very few records to ICNARC-CMP over the time period of this study. This means that for all cardiac surgeries where ICNARC-CMP data would have been available (post 2009 with a valid NHS number), only

16.5% (1193/7234) were linked to an associated CMP record.

HES/ONS

Out of all NHCDA procedure records (either surgical or catheter) with a valid NHS number and performed in an English public hospital, 95.6% (122 278/127 932) were linked to an associated HES record, mostly inpatient records. ONS age at death was provided for 7228 patients. In a total of 53 769 spells which included both NCHDA surgical procedures and an associated HES inpatient record, 94.6% of HES records had CHD ICD-10 diagnostic codes from online supplemental table S10, 3.8% had only acquired heart diagnoses (plausible miscoding of CHD) and 1.6% had other diagnostic codes.

These consistency checks provide assurance that, where linkage was theoretically possible, we achieved excellent linkage.

DISCUSSION

Principal findings

We have described a bespoke linkage algorithm, alongside quality, completeness and consistency checks, which we used to identify 96 041 unique patients across 143 862 NCHDA cardiac procedure records and to link their records to 65 797 PICU admissions, 4664 adult intensive care admissions and 6 167 277 HES (inpatient, OP and A&E) records.

While most of the linked records were identified using matching NHS numbers, a significant proportion (around 5%) was identified using other identifiers, highlighting the value of using additional identifiers. Close collaboration with each audit and NHS Digital meant that we could further check the quality of the linkage and further refine the identification of unique patients across records, improving the overall quality of the linked data set.

The quality of recorded identifiers used for linkage improved markedly over time as did the quality of resulting linkage. 90 678 (94.5%) patients had records that were linked to at least one other data set. We identified 4 908 153 spells of care for the 96 041 patients. The final linked data set (6 381 600 records) covers up to 20 years of life of patients, with a median (IQR) coverage of 12 (6,16) years for 87 735 patients with no known age of death, and 4 (1, 13) years for 8306 patients with known age of death.

Patients had a median (IQR) of 3.4 (1.8, 6.3) spells of care (either an inpatient stay or an OP event) per year. This frequent interaction with secondary and tertiary care outside of NCHDA procedures (only 2.6% spells of care included an NCHDA procedure) highlights the necessity and value of linking specialised validated procedure-based registry records (NCHDA) to other administrative and audit data sets to understand and potentially improve services for CHD.[21 22]

Strengths and weaknesses

All linked data sets were national established, high-quality, data sets. We designed a bespoke linkage method and data processors carefully prepared the identifiers for linkage in a consistent way to maximise matching. In our final data set, data consistency has been checked at patient level using year and month of birth, postcodes and diagnosis codes and also clinically sense checked at spell level for spells containing congenital heart procedures.

Each of the data sets used for linkage was available for different years. Additionally, PICANet's HRA CAG policy of data anonymisation restricted linkage feasibility for some patients, HES data only covered hospitals in England and ICNARC-CMP data set was of limited utility since many specialised adult cardiac intensive care units did not submit to ICNARC-CMP for most or all of the time period. More adult cardiac ICUs submit to ICNARC-CMP every year and so future linkage should be much more complete.

The linked data set covers at most 20 years of life of patients. While this represents an important step to understanding patient care for people with CHD, we do not yet have data on longer term adult follow-up for patients whose full CHD history is captured (ie, those born after 2000), since most cardiac procedures start in early life.

Comparison with other studies

In the UK, the Infant Heart Study linked an NCHDA cohort to PICANet data to explore risk factors for poor outcomes (1 year) after hospital discharge for infants undergoing heart surgery between years 2005 and 2010.[23 24] ONS mortality was included as part of NCHDA at that time, and the linkage to PICANet was carried out using just NHS number. A study looking at differences in access to Emergency Paediatric Intensive Care and care during Transport linked together PICANet, ICNARC-CMP and HES/ONS. NHS numbers were the primary identifiers used for matching.[25–27] Our bespoke linkage algorithm improved the approach based on NHS numbers, with 7.7% of the total NCHDA-PICANet matches obtained using agreement in other identifiers.

Implications for clinicians and policymakers

The NHCDA database is highly specialised and procedure based. The linked intensive care and hospital data sets provide a much wider and more complete picture of the interactions CHD patients have with secondary and tertiary care throughout their lives. In particular, the OP data means loss to follow-up in transition from child to adult services and/or during adulthood can be explored. The linked data of validated registries with administrative databases will facilitate the identification of appropriate outcomes for reporting and routine monitoring CHD services at all ages, including resource utilisation, and to develop methods of QI that take into account differences in risk across case mix.[28]

Unanswered questions and future research

The NCHDA data set only contains information for CHD patients that have at least one procedure. This means that when considering overall health service journeys of people living with CHD, we miss those who never have a procedure (either because disease is considered too mild or because it is too severe for correction). The ongoing CHAMPION project will use the National Congenital Anomaly and Rare Disease Registration Service (NCARDRS) data set to estimate the number of children born with CHD or that have an antenatal diagnosis but do not survive pregnancy (termination or in-utero death).[28 29] In future, linkage to NCARDRS might allow assessment of outcomes and healthcare journeys for the complete patient cohort.

Conclusion

We successfully linked five national data sets to achieve a large, high-quality combined data set spanning 20 years that will allow rich exploration of the healthcare journeys of patients with CHD. We hope that this detailed description will be useful to others looking to link national data sets to address important research priorities. While challenging, researchers, data controllers and data processors should continue to encourage and facilitate data linkage to enable generation of valuable new knowledge and insights.

Author affiliations
[1]Clinical Operational Research Unit, Department of Mathematics, University College London, London, UK
[2]Cardiorespiratory Division, NIHR Great Ormond Street Hospital Biomedical Research Centre, London, UK
[3]Intensive Care National Audit and Research Centre, London, UK
[4]Leeds Institute for Data Analytics, School of Medicine, University of Leeds, Leeds, UK
[5]Department of Paediatric Cardiology, Royal Brompton & Harefield NHS Foundation Trust, London, UK
[6]Institute of Health Informatics, University College London, London, UK
[7]Health Data Research UK, London, UK
[8]Department of Paediatric Cardiac Surgery, Birmingham Children's Hospital, Birmingham, UK

Twitter Ferran Espuny Pujol @ferranespuny, Christina Pagel @chrischirp, James C Doidge @StraightStats, Richard G Feltbower @rgfeltbower and Lee J Norman @Normmy

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

Contributors CP and SC conceived of and led the study. FEP, CP, KLB, JCD, RGF, RCF, AGI, DWG, LJN, JS, JAT and SC contributed to the design of the bespoke linkage algorithm. FEP holds an honorary contract at NICOR and assisted NICOR in preprocessing identifiers used for linkage and creating an internally linked NCHDA database. R code was developed by FEP for the processing, quality assessment and linkage of NCHDA records. Audit collaborators at PICANet (LJN) and ICNARC (JCD) adapted the code to perform the linkage. FEP, CP, KLB, JCD, RGF, RCF, AGI, DWG, LJN, JS, JAT and SC contributed to quality and consistency assurance of the linkage and data set. The clinical sense checking of linked records and spells of care was performed by RCF and KLB. FEP wrote the first draft of the manuscript. FEP, CP, KLB, JCD, RGF, RCF, AGI, DWG, LJN, JS, JAT and SC edited, commented and approved the final draft. CP and SC are responsible for the overall content of this work as guarantors.

Funding This study is supported by the Health Foundation, an independent charity committed to bringing about better health and health care for people in the UK (Award number 685009). Katherine L. Brown benefited from funding received by The Great Ormond Street Hospital NIHR Biomedical Research Centre.

Competing interests None declared.

Patient and public involvement Patients and/or the public were involved in the design, or conduct, or reporting, or dissemination plans of this research. Refer to the Methods section for further details.

Patient consent for publication Not applicable.

Ethics approval LAUNCHES received ethical approval from the Health Research Authority (reference: IRAS 246796) and the Confidentiality Advisory Group (reference: 18/CAG/0180). These are nationally collected routine data and as such it is not feasible to retrospectively ask for consent. We obtained CAG approval for the use of these non-consented data sets for this research study. Confidentiality Advisory Group reference: 18/CAG/0180.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data may be obtained from a third party and are not publicly available. This paper describes the linkage of five national data sets and does not present results based on analysis of that data. The linked data are held and processed in the Data Safe Haven under strict governance requirements and signed data sharing agreements. It cannot be shared with others without significant amendments to ethics, CAG and data sharing agreements. The R code developed by FEP for the processing, quality assessment and linkage of NCHDA records is publicly available (GitHub site: https://github.com/fespuny/LAUNCHESQI_linkage).

ORCID iDs
Ferran Espuny Pujol http://orcid.org/0000-0001-9085-7400
Christina Pagel http://orcid.org/0000-0002-2857-1628
Katherine L Brown http://orcid.org/0000-0002-0729-4959
James C Doidge http://orcid.org/0000-0002-3674-3100
Richard G Feltbower http://orcid.org/0000-0002-1728-9408
Arturo Gonzalez-Izquierdo http://orcid.org/0000-0002-0984-5830
Doug W Gould http://orcid.org/0000-0003-4148-3312
Lee J Norman http://orcid.org/0000-0002-8684-8787
Julie A Taylor http://orcid.org/0000-0002-7698-6991
Sonya Crowe http://orcid.org/0000-0003-1882-5476

REFERENCES

1 Rogers L, Brown KL, Franklin RC, *et al*. Improving risk adjustment for mortality after pediatric cardiac surgery: the UK PRAiS2 model. *Ann Thorac Surg* 2017;104:211–9.
2 Rogers L, Pagel C, Sullivan ID, *et al*. Interventions and outcomes in children with hypoplastic left heart syndrome born in England and Wales between 2000 and 2015 based on the National congenital heart disease audit. *Circulation* 2017;136:1765–7.
3 NHS England. New congenital heart disease review: final report, 2015. Available: https://www.england.nhs.uk/wp-content/uploads/2015/07/Item-4-CHD-Report.pdf
4 NICOR. Congenital heart disease in children and adults (congenital audit). Available: https://www.nicor.org.uk/national-cardiac-audit-programme/congenital-heart-disease-in-children-and-adults-congenital-audit/ [Accessed 15 May 2022].
5 Franklin R, Wang J, Ajayi S. National congenital heart disease audit. 2020 summary report (2018/19 data). Healthcare quality improvement programme (HQIP) 2020.
6 Franklin RCG, Anderson RH, Daniëls O, *et al*. Report of the coding Committee of the association for European paediatric cardiology. *Cardiol Young* 2002;12:1–8.

7 Universities of Leeds & Leicester. PICANet – paediatric intensive care audit network for the UK and ireland. Available: https://www.picanet.org.uk/ [Accessed 15 May 2022].
8 Harrison DA, Brady AR, Rowan K. Case mix, outcome and length of stay for admissions to adult, general critical care units in England, Wales and Northern Ireland: the Intensive Care National Audit & Research Centre Case Mix Programme Database. *Crit Care* 2004;9:cc3745
9 Herbert A, Wijlaars L, Zylbersztejn A, *et al*. Data resource profile: Hospital episode statistics admitted patient care (Hes APC). *Int J Epidemiol* 2017;46:1093–1093i.
10 Boyd A, Cornish R, Johnson L. *Understanding Hospital episode statistics (HES)*. London, UK: CLOSER, 2018. https://www.closer.ac.uk/wp-content/uploads/CLOSER-resource-understanding-hospital-episode-statistics-2018.pdf
11 Taylor JA, Crowe S, Espuny Pujol F, *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.
12 White O, Stickley J. National congenital heart disease audit. data manual for dataset version 6.1 – March 2020 revision 2020.
13 NHS Digital. Hospital episode statistics (HES). Available: https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics [Accessed 15 May 2022].
14 NHS Digital. Linked HES-ONS mortality data. Available: https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/linked-hes-ons-mortality-data [Accessed 15 May 2022].
15 Health and Social Care Information Centre. *A guide to linked mortality data from hospital episode statistics and the office for national statistics*. Health and Social Care Information Centre, 2015.
16 ICNARC. About the CMP. Available: https://www.icnarc.org/Our-Audit/Audits/Cmp/About [Accessed 15 May 2022].
17 NHS Digital. Hospital episode statistics data dictionary. Available: https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics/hospital-episode-statistics-data-dictionary [Accessed 15 May 2022].
18 Moser K, Hilder L. Assessing quality of NHS numbers for babies data and providing gestational age statistics. *Health Stat Q* 2008:15–23.
19 Primary Care Support England. Adoption and gender re-assignment processes. Available: https://pcse.england.nhs.uk/help/registrations/adoption-and-gender-re-assignment-processes/ [Accessed 15 May 2022].
20 PICANet. Policy of data anonymisation. Available: https://www.picanet.org.uk/wp-content/uploads/sites/25/2019/11/PICANet-ongoing-data-anonymisation.pdf [Accessed 15 May 2022].
21 Pasquali SK, Peterson ED, Jacobs JP, *et al*. Differential case ascertainment in clinical Registry versus administrative data and impact on outcomes assessment for pediatric cardiac operations. *Ann Thorac Surg* 2013;95:197–203.
22 Jacobs JP, Franklin RCG, Béland MJ, *et al*. Nomenclature for pediatric and congenital cardiac care: unification of clinical and administrative nomenclature - The 2021 International paediatric anc congenital cardiac code (IPCCC) and the eleventh revision of the International classification of diseases (ICD-11). *Cardiol Young* 2021;31:1057–188.
23 Crowe S, Ridout DA, Knowles R, *et al*. Death and emergency readmission of infants discharged after interventions for congenital heart disease: a national study of 7643 infants to inform service improvement. *J Am Heart Assoc* 2016;5:e003369.
24 Brown KL, Wray J, Knowles RL. *Infant deaths in the UK community following successful cardiac surgery: building the evidence base for optimal surveillance, a mixed-methods study*. Southampton, UK: NIHR Journals Library, 2016.
25 Ramnarayan P, Evans R, Draper ES, *et al*. Differences in access to emergency paediatric intensive care and care during transport (DEPICT): study protocol for a mixed methods study. *BMJ Open* 2019;9:e028000.
26 Seaton SE, Ramnarayan P, Davies P, *et al*. Does time taken by paediatric critical care transport teams to reach the bedside of critically ill children affect survival? A retrospective cohort study from England and Wales. *BMC Pediatr* 2020;20:301.
27 Seaton SE, Ramnarayan P, Pagel C, *et al*. Impact on 30-day survival of time taken by a critical care transport team to reach the bedside of critically ill children. *Intensive Care Med* 2020;46:1953–5.
28 CHAMPION project, NIHR PR-R20-0318-23001. Available: https://fundingawards.nihr.ac.uk/award/PR-R20-0318-23001 [Accessed 15 May 2022].
29 National congenital anomaly and rare disease registration service (NCARDRS). Available: https://www.gov.uk/government/collections/national-congenital-anomaly-and-rare-disease-registration-service [Accessed 15 May 2022].
30 What is an NHS number? 2018. Available: https://www.nhs.uk/using-the-nhs/about-the-nhs/what-is-an-nhs-number/ [Accessed 15 May 2022].
31 NHS number. Available: https://datadictionary.nhs.uk/attributes/nhs_number.html [Accessed 15 May 2022].
32 NHS Digital. Office for national statistics data. Available: https://digital.nhs.uk/services/organisation-data-service/data-downloads/office-for-national-statistics-data [Accessed 15 May 2022].
33 NHS Digital. Other NHS organisations. Available: https://digital.nhs.uk/services/organisation-data-service/data-downloads/other-nhs-organisations [Accessed 15 May 2022].

Linkage of National Congenital Heart Disease Audit data to hospital, critical care and mortality national data sets to enable research focused on quality improvement

Supplemental Tables

Table S1 NCHDA fields obtained.

NCHDA field name	Description
1.00 Data Set Version	The version of the data set the data was collected for and submitted to NICOR
1.01 Hospital identifier	The identifier allocated to the hospital by NICOR. The software should set this field without any user involvement.
1.07 Patient Gender	Identifies the genotypical sex of the patient.
1.08 Patient Ethnic Group	Identifies the patient's ethnic origin.
1.09 Patient Admin status	Type of admission, i.e. from UK public health service or other mode of entry to the service
2.01 Diagnosis	The preprocedural diagnosis of the patient
2.02 Previous Procedure	Relevant previous procedures
2.03 Weight	The patients weight in kg at the time of procedure to two decimal places.
2.03b Height	Height at time of procedure in cm
2.04 Antenatal Diagnosis	Diagnosis detected prior to birth from prenatal scans
2.05 Preprocedure seizures	Any preprocedural convulsions/seizures requiring medication
2.06b Comorbidity present	A comorbidity is the presence of one or more additional disorders (or diseases) co-occurring with a primary disease or disorder; or the effect of such additional disorders or diseases.
2.07 Comorbid Conditions	Identifies the specific comorbid condition
2.08 Preprocedure systemic ventricular ejection fraction	Categorises the percentage of the blood emptied from the systemic ventricle at the end of the contraction. Data may have been derived from angiography, echocardiography, nuclear imaging, magnetic resonance imaging etc. Use this metric to define ventricular function in patients with functionally single ventricle anatomy.
2.09 Preprocedure subpulmonary ventricular ejection fraction	Categorises the percentage of the blood emptied from the subpulmonary ventricle at the end of the contraction. Data may have been derived from angiography, echocardiography, nuclear imaging, magnetic resonance imaging etc. Do not use this metric for patients with functionally single ventricle anatomy.
3.01b Procedure urgency	Categorises the patient in terms of the urgency
3.01c Unplanned reoperation	Used to identify cases that aren't part of the planned pathway for that patient.
3.04 First operator grade	The grade of the secondary operator or assistant.
3.06 First assistant grade	The grade of the secondary operator or assistant.
3.07 Type of Procedure	Defines the group the procedure should be included in.
3.08 Sternotomy Sequence	Incremental count of the number of sternotomies that the patient has undergone.
3.09 Operation performed	The EPCC short codes that describe the procedure
3.10 Total bypass time	The total duration of cardiopulmonary bypass used during the procedure.
3.11 Total bypass cross clamp time	The total duration of aortic cross clamp during the procedure.
3.12 Total circulatory arrest time	The total duration of circulatory arrest during the procedure.
3.13 Catheter procedure duration	The operative time taken.
3.14 Total fluoroscopy time	The total time fluoroscopy was used during the procedure
3.15 Total fluoroscopy dose	The total fluoroscopy dose during the procedure

Espuny Pujol F, *et al. BMJ Open* 2022; 12:e057343. doi: 10.1136/bmjopen-2021-057343

3.16 Procedure Report Or Comment	Accompanying text that can help describe the procedure in cases where coding is thought to be inadequate.
4.03 Discharge status	The status of the patient at discharge from your hospital.
4.04 Discharge Destination	The immediate destination following discharge from your hospital
4.05 Postprocedure seizures	Any postprocedural convulsions/seizures requiring medication
4.07 Duration of postoperative intubation	Duration of postoperative intubation associated with a procedure.
4.08 Postoperative complications	Significant postoperative complications following surgery
4.09 Attribution of death	The attribution of death to a procedure
5.01 Device Manufacturer	The manufacturer of any implanted devices.
5.02 Device model	The model numbers of any implanted device.
5.04 Device Size	The size of any devices implanted.
6.01 Preprocedure NYHA status	The patient's preprocedural NYHA status.
6.02 Preprocedure smoking status	The patient's preprocedural smoking status
6.03 Preprocedure diabetes	The patient's preprocedural diabetes status
6.04 History of pulmonary disease	The patient's preprocedural pulmonary disease status
6.06 Preprocedural ischaemic heart disease	The patient's preprocedural ischaemic heart disease status
7.01 Preprocedural valve or septal defect or vessel size	The preprocedural size of the valve or septal defect or vessel size
7.02 Sizing balloon used for septal defect closure Y/N	Was a sizing balloon used for septal defect occlusion
7.03 Number of stents or coils	The number of stents and/or coils deployed
7.04 Catheterisation complication severity rating	Classifies the severity of the most major catheter complication.
7.05 Catheterisation complications	Significant postprocedural complications following a cardiac catheter
LAUNCHES derived field name	**Description**
LAUNCHESrecID	LAUNCHES record identifier
LAUNCHESpatID	LAUNCHES patient identifier
LAUNCHESpatID_rev	LAUNCHES patient identifier revised
Qcode	Record identifiers' quality code
dob.year	Year of birth derived from *1.06 Patient Date of Birth*
dob.month	Month of birth derived from *1.06 Patient Date of Birth*
country	Country of patient's residence derived from *1.10 Patient Post Code*
IMD2004.rank	IMD2004 rank of patient's residence derived from *1.10 Patient Post Code*
IMD2004.decile	IMD2004 decile of patient's residence derived from *1.10 Patient Post Code*
IMD2007.rank	IMD2007 rank of patient's residence derived from *1.10 Patient Post Code*
IMD2007.decile	IMD2007 decile of patient's residence derived from *1.10 Patient Post Code*
IMD2010.rank	IMD2010 rank of patient's residence derived from *1.10 Patient Post Code*
IMD2010.decile	IMD2010 decile of patient's residence derived from *1.10 Patient Post Code*
IMD2015.rank	IMD2015 rank of patient's residence derived from *1.10 Patient Post Code*
IMD2015.decile	IMD2015 decile of patient's residence derived from *1.10 Patient Post Code*
Raop	Age in years (4 decimal places) at procedure derived from *3.01 Date/Time procedure*
RFLAGtop	Record FLAG: time of procedure present?
Raodis	Age in years (4 decimal places) at discharge derived from *4.01 Date of Discharge*
Raod	Age in years (4 decimal places) at death derived from *4.02 Date of Death*

Note: The NCHDA linked data and linkage quality report were received on 23 September 2019. The revised ages and anonymised patient key identifiers were received on 27 February 2019.

Table S2 HES inpatient (HES APC) fields obtained.

HES Inpatient field name	**Description**
ADMIFLAG	Admission episode flag
ADMIMETH	Method of admission
ADMISORC	Source of admission
AEKEY	Record identifier
BIRSTAT_N	Birth status
CLASSPAT	Patient classification
DIAG_NN	All Diagnosis codes
DISDEST	Destination on discharge
DISFLAG	Discharge episode flag
DOMPROC*	Trust derived dominant procedure
ENCRYPTED_HESID	Patient identifier - HES generated (encrypted version)
EPIKEY	Record identifier
EPIORDER	Episode order
ETHNOS*	Ethnic category
ETHRAW*	Ethnic character (audit version)
FIRSTREG	First regular day or night admission
HRG_N.N	Healthcare resource group: version 3.1
HRGNHS	Trust derived HRG value
HRGNHSVN	Version No. of Trust derived HRG
IMD04	IMD Index of Multiple Deprivation
IMD04_DECILE	IMD Decile Group
IMD04RK	IMD Overall Rank
MAINSPEF	Main specialty
OPERTN_NN*	Primary Operative Procedure Codes
PROCODE5	Provider code (5 character)
PROCODET	Provider code of treatment
PROTYPE	Provider type
RESGOR	Government Office region of residence
RESGOR_ONS	Government office region of residence (ONS)
RURURB_IND	Rural/Urban Indicator
SEX*	Sex of patient
SEXBABY*	Sex of baby
SITEDIST	Distance between patient's LSOA and provider site code of treatment
SITETRET*	Site code of treatment
SPELEND	End of spell indicator
SUSCOREHRG	SUS generated Core Spell HRG
SUSHRG	SUS generated HRG
SUSHRGVERS	SUS generated HRG version number
TRETSPEF*	Treatment specialty
LAUNCHES derived field name	**Description**
MATCH_RANK	Quality code of the linkage between the NCHDA record and the HES ID
STUDY_ID	LAUNCHES record ID of the linked NCHDA record
MY_DOB	Date of Birth - month and year

age_admission	Age at admission to hospital to 4 decimal places (calculated from ADMIDATE and DOB)
age_epistart*	Age at episode start to 4 decimal places (calculated from ADMIDATE and DOB)
age_discharge	Age at discharge from hospital to 4 decimal places (calculated from DISDATE and DOB)

Note: The HES/ONS linked data and linkage quality report were received on 17 September 2020. Records with a star (*) were obtained on 08 February 2022.

Table S3 HES Outpatient (HES OP) fields obtained.

HES Outpatient field name	**Description**
ATENTYPE	Attendance type
ATTENDED	Attended or did not attend
ATTENDKEY*	Record identifier
DIAG_NN	Diagnosis
ENCRYPTED_HESID	Patient identifier - HES generated (encrypted version)
ETHNOS*	Ethnic category
ETHRAW*	Ethnic character (audit version)
FIRSTATT	First attendance
HRGNHS	Trust derived HRG value
HRGNHSVN	Version No. of Trust derived HRG
IMD04	IMD Index of Multiple Deprivation
IMD04_DECILE	IMD Decile Group
IMD04RK	IMD Overall Ranking
LOCTYPE	Location type
MAINSPEF	Main specialty
OUTCOME	Outcome of attendance
PRIORITY	Priority type
PROCODE5	Provider code (5 character)
PROCODET	Provider code of treatment
PROTYPE	Provider type
PROVDIST	Distance between patient's LSOA and provider
REFSOURC	Source of referral
RESGOR	Government Office region of residence
RESGOR_ONS	Government office region of residence (ONS)
RURURB_IND	Rural / urban indicator
SEX*	Sex of patient
SITEDIST	Distance between patient's LSOA and provider site code of treatment
SITETRET*	Site code of treatment
SUSHRG	SUS generated HRG
SUSHRGVERS	SUS generated HRG version number
TRETSPEF	Treatment specialty
LAUNCHES derived field name	**Description**
MATCH_RANK	Quality code of the linkage between the NCHDA record and the HES ID
STUDY_ID	LAUNCHES record ID of the linked NCHDA record
MY_DOB	Date of Birth - month and year
age_appointment_OP	Age at outpatient appointment date to 4 decimal places (calculated from APPTDATE and DOB)

Note: The HES/ONS linked data and linkage quality report were received on 17 September 2020.
Records with a star (*) were obtained on 08 February 2022.

Table S4 HES A&E (HES AE) fields obtained.

HES A&E field name	Description
AEARRIVALMODE	Arrival mode
AEATTEND_EXC_PLANNED	Attendances excluding planned
AEDEPTTYPE	Department type
AEKEY	Record identifier
DIAG_NN*	A&E diagnosis
DIAG2_NN	A&E diagnosis: 2 character
DIAGSCHEME*	Diagnosis Scheme in Use
DOMPROC	Dominant procedure
ENCRYPTED_HESID	Patient identifier – HES generated (encrypted version)
EPIKEY	Record identifier
ETHRAW*	Ethnic category
HRGNHS	Trust derived HRG value
HRGNHSVN	Version No. of trust derived HRG
IMD04	IMD Index of Multiple Deprivation
IMD04_DECILE	IMD Decile group
IMD04RK	IMD Overall rank
PROCODE5	Provider code (5 character)
PROCODET	Provider code of treatment
PROTYPE	Provider type
PROVDIST	Distance between patient's LSOA and provider
RESGOR	Government Office region of residence
RESGOR_ONS	Government office region of residence (ONS)
RURURB_IND	Rural/Urban Indicator
SEX*	Sex of patient
SITEDIST	Distance between patient's LSOA and provider site code of treatment
SITETRET*	Site code of treatment
SUSHRG	SUS generated HRG
SUSHRGINFO	SUS generated HRG for information
SUSHRGVERINFO	SUS generated HRG for information version number
SUSHRGVERS	SUS generated HRG version number
LAUNCHES derived field name	**Description**
MATCH_RANK	Quality code of the linkage between the NCHDA record and the HES ID
STUDY_ID	LAUNCHES record ID of the linked NCHDA record
MY_DOB	Date of Birth - month and year
age_arrival_AE	Age at arrival to A&E to 4 decimal places (calculated from ARRIVALDATE and DOB)

Note: The HES/ONS linked data and linkage quality report were received on 17 September 2020. Records with a star (*) were obtained on 08 February 2022.

Table S5 ONS mortality fields obtained.

ONS field name	Description
ID	Patient identifier for ONS data – HES generated (encrypted)
STUDY_ID	LAUNCHES record ID of the linked NCHDA record
ENCRYPTED_HESID	Patient identifier – HES generated (encrypted version)
HES_ONS_Match_Rank	Quality code of the linkage between the ONS and HES IDs
age_death*	Age of death to 4 decimal places
life_status*	Life status (A=Alive, D=Death)
age_life_status*	Age at life status to 4 decimal places
COMMUNAL_ESTABLISHMENT_NAME	Communal Establishment (with place of death from the communal establishment lookup reference data)

Note: The HES/ONS linked data and linkage quality report were received on 17 September 2020. Records with a star (*) were obtained on 08 February 2022, including an update on age of death.

Table S6 PICANet fields obtained.

PICANet field name	Description
AcuteNec	Acute Necrotising Enterocolitis (NEC) main reason for ICU admission 1 indicates a positive response
AdType	Type of admission to PICU
bp_PrimReason	Primary reason for admission, a PIM variable used as an input value for PIM2, The PICANet data set stopped collecting this variable when PIM3 data collection was initiated. This variable is used if present, if not present this value can be assumed as recovering from surgery if the PIM3 variables indicate that the patient was recovering from surgery
Cardiomyocarditis	Cardiomyopathy or myocarditis 1 indicates a positive response
CareAreaAd	Care area of admission - expected if Source of admission is 1 or 2
ClinicalCodeType	The type of clinical code supplied to PICANet
Diagnosis	Read CTV3 clinical code
DiagnosticGroup	As Per PICANet annual report data set reporting categories - only included for primary diagnosis
DisPalCare	Discharged for palliative care
EcmoDaysDays	Number of days that the patient received Extracorporeal membrane oxygenation (ECMO)
ElectiveAd	Elective admission, 1 indicates that the admission was elective
Ethnic	Recorded ethnic category of record from
FU30DisStatus	Follow up 30 days post discharge
FU30Location	Location at 30 days following discharge
Gest	Gestational age in weeks, expected between 24 and 42, 99 indicates that the value is unknown
HiFlowNasalDays	Number of days that the patient received High flow nasal cannula therapy
HypoPlas	Hypoplastic left heart syndrome 1 indicates a positive response
IcpDeviceDays	Number of days that the patient used ICP-intracranial pressure monitoring, Intraventricular catheter or external ventricular drain
InfInotropeDays	Number of days that the patient received Continuous infusion of inotrope, vasodilator or prostaglandin
IntTracheostomy	Tracheostomy performed
Intubation	Patient was intubated
InvasiveVentilationDay	Number of days that the patient was invasively ventilated based on the PICANet daily interventions fields Invasive ventilation via endotracheal tube, Invasive ventilation via tracheostomy tube, Advanced ventilatory support (jet ventilation) and Advanced ventilatory support (oscillatory ventilation)
IsReAd	This record is an emergency readmission
LvadDays	Number of days that the patient used Ventricular assist device (VAD)
MechVent	Mechanical ventilation
MedHistEvid	Is evidence available of past medical history?
Mult	Multiplicity of birth, Expecting a value between 1 and 4m 9 indicates that the value is unknown
NonInvasiveVentilationDay	Number of days that the patient received Non-invasive ventilatory support
PICANetPatientID	Unique identifier for each patient in the PICANet database - pseudonymised from PICANet PatientID
PICANetRecordID	Unique identifier for each PICANet Admission Event - pseudonymised from PICANet EventID
PICUOrg	PICANet Organisation identifier
PICUOrgName	Name of PICANet organisation
PIM	Original PIM score

PIM2	Original PIM2 score
PIM2r2011	PIM2 recalculated 2011
PIM2r2012	PIM2 recalculated 2012
PIM2r2013	PIM2 recalculated 2013
PIM2r2014	PIM2 recalculated 2014
PIM2r2015	PIM2 recalculated 2015
PIM2r2016	PIM2 recalculated 2016
PIM2r2017	PIM2 recalculated 2017
PIM3	Original PIM3 score
PIMr	PIM recalculated
PrecedCpr	Cardiac arrest before ICU admission 1 indicates a positive response
PreceHospCardArr	Cardiac arrest OUT of hospital 1 indicates a positive response
PrevIcuAd	Previous ICU admission
RenalSupportDays	Number of days that the patient has received peritoneal dialysis, Haemofiltration, Haemodialysis, Plasma filtration or Plasma exchange
Retrieval	Was the patient retrieved to the organisation
SpontCerebHaem	Spontaneous cerebral haemorrhage 1 indicates a positive response
TotalUnplannedExtubations	The total number of times during an admission that there was dislodgement of the ETT from the trachea, without the intention to extubate immediately and without the presence of airway competent clinical staff appropriately prepared for the procedure occurs
TrachDays	Number of Days that the patient had a Tracheostomy cared for by nursing staff
UnitDisDest	Destination following discharge
UnitDisStatus	Status at discharge
UZ01Z_days	number of days the patient received care at UZ01Z
XB01Z_days	number of days the patient received care at XB01Z
XB02Z_days	number of days the patient received care at XB02Z
XB03Z_days	number of days the patient received care at XB03Z
XB04Z_days	number of days the patient received care at XB04Z
XB05Z_days	number of days the patient received care at XB05Z
XB06Z_days	number of days the patient received care at XB06Z
XB07Z_days	number of days the patient received care at XB07Z
XB09Z_days	number of days the patient received care at XB09Z
LAUNCHES derived field name	**Description**
AgeYrAdmit	Age at admission in years to 4 decimal places
AgeYrDeath	Age at death in years to 4 decimal places
AgeYrDischarge	Age at discharge in years to 4 decimal places

Note: The PICANet linked data and linkage quality report were received on 17 October 2019.

Table S7 ICNARC-CMP fields obtained.

ICNARC-CMP field name	Description
AP2aps	APACHE II acute physiology score
AP2probIC3	APACHE II (ICNARC 2013) probability
AP2score	APACHE II score
CMPrecordID	CMP (ICNARC) record ID for LAUNCHES
IM2018prob	Recalibrated 2018 model predicted mortality probability
IMscore	ICNARC model physiology score
MEWS	Modified Early Warning Score
ODacid_v3	Metabolic acidosis organ dysfunction
ODcardio_v3	Cardiovascular organ dysfunction
ODhaem_v3	Haematological organ dysfunction
ODrenal_v3	Renal organ dysfunction
ODresp_v3	Respiratory organ dysfunction
POPscore	Pancreatitis Outcome Prediction (POP) Score
SIRShr_v3	SIRS tachycardia flag
SIRSrr_v3	SIRS tachypnoea
SIRStemp_v3	SIRS fever or hypothermia flag
SIRSwbc_v3	SIRS white blood cell flag
SOFAcardio_v3	SOFA cardiovascular score (0-2)
SOFAcoag_v3	SOFA coagulation score (0-2)
SOFAincr_v3	SOFA increase from baseline (0-2 per organ)
SOFAliver_v3	SOFA liver score (0-2)
SOFAneuro_v3	SOFA neurological score (0-2)
SOFAod_v3	SOFA number of organ dysfunctions (2+ each organ)
SOFArenal_v3	SOFA renal score (0-2)
SOFAresp_v3	SOFA respiratory score (0-2)
SOFAtot_v3	SOFA total score (0-2 per organ)
acsd	Calendar days of advanced cardiovascular support while in your unit
admtype	Type of admission
ahlos	All hospital length of stay, days
ahlosa	Any hospital length of stay after discharge from icu, days
ahlosb	Any hospital length of stay before admission to icu, days
ahsurv	Ultimate hospital survival
arsd	Calendar days of advanced respiratory support while in your unit
aulos	All unit length of stay, days
ausurv	Ultimate unit survival
bcsd	Calendar days of basic cardiovascular support while in your unit
bmi	BMI (kg/m2)
brsd	Calendar days of basic respiratory support while in your unit
bsdtp	Brainstem death declared
ccl0d	Days of level 0 care while in your unit
ccl1d	Days of level 1 care while in your unit
ccl2d	Days of level 2 care while in your unit
ccl3d	Days of level 3 care while in your unit

classns	Classification of surgery
cpr_v3	Cardiopulmonary resuscitation within 24 hours prior to admission to your unit
crpreg	Admission currently/recently pregnant
curb65	CURB 65 score
delay	Length of delay (days)
deldis12	Delayed discharges (12 hour delay)
deldis24	Delayed discharges (24 hour delay)
deldis24_exooh	Delayed discharges (24 hour delay), excluding night discharges
deldis4	Delayed discharges (4 hour delay)
dep	Dependency prior to admission to acute hospital
desc	ICNARC diagnostic category description (raicu1)
desth_v3	Destination following discharge from your hospital
dis	Status at discharge from your unit
dobest	Date of birth estimated
dsd	Calendar days of dermatological support while in your unit
ethnic	Ethnicity
gsd	Calendar days of gastrointestinal support while in your unit
hloca	Hospital housing non-transient location (in)
hlocd	Hospital housing non-transient location (out)
hrg	Healthcare Resource Group
htloca	Hospital housing transient location (in)
imd2015	Quintile of English IMD 2015/Welsh IMD 2014/NI MDM 2010
imd_error	Postcode available for derivation of IMD
infection_v3	Infection
itw_v3	Treatment withheld/withdrawn
kdigo_mdrd75	AKI stage (KDIGO)
korgfail	Number of Knaus organ system failures
leva	Highest level of care received in the first 24 hours in your unit
loca	Location (in)
locd	Location (out)
lsd	Calendar days of liver support while in your unit
nlb	Number of live births (babies) from recent pregnancy
npcs	Number of previous Caesarean sections excluding most recent pregnancy
nplsb	Number of live births (babies) or stillbirths from previous pregnancies
nsb	Number of stillbirths from recent pregnancy
nsd	Calendar days of neurological support while in your unit
nuaib	Number of unit-acquired infections present in blood
orgdys_v3	Number of organ dysfunctions
outrp	Outcome of recent pregnancy
ploca	Prior location (in)
raicu1	Primary reason for admission to your unit
rdis_v3	Reason for discharge from your unit
readearly	Early readmissions
readlate	Late readmissions
resa	Residence prior to admission to an acute hospital
resd	Residence post-discharge from your hospital

rsd	Calendar days of renal support while in your unit
sepsis3_v3	Sepsis-3
sepsis_v3	Sepsis
sirs_v3	SIRS criteria count
soha	Sector of other hospital (in)
sohd	Sector of other hospital (out)
sshock3_v3	Septic shock (Sepsis-3)
tnessd	Timeliness of discharge from your unit
typeiha	Type of adult ICU/HDU (in)
typeihd	Type of adult ICU/HDU (out)
version	ICMPDS version number
withinsh	Readmission within same hospital stay
wkg	Weight (kg)
yhlos	Your hospital length of stay, days
yhlosa	Your hospital length of stay after discharge from icu, days
yhlosb	Your hospital length of stay before admission to icu, days
yhsurv	Your hospital survival
yulos	Your unit length of stay, hours
yusurv	Your unit survival
LAUNCHES derived field name	**Description**
age_ah	Age at admission to hospital (not time-based, 4 decimal places)
age_aicu	Age at admission to ICU (time-based, 4 decimal places)
age_dbsd	Age at declaration of brainstem death (time-based, 4 decimal places)
age_dh	Age at discharge from hospital (not time-based, 4 decimal places)
age_dicu	Age at discharge from ICU (time-based, 4 decimal places)
age_oah	Age at original admission to hospital (not time-based, 4 decimal places)
age_oaicu	Age at original admission to ICU (not time-based, 4 decimal places)
age_od	Age at death (time-based, 4 decimal places)
age_tw	Age when decision to withdraw treatment made (time-based, 4 decimal places)
age_udicu	Age at death (time-based, 4 decimal places)
Country	Country
Mob	Month of birth
qcode_CMP	Quality of CMP record for linkage
Yob	Year of birth

Note: The ICNARC-CMP linked data and linkage quality report were received on 4 December 2019.

Table S8 NCHDA to PICANet lookup of organisations, used to match hospital patient identifiers.

NCHDA organisation name	PICANet organisation name
ACH. Alder Hey Hospital	PIC010. Liverpool Alder Hey
BCH. Birmingham Children's Hospital	PIC001. Birmingham Children's Hospital
BRC. Bristol Children's Hospital	PIC003. Bristol Royal Hospital for Children
FRE. Freeman Hospital	PIC021. Newcastle Freeman Hospital
GEO. St George's Hospital	PIC015. London St George's Hospital
GOS. Great Ormond Street Hospital for Children	PIC011/PIC039. London Great Ormond Street Hospital PICU_NICU/CCCU
GRL. Glenfield Hospital	PIC008. Leicester Glenfield Hospital
GUY. Evelina London Children's Hospital	PIC012. London Evelina Children's Hospital
HRI. Hull Royal Infirmary	PIC028. Hull Royal Infirmary
KCH. King's College Hospital	PIC013. London Kings College Hospital
LGI. Leeds General Infirmary	PIC006. Leeds General Infirmary
MRI. Manchester Royal Infirmary	PIC018. Manchester Royal Children's Hospital
NGS. Northern General Hospital	PIC025 / PIC029. Sheffield General NICU/PICU
NHB. Royal Brompton Hospital	PIC014. London Royal Brompton Hospital
RAD. John Radcliffe Hospital	PIC024. Oxford John Radcliffe Hospital
SBH. St Bartholomew's Hospital	PIC032. London The Royal London Hospital
SGH. Southampton General Hospital	PIC026. Southampton Children's Hospital
STM. St Marys Hospital, Paddington	PIC016. London St Mary's Hospital
STO. University Hospital of North Staffordshire	PIC027. Stoke on Trent - Royal Stoke University Hospital
UHW. University Hospital of Wales	PIC005. Cardiff Noah's Ark children's Hospital for Wales

Table S9 HES 8-step linkage method.

Match Rank	NHS number	DoB	Sex	Postcode	Extra Condition
1	Exact	Exact	Exact	Exact	
2	Exact	Exact	Exact		
3	Exact	Partial	Exact	Exact	
4	Exact	Partial	Exact		
5	Exact			Exact	
6		Exact	Exact	Exact	where NHS does not contradict the match and DOB is not 1 January and the POSTCODE is not in the 'ignore' list
7		Exact	Exact	Exact	where NHS does not contradict the match and DOB is not 1 January
8	Exact				

Note: Sex was not part of the NCHDA identifiers approved for linkage, so only steps 5 and 8 were used.

Table S10 ICD-10 diagnosis codes indicating congenital heart disease (CHD) or potential mistakenly coded acquired (non-rheumatic) heart disease.

ICD-10 congenital heart disease code	Description
Q20	Congenital malformations of cardiac chambers and connections
Q20.0	Common arterial trunk
Q20.1	Double outlet right ventricle
Q20.2	Double outlet left ventricle
Q20.3	Discordant ventriculoarterial connection
Q20.4	Double inlet ventricle
Q20.5	Discordant atrioventricular connection
Q20.6	Isomerism of atrial appendages
Q20.8	Other congenital malformations of cardiac chambers and connections
Q20.9	Congenital malformation of cardiac chambers and connections, unspecified
Q21	Congenital malformations of cardiac septa
Q21.0	Ventricular septal defect
Q21.1	Atrial septal defect
Q21.2	Atrioventricular septal defect
Q21.3	Tetralogy of Fallot
Q21.4	Aortopulmonary septal defect
Q21.8	Other congenital malformations of cardiac septa
Q21.9	Congenital malformation of cardiac septum, unspecified
Q22	Congenital malformations of pulmonary and tricuspid valves
Q22.0	Pulmonary valve atresia
Q22.1	Congenital pulmonary valve stenosis
Q22.2	Congenital pulmonary valve insufficiency
Q22.3	Other congenital malformations of pulmonary valve
Q22.4	Congenital tricuspid stenosis
Q22.5	Ebstein anomaly
Q22.6	Hypoplastic right heart syndrome
Q22.8	Other congenital malformations of tricuspid valve
Q22.9	Congenital malformation of tricuspid valve, unspecified
Q23	Congenital malformations of aortic and mitral valves
Q23.0	Congenital stenosis of aortic valve
Q23.1	Congenital insufficiency of aortic valve
Q23.2	Congenital mitral stenosis
Q23.3	Congenital mitral insufficiency
Q23.4	Hypoplastic left heart syndrome
Q23.8	Other congenital malformations of aortic and mitral valves
Q23.9	Congenital malformation of aortic and mitral valves, unspecified
Q24	Other congenital malformations of heart
Q24.0	Dextrocardia
Q24.1	Laevocardia
Q24.2	Cor triatriatum
Q24.3	Pulmonary infundibular stenosis
Q24.4	Congenital subaortic stenosis

Q24.5	Malformation of coronary vessels
Q24.6	Congenital heart block
Q24.8	Other specified congenital malformations of heart
Q24.9	Congenital malformation of heart, unspecified
Q25	Congenital malformations of great arteries
Q25.0	Patent ductus arteriosus
Q25.1	Coarctation of aorta
Q25.2	Atresia of aorta
Q25.3	Stenosis of aorta
Q25.4	Other congenital malformations of aorta
Q25.5	Atresia of pulmonary artery
Q25.6	Stenosis of pulmonary artery
Q25.7	Other congenital malformations of pulmonary artery
Q25.8	Other congenital malformations of great arteries
Q25.9	Congenital malformation of great arteries, unspecified
Q26	Congenital malformations of great veins
Q26.0	Congenital stenosis of vena cava
Q26.1	Persistent left superior vena cava
Q26.2	Total anomalous pulmonary venous connection
Q26.3	Partial anomalous pulmonary venous connection
Q26.4	Anomalous pulmonary venous connection, unspecified
Q26.5	Anomalous portal venous connection
Q26.6	Portal vein-hepatic artery fistula
Q26.8	Other congenital malformations of great veins
Q26.9	Congenital malformation of great vein, unspecified
Q28.8	Other specified congenital malformations of circulatory system
Q28.9	Congenital malformation of circulatory system, unspecified
Q87.4	Marfan syndrome
Q89.3	Situs inversus
ICD-10 acquired heart disease code	**Description**
I33	Acute and subacute endocarditis
I330	Acute and subacute infective endocarditis
I339	Acute and subacute endocarditis, unspecified
I34	Nonrheumatic mitral valve disorders
I340	Nonrheumatic mitral (valve) insufficiency
I341	Nonrheumatic mitral (valve) prolapse
I342	Nonrheumatic mitral (valve) stenosis
I348	Other nonrheumatic mitral valve disorders
I349	Nonrheumatic mitral valve disorder, unspecified
I35	Nonrheumatic aortic valve disorders
I350	Nonrheumatic aortic (valve) stenosis
I351	Nonrheumatic aortic (valve) insufficiency
I352	Nonrheumatic aortic (valve) stenosis with insufficiency
I358	Other nonrheumatic aortic valve disorders
I359	Nonrheumatic aortic valve disorder, unspecified
I36	Nonrheumatic tricuspid valve disorders

I360	Nonrheumatic tricuspid (valve) stenosis
I361	Nonrheumatic tricuspid (valve) insufficiency
I362	Nonrheumatic tricuspid (valve) stenosis with insufficiency
I368	Other nonrheumatic tricuspid valve disorders
I369	Nonrheumatic tricuspid valve disorder, unspecified
I37	Pulmonary valve disorders
I370	Nonrheumatic pulmonary valve stenosis
I371	Nonrheumatic pulmonary valve insufficiency
I372	Nonrheumatic pulmonary valve stenosis with insufficiency
I378	Other nonrheumatic pulmonary valve disorders
I379	Nonrheumatic pulmonary valve disorder, unspecified
I38	Endocarditis, valve unspecified
I39	Endocarditis and heart valve disorders in diseases classified elsewhere
I390	Mitral valve disorders in diseases classified elsewhere
I391	Aortic valve disorders in diseases classified elsewhere
I392	Tricuspid valve disorders in diseases classified elsewhere
I393	Pulmonary valve disorders in diseases classified elsewhere
I394	Multiple valve disorders in diseases classified elsewhere
I398	Endocarditis, valve unspecified, in diseases classified elsewhere
I40	Acute myocarditis
I400	Infective myocarditis
I401	Isolated myocarditis
I408	Other acute myocarditis
I409	Acute myocarditis, unspecified
I41	Myocarditis in diseases classified elsewhere
I410	Myocarditis in bacterial diseases classified elsewhere
I411	Myocarditis in viral diseases classified elsewhere
I412	Myocarditis in other infectious and parasitic diseases classified elsewhere
I418	Myocarditis in other diseases classified elsewhere
I42	Cardiomyopathy
I420	Dilated cardiomyopathy
I421	Obstructive hypertrophic cardiomyopathy
I422	Other hypertrophic cardiomyopathy
I423	Endomyocardial (eosinophilic) disease
I424	Endocardial fibroelastosis
I425	Other restrictive cardiomyopathy
I426	Alcoholic cardiomyopathy
I427	Cardiomyopathy due to drug and external agent
I428	Other cardiomyopathies
I429	Cardiomyopathy, unspecified
I43	Cardiomyopathy in diseases classified elsewhere
I430	Cardiomyopathy in infectious and parasitic diseases classified elsewhere
I431	Cardiomyopathy in metabolic diseases
I432	Cardiomyopathy in nutritional diseases
I438	Cardiomyopathy in other diseases classified elsewhere
I44	Atrioventricular and left bundle-branch block

I440	Atrioventricular block, first degree
I441	Atrioventricular block, second degree
I442	Atrioventricular block, complete
I443	Other atrioventricular block
I444	Left anterior fascicular block
I445	Left posterior fascicular block
I446	Other fascicular block
I447	Left bundle-branch block, unspecified
I45	Other conduction disorders
I450	Right fascicular block
I451	Other right bundle-branch block
I452	Bifascicular block
I453	Trifascicular block
I454	Nonspecific intraventricular block
I455	Other specified heart block
I456	Pre-excitation syndrome
I458	Other specified conduction disorders
I459	Conduction disorder, unspecified
I46	Cardiac arrest
I4.0	Cardiac arrest with successful resuscitation
I461	Sudden cardiac death, so described
I462	Cardiac arrest due to underlying cardiac condition
I468	Cardiac arrest due to other underlying condition
I469	Cardiac arrest, cause unspecified
I47	Paroxysmal tachycardia
I470	Re-entry ventricular arrhythmia
I471	Supraventricular tachycardia
I472	Ventricular tachycardia
I479	Paroxysmal tachycardia, unspecified
I48	Atrial fibrillation and flutter
I480	Paroxysmal atrial fibrillation
I481	Persistent atrial fibrillation
I482	Chronic atrial fibrillation
I483	Typical atrial flutter
I484	Atypical atrial flutter
I489	Atrial fibrillation and atrial flutter, unspecified
I49	Other cardiac arrhythmias
I490	Ventricular fibrillation and flutter
I491	Atrial premature depolarization
I492	Junctional premature depolarization
I493	Ventricular premature depolarization
I494	Other and unspecified premature depolarization
I495	Sick sinus syndrome
I498	Other specified cardiac arrhythmias
I499	Cardiac arrhythmia, unspecified
I50	Heart failure

I500	Congestive heart failure
I501	Left ventricular failure
I509	Heart failure, unspecified

Table S11 Total number of NCHDA records and number and percentage of those with a valid NHS number, broken down by country of residence derived from postcode (by NICOR).

Country of residence from patient postcode	Total NCHDA records	Valid NHS number (number of records)	Valid NHS number (percentage of records)
England	129,952	128,351	98.8%
Wales	6,514	6,457	99.1%
Crown Dependencies	531	342	64.4%
Scotland	681	241	35.4%
Northern Ireland	1,085	80	7.4%
Overseas	1,981	144	7.3%
Missing or invalid postcode	3,118	642	20.6%
Total	143,862	136,257	94.7%

Table S12 Quality of NHS numbers in each of the data sets (across all records before linkage): yearly total number of records and percentage of records with valid NHS numbers.

Financial Year	NCHDA *n (% valid)*	PICANet *n (% valid)*	ICNARC-CMP *n (% valid)*	HES Inpatient *n (% valid)*	HES Outpatient *n (% valid)*	HES A&E *n (% valid)*
1998				11,983,893 (74.3%)		
1999				12,196,270 (80.3%)		
2000	6,422 (85.1%)			12,264,676 (83.2%)		
2001	6,170 (92.5%)			12,337,724 (86.5%)		
2002	6,102 (93.7%)	2,427 (80.2%)		12,712,153 (90.5%)		
2003	7,446 (93.3%)	8,171 (76.9%)		13,295,166 (93.6%)	51,427,003 (93.8%)	
2004	6,936 (94.2%)	8,846 (75.7%)		13,706,450 (95.4%)	54,420,813 (95.6%)	
2005	7,720 (94.4%)	9,699 (75.9%)		14,423,506 (95.7%)	60,608,403 (96.5%)	
2006	8,122 (94.9%)	10,288 (84.4%)	81,752 (4.5%)	14,784,581 (96.4%)	63,217,226 (97.0%)	
2007	7,972 (95.4%)	11,146 (89.9%)	93,669 (56.7%)	15,359,062 (96.7%)	66,649,484 (97.8%)	12,318,051 (85.4%)
2008	8,324 (95.3%)	11,742 (90.6%)	101,926 (82.8%)	16,232,579 (97.1%)	74,853,493 (98.1%)	13,794,072 (88.8%)
2009	8,726 (96.2%)	12,646 (91.6%)	114,519 (90.4%)	16,806,196 (97.8%)	84,198,458 (98.2%)	15,569,736 (90.1%)
2010	8,969 (96.4%)	13,236 (93.1%)	135,880 (93.3%)	17,269,882 (98.2%)	87,998,505 (98.7%)	16,244,934 (92.0%)
2011	9,107 (96.9%)	13,390 (94.5%)	156,181 (95.0%)	17,465,425 (98.5%)	90,956,844 (98.8%)	17,619,708 (93.6%)
2012	8,994 (95.9%)	14,307 (94.6%)	160,737 (95.8%)	17,715,046 (98.7%)	94,091,748 (99.0%)	18,328,896 (95.1%)
2013	9,582 (95.3%)	14,551 (95.1%)	170,073 (96.1%)	18,163,101 (98.7%)	101,844,824 (99.0%)	18,517,381 (96.0%)
2014	9,641 (95.4%)	15,659 (94.8%)	191,585 (96.0%)	18,731,987 (98.8%)	107,188,423 (99.2%)	19,556,781 (95.2%)
2015	11,472 (94.8%)	16,898 (94.9%)	206,383 (96.1%)	19,239,608 (98.8%)	113,298,661 (99.2%)	20,457,805 (95.9%)
2016	12,157 (95.9%)	16,785 (96.6%)	217,496 (96.4%)	19,726,907 (98.5%)	118,578,912 (99.4%)	20,886,411 (96.9%)
2017			223,367 (96.7%)	20,030,870 (98.3%)	119,378,895 (99.4%)	21,278,504 (97.1%)
All years	*143,862 (94.7%)*	*179,791 (90.5%)*	*1,853.568 (88.7%)*	*314,445,082 (93.8%)*	*1,288,711,692 (98.0%)*	*194,572,279 (93.3%)*

Note: We do not know the quality of identifiers in ONS mortality data, which we obtained linked to HES data.

Table S13 Number of linked records in each data set before quality assurance, by estimated financial year

Financial Year	NCHDA	PICANet	ICNARC-CMP	HES Inpatient	HES Outpatient	HES A&E	Total
1998	0	0	0	16,498	0	0	16,498
1999	0	0	0	19,902	0	0	19,902
2000	6,421	15	2	29,235	0	0	35,673
2001	6,161	11	1	33,390	0	0	39,563
2002	6,137	952	0	37,128	0	0	44,217
2003	7,402	3,226	0	43,106	132,710	0	186,444
2004	6,968	3,464	0	45,657	150,065	0	206,154
2005	7,684	3,828	0	50,436	177,020	0	238,968
2006	8,152	4,052	6	52,256	196,331	0	260,797
2007	7,984	4,136	154	56,918	224,342	23,352	316,886
2008	8,294	4,275	215	60,251	255,702	27,597	356,334
2009	8,719	4,748	273	65,634	294,368	32,876	406,618
2010	8,987	4,891	388	69,485	323,853	35,989	443,593
2011	9,102	5,103	407	70,851	348,802	38,968	473,233
2012	9,013	5,176	411	71,279	370,113	41,745	497,737
2013	9,593	5,435	473	72,077	408,561	42,959	539,098
2014	9,639	5,435	447	73,027	442,349	45,018	575,915
2015	11,492	5,546	629	76,245	470,121	47,341	611,374
2016	12,114	5,504	686	73,112	478,591	47,060	617,067
2017	0	0	572	52,074	426,352	43,604	522,602
All years	*143,862*	*65,797*	*4,664*	*1,068,561*	*4,699,280*	*426,509*	*6,408,673*

Note: financial years (running from April to March) were estimated using the ages at events and the estimated date of birth (we took day 15th of the known month of birth as date of birth).