BMJ Open

Improving Data Quality in a UK Out-of-Hospital Cardiac Arrest Registry Through Data Linkage Between the Out-of-Hospital Cardiac Arrest Outcomes (OHCAO) Project and NHS Digital

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-017784
Article Type:	Research
Date Submitted by the Author:	16-May-2017
Complete List of Authors:	Rajagopal, Sangeerthana; University of Warwick Warwick Medical School, Warwick Clinical Trials Unit Booth, Scott; University of Warwick Warwick Medical School, Warwick Clinical Trials Unit Brown, Terry; University of Warwick Warwick Medical School, Warwick Clinical Trials Unit Ji, Chen; University of Warwick Warwick Medical School, Warwick Clinical Trials Unit Hawkes, Claire; University of Warwick Warwick Medical School, Warwick Clinical Trials Unit Siriwardena, Aloysius; University of Lincoln, School of Health and Social Care Kirby, Kim; South Western Ambulance Service NHS Foundation Trust Black, Sarah; South Western Ambulance Service NHS Foundation Trust, Spaight, Robert; East Midlands Ambulance Service NHS Trust Gunson, Imogen; West Midlands Ambulance Service NHS Trust Brace-McDonnell, Samantha; University of Warwick, Warwick Clinical Trials Unit Perkins, Gavin; University of Warwick, Warwick Clinical Trials Unit
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Emergency medicine
Keywords:	Cardiac arrest, Emergency medical services, Medical record linkage, Out-of-hospital cardiac arrest, Resuscitation

Improving Data Quality in a UK Out-of-Hospital Cardiac Arrest Registry Through Data Linkage Between the Out-of-Hospital Cardiac Arrest Outcomes (OHCAO) Project and NHS Digital

Sangeerthana Rajagopal[a,b], Scott J Booth[a], Terry P Brown[a], Chen Ji[a], Claire Hawkes[a], A Niroshan Siriwardena[c], Kim Kirby[d], Sarah Black[d], Robert Spaight[e], Imogen Gunson[f], Samantha J Brace-McDonnell[a,b], Gavin D Perkins[a,b], on behalf of OHCAO collaborators

[a]Warwick Clinical Trials Unit, University of Warwick, Coventry CV4 7AL, UK

[b]Heart of England NHS Foundation Trust, Birmingham B9 5SS, UK

[c]University of Lincoln, Lincolnshire LN6 7TS, UK

[d]South Western Ambulance Service NHS Foundation Trust, Exeter EX2 7HY, UK

[e]East Midlands Ambulance Service NHS Trust, Nottingham NG8 6PY, UK

[f] West Midlands Ambulance Service NHS Foundation Trust, Brierley Hill, DY5 1LX, UK

Correspondence to Professor Gavin D Perkins; G.D.Perkins@warwick.ac.uk

Word count 3941

ABSTRACT

Objectives: The Out-of-Hospital Cardiac Arrest Outcomes (OHCAO) project aims to understand the epidemiology and outcomes of out-of-hospital cardiac arrests (OHCA) across the UK. This study is a sub-project of OHCAO and aims to establish the feasibility of producing a registry of OHCAs by linking OHCAO data to National Health Service (NHS) patient demographic data and date of death data held on the Personal Demographics Service (PDS) database, via NHS Digital, to improve data quality and establish accurate 30-day survival outcomes for OHCA.

Design and setting: Data were collected from 1st January 2014 to 31st December 2014 as part of a prospective, observational study of OHCAs attended by ten English NHS Ambulance Services. 28,729 OHCA cases had resuscitation attempted by Emergency Medical Services and were included in the study. Of these, a randomly selected sample of 3120 cases were securely transferred to the NHS Digital list cleaning service to be matched using OHCAO patient demographic data to return previously missing data and provide Office for National Statistics (ONS) mortality data.

Results: A total of 80.5% of OHCAO cases were matched to the NHS PDS database. Using the linkage process, missing demographic data was retrieved for 72.7% of cases with incomplete data and confirmation of 30-day survival improved by 37.6% with a reduction in unknown 30-day survival status from 46.1% to 8.5%.

Conclusions: Data linkage was shown to successfully improve the quality of OHCA demographic data and survival status 30 days after OHCA.

ARTICLE SUMMARY

Strengths and limitations of this study

- Data points collected as part of the OHCAO project were based on established Utstein guidelines.
- The quality of demographic data collected by the OHCAO project was first improved through a list cleaning service provided by NHS digital.
- Following list cleaning, exact data matches with Office for National Statistics (ONS) data established 30-day survival status.
- Provision of NHS numbers from OHCAO and NHS digital provides potential for following long-term survival outcomes in OHCA patients through data linkage.
- Reliance on submission of good quality data to OHCAO project.

INTRODUCTION

Medical registries have been shown to initiate improvements in healthcare systems especially where data are fed back transparently to institutions providing the source data.[1] Every year in the United Kingdom (UK) there are around 60,000 out-of-hospital cardiac arrests (OHCA) attended by Emergency Medical Services (EMS) of which approximately 28,000 have resuscitation attempted.[2,3] This group suffer significant mortality and morbidity and since 2011, survival to hospital discharge rates have been part of the National Health Service (NHS) England Ambulance Quality Indictors (AQIs). Whilst survival to hospital discharge is easier to collect than 30-day survival, comparisons between international registries are limited by cultural differences (whether patients are discharged home to die or die predominantly in hospital) and health system differences (discharge process efficiency, long-term care provision in hospital versus care home settings).

Ensuring high data quality is essential as this forms the basis of decisions that ultimately impact on changes in care and healthcare resource allocation. Many issues require consideration when collecting OHCA datasets. Firstly, case identification occurs by many means including ambulance dispatch codes, crew reporting and database searches for arrest codes.[4] Secondly, outcomes are collected through different methods including coroner reports, emergency department records, central records centres or electronic records,[4] each with varying outcome definitions. This illustrates how different data collection processes affect data quality variability and reliability.

In the UK, significant variation exists between ambulance services in outcomes for patients with attempted resuscitation following OHCA. In 2011, return of spontaneous circulation (ROSC) was achieved in 13.3% to 26.7% of patients at hospital arrival and 2.2% to 12.0% of patients survived to hospital discharge.[4] Such differences have been observed worldwide.[5] However, Lilford et al highlighted that the greatest variation in reporting outcomes can be traced to the quality of data that outcome results are based on.[6] Data linkage may partly address the problem of missing data. This involves using multiple databases to fill missing data and merge data into a universal file. Thus, data

linkage also works towards solving the problem of using multiple data collection methods to source data as it results in the storage of information in one location.

Data linkage has been used successfully in several fields including maternal health records,[7] post market surveillance of medical devices,[8] cardiac rehabilitation[9] and hormone therapy in breast cancer.[10] It has also been utilised by regional and national OHCA databases to confirm survival status through mortality data linkage in Australia,[11] Denmark[12] and Canada.[13] Linkage can provide a centralised, high quality database for research and service appraisal and has the potential to allow longitudinal surveillance of patients with cardiac arrest to accurately determine survival.

The Out-of-Hospital Cardiac Arrest Outcomes (OHCAO) project is funded by the Resuscitation Council (UK), British Heart Foundation and University of Warwick. It is a prospective observational study aiming to investigate the epidemiology and outcomes of adults and children sustaining OHCAs across the UK.[14,15] This paper presents a sub-project of the OHCAO project aiming to establish the feasibility of linking OHCAO data to NHS patient demographic data and Office for National Statistics (ONS) mortality data through the NHS Digital list cleaning service.

METHODS

Setting

The ten English NHS ambulance services collecting data for the OHCAO project cover approximately 54 million people. Data were collected from 1st January 2014 to 31st December 2014 on 28,729 patients suffering OHCAs in whom resuscitation was attempted by EMS. This figure was reached after excluding individuals in whom resuscitation was not attempted as per national guidelines due to the presence of a do not attempt resuscitation order, signs incompatible with life or where resuscitation attempts would be futile.[16]

Aims & Objectives

The overall aim of this project was to investigate the feasibility of linking a sample of OHCAO 2014 data to NHS patient demographic data and date of death data held on the Personal Demographics Service (PDS) database, via the NHS Digital list cleaning service, to improve data quality and establish accurate 30-day survival outcomes for OHCA. The objectives were to (1) assess the success rate of OHCAO patient demographic variables (NHS number, surname, forename, date of birth (DOB), and home postcode) for matching to the NHS PDS database through NHS Digital list cleaning; (2) assess improvements in the completeness of OHCAO patient demographic variables through NHS Digital list cleaning; (3) create a linked OHCAO and NHS PDS database allowing analysis of 30-day survival from OHCA; (4) compare OHCA patient demographic, event, pre-EMS intervention, clinical, and outcome characteristics between cases that were matched to NHS PDS data compared to those that remained unmatched.

OHCAO project data collection

Detailed information about the OHCAO project is available in the study protocol.[14] EMS personnel identified OHCAs by searching case records for confirmed arrests, cases indicating treatment for cardiac arrests e.g. patients with no pulse or respiratory effort and 999 dispatch codes. Core and supplemental Utstein variables were collected encompassing demographic, system, process and outcome data.[17] The OHCAO project received ambulance service data uploads via a secure server and stored data on the OHCAO database at the University of Warwick.

OHCAO data sample

The analysis presented here represents a 10.9% sample of the 2014 data from the OHCAO database, selected using simple random sampling and stratified by ambulance service. The sample consisted of 3120 patients whose data were linked with NHS PDS data to establish the feasibility of linking to mortality data to confirm survival status.

OHCAO data linkage to ONS mortality data

OHCAO to NHS PDS data linkage approval was received after submitting an application to the NHS Digital Data Access Request Service; additional approval was obtained from ONS for the release of mortality data. OHCAO submitted 3120 cases to NHS Digital, via the NHS Digital secure transfer system, detailing the following patient demographic variables of varying completeness: NHS number, surname, forename, DOB and home postcode.

OHCAO used the NHS Digital list cleaning service which validates demographic data to ensure accuracy and improve data linkage outcomes. Validation is achieved by NHS Digital matching submitted demographic variables to NHS patient demographic data held on the PDS database. The PDS database is a national electronic database containing NHS patient demographic information, including NHS number, name and address. For each matched case NHS Digital were asked to provide the following patient demographic information: NHS number, surname, forename, and home postcode.

NHS Digital utilised both automatic and manual matching techniques, using a combination of deterministic and probabilistic matching methods.[18,19] Cases were initially submitted for automatic matching which uses a decision tree algorithm to provide matches. A subset of cases that failed automatic matching were resubmitted for manual matching. Where manual matching is required, NHS Digital operators use up to 20 search routes using the demographic variables to provide matches.

As part of the list cleaning service NHS Digital was also able to provide a date of death if the patient was deceased. The date of death data is held in the NHS PDS dabase and is sourced from ONS mortality data. OHCAO required information on deaths from 1st January 2014 until 31st January 2015. This was utilised to calculate 30-day survival. Where no date of death was provided the patient was categorised as alive.

ONS date of death data

ONS mortality data contains all deaths registered in England and Wales. Deaths are normally registered within 5 days of death dates collected from death certificates, coroner certificates and inquests. ONS mortality data is subject to validation and quality assurance processes and collected in line with the Statistics and Registration Service Act 2007.[20]

Analysis

An analysis was conducted to assess how particular demographic data points enabled linkage with NHS PDS data and how and if the data linkage process improved the completeness of patient demographic data. This was done descriptively with breakdowns of data linkage match rates for all combinations of the OHCAO demographic variables sent to NHS Digital for data linkage.

The combined linked dataset was analysed to investigate 30-day survival rates calculated by evaluating if patients were alive ≥30 calendar days from the EMS OHCA incident date. The analysis was carried out pre and post linkage, illustrating linkage effects. 30-day survival was calculated using OHCAO data where there was a date of death or date discharged >30 days after the OHCA incident date. Where there was an OHCAO date of death ≤30 days after the OHCA incident date or further ambulance service data indicating the patient status was deceased on the day of the OHCA incident date (e.g. hospital code indicating patient deceased and not conveyed to hospital) the patient was categorised as not surviving to 30 days. All other cases were categorised as unknown for patient 30-day survival status. For the combined linked dataset, cases that were linked to ONS mortality data were categorised as 30-day survival where there was no date of death or where a date of death was provided that was >30 days after the OHCA incident date. Where there was an ONS date of death ≤30 days after the OHCA incident date the patient was categorised as not surviving to 30 days. Where there was a contradiction in patient survival status between OHCAO data and ONS mortality data then ONS mortality data superseded OHCAO data.

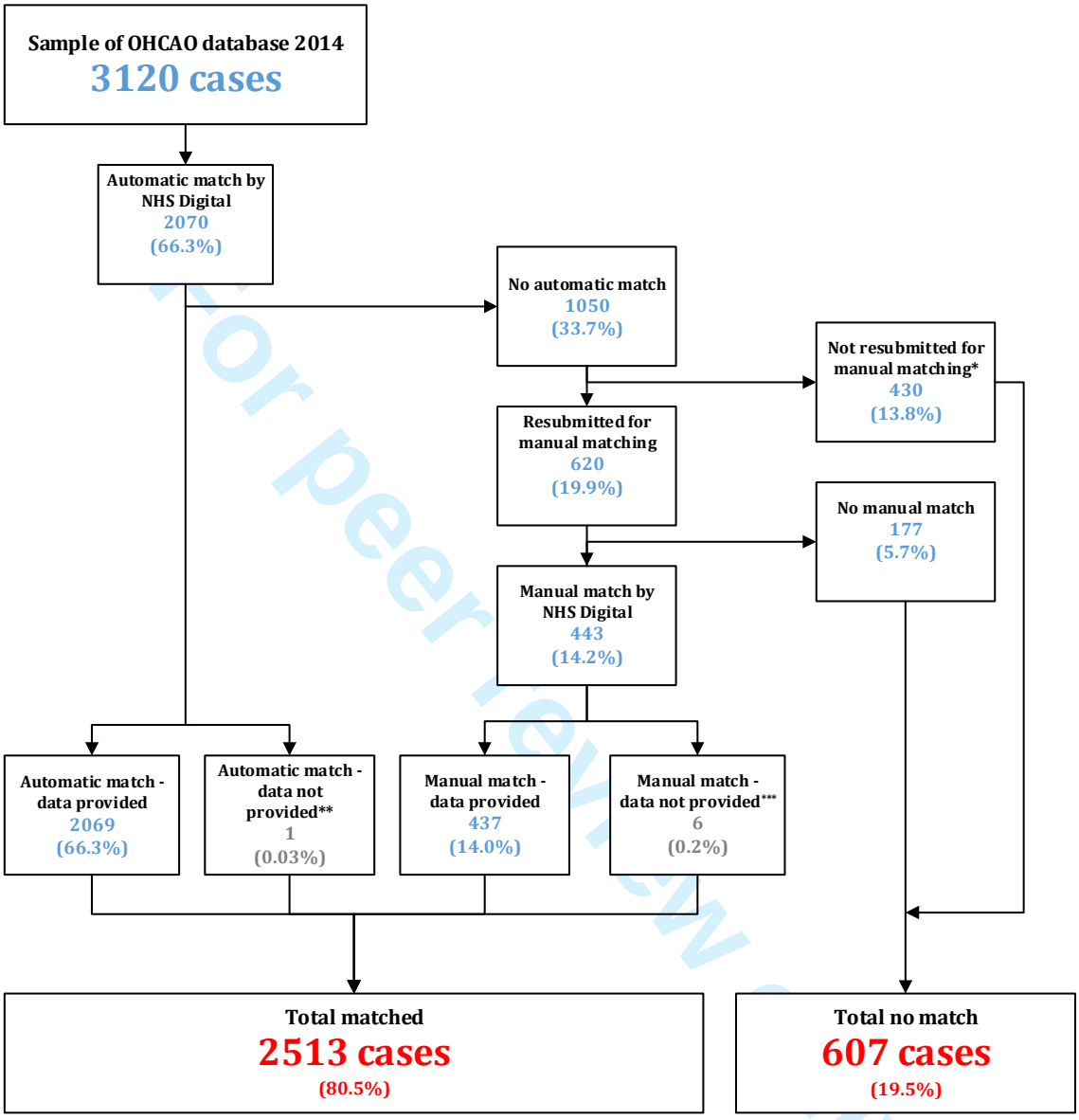
Descriptive data collected by the OHCAO project were generated for the combined OHCAO and NHS PDS dataset (NHS Digital matched cases and OHCAO non-matched cases). This included patient demographics, event data (arrest witness status), pre-EMS intervention (bystander CPR, public access defibrillation (PAD) use) and clinical information (initial rhythm, aetiology). Cases were allocated to 3 groups based on outcomes; *ROSC at anytime*, *survival to hospital discharge*, and *died*. Patients were categorised as ROSC at anytime if survival to hospital discharge was not confirmed but ROSC was recorded before or at hospital transfer. Patients were categorised as survival to hospital discharge where discharge was confirmed by OHCAO data. Finally, patients were categorised as died if there was no recorded ROSC (before or at hospital transfer) or survival to hospital discharge, with an ONS confirmed date of death. Within these outcome groups, the characteristics of matched and unmatched cases were compared descriptively to assess for differences. This assessed if match status was affected by these characteristics.

RESULTS

OHCAO data cleaning process

Of the 3120 cases transferred to NHS Digital, 2070 (66.3%) were automatically matched by the NHS Digital list cleaning algorithm while 1050 (33.7%) were not (Figure 1). 620 (19.9%) cases failing automatic matching were resubmitted for manual matching following which 437 (14.0%) were returned having been manually matched. 430 cases (13.8%) were not submitted for manual matching as there was little chance of a match due to missing data points (252 cases only had 1 data point out of surname, forename, DOB and home postcode and 178 cases did not have any data points). Overall, 2513 (80.5%) cases were matched of which 7 (0.2%) cases could not be released due to the patient being lost to follow-up (1 case, reason unknown) or the patient had registered a type 2 opt-out with NHS Digital, meaning that the patients' personal confidential data cannot be released by NHS Digital for reasons other than their own direct care (6 cases). 607 (19.5%) cases could not be matched due to insufficient data for matching (Table 2).

Figure 1: Data matching process linking OHCAO data with NHS PDS data through the NHS Digital list cleaning service



* Insufficient data from OHCAO database for NHS Digital to match these cases.
** Data not provided as patient lost to follow-up (reason unknown).
*** Data not provided due to patient registration of type 2 opt-out with NHS Digital.

Data points required for matching through NHS Digital list cleaning

The data point determining the highest match rate was NHS number (Table 1). 100% of cases with an NHS number were matched to the PDS database and therefore matched to

ONS mortality data with 99% of cases with an NHS number being automatically matched. However, only 31.7% of OHCAO cases had an NHS number.

Table 1: Total cases with each demographic data point collected by OHCAO project

		NHS No.	Surname	Forename	DOB	Postcode
Total OHCAO cases with data point (% of total 3120 cases)		989 (31.7%)	2699 (86.5%)	2693 (86.3%)	2700 (86.5%)	1626 (52.1%)
Match status (% of cases with specified data point)	Total matched	989 (100%)	2506 (92.8%)	2505 (93.0%)	2408 (89.2%)	1566 (96.3%)
	Auto match	979 (99.0%)	2070 (76.7%)	2070 (76.9%)	2070 (76.6%)	1364 (83.9%)
	Manual match	10 (1.0%)	436 (16.2%)	435 (16.2%)	338 (12.5%)	202 (12.4%)
	No match	0 (0.0%)	193 (7.2%)	188 (7.0%)	292 (10.8%)	60 (3.7%)

Approximately a quarter (27.8%) of the sample had all 5 data points allowing all of these to be automatically matched to NHS PDS data data (Table 2). 53.2% had 3 to 4 data points of which 93.4% and 95.7% were matched, respectively. Of these, all with NHS numbers were matched, whilst a combination of data points surname+forename+DOB+postcode and surname+forename+DOB resulted in match rates of 81.6% and 89.8%. However, cases where only 1 or 2 data points were provided were less likely to be matched (2.3% and 44.2% respectively). 178 (5.7%) cases had no OHCAO demographic data and therefore could not be matched.

Table 2: Combinations of data points required for linkage to the NHS PDS database

Number of data points (n cases, % of total)	Combinations	N (total 3120 cases)	
		Matched, 2513 (n, % of total in data point category)	Unmatched, 607 (n, % of total in data point category)
5 (868, 27.8%)	NHS+surname+forename+DOB+postcode	868 (100%)	0
4 (815, 26.1%)	NHS+surname+DOB+postcode	0	0
	NHS+forename+DOB+postcode	0	0
	NHS+surname+forename+postcode	3 (0.4%)	0
	NHS+surname+forename+DOB	112 (13.7%)	0
	surname+forename+DOB+postcode	665 (81.6%)	35 (4.3%)
	Total	**780 (95.7%)**	**35 (4.3%)**
3 (846, 27.1%)	NHS+surname+forename	1 (0.1%)	0
	NHS+surname+DOB	0	0
	NHS+surname+postcode	0	0
	NHS+forename+DOB	0	0
	NHS+forename+postcode	0	0
	NHS+DOB+postcode	0	0
	surname+forename+DOB	760 (89.8%)	44 (5.2%)
	surname+forename+postcode	27 (3.2%)	11 (1.3%)
	surname+DOB+postcode	1 (0.1%)	1 (0.1%)
	forename+DOB+postcode	1 (0.1%)	0
	total	**790 (93.4%)**	**56 (6.6%)**
2 (156, 5.0%)	NHS+surname	0	0
	NHS+forename	1 (0.6%)	0
	NHS+DOB	0	0
	NHS+postcode	0	0
	surname+forename	67 (42.9%)	82 (52.6%)
	surname+DOB	0	0
	surname+postcode	0	2 (1.3%)
	forename+DOB	0	1 (0.6%)
	forename+postcode	0	1 (0.6%)
	DOB+postcode	1 (0.6%)	1 (0.6%)
	total	**69 (44.2%)**	**87 (55.8%)**
1 (257, 8.2%)	NHS	4 (1.6%)	0
	surname	2 (0.8%)	18 (7.0%)
	forename	0	14 (5.4%)
	DOB	0	210 (81.7%)
	postcode	0	9 (3.5%)
	total	**6 (2.3%)**	**251 (97.7%)**
0 (178, 5.7%)	nil	0	178 (100%)

Data improvements after NHS Digital list cleaning and provision of ONS date of death data

Demographic improvements

After case matching, NHS Digital returned demographic data points (forename, surname, NHS number, home postcode) and ONS date of death if applicable. 1484 (47.6%) cases were not improved for any demographic data points (Table 3). These cases were those where complete demographic data were already collected by OHCAO (868 cases), matching failed (607 cases) or data could not be released by NHS Digital due to either the patient being lost to follow-up or the patient registering a type 2 opt-out with NHS Digital (6 of the 7 cases). All demographic data was already collected by OHCAO for 1 case out of these 7 cases and therefore was included in the aforementioned 868 cases. Lastly, for 3 cases OHCAO collected NHS+surname+forename+postcode and therefore these effectively could not be improved by matching as NHS Digital were not asked to provide DOB.

Of the 2249 cases with missing data, 1636 (72.7%) cases had demographic improvements following linkage. A quarter (25.8%) were improved by 1 demographic data point and a further quarter (26.4%) by 2 data points. Of the 7 that were improved by 3 data points (Table 3), OHCAO provided NHS number for 4 cases, surname for 2 cases and DOB, and postcode for 1 case.

Table 3: Number of data points added to OHCAO data after NHS Digital list cleaning

Number of demographic data points increased by list cleaning	Number of cases
0	1484 (47.6%)
1	804 (25.8%)
2	825 (26.4%)
3	7 (0.2%)

NHS Digital returned NHS numbers for 1518 (48.7%) cases in which it was not already collected by OHCAO (supplementary Table 1). OHCAO had already collected forename and surname in most cases (86.5% and 96.3% respectively) which were least improved following matching.

Survival data improvements following provision of ONS mortality data

30-day survival status (yes or no) using OHCAO data were confirmed for 1,682 (53.9%) cases (Table 4). 30-day survival was confirmed using OHCAO data if ambulance services provided a date of death or discharge date over 30 days after the OHCA incident date. Linking to ONS mortality data resulted in confirmed 30-day survival status (yes or no) for 2856 (91.5%) cases, a 37.6% improvement in 30-day survival status confirmation. Without ONS data, 30-day survival would be calculated as 0.4% while after linkage to ONS data the result was 9.3% overall 30-day survival.

Table 4: Comparison of 30-day survival calculation pre- and post- data linkage

30-day survival			
Dataset 1: OHCAO data		Dataset 2: Linked OHCAO and ONS data	
Yes	12 (0.4%)	Yes	290 (9.3%)
No	1670 (53.5%)	No	2566 (82.2%)
Unknown	1438 (46.1%)	Unknown	264 (8.5%)
Total	3120 (100%)	Total	3120 (100%)

Comparison of patient groups

All 3120 cases had resuscitation attempted. Table 5 displays data combining OHCAO and NHS PDS data. Cases were categorised in the following outcome group: *ROSC at anytime* (1184, 37.9%), *survival to hospital discharge* (253, 8.1%), and *died* (1584, 50.8%). For 99 cases (3.2%), there were insufficient data to be categorised in an outcome group and so these cases were categorised as *no group*.

All patients categorised as died or no group were matched to NHS PDS data while for ROSC at anytime cases 50.8% were matched and 90.1% of survival to hospital discharge cases were matched. Survival to hospital discharge patients were on average 8 years younger (55.2-55.7 years) than those categorised as ROSC at anytime (63.9 years or died (64.9 years). Additionally the bystander CPR rate was at least 10% higher for survival to hospital discharge patients compared to ROSC at anytime and died patients. The ROSC at anytime and died groups were comparable in all parameters as were the matched and unmatched cases within the ROSC at anytime group. However those not matched in the survival to hospital discharge groups showed comparatively higher

rates of defibrillator usage (20.0% vs. 5.7%) with lower rates of shockable rhythms (32.0% vs. 61.0%) compared to the matched group. However these figures are based on very few patients.

Table 5: Comparison of matched and unmatched OHCA case characteristics by outcome group

	Survival status as reported by ambulance services									
Variables	**ROSC at anytime** (1184, 37.9%)		**Survival to hospital discharge** (253, 8.1%)		**Died** (1584, 50.8%)		**No group** (99, 3.2%)		**Totals** (3120, 100%)	
	Match 602 (50.8%)	No match 582 (49.2%)	Match 228 (90.1%)	No match 25 (9.9%)	Match 1584 (100%)	No match (0.0%)	Match 99 (100%)	No match 0 (0%)	Match 2513 (80.5%)	No match 607 (19.5%)
Age (years)										
Mean (SD)	63.9 (24.0)	62.9 (23.2)	55.2 (24.3)	55.7 (18.0)	64.9 (24.4)	-	58.1 (28.1)	-	63.4 (24.7)	62.6 (23.1)
Median (IQR)	71.0 (27.5)	68.0 (31.0)	60.0 (52)	62.0 (20.3)	73.0 (28.0)	-	68.0 (40)	-	71.0 (29)	67.0 (30)
Sex										
Male	361 (60.0%)	338 (58.1%)	153 (67.1%)	12 (48.0%)	946 (59.7%)	0 (NA)	51 (51.5%)	0 (NA)	1511 (60.1%)	350 (57.7%)
Female	235 (39.0%)	168 (28.9%)	70 (30.7%)	5 (20.0%)	565 (35.7%)	0 (NA)	35 (35.4%)	0 (NA)	905 (36.0%)	173 (28.5%)
Missing	6 (1.0%)	76 (13.1%)	5 (2.2%)	8 (32.0%)	73 (4.6%)	0 (NA)	13 (13.1%)	0 (NA)	97 (3.9%)	84 (13.8%)
Witness status										
EMS	83 (13.8%)	61 (10.5%)	56 (24.6%)	3 (12.0%)	174 (11.0%)	0 (NA)	17 (17.2%)	0 (NA)	330 (13.1%)	64 (10.5%)
Bystander	206 (34.2%)	182 (31.3%)	84 (36.8%)	6 (24.0%)	418 (26.4%)	0 (NA)	26 (26.3%)	0 (NA)	734 (29.2%)	188 (31.0%)
Witness unspecified	57 (9.5%)	5 (0.9%)	3 (1.3%)	0 (0.0%)	78 (4.9%)	0 (NA)	4 (4.0%)	0 (NA)	142 (5.7%)	5 (0.8%)
Unwitnessed	171 (28.4%)	225 (38.7%)	58 (25.4%)	10 (40.0%)	635 (40.1%)	0 (NA)	26 (26.3%)	0 (NA)	890 (35.4%)	235 (38.7%)
Missing	85 (14.1%)	109 (18.7%)	27 (11.8%)	6 (24.0%)	279 (17.6%)	0 (NA)	26 (26.3%)	0 (NA)	417 (16.6%)	115 (18.9%)
Bystander CPR										
Yes	272 (45.2%)	272 (46.7%)	107 (46.9%)	15 (60.0%)	710 (44.8%)	0 (NA)	27 (27.3%)	0 (NA)	1116 (44.4%)	287 (47.3%)
No	254 (42.2%)	210 (36.1%)	86 (37.7%)	5 (20.0%)	637 (40.2%)	0 (NA)	41 (41.4%)	0 (NA)	1018 (40.5%)	215 (35.4%)
Missing	76 (12.6%)	100 (17.2%)	35 (15.4%)	5 (20.0%)	237 (15.0%)	0 (NA)	31 (31.3%)	0 (NA)	379 (15.1%)	105 (17.3%)
Bystander CPR rate*	52.4%	52.2%	62.2%	68.1%	50.4%	-	32.9%	-	51.1%	52.9%
Matched/unmatched Overall per group	52.3%		62.9%		50.4%		32.9%		51.5%	
PAD used										
Yes	12 (2.0%)	8 (1.4%)	13 (5.7%)	5 (20.0%)	13 (0.8%)	0 (NA)	0 (0.0%)	0 (NA)	38 (1.5%)	13 (2.1%)
No	430 (71.4%)	351 (60.3%)	154 (67.5%)	7 (28.0%)	1134 (71.6%)	0 (NA)	54 (54.5%)	0 (NA)	1772 (70.5%)	358 (59.0%)
Missing	160 (26.6%)	223 (38.3%)	61 (26.8%)	13 (52.0%)	437 (27.6%)	0 (NA)	45 (45.5%)	0 (NA)	703 (28.0%)	236 (38.9%)
Aetiology										
Cardiac	535 (88.9%)	408 (70.1%)	210 (92.1%)	18 (72.0%)	1428 (90.2%)	0 (NA)	84 (84.8%)	0 (NA)	2257 (89.8%)	426 (70.2%)
Traumatic	16 (2.7%)	38 (6.5%)	3 (1.3%)	1 (4.0%)	24 (1.5%)	0 (NA)	2 (2.0%)	0 (NA	45 (1.8%)	39 (6.4%)

Submersion	0 (0.0%)	2 (0.3%)	1 (0.4%)	0 (0.0%)	4 (0.3%)	0 (NA)	0 (0.0%)	0 (NA)	5 (0.2%)	2 (0.3%)
Drug Overdose	0 (0.0%)	2 (0.3%)	0 (0.0%)	0 (0.0%)	5 (0.3%)	0 (NA)	0 (0.0%)	0 (NA)	5 (0.2%)	2 (0.3%)
Asphyxia	20 (3.3%)	10 (1.7%)	6 (2.6%)	0 (0.0%)	36 (2.3%)	0 (NA)	3 (3.0%)	0 (NA)	65 (2.6%)	10 (1.6%)
Other	31 (5.1%)	122 (21.0%)	8 (3.5%)	6 (24.0%)	87 (5.5%)	0 (NA)	10 (10.1%)	0 (NA)	136 (5.4%)	128 (21.1%)
Initial Rhythm										
Asystole	228 (37.9%)	228 (39.2%)	35 (15.4%)	3 (12.0%)	818 (51.6%)	0 (NA)	18 (18.2%)	0 (NA)	1099 (43.7%)	231 (38.1%)
VF/VT	146 (24.3%)	86 (14.8%)	139 (61.0%)	8 (32.0%)	222 (14.0%)	0 (NA)	23 (23.2%)	0 (NA)	530 (21.1%)	94 (15.5%)
PEA	150 (24.9%)	75 (12.9%)	24 (10.5%)	3 (12.0%)	277 (17.5%)	0 (NA)	22 (22.2%)	0 (NA)	473 (18.8%)	78 (12.9%)
Bradycardia	1 (0.2%)	0 (0.0%)	1 (0.4%)	0 (0.0%)	7 (0.4%)	0 (NA)	1 (1.0%)	0 (NA)	10 (0.4%)	0 (0.0%)
Not recorded	13 (2.2%)	30 (5.2%)	11 (4.8%)	0 (0.0%)	40 (2.5%)	0 (NA)	3 (3.0%)	0 (NA)	67 (2.7%)	30 (4.9%)
Other	7 (1.2%)	17 (2.9%)	7 (3.1%)	0 (0.0%)	15 (0.9%)	0 (NA)	2 (2.0%)	0 (NA)	31 (1.2%)	17 (2.8%)
Missing	57 (9.5%)	146 (25.1%)	11 (4.8%)	11 (44.0%)	205 (12.9%)	0 (NA)	30 (30.3%)	0 (NA)	303 (12.1%)	157 (25.9%)

*Bystander CPR rate was calculated as the total number of patients receiving bystander CPR, divided by the total number of EMS treated OHCA events minus the EMS witnessed cardiac arrests.

Note: NA: not applicable.

Accuracy of OHCAO date of death data

In this sample OHCAO reported a date of death for 1178 (37.8%) cases from ambulance services, however for 64 (5.4%) of these cases ONS mortality data confirmed patients were still alive at the time of linkage. Of the 1942 (62.2%) cases where OHCAO could not confirm a date of death, 778 (40.1%) were confirmed alive at the time of linkage and 1164 (59.9%) had died (supplementary Table 2).

DISCUSSION

This study demonstrates the feasibility and value of using data linkage to significantly improve epidemiological data quality in an OHCA registry. Using a sample of 3120 OHCAO cases receiving EMS-attempted resuscitation, this study achieved an 80.5% match with NHS PDS and ONS mortality data. This enabled provision of registered death dates and survival status for more accurate calculation of 30-day survival. Results showed a 30-day survival rate of 9.3%, reducing unknown survival status from 46.1% to 8.5% (Table 4). Additionally, demographic data quality improved for 52.4% of cases (1636 cases). 1484 cases did not have improved demographic data after linkage but for 58.5% (868 cases) of these, complete data were already collected by OHCAO leaving 607 that could not be matched and 6 where data were not provided due to confidentiality or loss to follow up. NHS list cleaning provided NHS numbers for a further 1518 (48.7%) cases, increasing the potential to utilise data linkage to follow patients longitudinally after OHCA, e.g. by linking to Hospital Episode Statistics (HES) data. While the provision of postcodes for 942 (30.2%) cases allows comparison of OHCAs that occur at home verses those in public places and potentially assessment of multiple deprivation index effect on survival after OHCAs.

The variability of cardiac arrest survival is well documented.[4] Where data from ambulance services does not follow a standard procedure, data collection variability may have significant effects on data quality and comparability between services. Increasingly, core outcome sets for specific research areas are developed outlining minimum datasets for routine collection and create a level of standardisation to compare studies and allow formation of meta-analyses.[21] In the field of OHCAs, the Utstein guidelines have been developed.[17,22] However, a study investigating the level of

missing data within primary outcomes in 283 Cochrane Reviews of all areas of clinical practice found that over 50% of patient data were missing in 18% of reviews.[23] Furthermore, an analysis of 12 international OHCA registries' collection of data using Utstein templates found that although all registries collected core variables, there were differences in interpretation of the template and recorded 'unknown' for a mean of 4.8 variables and 'missing' for 1.9 variables.[24] Therefore minimum datasets are not sufficiently effective in reducing missing data.

The best data point provided by ambulance services to identify cases is the NHS number. It provides a unique identifier to resolve missing demographic data issues if no other demographic data are provided. 100% of OHCAO cases with an NHS number were matched to NHS PDS data, however, it was only available in a third (31.7%) of cases (Table 1). Logistical difficulties exist in ascertaining NHS number as it may not be available in the out-of-hospital setting. However, this study found that providing at least 3 to 4 demographic variables other than NHS number resulted in a match rate of up to 89.8%, depending on the combination and especially if forename and surname were provided. This also allowed provision of NHS number in 48.7% of cases where it was not collected by the OHCAO project (supplementary Table 1). If less than a threshold of 3 data points are provided, this study found a lower potential for data matching (0-44.2%, Table 2). These findings are in line with research showing that OHCA databases in different countries can successfully link OHCA patients to outcome databases where unique patient identifiers are readily available.[12,13] Whilst a study from the United States by Mumma and colleagues[25] showed limited feasibility for linking OHCA patients to longitudinal outcomes using probabilistic matching when there is no unique patient identifiers available and there is variability in completeness of patient demographic data. Furthermore, for 178 (5.7%) cases, no patient-identifiable demographic data were collected, preventing matching. Harron et al's[7] study linking mother and baby records using HES data created deterministic and probabilistic data links without using direct personal identifiers, instead using non-identifiable clinical and demographic variables to link cases. Such methods could be utilised where patient-identifiable demographic data are unavailable. Alternatively, Zwisler et al[9] implemented mandatory reporting to the Danish Cardiac Rehabilitation Database for all hospitals in Denmark providing cardiac rehabilitation, working towards creating a complete dataset.

In this sample of 3120 OHCA patients, survival to hospital discharge was 8.1% (Table 5), not dissimilar to the overall survival to hospital discharge rate of 8.7% in 2014 reported by the English ambulance services to the NHS England AQIs.[26] Half (50.8%) of OHCAs with attempted resuscitation died with no ROSC at anytime while 37.9% achieved ROSC at anytime but were not discharged from hospital. The characteristics of these patients were similar to those who achieved ROSC at anytime in both the matched and unmatched group. As survival status was not used to determine case allocation to the ROSC at anytime group (only ROSC status and lack of discharge confirmation), this may explain why match status does not confer any difference in characteristics of patients in this group. Survival to hospital discharge patients had 10% higher rates of bystander CPR with 58.1% having shockable rhythms. Initial shockable rhythms have been shown to independently predict survival to hospital discharge and therefore presents a significant survival advantage.[27] As 90% of discharged patients were matched, the data on characteristics of unmatched patients is based on only 9.9% of survival to hospital discharge patients and is insufficient to draw conclusions from.

The OHCAO project was able to collect a date of death for 1178 (37.8%) cases. Interestingly, a date of death was incorrectly stated for 64 cases where ONS mortality data confirmed patients were still alive at the time of linkage. This is an important finding as this shows the importance of data linkage to correct database errors. Such errors may lead to incorrect reporting of cardiac arrest survival as part of the NHS England AQIs which, in some cases, may lead to financial or other sanctions.

This study's strengths lie in its standardised procedures for OHCA case definition and data collection. The data points collected were based on established Utstein guidelines.[17] Exact data matches with NHS PDS and ONS mortality data established a high quality of case identification and survival status. The limitations were that 99 cases (3.2%) were matched but insufficient OHCAO project data were available to identify these patients as having achieved ROSC at anytime, survival to discharge, or as having died. Additionally, 178 cases collected by OHCAO had no demographic data points and therefore could not be matched.

CONCLUSIONS

This study shows the feasibility of linking data from the UK OHCAO project to NHS patient demographic and date of death data held on the PDS database, via the NHS Digital list cleaning service. This provided high quality data on survival status revealing a 30-day survival of 9.3% with reduction of unknown 30-day survival status from 46% to 8.5%. Knowledge of 30-day survival from OHCAs may be of use to the NHS in terms of resource planning and directing service provision. Demographic data were improved for over half of cases and can be used as a means of creating a registry of patients sustaining OHCA that may be followed longitudinally.

Missing NHS numbers are a significant obstacle in creating an OHCAO database as these enable linkage to other databases. This study found that if at least forename and surname is collected with one other demographic data point, there is still a high chance of retrieving NHS numbers. Future research could look at the utility of the linkage process in following patients longitudinally after cardiac arrest for example by linking to HES data. This may provide more reliable survival to hospital discharge outcome data.

Collaborators Dr Sukhdeep Dosanjh, Warwick Clinical Trials Unit, University of Warwick; Theresa Foster, East of England Ambulance Service NHS Trust; Frank Mersom, East of England Ambulance Service NHS Trust; Gurkamal Francis, London Ambulance Service NHS Trust; Michelle O'Rourke, North East Ambulance Service NHS Trust; Clare Bradley, North West Ambulance Service NHS Trust; Philip King, South Central Ambulance Service NHS Trust; Ed England, South Central Ambulance Service NHS Trust; Patricia Bucher, South East Coast Ambulance Service NHS Trust; Jessica Lynde, South Western Ambulance Service NHS Trust; Nancy Loughlin, South Western Ambulance Service NHS Trust; Jenny Lumley-Holmes, West Midlands Ambulance Service NHS Trust; Dr Julian Mark, Yorkshire Ambulance Service NHS Trust.

Contributions GDP designed the study. SJB, CJ, CH, ANS, KK, SB, RS, and IG contributed to data collection. SR analysed the data and wrote the initial draft of the paper. SR, GDP and SJB were involved in further drafting of the paper. All authors participated in interpreting the data, revising the paper for critically important intellectual content and gave final approval of the submitted version.

Acknowledgements National Ambulance Services Clinical Quality Group and the National Ambulance Research Steering Group.

Funding This work was supported by research grants from the British Heart Foundation and Resuscitation Council (UK).

Competing interests All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: SJB, TPB, CJ, CH, SJBM and GDP are employed by the University of Warwick, which receives grants from the British Heart Foundation and the Resuscitation Council (UK) for the conduct of the OHCAO project; no other relationships or activities that could appear to have influenced the submitted work.

Ethics approval Ethical approval for the OHCAO project was gained from the National Research Ethics Committee South Central, reference number 13/SC/0361. Confidential Advisory Group (CAG) approval was granted, reference number ECC 8-04(c)/2013, to collect identifiable patient information where it is not practical to obtain consent.

Data sharing statement Please refer to the OHCAO project website for information relating to data sharing requests: http://www2.warwick.ac.uk/fac/med/research/hscience/ctu/trials/ohcao/health/data/data_sharing/

REFERENCES

1. van der Veer SN, de Keizer NF, Ravelli AC, Tenkink S, Jager KJ. Improving quality of care. A systematic review of how medical registries provide information feedback to health care providers. Int J Med Inform. 2010;79(5):305-23.

2. Resuscitation Council (UK), NHS England, British Heart Foundation. Consensus paper on out-of-hospital cardiac arrest in England. Resuscitation Council; 2014.

3. Ambulance Service Association. National cardiac arrest audit report. London: Ambulance Service Association; 2006.

4. Perkins GD, Cooke MW. Variability in cardiac arrest survival: The NHS ambulance service quality indicators. Emerg Med J. 2012;29(1):3-5.

5. Berdowski J, Berg RA, Tijssen JG, Koster RW. Global incidences of out-of-hospital cardiac arrest and survival rates: Systematic review of 67 prospective studies. Resuscitation. 2010; 81:1479-87.

6. Lilford R, Mohammed MA, Spiegelhalter D, Thomson R. Use and misuse of process and outcome data in managing performance of acute medical care: Avoiding institutional stigma. Lancet. 2004;363(9415):1147-54.

7. Harron K, Gilbert R, Cromwell D, van der Meulen J. Linking data for mothers and babies in de-identified electronic health data. PLoS One. 2016;11(10):e0164667.doi:10.1371/journal. pone.0164667

8. Hickey GL, Bridgewater B, Grant SW, Deanfield J, Parkinson J, Bryan AJ, et al. National registry data and record linkage to inform postmarket surveillance of prosthetic aortic valve models over 15 years. JAMA Intern Med. 2017;177(1):79-86.

9. Zwisler AD, Rossau HK, Nakano A, Foghmar S, Eichhorst R, Prescott E, et al. The Danish cardiac rehabilitation database. Clin Epidemiol. 2016;8:451-56.

10. Roman M, Graff-Iversen S, Weiderpass E, Vangen S, Sakshaug S, Hofvind S, et al. Postmenopausal hormone therapy and breast cancer prognostic characteristics: A linkage between nationwide registries. Cancer Epidemiol Biomarkers Prev. 2016;25(11):1464-73.

11. Finn JC, Jacobs IG, Holman CD, Oxer HF. Outcomes of out-of-hospital cardiac arrest patients in Perth, Western Australia, 1996-1999. Resuscitation. 2001;51:247-55.

12. Hamilton A, Steinmetz J, Wissenberg M, Torp-Pedersen C, Lippert FK, Hove L, et al. Association between prehospital physician involvement and survival after out-of-hospital cardiac arrest: A Danish nationwide observational study. Resuscitation. 2016:108:95-101.

13. Shuvy M, Morrison LJ, Koh M, Qiu F, Buick JE, Dorian P, et al. Long-term clinical outcomes and predictors for survivors of out-of-hospital cardiac arrest. Resuscitation. 2017:112:59-64.

14. Perkins GD, Brace-McDonnell SJ, on behalf of the OHCAO Project Group. The UK out of hospital cardiac arrest outcome (OHCAO) project. BMJ Open. 2015;5:e008736.doi:10.1136/ bmjopen-2015-008736

15. Hawkes C, Booth S, Ji C, Brace-McDonnell SJ, Whittington A, Mapstone J, et al. Epidemiology and outcomes from out-of-hospital cardiac arrests in England. Resuscitation. 2017;110;133-40.

16. Joint Royal Colleges Ambulance Liaison Committee and Association of Ambulance Chief Executives. UK Ambulance Services Clinical Practice Guidelines 2016. Bridgwater: Class Professional Publishing. 2016.

17. Jacobs I, Nadkarni V, Bahr J, Berg RA, Billi JE, Bossaert L, et al. Cardiac arrest and cardiopulmonary resuscitation outcome reports: Update and simplification of the Utstein templates for resuscitation registries. Resuscitation. 2004;63(3):233-49.

18. Jaro MA. Probabilistic linkage of large public health data files. Stat Med. 1995;14(5-7):491-98.

19. Meray N, Reisma JB, Ravelli AC, Bonsel GJ. Probabilistic record linkage is a valid and transparent tool to combine databases without a patient identification number. J Clin Epidemiol. 2007;60:883-91.

20. User guide to mortality statistics. July 2016. Office for National Statistics. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/methodologies/userguidetomortalitystatistics

21. Kirkham JJ, Gorst S, Altman DG, Blazeby JM, Clarke M, Devane D, et al. Core Outcome Set-STAndards for Reporting: The COS-STAR statement. PLoS Med. 2016;13(10):e1002148.doi:1 0.1371/journal.pmed.1002148

22. Perkins GD, Jacobs IG, Nadkarni VM, Berg RA, Bhanji F, Biarent D, et al. Cardiac arrest and cardiopulmonary resuscitation outcome reports: update of the utstein resuscitation registry templates for out-of-hospital cardiac arrest. Resuscitation. 2015;96:328-40.

23. Kirkham JJ, Gargon E, Clarke M, Williamson PR. Can a core outcome set improve the quality of systematic reviews? — a survey of the co-ordinating editors of cochrane review groups. Trials. 2013;14(21):doi:10.1186/1745-6215-14-21

24. Nishiyama C, Brown SP, May S, Iwami T, Koster RW, Beesems SG, et al. Apples to apples or apples to oranges? International variation in reporting of process and outcome of care for out-of-hospital cardiac arrest. Resuscitation. 2014;85:1599-609.

25. Mumma BE, Diercks DB, Danielsen B, Holmes JF. Probabilistic linkage of prehospital and outcomes data in out-of-hospital cardiac arrest. Prehosp Emerg Care. 2015;19(3):358-64.

26. Ambulance Quality Indicators, 2015. http://www.england.nhs.uk/statistics/statistical-work-areas/ambulance-quality-indicators/ [Accessed 31 January 2017]

27. Wah W, Wai KL, Pek PP, Ho AF, Alaskaf O, Chia MY, et al. Conversion to shockable rhythms during resuscitation and survival for out-of-hospital cardiac arrest. Am J Emerg Med. 2017;35:206-13.

SUPPLEMENTARY DATA

Table 1: Specific data points increased following linkage

Data point	Number increased by linkage (n, % of total cases)
NHS	1518 (48.7%)
Surname	7 (0.2%)
Forename	8 (0.3%)
Postcode	942 (30.2%)

Table 2: Comparison of date of death confirmed by OHCAO and ONS data

		ONS confirmed survival status		Total (n, % total 3120 cases)
		Dead	Alive	
OHCAO project date of death provided	yes	1114	64	1178 (37.8%)
	No	1164	778	1942 (62.2%)
Totals (n, % total 3120 cases)		2278 (73.0%)	842 (27.0%)	3120 (100%)

The RECORD statement – checklist of items, extended from the STROBE statement, that should be reported in observational studies using routinely collected health data.

	Item No.	STROBE items	Location in manuscript where items are reported	RECORD items	Location in manuscript where items are reported
Title and abstract					
	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found	(a) PAGE 1 (b) PAGE 2	RECORD 1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included. RECORD 1.2: If applicable, the geographic region and timeframe within which the study took place should be reported in the title or abstract. RECORD 1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract.	1.1 PAGE 1 1.2 PAGE 2 1.3 PAGES 1-2
Introduction					
Background rationale	2	Explain the scientific background and rationale for the investigation being reported	PAGES 3-4		
Objectives	3	State specific objectives, including any prespecified hypotheses	PAGE 4		
Methods					
Study Design	4	Present key elements of study design early in the paper	PAGES 4-5		
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	PAGES 4-6		
Participants	6	*(a) Cohort study* - Give the		RECORD 6.1: The methods of study	

		eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study* - Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study* - Give the eligibility criteria, and the sources and methods of selection of participants *(b) Cohort study* - For matched studies, give matching criteria and number of exposed and unexposed *Case-control study* - For matched studies, give matching criteria and the number of controls per case	(a) PAGES 4-5 (b) N/A	population selection (such as codes or algorithms used to identify subjects) should be listed in detail. If this is not possible, an explanation should be provided. RECORD 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided. RECORD 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical display to demonstrate the data linkage process, including the number of individuals with linked data at each stage.	6.1 PAGE 6 6.2 PAGES 6-7 6.3 PAGE 9
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.	PAGE 7	RECORD 7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be reported, an explanation should be provided.	PAGES 7-8
Data sources/ measurement	8	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	PAGES 5-8		
Bias	9	Describe any efforts to address potential sources of bias	PAGE 5		

Study size	10	Explain how the study size was arrived at	PAGE 5		
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why	PAGES 7-8		
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) *Cohort study* - If applicable, explain how loss to follow-up was addressed *Case-control study* - If applicable, explain how matching of cases and controls was addressed *Cross-sectional study* - If applicable, describe analytical methods taking account of sampling strategy (e) Describe any sensitivity analyses	PAGES 7-8		
Data access and cleaning methods		..		RECORD 12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population. RECORD 12.2: Authors should provide information on the data cleaning methods used in the study.	12.1 PAGES 5-8 12.2 PAGE 4
Linkage		..		RECORD 12.3: State whether the	

				study included person-level, institutional-level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.	PAGES 6-7
Results					
Participants	13	(a) Report the numbers of individuals at each stage of the study (*e.g.*, numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed) (b) Give reasons for non-participation at each stage. (c) Consider use of a flow diagram	(a) PAGE 9 (b) PAGE 9 (c) PAGE 9	RECORD 13.1: Describe in detail the selection of the persons included in the study (*i.e.*, study population selection) including filtering based on data quality, data availability and linkage. The selection of included persons can be described in the text and/or by means of the study flow diagram.	PAGE 9
Descriptive data	14	(a) Give characteristics of study participants (*e.g.*, demographic, clinical, social) and information on exposures and potential confounders (b) Indicate the number of participants with missing data for each variable of interest (c) *Cohort study* - summarise follow-up time (*e.g.*, average and total amount)	(a) PAGES 15-16 (b) PAGES 15-16 (c) N/A		
Outcome data	15	*Cohort study* - Report numbers of outcome events or summary measures over time *Case-control study* - Report numbers in each exposure category, or summary measures of exposure *Cross-sectional study* - Report numbers of outcome events or	PAGE 13		

		summary measures			
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorized (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	(a) COVERED ACROSS PAGES 8-17 (b) N/A (c) N/A		
Other analyses	17	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses	PAGES 15-17		
Discussion					
Key results	18	Summarise key results with reference to study objectives	PAGE 17		
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	PAGE 19	RECORD 19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they pertain to the study being reported.	PAGE 19
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	PAGE 20		

Generalisability	21	Discuss the generalisability (external validity) of the study results	PAGE 20		
Other Information					
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	PAGE 21		
Accessibility of protocol, raw data, and programming code		..		RECORD 22.1: Authors should provide information on how to access any supplemental information such as the study protocol, raw data, or programming code.	PAGE 22

*Reference: Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM, the RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Medicine* 2015; in press.

BMJ Open

Data Quality and 30-day Survival for Out-of-Hospital Cardiac Arrest in the UK Out-of-Hospital Cardiac Arrest Registry: A Data Linkage Study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-017784.R1
Article Type:	Research
Date Submitted by the Author:	02-Aug-2017
Complete List of Authors:	Rajagopal, Sangeerthana; University of Warwick Warwick Medical School, Warwick Clinical Trials Unit Booth, Scott; University of Warwick Warwick Medical School, Warwick Clinical Trials Unit Brown, Terry; University of Warwick Warwick Medical School, Warwick Clinical Trials Unit Ji, Chen; University of Warwick Warwick Medical School, Warwick Clinical Trials Unit Hawkes, Claire; University of Warwick Warwick Medical School, Warwick Clinical Trials Unit Siriwardena, Aloysius; University of Lincoln, School of Health and Social Care Kirby, Kim; South Western Ambulance Service NHS Foundation Trust Black, Sarah; South Western Ambulance Service NHS Foundation Trust, Spaight, Robert; East Midlands Ambulance Service NHS Trust Gunson, Imogen; West Midlands Ambulance Service NHS Trust Brace-McDonnell, Samantha; University of Warwick, Warwick Clinical Trials Unit Perkins, Gavin; University of Warwick, Warwick Clinical Trials Unit
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Emergency medicine
Keywords:	Cardiac arrest, Emergency medical services, Medical record linkage, Out-of-hospital cardiac arrest, Resuscitation

Data Quality and 30-day Survival for Out-of-Hospital Cardiac Arrest in the UK Out-of-Hospital Cardiac Arrest Registry: A Data Linkage Study

Sangeerthana Rajagopal[a,b], Scott J Booth[a], Terry P Brown[a], Chen Ji[a], Claire Hawkes[a], A Niroshan Siriwardena[c], Kim Kirby[d], Sarah Black[d], Robert Spaight[e], Imogen Gunson[f], Samantha J Brace-McDonnell[a,b], Gavin D Perkins[a,b], on behalf of OHCAO collaborators

[a]Warwick Clinical Trials Unit, University of Warwick, Coventry CV4 7AL, UK

[b]Heart of England NHS Foundation Trust, Birmingham B9 5SS, UK

[c]University of Lincoln, Lincolnshire LN6 7TS, UK

[d]South Western Ambulance Service NHS Foundation Trust, Exeter EX2 7HY, UK

[e]East Midlands Ambulance Service NHS Trust, Nottingham NG8 6PY, UK

[f] West Midlands Ambulance Service NHS Foundation Trust, Brierley Hill, DY5 1LX, UK

Correspondence to Professor Gavin D Perkins; G.D.Perkins@warwick.ac.uk

Word count 3995

ABSTRACT

Objectives: The Out-of-Hospital Cardiac Arrest Outcomes (OHCAO) project aims to understand the epidemiology and outcomes of out-of-hospital cardiac arrests (OHCA) across the UK. This data linkage study is a sub-project of OHCAO. The aim was to establish the feasibility of linking OHCAO data to National Health Service (NHS) patient demographic data and Office for National Statistics (ONS) date of death data held on the NHS Personal Demographics Service (PDS) database to improve OHCAO demographic data quality and enable analysis of 30-day survival from OHCA.

Design and setting: Data were collected from 1st January 2014 to 31st December 2014 as part of a prospective, observational study of OHCA attended by ten English NHS Ambulance Services. 28,729 OHCA cases had resuscitation attempted by Emergency Medical Services and were included in the study. Data linkage was carried out using a data linkage service provided by NHS Digital, a national provider of health-related data. To assess data linkage feasibility a random sample of 3120 cases was selected. The sample was securely transferred to NHS Digital to be matched using OHCAO patient demographic data to return previously missing demographic data and provide ONS date of death data.

Results: A total of 2513 (80.5%) OHCAO cases were matched to patients in the NHS PDS database. Using the linkage process, missing demographic data were retrieved for 1636 (72.7%) out of 2249 OHCAO cases that had previously incomplete demographic data. Returned ONS date of death data allowed analysis of 30-day survival status. The results showed a 30-day survival rate of 9.3%, reducing unknown survival status from 46.1% to 8.5%.

Conclusions: In this sample, data linkage between the OHCAO registry and NHS PDS database was shown to be feasible, improving demographic data quality and allowing analysis of 30-day survival status.

ARTICLE SUMMARY

Strengths and limitations of this study

- Data points collected as part of the OHCAO project were based on established Utstein guidelines.
- The quality of demographic data collected by the OHCAO project was first improved through a list cleaning and patient status service provided by NHS Digital.
- Following list cleaning, exact data matches with Office for National Statistics (ONS) date of death data allowed calculation of 30-day survival status.
- Provision of NHS numbers from OHCAO and NHS digital provides potential for following long-term survival outcomes in OHCA patients through data linkage.
- Improved data linkage is reliant on improved data capture of patient demographic data by ambulance services.

INTRODUCTION

Every year in the United Kingdom (UK) there are around 60,000 out-of-hospital cardiac arrests (OHCA) attended by Emergency Medical Services (EMS) of which approximately 28,000 have resuscitation attempted.[1,2] This group suffers significant mortality and morbidity,[3,4] and improving outcomes from OHCA remains a worldwide research priority.[5]

Collecting high quality data is essential as this forms the basis of decisions that ultimately impact on changes in care and healthcare resource allocation. Since 2011, survival to hospital discharge rates for OHCA have been reported as part of the National Health Service (NHS) England Ambulance Quality Indictors (AQIs), with significant variation reported ranging from 2.2% to 12.0%.[6] Regional variation in survival rates have also been observed worldwide.[7-9] Lilford et al highlighted that an important source of variation in reporting outcomes can be traced to the quality of data that results are based on.[10] Collecting survival to discharge data in England is a challenging process for ambulance services as it involves tracking the patients survival status directly with

hospital emergency departments, which is time consuming and can be hindered by governance issues.[11,12]

Data collected in international OHCA registries enables comparisons of OHCA epidemiology and outcomes across different EMS systems.[13-15] The Utstein guidelines provide a structured template for collecting data on OHCA processes to support such comparisons.[16] To facilitate ease of reporting the updated Utstein guidelines recommend collecting either 30-day survival or survival to hospital discharge as a core outcome.[17] The research literature suggests most international registries are able to report either of these OHCA outcome measures.[13-15] A recent example is the EuReCa ONE study which aimed to benchmark OHCA incidence, process and outcomes across 27 European countries and reported a combined survival to discharge or 30-day survival rate which ranged between 1.1% and 30.8%.[15]

Data linkage methodology has increasingly been used in medical research to establish outcomes. It involves linking information together from different sources that belong to the same individual.[18] Data linkage has been utilised by regional and national OHCA databases to confirm survival status through linkage with mortality databases.[4,19,20] Data linkage can address missing data issues, providing a centralised, high quality database for research and service appraisal with the potential to allow longitudinal surveillance of OHCA patients.

The Out-of-Hospital Cardiac Arrest Outcomes (OHCAO) project is funded by the Resuscitation Council (UK), British Heart Foundation and managed by the University of Warwick. It is a prospective observational study investigating the epidemiology and outcomes of OHCA patients across the UK.[21,22] This paper presents a sub-project of the OHCAO project aiming to establish the feasibility of linking OHCAO registry data to NHS patient demographic data and Office for National Statistics (ONS) mortality data through the NHS Digital list cleaning and patient status service.

METHODS

Setting

The OHCAO project established a national UK OHCA registry to collect process and outcome data to facilitate OHCA research and quality improvement. Detailed information about the OHCAO project is available in the study protocol.[21] The ten English NHS ambulance services collecting data for the OHCAO project cover approximately 54 million people, equating to 99.7% of the England population and 83.9% of the UK population.[23] Data were collected from 1st January 2014 to 31st December 2014 on 28,729 patients suffering OHCA in whom resuscitation was attempted by statutory EMS (an incidence rate of 53.2 per 100,000 of the English population).[22] This figure was reached after excluding individuals who achieved ROSC before arrival of EMS (n=1711) and where resuscitation was not attempted as per national guidelines[24] due to the presence of a do not attempt resuscitation order (n=387), or signs incompatible with life or where resuscitation attempts would be futile (n=5403).

Aims & Objectives

The overall aim of this project was to investigate the feasibility of linking a sample of OHCAO 2014 data to NHS patient demographic data and ONS date of death data held on the NHS Personal Demographics Service (PDS) database, using the NHS Digital list cleaning and patient status service, to improve OHCAO demographic data quality and allow calculation of 30-day survival from OHCA. The objectives were to (1) assess the match rate of combinations of OHCAO patient demographic variables in the sample (NHS number, surname, forename, date of birth (DOB), and home postcode) for linking to the NHS PDS database through NHS Digital list cleaning; (2) assess improvements in the completeness of OHCAO patient demographic variables through NHS Digital list cleaning; (3) create a linked OHCAO and NHS PDS database allowing analysis of 30-day survival from OHCA.

OHCAO project data collection

Core and supplemental Utstein variables were collected encompassing demographic, system, process and outcome data.[16] Each ambulance service has their own methods for

OHCA case ascertainment e.g. electronic searches of patient report form databases for diagnostic codes indicating cardiac arrest. A trained member of the ambulance service clinical audit team entered eligible cases into a cardiac arrest database, followed by data cleaning and verification processes. Survival to hospital discharge data was collected directly from hospitals by the clinical audit team if data sharing protocols were in place. Each ambulance service uploaded their data via a secure server to the OHCAO registry which is stored at the University of Warwick.

OHCAO data sample

To assess feasibility whilst minimising costs associated with data linkage, the analysis presented here represents a 10.9% sample of the 2014 OHCAO data, comprising 3120 OHCA patients. To avoid selection bias the sample was selected using simple random sampling and stratified by ambulance service.

OHCAO data linkage to ONS mortality data

OHCAO to NHS PDS data linkage approval was received after submitting an application to the NHS Digital Data Access Request Service; additional approval was obtained from ONS for the release of mortality data. OHCAO submitted 3120 cases to NHS Digital, via the NHS Digital secure transfer system, detailing the following patient demographic variables of varying completeness: NHS number, surname, forename, DOB and home postcode.

NHS Digital is the national provider of data relating to health and social care in England. OHCAO used the NHS Digital list cleaning and patient status service. The list cleaning service was used to validate submitted demographic data to ensure accuracy and improve data linkage outcomes. Validation was achieved by NHS Digital matching submitted demographic variables to NHS patient demographic data held on the PDS database. The PDS database is a national electronic database containing NHS patient demographic information, including NHS number, name and address. For each matched case NHS Digital were asked to provide OHCAO with the following patient demographic information: NHS number, surname, forename, and home postcode. These

data were used to improve the percentage of missing data for these variables in the OHCAO sample.

NHS Digital utilised both automatic and manual matching techniques, using a combination of deterministic and probabilistic data linkage methods.[25,26] In deterministic data linkage it is decided *a priori* what combination of patient identifiers to match on (e.g. NHS number and DOB) and only complete agreement between records are considered a match. In probabilistic data linkage weights are assigned to different patient identifiers (based on their discriminatory power) to assess the probability that two records are a match.[27] Cases were initially submitted for automatic matching which used a decision tree algorithm to provide matches. A subset of cases that failed automatic matching were resubmitted for manual matching.

As part of the patient status service NHS Digital was also able to provide a date of death for deceased patients. The date of death data was held in the NHS PDS database and was sourced from ONS mortality data. OHCAO required information on deaths from 1st January 2014 until 31st January 2015. This was utilised to calculate 30-day survival. Where no date of death was provided the patient was categorised as alive.

ONS date of death data

ONS mortality data contains all deaths registered in England and Wales. When a person dies a formal medical certificate of death is produced, usually by a doctor, and which includes date of death. There is then a legal requirement for the death to be registered with the Registrar of Births, Deaths and Marriages through the local register office. The registration is typically performed by a close relative. The certification, and subsequent registration, of death may be delayed if the death is referred to a coroner for investigation (e.g. if cause of death is unknown). However, the majority of deaths in England and Wales are registered within 5 days of the death date.[28,29] ONS receives death data in electronic form directly from register offices. All data received is subject to both initial and routine data quality and validation processes and is collected in line with the Statistics and Registration Service Act 2007.[28]

Analysis

An analysis was conducted to assess how particular demographic data points enabled linkage with NHS PDS data and whether data linkage improved the completeness of patient demographic data. This was done descriptively with breakdowns of data linkage match rates for all combinations of the OHCAO demographic variables sent to NHS Digital for data linkage.

The combined linked dataset was analysed to investigate 30-day survival rates calculated by evaluating if patients were alive ≥30 calendar days from the EMS OHCA incident date. The analysis was carried out pre and post linkage, illustrating linkage effects. 30-day survival was calculated using OHCAO data where there was a date of death or date discharged >30 days after the OHCA incident date. Where there was an OHCAO date of death ≤30 days after the OHCA incident date or further ambulance service data indicating the patient was deceased on the day of the OHCA incident date (e.g. hospital code indicating patient deceased and not conveyed to hospital) the patient was categorised as not surviving to 30 days. All other cases were categorised as unknown for patient 30-day survival status. For the combined linked dataset, cases that were linked to ONS mortality data were categorised as 30-day survival where there was no date of death or where a date of death was provided that was >30 days after the OHCA incident date. Where there was an ONS date of death ≤30 days after the OHCA incident date the patient was categorised as not surviving to 30 days. Where there was a contradiction in patient survival status between OHCAO data and ONS mortality data then ONS mortality data superseded OHCAO data.

RESULTS

OHCAO data cleaning process

Of the 3120 cases transferred to NHS Digital, 2070 (66.3%) were automatically matched by the NHS Digital list cleaning algorithm while 1050 (33.7%) were not (Figure 1). 620 (19.9%) cases failing automatic matching were resubmitted for manual matching following which 437 (14.0%) were returned having been manually matched. 430 cases (13.8%) were not resubmitted for manual matching as there was little chance of a match due to missing data points (252 cases only had 1 data point out of surname, forename, DOB and home postcode and 178 cases did not have any data points). Overall, 2513 (80.5%) cases were matched of which 7 (0.2%) cases could not be released due to the patient being lost to follow-up (1 case, reason unknown) or the patient had registered a type 2 opt-out with NHS Digital, meaning that the patients' personal confidential data could not be released by NHS Digital for reasons other than their own direct care (6 cases). 607 (19.5%) cases could not be matched due to insufficient data for matching.

Data points required for matching through NHS Digital list cleaning

The percentage of each available demographic data point in the random sample of 3120 cases was similar to the percentage of each available demographic data point in all 28,729 cases for 2014 (Table 1). The data point determining the highest match rate was NHS number. 100% of cases with an NHS number were matched to the PDS database and therefore matched to ONS mortality data with 99% of cases with an NHS number being automatically matched. However, only 31.7% of OHCAO cases had an NHS number.

Table 1: Total cases with each demographic data point collected by OHCAO project

		NHS No.	Surname	Forename	DOB	Postcode
Total OHCAO cases with data point (% of total 28,729 cases)		9510 (33.1%)	24,814 (86.4%)	24,686 (85.9%)	24,956 (86.9%)	15,017 (52.3%)
Total OHCAO sample cases with data point (% of total 3120 cases)		989 (31.7%)	2699 (86.5%)	2693 (86.3%)	2700 (86.5%)	1626 (52.1%)
Match status (% of sample cases with specified data point)	Total matched	989 (100%)	2506 (92.8%)	2505 (93.0%)	2408 (89.2%)	1566 (96.3%)
	Auto match	979 (99.0%)	2070 (76.7%)	2070 (76.9%)	2070 (76.6%)	1364 (83.9%)
	Manual match	10 (1.0%)	436 (16.2%)	435 (16.2%)	338 (12.5%)	202 (12.4%)
	No match	0 (0.0%)	193 (7.2%)	188 (7.0%)	292 (10.8%)	60 (3.7%)

Approximately a quarter (27.8%) of the sample had all 5 data points allowing a match to NHS PDS data (Table 2). 53.2% had 3 to 4 data points of which 93.4% and 95.7% were matched, respectively. Of these, all with NHS numbers were matched, whilst a combination of data points surname+forename+DOB+postcode and surname+forename+DOB resulted in match rates of 81.6% and 89.8%. However, cases where only 1 or 2 data points were provided were less likely to be matched (2.3% and 44.2%, respectively). 178 (5.7%) cases had no OHCAO demographic data and could not be matched.

Table 2: Combinations of data points required for linkage to the NHS PDS database

Number of data points (n cases, % of total)	Combinations	N (total 3120 cases)	
		Matched, 2513 (n, % of total in data point category)	Unmatched, 607 (n, % of total in data point category)
5 (868, 27.8%)	NHS+surname+forename+DOB+postcode	868 (100%)	0
4 (815, 26.1%)	NHS+surname+DOB+postcode	0	0
	NHS+forename+DOB+postcode	0	0
	NHS+surname+forename+postcode	3 (0.4%)	0
	NHS+surname+forename+DOB	112 (13.7%)	0
	surname+forename+DOB+postcode	665 (81.6%)	35 (4.3%)
	Total	**780 (95.7%)**	**35 (4.3%)**
3 (846, 27.1%)	NHS+surname+forename	1 (0.1%)	0
	NHS+surname+DOB	0	0
	NHS+surname+postcode	0	0
	NHS+forename+DOB	0	0
	NHS+forename+postcode	0	0
	NHS+DOB+postcode	0	0
	surname+forename+DOB	760 (89.8%)	44 (5.2%)
	surname+forename+postcode	27 (3.2%)	11 (1.3%)
	surname+DOB+postcode	1 (0.1%)	1 (0.1%)
	forename+DOB+postcode	1 (0.1%)	0
	total	**790 (93.4%)**	**56 (6.6%)**
2 (156, 5.0%)	NHS+surname	0	0
	NHS+forename	1 (0.6%)	0
	NHS+DOB	0	0
	NHS+postcode	0	0
	surname+forename	67 (42.9%)	82 (52.6%)
	surname+DOB	0	0
	surname+postcode	0	2 (1.3%)
	forename+DOB	0	1 (0.6%)
	forename+postcode	0	1 (0.6%)
	DOB+postcode	1 (0.6%)	1 (0.6%)
	total	**69 (44.2%)**	**87 (55.8%)**
1 (257, 8.2%)	NHS	4 (1.6%)	0
	surname	2 (0.8%)	18 (7.0%)
	forename	0	14 (5.4%)
	DOB	0	210 (81.7%)
	postcode	0	9 (3.5%)
	total	**6 (2.3%)**	**251 (97.7%)**
0 (178, 5.7%)	nil	0	178 (100%)

Data improvements after NHS Digital list cleaning and provision of ONS date of death data

Demographic improvements

After case matching, NHS Digital returned demographic data (forename, surname, NHS number, home postcode) and ONS date of death if applicable. 1484 (47.6%) cases were not improved for any demographic data points (Table 3). These cases were those where complete demographic data were already collected by OHCAO (868 cases), matching failed (607 cases) or data could not be released by NHS Digital due to the patient either being lost to follow-up or registering a type 2 opt-out with NHS Digital (6 of the 7 cases). All demographic data were already collected by OHCAO for 1 case out of these 7 cases and therefore was included in the aforementioned 868 cases. Lastly, for 3 cases OHCAO collected NHS+surname+forename+postcode and therefore these effectively could not be improved by matching as NHS Digital were not asked to provide DOB.

Of the 2249 cases with missing data, 1636 (72.7%) cases had demographic improvements following linkage. A quarter (25.8%) were improved by 1 demographic data point and a further quarter (26.4%) by 2 data points. Of the 7 that were improved by 3 data points (Table 3), OHCAO provided NHS number for 4 cases, surname for 2, and DOB and postcode for 1.

Table 3: Number of data points added to OHCAO data after NHS Digital list cleaning

Number of demographic data points increased by list cleaning	Number of cases
0	1484 (47.6%)
1	804 (25.8%)
2	825 (26.4%)
3	7 (0.2%)

NHS Digital returned NHS numbers for 1518 (48.7%) cases in which it was not already collected by OHCAO (supplementary Table 1). OHCAO had already collected forename and surname in most cases (86.5% and 96.3% respectively) which were least improved following matching.

Survival data improvements following provision of ONS date of death data

30-day survival status (yes or no) using OHCAO data were confirmed for 1,682 (53.9%) cases (Table 4). 30-day survival was confirmed using OHCAO data if ambulance services provided a date of death or discharge date over 30 days after the OHCA incident date. Linking to ONS mortality data resulted in calculation of 30-day survival status (yes or no) for 2856 (91.5%) cases, a 37.6% improvement in 30-day survival status confirmation. The pre-linkage 30-day survival rate was calculated as 0.4% and post-linkage as 9.3%.

Table 4: Comparison of 30-day survival calculation pre- and post- data linkage

30-day survival			
Dataset 1: OHCAO data		Dataset 2: Linked OHCAO and ONS data	
Yes	12 (0.4%)	Yes	290 (9.3%)
No	1670 (53.5%)	No	2566 (82.2%)
Unknown	1438 (46.1%)	Unknown	264 (8.5%)
Total	3120 (100%)	Total	3120 (100%)

Accuracy of OHCAO date of death data

In this sample OHCAO reported a date of death for 1178 (37.8%) cases from ambulance services. In 7 (0.6%) cases death was not recorded with the ONS at the time of linkage. Of the 1942 (62.2%) cases where OHCAO could not confirm a date of death, 248 (12.8%) were recorded as alive at the time of linkage and 1137 (58.5%) had died according to ONS mortality data (Table 5).

Table 5: Comparison of date of death confirmed by OHCAO and ONS data

		ONS confirmed survival status			Total (n, % total 3120 cases)
		Dead	Alive	No	
OHCAO project date of death provided	yes	1114	7	57	1178 (37.8%)
	No	1137	248	557	1942 (62.2%)
Totals (n, % total 3120 cases)		2251 (72.1%)	255 (8.2%)	614* (19.7%)	3120 (100%)

*includes 7 matched cases where data not provided by NHS Digital (1 patient lost to follow-up (reason unknown), 6 patient registration of type 2 opt-out with NHS Digital).

DISCUSSION

This study demonstrates the feasibility of linking OHCAO data to NHS patient demographic data and ONS date of death data through NHS Digital. In this sample of 3120 OHCAO cases an 80.5% match rate was achieved and this enabled provision of registered death dates to calculate 30-day survival status. The results showed a 30-day survival rate of 9.3%, reducing unknown survival status from 46.1% to 8.5% (Table 4). Additionally, demographic data quality improved for 1636 (52.4%) cases, with NHS numbers being provided for 1518 (48.7%) cases and postcodes for 942 (30.2%) cases where this data were missing in the OHCAO database.

The variability of cardiac arrest survival across ambulance services in England has been previously highlighted.[6] Where data from ambulance services does not follow a standard procedure, data collection variability may have significant effects on data quality and comparability between services. Increasingly, core outcome sets for specific research areas are developed outlining minimum datasets for routine collection and create a level of standardisation to compare studies and allow formation of meta-analyses.[30] In the field of OHCA, the Utstein guidelines have been developed.[16,17] However, a study investigating the level of missing data within primary outcomes in 283 Cochrane Reviews of all areas of clinical practice found that over 50% of patient data were missing in 18% of reviews.[31] Furthermore, an analysis of 12 international OHCA registries collection of data using Utstein templates found that although all

registries collected core variables, there were differences in interpretation of the template and recorded 'unknown' for a mean of 4.8 variables and 'missing' for 1.9 variables.[32] Therefore minimum datasets are not sufficiently effective in reducing missing data.

The best data point provided by ambulance services to identify cases in the UK is the NHS number. It provides a unique identifier to resolve missing demographic data issues if no other demographic data are provided. 100% of OHCAO cases with an NHS number were matched to NHS PDS data, however, it was only available in a third (31.7%) of cases (Table 1). Logistical difficulties exist in ascertaining NHS number as it may not be available in the out-of-hospital setting. However, this study found that providing at least 3 to 4 demographic variables other than NHS number resulted in a match rate of up to 89.8%, depending on the combination and especially if forename and surname were provided. This also allowed provision of NHS number in 48.7% of cases where it was not collected by the OHCAO project (supplementary Table 1). If less than a threshold of 3 data points were provided, this study found a lower potential for matching (0- 44.2%, Table 2). Our findings support previous research showing that the ability to successfully link international OHCA databases to outcome data is dependent on the provision and completeness of patient identifiers. For example, the Danish Cardiac Arrest Registry was able to link to the Danish Civil Registration System to confirm 30-day survival for 100% of OHCA patients due to 100% provision of a unique Civil Registration Number.[20] Conversely, a study from the United States showed limited feasibility for linking OHCA patients to longitudinal outcomes when there was no unique patient identifiers available and there was variability in completeness of patient demographic data, resulting in a linkage rate of only 34.2%.[33]

NHS Digital list cleaning increased the number of OHCAO cases with a validated NHS number by 1518 (48.7%) to 2507 (80.4%) cases suggesting that data linkage is a feasible method for linking an OHCA dataset to the national mortality dataset. The current process for ambulance services in England to confirm survival to discharge from OHCA is challenging,[11,12] and utilising the NHS Digital list cleaning and patient status service to calculate 30-day survival from OHCA may be a viable alternative. The

results of this study also suggests the potential to utilise data linkage for further avenues of research relating to OHCA in the UK. Data linkage can be used to follow OHCA patients longitudinally, for example to investigate predictors of survival at 1 year, 5 years and beyond.[4,34] Furthermore, data linkage can be used to evaluate the complete patient care pathway by linking to existing routinely collected hospital data sources. For example, hospital interventions and hospital length of stay via Hospital Episode Statistics (HES), and intensive care interventions via the Intensive Care National Audit and Research Centre (ICNARC). NHS Digital also provided postcodes for a further 942 (30.2%) cases, which increases the potential to examine the influence of neighbourhood characteristics, such as population density and social deprivation, on OHCA incidence, whether an event is witnessed, and if they receive bystander CPR.[35,36]

The OHCAO project was able to collect a date of death for 1178 (37.8%) cases. Interestingly, a date of death was not recorded with the ONS for 7 of these cases indicating that the patients were still alive at the time of linkage. Such errors may lead to incorrect reporting of cardiac arrest survival as part of the NHS England AQIs. This is an important finding as this shows the importance of data linkage to correct database errors.

This study's strengths lie in its standardised procedures for OHCA case definition and data collection, with the data points collected based on established Utstein guidelines.[16] A further strength is that NHS Digital used both deterministic and probabilistic data linkage methods; they have different strengths and utilising both methods may enhance linkage performance.[37] Deterministic linkage methods have greater specificity but require exact matches between records, whilst probabilistic data linkage has greater sensitivity, working better with poorer quality data as it allows imperfect matches between records.[27] For example, the returned demographic data for the linked cases showed that 14 OHCAO cases with between 4 and 5 data points were linked despite having an erroneous NHS number. This allowed correction of the inaccurate NHS number in the OHCAO sample. Finally, successful data linkage enabled access to high quality national date of death data from ONS that is subject to rigorous data quality and validation processes.[28] Where no date of death was provided the patient was

categorised as alive. Absence of recorded death may mean registration of death has been delayed e.g. due to a coroner's inquest. However, NHS Digital did not commence data linkage until >12 months (March 2016) after the date (31st January 2015) where 30-day survival could be calculated for patients in the sample suffering an OHCA on 31st December 2014. ONS data for 2014-2015 shows that only 6.1% of deaths in England and Wales required a coroner's inquest[28] and the average time of an inquest was 24 weeks.[38,39] Whilst data from 2011 showed that overall 94% of deaths were registered within one month.[29] This suggests it is likely that ONS mortality data provided an accurate reflection of 30-day survival status in this study.

Limitations

This study had several limitations. Firstly, only 868 (27.8%) cases had all 5 OHCAO data points, whilst 178 (5.7%) cases had missing data for all OHCAO data points. Missing data is an issue in OHCA registries,[32] and improved data linkage in the OHCAO project is reliant on improved data capture of patient demographic data by ambulance services. Whilst NHS numbers were provided for only 989 (31.7%) OHCAO cases, one ambulance service provided NHS numbers for 100% of their cases. This suggests potential for the OHCAO project to work with ambulance services to increase provision of patient demographic data to improve data linkage. Secondly, following linkage 30-day survival status remained unknown for 264 (8.5%) cases. Data not missing completely at random can bias results.[40] For example, if those 264 patients survived to 30 days the overall 30-day survival rate would be 17.8% (584 cases) instead of 9.3% (290 cases). Finally, where the quality and completeness of data is variable data linkage errors can occur, which can bias reported outcomes.[41] Deterministic data linkage methods increase the likelihood of false negative matches (not matching to a correct match), whilst probabilistic data linkage increases the likelihood of false positive matches (matching to an incorrect match).[27] To quantify how data linkage errors may impact on study findings and outcomes a formal data linkage validation evaluation is required.[18] This was beyond the scope of this study but should be conducted if OHCAO establishes a data linkage programme.

CONCLUSIONS

This study shows the feasibility of linking data from the UK OHCAO project to NHS patient demographic and ONS date of death data using the NHS Digital list cleaning and patient status service. This enabled analysis of 30-day survival status which may be of use to the NHS in terms of resource planning and directing service provision. Missing NHS numbers are a significant obstacle to successful data linkage and this study found that if at least forename and surname is collected with one other demographic data point, there is a high chance of retrieving missing NHS numbers. Demographic data were improved for over half of cases and can be used as a means of creating a registry of OHCA patients to investigate post-resuscitation care and longitudinal outcomes.

Collaborators Dr Sukhdeep Dosanjh, Warwick Clinical Trials Unit, University of Warwick; Theresa Foster, East of England Ambulance Service NHS Trust; Frank Mersom, East of England Ambulance Service NHS Trust; Gurkamal Francis, London Ambulance Service NHS Trust; Michelle O'Rourke, North East Ambulance Service NHS Trust; Clare Bradley, North West Ambulance Service NHS Trust; Philip King, South Central Ambulance Service NHS Trust; Ed England, South Central Ambulance Service NHS Trust; Patricia Bucher, South East Coast Ambulance Service NHS Trust; Jessica Lynde, South Western Ambulance Service NHS Trust; Nancy Loughlin, South Western Ambulance Service NHS Trust; Jenny Lumley-Holmes, West Midlands Ambulance Service NHS Trust; Dr Julian Mark, Yorkshire Ambulance Service NHS Trust.

Contributions GDP designed the study. SJB, CJ, CH, ANS, KK, SB, RS, and IG contributed to data collection. SR analysed the data and wrote the initial draft of the paper. SR, GDP and SJB were involved in further drafting of the paper. All authors participated in interpreting the data, revising the paper for critically important intellectual content and gave final approval of the submitted version.

Acknowledgements National Ambulance Services Clinical Quality Group and the National Ambulance Research Steering Group.

Funding This work was supported by research grants from the British Heart Foundation and Resuscitation Council (UK).

Competing interests All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: SJB, TPB, CJ, CH, SJBM and GDP are employed by the University of Warwick, which receives grants from the British Heart Foundation and the Resuscitation Council (UK) for the conduct of the OHCAO project; no other relationships or activities that could appear to have influenced the submitted work.

Ethics approval Ethical approval for the OHCAO project was gained from the National Research Ethics Committee South Central, reference number 13/SC/0361. Confidential Advisory Group (CAG) approval was granted, reference number ECC 8-04(c)/2013, to collect identifiable patient information where it is not practical to obtain consent.

Data sharing statement Please refer to the OHCAO project website for information relating to data sharing requests: http://www2.warwick.ac.uk/fac/med/research/hscience/ctu/trials/ohcao/health/data/data_sharing/

REFERENCES

1. Resuscitation Council (UK), NHS England, British Heart Foundation. Consensus paper on out-of-hospital cardiac arrest in England. Resuscitation Council; 2014.

2. Ambulance Service Association. National cardiac arrest audit report. London: Ambulance Service Association; 2006.

3. Sasson C, Rogers MAM, Dahl J, Kellermann AL. Predictors of survival from out-of-hospital cardiac arrest. A systematic review and meta-analysis. Circ Cardiovasc Qual Outcomes. 2010;3:63-81.

4. Shuvy M, Morrison LJ, Koh M, Qiu F, Buick JE, Dorian P, et al. Long-term clinical outcomes and predictors for survivors of out-of-hospital cardiac arrest. Resuscitation. 2017;112:59-64.

5. Institute of Medicine Committee on the Treatment of Cardiac Arrest. Strategies to improve cardiac arrest survival. A time to act. National Academies Press; 2015.

6. Perkins GD, Cooke MW. Variability in cardiac arrest survival: The NHS ambulance service quality indicators. Emerg Med J. 2012;29(1):3-5.

7. Girotra S, van Diepen S, Nallamothu BK, Carrel M, Vellano K, Anderson ML, et al. Regional variation in out-of-hospital cardiac arrest survival in the United States. Circulation. 2016;doi:10.1161/circulationaha.115.018175

8. Okubo M, Kiyohara K, Iwami T, Callaway CW, Kitamura T. Nationwide and regional trends in survival from out-of-hospital cardiac arrest in Japan: A 10-year cohort study from 2005 to 2014. Resuscitation. 2017;115;120-28.

9. Stromsoe A, Svensson L, Axelsson AB, Claesson A, Goransson KE, Nordberg P, et al. Improved outcome in Sweden after out-of-hospital cardiac arrest and possible association with improvements in every link in the chain of survival. Eur Heart J. 2015;36;863-71.

10. Lilford R, Mohammed MA, Spiegelhalter D, Thomson R. Use and misuse of process and outcome data in managing performance of acute medical care: Avoiding institutional stigma. Lancet. 2004;363(9415):1147-54.

11. Fothergill R, Brace-McDonnell SJ, Perkins GD. Variation in epidemiology and outcomes from cardiac arrest. Resuscitation. 2014;85:1610-11.

12. Perkins GD, Lall R, Quinn T, Deakin CD, Cooke MW, Horton J, et al. Mechanical versus manual chest compression for out-of-hospital cardiac arrest (PARAMEDIC): A pragmatic, cluster randomised controlled trial. Lancet. 2015;385:947-55.

13. Berdowski J, Berg RA, Tijssen JGP, Koster RW. Global incidences of out-of-hospital cardiac arrest and survival rates: Systematic review of 67 prospective studies. Resuscitation. 2010;81:1479-87.

14. Ong ME, Shin SD, De Souza NN, Tanaka H, Nishiuchi T, Song KJ, et al. Outcomes for out-of-hospital cardiac arrests across 7 countries in Asia: The pan Asian Resuscitation Outcomes Study (PAROS). Resuscitation. 2015;96:100-08.

15. Grasner JT, Lefering R, Koster RW, Masterson S, Bottiger BW, Herlitz J, et al. EuReCa ONE 27 nations, ONE Europe, ONE registry: A prospective one month analysis of out-of-hospital cardiac arrest outcomes in 27 countries in Europe. Resuscitation. 2016;105:188-95.

16. Jacobs I, Nadkarni V, Bahr J, Berg RA, Billi JE, Bossaert L, et al. Cardiac arrest and cardiopulmonary resuscitation outcome reports: Update and simplification of the Utstein templates for resuscitation registries. Resuscitation. 2004;63(3):233-49.

17. Perkins GD, Jacobs IG, Nadkarni VM, Berg RA, Bhanji F, Biarent D, et al. Cardiac arrest and cardiopulmonary resuscitation outcome reports: Update of the Utstein resuscitation registry templates for out-of-hospital cardiac arrest. Resuscitation. 2015;96:328-40.

18. Moore, CL, Amin J, Gidding HF, Law MG. A new method for assessing how sensitivity and specificity of linkage studies affects estimation. Plos One. 2014;9(7):100-08. doi.org/10.1371/journal.pone.0103690

19. Elmer J, Rittenberger JC, Coppler PJ, Guyette FX, Doshi AA, Callaway CW. Long-term survival benefit from treatment at a specialty center after cardiac arrest. Resuscitation. 2016;108:48-53.

20. Hamilton A, Steinmetz J, Wissenberg M, Torp-Pedersen C, Lippert FK, Hove L, et al. Association between prehospital physician involvement and survival after out-of-hospital cardiac arrest: A Danish nationwide observational study. Resuscitation. 2016:108:95-101.

21. Perkins GD, Brace-McDonnell SJ, on behalf of the OHCAO Project Group. The UK out of hospital cardiac arrest outcome (OHCAO) project. BMJ Open. 2015;5:e008736.doi:10.1136/ bmjopen-2015-008736

22. Hawkes C, Booth S, Ji C, Brace-McDonnell SJ, Whittington A, Mapstone J, et al. Epidemiology and outcomes from out-of-hospital cardiac arrests in England. Resuscitation. 2017;110;133-40.

23. Office for National Statistics. Annual Mid-Year Population Estimates, 2014. http://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/2015-06-25. [Accessed 7 July 2017].

24. Joint Royal Colleges Ambulance Liaison Committee and Association of Ambulance Chief Executives. UK Ambulance Services Clinical Practice Guidelines 2016. Bridgwater: Class Professional Publishing. 2016.

25. Jaro MA. Probabilistic linkage of large public health data files. Stat Med. 1995;14(5-7):491-98.

26. Meray N, Reisma JB, Ravelli AC, Bonsel GJ. Probabilistic record linkage is a valid and transparent tool to combine databases without a patient identification number. J Clin Epidemiol. 2007;60:883-91.

27. Zhu Y, Matsuyama Y, Ohashi Y, Setoguchi S. When to conduct probabilistic linkage vs. deterministic linkage? A simulation study. J Biomed Inform. 2015;56:80-6.

28. Office for National Statistics. User guide to mortality statistics. July 2017. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/methodologies/userguidetomortalitystatisticsjuly2017

29. Office for National Statistics. Impact of registration delays on mortality statistics. 2011. http://webarchive.nationalarchives.gov.uk/20160106020016/http://www.ons.gov.uk/ons/guide-method/user-guidance/health-and-life-events/impact-of-registration-delays-on-mortality-statistics/index.html

30. Kirkham JJ, Gorst S, Altman DG, Blazeby JM, Clarke M, Devane D, et al. Core Outcome Set-STAndards for Reporting: The COS-STAR statement. PLoS Med. 2016;13(10):e1002148.doi:1 0.1371/journal.pmed.1002148

31. Kirkham JJ, Gargon E, Clarke M, Williamson PR. Can a core outcome set improve the quality of systematic reviews? — a survey of the Co-ordinating Editors of Cochrane review groups. Trials. 2013;14(21):doi:10.1186/1745-6215-14-21

32. Nishiyama C, Brown SP, May S, Iwami T, Koster RW, Beesems SG, et al. Apples to apples or apples to oranges? International variation in reporting of process and outcome of care for out-of-hospital cardiac arrest. Resuscitation. 2014;85:1599-609.

33. Mumma BE, Diercks DB, Danielsen B, Holmes JF. Probabilistic linkage of prehospital and outcomes data in out-of-hospital cardiac arrest. Prehosp Emerg Care. 2015;19(3):358-64.

34. Dumas F, Rea TD. Long-term prognosis following resuscitation from out-of-hospital cardiac arrest: Role of aetiology and presenting arrest rhythm. Resuscitation. 2012;83:1001-05.

35. Fosbol EL, Dupre ME, Strauss B, Swanson DR, Myers B, McNally BF, et al. Association of neighborhood characteristics with incidence of out-of-hospital cardiac arrest and rates of bystander-initiated CPR: Implications for community-based education intervention. Resuscitation. 2014;85:1512-17.

36. Moon S, Bobrow BJ, Vadeboncoeur TF, Kortuem W, Kisakye M, Sasson C, et al. Disparities in bystander CPR provision and survival from out-of-hospital cardiac arrest according to neighborhood ethnicity. Am J Emerg Med. 2014;32:1041-45.

37. Gomatam S, Carter R, Ariet M, Mitchell G. An empirical comparison of record linkage procedures. Stat Med. 2002;21:1485-96.

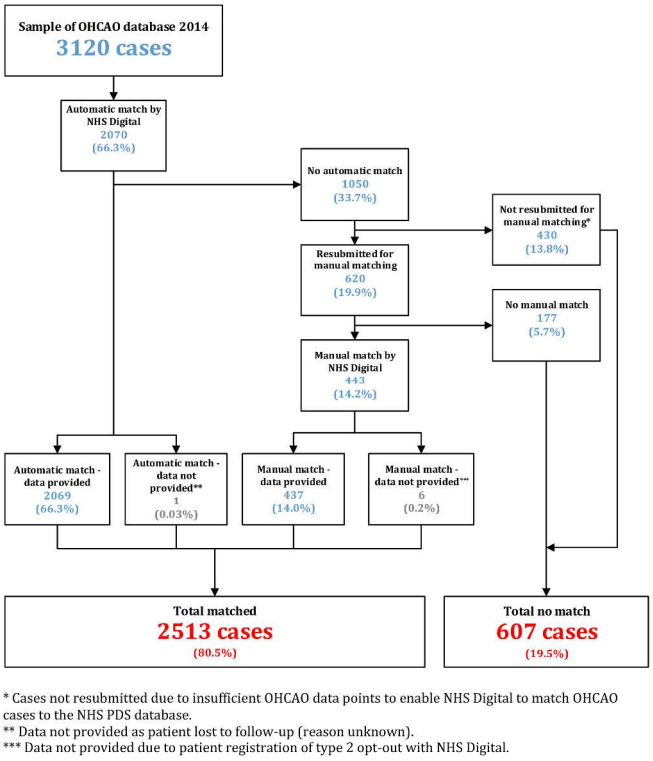
38. Ministry of Justice. Coroners Statistics 2014 England and Wales. May 2015. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/427720/coroners-statistics-2014.pdf

39. Ministry of Justice. Coroners Statistics 2015 England and Wales. May 2016. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/607728/coroners-statistics-2015.pdf

40. He Y. Missing data analysis using multiple imputation. Circ Cardiovasc Qual Outcomes. 2010;3:98-105.

41. Hagger-Johnson G, Harron K, Fleming T, Gilbert R, Goldstein H, Landy R, et al. Data linkage errors in hospital administrative data when applying a pseudonymisation algorithm to paediatric intensive care records. BMJ Open. 2015;5:e008118. doi:10.1136/bmjopen-2015-008118

Figure 1: Data matching process linking OHCAO data with NHS PDS data through the NHS Digital list cleaning and patient status service



* Cases not resubmitted due to insufficient OHCAO data points to enable NHS Digital to match OHCAO cases to the NHS PDS database.
** Data not provided as patient lost to follow-up (reason unknown).
*** Data not provided due to patient registration of type 2 opt-out with NHS Digital.

Figure 1: Data matching process linking OHCAO data with NHS PDS data through the NHS Digital list cleaning and patient status service

297x420mm (300 x 300 DPI)

SUPPLEMENTARY DATA

Table 1: Specific data points increased following linkage

Data point	Number increased by linkage (n, % of total cases)
NHS	1518 (48.7%)
Surname	7 (0.2%)
Forename	8 (0.3%)
Postcode	942 (30.2%)

The RECORD statement – checklist of items, extended from the STROBE statement, that should be reported in observational studies using routinely collected health data.

	Item No.	STROBE items	Location in manuscript where items are reported	RECORD items	Location in manuscript where items are reported
Title and abstract					
	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found	**(a) Page 1 & Page 2** **(b) Page 2**	RECORD 1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included. RECORD 1.2: If applicable, the geographic region and timeframe within which the study took place should be reported in the title or abstract. RECORD 1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract.	**(1.1) Page 1 & Page 2** **(1.2) Page 2** **(1.3) Page 1 & Page 2**
Introduction					
Background rationale	2	Explain the scientific background and rationale for the investigation being reported	**Pages 3 - 4**		
Objectives	3	State specific objectives, including any prespecified hypotheses	**Page 4**		
Methods					
Study Design	4	Present key elements of study design early in the paper	**Pages 5 - 6**		
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment,	**Pages 5 - 6**		

		exposure, follow-up, and data collection			
Participants	6	*(a) Cohort study* - Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study* - Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study* - Give the eligibility criteria, and the sources and methods of selection of participants *(b) Cohort study* - For matched studies, give matching criteria and number of exposed and unexposed *Case-control study* - For matched studies, give matching criteria and the number of controls per case	**(a) Page 5**	RECORD 6.1: The methods of study population selection (such as codes or algorithms used to identify subjects) should be listed in detail. If this is not possible, an explanation should be provided. RECORD 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided. RECORD 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical display to demonstrate the data linkage process, including the number of individuals with linked data at each stage.	**(6.1) Pages 5 - 8** **(6.2) N/A** **(6.3) Page 9**
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.	**Pages 6 - 7**	RECORD 7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be reported, an explanation should be provided.	**(7.1) Page 8**
Data sources/ measurement	8	For each variable of interest, give sources of data and details of methods of assessment (measurement).	**Pages 5 - 8**		

		Describe comparability of assessment methods if there is more than one group			
Bias	9	Describe any efforts to address potential sources of bias	**Page 6**		
Study size	10	Explain how the study size was arrived at	**Pages 5 - 6**		
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why	**Page 8**		
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) *Cohort study* - If applicable, explain how loss to follow-up was addressed *Case-control study* - If applicable, explain how matching of cases and controls was addressed *Cross-sectional study* - If applicable, describe analytical methods taking account of sampling strategy (e) Describe any sensitivity analyses	**(a) Page 8**		
Data access and cleaning methods		..		RECORD 12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population.	**(12.1) Pages 6 - 7** **(12.2) Page 5**

				RECORD 12.2: Authors should provide information on the data cleaning methods used in the study.	
Linkage		..		RECORD 12.3: State whether the study included person-level, institutional-level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.	**(12.3) Pages 6 - 7**
Results					
Participants	13	(a) Report the numbers of individuals at each stage of the study (*e.g.*, numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed) (b) Give reasons for non-participation at each stage. (c) Consider use of a flow diagram	**(a) Page 9** **(b) Page 9** **(c) Page 9**	RECORD 13.1: Describe in detail the selection of the persons included in the study (*i.e.*, study population selection) including filtering based on data quality, data availability and linkage. The selection of included persons can be described in the text and/or by means of the study flow diagram.	**(13.1) Page 9**
Descriptive data	14	(a) Give characteristics of study participants (*e.g.*, demographic, clinical, social) and information on exposures and potential confounders (b) Indicate the number of participants with missing data for each variable of interest (c) *Cohort study* - summarise follow-up time (*e.g.*, average and total amount)	**(a) N/A** **(b) Page 10** **(c) N/A**		
Outcome data	15	*Cohort study* - Report numbers of outcome events or summary measures over time	**Pages 13 - 14**		

		Case-control study - Report numbers in each exposure category, or summary measures of exposure *Cross-sectional study* - Report numbers of outcome events or summary measures			
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorized (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	**(a) Pages 9 - 14** **(b) N/A** **(c) N/A**		
Other analyses	17	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses	**Page 14**		
Discussion					
Key results	18	Summarise key results with reference to study objectives	**Page 14**		
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	**Page 17**	RECORD 19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they pertain to the study being reported.	**(19.1) Pages 16 - 17**

Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	**Pages 14 - 18**		
Generalisability	21	Discuss the generalisability (external validity) of the study results	**Page 15**		
Other Information					
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	**Page 19**		
Accessibility of protocol, raw data, and programming code		..		RECORD 22.1: Authors should provide information on how to access any supplemental information such as the study protocol, raw data, or programming code.	**(22.1) Page 20**

*Reference: Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM, the RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Medicine* 2015; in press.

BMJ Open

Data Quality and 30-day Survival for Out-of-Hospital Cardiac Arrest in the UK Out-of-Hospital Cardiac Arrest Registry: A Data Linkage Study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-017784.R2
Article Type:	Research
Date Submitted by the Author:	07-Sep-2017
Complete List of Authors:	Rajagopal, Sangeerthana; University of Warwick Warwick Medical School, Warwick Clinical Trials Unit Booth, Scott; University of Warwick Warwick Medical School, Warwick Clinical Trials Unit Brown, Terry; University of Warwick Warwick Medical School, Warwick Clinical Trials Unit Ji, Chen; University of Warwick Warwick Medical School, Warwick Clinical Trials Unit Hawkes, Claire; University of Warwick Warwick Medical School, Warwick Clinical Trials Unit Siriwardena, Aloysius; University of Lincoln, School of Health and Social Care Kirby, Kim; South Western Ambulance Service NHS Foundation Trust Black, Sarah; South Western Ambulance Service NHS Foundation Trust, Spaight, Robert; East Midlands Ambulance Service NHS Trust Gunson, Imogen; West Midlands Ambulance Service NHS Trust Brace-McDonnell, Samantha; University of Warwick, Warwick Clinical Trials Unit Perkins, Gavin; University of Warwick, Warwick Clinical Trials Unit
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Emergency medicine
Keywords:	Cardiac arrest, Emergency medical services, Medical record linkage, Out-of-hospital cardiac arrest, Resuscitation

Data Quality and 30-day Survival for Out-of-Hospital Cardiac Arrest in the UK Out-of-Hospital Cardiac Arrest Registry: A Data Linkage Study

Sangeerthana Rajagopal[a,b], Scott J Booth[a], Terry P Brown[a], Chen Ji[a], Claire Hawkes[a], A Niroshan Siriwardena[c], Kim Kirby[d], Sarah Black[d], Robert Spaight[e], Imogen Gunson[f], Samantha J Brace-McDonnell[a,b], Gavin D Perkins[a,b], on behalf of OHCAO collaborators

[a]Warwick Clinical Trials Unit, University of Warwick, Coventry CV4 7AL, UK

[b]Heart of England NHS Foundation Trust, Birmingham B9 5SS, UK

[c]University of Lincoln, Lincolnshire LN6 7TS, UK

[d]South Western Ambulance Service NHS Foundation Trust, Exeter EX2 7HY, UK

[e]East Midlands Ambulance Service NHS Trust, Nottingham NG8 6PY, UK

[f] West Midlands Ambulance Service NHS Foundation Trust, Brierley Hill, DY5 1LX, UK

Correspondence to Professor Gavin D Perkins; G.D.Perkins@warwick.ac.uk

Word count 3982

ABSTRACT

Objectives: The Out-of-Hospital Cardiac Arrest Outcomes (OHCAO) project aims to understand the epidemiology and outcomes of out-of-hospital cardiac arrests (OHCA) across the UK. This data linkage study is a sub-project of OHCAO. The aim was to establish the feasibility of linking OHCAO data to National Health Service (NHS) patient demographic data and Office for National Statistics (ONS) date of death data held on the NHS Personal Demographics Service (PDS) database to improve OHCAO demographic data quality and enable analysis of 30-day survival from OHCA.

Design and setting: Data were collected from 1st January 2014 to 31st December 2014 as part of a prospective, observational study of OHCA attended by ten English NHS Ambulance Services. 28,729 OHCA cases had resuscitation attempted by Emergency Medical Services and were included in the study. Data linkage was carried out using a data linkage service provided by NHS Digital, a national provider of health-related data. To assess data linkage feasibility a random sample of 3120 cases was selected. The sample was securely transferred to NHS Digital to be matched using OHCAO patient demographic data to return previously missing demographic data and provide ONS date of death data.

Results: A total of 2513 (80.5%) OHCAO cases were matched to patients in the NHS PDS database. Using the linkage process, missing demographic data were retrieved for 1636 (72.7%) out of 2249 OHCAO cases that had previously incomplete demographic data. Returned ONS date of death data allowed analysis of 30-day survival status. The results showed a 30-day survival rate of 9.3%, reducing unknown survival status from 46.1% to 8.5%.

Conclusions: In this sample, data linkage between the OHCAO registry and NHS PDS database was shown to be feasible, improving demographic data quality and allowing analysis of 30-day survival status.

ARTICLE SUMMARY

Strengths and limitations of this study

- Data points collected as part of the OHCAO project were based on established Utstein guidelines.
- The quality of demographic data collected by the OHCAO project was first improved through a list cleaning and patient status service provided by NHS Digital.
- Following list cleaning, exact data matches with Office for National Statistics (ONS) date of death data allowed calculation of 30-day survival status.
- Provision of NHS numbers from OHCAO and NHS digital provides potential for following long-term survival outcomes in OHCA patients through data linkage.
- Improved data linkage is reliant on improved data capture of patient demographic data by ambulance services.

INTRODUCTION

Every year in the United Kingdom (UK) there are around 60,000 out-of-hospital cardiac arrests (OHCA) attended by Emergency Medical Services (EMS) of which approximately 28,000 have resuscitation attempted.[1,2] This group suffers significant mortality and morbidity,[3,4] and improving outcomes from OHCA remains a worldwide research priority.[5]

Collecting high quality data is essential as this forms the basis of decisions that ultimately impact on changes in care and healthcare resource allocation. Since 2011, survival to hospital discharge rates for OHCA have been reported as part of the National Health Service (NHS) England Ambulance Quality Indictors (AQIs), with significant variation reported ranging from 2.2% to 12.0%.[6] Regional variation in survival rates have also been observed worldwide.[7-9] Lilford et al highlighted that an important source of variation in reporting outcomes can be traced to the quality of data that results are based on.[10] Collecting survival to discharge data in England is a challenging process for ambulance services as it involves tracking the patients survival status directly with

hospital emergency departments, which is time consuming and can be hindered by governance issues.[11,12]

Data collected in international OHCA registries enables comparisons of OHCA epidemiology and outcomes across different EMS systems.[13-15] The Utstein guidelines provide a structured template for collecting data on OHCA processes to support such comparisons.[16] To facilitate ease of reporting the updated Utstein guidelines recommend collecting either 30-day survival or survival to hospital discharge as a core outcome.[17] The research literature suggests most international registries are able to report either of these OHCA outcome measures.[13-15] A recent example is the EuReCa ONE study which aimed to benchmark OHCA incidence, process and outcomes across 27 European countries and reported a combined survival to discharge or 30-day survival rate which ranged between 1.1% and 30.8%.[15]

Data linkage methodology has increasingly been used in medical research to establish outcomes. It involves linking information together from different sources that belong to the same individual.[18] Data linkage has been utilised by regional and national OHCA databases to confirm survival status through linkage with mortality databases.[4,19,20] Data linkage can address missing data issues, providing a centralised, high quality database for research and service appraisal with the potential to allow longitudinal surveillance of OHCA patients.

The Out-of-Hospital Cardiac Arrest Outcomes (OHCAO) project is funded by the Resuscitation Council (UK), British Heart Foundation and managed by the University of Warwick. It is a prospective observational study investigating the epidemiology and outcomes of OHCA patients across the UK.[21,22] This paper presents a sub-project of the OHCAO project aiming to establish the feasibility of linking OHCAO registry data to NHS patient demographic data and Office for National Statistics (ONS) mortality data through the NHS Digital list cleaning and patient status service.

METHODS

Setting

The OHCAO project established a national UK OHCA registry to collect process and outcome data to facilitate OHCA research and quality improvement. Detailed information about the OHCAO project is available in the study protocol.[21] The ten English NHS ambulance services collecting data for the OHCAO project cover approximately 54 million people, equating to 99.7% of the England population and 83.9% of the UK population.[23] Data were collected from 1st January 2014 to 31st December 2014 on 28,729 patients suffering OHCA in whom resuscitation was attempted by statutory EMS (an incidence rate of 53.2 per 100,000 of the English population).[22] This figure was reached after excluding individuals who achieved ROSC before arrival of EMS (n=1711) and where resuscitation was not attempted as per national guidelines[24] due to the presence of a do not attempt resuscitation order (n=387), or signs incompatible with life or where resuscitation attempts would be futile (n=5403).

Aims & Objectives

The overall aim of this project was to investigate the feasibility of linking a sample of OHCAO 2014 data to NHS patient demographic data and ONS date of death data held on the NHS Personal Demographics Service (PDS) database, using the NHS Digital list cleaning and patient status service, to improve OHCAO demographic data quality and allow calculation of 30-day survival from OHCA. The objectives were to (1) assess the match rate of combinations of OHCAO patient demographic variables in the sample (NHS number, surname, forename, date of birth (DOB), and home postcode) for linking to the NHS PDS database through NHS Digital list cleaning; (2) assess improvements in the completeness of OHCAO patient demographic variables through NHS Digital list cleaning; (3) create a linked OHCAO and NHS PDS database allowing analysis of 30-day survival from OHCA.

OHCAO project data collection

Core and supplemental Utstein variables were collected encompassing demographic, system, process and outcome data.[16] Each ambulance service has their own methods for

OHCA case ascertainment e.g. electronic searches of patient report form databases for diagnostic codes indicating cardiac arrest. A trained member of the ambulance service clinical audit team entered eligible cases into a cardiac arrest database, followed by data cleaning and verification processes. Survival to hospital discharge data was collected directly from hospitals by the clinical audit team if data sharing protocols were in place. Each ambulance service uploaded their data via a secure server to the OHCAO registry which is stored at the University of Warwick.

OHCAO data sample

To assess feasibility whilst minimising costs associated with data linkage, the analysis presented here represents a 10.9% sample of the 2014 OHCAO data, comprising 3120 OHCA patients. To avoid selection bias the sample was selected using simple random sampling and stratified by ambulance service.

OHCAO data linkage to ONS mortality data

OHCAO to NHS PDS data linkage approval was received after submitting an application to the NHS Digital Data Access Request Service; additional approval was obtained from ONS for the release of mortality data. OHCAO submitted 3120 cases to NHS Digital, via the NHS Digital secure transfer system, detailing the following patient demographic variables of varying completeness: NHS number, surname, forename, DOB and home postcode.

NHS Digital is the national provider of data relating to health and social care in England. OHCAO used the NHS Digital list cleaning and patient status service. The list cleaning service was used to validate submitted demographic data to ensure accuracy and improve data linkage outcomes. Validation was achieved by NHS Digital matching submitted demographic variables to NHS patient demographic data held on the PDS database. The PDS database is a national electronic database containing NHS patient demographic information, including NHS number, name and address. For each matched case NHS Digital were asked to provide OHCAO with the following patient demographic information: NHS number, surname, forename, and home postcode. These

data were used to improve the percentage of missing data for these variables in the OHCAO sample.

NHS Digital utilised both automatic and manual matching techniques, using a combination of deterministic and probabilistic data linkage methods.[25,26] In deterministic data linkage it is decided *a priori* what combination of patient identifiers to match on (e.g. NHS number and DOB) and only complete agreement between records are considered a match. In probabilistic data linkage weights are assigned to different patient identifiers (based on their discriminatory power) to assess the probability that two records are a match.[27] Cases were initially submitted for automatic matching which used a decision tree algorithm to provide matches. A subset of cases that failed automatic matching were resubmitted for manual matching.

As part of the patient status service NHS Digital was also able to provide a date of death for deceased patients. The date of death data was held in the NHS PDS database and was sourced from ONS mortality data. OHCAO required information on deaths from 1st January 2014 until 31st January 2015. This was utilised to calculate 30-day survival. Where no date of death was provided the patient was categorised as alive.

ONS date of death data

ONS mortality data contains all deaths registered in England and Wales. When a person dies a formal medical certificate of death is produced, usually by a doctor, and which includes date of death. There is then a legal requirement for the death to be registered with the Registrar of Births, Deaths and Marriages through the local register office. The registration is typically performed by a close relative. The certification, and subsequent registration, of death may be delayed if the death is referred to a coroner for investigation (e.g. if cause of death is unknown). However, the majority of deaths in England and Wales are registered within 5 days of the death date.[28,29] ONS receives death data in electronic form directly from register offices. All data received is subject to both initial and routine data quality and validation processes and is collected in line with the Statistics and Registration Service Act 2007.[28]

Analysis

An analysis was conducted to assess how particular demographic data points enabled linkage with NHS PDS data and whether data linkage improved the completeness of patient demographic data. This was done descriptively with breakdowns of data linkage match rates for all combinations of the OHCAO demographic variables sent to NHS Digital for data linkage.

The combined linked dataset was analysed to investigate 30-day survival rates calculated by evaluating if patients were alive ≥30 calendar days from the EMS OHCA incident date. The analysis was carried out pre and post linkage, illustrating linkage effects. 30-day survival was calculated using OHCAO data where there was a date of death or date discharged >30 days after the OHCA incident date. Where there was an OHCAO date of death ≤30 days after the OHCA incident date or further ambulance service data indicating the patient was deceased on the day of the OHCA incident date (e.g. hospital code indicating patient deceased and not conveyed to hospital) the patient was categorised as not surviving to 30 days. All other cases were categorised as unknown for patient 30-day survival status. For the combined linked dataset, cases that were linked to ONS mortality data were categorised as 30-day survival where there was no date of death or where a date of death was provided that was >30 days after the OHCA incident date. Where there was an ONS date of death ≤30 days after the OHCA incident date the patient was categorised as not surviving to 30 days. Where there was a contradiction in patient survival status between OHCAO data and ONS mortality data then ONS mortality data superseded OHCAO data.

RESULTS

OHCAO data cleaning process

Of the 3120 cases transferred to NHS Digital, 2070 (66.3%) were automatically matched by the NHS Digital list cleaning algorithm while 1050 (33.7%) were not (Figure 1). 620 (19.9%) cases failing automatic matching were resubmitted for manual matching following which 437 (14.0%) were returned having been manually matched. 430 cases (13.8%) were not resubmitted for manual matching as there was little chance of a match due to missing data points (252 cases only had 1 data point out of surname, forename, DOB and home postcode and 178 cases did not have any data points). Overall, 2513 (80.5%) cases were matched of which 7 (0.2%) cases could not be released due to the patient being lost to follow-up (1 case, reason unknown) or the patient had registered a type 2 opt-out with NHS Digital, meaning that the patients' personal confidential data could not be released by NHS Digital for reasons other than their own direct care (6 cases). 607 (19.5%) cases could not be matched due to insufficient data for matching.

Data points required for matching through NHS Digital list cleaning

The percentage of each available demographic data point in the random sample of 3120 cases was similar to the percentage of each available demographic data point in all 28,729 cases for 2014 (Table 1). The data point determining the highest match rate was NHS number. 100% of cases with an NHS number were matched to the PDS database and therefore matched to ONS mortality data with 99% of cases with an NHS number being automatically matched. However, only 31.7% of OHCAO cases had an NHS number.

Table 1: Total cases with each demographic data point collected by OHCAO project

		NHS No.	Surname	Forename	DOB	Postcode
Total OHCAO cases with data point (% of total 28,729 cases)		9510 (33.1%)	24,814 (86.4%)	24,686 (85.9%)	24,956 (86.9%)	15,017 (52.3%)
Total OHCAO sample cases with data point (% of total 3120 cases)		989 (31.7%)	2699 (86.5%)	2693 (86.3%)	2700 (86.5%)	1626 (52.1%)
Match status (% of sample cases with specified data point)	Total matched	989 (100%)	2506 (92.8%)	2505 (93.0%)	2408 (89.2%)	1566 (96.3%)
	Auto match	979 (99.0%)	2070 (76.7%)	2070 (76.9%)	2070 (76.6%)	1364 (83.9%)
	Manual match	10 (1.0%)	436 (16.2%)	435 (16.2%)	338 (12.5%)	202 (12.4%)
	No match	0 (0.0%)	193 (7.2%)	188 (7.0%)	292 (10.8%)	60 (3.7%)

Approximately a quarter (27.8%) of the sample had all 5 data points allowing a match to NHS PDS data (Table 2). 53.2% had 3 to 4 data points of which 93.4% and 95.7% were matched, respectively. Of these, all with NHS numbers were matched, whilst a combination of data points surname+forename+DOB+postcode and surname+forename+DOB resulted in match rates of 81.6% and 89.8%. However, cases where only 1 or 2 data points were provided were less likely to be matched (2.3% and 44.2%, respectively). 178 (5.7%) cases had no OHCAO demographic data and could not be matched.

Table 2: Combinations of data points required for linkage to the NHS PDS database

Number of data points (n cases, % of total)	Combinations	N (total 3120 cases)	
		Matched, 2513 (n, % of total in data point category)	Unmatched, 607 (n, % of total in data point category)
5 (868, 27.8%)	NHS+surname+forename+DOB+postcode	868 (100%)	0
4 (815, 26.1%)	NHS+surname+DOB+postcode	0	0
	NHS+forename+DOB+postcode	0	0
	NHS+surname+forename+postcode	3 (0.4%)	0
	NHS+surname+forename+DOB	112 (13.7%)	0
	surname+forename+DOB+postcode	665 (81.6%)	35 (4.3%)
	Total	**780 (95.7%)**	**35 (4.3%)**
3 (846, 27.1%)	NHS+surname+forename	1 (0.1%)	0
	NHS+surname+DOB	0	0
	NHS+surname+postcode	0	0
	NHS+forename+DOB	0	0
	NHS+forename+postcode	0	0
	NHS+DOB+postcode	0	0
	surname+forename+DOB	760 (89.8%)	44 (5.2%)
	surname+forename+postcode	27 (3.2%)	11 (1.3%)
	surname+DOB+postcode	1 (0.1%)	1 (0.1%)
	forename+DOB+postcode	1 (0.1%)	0
	total	**790 (93.4%)**	**56 (6.6%)**
2 (156, 5.0%)	NHS+surname	0	0
	NHS+forename	1 (0.6%)	0
	NHS+DOB	0	0
	NHS+postcode	0	0
	surname+forename	67 (42.9%)	82 (52.6%)
	surname+DOB	0	0
	surname+postcode	0	2 (1.3%)
	forename+DOB	0	1 (0.6%)
	forename+postcode	0	1 (0.6%)
	DOB+postcode	1 (0.6%)	1 (0.6%)
	total	**69 (44.2%)**	**87 (55.8%)**
1 (257, 8.2%)	NHS	4 (1.6%)	0
	surname	2 (0.8%)	18 (7.0%)
	forename	0	14 (5.4%)
	DOB	0	210 (81.7%)
	postcode	0	9 (3.5%)
	total	**6 (2.3%)**	**251 (97.7%)**
0 (178, 5.7%)	nil	0	178 (100%)

Data improvements after NHS Digital list cleaning and provision of ONS date of death data

Demographic improvements

After case matching, NHS Digital returned demographic data (forename, surname, NHS number, home postcode) and ONS date of death if applicable. 1484 (47.6%) cases were not improved for any demographic data points (Table 3). These cases were those where complete demographic data were already collected by OHCAO (868 cases), matching failed (607 cases) or data could not be released by NHS Digital due to the patient either being lost to follow-up or registering a type 2 opt-out with NHS Digital (6 of the 7 cases). All demographic data were already collected by OHCAO for 1 case out of these 7 cases and therefore was included in the aforementioned 868 cases. Lastly, for 3 cases OHCAO collected NHS+surname+forename+postcode and therefore these effectively could not be improved by matching as NHS Digital were not asked to provide DOB.

Of the 2249 cases with missing data, 1636 (72.7%) cases had demographic improvements following linkage. A quarter (25.8%) were improved by 1 demographic data point and a further quarter (26.4%) by 2 data points. Of the 7 that were improved by 3 data points (Table 3), OHCAO provided NHS number for 4 cases, surname for 2, and DOB and postcode for 1.

Table 3: Number of data points added to OHCAO data after NHS Digital list cleaning

Number of demographic data points increased by list cleaning	Number of cases
0	1484 (47.6%)
1	804 (25.8%)
2	825 (26.4%)
3	7 (0.2%)

NHS Digital returned NHS numbers for 1518 (48.7%) cases in which it was not already collected by OHCAO (supplementary Table 1). OHCAO had already collected forename and surname in most cases (86.5% and 96.3% respectively) which were least improved following matching.

Survival data improvements following provision of ONS date of death data

30-day survival status (yes or no) using OHCAO data were confirmed for 1,682 (53.9%) cases (Table 4). 30-day survival was confirmed using OHCAO data if ambulance services provided a date of death or discharge date over 30 days after the OHCA incident date. Linking to ONS mortality data resulted in calculation of 30-day survival status (yes or no) for 2856 (91.5%) cases, a 37.6% improvement in 30-day survival status confirmation. The pre-linkage 30-day survival rate was calculated as 0.4% and post-linkage as 9.3%.

Table 4: Comparison of 30-day survival calculation pre- and post- data linkage

30-day survival			
Dataset 1: OHCAO data		Dataset 2: Linked OHCAO and ONS data	
Yes	12 (0.4%)	Yes	290 (9.3%)
No	1670 (53.5%)	No	2566 (82.2%)
Unknown	1438 (46.1%)	Unknown	264 (8.5%)
Total	3120 (100%)	Total	3120 (100%)

Accuracy of OHCAO date of death data

In this sample OHCAO reported a date of death for 1178 (37.8%) cases from ambulance services. In 7 (0.6%) cases death was not recorded with the ONS at the time of linkage. Of the 1942 (62.2%) cases where OHCAO could not confirm a date of death, 248 (12.8%) were recorded as alive at the time of linkage and 1137 (58.5%) had died according to ONS mortality data (Table 5).

Table 5: Comparison of date of death confirmed by OHCAO and ONS data

		ONS confirmed survival status			Total (n, % total 3120 cases)
		Dead	Alive	No	
OHCAO project date of death provided	yes	1114	7	57	1178 (37.8%)
	No	1137	248	557	1942 (62.2%)
Totals (n, % total 3120 cases)		2251 (72.1%)	255 (8.2%)	614* (19.7%)	3120 (100%)

*includes 7 matched cases where data not provided by NHS Digital (1 patient lost to follow-up (reason unknown), 6 patient registration of type 2 opt-out with NHS Digital).

DISCUSSION

This study demonstrates the feasibility of linking OHCAO data to NHS patient demographic data and ONS date of death data through NHS Digital. In this sample of 3120 OHCAO cases an 80.5% match rate was achieved and this enabled provision of registered death dates to calculate 30-day survival status. The results showed a 30-day survival rate of 9.3%, reducing unknown survival status from 46.1% to 8.5% (Table 4). Additionally, demographic data quality improved for 1636 (52.4%) cases, with NHS numbers being provided for 1518 (48.7%) cases and postcodes for 942 (30.2%) cases where this data were missing in the OHCAO database.

The variability of cardiac arrest survival across ambulance services in England has been previously highlighted.[6] Where data from ambulance services does not follow a standard procedure, data collection variability may have significant effects on data quality and comparability between services. Increasingly, core outcome sets for specific research areas are developed outlining minimum datasets for routine collection and create a level of standardisation to compare studies and allow formation of meta-analyses.[30] In the field of OHCA, the Utstein guidelines have been developed.[16,17] However, a study investigating the level of missing data within primary outcomes in 283 Cochrane Reviews of all areas of clinical practice found that over 50% of patient data were missing in 18% of reviews.[31] Furthermore, an analysis of 12 international OHCA registries collection of data using Utstein templates found that although all

registries collected core variables, there were differences in interpretation of the template and recorded 'unknown' for a mean of 4.8 variables and 'missing' for 1.9 variables.[32] Therefore minimum datasets are not sufficiently effective in reducing missing data.

The best data point provided by ambulance services to identify cases in the UK is the NHS number. It provides a unique identifier to resolve missing demographic data issues if no other demographic data are provided. 100% of OHCAO cases with an NHS number were matched to NHS PDS data, however, it was only available in a third (31.7%) of cases (Table 1). Logistical difficulties exist in ascertaining NHS number as it may not be available in the out-of-hospital setting. However, this study found that providing at least 3 to 4 demographic variables other than NHS number resulted in a match rate of up to 89.8%, depending on the combination and especially if forename and surname were provided. This also allowed provision of NHS number in 48.7% of cases where it was not collected by the OHCAO project (supplementary Table 1). If less than a threshold of 3 data points were provided, this study found a lower potential for matching (0- 44.2%, Table 2). Our findings support previous research showing that the ability to successfully link international OHCA databases to outcome data is dependent on the provision and completeness of patient identifiers. For example, the Danish Cardiac Arrest Registry was able to link to the Danish Civil Registration System to confirm 30-day survival for 100% of OHCA patients due to 100% provision of a unique Civil Registration Number.[20] Conversely, a study from the United States showed limited feasibility for linking OHCA patients to longitudinal outcomes when there was no unique patient identifiers available and there was variability in completeness of patient demographic data, resulting in a linkage rate of only 34.2%.[33]

NHS Digital list cleaning increased the number of OHCAO cases with a validated NHS number by 1518 (48.7%) to 2507 (80.4%) cases suggesting that data linkage is a feasible method for linking an OHCA dataset to the national mortality dataset. The current process for ambulance services in England to confirm survival to discharge from OHCA is challenging,[11,12] and utilising the NHS Digital list cleaning and patient status service to calculate 30-day survival from OHCA may be a viable alternative. The

results of this study also suggests the potential to utilise data linkage for further avenues of research relating to OHCA in the UK. Data linkage can be used to follow OHCA patients longitudinally, for example to investigate predictors of survival at 1 year, 5 years and beyond.[4,34] Furthermore, data linkage can be used to evaluate the complete patient care pathway by linking to existing routinely collected hospital data sources. For example, hospital interventions and hospital length of stay via Hospital Episode Statistics (HES), and intensive care interventions via the Intensive Care National Audit and Research Centre (ICNARC). NHS Digital also provided postcodes for a further 942 (30.2%) cases, which increases the potential to examine the influence of neighbourhood characteristics, such as population density and social deprivation, on OHCA incidence, whether an event is witnessed, and if they receive bystander CPR.[35,36]

The OHCAO project was able to collect a date of death for 1178 (37.8%) cases. Interestingly, a date of death was not recorded with the ONS for 7 of these cases indicating that the patients were still alive at the time of linkage. Such errors may lead to incorrect reporting of cardiac arrest survival as part of the NHS England AQIs. This is an important finding as this shows the importance of data linkage to correct database errors.

This study's strengths lie in its standardised procedures for OHCA case definition and data collection, with the data points collected based on established Utstein guidelines.[16] A further strength is that NHS Digital used both deterministic and probabilistic data linkage methods; they have different strengths and utilising both methods may enhance linkage performance.[37] Deterministic linkage methods have greater specificity but require exact matches between records, whilst probabilistic data linkage has greater sensitivity, working better with poorer quality data as it allows imperfect matches between records.[27] For example, the returned demographic data for the linked cases showed that 14 OHCAO cases with between 4 and 5 data points were linked despite having an erroneous NHS number. This allowed correction of the inaccurate NHS number in the OHCAO sample. Finally, successful data linkage enabled access to high quality national date of death data from ONS that is subject to rigorous data quality and validation processes.[28]

Limitations

This study had several limitations. Firstly, only 868 (27.8%) cases had all 5 OHCAO data points, whilst 178 (5.7%) cases had missing data for all OHCAO data points. Missing data is an issue in OHCA registries,[32] and improved data linkage in the OHCAO project is reliant on improved data capture of patient demographic data by ambulance services. Whilst NHS numbers were provided for only 989 (31.7%) OHCAO cases, one ambulance service provided NHS numbers for 100% of their cases. This suggests potential for the OHCAO project to work with ambulance services to increase provision of patient demographic data to improve data linkage. Secondly, following linkage 30-day survival status remained unknown for 264 (8.5%) cases. Data not missing completely at random can bias results.[38] For example, if those 264 patients survived to 30 days the overall 30-day survival rate would be 17.8% (584 cases) instead of 9.3% (290 cases). Thirdly, where no date of death was provided, cases were categorised as alive. However, absence of recorded death may mean registration of death has been delayed e.g. due to a coroner's inquest. Although it should also be noted that NHS Digital did not commence data linkage until >12 months (March 2016) after the date (31st January 2015) where 30-day survival could be calculated for patients in the sample suffering an OHCA on 31st December 2014. ONS data for 2014-2015 shows that only 6.1% of deaths in England and Wales required a coroner's inquest[28] and the average time of an inquest was 24 weeks.[39,40] Furthermore, ONS data from 2011 reports that overall 94% of deaths were registered within one month.[29] Finally, where the quality and completeness of data is variable data linkage errors can occur, and which can bias reported outcomes.[41] Deterministic data linkage methods increase the likelihood of false negative matches (not matching to a correct match), whilst probabilistic data linkage increases the likelihood of false positive matches (matching to an incorrect match).[27] To quantify how data linkage errors may impact on study findings and outcomes a formal data linkage validation evaluation is required.[18] This was beyond the scope of this study but should be conducted if OHCAO establishes a data linkage programme.

CONCLUSIONS

This study shows the feasibility of linking data from the UK OHCAO project to NHS patient demographic and ONS date of death data using the NHS Digital list cleaning and patient status service. This enabled analysis of 30-day survival status which may be of use to the NHS in terms of resource planning and directing service provision. Missing NHS numbers are a significant obstacle to successful data linkage and this study found that if at least forename and surname is collected with one other demographic data point, there is a high chance of retrieving missing NHS numbers. Demographic data were improved for over half of cases and can be used as a means of creating a registry of OHCA patients to investigate post-resuscitation care and longitudinal outcomes.

Collaborators Dr Sukhdeep Dosanjh, Warwick Clinical Trials Unit, University of Warwick; Theresa Foster, East of England Ambulance Service NHS Trust; Frank Mersom, East of England Ambulance Service NHS Trust; Gurkamal Francis, London Ambulance Service NHS Trust; Michelle O'Rourke, North East Ambulance Service NHS Trust; Clare Bradley, North West Ambulance Service NHS Trust; Philip King, South Central Ambulance Service NHS Trust; Ed England, South Central Ambulance Service NHS Trust; Patricia Bucher, South East Coast Ambulance Service NHS Trust; Jessica Lynde, South Western Ambulance Service NHS Trust; Nancy Loughlin, South Western Ambulance Service NHS Trust; Jenny Lumley-Holmes, West Midlands Ambulance Service NHS Trust; Dr Julian Mark, Yorkshire Ambulance Service NHS Trust.

Contributions GDP designed the study. SJB, CJ, CH, ANS, KK, SB, RS, and IG contributed to data collection. SR analysed the data and wrote the initial draft of the paper. SR, GDP and SJB were involved in further drafting of the paper. All authors participated in interpreting the data, revising the paper for critically important intellectual content and gave final approval of the submitted version.

Acknowledgements National Ambulance Services Clinical Quality Group and the National Ambulance Research Steering Group.

Funding This work was supported by research grants from the British Heart Foundation and Resuscitation Council (UK).

Competing interests All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: SJB, TPB, CJ, CH, SJBM and GDP are employed by the University of Warwick, which receives grants from the British Heart Foundation and the Resuscitation Council (UK) for the conduct of the OHCAO project; no other relationships or activities that could appear to have influenced the submitted work.

Ethics approval Ethical approval for the OHCAO project was gained from the National Research Ethics Committee South Central, reference number 13/SC/0361. Confidential Advisory Group (CAG) approval was granted, reference number ECC 8-04(c)/2013, to collect identifiable patient information where it is not practical to obtain consent.

Data sharing statement Please refer to the OHCAO project website for information relating to data sharing requests: http://www2.warwick.ac.uk/fac/med/research/hscience/ctu/trials/ohcao/health/data/data_sharing/

REFERENCES

1. Resuscitation Council (UK), NHS England, British Heart Foundation. Consensus paper on out-of-hospital cardiac arrest in England. Resuscitation Council; 2014.

2. Ambulance Service Association. National cardiac arrest audit report. London: Ambulance Service Association; 2006.

3. Sasson C, Rogers MAM, Dahl J, Kellermann AL. Predictors of survival from out-of-hospital cardiac arrest. A systematic review and meta-analysis. Circ Cardiovasc Qual Outcomes. 2010;3:63-81.

4. Shuvy M, Morrison LJ, Koh M, Qiu F, Buick JE, Dorian P, et al. Long-term clinical outcomes and predictors for survivors of out-of-hospital cardiac arrest. Resuscitation. 2017;112:59-64.

5. Institute of Medicine Committee on the Treatment of Cardiac Arrest. Strategies to improve cardiac arrest survival. A time to act. National Academies Press; 2015.

6. Perkins GD, Cooke MW. Variability in cardiac arrest survival: The NHS ambulance service quality indicators. Emerg Med J. 2012;29(1):3-5.

7. Girotra S, van Diepen S, Nallamothu BK, Carrel M, Vellano K, Anderson ML, et al. Regional variation in out-of-hospital cardiac arrest survival in the United States. Circulation. 2016;doi:10.1161/circulationaha.115.018175

8. Okubo M, Kiyohara K, Iwami T, Callaway CW, Kitamura T. Nationwide and regional trends in survival from out-of-hospital cardiac arrest in Japan: A 10-year cohort study from 2005 to 2014. Resuscitation. 2017;115;120-28.

9. Stromsoe A, Svensson L, Axelsson AB, Claesson A, Goransson KE, Nordberg P, et al. Improved outcome in Sweden after out-of-hospital cardiac arrest and possible association with improvements in every link in the chain of survival. Eur Heart J. 2015;36;863-71.

10. Lilford R, Mohammed MA, Spiegelhalter D, Thomson R. Use and misuse of process and outcome data in managing performance of acute medical care: Avoiding institutional stigma. Lancet. 2004;363(9415):1147-54.

11. Fothergill R, Brace-McDonnell SJ, Perkins GD. Variation in epidemiology and outcomes from cardiac arrest. Resuscitation. 2014;85:1610-11.

12. Perkins GD, Lall R, Quinn T, Deakin CD, Cooke MW, Horton J, et al. Mechanical versus manual chest compression for out-of-hospital cardiac arrest (PARAMEDIC): A pragmatic, cluster randomised controlled trial. Lancet. 2015;385:947-55.

13. Berdowski J, Berg RA, Tijssen JGP, Koster RW. Global incidences of out-of-hospital cardiac arrest and survival rates: Systematic review of 67 prospective studies. Resuscitation. 2010;81:1479-87.

14. Ong ME, Shin SD, De Souza NN, Tanaka H, Nishiuchi T, Song KJ, et al. Outcomes for out-of-hospital cardiac arrests across 7 countries in Asia: The pan Asian Resuscitation Outcomes Study (PAROS). Resuscitation. 2015;96:100-08.

15. Grasner JT, Lefering R, Koster RW, Masterson S, Bottiger BW, Herlitz J, et al. EuReCa ONE 27 nations, ONE Europe, ONE registry: A prospective one month analysis of out-of-hospital cardiac arrest outcomes in 27 countries in Europe. Resuscitation. 2016;105:188-95.

16. Jacobs I, Nadkarni V, Bahr J, Berg RA, Billi JE, Bossaert L, et al. Cardiac arrest and cardiopulmonary resuscitation outcome reports: Update and simplification of the Utstein templates for resuscitation registries. Resuscitation. 2004;63(3):233-49.

17. Perkins GD, Jacobs IG, Nadkarni VM, Berg RA, Bhanji F, Biarent D, et al. Cardiac arrest and cardiopulmonary resuscitation outcome reports: Update of the Utstein resuscitation registry templates for out-of-hospital cardiac arrest. Resuscitation. 2015;96:328-40.

18. Moore, CL, Amin J, Gidding HF, Law MG. A new method for assessing how sensitivity and specificity of linkage studies affects estimation. Plos One. 2014;9(7):100-08. doi.org/10.1371/journal.pone.0103690

19. Elmer J, Rittenberger JC, Coppler PJ, Guyette FX, Doshi AA, Callaway CW. Long-term survival benefit from treatment at a specialty center after cardiac arrest. Resuscitation. 2016;108:48-53.

20. Hamilton A, Steinmetz J, Wissenberg M, Torp-Pedersen C, Lippert FK, Hove L, et al. Association between prehospital physician involvement and survival after out-of-hospital cardiac arrest: A Danish nationwide observational study. Resuscitation. 2016:108:95-101.

21. Perkins GD, Brace-McDonnell SJ, on behalf of the OHCAO Project Group. The UK out of hospital cardiac arrest outcome (OHCAO) project. BMJ Open. 2015;5:e008736.doi:10.1136/ bmjopen-2015-008736

22. Hawkes C, Booth S, Ji C, Brace-McDonnell SJ, Whittington A, Mapstone J, et al. Epidemiology and outcomes from out-of-hospital cardiac arrests in England. Resuscitation. 2017;110;133-40.

23. Office for National Statistics. Annual Mid-Year Population Estimates, 2014. http://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/2015-06-25. [Accessed 7 July 2017].

24. Joint Royal Colleges Ambulance Liaison Committee and Association of Ambulance Chief Executives. UK Ambulance Services Clinical Practice Guidelines 2016. Bridgwater: Class Professional Publishing. 2016.

25. Jaro MA. Probabilistic linkage of large public health data files. Stat Med. 1995;14(5-7):491-98.

26. Meray N, Reisma JB, Ravelli AC, Bonsel GJ. Probabilistic record linkage is a valid and transparent tool to combine databases without a patient identification number. J Clin Epidemiol. 2007;60:883-91.

27. Zhu Y, Matsuyama Y, Ohashi Y, Setoguchi S. When to conduct probabilistic linkage vs. deterministic linkage? A simulation study. J Biomed Inform. 2015;56:80-6.

28. Office for National Statistics. User guide to mortality statistics. July 2017. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/methodologies/userguidetomortalitystatisticsjuly2017

29. Office for National Statistics. Impact of registration delays on mortality statistics. 2011. http://webarchive.nationalarchives.gov.uk/20160106020016/http://www.ons.gov.uk/ons/guide-method/user-guidance/health-and-life-events/impact-of-registration-delays-on-mortality-statistics/index.html

30. Kirkham JJ, Gorst S, Altman DG, Blazeby JM, Clarke M, Devane D, et al. Core Outcome Set-STAndards for Reporting: The COS-STAR statement. PLoS Med. 2016;13(10):e1002148.doi:1 0.1371/journal.pmed.1002148

31. Kirkham JJ, Gargon E, Clarke M, Williamson PR. Can a core outcome set improve the quality of systematic reviews? — a survey of the Co-ordinating Editors of Cochrane review groups. Trials. 2013;14(21):doi:10.1186/1745-6215-14-21

32. Nishiyama C, Brown SP, May S, Iwami T, Koster RW, Beesems SG, et al. Apples to apples or apples to oranges? International variation in reporting of process and outcome of care for out-of-hospital cardiac arrest. Resuscitation. 2014;85:1599-609.

33. Mumma BE, Diercks DB, Danielsen B, Holmes JF. Probabilistic linkage of prehospital and outcomes data in out-of-hospital cardiac arrest. Prehosp Emerg Care. 2015;19(3):358-64.

34. Dumas F, Rea TD. Long-term prognosis following resuscitation from out-of-hospital cardiac arrest: Role of aetiology and presenting arrest rhythm. Resuscitation. 2012;83:1001-05.

35. Fosbol EL, Dupre ME, Strauss B, Swanson DR, Myers B, McNally BF, et al. Association of neighborhood characteristics with incidence of out-of-hospital cardiac arrest and rates of bystander-initiated CPR: Implications for community-based education intervention. Resuscitation. 2014;85:1512-17.

36. Moon S, Bobrow BJ, Vadeboncoeur TF, Kortuem W, Kisakye M, Sasson C, et al. Disparities in bystander CPR provision and survival from out-of-hospital cardiac arrest according to neighborhood ethnicity. Am J Emerg Med. 2014;32:1041-45.

37. Gomatam S, Carter R, Ariet M, Mitchell G. An empirical comparison of record linkage procedures. Stat Med. 2002;21:1485-96.

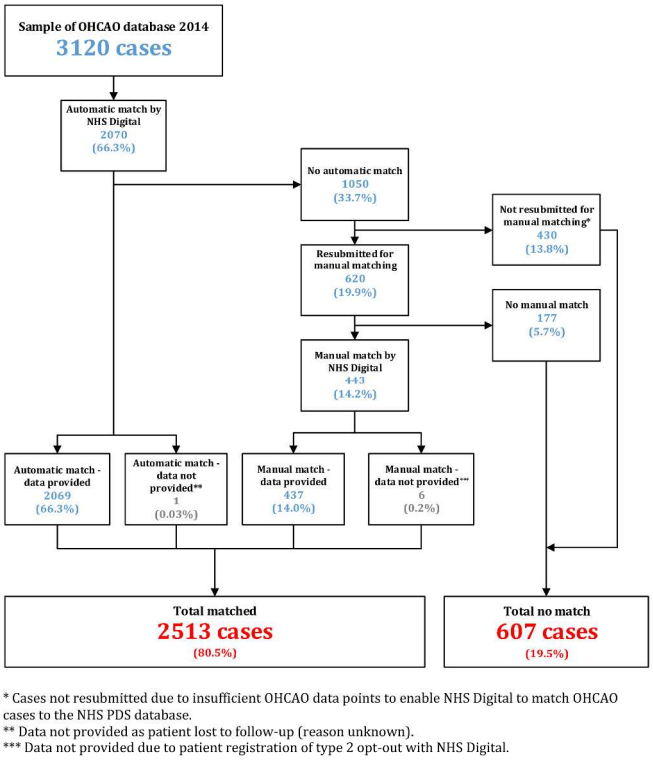
38. He Y. Missing data analysis using multiple imputation. Circ Cardiovasc Qual Outcomes. 2010;3:98-105.

39. Ministry of Justice. Coroners Statistics 2014 England and Wales. May 2015. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/427720/coroners-statistics-2014.pdf

40. Ministry of Justice. Coroners Statistics 2015 England and Wales. May 2016. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/607728/coroners-statistics-2015.pdf

41. Hagger-Johnson G, Harron K, Fleming T, Gilbert R, Goldstein H, Landy R, et al. Data linkage errors in hospital administrative data when applying a pseudonymisation algorithm to paediatric intensive care records. BMJ Open. 2015;5:e008118. doi:10.1136/bmjopen-2015-008118

Figure 1: Data matching process linking OHCAO data with NHS PDS data through the NHS Digital list cleaning and patient status service



* Cases not resubmitted due to insufficient OHCAO data points to enable NHS Digital to match OHCAO cases to the NHS PDS database.
** Data not provided as patient lost to follow-up (reason unknown).
*** Data not provided due to patient registration of type 2 opt-out with NHS Digital.

Figure 1: Data matching process linking OHCAO data with NHS PDS data through the NHS Digital list cleaning and patient status service

297x420mm (300 x 300 DPI)

SUPPLEMENTARY DATA

Table 1: Specific data points increased following linkage

Data point	Number increased by linkage (n, % of total cases)
NHS	1518 (48.7%)
Surname	7 (0.2%)
Forename	8 (0.3%)
Postcode	942 (30.2%)

The RECORD statement – checklist of items, extended from the STROBE statement, that should be reported in observational studies using routinely collected health data.

	Item No.	STROBE items	Location in manuscript where items are reported	RECORD items	Location in manuscript where items are reported
Title and abstract					
	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found	**(a) Page 1 & Page 2** **(b) Page 2**	RECORD 1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included. RECORD 1.2: If applicable, the geographic region and timeframe within which the study took place should be reported in the title or abstract. RECORD 1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract.	**(1.1) Page 1 & Page 2** **(1.2) Page 2** **(1.3) Page 1 & Page 2**
Introduction					
Background rationale	2	Explain the scientific background and rationale for the investigation being reported	**Pages 3 - 4**		
Objectives	3	State specific objectives, including any prespecified hypotheses	**Page 4**		
Methods					
Study Design	4	Present key elements of study design early in the paper	**Pages 5 - 6**		
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment,	**Pages 5 - 6**		

		exposure, follow-up, and data collection			
Participants	6	*(a) Cohort study* - Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study* - Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study* - Give the eligibility criteria, and the sources and methods of selection of participants *(b) Cohort study* - For matched studies, give matching criteria and number of exposed and unexposed *Case-control study* - For matched studies, give matching criteria and the number of controls per case	**(a) Page 5**	RECORD 6.1: The methods of study population selection (such as codes or algorithms used to identify subjects) should be listed in detail. If this is not possible, an explanation should be provided. RECORD 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided. RECORD 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical display to demonstrate the data linkage process, including the number of individuals with linked data at each stage.	**(6.1) Pages 5 - 8** **(6.2) N/A** **(6.3) Page 9**
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.	**Pages 6 - 7**	RECORD 7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be reported, an explanation should be provided.	**(7.1) Page 8**
Data sources/ measurement	8	For each variable of interest, give sources of data and details of methods of assessment (measurement).	**Pages 5 - 8**		

		Describe comparability of assessment methods if there is more than one group			
Bias	9	Describe any efforts to address potential sources of bias	**Page 6**		
Study size	10	Explain how the study size was arrived at	**Pages 5 - 6**		
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why	**Page 8**		
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) *Cohort study* - If applicable, explain how loss to follow-up was addressed *Case-control study* - If applicable, explain how matching of cases and controls was addressed *Cross-sectional study* - If applicable, describe analytical methods taking account of sampling strategy (e) Describe any sensitivity analyses	**(a) Page 8**		
Data access and cleaning methods		..		RECORD 12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population.	**(12.1) Pages 6 - 7** **(12.2) Page 5**

				RECORD 12.2: Authors should provide information on the data cleaning methods used in the study.	
Linkage		..		RECORD 12.3: State whether the study included person-level, institutional-level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.	**(12.3) Pages 6 - 7**
Results					
Participants	13	(a) Report the numbers of individuals at each stage of the study (*e.g.*, numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed) (b) Give reasons for non-participation at each stage. (c) Consider use of a flow diagram	**(a) Page 9** **(b) Page 9** **(c) Page 9**	RECORD 13.1: Describe in detail the selection of the persons included in the study (*i.e.*, study population selection) including filtering based on data quality, data availability and linkage. The selection of included persons can be described in the text and/or by means of the study flow diagram.	**(13.1) Page 9**
Descriptive data	14	(a) Give characteristics of study participants (*e.g.*, demographic, clinical, social) and information on exposures and potential confounders (b) Indicate the number of participants with missing data for each variable of interest (c) *Cohort study* - summarise follow-up time (*e.g.*, average and total amount)	**(a) N/A** **(b) Page 10** **(c) N/A**		
Outcome data	15	*Cohort study* - Report numbers of outcome events or summary measures over time	**Pages 13 - 14**		

		Case-control study - Report numbers in each exposure category, or summary measures of exposure *Cross-sectional study* - Report numbers of outcome events or summary measures			
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorized (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	**(a) Pages 9 - 14** **(b) N/A** **(c) N/A**		
Other analyses	17	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses	**Page 14**		
Discussion					
Key results	18	Summarise key results with reference to study objectives	**Page 14**		
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	**Page 17**	RECORD 19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they pertain to the study being reported.	**(19.1) Pages 16 - 17**

Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	**Pages 14 - 18**		
Generalisability	21	Discuss the generalisability (external validity) of the study results	**Page 15**		
Other Information					
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	**Page 19**		
Accessibility of protocol, raw data, and programming code		..		RECORD 22.1: Authors should provide information on how to access any supplemental information such as the study protocol, raw data, or programming code.	**(22.1) Page 20**

*Reference: Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM, the RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Medicine* 2015; in press.