Original citation:

Permanent WRAP URL:
http://wrap.warwick.ac.uk/95172

Copyright and reuse:
The Warwick Research Archive Portal (WRAP) makes this work of researchers of the University of Warwick available open access under the following conditions.

This article is made available under the Creative Commons Attribution-NonCommercial 4.0 (CC BY-NC 4.0) license and may be reused according to the conditions of the license. For more details see: http://creativecommons.org/licenses/by-nc/4.0/

A note on versions:
The version presented in WRAP is the published version, or, version of record, and may be cited as it appears here.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk
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

Sangeerthana Rajagopal,1,2 Scott J Booth,1 Terry P Brown,1 Chen Ji,1 Claire Hawkes,1 A Niroshan Siriwardena,3 Kim Kirby,4 Sarah Black,4 Robert Spaight,5 Imogen Gunson,6 Samantha J Brace-McDonnell,1,2 Gavin D Perkins,1,2 on behalf of OHCAO collaborators

Objectives The Out-of-Hospital Cardiac Arrest Outcomes (OHCAO) project aims to understand the epidemiology and outcomes of out-of-hospital cardiac arrest (OHCA) across the UK. This data linkage study is a subproject 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 Services (PDS) database to improve OHCAO demographic data quality and enable analysis of 30-day survival from OHCA.

Design and setting Data were collected from 1 January 2014 to 31 December 2014 as part of a prospective, observational study of OHCA attended by 10 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.

INTRODUCTION

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

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 Indicators (AQIs), with significant variation reported ranging from 2.2% to 12.0%. Regional variation in survival rates has also been observed worldwide. Lifford 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. Collecting survival to discharge data in England is a challenging process for...
ambulance services as it involves tracking the patient’s 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 enable 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 incidence of 729 patients suffering OHCA in whom resuscitation was attempted by statutory EMS (an incidence of 20.9 per 100,000 of the England population).23

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 It is a subproject of the OHCAO registry data collection to 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.

**Aims and 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 rates of combinations of OHCA 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 and (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 its own methods for OHCA case ascertainment, for example, 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 its data via a secure server to the OHCAO registry which is stored at the University of Warwick.

**OHCAO data sample**
To assess feasibility while 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 data with 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 was 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 used both automatic and manual matching techniques, using a combination of deterministic and probabilistic data linkage methods. In deterministic data linkage, it is decided a priori what combination of patient identifiers to match on (eg, NHS number and DOB) and only complete agreement between records is 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. 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 1 January 2014 until 31 January 2015. This was used 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 contain 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 (eg, if cause of death is unknown). However, the majority of deaths in England and Wales are registered within 5 days of the death date. ONS receives death data in electronic form directly from register offices. All data received are subject to both initial and routine data quality and validation processes and are collected in line with the Statistics and Registration Service Act 2007.

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-linkage and post-linkage, illustrating linkage effects. Thirty-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 (eg, 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’s 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). In total, 620 (19.9%) cases failing automatic matching were resubmitted for manual matching following which 437 (14.0%) were returned having been manually matched. Also, 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 (one case, reason unknown) or the patient had registered a type 2 opt-out confidential data could not be released by NHS Digital for reasons other than their own direct care (six cases). In total, 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
Table 1  Total cases with each demographic data point collected by Out-of-Hospital Cardiac Arrest Outcomes (OHCAO) project.

<table>
<thead>
<tr>
<th>National HealthService number</th>
<th>Surname</th>
<th>Forename</th>
<th>Date of birth</th>
<th>Postcode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total OHCAO cases with data point (% of total 28729 cases)</td>
<td>9510 (33.1%)</td>
<td>24814 (86.4%)</td>
<td>24686 (85.9%)</td>
<td>24956 (86.9%)</td>
</tr>
<tr>
<td>Total OHCAO sample cases with data point (% of total 3120 cases)</td>
<td>989 (31.7%)</td>
<td>2699 (86.5%)</td>
<td>2693 (86.3%)</td>
<td>2700 (86.5%)</td>
</tr>
<tr>
<td>Match status (% of sample cases with specified data point)</td>
<td>Total matched</td>
<td>989 (100%)</td>
<td>2506 (92.8%)</td>
<td>2505 (93.0%)</td>
</tr>
<tr>
<td>Auto match</td>
<td>979 (99.0%)</td>
<td>2070 (76.7%)</td>
<td>2070 (76.9%)</td>
<td>2070 (76.6%)</td>
</tr>
<tr>
<td>Manual match</td>
<td>10 (1.0%)</td>
<td>436 (16.2%)</td>
<td>435 (16.2%)</td>
<td>338 (12.5%)</td>
</tr>
<tr>
<td>No match</td>
<td>0 (0.0%)</td>
<td>193 (7.2%)</td>
<td>188 (7.0%)</td>
<td>292 (10.8%)</td>
</tr>
</tbody>
</table>

Figure 1  Data matching process linking Out-of-Hospital Cardiac Arrest Outcomes (OHCAO) data with National Health Service (NHS) Personal Demographics Service (PDS) data through the NHS Digital list cleaning and patient status service.
Table 2  Combinations of data points required for linkage to the National Health Service (NHS) Personal Demographics Service database

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

DOB, date of birth.

53.2% had 3–4 data points of which 93.4% and 95.7% were matched, respectively. Of these, all with NHS numbers were matched, while 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). In total, 178 (5.7%) cases had no OHCAO demographic data and could not be matched.

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. In total, 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 (six of the seven cases). All demographic data were already collected by OHCAO for one case out of these 7 cases and therefore were included in the aforementioned 868 cases. Lastly, for three cases OHCAO collected NHS+surname+forename+postcode and therefore these effectively could not be improved by matching as NHS Digital was 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 seven that were improved by 3 data points (table 3), OHCAO provided NHS number for four cases, surname and 96.3% respectively. In total, 178 (5.7%) cases had no OHCAO demographic data and could not be matched.

Survival data improvements following provision of ONS date of death data

Thirty-day survival status (yes or no) using OHCAO data was confirmed for 1682 (53.9%) cases (table 4). Thirty-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 2586 (91.5%) cases, a 57.6% improvement in 30-day survival status confirmation.

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 these data were missing in the OHCAO database.

**Table 3** Number of data points added to Out-of-Hospital Cardiac Arrest Outcomes data after NHS Digital list cleaning

<table>
<thead>
<tr>
<th>Number of demographic data points increased by list cleaning</th>
<th>Cases (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1484 (47.6%)</td>
</tr>
<tr>
<td>1</td>
<td>804 (25.8%)</td>
</tr>
<tr>
<td>2</td>
<td>825 (26.4%)</td>
</tr>
<tr>
<td>3</td>
<td>7 (0.2%)</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>30-day survival</th>
<th>Dataset: OHCAO data</th>
<th>Dataset 2: linked OHCAO and ONS data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Total (n, % total cases)</td>
<td>Yes</td>
<td>1114</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>1137</td>
</tr>
</tbody>
</table>

**Table 5** Comparison of date of death confirmed by Out-of-Hospital Cardiac Arrest Outcomes (OHCAO) and Office for National Statistics (ONS) data

<table>
<thead>
<tr>
<th>ONS confirmed survival status</th>
<th>Total (n, % total 3120 cases)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dead</td>
<td>Alive</td>
</tr>
<tr>
<td>OHCAO project date of death provided</td>
<td>Yes</td>
</tr>
<tr>
<td>No</td>
<td>1137</td>
</tr>
</tbody>
</table>

Total (n, % total 3120 cases) | 2251 | 255 | 614* | 3120 (100%)

*Includes seven matched cases where data not provided by NHS Digital (one patient lost to follow-up (reason unknown), six patient registration of type 2 opt-out with NHS Digital).
The variability of cardiac arrest survival across ambulance services in England has been previously highlighted. Where data from ambulance services do 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. In the field of OHCA, the Utstein guidelines have been developed. However, a study investigating the level of missing data within primary outcomes in 283 Cochrane Reviews of all areas of clinical practice found that >50% of patient data were missing in 18% of reviews. 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. 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. One hundred per cent 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 an NHS number as it may not be available in the out-of-hospital setting. However, this study found that providing at least 3–4 demographic variables other than an 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 an NHS number in 48.7% of cases where it was not collected by the OHCAO project (see online supplementary table 1). If less than a threshold of 3 data points were provided, this study found a lower potential for matching (0%–41.2%, 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. Conversely, a study from the USA 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%.

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, and using 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 suggest the potential to use 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. 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 and intensive care interventions via the Intensive Care National Audit and Research Centre. 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 cardiopulmonary resuscitation (CPR). 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 seven 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. A further strength is that NHS Digital used both deterministic and probabilistic data linkage methods; they have different strengths and using both methods may enhance linkage performance. Deterministic linkage methods have greater specificity but require exact matches between records, while probabilistic data linkage has greater sensitivity, working better with poorer quality data as it allows imperfect matches between records. 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.

**Limitations**

This study had several limitations. First, only 868 (27.8%) cases had all 5 OHCAO data points, while 178 (5.7%) cases had missing data for all OHCAO data points. Missing data is an issue in OHCA registries, and improved data linkage in the OHCAO project is reliant on improved data capture of patient demographic data by ambulance services. While 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. Second, following linkage 30-day survival status remained unknown for 264 (8.5%) cases. Data not missing completely at random can bias results. 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). Third, 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, for example, 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 (31 January 2015) where 30-day survival could be calculated for patients in the sample suffering an OHCA on 31 December 2014. ONS data for 2014–2015 show that only 6.1% of deaths in England and Wales required a coroner’s inquest and the average time of an inquest was 24 weeks. Furthermore, ONS data from 2011 report that overall 94% of deaths were registered within 1 month. Finally, where the quality and completeness of data is variable data linkage errors can occur, which can bias reported outcomes. Deterministic data linkage methods increase the likelihood of false negative matches (not matching to a correct match), while probabilistic data linkage increases the likelihood of false positive matches (matching to an incorrect match). To quantify how data linkage errors may impact on study findings and outcomes, a formal data linkage validation evaluation is required. 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 are 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 postresuscitation 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; Phillip 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.

Contributors 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.

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

Competing interests SJB, TPB, CJ, CH, SJB M 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.

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.

Provenance and peer review Not commissioned; externally peer reviewed.

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/ntscc/ctu/trials/ohcao/health/data/data_sharing/

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2017. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

REFERENCES


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, Scott J Booth, Terry P Brown, Chen Ji, Claire Hawkes, A Niroshan Siniwardena, Kim Kirby, Sarah Black, Robert Spaight, Imogen Gunson, Samantha J Brace-McDonnell and Gavin D Perkins

BMJ Open 2017 7:
doi: 10.1136/bmjopen-2017-017784

Updated information and services can be found at:
http://bmjopen.bmj.com/content/7/11/e017784

These include:

References
This article cites 32 articles, 5 of which you can access for free at:
http://bmjopen.bmj.com/content/7/11/e017784#BIBL

Open Access
This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/

Email alerting service
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Topic Collections
Articles on similar topics can be found in the following collections

Epidemiology (2224)

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/