Developing new ways of measuring the quality and impact of ambulance service care: the PhOEBE mixed-methods research programme

Janette Turner,[1] A Niroshan Siriwardena,[2,3]*
Joanne Coster,[1] Richard Jacques,[1] Andy Irving,[1]
Annabel Crum,[1] Helen Bell Gorrod,[1] Jon Nicholl,[1]
Viet-Hai Phung,[2] Fiona Togher,[2] Richard Wilson,[1]
Alicia O'Cathain,[1] Andrew Booth,[1] Daniel Bradbury,[1]
Steve Goodacre,[1] Anne Spaight,[3] Jane Shewan,[4]
Richard Pilbery,[4] Daniel Fall,[5] Maggie Marsh,[5]
Andrea Broadway-Parkinson,[5] Ronan Lyons,[6]
Helen Snooks[6] and Mike Campbell[1]

[1]School of Health and Related Research, University of Sheffield, Sheffield, UK
[2]Community and Health Research Unit (CaHRU), University of Lincoln, Lincoln, UK
[3]East Midlands Ambulance Service NHS Trust, Nottingham, UK
[4]Yorkshire Ambulance Service NHS Trust, Wakefield, UK
[5]Patient and public involvement, Sheffield, UK
[6]College of Medicine, Swansea University, Swansea, UK

*Corresponding author nsiriwardena@lincoln.ac.uk

Declared competing interests of authors: Anne Spaight reports grants from East Midlands Ambulance Service NHS Trust during the conduct of the study. Steve Goodacre is a member of the Health Technology Assessment (HTA) Clinical Trials Board, HTA Funding Boards Policy Group and HTA IP Methods Group. Helen Snooks is a member of National Institute for Health Research HTA and Efficacy and Mechanism Evaluation Editorial Board.

Published April 2019
DOI: 10.3310/pgfar07030

Scientific summary

The PhOEBE mixed-methods research programme

Programme Grants for Applied Research 2019; Vol. 7: No. 3
DOI: 10.3310/pgfar07030

NIHR Journals Library www.journalslibrary.nihr.ac.uk

Scientific summary

Background

Measurement of the performance of and quality of care provided by ambulance services has been dominated by response times and indicators for a small number of emergency conditions, such as cardiac arrest, heart attack and stroke. These measurements do not reflect the wide range of conditions that present to ambulance services and the variety of response options that are now available. This is largely because of the lack of previous research on what matters to patients and the absence of data on outcomes after patients are attended by ambulance services. The Prehospital Outcomes for Evidence Based Evaluation (PhOEBE) programme sought to address this.

Aim and objectives

The PhOEBE programme aimed to explore the development of new ways of measuring the impact of prehospital ambulance service care. These new ways of measuring could provide better information about the effectiveness and quality of the different types of care delivered to the range of patients attended by ambulance services to support quality improvement, audit and evaluation of future service changes.

The objectives of the programme were to:

1. review, assess and synthesise policy and research literature on prehospital ambulance outcome measures, conduct a qualitative study of ambulance service users' experiences, and use consensus methods to identify measures relevant to NHS patients, providers, practitioners and commissioners that had potential for further development
2. create an information data set that could be used for measuring ambulance service care by linking routinely collected prehospital ambulance data, hospital episode data and mortality data using processes that were acceptable to patients and complied with information legislation to provide information on outcomes
3. use the linked data to develop case-mix adjustment models that measure mortality and non-mortality outcomes for assessing ambulance service quality and detecting change over time
4. explore the practical use of the linked data set and risk adjustment models to measure the effectiveness and quality of ambulance service care and assess how the linked data set and models could be best used to support quality improvement strategies.

Methods

We used a series of linked studies using multiple methods to investigate the aim and objectives. The study was conducted in the East Midlands and Yorkshire regions of England and involved ambulance services, patients who had used ambulance services, patient and public involvement (PPI) groups, emergency care clinical academics, commissioners and policy-makers between 2011 and 2017.

To address objective 1 we undertook two systematic reviews of relevant literature: first, a review of published policy documents that were relevant to actual and aspirational quality and outcomes of prehospital ambulance services and care; and, second, a review systematically synthesising performance and outcome measures reported in the published prehospital research literature. We also conducted a qualitative interview study of a purposive sample of recent ambulance service users to investigate aspects of emergency ambulance service care valued by patients.

Using the outputs from the systematic reviews and interviews, we used a three-stage multimethod consensus study comprising multistakeholder consensus event using a modified nominal group technique, modified Delphi study, and patient and public consensus workshop to prioritise and rate potential measures. An expert panel used the results of the consensus work and further assessed 20 potential measures and identified a small number of candidate indicators for further development.

To address objective 2 we created a data set linking ambulance, hospital and mortality data using ambulance call-and-dispatch and electronic patient report forms, Hospital Episode Statistics (HES) accident and emergency and inpatient data, and Office for National Statistics mortality data.

To address objectives 3 and 4 we conducted a statistical analysis of the linked data set to explore use of case-mix-adjusted methods for the candidate indicators. Case-mix adjustment was used to explore factors other than ambulance service care that may be influencing processes or outcomes. These included patient characteristics, for example age and the type of condition, and external factors, such as the time of day an incident happens, the location and which hospital provided subsequent care. For each indicator we built a statistical model that allowed us to assess which factors were important in predicting the process or outcome and determine if case-mix adjustment would improve the robustness of the indicator as a measure of performance and quality.

We also used the linked data and the results of two indicators to conduct an economic analysis assessing the costs of different types of ambulance response and the economic consequences of making incorrect decisions about whether to take people to hospital or leave them at home.

Results

Seventy-two candidate measures were generated from systematic reviews in four categories: (1) ambulance service operations ($n = 14$), (2) clinical management of patients ($n = 20$), (3) impact of care on patients ($n = 9$) and (4) time measures ($n = 29$). Forty-two potential measures (excluding time measures) were presented at the consensus conference and nine measures or principles were highly prioritised by over three-quarters of consensus event participants, including measures relating to pain, patient experience, accuracy of dispatch decisions and patient safety.

Twenty experts participated in two Delphi rounds to refine and prioritise measures; 20 measures scored $\geq 8/9$ points, indicating good consensus, including the proportion of calls correctly prioritised, time to definitive care and pain measures. Eighteen patient and public representatives attending a consensus workshop identified six measures as important including time to definitive care, response time, reduction in pain score, calls correctly prioritised to appropriate levels of response and survival to hospital discharge for treatable emergency conditions.

A total of 187,287 cases were retrieved and a linked data set of 154,927 cases was created for the 6-month period of January to June 2013 from one ambulance service. For patients attended by the ambulance service, 85% of records were successfully linked. For cases with no ambulance response (telephone advice only), the linkage rate was poor at 24%.

Using the outputs from the consensus work the expert panel identified six candidate indicators. These were (1) mean change in pain score; (2) proportion of serious emergency conditions correctly identified at the time of the 999 call; (3) response time; (4) proportion of decisions to leave a patient at scene ('hear and treat' and 'see and treat') who subsequently re-contact a service or die within 3 days (potentially inappropriate decision); (5) proportion of patients transported to ED by 999 emergency ambulance, but who were discharged to usual place of residence or care of their general practitioner without treatment or investigation(s) that needed hospital facilities (unnecessary transports); and (6) proportion of ambulance patients with a serious emergency condition who survive to admission, and to 7 days post admission.

Response time was not case-mix-adjusted as this is entirely in the control of the ambulance service. Of the other five, only the mean change in pain score was not affected by other factors. For the four case-mix-adjusted indicators, age, deprivation (measured by the Index of Multiple Deprivation) and condition or call reason were important, and sex affected all indicators except survival. Receiving hospital was an important factor in the indicators for unnecessary transports and survival at 7 days post incident. The overall accuracy rate for identifying 16 emergency conditions at the time of the 999 call was 61%, although this varied substantially for individual conditions (10–85%). The rate for potentially inappropriate decisions to not transport a patient to hospital varied from 5 to 10.2 cases per 100 calls across 22 Clinical Commissioning Groups, and unnecessary transports varied from 2.4 to 8 cases per 100 calls. Survival to hospital admission ranged from 90.5 to 97.3 cases per 100 calls and survival at 7 days ranged from 90.5 to 98 cases per 100 calls. This confirms that the proportion of calls to the ambulance service that are truly life-threatening is very small. Indicators not affected by external hospital factors could be used to measure differences between different services or areas as well as monitoring over time. Indicators for which hospital has an effect can be used for within-service comparisons and trends but would not make a fair judgement between services as the outcome is influenced by service factors outside ambulance service control.

Early in the programme we identified one other potential indicator – survival rates for patients who call for urgent conditions – but could not reliably identify all of the relevant patient population. Instead, we conducted a retrospective case review of a sample of patients we identified from the linked data as having a low risk of death. We assessed 153 cases of patients who died within 3 days and identified 29 patients who were not taken to hospital at the time of the incident. These 29 cases were reviewed using structured judgement review. Overall, 8 cases out of 29 (27.6%) were judged definitely avoidable, 8 (27.6%) probably avoidable and 13 (44.8%) slightly or definitely not avoidable.

The economic study analysed 182,566 cases, of which 10,151 (5.6%) were ‘hear and treat’ so received no ambulance face-to-face response, 51,223 (28.0%) received a ‘see and treat’ response and 121,192 (66.4%) were taken to hospital. The total mean cost of a ‘hear and treat’ call was £125, ‘see and treat’ call was £415 and ‘see and convey’ call £1745. The main reasons for these cost differences are the differences in inpatient costs. We also compared costs for correct and incorrect decisions using the indicators for potential inappropriate non-conveyance and unnecessary transports to identify calls with correct and incorrect decisions. We found that the mean total cost of an incorrect non-conveyance decision is £313 more than that of a correct conveyance decision. The mean total cost of an incorrect conveyance decision is £237 more than that of a correct non-conveyance decision. There are potential costs savings if decision-making can be improved.

Limitations

A serious limitation to this programme was the delays caused by difficulties in obtaining the linked data. The data linkage component was contracted to the then Health and Social Care Information Centre (HSCIC) (now NHS Digital). NHS Digital holds the central HES data and also provided a trusted data linkage service. We planned to use this service to obtain the first set of linked data by the end of year 2 and a second set in year 3 (2013–14). During 2013 major data security issues arose at NHS Digital and, as a result, there was a major review and restructuring of the organisation. This meant that no data were released for (any) research use for almost 2 years. As a consequence, we did not receive a linked data set until October 2016, which was 4 months after the original expected end date of the programme. These delays meant that we have been unable to conduct the work to further test and validate the indicators set out in objective 4.

The complexities of data management and linkage also proved to be costly and time-consuming. For some groups of patients (e.g. those who do not receive an ambulance response), there is insufficient information held to achieve high levels of matching and data linkage, so this patient group is not well represented in some of the analyses. The construction of the case-mix-adjusted indicators was also complex. Although processes to do this have now been established, we have used data from one ambulance service only.

The results showed that the rates measured vary by geographical area and so the indicators would have to be recalibrated for each individual ambulance service. The inability to test the indicators and their usefulness in a 'real-world' setting is also a limitation to further use.

Conclusions

We identified and prioritised, through systematic reviews of the literature followed by a series of formal consensus processes with a wide range of stakeholders, a set of potential ambulance service quality measures that reflect the preferences of both services and users. We also created a comprehensive linked data set providing information for individual calls that extends beyond the prehospital component of care, although this proved to be a complex and time-consuming process. Six candidate indicators were developed using case-mix adjustment and, of these, four were found to need adjustment to make fair comparisons. Hospital was found to have a substantial effect on the process or outcome for two indicators, which means that these are suitable for use only at an individual service or system level. Other indicators could be used to make comparisons between regions or services. The complexities of both creating linked data and constructing the indicators means that, at present, they are of limited value as it would not be possible to measure them routinely. There is a pressing need to improve information systems and data linkage processes if more sophisticated case-mix-adjusted performance and quality measurement is to be implemented. Currently, no national, centralised ambulance service data are collected. Development of an ambulance equivalent to the Emergency Care Dataset that could be linked to this and HES data would produce real benefits for better performance measurement and associated clinical audit. The different studies conducted as part of the PhOEBE programme open up a wide range of potential future research:

- The measures prioritised through the consensus studies should be further developed, validated and examined to investigate their importance, validity, feasibility, relevance and sensitivity to differences in services, service changes and quality improvement efforts in practice.
- New measures, such as patient-related experience measures, should be developed based on our understanding of what is important to patients using ambulance services.
- Future data sets linking ambulance, hospital and mortality data could be used to investigate the effects, safety and costs of different pathways and processes for a variety of clinical conditions and patient outcomes.

Funding

Funding for this study was provided by the Programme Grants for Applied Research programme of the National Institute for Health Research.

Programme Grants for Applied Research

ISSN 2050-4322 (Print)

ISSN 2050-4330 (Online)

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: journals.library@nihr.ac.uk

The full PGfAR archive is freely available to view online at www.journalslibrary.nihr.ac.uk/pgfar. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

Criteria for inclusion in the *Programme Grants for Applied Research* journal

Reports are published in *Programme Grants for Applied Research* (PGfAR) if (1) they have resulted from work for the PGfAR programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Programme Grants for Applied Research programme

The Programme Grants for Applied Research (PGfAR) programme, part of the National Institute for Health Research (NIHR), was set up in 2006 to produce independent research findings that will have practical application for the benefit of patients and the NHS in the relatively near future. The Programme is managed by the NIHR Central Commissioning Facility (CCF) with strategic input from the Programme Director.

The programme is a national response mode funding scheme that aims to provide evidence to improve health outcomes in England through promotion of health, prevention of ill health, and optimal disease management (including safety and quality), with particular emphasis on conditions causing significant disease burden.

For more information about the PGfAR programme please visit the website: http://www.nihr.ac.uk/funding/programme-grants-for-applied-research.htm

This report

The research reported in this issue of the journal was funded by PGfAR as project number RP-PG-0609-10195. The contractual start date was in June 2011. The final report began editorial review in November 2017 and was accepted for publication in November 2018. As the funder, the PGfAR programme agreed the research questions and study designs in advance with the investigators. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The PGfAR editors and production house have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the final report document. However, they do not accept liability for damages or losses arising from material published in this report.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).