

The  
University  
Of  
Sheffield.

# ***D*evelopment *A*nd *V*alidation of *R*isk-adjusted *O*utcomes for *S*ystems of Emergency Medical Care**

## ***The DAVROS Project*** ***“Improving emergency care”***

Medical Care Research Unit, School of Health and Related Research

**“Protocol”**

# **DAVROS: Development and validation of risk-adjusted outcomes for systems of emergency medical care**

## **1) Background**

### **1.1) The need for evaluation of systems of emergency care**

There are around 1 million emergency medical hospital admissions each year in the United Kingdom, with about 7% of these resulting in death whilst in hospital (Baker). The development of time-critical interventions for common, life-threatening conditions, such as stroke and myocardial infarction, has emphasised the need for systems of emergency care that can deliver these interventions in an optimal manner. Substantial research has been undertaken, in the form of clinical trials, to establish the efficacy of emergency treatments. Evaluation of systems of emergency care is now required to ensure that these treatments realise their full potential.

Emergency patients are rarely managed by a single service, but by a succession of services (e.g. call operators, ambulance services, emergency department). Hence they are managed by a system of services leading from first contact to definitive care. We define a system of emergency care as consisting of the medical and administrative services that provide and organise care from the moment emergency medical help is requested to the point at which the patient receives definitive treatment for their emergency complaint. The system will thus encompass ambulance services, emergency departments, and inpatient medical care, including operating theatre and intensive care, and their organisation. In some settings general practitioners may also be involved, although their role in out-of-hours emergency care is generally diminishing. The most practical definition of a "unit" for a system of emergency care is likely to be the ambulance service, although this does not prevent analysis of different elements of the system.

There have been few formal evaluations of systems of emergency care. Randomised trials or other experimental methods are usually impractical or unethical, so most evaluations have used weaker methods, such as case studies or uncontrolled before-and-after intervention studies, and have focussed on process of care measures, such as waiting times. Meanwhile, comparison of crude mortality between systems is limited by differences in case mix and differences in the way people use the emergency care system.

Elements of emergency care systems have also been the focus for recent national audits. The National Patient Safety Agency audit of medical intensive care admissions revealed a number of shortcomings in emergency care and made recommendations for improvement (NCEPOD). However, this study did not compare outcomes between services and it was not clear what evidence the recommendations were based upon. The National Audit Office surveyed waiting times and patient satisfaction across emergency departments and found that waiting had recently shortened and this was associated with some evidence of improved patient satisfaction (National Audit Office). However, it is not known how these changes have affected quality of care, mortality and morbidity.

Given the lack of reliable evidence in this area it is not surprising that there are substantial variations between different systems. Characteristics that vary between systems and may influence patient outcomes are outlined below. We are currently undertaking research to characterise emergency care systems in the United Kingdom, which can differ markedly. However, there is much greater variation between the emergency medical care systems of different countries, and we know nothing about which of these very different systems is the most effective. Research is required to determine which systems of emergency care are associated with improved patient outcomes. This will only be possible if we can develop reliable methods for evaluation.

### **1.2) System characteristics that may influence outcomes in emergency medical care**

1. Funding
2. Patient volume

3. Organisation: e.g. centralised or dispersed expertise, use of referral networks, use of inter-hospital transfers, practitioner to patient or patient to practitioner delivery of care
4. Facilities: e.g. number and type of ambulances and response vehicles, emergency department (resuscitation, high-dependency), observation facilities (clinical decision unit, chest pain unit), inpatient (intensive, high-dependency or coronary care units), respiratory support unit & telemetry beds, emergency admission ward, diagnostic services
5. Staffing: e.g. seniority, availability of specialist staff, paramedical staff, specialist nursing staff & nurse practitioners, staff:patient ratios, staff rostering (night-time cover), emergency physicians, acute physicians
6. Timing: e.g. time to definitive care, ambulance response times, time spent in emergency department, inter hospital distances and transfer times
7. Specialist services: e.g. cardiac catheterisation, surgical specialities, emergency endoscopy, intensive care outreach
8. Staff training: e.g. advanced life support, advanced trauma life support & other training courses, qualifications, revalidation and re-accreditation, proportion of time spent in training
9. Audit: e.g. participation in national audit, adverse incident reporting
10. Use of guidelines

### **1.3) Rationale for our proposed approach to evaluation**

When experimental methods are impractical the presence of variation in the way services are organised and delivered can allow services to be compared using observational methods. However, before this approach is taken we need to ensure that these methods will be valid and appropriate. This requires development of methods which: a) take a population-wide approach, b) focus upon important outcomes, and c) take differences in case mix into account.

A population-wide approach is required because although emergency medical patients have a variety of complaints, often with multiple pathologies and co-morbidities, which may be managed by a variety of different specialties or professions in a variety of different settings, they are all managed by the same system. Changes to the system, such as introducing paramedic practitioners or fast-track systems for specific patients, affect all or nearly all emergency patients. Any attempt to restrict evaluation to a single group of patients on the basis of pathology will focus on an unrepresentative minority and fail to assess the whole system. For example, evaluation limited to patients with myocardial infarction will favour systems of care that prioritise patients with myocardial infarction, possibly at the expense of other patients, and may even favour systems of care that prioritise patients with a clear diagnosis and no co-morbidities over those with an unclear diagnosis or multiple pathologies.

Evaluations of emergency care should focus on outcomes. To date, most population evaluations of emergency care have focussed on processes of care, such as ambulance service response times and emergency department waiting times. Evaluations of emergency care in specific groups, such as trauma and cardiac patients, have often focussed on mortality but rarely on quality of life, disability or other morbidity. We propose to develop a methodology for evaluation that will focus on mortality in the general emergency care population and to explore the potential for measuring quality of life.

Differences in outcome between emergency care systems not due to random variation or differences in data collection, may be due to differences in population characteristics (age, gender balance, socio-economic status, ethnicity and prevalence of underlying disease), differences in the way in which the population uses or is admitted to the emergency services (e.g. an emergency care system that is frequently used by patients with minor ailments is likely to have better unadjusted outcomes than one that is only used by critically ill patients), or differences in the quality of care provided. Any attempt to compare the performance of systems of emergency medical care therefore needs to use a robust and reliable method for adjusting for differences in population characteristics and the way the service is used. We propose to use variables measured at presentation to the emergency service that could act as surrogate measures for emergency care population morbidity and use of emergency services to provide risk-adjusted estimates of outcomes for systems of emergency care.

#### **1.4) Past and current research relevant to the proposal**

Risk-adjustment methods have been developed in a number of fields, such as trauma care (Champion), adult, paediatric and neonatal intensive care (Harrison, Brady, Parry), and acute coronary syndromes (Antman), and have been extensively used for audit and research. Important research into trauma services, such as examining the effect of trauma teams, senior doctors and access to neurosurgical services, and into intensive care, examining the capacity of neonatal units and the outcome of nighttime discharges from adult intensive care, could not have been undertaken without risk-adjustment methods. Methods for development are therefore well established (Iezzoni), although they are by no means perfect. Problems include failure to systematically identify all potentially useful covariates, inappropriate use of coefficients developed in a different setting, and assuming that differences between expected (risk-adjusted) and observed outcomes are due to deficient care, without considering other potential causes (Lilford). We intend to use our experience of using and developing risk-adjustment methods to address these potential problems.

Attempts have been made to develop a risk adjustment score for emergency medical patients. The Rapid Acute Physiology Score (RAPS) (Rhee) was developed by selecting parameters from the APACHE score that can be feasibly collected in the pre-hospital setting (blood pressure, pulse rate, respiratory rate, and Glasgow Coma Score). The Rapid Emergency Medicine Score (REMS) (Olson) added age and peripheral oxygen saturation to the parameters used in RAPS. Studies in a Swedish hospital showed that REMS had better discriminatory power than RAPS. We confirmed this finding in emergency medical admissions transported to English hospitals by ambulance (Goodacre), but also found that only age, oxygen saturation and Glasgow Coma Scale were independent predictors of mortality. Meanwhile, the Modified Early Warning Score (MEWS) (Subbe), which uses the same parameters, but with temperature included, has been shown to predict need for intensive care and is now used to identify high-risk medical admissions.

These scores suggest that physiological variables may be useful in a risk-adjustment score but they share a number of limitations. Variables measured on arrival at hospital may be influenced by pre-hospital treatment, so variables should ideally be measured before treatments are initiated (otherwise they may merely reflect the effect of treatment rather than the severity of the underlying disease). However, variables can also be measured too early, at the first signs of an acute episode and hence again fail to reflect the true risk of the outcome. Ideally, therefore, variables should be measured as late as possible prior to medical intervention. This can be approximated as the time of entry to the emergency care system, although occasionally self-medication (eg. with bronchodilators, aspirin) or bystander intervention may have occurred already.

Limiting potential predictor variables to those currently routinely collected at presentation means that other potentially useful predictors will not have been considered. We have outlined some examples of variables that may be useful for risk-adjustment below. Finally, by moving directly from identification of a few significant predictors to derivation of a score, the researchers may have failed to develop an optimal score, and will have missed an opportunity to identify all the key prognostic variables that should be recorded and measured in emergency care research. For purposes of research, rather than audit, identification of individual predictors may be more useful than derivation of a score. Failure to appreciate the importance of this step in the development of the TRISS methodology for trauma scoring has led to misunderstandings about use of TRISS.

The Emergency Medical Services Outcomes Project (EMSOP), currently being developed in the United States (Keim), has identified a number of potential predictor variables for risk-adjustment of outcomes in out-of-hospital outcomes research and has recommended that these variables be systematically evaluated (Spaite). However, this project will only evaluate out-of-hospital emergency care and will therefore not evaluate the in-hospital elements of emergency medical care.

### **1.5) Examples of variables that could be used in risk-adjustment**

1. Basic characteristics: age, gender, post code (to derive deprivation scores)
2. Physiological variables: e.g. heart rate, respiratory rate, blood pressure, Glasgow Coma Scale, oxygen saturation
3. Rapidly identifiable co-morbidities or underlying diseases: e.g. active malignancy, chronic respiratory disease, diabetes, coronary heart disease
4. Presenting complaint: e.g. chest pain, collapse, reduced consciousness, breathing difficulties, fitting, abdominal pain, fall (elderly)
5. Location: own home, nursing care, sheltered accommodation, public place
6. Time of presentation: time of day, day of week
7. Current regular medication: number of daily medications prescribed, aspirin, GTN, warfarin, home oxygen therapy
8. Crude assessments of current health status: exercise tolerance, cognitive function, mobility
9. Paramedic's subjective impression of risk (this would not be appropriate for use in a risk-adjustment model for comparing system performance, but would be valuable in analysing and interpreting the performance of other variables)

### **2) Aims and objectives**

We aim to develop a method for evaluating the performance of systems of emergency medical care for patients attended by emergency ambulance, using the following steps:

1. Identification of variables that could potentially be used to adjust for differences in population characteristics and use of emergency services.
2. Selection of the most promising variables on the basis of frequency, feasibility for routine data collection and association with outcome (7-day mortality).
3. Development of a risk-adjustment tool from the variables identified in stage 2.
4. Measurement of the predictive validity of the risk-adjustment tool in a new population.
5. Measurement of the attributional validity of the risk-adjustment tool by comparison of observed outcomes to those predicted by the tool in a variety of different systems of emergency care
6. Exploration of potential reasons for discrepancies between observed and predicted outcome in different systems and refinement of the risk-adjustment tool

The principal objective of the project is to develop a risk-adjustment tool that can be used in a variety of settings to predict 7-day mortality in a heterogeneous population of emergency medical patients. Secondary objectives are: 1) to identify the key variables which predict mortality (7-day, 30-day and in-hospital), quality of life and length of stay in emergency medical patients, and should thus be measured in any non-randomised evaluation of these outcomes in emergency care, and 2) to evaluate the tool and the key variables in specific patient groups.

### **3) Plan of investigation**

The project will involve seven phases. Phase 1 will consist of systematic literature review and expert panel review. The settings, population, outcomes, data collection and data analysis for phases 2 to 6 are outlined below.

#### **3.1) Phase 1**

We will undertake a systematic review of the literature to identify all published studies that either: a) investigate the predictive value of variables in emergency medical patients; b) evaluate or attempt to develop a risk score for emergency medical patients; or c) measure potential confounding factors as part of a non-randomised evaluation in emergency medical care. Studies will be selected that measure the association between one or more predictor variables and mortality. We will record the variable(s) concerned, the strength of association demonstrated, the method of data collection, the setting and the population characteristics.

We will then assemble a panel of experts in emergency medical care who will review the list of potential predictor variables and add any additional potential variables that have not been identified. The panel of experts will then be asked to identify which variables can clearly not be feasibly collected in routine practice. Any variables that the panel unanimously agree are not feasible to collect will be excluded from further consideration. The experts will also be asked whenever possible to represent graphically the expected shape of the relationship between the selected predictor variables and the outcomes (paying particular attention to monotonicity and asymptotic behaviour), identify any potential interactions between variables, and identify potential reasons for missing data for each variable (e.g. would they expect data to be missing if the patient was critically ill or if they were very stable?).

### **3.2) Phase 2**

We will explore the feasibility of collecting variables by asking ambulance staff to record, in addition to routinely recorded variables, up to five additional variables when they attend a patient with a medical emergency. Different crews will collect different sets of additional variables. We will measure the proportion of missing data for each variable. Mid-way through this phase we will undertake a focus group discussion with ambulance staff to identify potentially remediable problems with collection of specific variables. We will also measure the association of each variable with each of the outcomes.

Based upon suggestions from the expert panel, we will examine potential interactions between variables, explore appropriate methods for handling missing data, and develop methods for transforming variables so as to maximise the predictive value of each variable. The expert panel will then use a consensus approach to select the most potentially useful variables on the basis of: a) measured strength of association with outcome, b) the frequency of the characteristic in the population, and c) the proportion of missing data for the variable.

### **3.3) Phase 3**

We will measure the univariate association between each of the selected variables and each outcome in a sample of patients using emergency care systems in South Yorkshire. Ambulance staff will be asked to record the subset of selected variables as soon as possible after attending the patient. All variables will then be entered into a multivariate model for each outcome to identify which variables are independent predictors of outcome. First-order interactions between variables will be measured. We will thus derive a risk-adjustment model for 7-day mortality based upon a limited number of variables that are feasible to collect and are independent predictors of outcome, and will identify which variables independently predict the other outcomes. Secondary analyses will explore the performance of individual variables and the overall model in specific patient groups to determine a) whether the model performs in a consistent manner across different patient groups, and b) whether individual variables or a different model may perform better in specific groups.

### **3.4) Phase 4**

To measure the predictive validity of the risk-adjustment model (lezzioni), we will test the model and measure the coefficients of its' constituent variables in a new sample from the same population as phase 3. The data for phases 3 and 4 will be collected together and then split, so that every first and second day of data collection is used for phase 3 and every third day is used for phase 4. This will ensure a systematic split between the samples for the two phases that is unlikely to be confounded by other factors. We plan to use a systematic split rather than a random split because the latter merely replicates the performance of the model in both datasets. Again, secondary analyses will explore the performance of individual variables and the overall model in different patient groups.

### **3.5) Phase 5**

We will use the model in six new settings to compare mortality predicted by the model to observed mortality in each setting. This phase is often considered to be the phase at which a risk-adjustment tool is applied to other populations, and any discrepancy between expected and observed outcome is assumed to be due to variation in the standard of care provided. However, this assumption may be inappropriate. The discrepancy may actually be due to differences between settings that are unrelated to quality of care, but have not been accounted for by risk-adjustment. This has been described as the attributional validity of the model (Iezzoni) – whether the differences between observed and predicted outcome can be appropriately attributed to differences in quality of care. The next step in evaluation will therefore be to explore the potential reasons for any discrepancies between expected and observed outcomes identified in this phase.

### **3.6) Phase 6**

For the attributional validity phase of the study we will select a sample of cases which exhibit the maximum discrepancy between the observed outcome and the outcome predicted by the model. Predicted survival is defined as the probability of survival generated by the model and theoretically may range between 0 (i.e. certain to die) and 1 (i.e. certain to live). Actual survival is defined as either 0 (death) or 1 (survival).

A total of no more than 200 cases will be selected, drawn from Sheffield and the phase five sites. This breaks down into 50 Sheffield cases and 25 cases from each of the six phase five sites. Cases which feature missing data or obvious data errors will be excluded from phase six.

At each site the local investigator will review the selected cases to identify the reason for the discrepancy between observed and predicted outcome. The existence of a discrepancy will be because either 1) the data used in the model have not been recorded accurately, so the model may be functioning correctly but is operating on misleading information; or, 2) that the model fails to predict illness severity appropriately, in other words, the model is at fault; or, 3) that a health care intervention causes an unpredicted outcome, i.e. that an action taking place within the emergency care system is directly responsible for the survival or failure to survive.

If the model were performing perfectly then any discrepancy between the predicted and the observed outcome would be directly attributable to quality of care. However, in reality there are always cases in which the model fails to predict illness severity appropriately and in those instances the discrepancy cannot be put down to health care. The aim of this phase therefore is to identify the common reasons why the model fails to predict illness severity appropriately.

To guide the local review process and ensure reasonable comparability across the reviewed cases we will issue the local investigators with electronic forms into which they can enter data. Local investigators will be asked to attend to the following questions when reviewing each case.

For those cases who have unexpectedly died we will seek answers to the following questions:

- 1) What was the cause of death?
- 2) Was there any potential manifestation of the cause of death in the model variables recorded at presentation?
- 3) Was there any potential manifestation of the cause of death in any presentation variable?
- 4) Did any intervention potentially contribute to death?
- 5) Could death have been prevented by intervention?
- 6) Why did the patient die when the model predicted that they would survive?

For those cases of unexpected survival the questions will be:

- 1) What was the reason for the admission?
- 2) What was the cause of any abnormal physiological variables in the model?
- 3) Did the patient receive a life-saving intervention?
- 4) Why did the patient survive when the model predicted that they would die?

Cases will then be analysed to determine if there are patterns in the reasons identified. Should any patterns so identified relate to the performance of the emergency care system then this will provide evidence for the attributional validity of the model. On the other hand, if the trends relate

to other factors, such as the patients' co-morbidity or particular use of the system, then we will collect data relating to these factors and seek to refine the model take these factors into account. One particular issue that we will attend to is the possibility that risk factors are associated with a different level of risk in different settings. We have termed this the "constant risk fallacy".

#### **4) Patient population**

The population will be adults (i.e. over eighteen) who are alive and not in cardiac arrest when attended by an emergency ambulance, and are then either admitted to hospital or die in the ambulance or emergency department. We will exclude children, injured (aged under 65), obstetric, and psychiatric patients. The TRISS methodology and the UK Trauma Audit and Research Network (UK-TARN) currently evaluate of the management of seriously injured patients and have developed extensive expertise in risk-adjustment for these patients. The systems of emergency care and important outcomes for obstetric and psychiatric patients are substantially different from the main body of patients requiring emergency care, and should therefore be evaluated separately.

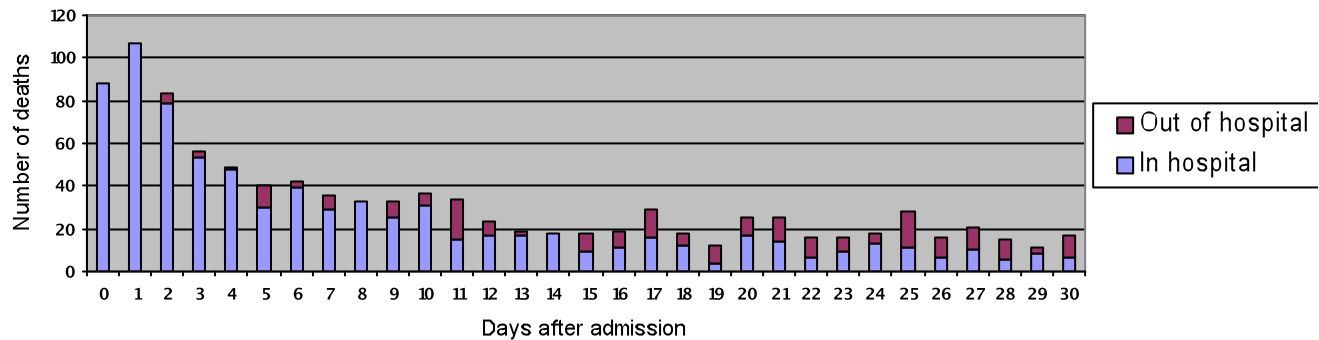
This population is deliberately defined to include a broad range of patients. This is appropriate when evaluating systems of care. Although it may be appropriate to use a narrowly defined population to evaluate patient-level interventions, this approach carries the risk in systems evaluation that systems favouring a specific group will appear to be better, even though they may fail the wider population. However, as outlined above, secondary analyses will focus upon specific patient groups, defined by presenting complaint or age group (e.g. adult or elderly).

By selecting patients brought to hospital by emergency ambulance we can select the population at a clear point of entry into the system. This also has the advantage of ensuring rigorous data collection at a specific point in time, before most emergency interventions will have been instituted. We recognise that a proportion of patients enter the system by direct presentation to hospital. Reliable data collection is likely to be problematic in these patients and the different route of presentation means that different variables are likely to be predictive, compared to those arriving by emergency ambulance. We have therefore decided to focus upon developing a model for those arriving by emergency ambulance.

#### **5) Outcome**

The primary outcome will be 7-day mortality, which will form the basis of the risk-adjustment model. Secondary outcomes will include survival to hospital discharge, outcome (home, in hospital or dead) at 30 days, quality of life at 30 days (measured by the EQ-5D), and length of hospital stay. The following figure shows the timing of in-hospital and out-of-hospital deaths up to 30 days in 1002 patients who died after emergency medical admission (Baker), and hence the relationship between the three potential choices of mortality outcome. We have selected 7-day mortality as the primary outcome because early deaths are most likely to be directly related to the quality of emergency care. In-hospital deaths may be easier to identify and compare between systems, but may be influenced by factors unrelated to quality of emergency care, such as access to palliative care outside hospital.

## Timing of death up to 30 days after emergency medical admission



We will only develop a risk-adjustment model for the primary outcome. The purpose of collecting secondary outcome data is to identify which covariates have an important association with each outcome and should therefore be adjusted for in future studies reporting this outcome. Quality of life is increasingly recognised to be an important outcome in evaluations of emergency care. However, previous experience of measuring quality of life by questionnaire follow-up to unselected emergency patients (the ESCAPE study & UK-TARN) has produced a response rate of only 30-40% despite all attempts to maximise responses. We feel that, although a low response rate will produce a biased sample, the importance of measuring quality of life is such that it should be included in future evaluations and it is worthwhile identifying key covariates, whilst allowing for an anticipated low response rate.

### 5.1) Setting

Phases 2 to 4 of the study will take place in one system of emergency care (South Yorkshire). Phases 5 and 6 will take place in six further systems of emergency care. To enhance the international generalisability of the tool we will include an Australian and a Canadian centre as two of the six new settings.

### 5.2) Identification of the study population

Patients will be identified by review of hospital computer systems. Any patient who was transported to hospital by emergency ambulance and then either died or was admitted to hospital will be included in the study. Patients who had no vital signs on ambulance arrival at the patient will be excluded (see paragraph 12, page 9, for a more detailed discussion of ethical issues raised by this project).

## 6) Planned data collection

Data collection methods will be developed during phase 2. Ambulance staff will record potential predictor variables using standard ambulance report forms that will be adapted for the study to allow collection of up to five additional variables. All staff in participating services will undergo a brief training session to ensure that: a) data collection is standardised as far as possible, and b) data collection does not delay transfer or interfere with providing emergency care. We intend that data collection will be standardised between systems to optimise reliability. However, formal evaluation of inter-rater reliability will not be feasible.

Research staff will identify eligible patients, record predictor variables from the ambulance report form and follow-up patients using the hospital computer system to identify whether the patient died in hospital or survived to discharge. A random sample of patients at each hospital will be followed up by case note review to check the accuracy of computer records. The only predictor variables not abstracted from the ambulance report form will be those that are routinely recorded on hospital computer systems and could not be influenced by the emergency care system, such as age, gender and post code. At 30 days after initial presentation all patients not known to have died will be mailed the EQ-5D questionnaire to measure self-reported quality of life.



## 7) Planned data analysis

### Development - Univariable analyses.

The form of the relationship between each potential predictor variable and the primary outcome (death at 7 days) will be examined using graphical and analytical techniques taking into account the understanding of the expert panel with regard to the form of the relationship with respect to two key characteristics - monotonicity and asymptomatic behaviour. For non-linear relationships between quantitative predictors, such as blood pressure and respiratory rate, categorisations and fractional polynomials which reflect medical understanding will be explored to describe the relationship (Sauerbrei). The strength of the relationship will be measured in terms of the area under the ROC curve (the 'c' statistic) generated by the variable.

Data items which are supposed to be routinely recorded in clinical records may be missing a) by mistake so that they can be assumed to be missing at random, b) because they are assumed to be 'normal', typically because they are not judged by the clinician to be relevant to the particular type of patient, or c) because of systematic reasons, such as time pressures in emergency care for patients in a critical condition who need immediate care.

These three types of reason for missing data lead to three distinct ways of handling missing values. If data items are missing at random, then in population case-mix adjustment models, as opposed to patient level prognostic indices, the cases with missing data can simply be omitted. If data items are missing because they were assumed to be normal then the missing items can be replaced with the normal value. If data items are missing for systematic reasons, such as the condition of the patient, then the optimum approach is to replace them with values imputed from the known values of the recorded items. In the example of critically ill patients who need immediate care, these known values will also be deranged and the known values will typically 'predict' a deranged value for the missing item.

We therefore plan to examine the different approaches for each variable in the development sample. On the basis of the focus group discussions and with the advice of the expert panel, we will identify for each variable which of the three approaches which may be appropriate. For each appropriate approach we will fit models with missing values replaced as above (together with an indicator variable for whether the value was recorded or replaced) and examine the relative performance of the different approaches. This will be done univariately, that is for each variable one at a time, and a missing value replacement strategy chosen for each variable. Imputing values will only be chosen over the simpler methods of replacement with normal values or the average of known values if this approach is shown to be significantly better.

Taking into account the form of the relationship and the method of dealing with missing values the first order interactions of all variables which are significantly related at  $p < 0.15$  to the primary outcome (death at 7 days) will be examined. Higher order interactions will be ignored.

### Development - Multivariable analyses.

All single variables and interactions which are significantly related to the primary outcome at  $p < 0.15$  will be entered into the multivariable model building analysis. Using a backward elimination strategy based on the methods described by Sauerbrei for fractional polynomials (Sauerbrei) and with missing values replaced as described above with terms removed at  $p > 0.05$ , the subset of terms independently predicting the outcome will be selected. The performance of the model in terms of the ROC curve for death and  $R^2$  for quality of life and length of stay will be examined in subsets of patients with different clinical conditions to examine whether the model performs in a consistent manner across different patient groups, and to examine how individual terms perform in sub-groups.

### Validation

Validity will be assessed using three standard criteria:

1. *Discriminant ability.* Receiver operator characteristic (ROC) curves will be used to test the ability of the score to discriminate between patients who die and patients who survive.

2. *Likelihood.* The risk score will be fitted as an explanatory variable in a logistic regression model with mortality as the outcome. The log likelihood of the model will indicate the explanatory power of the score, while an indication of the goodness-of-fit of the model will be given from the residuals by calculating the Pearson Chi-square statistic.

3. *Calibration.* The accuracy of the predictions of the risk score will be assessed across the whole range of scores using techniques described by Hosmer and Lemeshow.

## 8) Sample size

### Development sample (phase 3)

The sample should be large enough to be confident that none of the covariates which have been measured and which have an important influence on the outcome will be missed because of lack of evidence of the association. Two of the outcomes (quality of life and length of stay) are semi-quantitative and some of the covariates (such as the physiological measures) are quantitative. However, the usual considerations mean that for a given sample size the covariates least likely to be detected are categorical covariates or factors (such as sex or day of week) associated with the categorical outcome, death. We have therefore set the sample size so that there will be at least an 80% chance of detecting as statistically significant an association between a factor (present in 25% of the population) and death assuming that the risk of death in patients with the factor relative to those without is  $0.075/0.05 = 1.5$ .

For calculating the required sample size the significance level used for including terms in the model must also be specified. Arguments and evidence for using significance levels between 0.01 and 0.16 have been put forward. The required sample sizes for different sizes of test are shown in the table. A smaller size of test for inclusion or retention in the model will of course result in a smaller model, which is an important consideration in developing individual risk predictor models for use in clinical practice eg. for prognostic indices or therapeutic decisions.

Significance level	n <sub>1</sub>	n <sub>2</sub>	Total
0.01	1520	6082	7602
0.02	1307	5227	6534
0.05	1022	4088	5110
0.10	804	3219	4023
0.16	657	2627	3284

However, we are constructing a risk-adjustment model for use in research in populations and the size or complexity of the model is much less of an issue. Nevertheless, we expect to have up to 20 possible predictors, various transforms of the continuous covariates, and a few first-order interactions to investigate, and given the large number of terms it is important to be able to avoid over-fitting. We are therefore planning to use a 'conventional' significance level of 0.05 in a backwards elimination strategy to develop the model.

### Validation sample (phase 4)

The developed model will be validated using the discriminant ability, likelihood and calibration techniques described earlier. The validation sample will be chosen by systematically splitting the total sample using data from some days in the development sample and data from other days in the validation sample. There are arguments about what proportions of the sample should be used for development and validation. We plan to split the sample 2:1 so that data from two consecutive days will be used for development, the third day for validation, and so on. Thus the total sample will be systematically rather than randomly split, but in a way which is very unlikely to be confounded by other differences.

The sample sizes for each phase will therefore be:

Phase 1, Identification of variables: no sample required

Phase 2, Selection of variables: 5000 patients in one emergency care system

Phase 3, Development: 5000 patients in one system

Phase 4, Assessment of predictive validity: 2500 patients in one system

Phase 5, Validation in other settings: 9000 patients in six new systems

Phase 6, Analysis of attributional validity: expert panel consensus review

Pilot data from suggest that it will be feasible to identify at least 50 eligible cases per day with anticipated 7-day mortality of 5% (Clancy). We estimate that each Research Assistant will initially be able to manage, on average, 40 patients per working day (identifying patients, abstracting predictive data and recording outcomes), but should become more efficient as the project progresses.

## 9) Timescale

Phase 1, Identification of variables: 6 months.

Phase 2, Selection of variables: 6 months

Phases 3 & 4, Derivation and assessment of predictive validity: 9 months

Phase 5, Validation in other settings: 9 months

Phase 6, Analysis of attributional validity: 6 months

Dissemination: 6 months

Total = 42 months

	Month													
	1 to 3	4 to 6	7 to 9	10 to 12	13 to 15	16 to 18	19 to 21	22 to 24	25 to 27	28 to 30	31 to 33	34 to 36	37 to 39	40 to 42
Phase	1		2		3 & 4			5			6			
Project staff														
Project Manager	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Research Assistant 1		X	X	X	X	X	X	X	X	X	X	X	X	
Research Assistant 2		X	X	X	X	X	X	X	X	X	X	X	X	
Research Assistant 3								X	X	X	X	X		
Clerical Assistant	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Project centres														
South Yorkshire			X	X	X	X	X							
Other UK centres								X	X	X	X	X		
Australia & Canada								X	X	X	X	X		

## 10) The Research Team

This project will involve collaboration between health service researchers and clinicians involved in delivering emergency medical care. Steve Goodacre, the Principal Investigator, has training in both health service research and emergency medicine and has experience of leading multi-centre investigations in emergency care. He is currently principal investigator for both the ESCAPE multi-centre evaluation of chest pain units in the NHS and the evaluation of the National Infarct Angioplasty Pilots. Jon Nicholl is Director of the Medical Care Research Unit (MCRU) and has undertaken many internationally important investigations in emergency care, including evaluations of trauma centres, helicopter emergency services and paramedic interventions. David Harrison, a statistician from the Intensive Care National Audit and Research Centre (ICNARC), has specialist expertise in developing and using methods of risk-adjustment. Richard Wilson, the Project Manager, has experience of managing health care evaluations in the NHS with the Review Body for Interventional Procedures (ReBIP).

The research team will collaborate with leading emergency clinicians from the UK (Mike Clancy, Jim Wardrope and Alison Walker), Australia (Anne-Maree Kelly) and Canada (Michael Schull) to ensure that the investigation is feasible in the emergency setting, relevant to emergency practice and internationally important. Alison Walker is Medical Director of Yorkshire Ambulance Service and thus well placed to facilitate data collection. Jim Wardrope is President of the College of Emergency Medicine and thus well placed to represent expertise in emergency medicine.

## 11) Project management

A Project Management Group consisting of the UK researchers, the Project Manager and the Research Assistants will meet every three months to oversee the project. A Steering Group will be assembled to meet every six to twelve months to provide independent advice to the Project Management Group.

## **12) Ethical and Other Implications**

Two key ethical issues must be considered in this study: 1) Potential risk to patients if data collection interferes with care or delays transport to hospital; and 2) The need to use patient data in a situation where gaining informed consent is impractical. These two issues interact to a certain extent. The main barrier to gaining informed consent to the use of data is that this process will delay transfer to hospital and interfere with providing care.

We will minimise the risk to patients by: a) providing training for all participating ambulance staff to emphasize that data collection should not interfere with providing care or transport to hospital, and b) ensuring that the variables recorded can be easily measured in the emergency situation and are those normally required to provide appropriate care. We will ask an independent panel, including a patient representative and an expert in pre-hospital care, to review the data collection requirements and ensure that they will not compromise patient care.

Under the Data Protection Act (1998) confidential patient data may be used for medical research without patient consent if it is not practical to obtain consent from each individual patient. We believe that this is clearly the case in this study. However, we will take steps to ensure that patients are aware that their data may be used for research purposes. Posters and leaflets will be displayed in emergency departments informing patients that this study is taking place. The leaflets will contain forms by means of which patients can withdraw from the study. We will also ask the ethics committee to review this specific issue<sup>1</sup>.

## **13) Public Engagement in Science**

Communication with the public has an important role to play in two aspects of this project. Firstly, since the project requires patient data to be collected for research purposes without explicit individual patient consent it is an ethical requirement that the public are informed of the process and the purpose of data collection. Secondly, the ultimate aim of this project is to produce a risk-adjustment method that will allow the performance of individual systems of emergency care to be evaluated and compared. The public has a right to know how their local emergency care system compares to others, but in providing this information it is crucial that data are not presented in a misleading way that may inappropriately undermine confidence in local health services and demoralise staff.

We will create a user group consisting of representative members of the public, assisted by experts in pre-hospital care, who will review the data collection processes for the project and the planned output at each phase. During the validation phases (4 to 6) we will develop methods to present outcome data to the public and pilot these with the user groups to ensure that the output of this project is interpreted appropriately.

## **14) Exploitation and Dissemination**

This research is unlikely to generate commercially exploitable results.

The principal outputs of this project will be:

1. A risk-adjustment tool that can be used in a wide range of different settings to compare mortality among patients with medical emergencies.
2. Identification of the key predictor variables that should be collected and used in the analysis of any non-randomised study of emergency medical care.

---

<sup>1</sup> Support under Section 60 of the Health and Social Care Act 2001 for the use of identifiable information without prior consent has been sought from the Patient Information Advisory Group. Support granted on 12<sup>th</sup> July, 2007 (ref PIAG 1-05(FT1)/2007).

3. Estimates of the predictive coefficients for each variable in the overall population and in different subgroups of patient.

These outputs will be valuable to researchers, clinicians, managers and policy-makers in emergency medical care. Dissemination will aim to maximise awareness in the research community of which variables should be recorded in any evaluation of emergency care. This will include publication of guidelines and summaries that outline how to undertake evaluations in emergency care, workshops and presentations at scientific meetings, and dissemination of a summary of our findings to key researchers and decision-makers. We will specifically target those responsible for commissioning and peer reviewing proposals for evaluations in emergency care to ensure that our findings are taken into account in future evaluations.

In addition to these outputs, we anticipate that this project will result in a method to allow risk-adjusted comparison of systems of emergency care throughout the UK and other countries. We therefore plan to focus dissemination upon those responsible for planning emergency care and promote routine use of the risk-adjustment method. Ultimately we anticipate that this project will lead to a national network of audit and research that will compare systems of emergency care, similar to those already established to evaluate trauma care and intensive care.

## References

- Baker M, Clancy M. Can mortality rates for patients who die within the emergency department be easily measured? *Emerg Med J*, in press.
- National Confidential Enquiry into Patient Outcome and Death (NCEPOD). An acute problem? NCEPOD, 2005.
- National Audit Office. Improving emergency care in England. London, the Stationary Office, 2004.
- Champion HR, Sacco WJ, Copes WS, Gann DS, Gennarelli TA, Flannagan MA. A revision of the trauma score. *J Trauma* 1989;29:623-9.
- Harrison DA, Brady AR, Parry GJ, et al. Recalibration of risk prediction models in a large multicentre cohort of admissions to adult, general critical care units in the United Kingdom. *Crit Care Med*, in press.
- Brady AR, Harrison D, Black S, et al. Assessment and optimization of mortality prediction tools for admissions to pediatric intensive care in the United Kingdom. *Pediatrics*, in press.
- Parry GJ for UK Neonatal Staffing Study Group. A prospective evaluation of patient volume, staffing and workload in relation to risk-adjusted outcomes in a random, stratified sample of all UK neonatal intensive care units. *Lancet* 2002; 359:99-107.
- Antman EM, Cohen M, Berninck PJ et al. The TIMI risk score for unstable angina/non ST elevation myocardial infarction: a method for prognostication and therapeutic decision-making. *JAMA* 2000;284:835-42.
- Iezzoni LI (Ed). Risk adjustment for measuring healthcare outcomes, 2<sup>nd</sup> ed. Chicago IL: Health Administration Press; 1997.
- Lilford R, Mohammed MA, Spiegelhalter D, Thomson R. Use and misuse of outcome data in managing performance of acute medical care: avoiding institutional stigma. *Lancet* 2004;363:1147-54.
- Rhee K, Fisher C, Willitis N. The Rapid Acute Physiology Score. *Am J Emerg Med* 1987;5:278-86.
- Olsson T, Lind L. Comparison of the Rapid Emergency Medicine Score and APACHE II in nonsurgical emergency department patients. *Acad Emerg Med* 2003;10:1040-8.
- Olsson T, Terent A, Lind L. Rapid Emergency Medicine Score: a new prognostic tool for in-hospital mortality in nonsurgical emergency department patients. *J Intern Med* 2004;255:579-587.
- Olsson T, Terent A, Lind L. Rapid Emergency Medicine Score can predict long-term mortality in nonsurgical emergency department patients. *Acad Emerg Med* 2004;11:1008-13.
- Goodacre S, Turner J, Nicholl JP. Prediction of mortality among emergency medical admissions. *Emerg Med J*, in press.
- Subbe CP, Kruger M, Rutherford P and Gemmel L. Validation of a modified Early Warning Score in medical admissions. *Q J Med* 2001; 94: 521-526.
- Spaite DW, Maio R, Garrison HG et al. Emergency Medical Services Outcome Project (EMSOP) II: Developing the foundation and conceptual models for out-of-hospital outcomes research. *Ann Emerg Med* 2001;37:657-633.

Spaite DW, Maio R, Garrison HG et al. Emergency Medical Services Outcome Project (EMSOP) III: The role of risk adjustment in out-of-hospital outcomes research. *Ann Emerg Med* 2002;40:79-88.

Keim SM, Spaite DW, Miao RF et al. Establishing the scope and methodological approach to out-of-hospital outcomes and effectiveness research. *Acad Emerg Med* 2004;11:1067-1073.

United Kingdom Department of Health. Hospital Episodes Statistics 2003-4; <http://www.hesonline.nhs.uk> (accessed 27/06/05).

Sauerbrei W, Royston P. Building multivariable prognostic and diagnostic models: Transformation of the predictors by using fractional polynomials. *J R Statist Soc* 1999;162:71-94.