# Title page

## Working title

Mortality among deteriorating ward patients referred to critical care: a prospective observational cohort study in 48 NHS hospitals

## Authors

* Steve Harris
* Mervyn Singer
* Colin Sanderson
* David Harrison
* Kathy Rowan

# Abstract

## Background

Identifying and responding to deterioration among ward patients includes early access to critical care. However, critical care provision in the NHS is constrained, and the effects of this on delay to critical care admission, and patient outcomes are poorly understood.

## Methods

We conducted a prospective cohort study of consecutive deteriorating ward patients referred to critical care in 48 NHS hospitals (1 November 2010 — 31 December 2011). We recorded both the assessor’s recommendation for critical care, and the decision to admit. Admissions to critical care within one week, and deaths within one year were defined by linking to national registries. Incidence models were stratified by the NHS National Early Warning Score (NEWS) risk class, and used generalised estimating equations. Decision making and survival were modelled with random effects for the hospital using logistic regression and proportional hazards models respectively.

## Findings

Critical care teams assessed 15158 patients of whom 6759 (45%) were in the highest NEWS risk class giving an incidence of 17 NEWS high risk patients (95%CI 17–18) per hospital per month. 5164 (34%) patients were already in established organ failure with only 870 patients (6%) already on organ support. Sepsis was reported in 9296 (61%) patients.

2141 (14%) patients with treatment limitation orders were declined critical care. 7-day, 90-day, and 1-year mortalities were 41%, 65%, and 76% respectively.

Of the remainder, the bedside assessor recommended 4976 (38%) to critical care. 3375 (68%) were immediately accepted. The median delay between asssessment and admission was 2 hours (IQR 1 to 4). Patients over 80 years were less likely to be accepted (OR 0.60 95%CI 0.53–0.69). Despite adjustment for patient specific risk factors, decision making varied markedly between hospitals (median inter-hospital OR 2.11, 95%1.81–2.42).

Of the 1601 (32%) patients recommended but initially refused, 1021 (64%) were admitted later with a median additional delay of 6 hours (IQR 5–7). A further 179 (11%) patients recommended and refused died without admission.

Excluding patients with treatment limits, 7-day, 90-day, and 1-year mortalities were 14%, 30%, and 39% respectively. Mortality occured early with 964 of first week deaths (53%) within two days. Survival also varied between hospitals with a median interhospital hazard ratio of 1.29 (95%CI, 1.22–1.35).

The critical care unit was full at the time of 1198 (8%) assessments. The patients concerned were less likely to be accepted to critical care (OR 0.72 [95%CI 0.59–0.88]), and less likely to be admitted promptly (OR 0.27 [95%CI 0.19–0.37]). Increasing occupancy was associated with greater physiological deterioration pending admission (p=0.01). We could not exclude an effect of occupancy on 90-day survival (HR 1.07, 95%CI 1.00–-1.15).

## Conclusion

Deteriorating ward patients referred to critical care are vulnerable with a high initial mortality. Despite clinical recommendation for admission a substantial minority die without admission. High critical care occupancy both prejudices and delays admission.

## Funding

Wellcome Trust, NIHR Service Support Costs, and the Intensive Care National Audit & Research Centre

# Introduction

Around 200 acute hospitals in England that care for more than 11 million overnight hospital admissions per annum. Each patient spends an average of 5 days on a hospital ward where they undergo a process of continual triage, and those who deteriorate are referred to critical care.[@Anonymous:tc] This interface between the ward and critical care has been a priority area for the English National Health Service (NHS), but available data derive from qualitative work, small retrospective studies, or voluntary reporting systems [@Luettel:2007va; @McQuillan:1998uj; @Cullinane:2005vg].

Recent international reports suggest that critical care capacity can affect decision making for these patients.[@Chen:2012hu; @Stelfox:2012iy; @Robert:2012ep] The last significant funding increase for critical care in the NHS was in 2000, and, in 2010, the United Kingdom (UK) was still ranked 24 out of 28 European countries in terms of critical care provision.[@Rhodes:2012kb] Similar results are found when for comparisons with North American health care.[@Wunsch:2008kk] This implies that access to critical care in England may be relatively constrained, and that more referrals may be delayed or refused.

These constraints will particularly affect ward patients referred to critical care that already suffer an inpatient mortality two to three times higher than average.[@Simchen:2007bn; @Buist:2002kc; @Clark:2014hu]

We set out to investigate the impact of, and circumstances surrounding delays in admission to critical care. Previous studies have typically limited themselves to comparisons of early versus late admissions, and have excluded by design those never admitted. This introduces survivorship bias (those who die before late admission), and an exclusion bias (those who survive without admission). Instead, we have prospectively followed all patients referred to critical care, traced subsequent critical care admission within the first week after referral (the ‘efferent limb’ of a rapid response system), and performed survival follow-up to one-year.

# Methods

## Study design and participants

The (SPOT)light study was a prospective observational cohort study of the deteriorating ward patient referred for ward assessment by critical care. The physiological status of the patient at the time of the first bedside assessment by critical care was prospectively recorded along with the recommendation made at the end of the assessment. By linking the records generated at the time of the bedside assessment, to records in the Intensive Care National Audit & Research Centre’s Case Mix Programme database (ICNARC CMPD), the fact and timing of admission to critical care were identified. By linking to the NHS Information Service then survival status up to one year was obtained.

Patients were eligible if they were inpatients on general hospital wards who had been referred to, and assessed by, critical care. The ward assessment had to be performed at the bedside by a member of the critical care team. This team was defined broadly to include members of the critical care outreach team (CCOT), or members of the critical care medical or nursing staff. Only the first ward assessment for a given episode of illness was eligible; cardiac arrests, planned critical care admissions, and visits by the team solely for the purpose of retrieving a patient (where a decision to admit had already been made) were excluded.

Demographic information, the date, time and location of the visit, and the level of care at the time of the visit were recorded.[@Eddleston:2009wy] Patient physiology (vital signs, arterial blood gas and laboratory measurements) at the time of, or immediately preceding, the ward assessment was abstracted along with organ support, antibiotic therapy, and a subjective assessment of the likelihood of sepsis, and its source. The assessor finally reported the level of care he or she recommended, and the actual outcome of that recommendation at that assessment as either acceptance or refusal. Treatment limitation orders were recorded for those refused critical care.

## Procedures

The study was registered on the National Institute for Health Research (NIHR) Clinical Research Network portfolio, and only hospitals participating in the CMP were eligible. Research teams at each hospital attended a Dataset Familiarization Course, and a data collection manual (containing definitions of items to be collected) was provided. The Clinical Trials Unit at ICNARC co-ordinated and provided support for research queries during the study.

Hospitals were asked to report all consecutive ward referrals to the critical care team. Contemporaneous data collection was promoted, but hospitals were also requested to identify and submit any missed referrals. Reporting was via a secure online web portal which performed real-time field and record level validation. Further on-line validation reports were completed by all hospitals before the database was locked in September 2012. Fact and date of death were then requested from the NHS Information Service. CCOT provision was reported by participating hospitals, and contemporaneous CMP data and Hospital Episode Statistics (HES) were used to define critical care provision, occupancy, and hospital characteristics. To inform completeness of capture of ward referrals and to quality control the study, we used the proportion of emergency ward admissions in the CMP successfully linked to the (SPOT)light database. Data quality was judged on a monthly basis, and only those months were linkage exceeded 80% were included.

## Statistical analysis

Survival was evaluated at 90-days. Sample size was calculated to evaluate mortality increases from delay to admission using estimates from 2007 ICNARC CMP data. The target sample size was 12,075–20,125 patients referred to critical care which allowed for delay to admission to occur in 10–40% of referrals and for mortality effect sizes between 5–10%.

From the physiology measurements at ward assessment, the ICNARC physiology score, the NHS National Early Warning Score (NEWS) and the Sequential Organ Failure Assessment (SOFA) score were calculated with missing values given zero weights as recommended.[@Anonymous:2012un; @Harrison:2007jt; @Vincent:1996vs] The NEWS score can be used to define three risk classes (Low, Medium, and High) designed to trigger an escalating clinical response.

Prompt admission to critical care was defined as one occuring within four hours of ward assessment.[@Anonymous:2013up]

The indicator of critical care unit occupancy was the difference between the maximum number of beds reported to ICNARC, and the number of actively treated patients occupying those beds at the time the ward patient was assessed. Bed pressure (occupancy) was defined as being high (zero or fewer beds available), medium (one or two beds available), or low (three or more beds available).

Incidence models were stratified by NEWS risk class. The unit of analysis was a study day so that daily fluctuations in lagged critical care occupancy could be examined. Estimation was via generalised estimating equations (GEE) with each hospital as a cluster, and day-by-day correlations modelled using a first order auto-regressive structure. Decision to admit to critical care, and promptness of admission, were modelled using multi-level logistic regression with patients nested within hospitals. Cox proportional hazards were used to model survival with a shared frailty factor for hospitals. The proportional hazards assumption was checked by inspecting plots of smoothed exponentiated standardised Schoënfeld residuals, and re-entering terms using time-varying co-efficients where necessary. Random effects are reported using the Median Odds Ratio (MOR), and the Median Hazard Ratio (MHR) using the bootstrap to generate 95% confidence intervals. These statistics represent the median difference when comparing patient outcomes from any two randomly selected hospitals.[@Bengtsson:2010ib]

Categorical data were reported as counts and percentages, and continuous data as mean (SD) or median (IQR) values. Effect measures are reported with their 95% confidence intervals.

## Role of the funding source

The study was centrally funded by the Wellcome Trust, sponsored by ICNARC, and supported at NHS hospitals through the NIHR service support costs. The funders of the study had no role in the study design; data gathering, analysis, and interpretation; writing of the report; and decision to submit for publication. The corresponding author had full access to all the data (including statistical reports and tables), takes responsibility for the integrity of the data and accuracy of the data analysis, and takes final responsibility for the decision to submit for publication.

# Results

48 hospitals reported 20,893 visits for ward assessment over 435 study months. 2,694 visits (12.9%) did not meet the inclusion criteria including 1,860 (8.9%) repeat assessments, and 586 (2.8%) assessments for recent critical care discharges. Data linkage did not meet the quality control level (> 80%) for 66 (15%) study-months excluding a further 2,440 (11.7%) visits. Of the 15,759 patients remaining, 15,158 (96.1%) completed follow-up without error and were available for analysis (Figure 1). Final data linkage (ward visits to critical care admissions) was 93% complete.

## Participating hospitals

Participating hospitals comprised 10 teaching and 38 general hospitals collecting data for a median of 8 months (IQR 5 to 9 months) between September 2010 and December 2011. Each contributed a median of 252 patients (IQR 162 to 380). CCOTs operated 24 hours/day and 7 days/week in 14 (29%) hospitals, less than 24 hours/day in 19 (40%) hospitals, and less than 7 days/week in 13 (27%) hospitals. Two hospitals had no CCOT.

There was a median of 12 (IQR 9 to 18) adult general critical care beds per hospital (mixed Level 2 [typically intensive monitoring or single organ support], and Level 3 [ventilated or multiple organ support]), most often co-located in a single physical location (45 hospitals).

Bed pressure was high (zero or fewer available beds) at the time of 1198 (8%) ward assessments, medium (one or two beds available) for 3757 (25%) assessments, and low (three or more beds available) for the remaining 10197 (67%) assessments. Critical care occupancy fluctuated with time of the day, day of the week, and season of the year (supplementary Figure 1).

## Incidence of referrals for ward assessment by critical care

The mean baseline incidence of referrals to critical care (for a non-teaching hospital with 60,000 admissions per year and 24/7 CCOT provision) was 46 (95%CI 50 to 54) patients per month of whom 17 (95%CI 17 to 18) patients met the NEWS high risk criteria at assessment. This is equivalent to 8 unselected referrals or 3 NEWS high risk referrals per 1,000 overnight admissions.

With decreasing provision of critical care outreach, the number of patients assessed also fell (supplementary Table 1). Winter was busier (IRR 1.22, 95%CI 1.14 to 1.31), and weekends quieter (IRR 0.87, 95%CI 0.82 to 0.92) than the rest of the year. When a measure of case finding was included in the models (cases assessed per 1000 overnight hospital admissions), referral incidence increased initially but then began to plateau for those hospitals with referral rates in the highest quartile (supplementary Figure 2).

## Patient characteristics and outcomes

Table 1 presents the baseline data for all ward patients assessed. Sepsis was reported in 9296 patients (61%). Of these, the respiratory system was considered to be the source about half (4772, 51%). Organ failure, defined as a SOFA score greater than or equal to two, was present in 5164 of patients (34%). 1427 patients (9%) were in respiratory failure, 2931 (19%) were in renal failure, and 4636 (31%) were in cardiovascular shock. Organ support at the time of assessment was uncommon (870 patients, 6%).

2708 (18%) patients died during the 7-days following ward assessment. Mortality was heavily front-loaded with 1539 (57%) of these deaths occuring within the first 48 hours (supplementary Figure 3). There was a clear correlation between physiological severity and early (7-day) mortality using either ward based (NEWS) or critical care scoring systems (SOFA, ICNARC) (supplementary Figure 3). As an example, the 7-day mortality was 9% (328 deaths), 15% (629 deaths), and 26% (1734 deaths)for NEWS low, medium and high risk classes respectively.

The critical care assessors judged that 5321 patients (35%) required critical care. These patients had a higher physiological severity of illness (ICNARC physiology score 17.6 versus 14.3, 95% confidence interval for difference 3.0 to 3.5), and a greater 7-day mortality (19.6% versus 17.0%, difference 1.3% to 4.0%). There was a clear correlation between measured severity and the assessors’ judgements of need (supplementary Figure 5).

Overall mortality at 90-days was 35% (5337 patients), and at one year was 44% (6703 patients).

## Patient pathways following ward assessment by critical care

Patients were classified into three groups following the initial ward assessment: 2141 patients (14%) declined admission with treatment limits (pre-existing or newly-placed); 9471 patients (62%) declined admission without treatment limits (Ongoing ward care); and 3546 patients (23%) offered immediate critical care.

### Declined critical care with treatment limits

The 2141 patients with treatment limits had a 7-day mortality of 41% (881 deaths). The initial decision to decline admission was reversed in just 76 patients (4%) of whom 26 (34%) died within the week. Although the final 90-day mortality was substantial 65% (1402 deaths), 506 patients (24%) survived for at least year despite the decision.

Patients declined critical care with treatment limits were older (77 versus 66 years, 95%CI for difference 11 to 12 years), and more acutely unwell (17.1 versus 13.9 ICNARC physiology points, 95%CI for the difference 2.8 to 3.6) than those assigned Ongoing ward care. Critical care occupancy did not affect the proportion of patients declined admission with treatment limits (Table 2).

### Ongoing ward care

The 9471 patients for ongoing ward care had a 7-day mortality of 12% (1102 deaths). Most deaths (799 deaths 73%) occured on the ward but 303 deaths (27%) followed delayed critical care admission. The initial decision to decline critical care was reversed within the week for 1745 patients (18%), so a total of 2544 (27%) patients died or were admitted to critical care.

The ongoing ward care group included 1601 (17%) patients who had nonetheless been recommended critical care by the assessor. These patients had a higher 7-day mortality (18% versus 10%, 95% confidence interval for difference 5% to 9%), and were more likely to have the initial refusal reversed (36% versus 15%, risk difference 19% to 24%).

### Immediate critical care

The 3546 patients immediately accepted to critical care had a 7-day mortality of 20% (725 deaths). Just 42 (6%) of those deaths occurred before admission was arranged, but a further 254 patients (9%) were never admitted but survived nonetheless.

Those offered immediate admission were marginally younger (64.1 versus 65.6 years, 95%CI 0.8 to 2.2 years), but distinctly more unwell (18.1 versus 13.9 ICNARC physiology points, 95%CI 3.9 to 4.5) than the ongoing ward care group. As critical care occupancy at the time of the ward assessment increased, patients were less likely to be immediately accepted (Table 2, Cochran-Armitage test for trend p<0.0001).

## Delay to admission to critical care

The median delay between assessment and admission for patients immediately accepted was 2 hours (IQR 1 to 4) compared to 12 hours (IQR 5 to 29) for those whose initial refusal was subsequently reversed (median additional delay 9 hours, IQR 9 to 10) (Figure 2a). Thus prompt admission (within 4 hours) was delivered for 2277 patients (74%) when immediately accepted versus 256 (16%) when initially declined (risk difference 58%, 95%CI 56% to 60%).

For the subgroup of 580 ongoing ward care patients who had been recommended for critical care (by their assessor), but were initially declined, and later had that refusal reversed, the median delay to admission was 8 hours (IQR 3 to 22).

Increasing occupancy at assessment increased the median delay from 3 (low bed pressure), to 4 (medium pressure), to 6 hours (high pressure, Figure 2b, Jonckheere-Terpstra test for trend p=0.0004).

## Determinants of a decision to admit

We built a multi-level (patients nested within hospitals) logistic regression model to examine factors associated with a decision to admit for patients without treatment limits (Table 3). As with the univariate comparisons above, older patients were less likely to be admitted (patients over 80 years: odds ratio 0.60, 0.53 to 0.69), and more acutely unwell patients were more likely to be admitted (OR 1.07 per ICNARC physiology point, 95% confidence interval 1.06 to 1.07). Similarly, patients already receiving organ support (1.83, 1.55 to 2.16), or clinically judged to be peri-arrest (6.32, 5.18 to 7.70) were also more likely to be admitted.

Patients referred out-of-hours (7pm-7am), during the weekend, or during the winter were more likely to be offered critical care (odds ratios between 1.04 to 1.33), but those assessed when bed pressures were high (OR 0.70, 0.57 to 0.86), or medium (0.87, 0.77 to 0.98) were less likely to be accepted. We estimated, that in this sample, an additional 122 patients (95%CI 53 to 186) would have been immediately accepted had there been no limitations on critical care capacity.

The model also demonstrated significant hospital level variation with a MOR of 2.11 (95% confidence interval 1.81 to 2.42) which differed little to that estimated excluding patient predictors (MOR 2.18, 1.82 to 2.60). The MOR summarises the differences when comparing decision making for similar patients from any two randomly selected hospitals, and consistency when excluding patient level predictors suggests that it is a true hospital level difference.

## Determinants of prompt admission

The modelling was repeated but now with the delivery of admission to critical care within 4 hours (a prompt admission) as the outcome, and the decision to admit as an additional predictor (Table 3). In this analysis, we also excluded 358 (2.4%) patients whose admission was inevitably delayed urgent surgery.

The patient-level predictors of prompt admission were broadly similar to those for decisions to admit, with younger and sicker patients being admitted more promptly. However, patients assessed during the winter, while being more likely to be offered critical care, were less likely to be admitted promptly (OR 0.76, 0.64 to 0.90). The strongest predictor of prompt admission was a decision to admit at the initial bedside assessment (OR 69, 59 to 81). Even though the decision to admit was included in the model, bed pressure still had a marked effect (high pressure: 0.27, 0.19 to 0.37) and hospital level variation persisted (MOR 1.89, 1.63 to 2.21).

## Determinants of 90-day mortality

Amongst patients without treatment limits, there were 372 deaths (3%) by the end of the first day, 1742 (13%) by the end of the first week, 3130 (24%) by the 30 days, and 3946(30%) by 90 days (proportions from Kaplan-Meier failure function).

A series of models were fitted with 90-day survival as the dependent variable for patients without treatment limits. The final best model (Table 4) incorporated a time-varying effect for measured physiological severity and reported peri-arrest status such that their effects were attenuated after the first week (supplementary Figure 6).

Other patient level risk factors were consistent with the existing literature on outcomes in similar patients: older patients, and those with sepsis (other than genito-urinary) had worse survival.[@Harrison:2004brb] Patients assessed during the winter months, over the weekend, and out-of-hours did not have a significantly worse adjusted survival than baseline.

Critical care occupancy did not affect adjusted mortality in the multi-level model (high pressure: hazard ratio 1.03, 95% confidence interval 0.90 to 1.17). A single level model (supplementary Table 2), constructed in case occupancy was mediated through rather than confounded by the effect of the hospital, similarly could not exclude a null effect (hazard ratio 1.07, 1.00 to 1.15, p=0.06).

The full multi-level model demonstrated significant hospital level variation in survival (MHR 1.28, 1.22 to 1.34) which was little altered by adjustment for patient level risk factors (MHR 1.29, 1.22 to 1.35).

Repeating the survival model in the subgroup recommended to critical care at the initial assessment produced similar effects albeit with less precision (supplementary Table 2).

# Discussion

We describe the events following initial bedside assessment by critical care of more than 15000 ward patients in 48 acute NHS hospitals. Nearly half (45%) of these patients were defined as being at high risk by current guidelines[@Anonymous:2012un], and one-third (33%) were assessed when the critical care unit was under-strain (two or fewer beds available). One in twelve (8%) were assessed when the unit was completely full.

Critical care capacity affected the decision to admit, and subsequent patient pathways. Moreover, the decision making varied between hospitals. Bias against admitting the elderly provided additional evidence of rationing.

The consequences of critical care strain, and inter-hospital variation in decision making were two-fold. Firstly, affected patients had their admission delayed either directly (delayed immediate critical care), or indirectly (declined immediate admission before late reversal of the initial decision). Secondly, patients were simply less likely to ever be cared for critical care during the week following assessment. Correspondingly, they were more likely to die on the ward without critical care. This was true even for the cohort of patients for whom the bedside assessor thought critical care was justified.

Despite an immediate decision to admit, one-quarter of patients were still not admitted within the fours hours recommended. Compared to the international literature this four hour target is not strict.[@Stelfox:2012iy; @Robert:2012ep]

The mortality rate in these patients is high. Around one in three patients assessed die within 90-days, and nearly one half do not survive a year. In fact, the mortality is not dissimilar to that seen in the unselected critical care population even though only a mintority of these ward patients were admitted.[@Harrison:2014ei; @Harrison:2007jt]

Importantly, early (7-day) mortality is still elevated either for patients defined as objectively low risk by NEWS class, or subjectively low risk by the bedside assessor. In other words, mere referral for ward assessment by critical care is already effectively identifying a high risk population. This risk is heavily front-loaded with around half of deaths in the first week, and half of those deaths in the first 48 hours. This implies that the time window for successful intervention has either passed, or is narrow.

Hospital level variation is a significant factor. We could not show a direct effect of occupancy on mortality in this model. However, if part of the effect of being in a ‘good’ hospital is mediated through how critical care occupancy is managed then this difference would have been inappropriately ‘adjusted away’. In our single level model, the effect of occupancy on mortality was greater; thus we cannot confidently exclude occupancy as having a clinically important effect. The observation that, as critical care capacity becomes more constrained, those patients eventually admitted both waited longer, and deteriorated further, would be consistent with this hypothesis.

Finally, although referrals for assessment by critical care were more frequent during winter, the weekend, and out-of-hours, there was no independent decrease in risk adjusted survival in these groups.

## Comparison with other studies

There are no similar studies of ward patients referred to critical care in the UK. However it would appear that the incidence of referrals we report is much lower than in similar health care systems. We observed around 8 unselected referrals per 1,000 inpatient admissions whereas others report in the range of 25–50.[@Jones:2011gy; @Buist:2007oq; @Bell:2006gr; @Simchen:2004vf]

Despite our smaller and presumably more selected population, the mortality we observe is similar — although direct comparisons remain difficult because the reporting metrics differ (hospital mortality versus 30-day survival). For additional context, it is worth comparing the overall 61% one year survival for the ward patients referred to critical care to that for UK patients diagnosed with lung (30%), colorectal (75%) and breast cancer (95%).[@Coleman:2011ip]

Finally, while critical care occupancy rates in the UK vastly exceed those seen in a recent report from the US Veteran’s Affairs hospitals, they are not dissimilar to those in France or Canada.[@Chen:2012hu; @Stelfox:2012iy; @Robert:2012ep] However, since our delay to admission is markedly worse than these studies (between 1 and 5 hours for direct and indirect admissions versus 2 and 12 hours in this study), it is possible that the consequences of occupancy also vary. Delayed discharges because of lack of inpatient ward capacity are increasingly common in the NHS, and present a more significant impediment to ICU admission than a unit that chooses to defer discharge until the need arises but has no problem doing so when required.

## Strengths and limitations

This is the largest prospective study of deteriorating ward patients to date. The hospitals contributing are representative of the full spectrum of those in the NHS. Most importantly, we have followed up all patients referred to critical care rather than just the subset offered admission.[@OCallaghan:2012da]

This ‘denominator’ data allows us to examine the demand for critical care arising from inpatient work. Critical care is normally judged based on the patients it admits, but here we can see the wider context, and the ongoing triage of an at-risk population. Crucially, follow-up of all patients assessed generates the control group that enables us to understand the consequences of this triage process. With two notable exceptions[@Stelfox:2012iy; @Robert:2012ep], most studies of ward patients referred to critical care only follow those eventually admitted. Evaluations of decision making are therefore limited to early versus late admission, and suffer from survival and exclusion bias. They cannot report on patients who die without critical care admission, or survive despite initial refusal. Our work demonstrates that such patients are numerous. In fact in the first week, even after excluding patients with treatment limits, most deaths occur on the ward without critical care.

An additional strength has been our ability to link the precise time of the bedside assessment to the contemporaneous occupancy of the critical care unit. This has allowed us to observe the pressures on decision makers, and the effect that limiting resources has on patient pathways.

Finally, we have completed follow-up to one year. While mortality is substantial, it is predominantly an early problem thereby supporting the concept of early intervention. In particular, our observation that one in four patients with a treatment limitation orders survive a year without critical care suggests a need for humility before refusing early intervention on the grounds of long term prognosis.

Some limitations also deserve highlighting. Firstly, we have used real time data collection in order to capture assessments and decisions. Despite limiting the data request, not all hospitals managed to submit complete data at all times.We used the proportion of emergency ward admissions in the CMP successfully linked to the (SPOT)light database to define completeness. Reassuringly, we tested our findings by raising the threshold for judging data capture to 90% so that the median proportion of eligible admissions was 97%. We found no consistent difference in any result other than a fall in precision as the quality threshold increased, and the sample size inevitably fell.

The second weakness is structural rather than operational, and is that, in defining our population as those referred, we are blind to the process that leads to that referral. The literature typically names the ward monitoring process leading to referral as the afferent limb, and response of critical care to that referral as the efferent limb of a rapid response system. There is clearly both an unobserved period of deterioration prior to referral for patients in the study, and an unobserved population that might have been referred. Both these afferent components are as valid targets for intervention, as the observed efferent components of the pathway that we have discussed. However, observing the wider group of potential referrals would require a much briefer, narrower study.[@Simchen:2007bn]

## Conclusions and implications for practice and future research

Aspects of the study stand in spite of these limitations. Ward patients referred to critical care numerous, and vulnerable. The opportunity for intervention is brief, and there is important variation in practice between hospitals. The bedside assessment is an effective but imperfect triage tool, as the mortality in those initially declined admission is high. Given that we already excluded patients with treatment limitations, it is of concern that around half of these early deaths occur without a trial of critical care.

A substantial proportion of patients assessed and recommended for critical care are not offered a bed, and this proportion increases when capacity is limited. Expanding critical care bed numbers would first and foremost benefit this group. However, it might also create a virtuous circle. Earlier admission may lead to shorter stays thereby improving flow through critical care as well as outcomes. Identifying those patients who should be admitted promptly is already the top priority for both clinicians and patients.[@Reay:2014vm] What we have contributed we hope, is firm evidence in support of this.