Red-flag sepsis and SOFA identifies different patient population at risk of sepsis-related deaths on the general ward

Maja Kopczynska¹, Ben Sharif¹, Sian Cleaver¹, Naomi Spencer¹, Amit Kurani¹, Camilla Lee¹, Jessica Webster¹, Carys Durie¹, Jude Joseph-Gubral¹, Angelica Sharma¹, Lucy Allen¹, Billie Atkins¹, Alex Gordon¹, Llewelyn Jones¹, Amy Noble¹, Matthew Bradley¹, Henry Atkinson¹, Joy Inns¹, Harriet Penney¹, Carys Gilbert¹, Rebecca Walford¹, Louise Pike¹, Ross Edwards¹, Robyn Howcroft¹, Hazel Preston¹, Jennifer Gee¹, Nicholas Doyle¹, Charlotte Maden¹, Claire Frances MacLaughlin Smith¹, Nik Syakirah Nik Azis¹, Navritha Vadivale¹, Tamas Szakmany¹,² on behalf of the Welsh Digital Data Collection Platform Collaborators

¹Department of Anaesthesia, Intensive Care and Pain Medicine, Division of Population Medicine, Cardiff University, Heath Park Campus, Cardiff, CF14 4XN, UK

²Anaesthetic Directorate, Aneurin Bevan University Health Board, Royal Gwent Hospital, Cardiff Road, Newport, Gwent, NP20 2UB, UK

*Corresponding author:
Dr Tamas Szakmany MD, PhD, EDIC, DESA, FRCA, FFICM
Senior Lecturer in Intensive Care, Department of Anaesthesia, Intensive Care and Pain Medicine, Division of Population Medicine, Cardiff University, Heath Park Campus, Cardiff, CF14 4XN, UK

Tel: +442920723112
e-mail: szakmanyt1@cardiff.ac.uk
INTRODUCTION: Controversy exists regarding the best diagnostic and screening tool for sepsis outside the ICU. SOFA score has been shown to be superior to SIRS criteria, however the performance of “Red Flag sepsis criteria” has not been tested formally.

OBJECTIVES: The aim of the study was to investigate the ability of Red Flag sepsis criteria to identify the patients at high risk of sepsis-related death in comparison to SOFA based sepsis criteria. We also investigated the comparison of Red Flag sepsis to qSOFA, SIRS and NEWS scores and factors influencing patient mortality.

METHODS: Patients were recruited into a 24-hour point-prevalence study on the general wards and emergency departments across all Welsh acute hospitals. Inclusion criteria were: clinical suspicion of infection and NEWS 3 or above in-line with established escalation criteria in Wales. Data on Red Flag sepsis and SOFA criteria was collected together with qSOFA and SIRS scores and 90-day mortality.

RESULTS: 459 patients were recruited over a 24-hour period. 246 were positive for Red Flag sepsis, mortality 33.7% (83/246); 241 for SOFA based sepsis criteria, mortality 39.4% (95/241). 54 for qSOFA, mortality 57.4% (31/54), and 268 for SIRS, mortality 33.6% (90/268). 55 patients were not picked up by any criteria. We found that older age was associated with death with HR (95%CI) of 1.03 (1.02 - 1.04); higher frailty score 1.24 (1.11 - 1.40); DNA-CPR order 1.74 (1.14 - 2.65); ceiling of care 1.55 (1.02 - 2.33); and SOFA score of two and above 1.69 (1.16 - 2.47).

CONCLUSIONS: The different clinical tools captured different subsets of the at-risk population, with similar sensitivity. SOFA score 2 or above was independently associated with increased risk of death at 90 days. The sequelae of infection related organ dysfunction cannot be reliably captured based on routine clinical and physiological parameters alone. SOFA, but not ‘Red Flag Sepsis’ based clinical criteria identify patients with suspected infection at high risk of death on the general ward. SOFA score 2 or above is associated with increased risk of death at 90 days. The sequelae of infection related organ dysfunction cannot be reliably captured based on routine clinical and physiological parameters alone.
INTRODUCTION

Sepsis is defined as dysregulated host response to infection, resulting in acute organ dysfunction. This condition has been thoroughly studied in the Intensive Care Unit (ICU), however, data from the general ward and emergency department (ED) setting is sparse. In addition, it is increasingly recognised that sepsis is even more prevalent and may be associated with greater mortality burden on the general wards. Controversy in sepsis research exists also regarding the best diagnostic and screening tool for sepsis outside the ICU. Development of a reliable tool is crucial as this condition still represents a major cause of morbidity and mortality.

In addition to unclear sepsis prevalence and inaccurate identification tools there has also been considerable debate regarding validity of sepsis definition used. We previously reported the results of a point prevalence study of all Welsh centres using the 2001 International Consensus Criteria for sepsis (SEPSIS-1) as well as comparison between SEPSIS-1 and the 3rd International Consensus Definitions for sepsis (SEPSIS-3) utilizing electronic data collection and real-time data monitoring. We found that between four and five percent of hospitalized patients had sepsis. Strikingly, different sepsis criteria identified different patient populations with a different 30-day mortality risk. While Sequential Organ Failure Assessment (SOFA) tool was found to be superior to systemic inflammatory response syndrome (SIRS) both in patient identification and predication of mortality outcome, it requires complete blood test analysis, potentially creating delays in patient treatment and some or many elements unavailable in resource constricted environments. The simplified quick SOFA (qSOFA) which was suggested for a ward-based use, however, was only able to identify about 10% of the at-risk population.

Following the 2016 publication of NICE guidance NG51, the UK Sepsis Trust in communication with NICE launched a new screening tool based on NG51’s “high” and “moderate” risk categories, which they have termed “Red Flag” and “Amber Flag” criteria respectively. The premise is that by identifying the sickest patients with the highest mortality risk early, treatment can be delivered as soon as sepsis-related organ failures are recognised, potentially reducing the mortality and the morbidity from sepsis. The Red Flag sepsis is based on clinical features and aims at triggering Sepsis Six bundle while confirmatory blood results are pending. The performance of Red Flag sepsis criteria has not been tested formally and analysis of their performance compared to SIRS, SOFA and a well-established track and trigger tool, the National Early Warning Score (NEWS) is needed.

The primary objective of the study was to investigate the ability of Red Flag sepsis criteria to identify the patients at high risk of sepsis-related death in comparison to SOFA based sepsis criteria. Secondary
The objective was to compare the Red Flag sepsis to qSOFA, SIRS and NEWS scores and investigate patient characteristics to find factors influencing patient mortality.

METHODS

Study design and participants

This was a large multi-centre, point-prevalence study of at-risk populations of patients on general wards and ED. Fourteen hospitals in Wales, United Kingdom with 24/7 consultant-level Emergency Department supervision and the facility to admit and treat any acutely unwell patient participated in the study. The study covered one 24-hour period (0800 to 0759 hours the following day) on 18th October 2017, during which we screened all patients presenting to the ED and being cared for in wards outside intensive care, paediatrics and psychiatry units. Our previous study was performed on the 19th October 2016 which enabled us to analyse the validity of the study. The medical student data collectors recruited all patients with NEWS ≥3 in whom the treating clinical teams had a high clinical suspicion of an infection (documented as such in the medical or nursing notes). Patients were excluded if they were less than 18 years of age.

The details of the digital data collection platform developed for this study as well as description of the data collector training and performance during the study have been published previously. The data collected were obtained from medical and nursing records and included patients demographics and pre-admission characteristics, frailty (according to the Dalhousie Clinical Frailty Scale), physiological and laboratory data, presence of components of Red Flag, SOFA (where PaO2/FiO2 ratio was substituted with SpO2/FiO2 ratio as described by Pandharipande et al.) and SIRS sepsis criteria as well as input from the treating teams.

The complete list of variables is presented in Supplementary Table 1. We have collected data on the use of screening tools by the patients’ own team and the delivery of ‘Sepsis Six’ bundle. The results of any microbiology investigations (blood, urine, respiratory and wound cultures) and radiology imaging performed within the 48 hour period of the study were collected, together with data on antimicrobial prescription. Microbiology results and radiology results were used to determine if infection was present.

The amount of missing data was low and no assumptions were made for the missing data in line with similar recent critical care studies.

Patients were grouped by the following clinical criteria: qSOFA group: qSOFA 2 or above; SOFA group: SOFA 2 or above; SIRS group: presence of 2 or more SIRS criteria; Red Flag sepsis group: presence of one or more Red Flag sepsis criteria. Due to the composition of the different clinical tools, patients could be grouped into none or more than one group. Our primary outcome was 90-day all-cause mortality.
The project was approved by the South Wales Regional Ethics Committee (16/WA/0071) and patients or legal representatives gave written informed consent. To facilitate linkage to national databases for the collection of follow-up data, patient identifiable data was collected and entered on to the secure data collection tool. The Defining Sepsis on the Wards project was prospectively registered with an international trial registry (ISRCTN86502304).

**Statistical analysis**

Categorical variables are described as proportions. Continuous variables are described as median and interquartile range. We compared the distribution of clinical and biochemical variables between survivors and non-survivors using Mann-Whitney U test or Chi-square test, as appropriate. A two-tailed p-value <0.05 was considered statistically significant. To assess the performances of the sepsis criteria to predict the primary end point, we calculated diagnostic performances (sensitivity, specificity, negative and positive predictive values). We estimated the respective odds ratios (ORs) for the primary outcome within 90 days with a binary logistic regression with backward elimination model using mortality as a dependent variable. We determined goodness-of-fit of the model using the Hosmer-Lemeshow test. All statistical tests were calculated using SPSS 23.0 (SPSS Inc., Chicago, IL).
RESULTS

In our study we screened 7055 patients over the 24-hour study period in the 14 Welsh hospitals. Four hundred fifty-nine patients had NEWS $\geq 3$ and documented clinical suspicion of infection and were recruited in the study. Baseline characteristics are summarised in Table 1.

Different sepsis criteria identify different patient populations

Out of 459 patients with NEWS $\geq 3$ and high suspicion of infection, Red Flag criteria were present for 246 patients (53.6%), SOFA for 241 (52.5%), qSOFA for 54 (11.8%) and SIRS for 268 (58.4%). Some patients were identified by more than one criteria and 55 patients were not picked up by any criteria (Figure 1). 131 patients had clinically proven infection. 69/241 (28.3%) patients had SOFA 2 or above, 75/246 (30.8%) had Red Flag criteria, 50/268 (18.7%) had 2 or more SIRS criteria and 13/54 (24.1%) had qSOFA 2 or above. Out of the 55 patients, who didn’t fulfil any of the sepsis criteria, 21 had clinically proven infection.

Different sepsis criteria were associated with different patient mortality

We analysed the 90-day mortality for patients identified by each tool. Firstly, we calculated the mortality for each score. The mortality was highest for patients identified by qSOFA (31/54, 57.4%), followed by SOFA (95/241, 39.42%), then Red Flag Sepsis (83/246, 33.7%) and SIRS (90/268, 33.6%). We calculated the diagnostic performances of each sepsis tool to predict death at 90 days in Table 2.

We also analysed the mortality stratifying patients into groups of patients identified by one scoring criteria only, two scoring criteria and three criteria (Figure 2). The mortality was highest for “Red Flag + SOFA” group (19/37, 51.4%), followed by “Red Flag + SOFA + SIRS” (43/103, 41.8%), “SOFA only” (14/42, 33.3%), “SOFA + SIRS” (19/59, 32.2%), “SIRS only” (17/57, 29.8%), “SIRS + Red Flag” (11/49, 22.5%). Interestingly “Red Flag only” mortality was (10/57, 17.5%), which was lower even than the mortality of patients not picked up by any criteria (11/55, 20%). We performed a separate analysis of the mortality for the “qSOFA only” group - patients identified by qSOFA but not SIRS or Red Flag criteria, not taking the presence of SOFA criteria into consideration. This revealed that qSOFA identified only 4 patients with 100% mortality rate.

We investigated whether lower mortality of the screening tool is associated with earlier recognition and treatment of sepsis. The Sepsis Six completion was low for all the screening tools: Red Flag Sepsis 15.5% (38/256), SOFA 18.7% (45/241), qSOFA: 20.4% (11/54) and SIRS: 16.4% (44/268).
We also explored the rate of microbiology confirmed infection for each screening tool. We obtained information about the blood, sputum, urine, wound and CSF cultures for 405 out of 459 patients. There were 152 blood cultures, 39 sputum cultures, 133 urine cultures and 20 wound swab cultures performed. The rate of positive cultures was similar for each screening tool: Red Flag Sepsis 9.5% (21/222), SOFA 10.7% (21/196), qSOFA 8.2% (4/48), SIRS 10.1% (24/237).

Survival analysis
We also aimed at identifying potential factors that could have an impact on the survival. In addition to baseline patient characteristics presented in Table 1 we analysed patient clinical results, patient reserve (clinical frailty score and implemented limitations of care) and input from the treating team (Table 3).

Analysis of patient observations and laboratory results identified some parameters as statistically significant but the differences in the respective medians were too small to acknowledge the analytes as clinically significant (See Table 2, Supplemental Content, which shows patient observations and laboratory results).

We used a binary logistic regression model to independently assess variables that in a univariate analysis were associated with mortality and we felt were clinically important. All selected variables described both patient pre-admission characteristics and the screening tools most predictive of patient mortality. Consequently, we included age, frailty score, DNA-CPR, ceiling of care order and SOFA ≥2.

We found that older age was associated with death with OR (95%CI) of 1.03 (1.02 - 1.04); higher frailty score 1.24 (1.11 - 1.40); DNA-CPR order 1.74 (1.14 - 2.65); ceiling of care 1.55 (1.02 - 2.33); and SOFA score of two and above 1.69 (1.16 - 2.47). The result of the Hosmer-Lemeshow test indicated good fit of the model.

DISCUSSION
To our knowledge this is the first prospective study comparing the diagnostic performance of “Red Flag sepsis criteria” to SEPSIS-1 and SEPSIS-3. We found that SEPSIS-1 criteria identified the most patients in at-risk population (58.4%), followed by Red Flag (53.6%) and SEPSIS-3 (52.5%). There was a significant overlap between the criteria, although 55 patients (12%) with high NEWS scores, clinical suspicion of infection and 20% 90-day mortality were missed even after application of all scoring criteria, similarly to previous studies. Diagnostic performances of different combinations of sepsis screening tools revealed once again the low sensitivity and high specificity of qSOFA.
Red Flag sepsis criteria were developed to aid healthcare providers in the ED and on the general wards, as well as in the community to identify patients at high risk of deterioration. Notably, this was done without a thorough analysis of a robust clinical dataset, similar to the SEPSIS-1 SIRS criteria and in stark contrast to the SOFA score used in the SEPSIS-3 definition.1,12 Following the publication of the NICE guidance, many organisations in the UK started to use the Red Flag sepsis tool as their screening for patients who might have organ dysfunction due to infection and are high risk of death.13 Our results cast some doubt on the potential effectiveness of this approach. SOFA score 2 or above either on its own or in conjunction with other tools was consistently associated with high risk of death. Red Flag Sepsis criteria failed to identify almost half of this population, signalling serious deficit in its clinical utility to deliver the intended goal of using it as a sensitive tool to identify patients at high risk of deterioration. Furthermore, out of the four different clinical tools, only SOFA score 2 or above was independently associated with increased risk of death at 90 days. We argue, that based on our results, the Red Flag Sepsis tool should not be used on its own as the primary triggering tool for Sepsis Six bundle as it can potentially miss up to 45% of patients who are at high risk of death following an infectious episode and also it is not independently associated with adverse outcome. Patients with Red Flag signs didn’t have more formal sepsis screening or more reliable Sepsis 6 bundle delivery, nor had more microbiologically confirmed infections compared to patients identified by other clinical tools. Although it can be argued that the currently operational sepsis screening in Wales are based on the original SEPSIS-1 definition and use the SIRS criteria, our data suggest, that regardless of the clinical criteria used, patients at high risk of deterioration are not reliably screened. It also appears that despite clear warning signs present, their treatment is not universal, although appears to be improved compared to previous years.10,11

Investigating diagnostic performance of the different clinical tools is crucial as failure to recognize sepsis could lead to excess deaths but conversely, over-triage of suspected sepsis is likely to burden general wards and ED and risks detracting from care provided to other patients. As there is still no gold standard diagnostic test, the clinicians are required to interpret a constellation of nonspecific physiological and laboratory abnormalities among patients with suspected or definite infection.20 Regardless of the clinical tool used, either based on established organ failure scores or more intuitive clinical categories, less than a third of the patients who fall in to these categories had clinically proven infection. This could be partly explained by the relatively low rate of microbiological sampling and radiological examinations. Importantly, patients who didn’t score on any of the clinical tools, had similar rate of clinically proven infection. These results emphasise the need for more sensitive screening tools which could highlight the patients at true risk of deterioration secondary to an infectious insult. In our study, however, none of those individual physiology parameters were clinically significant in predicting patient mortality. Interestingly, a recent multicentre study was able to accurately utilise an algorithm incorporated to an electronic health record system, to
predict the development of sepsis in multiple clinical environments.\textsuperscript{21} However, the true heterogeneity of sepsis which can present opportunities to intervene earlier to prevent the sequelae can only probably be captured with sophisticated analysis of the host response on a molecular level.\textsuperscript{22,23} Creating a tool which could be useful for identification of sepsis and prediction of mortality due to this condition has important implications for clinical care, epidemiologic and clinical studies, public health surveillance, and quality improvement programs.\textsuperscript{24}

Analysis of the survival revealed that completion of a screening tool and Sepsis Six bundle is associated with improved outcome, in line with previous findings, but the numbers are too small for a definitive comparison\textsuperscript{25,26} It is apparent that pre-admission trajectories are the most clinically significant predictors of patient survival, similar to recent findings in the critically ill population.\textsuperscript{27} In the ward and ED setting, older age and decreased patient reserve and expectation of the treating physicians (higher frailty score, higher rate of DNA-CPR and higher rate of ceiling of care implementation) strongly predicted patient mortality.

The strengths of our study include wide participation of centres and prospectively collected patient information. Our study has high internal validity as our previous two studies applied the same methodology and recruited similar number of patients in the same centres.\textsuperscript{10,11} Moreover, to our knowledge this is the first study to investigate the diagnostic performance of Red Flag sepsis criteria and their real-life utility in clinical practice.

Our study has some limitations. Firstly, our dataset could have missed some determinants of sepsis due to the necessity to be small enough to maintain data collector participation and data reliability. Secondly, we recruited patients who had NEWS score of 3 and above, meaning that we could also have missed patients with sepsis who had a lower score.\textsuperscript{6,28} However, recent data suggest that NEWS cut-off of 3 may be the best trigger to screen patients for sepsis in the ED.\textsuperscript{29} Score of 3 is also recommended as an escalation trigger by NICE and used in the Sepsis Trust’s Red Flag Sepsis pathways\textsuperscript{12}.

CONCLUSION

The different clinical tools captured different subsets of the at-risk population, with similar sensitivity and with similar precision for confirmed infection. SOFA, but not ‘Red Flag Sepsis’ or SIRS based clinical criteria identify patients with suspected infection at high risk of deaths on the general ward. SOFA score 2 or above is independently associated with increased risk of death at 90 days. The sequelae of infection related organ dysfunction cannot be reliably captured based on routine clinical and physiological parameters alone.
ACKNOWLEDGEMENTS

REFERENCES


5. Nafsi T, Russell R, Reid CM, Rizvi SMM. Audit of deaths less than a week after admission through an emergency department: how accurate was the ED diagnosis and were any deaths preventable? *Emergency Medicine Journal : EMJ*. 2007;24(10):691-695.


Figure 1. Patients identified using different scoring criteria. ● Red Flag sepsis; ● SOFA, Sequential Organ Failure Assessment score; ● SIRS, systemic inflammatory response syndrome criteria. SEPSIS-1 is defined by SIRS ≥2. SEPSIS-3 is defined by SOFA ≥2; Red Flag is defined by ≥1 Red Flag criteria. qSOFA was omitted in the diagram as patients identified by this criteria were also captured by SOFA score.

Figure 2. Patient mortality depending on the scoring criteria. RF, Red Flag sepsis; SOFA, Sequential Organ Failure Assessment score; SIRS, systemic inflammatory response syndrome criteria.

Supplementary Table 1. Variables collected during the study day. COPD, Chronic Obstructive Pulmonary Disease; BP, blood pressure; GCS, Glasgow Coma Scale; WCC, white cell count; CRP, c-reactive protein. INR, international normalized ratio; APTT, activated partial thromboplastin time; CSF, cerebrospinal fluid; DNA-CPR, do not attempt cardiopulmonary resuscitation.

Supplementary Table 2. Patient observations and laboratory results. BP, blood pressure; GCS, Glasgow Coma Scale; WCC, white cell count; CRP, c-reactive protein. Values are median (range). P-value of less than 0.05 is bold and underlined.