

# **Title: “Death is not the answer” The challenge of measuring the impact of early warning systems**

Susan M Chapman,<sup>1,2</sup> Jo Wray,<sup>3</sup> Kate Oulton,<sup>3</sup> and Mark J Peters<sup>2,4</sup>

<sup>1</sup>*International and Private Patients Division, Great Ormond Street Hospital for Children NHS Foundation Trust, London WC1N 3JH*

<sup>2</sup>*UCL Great Ormond Street Institute of Child Health, London (UK) WC1N 1EH*

<sup>3</sup>*Centre for Outcomes and Experience Research in Children's Health, Illness and Disability (ORCHID). Great Ormond Street Hospital for Children NHS Foundation Trust, London (UK) WC1N 3JH*

<sup>4</sup>*PICU, Great Ormond Street Hospital for Children NHS Foundation Trust, London (UK) WC1N 3JH*

Corresponding author: Susan Chapman

Corresponding author address:

International and Private Patients Division, Great Ormond Street Hospital for Children, London, UK,  
WC1N 3JH

[E: Sue.Chapman@gosh.nhs.uk](mailto:Sue.Chapman@gosh.nhs.uk)

**T +44 207 405 9200**

**Word count:** 1177

**Key words:** early warning scores, deteriorating children, methodology, outcomes

We can all remember individual children in whom a deterioration went unrecognised. Sometimes fatally. Our defences were little more than the pearls offered by senior colleagues of grave warning signs: 'beware grunting in an infant' or 'watch out for a tachycardia after the temperature has fallen'. But this advice was unstructured, and children are so different, and their co-morbidities so broad we failed some of them. Paediatric Early Warning Systems (PEWS) are serious attempts to reduce the unacceptable and dangerous variability in this recognition and response process. Scoring systems should provide age-appropriate thresholds for concern for single parameters or aggregated abnormal physiology and prompt standardised responses. The idea has such natural appeal that PEWS use was soon advocated by a number of national bodies [1,2] without evidence. This may have been a mistake. Many of the scores in widespread use were not calibrated nor validated. When formally assessed, most had poor predictive performance.[3] This is not a trivial problem because staff may choose not to raise an alarm in the absence of a raised score or may choose to ignore a score 'because it never works for him/her'.

Other than optimism, the main reason for the lack of evidence was the low event rates of critical deterioration or death within individual centres. An adequately powered trial was therefore a huge challenge. Fortunately, Parshuram and colleagues took on this challenge with the study Effect of a PEWS on All-Cause Mortality in Hospitalized Pediatric Patients – EPOCH.[4] This trial has many strengths. It assessed a strong candidate for a score: the BedsidePEWS. The BedsidePEWS is notable for having significant prior validation.[5,6] It was one of the highest performing of 18 scores in head to head comparisons.[3] Further, it addressed a relevant and very large population: 144,539 patients in 21 hospitals in seven countries. The intervention was well designed and formally implemented. The validated severity of illness score was implemented alongside an education program with inter-professionally designed documentation and structured escalation/de-escalation care recommendations. The cluster-randomised design was efficient and the only feasible approach to avoid contamination. The outcome measures were relevant to patients and families: the primary being all-cause hospital mortality. Secondary measures included occurrence of a significant clinical deterioration event, a composite measure felt to reflect late admission to intensive care (Table 1). Ten hospitals were randomized to the BedsidePEWS intervention. The remaining 11 hospitals did not have a PEWS, although four did have a rapid response team in situ.

Table 1: Significant Clinical Deterioration Event component outcomes

Within the 12 hours before intensive care admission	Within first hour after intensive care admission
Death	Death
Cardiopulmonary resuscitation	Cardiopulmonary resuscitation
Tracheal intubation	Tracheal intubation
Administration of vasoactive medication	Extracorporeal membrane oxygenation
Fluid boluses of 60 mL/kg or greater	

Occurrence of one or more component outcomes would be considered a significant clinical deterioration event

Despite all these strengths, EPOCH was a ‘negative trial’. Implementing the BedsidePEWS did not significantly decrease all-cause mortality. Of the 22 secondary outcome measures the only positive finding was a reduction in significant clinical deterioration events (relative risk 0.77 (0.61 to 0.97),  $p=0.03$ ).

So what is to be made of this? Should governments and organisations no longer recommend PEWS? Should hospitals consider removing a PEWS which is already in place or not go forward with implementing a PEWS in the future?

If we look a little more closely at the main outcome measure of all-cause hospital mortality, we can see that this presents some challenges, particularly as death in childhood remains, thankfully, a relatively rare event. The power calculations were based on 2007-2009 data with baseline all-cause hospital mortality of 5.1 deaths per 1000 hospital discharges and anticipated a reduction of 1/1000 hospital discharges. However there has been a general reduction in mortality over time [7] which invalidated these power calculations. The trial observed fewer than 2 deaths per 1000 hospital discharges in both groups and roughly half of these deaths were in the context of ‘do not attempt resuscitation’ orders (DNAR). Exclusion of children with a DNAR did not alter the findings. The study was therefore underpowered despite the 21 hospitals and 144,539 patients. Assuming mortality fell no further, to detect a 10% relative risk reduction (from 2/1000 to 1.8/1000) with 90% power of in-hospital deaths in the absence of a DNAR would require a study of >2 million *individually randomised* paediatric admissions. And individual randomisation is not feasible for an intervention such as this. The necessary cluster randomisation would require many hospitals and a significantly higher number of patients.[8] Also, children die for different reasons than they did 10-15 years ago.[9] Severe or multiple co-morbidities are the dominant causes of in-hospital death and the contribution from

severely deranged acute physiology is less than it ever was.[10] So, a potential effect size of 10% may be unrealistic. Therefore the required sample size increases into many millions.

Whilst no-one would argue that mortality is not important, is it the most appropriate outcome to assess PEWS efficacy? Some cases remain resistant to intervention and death will occur regardless of whether a PEWS is truly effective or not. If mortality is not the optimal outcome, what are the alternatives? The EPOCH team examined potentially preventable cardiac arrests. This has significant appeal but has been challenged as subjective. Initial efforts were hampered by low levels of reviewer agreement, prompting amendments to the previous agreed protocol.

Ensuring rapid intensive care access for those who would benefit from it may be a better marker, given the purpose of PEWS. Within EPOCH the authors used significant clinical deterioration events as a marker of late transfers but closer examination reveals components associated with critical illness that is well established. The thresholds may be inappropriately high and children deemed to be a 'timely' transfer may in fact have been inappropriately classified. This may have masked the true effect of PEWS. Some of these critical events, such as cardiopulmonary resuscitation, tracheal intubation and death occurring outside or immediately on arrival to intensive care may indicate a lost opportunity for preventative action.

As well as challenges relating to the appropriateness of the outcome measures, the EPOCH authors also faced the complexities of conducting a randomised trial in a real-world setting, where organisational culture and human interaction are pivotal. The successful deployment of a PEWS is dependent upon the complex interplay between multiple factors such as leadership, culture, teamwork, nurse and family empowerment, safe staffing levels, effective communication and continuity of care [11], which take time and effort to be embedded in practice. Data collection for the EPOCH trial may not have continued long enough, following the introduction of PEWS, for the long-term benefits (or harms) to be identified.

Whilst detecting and managing deterioration in children is complex, the need for systems to deal with avoidable deterioration is not disputed. For this we need to agree robust, valid and clinically meaningful outcomes to evaluate patient safety efforts, not matter how challenging that might be. The BedsidePEWS is better validated than other systems, performs very well in head to head comparisons and with EPOCH has clear 1A evidence that it is not harmful. Determining a mortality effect may have been unrealistic given the many steps between deterioration and death, but we should take serious note of the significant critical deterioration signal. The EPOCH trial remains a

valuable study and no other PEWS has anything approaching this level of evidence to support it. In evaluating PEWS, death is not the answer.

## References:

- 1 National Confidential Enquiry into Patient Outcome and Death. Are we there yet? A review of organisational and clinical aspects of children's surgery. <http://www.ncepod.org.uk/reportdownloads/SICfullreport.pdf>. 2011. doi:10.1016/j.ijvsu.2013.06.644
- 2 Pearson GA. Why Children Die: A Pilot Study. London: : CEMACH 2008.
- 3 Chapman SM, Wray J, Oulton K, *et al.* "The Score Matters": wide variations in predictive performance of 18 paediatric track and trigger systems. *Arch Dis Child* 2017;**102**:487–95. doi:10.1136/archdischild-2016-311088
- 4 Parshuram CS, Dryden-Palmer K, Farrell C, *et al.* Effect of a Pediatric Early Warning System on All-Cause Mortality in Hospitalized Pediatric Patients: The EPOCH Randomized Clinical Trial. *JAMA: The Journal of the American Medical Association* 2018;**319**:1002–12. doi:10.1001/jama.2018.0948
- 5 Parshuram CS, Hutchison J, Middaugh K. Development and initial validation of the Bedside Paediatric Early Warning System score. *Crit Care* 2009;**13**:R135. doi:10.1186/cc7998
- 6 Parshuram CS, Duncan HP, Joffe AR, *et al.* Multicentre validation of the bedside paediatric early warning system score: a severity of illness score to detect evolving critical illness in hospitalised children. *Crit Care* 2011;**15**:R184.
- 7 Joffe AR, Anton NR, Burkholder SC. Reduction in hospital mortality over time in a hospital without a pediatric medical emergency team: limitations of before-and-after study designs. *Arch Pediatr Adolesc Med* 2011;**165**:419–23.
- 8 Rutterford C, Copas A, Eldridge S. Methods for sample size determination in cluster randomized trials. *Int J Epidemiol* 2015;**44**:1051–67. doi:10.1093/ije/dyv113
- 9 Pearson GA, Ward-Platt M, Kelly D. How children die: classifying child deaths. *Arch Dis Child* 2011;**96**:922–6. doi:10.1136/adc.2009.177089
- 10 Namachivayam P, Shann F, Shekerdemian L, *et al.* Three decades of pediatric intensive care: Who was admitted, what happened in intensive care, and what happened afterward. *Pediatr Crit Care Med* 2010;**11**:549–55. doi:10.1097/PCC.0b013e3181ce7427
- 11 Royal College of Paediatrics and Child Health, NHS Improvement. A safe system for recognising and responding to children at risk of deterioration. 2016.<http://www.rcpch.ac.uk/safer-system-children-risk-deterioration> (accessed 19 Jun2016).