**RAND EUROPE**

# Paediatric Early Warning Systems: a scoping study

Lessons from a rapid review

Sarah Ball, Sarah Parkinson and Sonja Marjanovic*

*Senior and corresponding author

# THIS.Institute

The Healthcare
Improvement
Studies Institute

For more information on this publication, visit www.rand.org/t/RRA1105-1

RAND Europe is a not-for-profit research organisation that helps to improve
policy and decision making through research and analysis. RAND's publications
do not necessarily reflect the opinions of its research clients and sponsors.

**Support RAND**
Make a tax-deductible charitable contribution at
www.rand.org/giving/contribute

www.rand.org

www.randeurope.org

# Summary

## Context and aims

RAND Europe was asked by The Healthcare Improvement Studies (THIS) Institute at Cambridge University to conduct a scoping study to identify and synthesise evidence relating to the use of paediatric early warning (PEW) scores and systems in the United Kingdom (UK). RAND Europe was also asked, to the extent possible, to note any insights related to issues of standardisation of scores and systems in use and the de-implementation of old practices. The insights from this scoping study are meant to inform potential future research into the process of de-implementing existing practices and services in healthcare using learning from the efforts of NHS England, NHS Improvement and the Royal College of Paediatrics and Child Health to develop and introduce a standardised PEW system in England as a key empirical case.

Early warning (EW) systems are used in hospitals and other healthcare settings to help healthcare staff identify early signs of clinical deterioration in patients and facilitate timely intervention. They take the form of pre-specified alert criteria intended to trigger additional care when needed, monitored through observation charts. In addition to their longstanding use in adult populations, these systems are increasingly used in paediatric settings. The focus of this report is specifically on *paediatric* EW scores and systems.

## Methods

The insights presented in this report are based primarily on a literature review, complemented with three interviews with five experts on PEW scores and systems in the UK (one interview conducted with three experts simultaneously). The literature review focused on articles covering the use of PEW scores and systems in UK settings. However, review articles including UK studies alongside studies in other country settings were also included. Overall, we analysed 19 journal articles and five documents relating to the development of a standardised PEW system in England (the latter were provided to the study team by NHS England and NHS Improvement). Further detail on the methods used and on associated caveats is provided in the full report and Annex A.

## Key insights

### The nature of PEW scores and systems:

- The terms *PEW systems*, *paediatric track-and-trigger systems*, and *paediatric alert criteria* are used relatively interchangeably in the literature. In general, PEW systems comprise PEW *scores, response algorithms and criteria* under which a response or set of responses should be initiated, as well as *wider governance and management arrangements*, including the procedures and processes that support the monitoring

and scoring of deterioration and the responses they are intended to initiate. *Responses* occur once deterioration is detected. These elements operate within a wider social, organisational and cultural context, spanning safety cultures, leadership, training, relationships, organisational management, and governance structures and practices.

- PEW scores provide a formal framework for evaluating various parameters and observational data to identify early deterioration in paediatric patients and produce a single descriptor of the degree of deterioration. Observations are recorded systematically in PEW charts. The term PEW *score* refers both to the scoring mechanism used (i.e. the parameters included and associated scoring thresholds) and the overall score (value) assigned to a given patient. A wide variety of parameters can be included in a PEW score. The range of potential values and their associated scores for any given parameter can also vary between different PEW scores and systems. Common parameters used in the UK include heart rate, respiratory rate, oxygen saturation, abnormal consciousness, respiratory effort, nursing concern, systolic blood pressure and oxygen therapy. Some PEW scores also have different scoring systems based on a child's age.

- PEW systems can be score-based or trigger-based. In score-based PEW systems, an overall score is calculated by summing up the scores for individual parameters. A response or escalation algorithm indicates the particular action to be taken by the healthcare professional for a given score. Responses are graduated depending on the value of a PEW score. Children at increased risk of deterioration or with more severe clinical symptoms

receive a different clinical response compared to children with less serious deterioration, and are escalated to staff with specific required skills and expertise. By contrast, trigger-based PEW systems do not calculate an overall score. Instead, they are based on reaching a particular threshold in one or more parameters, at which time a specific (non-graduated) response is triggered. Response algorithms or alert criteria are used to determine the circumstances under which a response is initiated, i.e. the thresholds that must be met to trigger a response. Variables such as the pre-specified responses, who responds, who observes and how frequently, may be set out as part of the PEW system.

- Healthcare staff score patients according to a regular schedule. If a score threshold is reached, a review may be required from more senior and specialised staff than those doing the regular observations to prevent further deterioration. In addition to offering a validated measure of patients' current clinical condition in the form of a score, PEW systems are also intended to provide a rapid means of communication across healthcare teams and ensure appropriate prioritisation and focus on patients. Most PEW systems are currently paper-based, but the potential for electronic and automated observations is of increasing interest to policy and practice.

## The diversity of PEW scores and systems in use in the UK:

- PEW scores and systems are widely used in acute paediatric settings in the UK, with various scores and systems used in different clinical settings. They are commonly used in paediatric inpatient settings, including those in children's hospitals, oncology units and Paediatric

Intensive Care Units (PICU), as well as district hospitals, specialist hospitals and general hospital wards with paediatric patients. PEW scores and systems have also been introduced in some emergency departments. While Northern Ireland and Scotland have standardised national PEW systems, England and Wales do not. Examples of PEW scores used in the UK and frequently discussed in the literature include the Bedside PEW system, the Bristol PEW system tool, the Cardiff and Wales PEW system, the MET (Medical Emergency Team) activation criteria, the NHS Improvement PEW system, the PEW Score, the PEW System Score, the Brighton PEW score, the Melbourne Activation Criteria (MAC) score, the Paediatric Observation Priority Score (POPS), and the Paediatric Advanced Warning Score (PAWS). For more information on the details of these specific PEW scores and systems, please see Annex C.

- PEW scores and systems can be introduced for a range of purposes, which may vary between different clinical settings (e.g. inpatient, emergency department). In addition to their core utility for identifying children at risk of deterioration and supporting appropriate responses, PEW systems can also help support patient triage or assess risk for specific conditions such as sepsis.

- PEW scores and systems in use in the UK vary according to the number and types of parameters included and how they are scored. The most complex system identified is the Paediatric Early Warning System Score I, which contains 19 parameters. Other PEW scores contain anywhere from 5-14 parameters. There is also variety in the extent to which the scores are age-dependent and the age bands.

## Do PEW scores and systems make a difference?

- There is a lack of evidence that PEW scores and systems are associated with improved patient outcomes (e.g. reduced mortality rates or reduced rates of cardiopulmonary arrest), in the absence of additional accompanying interventions such as rapid response teams (RRTs). However, the existing evidence consists primarily of studies of limited quality and hence needs strengthening.

- There is some evidence for PEW scores and systems' benefits in areas other than clinical outcomes, such as improvements in situational awareness, staff empowerment, communication and ability to recognise deterioration.

- However, there is also some evidence that the use of PEW scores and systems can have unintended consequences related to unnecessarily-increased staff workloads (and hence reluctance to use the PEW score), and potentially to patient outcomes too (i.e. risk of over or undertreatment).

- It is unclear whether any single PEW score or system is better than others in terms of identifying deteriorating patients and avoiding negative outcomes; most have not been validated or rigorously evaluated

## What influences the implementation and use of PEW scores and systems?

Despite the lack of strong evidence regarding the relative merits of different PEW scores and systems, there is some information in the literature about features that make PEW scores and systems more or less difficult to implement and use, and about features that influence their effectiveness in identifying deteriorating patients and escalating care appropriately. Key influences on implementation that relate to the PEW score

and system (as opposed to its wider social, cultural and organisational factors) include:

**Whether the PEW system is score-based or trigger-based**: Some evidence suggests that score-based systems that allow for graduated responses are better in identifying deterioration and triggering appropriate responses. However, they are more complicated to implement and use than trigger-based ones.

**Complexity and ease of use, including the system's visibility on healthcare providers' wards:** More complex mechanisms are more prone to error, and come with higher staff time demands. User-friendly formats for recording and displaying data (e.g. using graphical and numerical displays or coded traffic light systems) affect how easily score values can be interpreted. Increasing the visibility of PEW scores and systems on wards (e.g. laminated charts at nursing and computer stations) can help facilitate their use.

**How observations are recorded and escalation requirements communicated:** Implementation of paper-based systems is considered to be more straightforward, less resource-intensive, and subject to fewer delays than digital systems, given the digital infrastructure in the NHS that is available at present. However, there is some evidence that electronic-based PEW scores and systems could be more accurate than paper-based ones at detecting deterioration and channeling appropriate escalation. This evidence is limited, however, and there are also concerns about higher false-positive rates – with automated alarm systems resulting in alarm fatigue. The use of digital systems could also facilitate data collection to improve scores and enable

monitoring and evaluation, ultimately helping to improve their effectiveness.

**The predictive validity of the PEW score:** The degree to which the PEW score or system can distinguish between deteriorating and non-deteriorating patients influences its effectiveness. There is a need to balance sensitivity and specificity.

**The degree to which the score integrates subjective criteria and accommodates scope for clinical judgement:** The degree to which the definition of a normal value for an objectively measured parameter (such as heart rate) is specified or open to a clinician's judgement influences the reliability of the PEW score. There are diverse opinions about whether subjective criteria relating to clinician/nurse concern should be included in PEW scores. The value of incorporating parent, family or carer concern is also debated. The evidence base on these issues is scarce and inconclusive. However, there is a widely-acknowledged need to ensure that PEW scores and systems do not override clinical judgement.
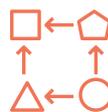
**The ability to adapt a particular PEW score and system to local contexts during implementation:** Such adaptiveness is important in selecting which PEW score and system to implement in a particular setting, given the different patient populations, illness cohorts and clinical settings.

Influences related to the wider social, cultural and organisational context in which PEW systems are deployed have also been identified as influencing implementation and use. These include:

**Staff engagement, teamwork and leadership:** Early engagement

of staff across professional groups and hierarchies can increase buy-in and build a supportive safety culture to aid local implementation and effective use of a PEW score and system. However, a lack of clear lines of accountability, issues of hierarchy, and inappropriate communication channels between professions can also pose challenges to effective implementation and use.

**Levels of clinical competence and situational awareness, and investment into their development through training and education:** Appropriate training and supervision are essential to ensure key skills and competencies can be demonstrated. Supporting communication and teamwork throughout the implementation process, and empowering patients, carers and families as well as staff to escalate concerns, can contribute to improved situational awareness.
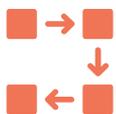
**A range of standardised processes, roles and structures to support staff situational awareness:** For example, huddles and 'watch-stander roles' (i.e. a nurse or clinician who is responsible for knowing which patients are at high risk of deterioration) to proactively plan for and identify risks; structures to support continuity of care and communication during handovers (e.g. checklists and cognitive aids); and structures to support adequate workload and staffing (e.g. adequate staff-to-patient ratios).

**Specific measures introduced to support the response to identified clinical deterioration:** For example, PEW systems incorporating RRTs (i.e. teams of specialist personnel to respond rapidly to critical illness) may be more effective than those without (although the evidence is mixed).

**Resource availability (staff, funding and infrastructure):** The availability and appropriate balance of staff influences how PEW scores and systems are implemented (e.g. management staff versus senior nursing staff on each shift). Other important resources include PICU beds, IT infrastructure and equipment.

**Evaluation and learning:** PEW scores and systems must be subject to ongoing evaluation to ensure continual improvements in their ability to identify deteriorations in children and ensure a timely and accurate response. Evaluation of PEW scores and systems is currently insufficient, especially as complex interventions, and relevant process and outcome measures need to be identified.

## Insights related to standardisation

Standardised PEW systems have been introduced in Northern Ireland and Scotland but are not yet in place in England or Wales. Work to develop a national PEW system for England is in progress. To date, efforts in Northern Ireland and Scotland have focused on implementation in the acute care setting. In Scotland, however, the national PEW system is being trialled by the Scottish Ambulance Service and by general practitioners (GPs) in certain situations. Similarly, plans for developing a standardised PEW system in England will initially focus on hospital settings before being adapted for use in primary care and community settings.

Key drivers of standardisation of PEW scores and systems (i.e. why decisions to standardise are pursued) can include:

- A strong desire for standardisation from within the paediatric community, in part due to dissatisfaction with existing approaches and concern about the ability

to use different systems effectively when healthcare staff and patients move between hospitals.

- The potential to optimise resource use (avoiding the duplication of time, effort and resources necessary if each healthcare setting designs its own system) and to share learning about effective practice between different settings.

- A desire to overcome the challenges of evaluating PEW systems associated with the large variety of systems in use) and maximise learning opportunities about what works and how.

## We also identified key influences on the evolution of standardisation processes:

- Key influences broadly revolve around three factors: the ability to secure early buy-in from stakeholders by engaging people across different professional groups and ensuring open lines of communication; the ability to sustain engagement over time; and the ability to ensure the standardised system is adaptable to local implementation settings to some degree.

- The lack of compelling evidence on the impact of PEW systems on patient mortality and morbidity can be a barrier to securing buy-in, and decision-makers have needed to frame standardisation efforts around the benefits it could bring to improving organisational processes, e.g. communication and system-related difficulties in identifying deteriorating children.

- De-implementation of incumbent PEW systems is also a particular challenge. Based on evidence from the interviews we conducted, it may be easier to secure buy-in for standardisation in contexts

where there is no predecessor system (i.e. where a PEW system is being introduced from scratch) or where an existing system is not highly valued or acceptable to users (and where there is thus a greater openness to de-implementation). However, when locally-developed PEW systems designed to benefit a specific cohort of children are already in place, de-implementation-related challenges may be accentuated.

- Despite the challenges, a range of engagement activities can support buy-in for standardisation. These include national engagement events; targeted communication plans to encourage engagement across a geographical region the development and nurturing of professional networks to engage with the standardisation processes and as a way of enabling regular communication and feedback on planning and progress; education about the need for standardisation; and the use of local champions to pilot and stage the roll-out of a standardised system. Continuing dialogue is needed to sustain engagement and support effective implementation and use of a standardised system through time.

- As with the implementation of PEW scores and systems in general (i.e. regardless of whether they are standardised or not), there is a need to consider the adaptability of the standardised system to different adoption settings (which may differ with respect to available resources, infrastructure or staffing models and the patient populations they serve). What is considered a normal threshold for a patient's parameter value can also vary between settings (such as between a critical care ward and primary care).

The limited evidence on the impacts of standardising PEW scores and systems needs to be borne in mind when contemplating an approach to the roll-out of a standardised system:

- The evidence base on the benefits of standardisation of PEW scores and systems is scarce and inconclusive, with only limited anecdotal evidence of the impact of PEW system standardisation on patient outcomes.

- There are some reports (by interviewees) of observed improvements to processes for identifying deteriorating children, such as reductions in unnecessary data collection, a better interface between patient safety teams and people at the front line, a reduced need for staff to relearn new processes and procedures when they move between different healthcare provider organisations, and improved communication between healthcare professionals.

- Better evaluative evidence could test the case for standardisation. It would require evaluations that capture learning about both processes and outcomes, with evidence gathered from the experience of implementing standardisation in diverse contexts.

# Table of contents

# Acknowledgements

# Abbreviations

| | |
|---|---|
| ED | Emergency department |
| EW | Early warning |
| GP | General practitioner |
| ICU | Intensive Care Unit |
| IT | Information technology |
| NHS | National Health Service |
| NICE | National Institute for Clinical Excellence |
| PEW | Paediatric early warning |
| PICU | Paediatric Intensive Care Unit |
| POPS | Paediatric Observation Priority Score |
| RCPCH | Royal College of Paediatrics and Child Health |
| RRT | Rapid response team |
| TPR | Temperature, pulse and respiration |
| UK | United Kingdom |

# 1 Background and aims of scoping exercise

## 1.1. The challenge of detecting deterioration among children and enabling timely intervention

Clinical deterioration in children often follows a period of physiological instability, which can be detected by monitoring vital signs and other clinical observations. When such changes are recognised, timely intervention can prevent adverse events such as cardiac arrest and death. However, early warning signs can be overlooked by healthcare staff or may fail to trigger an appropriate response [1, 2]. Paediatric early warning (PEW) scores and systems are a mechanism to help healthcare staff identify children at risk of deterioration, and prompt increased monitoring and escalation to staff with appropriate skills in emergency and critical care [1, 3]. PEW systems are comprised of multiple components. These include PEW scores (generally a composite of multiple vital signs and other clinical indicators, monitored through observation charts), response algorithms and criteria under which a response or set of responses should be initiated, and wider governance and management arrangements [4].

Early warning (EW) systems are widely used in adult populations [5], but while their use in paediatric settings is increasing – and although PEW scores and systems appear to be used widely in acute paediatric settings – they are less well established across the entire healthcare system. In England, a National Early Warning Score (NEWS) for use in adult populations was developed by NHS England and the Royal College of Physicians in 2012 and updated (as NEWS2) in 2017, aiming to achieve 100 per cent adoption by March 2019 [6]. At a national level, standardisation of EW systems across trusts has been seen as crucial in order to ensure a common understanding of EW systems across healthcare staff and minimise the risk of misinterpretation when staff move between organisations [6]. However, while standardised EW systems for use in paediatric populations have been introduced in some regions of the United Kingdom (UK) – namely in Northern Ireland and Scotland – there is no standardised PEW system in place in England. NHS England and NHS Improvement and the Royal College of Paediatrics and Child Health (RCPCH) began work to develop such a system in 2018, and this work is ongoing. Given the diversity of PEW systems that currently exist [7-9], any roll-out of a standardised PEW system is likely to require the de-implementation of scores and systems that are already in use.

## 1.2. Aims of scoping exercise

RAND Europe was asked by The Healthcare Improvement Studies (THIS) Institute at Cambridge University to conduct a scoping

study to identify and synthesise evidence relating to the use of PEW scores and systems in the UK. RAND Europe was also asked, to the extent possible, to note any insights related to issues of standardisation of scores and systems in use and the de-implementation of old practices. The insights from this scoping study are meant to inform potential future research into the process of de-implementing existing practices and services in healthcare, using learning from the efforts of NHS England and NHS Improvement and the Royal College of Paediatrics and Child Health to develop and introduce a standardised PEW system in England as a key empirical case.

RAND Europe was asked to examine a sample of published articles from the literature on PEW scores and systems and to conduct a small number of scoping interviews. The work presented in this report outlines key insights related to:

- The prevalence and diversity of the PEW scores and systems used in the UK and evidence for their effectiveness

- Influences on the implementation and effective use of PEW scores and systems (including those related to the features of the scores and systems themselves and those related to the wider social, cultural and organisational context), and

- The standardisation of PEW scores and systems, as well as the de-implementation of existing scores and systems.

The remainder of this report is structured as follows: Section 2 provides an overview of the methodology adopted for this scoping exercise; Section 3 presents insights on key features of PEW scores and systems, including insights related to their use in the UK; Section 4 presents insights relating to influences on their implementation and effective use, and Section 5 presents insights on the standardisation of PEW scores and systems at the national level and de-implementation of existing scores and systems.

# 2 Methodology

## 2.1. Overview of the methodological approach

The insights presented in this report are based primarily on a literature review. The review focused on articles covering the use of PEW scores and systems in UK settings. However, review articles that included UK studies alongside studies in other country settings were also included. We also consulted documents relating to the development of a standardised PEW system in England that were shared with us by NHS England and NHS Improvement.

The review's findings are complemented with three semi-structured interviews with five experts who have experience in the development, use and standardisation of PEW scores and systems from three UK regions. One interview was conducted with three experts simultaneously. Throughout the report, we reference information from interviewees by referring to them as Int. 1, Int. 2, Int. 5, Int. 4, and Int. 5.

In addition to these interviews, we also engaged in telephone-based communications with NHS England and NHS Improvement representatives and a clinical expert involved in the development and roll-out of a standardised PEW system for England.

For more information on methods, please see Annex A.

## 2.2. Profile of reviewed literature

Overall, we analysed 19 journal articles on the use of PEW scores and systems, and an additional 5 documents shared by NHS England and NHS Improvement relating to the development of a national PEW system in England. Among the journal articles, 11 were review articles (including 4 systematic reviews, 1 hermeneutic systematic review – involving an iterative process that integrates analysis and interpretation of evidence with literature searching – and 6 informal literature reviews/commentary pieces with no reported methods). In addition, the articles reviewed included one editorial and one abstract relating to a prospective mixed-methods before-and-after study.

Six articles reported on the findings of primary studies with a range of designs. This included one qualitative focus group study, one survey-based study, one cluster randomised controlled trial, one retrospective observational study of administrative data, one descriptive account of implementing a PEW system and one pilot study.

The systematic reviews included in this scoping exercise assessed the quality of the evidence reviewed. They consistently highlighted a lack of evidence supporting the validity, reliability and utility of PEW scores. Few primary studies reported on these issues

and those that did were often limited by methodological concerns.

## 2.3. Caveats

As introduced earlier, this report sets out key insights to inform the development of a potential future study that would examine how efforts to standardise the PEW system in England and de-implement incumbent practices unfold, in order to draw learning relevant to the wider theme of de-implementation and standardisation in healthcare. The literature review conducted is not a comprehensive review of the evidence but based on a sample of journal articles selected in collaboration with THIS Institute according to their relevance to the proposed focus of future work and reports shared by NHS England and NHS Improvement. It is possible that the articles covered have not captured the wider literature's full range of insights relating to PEW scores and systems

used in the UK or about influences on their effective use and implementation. In addition, the sources reviewed are limited to academic publications and those shared by NHS England and NHS Improvement and do not extend to the wider grey literature. Therefore, it is possible that the findings are not representative of the full range of activity undertaken in the development and standardisation of PEW systems in the UK. However, we have taken a systematic approach to searching the literature and screening potentially relevant records, and the articles reviewed include a number of systematic reviews on the topic. In addition, we sought the advice of an expert clinician in identifying key articles and complemented review findings with expert insights from scoping interviews. A further limitation of this scoping work relates to the relatively low-quality evidence on the validity of PEW scores and systems, their effectiveness in detecting and responding to deteriorating children, and their impact on patient outcomes.

# 3 Key insights on paediatric early warning (PEW) scores and systems used in the UK

## 3.1. What is a PEW score and system? Deciphering the plurality of terms and concepts used in the literature

EW systems are used in hospitals and other healthcare settings to help healthcare staff identify early signs of clinical deterioration in patients and facilitate timely intervention [5]. They take the form of pre-specified alert criteria intended to trigger additional care when needed, monitored through observation charts [10]. In addition to their longstanding use in adult populations, these systems are increasingly used in paediatric settings. The focus of this report is on *paediatric* EW (PEW) systems and their components.

### 3.1.1. There is a lack of clarity about the boundaries and relationships between *PEW systems* and other related terms such as *paediatric track-and-trigger systems* and *paediatric alert criteria* in the literature we reviewed. Most of the terms seem to be used relatively interchangeably.

For example, Chapman et al. (2016) define track-and-trigger systems as systems that 'support clinical decision-making by tracking a child's condition by monitoring certain clinical signs and triggering a request for appropriate review when pre-determined

criteria are breached' [8]. Roland et al. (2014) [10] define EW systems in a very similar way, as an umbrella term for 'the implementation of predefined alert criteria within observation charts which trigger additional nursing or medical involvement'. In an earlier review, Chapman et al. (2010) define the term *paediatric alert criteria* as an umbrella term including EW scores and systems and trigger/ activation criteria for response [7].

Most of the reviews we examined either used the term *PEW systems* or *track-and-trigger systems*. Some explicitly acknowledge that they use the terms interchangeably due to a lack of clarity in the literature about definitions [9, 11]. The articles covered by our scoping exercise include those that focus on PEW systems, track-and-trigger systems and paediatric alert criteria.

### 3.1.2. PEW systems comprise multiple components, including PEW scores, response algorithms and criteria under which a response or set of responses should be initiated, and wider governance and management arrangements.

Jacob et al. (2018) [4] conceptualise safety systems, including PEW systems, as the product of four integrated components working together: (i) an *afferent* component which detects deterioration and triggers action; (ii) an *efferent* component consisting of the people

and resources responding; along with (iii) a *process improvement* component (including system auditing and monitoring); and (iv) an *administrative* component (including organisational leadership and education required to implement the system and sustain it). The *afferent* component of a PEW system can be further broken down into three sub-systems: a detection sub-system to identify that a child is at risk of deterioration and early signs of deterioration; a preparation sub-system to ensure clinical teams are ready to respond in the event of deterioration (e.g. through situational awareness, group decisions and planning); and an action sub-system to allow for appropriate response actions to be initiated [4, p.11].

Similarly, the Safe System Framework for children at risk of deterioration, developed by NHS Improvement and the Royal College of Paediatrics and Child Health (referenced in the commentary by Roland (2017) [11]) conceptualises the PEW system in terms of nested levels of support for the identification of and response to the deteriorating child [11]. This model places the infant, child or young person at the centre, supported by family or carers, clinicians providing direct care, the wider clinical team, service providers and regional and national networks. The framework specifies six core elements of a safe system requiring inter-group engagement: (i) developing a patient safety culture; (ii) creating partnerships with patients and families; (iii) recognising deterioration; (iv) responding to deterioration; (v) open and consistent learning through evaluation; and (vi) education and training (see Annex D for a figure of the framework conceptualisation). NHS England has proposed the Safe Systems Framework as a potential framework to guide the development of a standardised PEW system for England [12]. The experts we interviewed for this scoping research

also supported the importance of adopting a systems view (i.e. going beyond consideration of just a score mechanism in and of itself). As one interviewee put it, 'just the score doesn't work on its own' without people who know how to use it well, and those invested in teaching, training, educating and responding around it [Int. 3]. Another interviewee commented on the importance of the wider culture and creating a psychologically safe interface between the frontline and busy medics' [Int. 5].

Sensitised by these classifications, and to provide further detail, the term *PEW system* as we use it in this report refers to a system made up of the following key elements:

1.  **PEW scores** – also referred to in the literature as *paediatric early warning tools* **or** *paediatric track-and-trigger tools* **–** have been described as a 'formal framework for evaluating routine physiological, clinical and observational data for early indicators of deterioration in paediatric patients' [9, p.2] and the 'amalgamation of various physiological and observational parameters into a singular descriptor of their extent of deviation from normal' [11, p.242]. PEW scores are a key component of many (but not all) PEW systems, with a distinction made between 'score-based' systems based on calculating a single score, and 'trigger-based' systems that do not calculate a score (see Section 3.1.3). There are also PEW systems that integrate elements of score-based *and* trigger-based systems, a matter we return to in Section 3.1.3. The term *PEW score* is used in the literature to refer to both the *scoring mechanism* used in the PEW system (i.e. the parameters included and the way these parameters are scored – such as the score a heart rate of X would receive, how many points are added for criteria such as nurse concern or respiratory support, and whether the scoring is age-dependent) and

the *value* assigned to a given patient (e.g. child X has a PEW score of Y).

- **A wide range of parameters can be included in PEW scores and various methods used to assign values to constituent parameters.** For example, different PEW scores may have different cut-off values for heart rate values classified as high, normal or low. Parameters included in PEW scores can include objective findings, subjective clinical findings, clinical events, intuitive parameters and diagnoses [7]. Frequently considered parameters include heart rate and temperature (objective findings), skin colour, fatigue and pain (subjective clinical finding), loss of consciousness (event parameter), nurse concern (intuitive parameter), or a diagnosis of sepsis (diagnostic parameter) [7]. The number and nature of the parameters included in a PEW score may have implications for how easy the system is to use and the number of false positives/ false negatives generated by the score. Please see Section 4.1 for more information on how different parameters may influence the effectiveness of PEW scores and systems.

- **PEW scores can also differ in how they deal with children of different ages.** Some PEW scores have different scoring systems based on a child's age, as acceptable values for parameters such as respiratory rate, heart rate and blood pressure vary with age. However, this variability complicates the standardisation of PEW systems, e.g. different age-bands require different types of paper-based forms [13]. In a review of 33 track-and-trigger systems, 14 were identified as being age-dependent and 19 as age-independent [8]. Similarly, in a review of ten sets of paediatric-alert criteria, half were found to be age-dependent, and the age bands varied across different tools [7].

2. **PEW charts** are used to record clinical observations systematically [Int. 1, Int. 2, Int. 3, Int. 5] [10, 14]. These observation charts are used for monitoring, recording and detecting patient deterioration [4] and can be presented in various formats (e.g. colour-coding and traffic-light systems using green, amber and red to indicate acceptable ranges for a child of a particular age bracket) [Int. 2, Int. 3, Int. 5]. PEW charts also vary in terms of how information is entered and recorded; while most are currently paper-based, the potential for electronic and automated observations is of increasing interest to policy and practice. In a 2017 commentary, Roland highlights the promise of electronic observation on hand-held and portable devices, suggesting these will become increasingly relevant [11]. A 2019 systematic review by Jacob et al. also notes the growing use of technology and refers to a number of evaluations of new technologies such as electronic recording of vital signs [4]. Progress in efforts to develop digital systems was also noted by one of the experts we consulted [Int. 4].

3. **Response algorithms or alert criteria** are used to determine the circumstances under which a response is initiated (i.e. the thresholds that need to be met to trigger a response) [7, 8, 14, 15].

4. **A response or graduated set of responses** (with responses varying dependent on the risk of deterioration or severity of clinical symptoms) is triggered when pre-defined criteria are met. These responses could include additional nursing or medical

involvement, either from the assigned clinical team or from an RRT, of individuals with enhanced critical care skills [3, 9, 10]. Details such as who observes (and how frequently), who responds and how, and pre-specified responses may be set out as part of the system.

5. **A wider context of social and organisational influences:** The elements outlined above operate within a wider system of influences on PEW-system implementation and effectiveness. These are considered in detail in Section 4 and relate to safety cultures, leadership, training, relationships, organisational management, and governance structures and practices, among other factors.

It is worth noting the lack of clarity and consistency in the literature about how the acronym PEW' is used [3, 11], sometimes referring only to the *scores* and sometimes to the wider PEW *systems* they are embedded in (i.e. to the scores, responses and actions they lead to, as well as wider organisational influences, etc.). For clarity, we refer specifically to *PEW scores* or *PEW systems* (as defined in this section) throughout this report, avoiding the ambiguity of the PEW acronym.

### 3.1.3. PEW systems can be classified according to whether and how overall scores are calculated and how responses are initiated

PEW systems can be classified into two broad types: *score-based* and *trigger-based* [1, 8], although some composite systems blur these boundaries.

- **Score-based PEW systems** calculate an overall score based on a summation of scores across individual parameters. Scores are associated with a response or escalation algorithm, which indicates the particular action to be taken by the healthcare professional for a given score. This leads to graduated responses depending on the value of a PEW score [1, 8], meaning that children at increased risk of deterioration or more severe clinical symptoms receive a different clinical response to children with less serious deterioration.

- **Trigger-based PEW systems** do not calculate an overall score. They are based instead on reaching a particular threshold in at least one parameter (not excluding the possibility that a threshold can be exceeded in more than one parameter). A specific response is triggered once a threshold for any single parameter is reached. If the observed parameter remains under the specified threshold, the response does not occur. In this system, one or more abnormal indicators identify children at risk and trigger a specified action [1, 8]. Using the modified Bristol PEW system, for example, a response is triggered if oxygen saturation levels fall below a specific threshold, *or* heart rate or respiratory rate are outside normal expectations, *or* the patient is fitting *or* staff are concerned about the patient's condition. Although a response is triggered when any single parameter reaches a specific threshold, multiple parameters could simultaneously reach the threshold for triggering a response [8]. This response involves multiple elements, including alerting the nurse in charge, increasing observation frequency, notifying the patient's medical team, and discussing observations with the registrar.

- **PEW systems that blur the boundaries between score- and trigger-based systems** are also in use. While the *score-based* vs. *trigger-based* distinction is helpful and commonly used, more detailed distinctions can be made between

PEW systems – since some integrate elements of both types. An example is the standardised PEW system introduced in Scotland, which includes eight parameters (each with a score of 0-3). A score of 3 on any indicator triggers a response on the escalation plan, while a summed score of 3 across all indicators also triggers a response and a score of 5 results in higher-level escalation [Int. 5]. Furthermore, as well as single-parameter trigger-based systems – where a response is triggered when a value for any single parameter passes a certain threshold – there can be EW systems where responses are triggered only when multiple parameters reach specific thresholds at the same time (e.g. a set number of thresholds are breached).

- In addition to the distinction between score-based and trigger-based systems, the National Institute for Health and Care Excellence (NICE) has classifications for different types of adult EW systems [NICE Short Clinical Guidelines Technical Team, 2007: cited in 11, 14, 15] and some of the classifications used in the adult sphere may also provide insights on the diversity of potential systems used in a paediatric setting. While these distinctions are useful, they do not necessarily align with the well-established distinction specifically used in the literature on PEW systems. The NICE adult classifications distinguish between:

  - **Single-parameter systems:** where specific vital signs are periodically checked against pre-determined thresholds, and a response is triggered when any single escalation criterion is met.

  - **Multiple-parameter systems:** consisting of multiple parameters with different thresholds, where a response is triggered if the thresholds for more than one parameter are met simultaneously

or responses differ depending on the number of criteria met.

- **Aggregate scoring systems:** where weighted scores are assigned to ranges of values for different parameters. A response is triggered when the total aggregated score (summed across parameters) exceeds a pre-defined threshold.

- **Combination systems:** consisting of a variety of single-parameter, multiple-parameter and aggregate-scoring systems combined into one system to trigger responses. These are the most complex.

## 3.2. The prevalence and diversity of PEW scores and systems in use

### 3.2.1. PEW scores and systems are widely used in acute settings in the UK, with various scores and systems used in different clinical settings.

A wide variety of PEW scores and/or systems are in use throughout the UK, and their use is increasing in acute settings [8, 10, 14]. They are commonly used in paediatric inpatient settings [4, 10, 16] [Int. 2], including children's hospitals [3, 7, 17, 18], oncology units and Paediatric Intensive Care Units (PICU) [9], as well as general, district and specialist hospital wards in the UK with paediatric patients [13, 19]. PEW scores and systems have also been introduced in the emergency department (ED), although their use is less well established in this setting [15, 20, 21] [Int. 3].

According to a 2014 study by Roland et al., the majority (85 per cent) of hospitals in the UK had at least one PEW score or system in place. However, only a third of those were based on established PEW scores or systems (i.e. their details had been published in the literature), while the remainder were bespoke

systems, including those adapted from other hospitals or specifically designed for use in a particular hospital [10]. One reason for this may be that clinical departments and trusts avoid using systems borrowed from others without customising them to fit their specific needs [14]. That said, there are many similarities between the diverse tools in use [Int. 2, discussions with individuals involved with the development of NHS England and NHS Improvement PEW scores and systems standardisation plans].

Historically, the absence of a coordinated approach to developing and implementing PEW scores and systems led to a lack of standardisation across the UK [3, 20]. However, the degree to which this is still the case varies between different nations in the UK. While Northern Ireland and Scotland have standardised national PEW systems in place (at least for use in paediatric inpatient settings) [Int. 1, Int. 5], this is not the case for England [22, 23] and Wales [Int. 2] where multiple different PEW systems are in place. We discuss the theme of standardisation further in Section 5.

For more information on specific PEW scores and systems used in England and the UK, please see Annex C.

PEW scores and systems in use in the UK vary according to:

- **The number and type of parameters included and how they are scored:** A 2014 review of PEW scores in hospitals in Great Britain by Roland et al. (based on a cross-sectional survey) found that 47 different component parameters were in use. Of these, heart rate and respiratory rate were most frequently used (in about 90 per cent of PEW scores), followed by oxygen saturation, abnormal consciousness and respiratory effort (all used in approximately 60-75 per cent of PEW scores) [10]. Other frequently-used parameters (in at least 50 per cent of UK units using PEW scores)

were nurses' concern, systolic blood pressure and oxygen therapy [10]. Similarly, a systematic review of ten PEW systems conducted by Chapman et al. (2010) found that the most common parameters were respiratory rate, heart rate, oxygen saturation and consciousness. The least common was temperature, which was only used in one tool [7]. The most complex system identified was the Paediatric Early Warning System Score I, consisting of 19 parameters, while other PEW scores contained anywhere from 5-14 parameters [7, 8]. A 2019 review of PEW charts in England (shared with the study team by NHS England and NHS Improvement) found similar common indicators, but also highlighted variation in whether blood pressure was included as part of the score and how behavioural changes and parental and nurse concern were scored [22, 23]. In the POPS chart, for example, the 'gut feeling' parameter is scored from 0-2 (where 0 corresponds with 'well', 1 corresponds with 'low level concern' and 2 corresponds with 'high level of concern' or 'child looks unwell'), and clinicians might score this differently depending on their experience and context [21]. Three of the experts we spoke with as part of our scoping exercise highlighted the importance of including parental concern, acknowledging that it is currently underused [Int. 1, Int. 2, Int. 3]. In a systematic review of paediatric alert criteria by Chapman, Grocott and Franck (2009), the authors distinguished between intuitive parameters (defined by knowledge without the need for rational or conscious reasoning, such as 'worried') and subjective parameters (such as 'increased work breathing'). Although the literature highlighted that intuitive findings and 'gut feeling' parameters seem important, more research is needed to understand how they should be incorporated into systems [7, 10, 21].

- **Whether the scores are age-dependent:** A review of PEW charts in England (shared with the study team by NHS England and NHS Improvement) found that the most common age brackets were from 0-1, 1-4, 5-12 and 12+ years, with considerable variation between sites as to whether they split the 0-1 age group into more specific categories [22, 23].

- **What they are used for and how:** PEW scores and systems can be introduced for a range of purposes, which may vary between settings, e.g. paediatric inpatient settings versus ED settings [14]. As well as identifying children at risk of deterioration and supporting appropriate responses, these include assisting with other aspects of healthcare delivery such as patient triage [Int. 1, Int. 3] or assessing risk for specific conditions such as sepsis [24]. In addition, PEW scores and systems are often introduced to help standardise practice and improve communication in a given setting [Int. 1, Int. 2]. The variation in how PEW scores and systems are used in practice is highlighted in a 2019 survey of UK NHS trusts (shared with the study team by NHS England and NHS Improvement), which showed that all respondents used PEW scores or systems (compared to 85 per cent of hospitals found to have at least one PEW score or system in place in a 2014 survey [10]). The PEWS scores and systems are used in different ways across trusts. For example, 55 per cent of responding UK trusts in the 2019 survey used PEW scores as a sepsis trigger tool (amongst other uses), while others did not, and 20 per cent of trusts used PEW scores to initiate a response from an RRT [24]. How the remainder determined their response is not clear from the literature. Interviewees noted that most PEW systems are about tracking deterioration

and initiating a response rather than patient triage prioritisation [Int. 1, Int. 3]. This was seen to potentially limit their usefulness from an ED perspective, where prioritisation is a key concern. However, PEW scores and systems may have value in EDs for identifying patients who are not at risk of deterioration and can be moved home or to a less resource-intensive ward [15, 21]. The Paediatric Observation Priority Score (POPS) is increasingly common in ED settings in the UK. However, there is some debate over whether this is, strictly speaking, a PEW score [Int. 1, Int. 3] or a triage tool.

## 3.3. Do PEW scores and systems make a difference?

This section presents what is known about the outcomes and impacts of PEW scores and systems. It is important to note, however, that most PEW scores and systems have not been robustly evaluated (see Section 3.4.1). Given limitations in the quality of the studies informing evidence on these issues, caution is therefore advised when interpreting insights from the literature on the impacts of PEW scores and systems.

### 3.3.1. There is a lack of evidence that PEW scores and systems are associated with improved patient outcomes in the absence of additional accompanying interventions. However, the existing evidence is based primarily on studies of limited quality, and further research is needed.

Based on the literature available, there is a lack of evidence that using PEW scores and systems improves patient outcomes [1, 3, 4, 7-10, 12, 15]. For example, a 2016 systematic review of paediatric track-and-trigger systems found that implementation does not lead to significant reductions in cardiac

arrest, respiratory arrest or mortality without accompanying interventions (such as the implementation of RRT rather than escalation by the usual care team). However, the authors of the systematic review highlighted that the quality of all reviewed studies was low to very-low quality, with serious methodological concerns and few statistically significant findings. The same review found that when the implementation of an RTT accompanied that of a paediatric track-and-trigger system, there was moderate evidence for improved patient outcomes in relation to cardiac and respiratory arrest, based on statistically significant findings [8]. Similarly, a more recent systematic review by Trubey et al. (2019) [9] identified only a small number of studies on the effectiveness of track-and-trigger tools (including PEW scores and RRT activation criteria) that reported statistically-significant decreases in mortality, arrests or code calls, but these studies were also limited by methodological concerns. That said, it is important to note that statistical significance is hard to achieve in assessments of a PEW system's impact due to the relative rarity of adverse events (e.g. cardiac arrest, respiratory arrest and mortality) in paediatric populations [8, 12, 13]. Findings from the literature and interviews indicate that evaluation efforts are hampered by a lack of robust, valid and clinically-meaningful outcomes and the wide variety of different systems in use [1, 12] [Int. 3, Int. 5].

The Bedside PEW System has been consistently identified in the reviews as the PEW system that has undergone the most rigorous evaluation [1, 3, 9]. It has recently been studied in an international multi-centre randomised control trial that assessed its impact on mortality, late admission to an Intensive Care Unit (ICU), cardiac arrest and ICU resource use [16]. In this 2018 study by Parshuram et al., the system's implementation was not found to significantly decrease all-cause mortality in

hospitalised paediatric patients, nor did it lead to significant decreases in cardiac arrest or preventable cardiac arrest, urgent or unplanned ICU admissions, hospital re-admission or ICU resource use. It did, however, lead to a decrease in late ICU admission after significant deterioration events [16].

### 3.3.2. However, there is some evidence that PEW scores and systems have benefits beyond clinical outcomes, such as improvements in situational awareness, staff empowerment, communication and recognition of deterioration.

Although there is a lack of compelling evidence about how PEW scores and systems improve patient outcomes, there is some evidence that they may have benefits outside of clinical outcomes – although assessing the strength of this evidence would require further research. For example, a recent systematic review found qualitative evidence to suggest that the use of PEW scores and systems may improve individual and team situational awareness, enabling clinicians to have a 'birds-eye' view over admitted patients [Vries et al., 2017: cited in 4]. This finding was reinforced by another more recent (2019) review by Chapman and Maconochie (no method stated), which highlights specific improvements to the situational awareness of junior staff or those new to paediatrics [1]. This benefit was also highlighted by one of the experts we consulted for the scoping work [Int. 4], and a clinician in England involved with NHS England and NHS Improvement PEW scores and systems' standardisation efforts.

Some reviews included in our scoping exercise (including two recent systematic reviews) suggest that PEW scores and systems can also support multidisciplinary teamwork, communication and staff confidence in recognising deterioration and making decisions around escalation [3, 4, 15], particularly

by allowing the risk of deterioration to be communicated more effectively and efficiently between healthcare staff members. The evidence reviewed consisted of qualitative interview studies, staff survey findings, quality improvement projects and descriptive case studies (such as that by Monaghan et al. [19]). This finding was reinforced by one of the experts we interviewed [Int. 3]. Improved communication around deterioration can be especially important during handovers, as they can help ensure that deterioration does not go unnoticed during this high-risk time for patients [21]. PEW scores and systems may also empower all staff to summon help when a child is deteriorating [3, 14], which may help flatten the hierarchies in medical settings that prevent junior staff from taking action. Notably, while improved situational awareness, teamwork and communication are identified as benefits of PEW scores or systems in the studies referenced above, they are also considered to be modifying influences on the implementation and effective use of PEW scores and systems in the wider literature. This is explored in Section 4.2 below.

### 3.3.3. There is also some evidence that the use of PEW scores and systems can have unintended consequences related to unnecessarily increased staff workload, as well as potentially to patient outcomes

Roland (2012) cites anecdotal reports that PEW scores can increase workload rather than improve patient safety [14]. In a later commentary piece, Roland (2017) also discusses a study conducted in an ED that found that review activations may have generated an additional 7060 minutes of work for no net benefit [11]. Expectations of increased workload may contribute to staff distrust and a reluctance to use PEW scores [11, 14]. A retrospective study in the United States found that implementing PEW

systems could overwhelm hospital resources due to delayed discharge and inappropriate readmissions for children incorrectly identified as deteriorating [Mandell et al. 2015; cited in 1]. Furthermore, based on a review of PEW scores and systems used in emergency departments, Oldroyd et al. (2011) found that many tools are complicated and time-consuming to use [20] (although it should be noted that no method is described for the review and hence the comprehensiveness of the review is unclear).

The potential harm to patients from PEW scores and systems is associated with limitations in score sensitivity or specificity [3, 4, 8, 9, 15] and the associated over-treatment of children identified through a PEW system who are not actually deteriorating (false positives) or the under-treatment of children identified through a PEW system who are actually deteriorating (false negatives). This risk has also been highlighted in the literature [10] and reinforced by interview findings [Int. 1, Int. 3]. A systematic review by Lambert et al. (2014) included a qualitative study by Brady and Goldenhar (2014) suggesting that standardised processes and procedures can result in 'task fixation' for less experienced staff, preventing them from seeing the bigger picture of good care provision [3].

## 3.4. Are certain PEW scores or systems better than others?

### 3.4.1. It is not clear whether any particular PEW score or system is better than others at identifying deteriorating patients and avoiding negative outcomes. Most have not been validated or rigorously evaluated.

The predictive accuracy of most PEW scores and systems used in the UK and/or internationally to assess the risk of adverse outcomes in children (e.g. the Brighton PEW score, the Yale System, the Cardiff and Vale

PEW system, and the Melbourne Activation Criteria) have not been validated. In addition, some have only been tested in single-centre studies or studies looking at homogenous patient cohorts in similar healthcare settings [1, 3, 9, 20]. Based on an international multi-centre randomised controlled trial, the evaluation of the Bedside PEW system (as discussed in Section 3.3.1 above) is somewhat of an exception. Therefore, the ability of many PEW scores and systems to successfully identify children at risk of deterioration is largely untested. A systematic review by Lambert et al. (2014) highlights the lack of studies evaluating PEW systems as a complex health intervention, so it is unclear what the 'active ingredients' of implementing a PEW score or system are, even when positive outcomes

have been demonstrated [3]. There have also been very few studies looking at the inter-rater reliability of PEW scores [14], meaning that their reproducibility is mostly unknown.

For those PEW scores and systems that have been evaluated, there is very limited evidence of one particular score or system's advantages over another. The PEW System Score (created in Toronto) has been validated retrospectively using a frequency-matched case-control design in a Canadian University-affiliated paediatric hospital. It has been shown to be a stronger predictor of cardiopulmonary deterioration than both the *PEW Tool* and *Bedside PEW systems*. However, the large number of parameters means that staff are less likely to complete the chart in full compared to tools that use simpler scoring mechanisms [3].

# 4 Influences on the implementation and use of PEW scores and systems

This section focuses on influences on the implementation and use of PEW scores and systems as they apply to their initial introduction/implementation processes and continued use.

We first consider the influence of specific features of PEW scores and operational systems as introduced in Section 3 (by 'operational systems', we mean the scoring mechanism in the context of its larger system of response-algorithms and triggered responses). We then focus on the influence of factors related to the wider social, cultural and organisational context. We are interested in influences on both how well the *process* of introducing and using PEW scores and systems works and on the *performance* of the PEW score and system in identifying patients at risk of deterioration and escalating care. Sometimes these distinctions are clear in the literature, while in other cases, influences are reported in more general terms. Where possible, we have identified these distinctions in the influence described.

## 4.1. Features of PEW scores and systems that influence their implementation, use and effectiveness

Despite the lack of strong evidence regarding the relative merits of one particular PEW score or system over another, there is some information in the reviewed literature about features that make PEW scores and systems more or less difficult to implement and use and that influence their effectiveness for identifying deteriorating patients and escalating care appropriately.

Such key influences on implementation (related to the PEW score and system itself as opposed to the wider social, cultural and organisational context) include:

- Whether the PEW system is score-based or trigger-based

- The complexity and ease of use of the PEW score and system, including the visibility of the system on wards in healthcare provider organisations

- How observations are recorded and how escalation requirements are communicated: paper-based vs. digital systems

- The predictive validity of the PEW score (specificity and sensitivity)

- The degree to which the score integrates subjective criteria and accommodates scope for clinical judgement, and

- How tailored and adaptable the PEW score and system is to a particular use and setting.

### 4.1.1. Trigger-based versus score-based systems

**There is some evidence to suggest that score-based PEW systems outperform trigger-based systems in identifying deteriorating patients and triggering appropriate responses.** For example, based on a systematic review including articles relating to 21 score-based and 12 trigger-based PEW systems used in paediatric inpatient settings, Chapman et al. (2016) concluded that 'score-based systems may have benefits over trigger systems' because the opportunity they offer for a graduated response 'may be a better use of resources and expertise' and they have been evaluated more and demonstrated better sensitivity [8 , p.107].

There is also some evidence to suggest that score-based systems perform better in the ED setting. A study comparing the performance of ten PEW systems (six score-based and four trigger-based) in a single ED in the Netherlands found that score-based PEW systems outrank trigger-based systems in terms of their ability to predict children who will be admitted to PICU [Seiger et al., 2013; cited in 1] and in [3]]. NICE guidelines relating to the use of EW/track-and-trigger systems in general (not specific to paediatrics) also recommend the use of EW systems that enable a graduated response and have better predictive value (those with multiple parameters and aggregate weighted scoring) over EWs that are based on single parameters or do not allow for graduated score-based escalation of patients [7].

However, some of the interviewees consulted in this scoping research also identified advantages associated with trigger-based systems, including their relative simplicity compared to score-based systems (i.e. there is no requirement to add up a score, therefore less room for error) [Int. 3, Int. 5]. Section 4.1.2 discusses complexity further. One interviewee also noted that while one advantage of

score-based systems is that that they require the completion of all specified observations in order to generate a score, this may also be a limitation if it is not possible to obtain a particular measurement (e.g. the blood pressure of a distressed child) [Int. 2].

Thus, whereas score-based systems may allow graduated responses and better sensitivity, trigger-based systems may be more practical to use and further research is needed for a more comprehensive evidence base.

### 4.1.2. The complexity and ease of use of the scoring mechanism and how observation data is communicated and made visible to staff

**There is evidence to suggest that the complexity of the scoring mechanism and the degree to which displays are user-friendly and easily visible to staff can influence the implementation of a PEW system and its effectiveness in several ways. These include the following:**

- **Scoring mechanism complexity influences how clearly a PEW score can be interpreted and the degree to which it may be prone to error.** Scoring mechanisms that are simple and easy to use may support more effective identification of deteriorating patients [3, 12]. For example, some systems require clinicians to calculate one score for a parameter based on several sub-scores (e.g. a cardiovascular parameter measured by skin colour, capillary-refill time and heart rate), which can complicate score-assignment when sub-scores indicate different levels of severity [8]. More complex scoring mechanisms may, therefore, be more prone to error [7]. One interviewee highlighted that the ability to count and add up can be impaired even under minimal stress [Int. 5].

- **More complex scoring mechanisms affect the time and effort required by staff to**

complete the required observations, the acceptability of the PEW score or system by healthcare staff, and the likelihood of wider adoption [7, 10]. However, Monaghan (2005) [19] notes that as staff become more familiar with PEW scores and systems, the required time and effort can decrease, leading to increased acceptability.

- **User-friendly formats for recording and displaying data can facilitate the interpretation of score values.** The inclusion of graphical and numerical displays in PEW charts can decrease staff cognitive workload when completing PEW score observations, helping facilitate their use and more effective identification of children at risk of deterioration [21]. Interviewees highlighted the use of colour-coded 'traffic light' systems (with scores within the red range of a chart indicating a need for action) as a widely-used and useful approach [Int. 3], instilling users with a greater level of confidence about escalation requirements [Int. 5].

- **Increasing the visibility of PEW scores and systems on healthcare wards may help staff use them effectively.** For example, placing laminated PEW charts at all nursing and computer stations or placing whiteboards with coloured circles to indicate a patient's PEW score at each ward bed have been implemented to facilitate effective use [3, 4].

### 4.1.3. How observations are recorded and escalation requirements communicated: paper-based vs. digital systems

The formats for completing observations and communicating the need for escalation influence both the implementation of PEW scores and systems and their effectiveness:

- **Implementation of paper-based systems is considered more straightforward and**

less resource-intensive and subject to fewer delays than would be the case with digital systems (given current digital infrastructure in the NHS). Interviewees observed that the introduction of a digital PEW system would have greater resource and governance implications, thus increasing the need for senior buy-in [Int.1]. The key resource requirement for paper-based systems relates to printing costs for PEW charts [Int. 1] and roll-out is not dependent on compatibility with existing IT systems, nor is it subject to IT-related delays [Int. 3].

- **There is some evidence that electronic-based PEW scores and systems could detect deterioration and channel appropriate escalation more accurately than paper-based systems. However, this evidence is limited, and there are also concerns about higher false-positive rates with automated alarm systems, resulting in alarm fatigue.** Electronic-based PEW systems can enable new functionalities, including automated escalation of patients and the use of machine learning to predict the need for escalation using a patient's baseline data [1]. Similarly, incorporating electronic vital-sign monitoring into PEW systems is associated with positive outcomes for the timeliness and accuracy of detecting deterioration [4]. According to one interviewee, paper-based systems do not adequately deal with the range of age-based physiological variation, and digital systems could allow for narrower age bands to assess physiological variation – not least because paper-based systems require the use of multiple age-dependent charts and paper forms [Int. 1]. It was also noted that digital systems could mitigate human error in translating observations into scores [Int. 1]. However, there is also some evidence (both anecdotal and based on evaluations of automated systems that

trigger audible alarms) that they can also increase alarm-fatigue in hospitals, leading to a decreased response to deteriorating patients when an alarm does sound [4, 11] [Int. 4]. This happens when such systems lack specificity, leading to high numbers of false positives. One interviewee described how a newly introduced system 'beeped' so frequently that it was turned off within 24 hours [Int. 5].

- **The use of digital systems may facilitate data collection to improve scores and enable monitoring and evaluation that ultimately help improve outcomes.** For example, the automated collection of data and incorporation of PEW systems into electronic health records may help evaluate PEW systems in real-time and improve how scores and responses are constructed [11] [Int. 1].

### 4.1.4. The predictive validity of a PEW score within the particular settings in which it is used

**The degree to which the PEW score or system can distinguish between deteriorating and non-deteriorating patients influences its effectiveness. There is a need for balance between the sensitivity and specificity of any PEW score. Considerations are:**

- **The most effective PEW screening tool would balance high specificity and high sensitivity** [3]. In any PEW score, there will be a trade-off between sensitivity (the ability of a score to correctly identify children who are deteriorating) and specificity (the ability to correctly identify children who are not deteriorating). Low sensitivity results in a high number of false-negatives (i.e. high numbers of children who are deteriorating but not identified as such), while low specificity results in a high rate of false-positives (i.e. children who are not deteriorating but who are

falsely identified as deteriorating) [3]. As sensitivity increases, specificity tends to decrease, and vice versa. Hence, there is a need to consider how to balance these requirements when determining the threshold values for each parameter that triggers responses within a PEW system [3, 7].

- **Scores with low specificity (i.e. with a high rate of false-positives) can have unintended consequences on resource use and staff response (i.e. alarm fatigue,** as discussed earlier [3, 9])**.** One interviewee described how this can lead to frustration for senior clinicians and a lack of confidence in the processes [Int. 1].

- **PEW scores with low sensitivity (i.e. resulting in a high rate of false-negatives) are associated with a risk of under-treating deteriorating children who are not correctly identified by a PEW system** [Int. 1]. As one interviewee highlighted, 'It just takes that one child with a death or with significant morbidities to wipe out all the other good stuff that has been done to improve your system' [Int. 5].

### 4.1.5. The degree to which PEW scores and systems incorporate subjective criteria and allow for clinical judgement

The evidence base on the relative merits and limitations of including subjective observations, judgements and concerns of clinicians, nurses, parents and carers into PEW scores and systems is inconclusive:

- **The reliability of a PEW score is influenced by the degree to which the definition of a normal value for an objectively measured parameter (such as heart rate) is explicitly specified or open to a clinician's judgement.** Some systems require clinicians to judge what constitutes a 'normal' range of values for parameters

such as heart rate and respiratory rate [8], which may make scores more error-prone and less reproducible.

- **There are diverse opinions as to whether subjective criteria relating to clinician or nurse concern/gut instinct should be included in PEW scores.** Their inclusion may help ensure that PEW scores and systems do not risk overriding clinical judgement and support increased situational awareness [11]. However, inclusion can also make PEW scores prone to error and misinterpretation based on there being more than one way to complete a PEW score observation, especially during busy times on wards [19].

- **Similarly, the utility of incorporating parent, family or carer concerns is a matter for debate.** There is insufficient evidence to support the effectiveness of including this parameter [4]. However, two interviewees flagged that parental input was particularly important for children with complex backgrounds for whom 'normal' observations may not be anticipated [Int. 2, Int. 3].

- **There is a widely acknowledged need to ensure that PEW scores and systems do not override clinical judgement.** Chapman and Maconochie (2019) [1] report that failure to integrate clinical judgement impedes the use of PEW scores and systems in clinical practice. A number of articles highlight the need to ensure that PEW scores are not the sole mechanism that is relied on, with professional judgement continuing to play an important role in recognising and responding to signs of deterioration [4, 7, 15] and physicians being trained to recognise signs of deterioration without PEW scores and systems [1, 15]. In a qualitative study focused on escalating care, Gawronski et al. (2018) [17] warn that introduction

of PEW scores and systems can result in tunnel vision around the reasons for deterioration, further highlighting the need for clinical judgement to be part of the decision-making process.

## 4.1.6. The extent to which the score and system aligns with the specific needs of a local adoption context: adaptability

The effectiveness of a PEW score and system may be improved by its ability to adapt to local contexts during implementation.

- **Because patient populations, illness cohorts and clinical settings are diverse, there may not be a single PEW score or system that works well for all settings and patients** [1]. For example, a review of the literature around PEW scores and systems has shown that some parameters included in PEW scores may be of limited value immediately after surgery, when acceptable values of vital signs tend to vary [4]. Similarly, a review of PEW scores and systems used in emergency departments found that the specificity of PEW scores in this setting was particularly low, yielding high numbers of false-positives [15]. This means that threshold values may need to be adapted to particular clinical contexts for scores to accurately capture acceptable variation in physiological parameters within a particular clinical context [15]. Different PEW scores and systems may be more suited for particular settings due to factors such as the acuity level associated with a particular illness or setting [1]. In light of these considerations, some authors have argued that the functions, processes and goals of PEW scores and systems should be standardised, rather than the exact components that make up a PEW score or system [13].

- **Both the initial choice of a PEW system and consideration of how it could**

be adapted to the local context have been highlighted as important factors influencing PEW system effectiveness within a given setting. Based on the findings of their systematic review, Chapman et al. (2010) [7] recommend that hospitals considering introducing a PEW score or system should consider which system (amongst the many available) best meets their local needs and patient population, and, once a system is in place, undertake ongoing performance monitoring and modifications to improve performance [12]. For example, PEW scores and systems may be adapted to the local context by customising the language used in PEW charts to match the way healthcare teams talk about escalation [3, 16].

## 4.2. Factors related to the wider social, cultural and organisational context

In addition to features of the PEW scores and systems themselves, the wider social, cultural and organisational context in which they operate impacts implementation processes and score/ system effectiveness. Useful models conceptualising how PEW scores and systems are situated within this broader context include the Safe System Framework for children at risk of deterioration [12], and the model put forward by Jacob et al. (2018) [4] based on their hermeneutic systematic review (as described in Section 3.1.2 above). Sensitised by these frameworks, we consider the relevant wider social, cultural and organisational factors identified in the broader literature in the follow sections. These influences include:

- Staff engagement, teamwork and leadership

- Ensuring staff have appropriate skills and capabilities

- The existence of complementary structures, processes or procedures to support the recognition of and response to deterioration

- Resource availability (staff, funding and infrastructure), and

- Ongoing monitoring and evaluation of PEW scores and systems.

### 4.2.1. Staff engagement, teamwork and leadership to support a culture that enables the implementation and effective use of PEW scores and systems

- **Engaging staff in developing and implementing PEW scores and systems can help increase buy-in and build a supportive safety culture to aid local implementation and effective use.** Staff perceptions regarding the applicability of PEW scores and systems within their particular setting can influence the acceptability of, or resistance to, their uptake. For instance, senior medical and nursing staff may potentially reject PEW scores and systems because they feel their patient or illness cohort is unique and that PEW scores and systems do not work in specialised wards with specific patient populations [11]. The introduction of a PEW system may also be met with scepticism by staff who feel capable of recognising and responding to signs of deterioration without additional tools or external escalation resources [19], particularly in the light of limited evidence for the effectiveness of PEW scores for improving patient outcomes [Int. 3, Int. 4]. Local champions of PEW scores and systems who can lead implementation efforts and help secure buy-in from staff, management and ward leaders can help ease the implementation process [3]. Staff engagement is important in aiding the adjustment and refinement of scores

and systems to meet local requirements [14]. The inclusion of a multidisciplinary team to assist in the implementation of PEW systems can help staff discuss the potential challenges that will arise during implementation and come up with solutions to them. Incorporating PEW score observations into existing observational charts can reduce the need for additional paperwork, making the tool more acceptable among staff members [19]. To ease implementation, PEW scores and systems can be rolled out gradually. For example, they may be rolled out in particular wards on weekdays during day shifts before being rolled out as a 24/7 process. Once PEW scores and systems are implemented, regular stakeholder meetings can help troubleshoot challenges as they arise [3].

- **Leadership and clinical buy-in at a senior level are important in facilitating PEW system implementation** [25]. A commitment at an organisational level to overall improvement in patient safety, prioritisation of safety over other concerns and executive accountability have been identified as important facilitators to PEW system implementation [11]. The requirement for senior-level buy-in is of particular concern if PEW system implementation has significant resource implications [Int. 1, Int. 3].

- **Unclear lines of accountability, issues of hierarchy, and inappropriate communication channels between different professions can pose challenges to the escalation of care through designated mechanisms.** For example, if the culture of a healthcare team discourages nurses from calling external staff (i.e. to the ward/primary healthcare team) to escalate issues, or if staff are criticised for escalating patients that end up not needing additional care, this may disincentivise them from calling RRTs or taking pre-specified actions once patients are identified through PEW systems [3]. Similarly, if track–and-trigger systems are seen primarily as a nursing tool rather than a tool for clinicians, nurses may struggle to summon appropriate responses when they escalate a patient based on their PEW scores [4]. In order for nurses and clinicians to appropriately respond to children who are identified through a track-and-trigger system as potentially deteriorating, staff need to be empowered to call for additional assistance through designated channels [8]. Concerns about appearing inadequate in front of colleagues and feelings of not wanting to seek help can impede the effective escalation of patients through PEW scores and systems, especially when staff feel that the situation is under control or the area of deterioration is within the expertise of the team treating the patient at the time [4]. Further barriers to effective escalation include unclear accountability concerning the escalation process and lack of effective communication [17] [Int. 1].

## 4.2.2. Whether staff have appropriate skills and capabilities to support the implementation and effective use of PEW scores and systems (including the importance of situational awareness)

- **The degree to which staff have the required clinical competencies to use PEW scores and systems effectively to identify deteriorating patients and respond appropriately is an important factor influencing implementation and use.** The degree of clinical competence amongst staff for identifying and responding to deteriorating children will vary across settings and impact the interpretation of PEW scores and whether appropriate

action is taken [Int. 1]. For example, there are particular challenges associated with implementing PEW scores and systems in general EDs compared to specialist paediatric wards. ED nurses, who may work with a small number of paediatric patients each day, may lack the clinical competence to identify deteriorating children, whose deterioration is clinically different from adults [20]. In a qualitative study exploring factors influencing effective escalation of care in an inpatient setting, Gawronski et al. (2018) found that staff may lack either the clinical skills to respond or the confidence in their abilities to do so, depending on exposure to relevant training, past experiences in escalating care for deteriorating children and mentoring by peers [17]. However, this lack of experience or clinical competence can also provide an argument for the introduction of PEW systems, which may be introduced in part with educational intent. There is often particular demand for their introduction from those frontline staff who are not used to dealing with children since PEW systems have an important role in supporting less experienced staff [Int. 3]. The importance of clinical judgement in recognising and responding to signs of deterioration and the acknowledged need to ensure that PEW scores and systems do not override this are considered in Section 4.1.5 above.

- **Situational awareness among healthcare staff and teams is an important enabler for the effective use of PEW scores and systems to identify and respond to deteriorating children. This can be developed through training, communication and teamwork, and by empowering patients, carers and families as well as staff to escalate concerns.** Situational awareness, a concept with origins in aviation training, refers to

being fully engaged and aware of one's environment and any dangers (current or potential) to health and safety at that moment. The importance of embedding PEW scores and systems into a wider approach that seeks to support and improve situational awareness in individual healthcare staff and teams is widely acknowledged [1, 4]. In a systematic review of PEW systems, Lambert et al. (2014) identified a range of enablers to situational awareness, based on findings of a qualitative study [Brady and Goldenhar (2013); cited in 3] ). Among these were efforts classified as relating to 'team-based care', including efforts to empower families to escalate concerns to nurses and physicians when they observe deterioration, empower nurses to escalate concerns to senior staff, and promote a unit culture that supports teamwork, open communication, trust, accountability and safety. These findings are echoed by Gawronski et al. (2018) [17] in a qualitative study exploring factors influencing escalation.

- **Training and education are important elements of the wider PEW system to ensure staff have appropriate skills and capabilities to support their implementation and effective use.** For example, the report on a pilot study to assess the feasibility of using confidential enquiries to investigate children's deaths in the UK [2] recommends that all healthcare professionals who treat sick children should have appropriate training and supervision to ensure that their key skills and competencies can be demonstrated, standards maintained and performance assured. Some interventions to build staff skills and capabilities can happen before introducing a PEW score and system, such as training to improve situational awareness or training for staff across

an entire ward to ensure consistent use of language regarding patient risk [3]. For example, one interviewee noted the importance of PEW systems 'not just being the score, but having people invested in teaching, training, educating and responding around it' [Int. 3]. Real-life scenario-based learning has been reported to be particularly useful when implementing PEW systems – staff can rehearse scoring, escalation and handovers using PEW systems to help clarify the role of each professional group (e.g. nurses, clinicians) [3]. There is also evidence that self-directed e-learning and peer training models can be effective in training staff to use PEW systems [3]. According to Monaghan et al. (2005), in an article describing the implementation of a PEW system, bringing in a team of intensive care staff to assess and care for children identified as deteriorating and arrange transfer to intensive care was perceived as de-skilling existing staff and met with scepticism. The authors found that a change in approach to a more education-based model that focuses on up-skilling existing ward staff and training them to better respond to deteriorating patients improved the acceptability of introducing a PEW score and system [19].

### 4.2.3. The existence of complementary structures, processes or procedures to support the recognition of and response to deterioration through the implementation of PEW scores and systems

A range of standardised processes, roles and structures can also support staff situational awareness and improve the effectiveness of PEW scores and systems [4, 12]. These include:

- **Structures and roles to proactively identify and plan for risk** [3]. For example, studies

identified in a recent systematic review found that the introduction of 'huddles' (multidisciplinary discussions at specified times of the day on risk factors around deterioration) can facilitate situational awareness [4] [Int.3]. The combination of huddles with a 'watchstander role' (a nurse or clinician who is responsible for knowing which patients are at a high risk of deterioration) was found to be associated with an almost 50 per cent reduction in transfers from acute to intensive care due to unrecognized situational awareness events [4]. The use of the 'watcher' patient category (patients identified by staff as at risk of deterioration) has been identified as a further helpful strategy [4], and consistent processes for the ward rounds and continuous observation of patients as important for recognising deterioration [17].

- **Structures to support communication during handovers and continuity of care** [3]. In a systematic review of ways to optimise EW scores for paediatric inpatients, Jacob et al. (2018) identified growing evidence that structured approaches such as checklist systems and cognitive aids improve the quality of information sharing at handover and contribute to situational awareness, with most of the published interventions being variations of the Situation, Background, Assessment, and Recommendation tool (SBAR) [4]. They also identified literature to support the use of common information spaces such as whiteboards to aid situational awareness, all of which can make PEW systems more effective [4].

- **Structures to support adequate workload and staffing.** A qualitative study by Brady and Goldenhar (2013) [cited in 3] identified a range of staffing and workload measures that impacted situational awareness, including an improved staff-to-patient

ratio to ensure that patient monitoring is appropriate and timely, an experienced and diverse team of providers available on all shifts, and availability of extra resources if needed.

**In addition, specific measures may be introduced to support the response to an identified clinical deterioration [4],** e.g. RRTs (teams of specialist personnel to rapidly respond to critical illness). A 2012 cross-sectional survey of hospitals in Great Britain with inpatient paediatric services found that a quarter of the units surveyed had an RRT in place [10]. However, the evidence on their effectiveness is inconclusive. For example, a systematic review by Chapman et al. (2016) found that while there is limited evidence on the effect of PEW systems as a single intervention, there is moderate evidence of its impact on mortality and cardiac and respiratory arrests when delivered alongside an RRT [8]. There is also some qualitative evidence to suggest that RRTs could act as a barrier to escalating care. A qualitative study exploring factors influencing the escalation of care of deteriorating children in a children's hospital [17] highlighted the RRT as a potential barrier to escalation due to staff reluctance to call the RRT unless clinical deterioration is extreme.

## 4.2.4. Resource availability: staff, funding and infrastructure

**The availability and appropriate balance of staff influences how PEW scores and systems are implemented** (e.g. management staff, senior nursing staff on each shift [17]). In particular, issues with high workload and reduced senior staffing have been identified as factors impeding the escalation of care [17][Int. 1]. For example, one interviewee noted that 'there are some units that have registrars available 24/7, but there are other units where that is more challenging, not just from an operational vacancy point of view

but [due to] the model of care', and this can influence responses to PEW scores [Int. 1]. Another commented that 'There is a resource and infrastructure necessity for PEWS […] you can have all these things sitting there looking nice and brightly coloured, but if you don't have time to do it then is not going to work' [Int. 5]. A perceived lack of PICU capacity or other staff involved in escalation processes through PEW systems may discourage staff from appropriately escalating patients [4]. High patient loads during particular seasons or times of day are also a particular challenge [21].

**The availability of other resources also influences how the implementation of PEW scores and systems can unfold. These include, for example, PICU beds [17], IT infrastructure [4] and equipment [20].** Larger hospitals may have more resources available to escalate patients, for example, leading to better outcomes in those settings [14]. There may also be particular demands on resources associated with PEW-system implementation in ED settings. For example, children require different-sized instruments (e.g. blood pressure cuffs and pulse oximetry probes) compared to adults, and staff may lack the appropriate equipment to take measurements for PEW scores [20]. The availability of appropriate IT infrastructure has also been discussed in the literature. In their systematic review, Jacob et al. (2018) highlight that while it is likely that technology will be increasingly incorporated into PEW systems, the extent to which this is feasible across the NHS merits further research into potential unintended consequences [4]. They highlight that the successful integration of technology into PEW systems will depend on local contexts. For example, if insufficient technology is available to complete a digital-based PEW system, staff may 'batch' the input of PEW-related data into the electronic system, delaying the recognition and response to deteriorating patients.

IT-system interoperability within and across healthcare settings was also highlighted as a challenge [Int. 3].

## 4.2.5. Ongoing monitoring and evaluation of PEW score or system

**PEW scores and systems must be subject to ongoing evaluation to ensure continual improvements in their ability to identify a deteriorating child and ensure a timely and accurate response** [3, 7, 11, 25]. There is currently insufficient evaluation of PEW scores and systems [1, 3, 8], especially as complex interventions [3]. One interviewee notes, 'We need to have more data gathering, which is good from a research point of view, but we actually need to translate some of the knowledge and learning into a very focused programme whereby we're developing much more specific and sensitive tools that change outcomes' [Int. 1]. Continuous evaluation should ensure that PEW systems achieve the correct balance between complexity/ ease of use, diagnostic accuracy and required monitoring frequency [7] [Int. 1], and it is recommended that such evaluations should be conducted in each setting where PEW systems are implemented to ensure that they are being used effectively [3] (see also Section 4.1.4).

**Relevant process and outcome measures need to be identified to allow meaningful evaluation of PEW systems** [25 ; Int. 2; Int. 3]. Mortality and cardiac arrests are rare in the paediatric population, so while these outcomes are useful in evaluating EW systems' effectiveness in adult settings, they are less so in paediatric settings [25]. Although further research is needed to identify the best alternatives, some examples mentioned in the literature include the time between cardiac arrests as an interim outcome [3] and 'emergency transfer' rates as a measure of deterioration that leads to worse outcomes. The latter metric incorporates transfer to ICU, where a patient receives intubation, inotropes or three or more fluid boluses within an hour [25]. With respect to process measures, discussions with individuals involved with PEW system standardisation plans in NHS England and NHS Improvement suggest that there is scope for trusts to self-assess the maturity of standardisation processes for PEW scores and systems in their localities, with possible maturity levels ranging from Level 1 (use of standardised observation chart, regular audit of practice and use in all paediatric inpatient units) to Level 4 (system-wide electronic data collection and display, system learning and adjustment in real-time, and system interoperability across healthcare settings) [12]. An interviewee we consulted also described an evaluation system they use, which assesses whether five elements are in place to support the detection of deterioration in children through a PEW system: governance, structures, processes, outcomes and training [Int. 4].

# 5 Key insights on the standardisation of PEW scores and systems at the national level and de-implementation of existing scores and systems

In this section, we set out the key insights gained through our scoping exercise on issues related to the standardisation of PEW scores and systems and the de-implementation of existing ones. We found some limited reference to these issues within the literature. However, most of the findings presented relate to interviews with experts in Northern Ireland, Wales and Scotland, and communications with individuals in England involved with developing NHS England and NHS Improvement's PEW scores and systems standardisation plans for England.

Standardised PEW systems have been introduced in Northern Ireland and Scotland, but not yet in England or Wales. Work to develop a national PEW system for England is in progress. Efforts in Northern Ireland and Scotland have focused on the acute-care setting to date [Int. 1, Int. 5], although in Scotland the national PEW system is being trialled by the Scottish Ambulance Service and by general practitioners (GPs) in certain situations [Int. 5]. Similarly, plans for developing a standardised PEW system in England will focus on its use in hospital settings initially [Source: communications with individuals involved with developing NHS England and NHS Improvement PEW scores and systems standardisation plans].

In both Northern Ireland and Scotland, existing PEW systems needed to be de-implemented for a standardised PEW system to be introduced. PEW systems in use before standardisation in Northern Ireland were limited to a small number of hospitals using PEW charts developed elsewhere (with the majority using rudimentary temperature, pulse and respiration (TPR) charts [Int. 1]). This was not the case in Scotland, however, where PEW charts were widely used, and there had been significant investment in the development of local PEW systems in many hospitals [Int. 5]. The latter mirrors the current landscape in England and Wales.

In the contents that follow, we draw on insights from standardisation efforts that have already taken place (e.g. in Northern Ireland and Scotland) and from expert views on issues of standardisation or de-implementation about what might happen in the future in England. The insights presented in this section relate to the drivers behind standardisation, the factors influencing the standardisation process, and evidence for the impact of standardisation. These are considered in turn below.

## 5.1. Drivers of standardisation: Why standardise a PEW score and system?

A number of drivers for the development of a standardised national PEW system were identified in the literature and described by the experts we interviewed. We focus on drivers specifically related to *standardisation* (as opposed to the *implementation* of any local PEW score or system). These include:

- **A strong desire for standardisation among paediatric clinicians across district and regional hospitals and dissatisfaction with existing approaches to identifying and responding to deteriorating children** [Int. 1]. This is of particular concern where levels of migration among healthcare staff and patients are high, and there is frequent movement between hospitals – particularly when the lack of a common language or scoring system means that a particular PEW score might indicate a need for concern in one unit but not another [Int. 5]. Sources highlight that nurses have advocated for a standardised PEW system in England and related training to be integrated into the nursing curriculum (according to communications with individuals involved with the development of NHS England and NHS Improvement PEW scores and systems standardisation plans). Difficulties in tracking ill children across healthcare settings (i.e. from GPs to district hospitals to ambulance services to ICUs) were also identified as a potential reason for inter-setting standardisation by one of the experts we interviewed [Int. 5].

- **The potential to optimise the use of resources via a single approach rather than duplicating effort** [1]. This also enables learning to be shared across UK regions [Int. 1].

- **A desire to overcome the challenges in evaluating PEW systems that such a large variety of systems presents.** The development and implementation of a standardised PEW system facilitates a coordinated national evaluation of its implementation, impact and effectiveness [10].

## 5.2. Influences on the standardisation process

A range of influences on standardisation was also identified in our scoping exercise. These broadly revolve around the ability to secure buy-in from stakeholders early on (and sustain it throughout the implementation of a standardised PEW score and system) and the adaptability of the standardised system to local settings.

**As per the process of introducing PEW scores and systems in general, standardisation processes call for system-wide stakeholder engagement early on in the process in order to secure buy-in.** This includes engagement with senior leadership (clinical, managerial and executive leaders) and frontline staff who will use the standardised PEW system on a day-to-day basis. One interviewee stressed the importance of genuinely listening to and engaging with people's concerns as part of the effort to secure buy-in for standardisation, as well as feeding back on how concerns have been considered in standardisation plans [Int. 5]. Engaging staff who will be using the PEW system *early* in the development process rather than imposing a new policy in a top-down fashion was seen as key to successful engagement with standardisation efforts [Int. 2].

**The lack of evidence showing that PEW systems decrease mortality and morbidity can be a barrier to securing buy-in.** In this context, individuals involved with the development of NHS England and NHS Improvement PEW

scores and systems standardisation plans reflected that they have needed to frame their communications around standardisation in the context of the evidence relating to its potential benefits. Rather than focusing on benefits of standardisation in terms of patient outcomes(which is unsupported by the available evidence), they have instead focused on improving communication and addressing system difficulties in identifying unwell children. The lack of evidence for PEW systems' effectiveness was also seen as an impediment to standardisation in Wales, and associated with some hospitals' reluctance to adopt a standardised PEW system [Int. 3]. However, two interviewees (during a joint interview) agreed that delaying introducing a new national PEW system whilst waiting for evidence may mean that existing processes become more embedded and difficult to shift [Int. 3, Int. 4].

**The appetite for change also depends on how valued and acceptable an existing system is seen to be.** This has implications for introducing a standardised system and for the associated de-implementation of existing processes and procedures. Discussions with individuals involved with the development of NHS England and NHS Improvement PEW scores and systems standardisation plans have emphasised that a key challenge in standardisation will be for hospitals to let go of systems they have already developed. However, interviews suggest that the extent of this challenge depends on the acceptability of the current approaches in use within a region [Int. 1, Int. 3, Int. 5]. If existing systems are not highly valued, then de-implementing incumbent practices may be easier [Int. 1]. Conversely, when locally developed PEW systems that have been designed to benefit a specific cohort of children are already in place, de-implementation-related challenges may be accentuated [Int. 3, Int. 5].

Challenges notwithstanding, interviewees described a range of engagement activities that can support buy-in, including:

- **National engagement events that can provide a forum for discussing plans for standardisation** [Int. 5].

- **Targeted communication plans to encourage engagement across a geographical region** [Source: Communications with individuals involved with the development of NHS England and NHS Improvement PEW scores and systems standardisation plans]

- **Development of professional networks to engage with standardisation processes and enable regular communication and feedback on planning and progress:** In Northern Ireland, establishing a paediatric network across the area, along with regular meetings with commissioners and other managers, enabled progress [Int.1], including the iterative development of PEW forms [Int. 1]. However, communications with individuals involved with the development of NHS England and NHS Improvement PEW scores and systems standardisation plans highlighted that communication with some groups can be more challenging than with others. For example, it is acknowledged that further work is required to engage generalist paediatricians.

- **Education efforts relating to the need for standardisation** are perceived to be an important enabler of implementation [Int. 5]. One interviewee highlighted that early education helped develop an appetite for introducing the standardised PEW system in Scotland, noting that there were few barriers to its roll-out – despite the majority of hospitals having existing PEW charts in use [Int. 5].

- **Local champions of the standardisation effort:** For example, one interviewee highlighted the importance of having early adopters in district general hospitals and tertiary care centres, as well as larger teaching hospitals, 'because if you go to smaller boards or district general hospitals saying that the big ivory towers are using it and therefore you should use it, that's not going to work.' [Int. 5]. Communications with individuals involved with the development of NHS England and NHS Improvement PEW scores and systems standardisation plans have highlighted that involvement of local champions also matters, based on experience with the roll-out of an adult score and system. Having people involved who identify as patient safety champions can also facilitate discussion, helping to take views forward [Int. 5].

- **Piloting and staged roll-out are important considerations:** Piloting in different settings (and feeding back on the results) is considered to be particularly important to secure buy-in from across the range of organisations for which standardised PEW systems are planned [Int. 5]

**Staff involvement and engagement need to continue beyond the early phases of the standardisation effort, and be sustained throughout and beyond the implementation process, in order to promote the effective use of the standardised PEW system.** There is a need for continuing dialogue to sustain engagement and support effective implementation and use of a standardised system through time. One interviewee described how healthcare staff were encouraged to produce and share videos about how the PEW system is used locally so that, rather than an external body coming in and telling people how to do things, there is an

effort to facilitate a conversation on the front line about effective use [Int. 5].

**As is the case with the implementation of PEW scores and systems in general (i.e. regardless of whether they are standardised or not), contexts pursuing standardisation also need to consider the adaptability of the standardised system to different adoption settings (which may differ with respect to available resources, infrastructure or staffing models and the patient populations they serve):** The lack of resources can impede the roll-out of a standardised PEW system in some settings. For example, in remote or rural areas, units may lack the required paediatricians and paediatric facilities to support the standardised PEW system [Int. 5]. In addition, different staffing models across units mean that escalation plans specified at a national level may not be appropriate. For example, smaller units may not employ an individual in the role specified in the plan [Int. 1, Int. 5]. The escalation system needs to be broad enough to work across multiple settings, e.g. by escalating to certain levels of experience rather than individual roles [Int. 2]. A setting's infrastructure and capabilities may also affect its ability to adopt standardised PEW scores and systems. For units already using digital platforms, for instance, the introduction of a paper-based standardised system may present a particular challenge, depending on the degree to which paper-based PEW systems can be translated into digital contexts [Int. 1]. To recognise the need for adaptability, it has been proposed that a standardised PEW system should incorporate a common 'core' data set with the ability to add additional observations where required [10]. Other practical considerations also impact roll-out. For example, one interviewee described how issues with printing PEW charts led to delays in the roll-out of a standardised PEW system in Scotland [Int. 5].

**Extending the use of a standardised PEW system developed and rolled out in one clinical setting (such as the inpatient setting) to other settings (such as the ED or primary care) can require significant exploratory work.** For example, there is a need to consider which children would benefit, and balance this against the additional burden placed on already overstretched services. According to one interviewee, using a PEW chart with every child in a GP setting would not be feasible (taking 3 minutes out of a 10-minute appointment), but use in certain situations may be appropriate (e.g. out of hours practice) [Int. 5]. Availability of the required equipment for observations may also be an issue in the primary care setting, which impedes roll-out (i.e. if blood pressure cuffs for two-year olds are not available). This was illustrated by an interviewee: 'Initially we thought this would be brilliant, and quite a few GPs said "this would be great for us". But, actually further exploratory work has suggested that this will be a much bigger deal' [Int. 5].

## 5.3. Pros and cons of introducing a standardised PEW system, and evidence on the impact of standardisation

In a recent commentary piece, Chapman and Maconochie (2019) conclude that based on a review of the literature, it is uncertain whether a unified/standardised system delivers benefits in terms of outcomes or financial savings [1]. We found limited anecdotal evidence of the impact of PEW standardisation on patient outcomes in our scoping exercise. One expert we interviewed reported that, based on data available to decision-makers, the introduction of the standardised system was associated with a reduction in the number of cardiopulmonary arrests outside of intensive care in Scotland [Int. 5].

However, a number of observed process improvements (for identifying deteriorating children specifically and ensuring patient safety more broadly) have been shared by experts we interviewed. These include: ceasing data collection efforts that were no longer useful, thus reducing unnecessary data collection burden [Int. 5]; creating a better interface between people at the front line and patient safety teams [Int. 5]; reducing the need for staff to re-learn new processes and procedures when they move workplace [Int. 1, Int. 5]; and improved communication between healthcare professionals, with everyone speaking a common language [Int. 2, Int. 3].

# 6 Conclusion

Although a targeted and relatively small scoping exercise, this research has provided a diversity of insights on the use of PEW scores and systems as an improvement intervention. We considered influences on PEW system introduction, use and standardisation as they relate both to the 'technical' nature of a PEW score and system and the wider social, cultural and organisational context in which they are deployed.

At the outset, we mentioned that this scoping research was conducted to inform important issues for consideration in a potential fuller study that THIS Institute and RAND Europe are considering conducting. The insights in this scoping report may have some implications for future research on the process of standardising PEW scores and systems in England being proposed by NHS England and NHS Improvement, and the associated de-implementation of incumbent practices.

First, the attitudes to de-implementation of incumbent practices in healthcare are likely to be linked to the nature of the intervention being implemented and standardised, at least in part. Given the lack of evidence about the positive impact of PEW scores and systems on patient outcomes and the poor evidence evaluating different PEW scores and systems more generally, any study that seeks to learn from the experience of efforts to standardise PEW scores and systems should not generalise insights about the de-implementation process to other healthcare interventions. However, the learning gained from such a study could enrich our understanding of how the evidence base behind an intervention influences attitudes to and processes of de-implementation and standardisation.

Second, PEW systems are complex interventions with many interrelated components and dependencies between the intervention and the social and organisational context in which it is deployed. Thus, research that seeks to understand standardisation and de-implementation efforts associated with PEW scores and systems should be framed around an understanding of de-implementation and standardisation in the context of complex interventions The nature of standardisation and de-implementation efforts and the learning gained may be different for simpler interventions, where there are fewer constituent parts or less interdependency between the intervention and the adoption context.

Any effort to learn about de-implementation processes that might accompany the development and introduction of a standardised PEW system will also need to ensure a diversity of geographical and clinical settings and account for diverse patient populations. Understanding implementation fidelity of the standardised system, as well as scope for adaptability, is also likely to be

an important concern as it may influence which aspects of incumbent practices are de-implemented (or not) and the feasibility of meeting de-implementation needs.

There are undoubtedly other methodological and design considerations relating to sampling, recruitment of participant sites, stakeholder engagement, and relevant research methods that would need to be considered in any potential future research on de-implementing and standardising healthcare practices and learning from the example of PEW system-related efforts. We hope that this scoping study has laid a foundation for future research on the topic.

# References

1. Chapman, S.M. and I.K. Maconochie, *Early warning scores in paediatrics: an overview.* Archives of Disease in Childhood, 2019. 104(4): p. 395-399.

2. Pearson, G.A., *Why Children Die: A Pilot Study 2006; England (South West, North East and West Midlands), Wales and Northern Ireland*, G.A.E. Pearson, Editor. 2008, CEMACH: London.

3. Lambert, V.e.a., *A systematic literature review to support the development of a National Clinical Guideline – Paediatric Early Warning System (PEWS).* 2014, School of Nursing and Human Sciences (SNHS).

4. Jacob, N., et al., *Optimising Paediatric Afferent Component Early Warning Systems: A hermeneutic systematic literature review and model development.*

5. Smith, M.E.B., et al., *VA Evidence-based Synthesis Program Reports*, in *Early Warning System Scores: A Systematic Review.* 2014, Department of Veterans Affairs: Washington (DC).

6. Wynn, A., et al., *Sexually Transmitted Infections in Pregnancy and Reproductive Health: Proceedings of the STAR Sexually Transmitted Infection Clinical Trial Group Programmatic Meeting.* Sex Transm Dis, 2019.

7. Chapman, S.M., M.P. Grocott, and L.S. Franck, *Systematic review of paediatric alert criteria for identifying hospitalised children at risk of critical deterioration.* Intensive Care Medicine, 2010. 36(4): p. 600-11.

8. Chapman, S.M., et al., *Systematic review of paediatric track and trigger systems for hospitalised children.* Resuscitation, 2016. 109: p. 87-109.

9. Trubey, R., et al., *Validity and effectiveness of paediatric early warning systems and track and trigger tools for identifying and reducing clinical deterioration in hospitalised children: a systematic review.* BMJ Open, 2019. 9(5): p. e022105.

10. Roland, D., et al., *Use of paediatric early warning systems in Great Britain: has there been a change of practice in the last 7 years?* Archives of Disease in Childhood, 2014. 99(1): p. 26-9.

11. Roland, D., *Paediatric Early Warning Systems: myths and muses.* Paediatrics and Child Health, 2017. 27(5): p. 242-246.

12. Roland, D. and M. Sookee, *Towards a National Paediatric Early Warning System (PEWS): A principles and proposal document.* 2019, Royal College of Paediatrics and Child Health (RCPCH), Royal College of Nursing (RCN), NHS England/Improvement and clinicians from across the NHS.

13. Thomas-Jones, E., et al., *A prospective, mixed-methods, before and after study to identify the evidence base for the core components of an effective Paediatric Early Warning System and the development of an implementation package containing those core recommendations for use in the UK: Paediatric early warning system−utilisation and mortality avoidance−the PUMA study protocol.* BMC pediatrics, 2018. 18(1): p. 244.

14. Roland, D., *Paediatric Early Warning Scores: Holy Grail and Achilles' heel.* Archives of Disease in Childhood Education & Practice, 2012. 97(6): p. 208-15.

15. Challen, K. and D. Roland, *Early warning scores: a health warning.* Emergency Medicine Journal, 2016. 33(11): p. 812-817.

16. Parshuram, C.S., et al., *Effect of a Pediatric Early Warning System on All-Cause Mortality in Hospitalized Pediatric Patients: The EPOCH Randomized Clinical Trial.* Jama, 2018. 319(10): p. 1002-1012.

17. Gawronski, O., et al., *Qualitative study exploring factors influencing escalation of care of deteriorating children in a children's hospital.* BMJ Paediatrics Open, 2018. 2(1): p. e000241.

18. Joffe, A.R., N.R. Anton, and S.C. Burkholder, *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(5): p. 419-23.

19. Monaghan, A., *Detecting and managing deterioration in children.* Paediatr Nurs, 2005. 17(1): p. 32-5.

20. Oldroyd, C. and A. Day, *The use of pediatric early warning scores in the emergency department.* Journal of Emergency Nursing, 2011. 37(4): p. 374-6; quiz 424.

21. Roland, D., K. McCaffery, and F. Davies, *Scoring systems in paediatric emergency care: Panacea or paper exercise?* Journal of Paediatrics & Child Health, 2016. 52(2): p. 181-6.

22. Phillips, J.E., et al., *Assay Integrity of a PCR Influenza Point-of-Care Test Remains Following Artificial System Contamination.* J Appl Lab Med, 2019. 4(3): p. 422-426.

23. Mole, G. and M. Sookee, *Review of PEWS charts from 18 sites in England, Scotland, Wales and Australia.* 2019.

24. Cook, A., J. Wheway, and D. Roland, *PEWS Survey 2019.* 2019, NHS England and NHS Improvement.

25. Roland, D., *Beyond Mortality: Improving Outcomes for Children Who Deteriorate in Inpatient Settings.* J Hosp Med, 2019. 14(8): p. 512-513.