Nurses' judgments of patient risk of deterioration at change-of-shift handoff:

Agreement between nurses and comparison with early warning scores

Patrick Lavoie, RN, Ph. D.[a-b]
patrick.lavoie.1@umontreal.ca

Sean P. Clarke, RN, Ph. D., FAAN[c]
sc7372@nyu.edu

Christina Clausen, RN, Ph. D.[d-f]
christina.clausen@mail.mcgill.ca

Margaret Purden, RN, Ph. D.[d-e]
margaret.purden@mcgill.ca

Jessica Emed, RN, M.Sc. (A)[e-f]
jemed@jgh.mcgill.ca

Tanya Mailhot, RN, Ph. D.[a]
t.mailhot@umontreal.ca

Guillaume Fontaine, RN, Ph. D.[a-b]
guillaume.fontaine@umontreal.ca

Valerie Frunchak, RN, M. Sc. (A)[e-f]
vfruncha@jgh.mcgill.ca

a. Faculty of Nursing, Université de Montréal
   C.P. 6128, succ. Centre-Ville, Montreal, QC H3C 3J7, Canada
b. Montreal Heart Institute Research Center
   5000 rue Bélanger, Montreal, QC H1T 1C8, Canada
c. Rory Meyers College of Nursing, New York University
   433 First Avenue, New York, NY 10010, USA
d. Center for Nursing Research, Jewish General Hospital, Montreal, Canada
   3755 ch. Côte-Ste-Catherine, Montreal, QC H3T 1E2, Canada
e. Ingram School of Nursing, McGill University, Montreal, Canada
   680 Sherbrooke West #1800, Montreal, QC H3A 2M7, Canada
f. Department of Nursing, Jewish General Hospital, Montreal, Canada
   3755 ch. Côte-Ste-Catherine, Montreal, QC H3T 1E2, Canada

Corresponding author:

Patrick Lavoie, patrick.lavoie.1@umontreal.ca

Acknowledgement:

This study was funded by the McGill Nursing Collaborative for Education and Innovation in Patient- and Family-Centered Care. During the study, the first author received a postdoctoral fellowship from the William F. Connell School of Nursing at Boston College.

ABSTRACT

Background: Nurses begin forming judgments regarding patients' clinical stability during change-of-shift handoffs.

Objectives: To examine the agreement between incoming and outgoing nurses' judgments of deterioration risk following handoff and compare these judgments to commonly used early warning scores (MEWS, NEWS, ViEWS).

Methods: Following handoffs on three medical/surgical units, nurses completed the Patient Acuity Rating. Nurse ratings were compared with computed early warning scores based on clinical data. In follow-up interviews, nurses were invited to describe their experiences of using the rating scale.

Results: Sixty-two nurses carried out 444 handoffs for 158 patients. While the agreement between incoming and outgoing nurses was fair, correlations with early warning scores were low. Nurses struggled with predicting risk and used their impressions of differential risk across all the patients to whom they had been assigned to arrive at their ratings.

Conclusion: Nurses shared information that influenced their clinical judgments at handoff; not all of these cues may necessarily be captured in early warning scores.

Keywords: clinical judgment, nursing handoff, patient deterioration, early warning scores

INTRODUCTION[1]

In acute care settings, nurses are expected to determine whether patients are stable or are deteriorating—i.e., if patients are experiencing changes that could lead to a cardiac arrest or an unplanned transfer to an intensive care unit (ICU).[1] Nurses monitor patients and interpret the data they gather to arrive at clinical judgments and set priorities in care. The term 'clinical judgment' refers to a nurse's understandings or conclusions about a patient's health, care needs, or concerns.[2] Nurses begin to form clinical judgments the moment they accept responsibility for their assigned patients during a process known as 'handoff,' which generally occurs at the beginning of a scheduled shift and consists of the exchange of information regarding one (or more) patient(s)[3]. Typically, one nurse presents details to the colleague who will oversee the patient's care on the following shift. In handoffs, nurses share crucial information regarding the condition of their patients and develop a shared picture of patient needs and priorities in care.

Although there is much published research on nursing handoffs[4, 5], little is known about how handoffs relate to nurses' clinical judgments. Broadly speaking, research has shown that when presented with the same patient data, nurses will typically arrive at different judgments depending on factors such as their knowledge and experience and on characteristics of the information communicated (e.g., complexity, ambiguity, quantity).[6-10] To safeguard against potential errors—i.e., failure to appreciate when a patient requires urgent care—and help members of the healthcare team, including nurses, identify deteriorating patients, early warning scoring systems have been developed. Based on studies of patients' trajectories before in-hospital cardiac arrests or unplanned admissions to an ICU,[11, 12] these tools first assign scores to deviations in various clinical

[1] COPD: Chronic Obstructive Pulmonary Disorder. ICU: Intensive Care Unit; MEWS: Modified Early Warning Score; NEWS: National Early Warning Score; PAR: Patient Acuity Rating; ViEWS: ROC: Receiver Operating Characteristic; VitalPAC™ Early Warning Score.

parameters—e.g., each of the vital signs, level of consciousness, and indicators of oxygenation. The scores are summed, and the total is used to determine whether the patient's physiological instability suggests a need to investigate causes and possibly escalate treatment.

While research regarding the prediction of patient deterioration continues,[13-15] the Modified Early Warning Score (MEWS),[16] the National Early Warning Score (NEWS),[17] and the VitalPAC™ Early Warning Score (ViEWS)[18] remain the most commonly used and studied early warning scores.[19,20] As shown in Table 1, they consider different parameters. One obvious difference is that the NEWS and ViEWS include oxygen saturation and use of supplemental oxygen, whereas the MEWS does not. Although optimal cutoff scores have not been established, a MEWS of 3-4 or above, a NEWS of 5-6 or above, and a ViEWS of 5 or above have shown good predictive value for adverse outcomes in hospitalized patients, including cardiac arrest and death within 24 to 48 hours.[19, 20]

Despite wide uptake and current use, some nurses are reluctant to use early warning scores and are dubious about their value.[21-24] This may be explained by nurses' use of criteria beyond vital sign abnormalities (that are difficult to score objectively and thus are absent from early warning scoring schemes) to make judgments regarding a patient's risk of deterioration. Among these is a sense of 'worry or concern' on a nurse's part—an intuitive feeling that something is wrong with a patient. A systematic review has linked subjective impressions of risk with ten indicators of patient deterioration: changes in breathing, changes in circulation, body temperature, impaired mentation, agitation, pain, a failure of a patient to improve/progress, patient sensations/reports, subjective nurse observations (e.g., patient looks unwell), and nurse "intuition" (i.e., knowing without a rationale[25]). It is important to note that these indicators were identified in studies with very heterogeneous designs that were conducted in a variety of settings; the utility of such indicators across patient populations is therefore unknown. Nevertheless, studies have shown that some of

these cues appear before any vital sign abnormalities are observable[26] and that they have shown significant associations with mortality and ICU admissions.[27, 28]

Nurses do not appear to base judgments of patient risk of deterioration solely on objective clinical criteria. As discussed, nurses begin forming impressions of stability or risk of deterioration at the time they receive a handoff. These impressions play a critical role in adjusting the monitoring of patients, can trigger rescue interventions, and shape communication with other professionals, all of which may prevent cardiac arrests or unplanned transfers to the ICU.[29] Nonetheless, little research has examined how handoffs relate to nurse clinical judgments.

The purpose of this study was to examine acute care nurses' judgments of patient risk of deterioration following a change-of-shift handoff. Specifically, we examined the degree of agreement between nurses in their judgments of stability/risk and compared these judgments to "objective" numerical ratings of risk reflected in commonly used early warning scores (MEWS, NEWS, ViEWS). In addition, this study explored nurses' experiences of using a rating scale to express their judgments of patient risk of deterioration.

METHODS

Participants and Setting

Data for this study were drawn from a larger study of the relationship between nurses' handoffs and judgments of patient risk of deterioration, the design of which has been detailed elsewhere.[30] Data in this prospective descriptive correlational study were collected from acute care nurses on one surgical unit (A) and two medical units (B and C) at a tertiary acute care bilingual (English and French) university-affiliated hospital in Montreal, Canada from September to December 2017. In fiscal year 2016-2017, the units received 884, 805, and 888 admissions, and mean lengths of stay were 8.2, 14.2, and 12.3 days, respectively. The most common reasons for admission were bowel obstruction, gynecologic neoplasm, and cholecystitis on Unit A, pneumonia, septicemia, and

congestive heart failure on Unit B, and chemotherapy, neutropenia, and pneumonia on Unit C. Of note, early warning scoring systems had not been implemented on participating units at the time of the study.

A convenience sampling strategy was used. All nurses from these units were eligible to participate in the study if they worked at least two shifts during the data collection period. Enrollment was voluntary and informed consent was obtained from all participants. After recruitment, the research team screened unit schedules for four consecutive weeks to select ten days when the probability of being able to record handoffs from nurses who agreed to participate was the highest. Before data collection began, the institution's Research Ethics Committee approved the protocol.

Data Collection

Upon enrollment, nurses completed a sociodemographic questionnaire (age, gender, first language, work status, work experience, highest earned degree). Data collection was undertaken on the days when participating nurses were most heavily represented among the scheduled staff. When the incoming and outgoing nurses caring for a particular patient had both provided consent for the study, nurses were asked to record their handoffs. Immediately after the handoff, both nurses completed separate questionnaires that included a question about their judgment of the particular patient's risk of deterioration using the Patient Acuity Rating (PAR).[31] The PAR is a 7-point scale used to quantify judgments of a patient risk of deterioration where the respondent indicates the likelihood of the patient being transferred to an ICU or experiencing a cardiac arrest within the next 24 hours (from 1-Extremely unlikely to 7-Extremely likely). In a validation study of the PAR with 140 physicians, nurse practitioners, and physician assistants who rated 1663 patients,[31] the area under the receiver operating characteristic (ROC) curve ranged from 0.69 to 0.85, which the authors interpreted as suggesting good accuracy for predicting ICU transfers and cardiac arrests.

Weighted kappa ranged from 0.32 to 0.43 for the same patients, which the authors interpreted as suggestive of moderate inter-rater reliability. For this study, nurses were asked to complete the PAR individually and avoid discussing their ratings with their colleagues.

Subsequently, research assistants collected the most recent clinical parameters before the handoff from the patients' medical records in order to calculate three early warning scores: a MEWS, a NEWS, and a ViEWS (see Table 1). After the data collection period, sociodemographic data for patients handed off during the study (age, sex, length of stay, discharge destination) and cardiac arrests, ICU information about patients who had deteriorated on the units during the study (i.e., cardiac arrests, ICU transfers, and deaths) were obtained from hospital clinical information systems. Furthermore, we identified all cardiac arrest calls ('code blues') on the units from the paging system records—since the hospital did not have a rapid response team at the time of the study, the code blue team responded to all medical emergencies, even if they did not involve cardiac arrests or transfers to the ICU.

Upon completion of the study, all nurses who recorded a handoff were invited to focus groups to comment on a number of topics, including their experience of expressing their judgments of patient risk of deterioration using the PAR. Focus group discussions were audio-recorded and transcribed verbatim.

Data Analysis

Sociodemographic data regarding nurses and patients, as well as clinical characteristics of the patients, were summarized using descriptive statistics (means and standard deviations, median and interquartile range, or frequencies and percentages). For each handoff, the following data were summarized: two ratings of the patient risk of deterioration—one from the outgoing nurse giving handoff (PAR_{OUT}) and one from the incoming nurse receiving handoff (PAR_{IN})—and three early warning scores (MEWS, NEWS, and ViEWS). All incidents for the patients in the 24 hours

following the handoffs (specifically, calls to the 'code blue' team, cardiac arrests, deaths, and ICU transfers) were also tabulated.

Agreement between outgoing and incoming nurses in their risk judgments for the same patients was examined with descriptive statistics (average difference between PAR_{OUT} and PAR_{IN}) as well as weighted Kappa statistics, Pearson correlations and intraclass correlation coefficients (ICC; two-way random model for absolute agreement). We used *t*-tests to determine if nurses' level of agreement (the difference between PAR_{OUT} and PAR_{IN}) varied whether nurses had the same first language or not and whether they held diplomas or university degrees. Multiple linear regression models were used to assess whether nurse experience predicted level of agreement in risk ratings. For all tests, the significance level was set at 0.05. Pearson correlations were used to assess the associations between nurses' risk ratings and early warning scores. Analyses were conducted using IBM® SPSS® Statistics version 24.0.

Focus group transcripts were subjected to a thematic analysis[32]. Meaningful units (words and phrases) related to nurses' experience of rating patients' risk of deterioration using the PAR were coded and categorized. Within the categories, codes were arranged into themes that reflected nurses' accounts. An audit trail was built throughout the analysis, which was conducted by two researchers (PL, TM). Focus group data were managed using Microsoft Word©.

RESULTS

Participants and Patients

Over the course of the study, 62 out of 108 eligible nurses on the three study units (57.4%) carried out 444 handoffs for 158 patients; on average, each patient appeared 2.9 (± 2.1) times in the dataset. Nurses' characteristics are presented in Table 2 and patients' characteristics are presented in Table 3. Nurses were mostly female, held bachelor's degrees as their highest credentials in nursing, and had under five years of experience on average. As shown in Table 3,

the most frequent patients' diagnoses were similar to the most frequent reasons for admission on the units. Patient lengths of stay were not normally distributed. Approximately 10% of each unit's patients were outliers with long stays, which was reflected by higher mean lengths of stay (respectively, 26.1 (± 27.6), 18.2 (± 14.7), and 28.7 (± 18.2) days) in comparison with unit means for the preceding year.

Nurses' Judgments and Early Warning Scores

Across the 444 handoffs, nurses' judgments of patient risk of deterioration (PAR) averaged 2.9 ± 1.6 on a seven-point scale, indicating that nurses judged that, on average, patients were unlikely to experience a cardiac arrest or a transfer to the ICU in the next 24 hours. In total, there were 125 handoffs (28.2%) where at least one nurse judged that the patient was at high risk of deterioration (PAR ≥ 5): 53 (30.6%), 43 (23.5%), and 29 (33.0%) on Units A, B, and C, respectively. Patients' early warning scores generally fell below the thresholds designated as warranting escalation of care: the average MEWS was 1.5 ± 0.9, the average NEWS was 2.1 ± 2.1, and the average ViEWS was 2.3 ± 2.3. Table 4 presents summary statistics by unit.

Level of Agreement

Across the 444 handoffs, the agreement between nurses for the same patients was fair to moderate: weighted Kappa=0.27, r=0.38 (p<0.001), ICC for single measures=0.38 (95% CI: 0.29-0.45). On average, incoming nurses' PARs were 0.19 points higher than outgoing nurses' PARs. In 95% of handoffs, the incoming nurse's PAR was within 3.25 points below or 3.63 points above the outgoing nurse's.

There was no difference in nurses' level of agreement whether they had the same first language (M=-0.24 ± 1.76) or not (M=-0.17 ± 1.76) (t(442)=-0.374, p=0.71). There was no difference in level of agreement whether the outgoing nurse had a university degree (M=-0.28 ± 1.64) or not

(M=-0.02 ± 2.00), or whether the incoming nurse had a university degree (M=-0.14 ± 1.75) or not (M=-0.36 ± 1.78) ($t(442)$=-1.641, p=0.10 and t(442)=1.137, p=0.26, respectively). Multiple linear regression models found no evidence of associations between either outgoing or incoming nurses' years of experience with level of agreement ($F(2, 441)$=2.85, p=0.06, R^2=0.01).

The correlations between nurses' judgments on the PAR and patients' early warning scores were low; all were between 0.20 and 0.22 (p<0.001). Statistics by unit are presented in Table 4. Of the 158 patients involved in the study, none experienced a cardiac arrest team call ('code blue'), a cardiac arrest, or death in the 24 hours following a recorded handoff. Only one patient from Unit B who presented signs of sepsis was transferred to the ICU. Before transfer, both nurses assigned the patient a PAR of 6; the patient's MEWS was 4, and both the NEWS and ViEWS were 12.

Nurses' Experience of Rating Patient Risk of Deterioration

In total, 16 focus groups with 44 participants were held. Thematic analysis of the transcripts revealed that expressing their judgments of patient risk of deterioration using a rating scale made sense to most nurses, even if it was not an exercise to which they were accustomed. Three themes were identified from the focus group data: 1) nurses' discomfort with probabilities, 2) a tendency to use comparisons across patients to rate risk of deterioration, and 3) unpredictability of deterioration in medical (versus surgical) patients. In the first theme, nurses appeared uncomfortable with the concept of probability; they reported that the scale had too many levels and that the difference between levels was subjective, which made it difficult to rate patients' risk: "What's a 1? What's a 3? What's a 7? We are not statisticians…" (Participant from Unit A). To overcome this problem, nurses reported that they ranked patients in their assignments on a particular shift from the least at risk to the most at risk and assigned ratings accordingly. This reasoning process refers to the second theme identified (i.e., comparative judgments).

In terms of the third theme, some participants from medical units (B and C) doubted that a cardiac arrest or transfer to the ICU was predictable. According to participants on Unit B, all patients were at risk of deterioration by virtue of being hospitalized; nurses felt that anything could happen to patients at any time, even when they considered patients to be "stable." Participants from Unit C believed that deterioration of hematology/oncology patients is unpredictable, and that patient deterioration would sometimes surprise nurses. Nurses from Unit C believed that relative to medical patients it was easier to predict whether surgical patients would deteriorate.

DISCUSSION

In this study, we examined nurses' judgments of patient risk of deterioration at change-of-shift handoffs. Across the 444 handoffs, nurses' level of agreement was slightly stronger agreement than found in previous studies that used the PAR with internal medicine residents,[33] and slightly weaker than the agreement found in the original PAR validation study with physicians, nurse practitioners, and physician assistants.[31] Bearing in mind the limited points of comparison available in the literature, these results suggest that overall, nurses' level of agreement in ratings of the same patients between each other in this study was comparable to agreement in ratings by other health professionals in earlier research.

There were important differences in agreement between nurse ratings at the unit/specialty level, however. Agreement between nurses was higher on Units A and B than on Unit C, which could be attributable in part to the smaller sample size on Unit C. More importantly, the correlation of nurses' judgments with early warning scores (especially the NEWS and ViEWS) was stronger on surgical Unit A than on medical Units B and C. This suggests that surgical nurses used criteria more similar to those of early warning scores to determine if patients are at risk of deterioration—i.e., abnormal vital signs, level of consciousness, oxygen saturation, and supplemental oxygen—while medical nurses appeared to rely on different criteria. This is an interesting finding,

considering that early warning scores have been validated in both medical and surgical populations[20, 34] and that the NEWS was originally designed for medical patients.[19] Moreover, focus group interviews revealed that medical nurses doubted the predictability of patient deterioration, which echoes previous research showing that nurses sometimes doubt the value of early warning scores.[21-24] Medical nurses' risk ratings showed greater agreement with each other than with early warning scores. This suggests that medical nurses rely on a common set of cues to detect patients at risk of deterioration and further, that these cues may not be captured by early warning scores. Further investigation of the content of the handoffs on medical units might reveal more about the nature of those cues, which may resemble the subjective indicators of patient deterioration that were identified in previous research[25] or more population-specific criteria.

Contrary to expectations, nurses' level of agreement did not differ depending on their years of experience and educational level. This was surprising, considering that experience and knowledge are commonly thought of as linked with skilled clinical judgments in nurses[2] and other health professionals.[35] However, it should be noted that this study used a convenience sample of nurses and was neither designed to obtain a full range of nurses' experience and education levels or to sample handovers to detect such differences. Nevertheless, the place of education and experience in judgment merits further explorations.

Another interesting point was nurses' struggle with the concept of probability when predicting patients' risk of deterioration. In the focus groups, nurses discussed their uncertainties regarding the levels of the PAR scale in terms of what they meant and how they differed from one another. The fact that nurses ranked their assigned patients against each other to guide their ratings of risk showed that they engaged in comparisons and were more comfortable thinking about patients' relative risk (compared to other patients) than absolute risk. While comparative thinking is a known feature of human judgment, it is also known that the basis used for comparison—in this case a

nurse's assignment—affects judgment[36]. For example, the same patient might be judged more at risk if other patients in a nurse's assignments are at low-risk and will be judged less at risk if other patients are at high-risk. This contrast effect entails that the exact value of nurses' PAR may not be as informative as the position of the rating above or below the neutral point on the scale. It is also important to keep in mind that nurses' judgments were 3.25 points below or 3.63 points above each other's in 95% of cases, which is equivalent to the entire range of the scale.

This study has several strengths and limitations. In terms of strengths, three units with different specialties were involved, the majority of eligible nurses agreed to participate, and data collection was prospective in nature. In terms of limitations, this was a single-center study and the days of data collection were neither sequential, nor selected at random. The markedly longer mean lengths of stay for patients in the study (in comparison with unit mean lengths of stay for the preceding year) can be attributed in part to the sampling strategy for handoffs used—because data were collected on ten days over a four-week period, patients with longer lengths of stay had more opportunities to be the subjects of recorded handoffs than the average patient. Potentially confounding variables that might influence risk of patient deterioration (and its assessment) such as acuity of care needs, multiple comorbidities, and polypharmacy were not accounted for in the study design and should be considered in future research. The statistical methods employed did not account for the nested structure of the data; collecting a larger dataset would allow use of more appropriate multilevel statistical models. The small number of incidents (code blue, cardiac arrest, death, ICU transfer) observed in our sample precluded the analysis of the predictive power of nurses' clinical judgments (such as ROC curve analysis)—clearly, future studies seeking to examine prediction of events must be designed using accurate estimates of prevalence of adverse outcomes. Even though the only event that occurred during the study was predicted by both nurses, it is important to note that one or both nurses judged that the patient was at risk of deterioration in

approximately one out of four handoffs. Considering the severity of adverse events that may result from patient deterioration, it might be desirable for nurses to maintain a higher index of suspicion with respect to patient risk of deterioration. However, our results are in no way indicative of the predictive value of the PAR when completed by nurses and further studies should be conducted in that respect.

Conclusion

To our knowledge, this was the first study to employ the PAR as a measure of judgment of patient risk of deterioration in nurses. Although we were not able to assess the predictive power of nurses' judgments of patient risk of deterioration, results show that nurse's ratings showed higher interrater agreement between themselves than agreement with common early warning scores. This is consistent with the notion that nurses share information with each other that influences clinical judgments of patients' risk of deterioration at handoff and that not all of these cues may be captured by early warning scores. This idea should be considered in future research on rapid response systems and nurses' roles within them.

REFERENCES

1 Lavoie P, Pepin J, Alderson M. Defining patient deterioration through acute care and intensive care nurses' perspectives. *Nurs Crit Care;* 2016;21:68-77. https://doi.org/10.1111/nicc.12114

2 Tanner CA. Thinking like a nurse: A research-based model of clinical judgment in nursing. *J Nurs Educ*; 2006;45:204-11. https://doi.org/10.3928/01484834-20060601-04

3 Cohen MD, Hilligoss PB. The published literature on handoffs in hospitals: Deficiencies identified in an extensive review. *Qual Saf Health Care*; 2010;19:493-7. https://doi.org/10.1136/qshc.2009.033480

4 Staggers N, Blaz JW. Research on nursing handoffs for medical and surgical settings: An integrative review. *J Adv Nurs*; 2013;69:247-62. https://doi.org/10.1111/j.1365-2648.2012.06087.x

5 Riesenberg LA, Leitzsch J, Cunningham JM. Nursing handoffs: A systematic review of the literature. *Am J Nurs*; 2010;110:24-34. https://doi.org/10.1097/01.naj.0000370154.79857.09

6 Ericsson KA, Whyte J, Ward P. Expert performance in nursing: Reviewing research on expertise in nursing within the framework of the expert-performance approach. *Adv Nurs Sci*; 2007;30:E58-E71. https://doi.org/10.1097/00012272-200701000-00014

7 Hammond KR. Human judgment and social policy: Irreducible uncertainty, inevitable error, unavoidable injustice. New York, NY: Oxford University Press; 1996.

8 Stamp KD. How nurse practitioners make decisions regarding coronary heat disease risk: A social judgment analysis. *Int J Nurs Knowl*; 2012;23:29-40. https://doi.org/10.1111/j.2047-3095.2011.01196.x

9 Thompson C, Bucknall T, Estabrookes CA, Hutchinson A, Fraser K, de Vos R, et al. Nurses' critical event risk assessments: A judgement analysis. *J Clin Nurs*; 2007;18:601-12. https://doi.org/10.1111/j.1365-2702.2007.02191.x

10 Yang H, Thompson C. Capturing judgement strategies in risk assessments with improved quality of clinical information: How nurses' strategies differ from the ecological model. *BMC Med Inform Decis Mak*; 2016;16:7. https://doi.org/10.1186/s12911-016-0243-1

11 Buist MD, Jarmolowski E, Burton PR, Bernard SA, Waxman BP, Anderson J. Recognising clinical instability in hospital patients before cardiac arrest or unplanned admission to intensive care: A pilot study in a tertiary-care hospital. *Med J Aust*; 1999;171:22-5. https://doi.org/10.5694/j.1326-5377.1999.tb123492.x

12 Hodgetts TJ, Kenward G, Vlackonikolis I, Payne S, Castle N, Crouch R, et al. Incidence, location and reasons for avoidable in-hospital cardiac arrest in a district general hospital. *Resuscitation*; 2002;54:115-23. https://doi.org/10.1016/s0300-9572(02)00098-9

13 Churpek MM, Adhikari R, Edelson DP. The value of vital sign trends for detecting clinical deterioration on the wards. *Resuscitation*; 2016;102:1-5. https://10.1016/j.resuscitation.2016.02.005

14 Jeffery AD, Dietrich MS, Fabbri D, Kennedy B, Novak LL, Coco J, et al. Advancing in-hospital clinical deterioration prediction models. *Am J Crit Care*; 2018;27:381-391. https://doi.org/10.4037/ajcc2018957

15 Kipnis P, Turk BJ, Wulf DA, LaGuardia JC, Liu V, Churpek MM, et al. Development and validation of an electronic medical record-based alert score for detection of inpatient deterioration outside the ICU. *J Biomed Inform*; 2016;64:10-19. https://doi.org/10.1016/j.jbi.2016.09.013

16 Subbe C, Kruger M, Rutherford P, Gemmel L. Validation of a modified Early Warning Score in medical admissions. *QJM*; 2001;94:507-10. https://doi.org/10.1093/qjmed/94.10.521

17 Royal College of Physicians. National Early Warning Score (NEWS): Standardising the assessement of acute-illness severity in the NHS. London: Royal College of Physicians; 2012.

18 Prytherch DR, Smith GB, Schmidt PE, Featherstone PI. ViEWS—Towards a national early warning score for detecting adult inpatient deterioration. *Resuscitation*; 2010;81:932-7. https://doi.org/10.1016/j.resuscitation.2010.04.014

19 Downey CL, Tahir W, Randell R, Brown JM, Jayne DG. Strengths and limitations of early warning scores: A systematic review and narrative synthesis. *Int J Nurs Stud*; 2017;76:106-19. https://doi.org/10.1016/j.ijnurstu.2017.09.003

20 Smith MEB, Chiovaro JC, O'Neil M, Kansagara D, Quinones AR, Freeman M, et al. Early warning system scores for clinical deterioration in hospitalized patients: A systematic review. *Ann Am Thorac Soc*; 2014;11:1454-65. https://doi.org/10.1513/annalsats.201403-102oc

21 Braaten JS. (2015). Hospital system barriers to rapid response team activation: A cognitive work analysis. *Am J Nurs*; 2015;115:22-33. https://doi.org/10.1097/01.NAJ.0000460673.82070.af

22 Donohue LA, Endacott R. Track, trigger and teamwork: Communication of deterioration in acute medical and surgical wards. *Intensive Crit Care Nurs*; 2010;26:10-7. https://doi.org/10.1016/j.iccn.2009.10.006

23 Kitto S, Marshall SD, McMillan SE, Shearer B, Buist M, Grant R, et al. Rapid response systems and collective (in)competence: An exploratory analysis of intraprofessional and interprofessional activation factors. *J Interprof Care*; 2015;29:340-6. https://doi.org/10.3109/13561820.2014.984021

24 Odell M. Detection and management of the deteriorating ward patient: An evaluation of nursing practice. *J Clin Nurs*; 2015;24:173-82. https://doi.org/10.1111/jocn.12655

25 Douw G, Schoonhoven L, Holwerda T, Huisman-de Waal G, van Zanten AR, van Achterberg T, et al. Nurses' worry or concern and early recognition of deteriorating patients on general wards

in acute care hospitals: A systematic review. *Crit Care*; 2015;19:230. https://doi.org/10.1186/s13054-015-0950-5

26 Hodgetts TJ, Kenward G, Vlachonikolis IG, Payne S, Castle N. The identification of risk factors for cardiac arrest and formulation of activation criteria to alert a medical emergency team. *Resuscitation*; 2002;54:125-31. https://doi.org/10.1016/s0300-9572(02)00100-4

27 Jacques T, Harrison GA, McLaws M-L, Kilborn G. Signs of critical conditions and emergency responses (SOCCER): A model for predicting adverse events in the inpatient setting. *Resuscitation*; 2006;69:175-83. https://doi.org/10.1016/j.resuscitation.2005.08.015

28 Buist MD, Moore GE, Bernard SA, Waxman BP, Anderson JN, Nguyen TV. Effects of a medical emergency team on reduction of incidence of and mortality from unexpected cardiac arrests in hospital: Preliminary study. *Br Med J*; 2002;324:1-6. https://doi.org/10.1136/bmj.324.7334.387

29 Massey D, Chaboyer W, Anderson V. What factors influence ward nurses' recognition of and response to patient deterioration? An integrative review of the literature. *Nurs Open*; 2017. https://doi.org/10.1002/nop2.53

30 Lavoie P, Clarke SP, Clausen C, Purden M, Emed J, Mailhot T, et al. Acceptability and feasibility of recruitment and data collection in a field study of hospital nurses' handoffs using mobile devices. *Pilot Feasibility Stud*; 2018;4:163. https://doi.org/10.1186/s40814-018-0353-x

31 Edelson DP, Retzer E, Weidman EK, Woodruff J, Davis AM, Minsky BD, et al. Patient acuity rating: Quantifying clinical judgment regarding inpatient stability. *J Hosp Med*; 2011;6:480-8. https://doi.org/10.1002/jhm.886

32 Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol*; 2006;3:77-101. https://doi.org/10.1191/1478088706qp063oa

33 Ratelle JT, Kelm DJ, Halvorsen AJ, West CP, Oxentenko AS. Predicting and communicating risk of clinical deterioration: An observational cohort study of internal medicine residents. *J Gen Intern Med*; 2015;30:448-53. https://doi.org/10.1007/s11606-014-3114-4

34 Kovacs C, Jarvis SW, Prytherch DR, Meredith P, Schmidt PE, Briggs JS, et al. Comparison of the National Early Warning Score in non-elective medical and surgical patients. *Br J Surg*; 2016;103:1385-93. https://doi.org/10.1002/bjs.10267

35 Spengler PM, White MJ, Ægisdóttir S, Maugherman AS, Anderson LA, Cook RS, et al. The meta-analysis of clinical judgment project effects of experience on judgment accuracy. *Couns Psychol*; 2009;37:350-99. https://doi.org/10.1177/0011000006295149

36 Mussweiler T. Comparison processes in social judgment: Mechanisms and consequences. *Psychological Rev*; 2003;110:472-89. https://doi.org/10.1037/0033-295x.110.3.472

Table 1. Early warning scores

	3	2	1	0	1	2	3
MEWS[16]							
Systolic blood pressure	<70	71-80	81-100	101-199		≥200	
Heart rate		<40	41-50	51-100	101-110	111-129	≥130
Respiratory rate		<9		9-14	15-20	21-29	≥30
Temperature (°C)		<35		35-38.4		≥38.5	
AVPU score				Alert (A)	Voice (V)	Pain (P)	Unresponsive (U)
NEWS[17]							
Systolic blood pressure	≤90	91-100	101-110	111-219			≥220
Heart rate	≤40		41-50	51-90	91-110	111-130	≥131
Respiratory rate	≤8		9-11	12-20		21-24	≥25
Temperature (°C)	≤35.0		35.1-36.0	36.1-38.0	38.1-39.0	≥39.1	
AVPU score				A			V, P, U
Oxygen saturation	≤91	92-93	94-95	≥96			
Supplemental oxygen		Yes		No			
ViEWS[18]							
Systolic blood pressure	≤90	91-100	101-110	111-249	≥250		
Heart rate		≤40	41-50	51-90	91-110	111-130	≥131
Respiratory rate	≤8		9-11	12-20		21-24	≥25
Temperature (°C)	≤35.0		35.1-36.0	36.1-38.0	38.1-39.0	≥39.1	
AVPU score				A			V, P, U
Oxygen saturation	≤91	92-93	94-95	≥96			
Supplemental oxygen				Air			Any O_2

Table 2. Sociodemographic characteristics of the nurse participants (*N*=62)

	n (%) or M (SD)
Age (years)[1]	30.8 (6.7)
Gender (female)[2]	50 (80.6)
First language[2]	
English	28 (45.2)
French	24 (38.7)
Other	10 (16.1)
Full-time[2]	34 (54.8)
Nursing experience (years)[1]	4.7 (4.1)
Experience on unit (years)[1]	3.6 (3.3)
Highest degree[2]	
Diploma	16 (25.8)
Bachelor's	42 (67.7)
Master's	4 (6.5)

NOTE. [1]Means (standard deviations). [2]Numbers of participants (percentages).

Table 3. Sociodemographic characteristics, most frequent diagnoses, length of stay, and discharge destination for patients handed off during the study (*N*=158)

	Unit A (*n*=53)	Unit B (*n*=67)	Unit C (*n*=38)
Age (years)[1]	66.4 (18.5)	75.4 (14.4)	65.9 (15.2)
Gender (female)[2]	32 (60.4)	31 (46.3)	13 (34.2)
Most frequent diagnoses[2]			
Bowel obstruction	7 (13.2)		
Intestinal resection	6 (11.3)		
Ileostomy closure	3 (5.7)		
Hernia repair	3 (5.7)		
Septicemia		7 (10.4)	
Pneumonia		5 (7.5)	3 (7.9)
Pleural effusion		4 (6.0)	
COPD exacerbation		3 (4.5)	
Seizure or convulsion		3 (4.5)	
Weakness or fatigue		3 (4.5)	
Chemotherapy			8 (21.1)
Fever			3 (7.9)
Leukemia			3 (7.9)
Length of stay (days)[3]	18.2 (44.8)	13.9 (23.3)	31.2 (27.7)
Without outliers	13.2 (27.6)	12.7 (19.8)	29.5 (28.1)
Outliers (*n*)[4]	6	7	4
Discharge destination[2]			
Home	33 (62.3)	29 (43.3)	20 (52.6)
Long-term care	10 (18.9)	20 (31.7)	3 (7.9)
Remained in hospital at 2 months after study	1 (1.9)	2 (3.0)	1 (2.7)
Death	2 (3.8)	12 (17.9)	6 (15.8)
Unknown	7 (13.2)	4 (6.0)	8 (21.1)

NOTE. [1]Mean years (standard deviations) and exclude outliers. [2]Numbers of patients (percentages). [3]Medians (interquartile range). [4]Includes patient still hospitalized two months after the end of the study and patients whose length of stay exceeded their unit's third quartile by 1.5 interquartile range.

Table 4. Early warning scores, nurses' judgments of risk, agreement between nurses' judgments, and correlations of risk judgments with early warning scores

	Unit A Surgical (*n*=173)	Unit B Medical (*n*=183)	Unit C Medical (*n*=88)
Early warning scores[1]			
MEWS	1.5 (0.9)	1.6 (0.9)	1.4 (0.6)
NEWS	1.8 (1.8)	2.7 (2.4)	1.6 (1.9)
ViEWS	1.9 (2.0)	2.9 (2.6)	1.8 (2.1)
Nurses' judgments[1]			
PAR_{OUT}	2.6 (1.6)	2.8 (1.5)	3.1 (1.5)
PAR_{IN}	3.0 (1.7)	3.0 (1.5)	3.2 (1.6)
Nurses' agreement			
Kappa	0.31	0.28	0.15
PAR_{OUT}-PAR_{IN} (r)	0.41**	0.39**	0.26*
ICC (95% CI)	0.41 (0.28-0.53)**	0.39 (0.26-0.51)**	0.25 (0.05-0.44)*
Correlations			
PAR_{OUT}-MEWS	0.30**	0.20**	0.08
PAR_{IN}-MEWS	0.22*	0.20**	0.16
PAR_{OUT}-NEWS	0.38**	0.15*	0.13
PAR_{IN}-NEWS	0.35**	0.15*	0.19
PAR_{OUT}-ViEWS	0.40**	0.13	0.16
PAR_{IN}-ViEWS	0.37**	0.14	0.19

NOTE. [1]Mean years (standard deviations). *Significant at the p <0.05 level. **Significant at the p <0.001 level.