Development and validation of a methodology to measure time involved in vital signs observations by hospital nurses

Chiara Dall’Ora; PhD, MSc, RN; Research Fellow in Nursing Workforce, University Of Southampton; School of Health Sciences, University Road, Southampton, SO17 1BJ; [C.Dallora@soton.ac.uk](mailto:C.Dallora@soton.ac.uk); +44(0)23 80598249

Jo Hope; PhD, MSc, BSc; Research Fellow; University Of Southampton; School of Health Sciences, University Road, Southampton, SO17 1BJ; [J.L.Hope@soton.ac.uk](mailto:J.L.Hope@soton.ac.uk); +44 (0)23 8059 5292

Jackie Bridges; PhD, MSN, B.Nurs(Hons), RN, PG Cert; Professor of Older People’s Care; University Of Southampton; School of Health Sciences, University Road, Southampton, SO17 1BJ; [Jackie.Bridges@soton.ac.uk](mailto:Jackie.Bridges@soton.ac.uk); +44 (0)23 8059 8282

Peter Griffiths; PhD, BA, RN, FEANS, FHEA; Chair of Health Services Research; University Of Southampton; School of Health Sciences, University Road, Southampton, SO17 1BJ; [peter.griffiths@soton.ac.uk](mailto:peter.griffiths@soton.ac.uk); +44 (0)23 8059 7877

CONFLICT OF INTEREST

The authors have no conflicts of interest to declare.

FUNDING SOURCES

This study is funded by the National Institute for Health Research (NIHR) [Health Services & Delivery programme (grant number 17/05/03)]. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

# ACKNOWLEDGMENTS

We wish to thank University of Southampton Bachelor of Nursing students Natasha Anderson, Mary Fleming, and Suzannah Hyde for acting as nurses or healthy patients during our inter-rater reliability sessions.

ABSTRACT

Background: Several time-and-motion studies have sought to quantify the nursing work involved in undertaking patients’ vital signs observations; however, none of these studies offered a validated methodology that can be replicated. This is reflected in the high variation in mean times to measure and record vital signs observations found in previous studies.

Aim: To describe the development and inter-rater reliability of a methodology to observe the nursing time and workload involved in measuring and recording patients’ vital signs observations.

Discussion: We developed a methodology using the Quality of Interactions (QI) Tool, to record and measure start and finish times of nurses taking vital signs observation rounds and individual observations clustered within a round. Two raters concurrently documented their observations of nurses undertaking patient observations in a simulated setting. Our tool and associated documentation were found to be easy to use, and there was a high level of agreement in measurements by different observers, according to the ‘Bland and Altman’ criteria.

Conclusion: Our approach can be used to reliably measure the time involved in taking vital signs observations.

Implications for practice: Use of our tool may increase the level of precision when timing and classifying nursing activities around vital signs observations. We anticipate that our tool could be effectively adapted to measure several other nursing activities, hence it will support researchers interested in capturing different aspects of nurses’ work.

KEYWORDS: Bland & Altman; Inter-Rater Reliability; Nursing; Vital Signs; Measurement; Time and Motion studies

INTRODUCTION

Vital signs observations help nurses and clinicians to detect patient deterioration and are considered a fundamental part of nurses’ work (Wood et al., 2019). Evidence about the time and workload involved with nurses’ vital signs observation activities is scant and inconclusive: when using pen and paper methods to record, mean times to take a full set of vital signs observations range from 3.58 minutes (Wong et al., 2017) to 5.80 minutes (Zeitz, 2005, Zeitz and McCutcheon, 2006). Estimates of mean times for measurement and documentation are lower when documentation is performed using electronic systems, with an average of 2.50 minutes (Bellomo et al., 2012, Wong et al., 2017). Previous time-and-motion studies used different methodologies to capture: the time nurses spend measuring patients’ vital signs; recording and documenting patients’ vital signs; and calculating Early Warning Scores (EWS). Some studies reported no details of the method used to capture nurses’ vital signs activities (Bellomo et al., 2012).

In the UK, current guidance advises vital signs observations are taken for hospital patients at least every 12 hours. Protocols on the frequency of vital signs observations are based on expert opinion rather than empirical evidence (Royal College of Physicians, 2017, Smith et al., 2017), and it is assumed that more frequent vital signs observations will lead to better patient outcomes. However, the UK, in common with other countries, is experiencing a nursing workforce crisis (Buchan et al., 2017), so increasing the workload for current nurses may be infeasible. Establishing how much nursing time is required for vital signs activities is therefore essential before any change to national guidance is implemented. This would allow an accurate assessment of the feasibility of any proposed changes and the resources required.

This paper reports a methodology that has been developed to time nurses’ vital signs observations activities. It is part of a wider parent study “Safer and more efficient vital signs monitoring to identify the deteriorating patient: an observational study towards deriving evidence-based protocols for patient surveillance on the general hospital ward”, ISRCTN10863045 (http://www.isrctn.com/ISRCTN10863045). The parent study aims to develop a validated vital signs monitoring protocol for patients on general medical and surgical wards that maximises the detection of patient deterioration while remaining feasible in terms of nursing workload. In this article, we aim to describe the development of methods to record nurses’ vital signs activities, and to report the inter-rater reliability for two observers using this method. We use the Bland and Altman procedure for establishing inter-rater reliability (Bland and Altman, 2010). This is a standard approach to establishing agreement between two measurements, raters, or methods of assessment, but is often neglected in establishing inter-rater reliability in nursing research (Griffiths and Murrells, 2010).

METHODS

The inter-rater reliability testing took place in May 2019. The sections below detail each step of the study.

*Data collection tool: the QI Tool*

We used the QI Tool to record timings and frequency of vital signs activities. The QI Tool is a tablet-based time sampling tool that enables users to enter data in real-time. The QI Tool software was originally designed for use in the Creating Learning Environments for Compassionate Care (CLECC) study (Bridges et al., 2018). The QI Tool includes the Quality of Interactions Schedule (QuIS), a validated and widely used research instrument which captures the quality of staff-patient interactions (Dean et al., 1993). The QI tool was originally designed to allow observers to record the length, quality and frequency of interactions between staff and patients.

The QI tool also enables the collection of ward-based contextual data, including the total number of patients on the ward, and numbers of registered nurses (RNs), nursing assistants and student nurses on shift during the observation session. When the observation session is complete, data are uploaded wirelessly to the central database.

*Adaptation and protocol*

Our focus was on length and frequency of nursing vital signs observations rather than quality of patient-staff interactions; we therefore designed checklists and flow-charts that would allow us to use the QI Tool to capture vital signs-related activities. Firstly, we produced a list of operational definitions related to observations of staff monitoring patients’ vital signs.

These are as follows:

* Vital Signs: one or more of the six physiological parameters that form the basis of the NEWS2 scoring system (Royal College of Physicians, 2017): respiration rate, oxygen saturation, systolic blood pressure, pulse rate, level of consciousness or new confusion and temperature.
* Vital Signs Equipment: timer, sphygmomanometer, manual blood pressure cuff, stethoscope, oxygen pulse oximeter, thermometer, electronic recording system of vital signs, ward tablets.
* Vital Signs Documentation: observing previous vital signs trends electronically or on paper, patient notes, track and trigger charts, other vital signs observations charts.
* Vital Signs Round: captures all parts of taking vital signs observations including using ‘Vital Signs Documentation’ and ‘Vital Signs Equipment’ as well as carrying out ‘Individual Vital Signs’. One or more sets of individual vital signs observations taken sequentially by a single nurse are clustered in a vital signs round.
* Individual Vital Signs Observations: the act of taking and recording the measurements of vital signs and sourcing vital signs equipment per patient.
* Complete ‘Individual Vital Signs’ means that there is no reason for the observer to suppose that the six physiological parameters of vital signs as outlined in NEWS2 have not all been completed.
* Incomplete ‘Individual Vital Signs’ means that it is obvious to the observer that not all 6 physiological vital signs have been completed.
* Interruption: including anything outside of the process of ‘Vital Signs’, ‘Vital Signs Equipment’ and ‘Vital Signs Documentation’ i.e. provision of personal care, leaving the patient to speak to someone else, or to attend to other patients.

For reasons of feasibility, only one vital signs round was observed at a time which meant, in effect, a single nurse was to be the focus of observation for the duration of the round. If two nurses worked in tandem for any period of the round, we made a note of this.

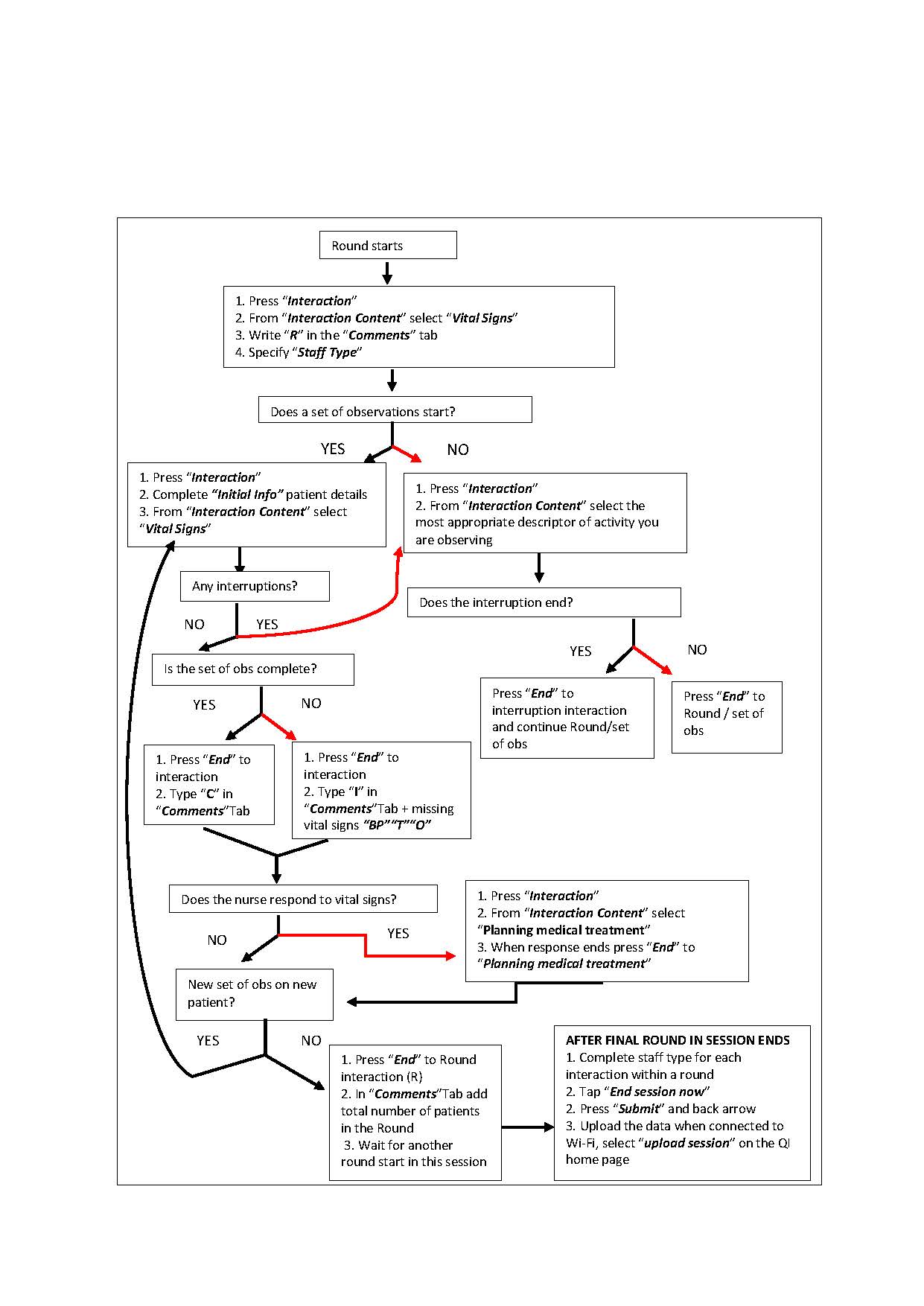
Secondly, we established clear boundaries of start and stop times of vital signs rounds, of individual vital signs observations, and of interruptions. These criteria can be found in Table 1.

Table 1 Criteria for start and stop of vital signs observations rounds, individual observations, and interruptions

|  |  |  |
| --- | --- | --- |
|  | **Timing of Observations** | |
| **Press start when…** | **Press End when…** |
| **‘Observations Round’** | The nurse sources ‘Vital Signs Equipment’ or ‘Vital Signs Documentation’ | The nurse has taken one or more sets of observations and leaves the patients then replaces the ‘Vital Signs Equipment’ or ‘Vital Signs Documentation’ |
| **‘Individual Observation’** | The nurse enters the bed space and measures one or more of the six physiological parameters of the NEWS2 scoring system. | The nurse leaves the bed space and it is clear that the specific measuring of these vital signs as defined by NEWS2 has finished. |
| **Interruption** | The Nurse is interrupted from the process of taking and recording an ‘Individual Observation’ or ‘Observation Round’, stops and starts a different task or conversation | The Nurse reverts back to the process of taking and recording the an ‘Individual Observation’ or ‘Observation Round’ |

We then applied these criteria as a protocol in practice, during four training sessions, which enabled the two observers to familiarise themselves with the QI Tool. We used the content of Table 1 as the basis of a checklist and flow chart, which we adjusted and developed iteratively. The final version of the flowchart can be found in Figure 1.

Figure 1 Flowchart for observing staff undertaking vital signs observations



Originally, the term “Interaction” in the QI Tool referred to any instance where there was an interaction between a staff member and a patient. For the purposes of our study, we restricted “interaction” to observation rounds, individual sets of observations, interruptions during a round, and responses to observed vital signs.

Inter-rater reliability testing

After a number of iterations the study team agreed the final version. We then undertook inter-rater reliability testing in May 2019 over four hours of observation in a clinical skill classroom that was laid out and equipped as a hospital ward. Two student nurses acted as nursing staff undertaking vital signs observations on healthy volunteers who acted as patients. The observers were two research fellows, one of whom is a registered nurse and one a social scientist. Eight different scenarios were devised in conjunction with clinical academic staff, to include potential disruptors such as interruptions or incomplete measurement. Observers were located so that they could not see each other inputting data, which could have influenced the timing of initiation and ending for each observation made.

We applied the Bland and Altman criteria (Bland and Altman, 2010) to describe the agreement between measurements of time involved in nurses’ vital signs activities by these two observers. Bland and Altman (2010) developed a method to quantify agreement between two measurements by establishing limits of agreement. These statistical limits are calculated through the mean and the standard deviation of the differences between two measurements.

RESULTS

The four hours of observation yielded 44 individual vital signs observations sets. Based on averaging the two observers’ ratings, the mean time to take and record a set of vital signs observations was 3 minutes 47 seconds, and the median was 3 minutes 25 seconds.

Over the four hours of observation observer A recorded a total of 2 hours 44 minutes of vital signs measurements compared to observer B who recorded 2 hours 42 minutes. The mean difference per set of vital signs observations was 3 seconds, (standard deviation 9 seconds). According to Bland and Altman (2010), all calculations should be preceded by a visualisation of the results of one measurement against the other on a scatter plot. The scatter plot indicates that timings recorded by the two raters were closely associated (Figure 2).

Figure 2 Scatterplot of observer measurements with linear regression line

Bland and Altman (2010) argued that regression should not be adopted as an indication of agreement between measurements, as all data points usually will cluster near the regression line. They suggested plotting the difference between the measurements against their mean. This is reported in Figure 3.

Figure 3 Difference against the mean for vital signs length data

Two standard deviations from the mean are defined as limits of agreement; in this case, +19 seconds and -13 second are the limits of agreement. Considering clinical measurements, Bland and Altman argued that limits of agreement should be “clinically” acceptable, and that 95% of the differences should fall within the limits of agreement. In our case, only one of the 44 differences (2%) was outside (35 seconds) of limits of agreement, suggesting that observers could reliably record times of nurses’ vital signs observations.

Bland and Altman (2010) advise that this approach only estimates the values which apply to the whole population, hence they recommend using standard errors and confidence intervals of the mean difference and of the limits of agreement to check how precise these estimates are. The standard error of the mean is calculated as follows:

Where n is the sample size and sd is the standard deviation. The standard error of the lower and upper limits of agreement is:

The appropriate point on the t distribution was identified with (sample size – 1) degrees of freedom, two-tailed at 0.05. The confidence intervals of the mean were derived from the mean minus t standard errors to mean plus t standard error. The same procedure was applied to the limits of agreement. Calculations from our sample are reported in Table 3.

Table 3 Precision of estimated limits of agreement (seconds)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Parameter** | **Value** | **Standard Error** | **t-value** | **Confidence Interval** |
| Sample size (n) | 44 (observations) |  |  |  |
| degrees of freedom  (n-1) | 43 (observations) |  |  |  |
| Mean | 3 | 1 | 2 | 1 to 5 |
| Standard deviation | 9 |  |  |  |
| Lower limit of agreement (mean -1.96\*standard deviation) | -13 | 2 | 2 | -17 to -9 |
| Upper limit of agreement (mean +1.96\*standard deviation) | 19 | 2 | 2 | 15 to 23 |

The lower limit of agreement was -13 seconds, with confidence intervals ranging from -17 and -9 seconds. The upper limit of agreement was 19 seconds, with confidence intervals ranging from 15 to 23 seconds.

DISCUSSION

We developed a methodology to record time and frequency of vital signs observations and found researchers were in agreement when applying it. The two observers were in close agreement with the estimated mean difference per set of vital signs observations of less than 5 seconds. Average time to take a set of vital signs was 3 minutes 25 seconds. The limits of agreement were narrow and no more observations than expected fell outside them. The differences between observers were small and of little if any material importance. When totalling time up across all the observations, the absolute difference was less than 2%.

Limitations

Although our study was limited by the relatively small sample size (n= 44 observations), the confidence intervals of the limits of agreement were not wide. We therefore conclude this sample is sufficient to ascertain reliability. We tested reliability in a simulated setting where volunteers were healthy. To aid relevance to real life situations, a clinical expert designed deliberately disruptive and realistic scenarios. Examples included patients asking to be taken to the toilet, or colleagues requesting immediate assistance. We suggest further reliability testing takes place in a real clinical setting. Our observers were trained in applying the methodology, although it should be noted that the training sessions also served as opportunities to refine the methodology. While we tested the reliability of two raters, they had diverse backgrounds and this study demonstrates it is possible to train raters to achieve high reliability with this tool.

CONCLUSION

Our time-and-motion methodology can be used by properly trained observers in future studies to reliably measure the nursing time involved in undertaking vital signs observations. As with other observational measures, before using this tool in future studies the reliability of observers should be tested after training and practice. This is because reliability is not an intrinsic property of a measure. However, we have demonstrated that reliability can be achieved. Our methodology will be of particular importance to health services researchers. This is because there is no consistent evidence on the time and workload involved with vital signs observations for nurses. Any changes to protocols of recommended frequency of vital signs observations should occur only when the input required from nurses is understood. Our reliable methodology offers the opportunity to fill this evidence gap.

REFERENCES

BELLOMO, R., ACKERMAN, M., BAILEY, M., BEALE, R., CLANCY, G., DANESH, V., HVARFNER, A., JIMENEZ, E., KONRAD, D., LECARDO, M., PATTEE, K. S., RITCHIE, J., SHERMAN, K., TANGKAU, P., VITAL SIGNS TO IDENTIFY, T. & ASSESS LEVEL OF CARE STUDY, I. 2012. A controlled trial of electronic automated advisory vital signs monitoring in general hospital wards. *Crit Care Med,* 40**,** 2349-61.

BLAND, J. M. & ALTMAN, D. G. 2010. Statistical methods for assessing agreement between two methods of clinical measurement. *International Journal of Nursing Studies,* 47**,** 931-936.

BRIDGES, J., PICKERING, R. M., BARKER, H., CHABLE, R., FULLER, A., GOULD, L., LIBBERTON, P., MESA-EGUIAGARAY, I., RAFTERY, J., AIHIE SAYER, A., WESTWOOD, G., WIGLEY, W., YAO, G., ZHU, S. & GRIFFITHS, P. 2018. Creating Learning Environments for Compassionate Care (CLECC): a feasibility study. *Health Services and Delivery Research*.

BUCHAN, J., SECCOMBE, I., GERSHLICK, B. & CHARLESWORTH, A. 2017. In short supply: pay policy and nurse numbers - Workforce profile and trends in the English NHS. The Health Foundation.

DEAN, R., PROUDFOOT, R. & LINDESAY, J. 1993. The quality of interactions schedule (QUIS): Development, reliability and use in the evaluation of two domus units. *International Journal of Geriatric Psychiatry,* 8**,** 819-826.

GRIFFITHS, P. & MURRELLS, T. 2010. Reliability assessment and approaches to determining agreement between measurements: Classic methods paper. *International Journal of Nursing Studies,* 47**,** 937-938.

ROYAL COLLEGE OF PHYSICIANS 2017. National Early Warning Score (NEWS) 2 - Standardising the assessment of acute-illness severity in the NHS.

SMITH, G. B., RECIO-SAUCEDO, A. & GRIFFITHS, P. 2017. The measurement frequency and completeness of vital signs in general hospital wards: An evidence free zone? *Int J Nurs Stud,* 74**,** A1-A4.

WONG, D., BONNICI, T., KNIGHT, J., GERRY, S., TURTON, J. & WATKINSON, P. 2017. A ward-based time study of paper and electronic documentation for recording vital sign observations. *J Am Med Inform Assoc,* 24**,** 717-721.

WOOD, C., CHABOYER, W. & CARR, P. 2019. How do nurses use early warning scoring systems to detect and act on patient deterioration to ensure patient safety? A scoping review. *Int J Nurs Stud,* 94**,** 166-178.

ZEITZ, K. 2005. Nursing observations during the first 24 hours after a surgical procedure: what do we do? *J Clin Nurs,* 14**,** 334-43.

ZEITZ, K. & MCCUTCHEON, H. 2006. Observations and vital signs: ritual or vital for the monitoring of postoperative patients? *Appl Nurs Res,* 19**,** 204-11.