JAN JOURNAL OF ADVANCED NURSING

ORIGINAL RESEARCH

Nurses' risk assessment judgements: a confidence calibration study

Huiqin Yang & Carl Thompson

Accepted for publication 16 July 2010

Correspondence to H. Yang:
e-mail: hy505@york.ac.uk

Huiqin Yang MSc PhD RN
Research Fellow
Centre for Reviews and Dissemination
University of York, UK

Carl Thompson PhD RN
Professor
Department of Health Sciences
University of York, UK

YANG H. & THOMPSON C. (2010) Nurses' risk assessment judgements: a confidence calibration study. *Journal of Advanced Nursing* **66**(12), 2751–2760.
doi: 10.1111/j.1365-2648.2010.05437.x

Abstract

Aim. This paper is a report of a study of the relationship between nurses' clinical experience and calibration of their self-confidence and judgement accuracy for critical event risk assessment judgements.

Background. Miscalibration (i.e. under-confidence or over-confidence of confidence levels) has an important impact on the quality of nursing care. Despite this, little is known about how nurses' subjective confidence is calibrated with the accuracy of their judgments.

Methods. A sample of 103 nursing students and 34 experienced nurses were exposed to 25 risk assessment vignettes. For each vignette they made dichotomous judgements of whether the patient in each scenario was at risk of a critical event, and assigned confidence ratings (0–100) to their judgement calls. The clinical vignettes and judgement criteria were generated from real patient cases. The methodology of confidence calibration was used to calculate calibration measures and generate calibration curves. Data were collected between March 2007 and January 2008.

Findings. Experienced nurses were statistically significantly more confident than students but no more accurate. Whilst students tended towards under-confidence, experienced nurses were over-confident. Experienced nurses were no more calibrated than students. Experienced nurses were no better at discriminating between correct and incorrect judgements than students. These patterns were exacerbated when nurses and students were extremely over-confident or extremely under-confident.

Conclusion. Nurses were systematically biased towards over/under-confidence in their critical event risk assessment judgements. In particular, experienced nurses were no better calibrated than their student counterparts; with student under-confidence countered by experienced nurses' greater susceptibility to over-confidence.

Keywords: clinical judgement, confidence calibration, critical event risk assessment, nursing, over-confidence, under-confidence

Introduction

Over-confidence and under-confidence in judgement and decision-making represent a major reasoning bias (Lichtenstein & Fischhoff 1980, 1982, Juslin 1993, 1994, Petrusic & Baranski 1997). The ability of nurses in all nursing clinical settings to think reflexively but appropriately about the interaction between confidence and clinical judgement is crucial. Critical event risk assessment by nurses in acute care is a major focus for quality improvement nurses in a variety of international developed healthcare systems (DeVita *et al.* 2006). A nurse who is inappropriately over-confident in their critical event risk assessments may omit or delay remedial or proactive clinical interventions. Confidence, reasoning, judgement and the actions that promote good quality healthcare and better patient outcomes are intricately linked.

Background

The relationship between confidence and judgemental correctness is captured by the term confidence *calibration* (Lichtenstein & Fischhoff 1982, Keren 1991). Calibration refers to the correspondence between a subjective probability estimate of the occurrence of a particular event and empirical (observable and objective) probabilities of the occurrence of that same event (Juslin & Olsson 1997, Petrusic & Baranski 1997). Calibration studies are primarily concerned with the appropriateness of these subjective probability estimates, which is determined by the degree of concordance between the perceived confidence in an event occurring and the probability of its actual occurrence.

In medicine, evidence shows that doctors are often prone to over- or under-confidence in their estimates of the correctness of their judgements. Hausman *et al.* (1990) investigated the confidence calibration of 33 paediatric resident physicians on the results of multiple choice examinations, and found a significant association between increasing over-confidence and lower examination scores, with participants tending towards over-confidence in the correctness of their judgements. This finding was reinforced by Friedman *et al.* (2005), who investigated the alignment between physicians' confidence in their diagnoses and their diagnostic correctness. The results suggested that residents were over-confident in 41% cases, educators in 36% and students in 25%. Experienced clinicians can be unaware of the correctness of their diagnoses and prone to over-confidence. Experience may even lead to greater over-confidence than that seen in less experienced clinicians.

In nursing, the need to handle complex information in conditions of irreducible uncertainty (Eddy 1988) means the potential for over-confidence is always present (Thompson 2003). McMurray (1992) showed that experienced nurses were indeed prone to being highly confident in their judgements. Baumann *et al.* (1991) went further and found that critical care nurses were very confident in their judgements, despite a lack of consensus on the optimal choice of intervention. Clinical experience and confidence in their actions are often viewed as indicators of nurses' competency in clinical activities. In turn, accumulated clinical experience has long formed the foundation for building confident nurse judges (McMurray 1992, Haffer & Raingruber 1998, Crooks *et al.* 2005). This is hardly surprising, as it is experiential knowledge that nurses primarily use in their daily judgments (Thompson 2003). Whilst these studies imply that the presence of over-confidence might undermine the validity of nurses' judgements, and by implication the quality of care delivered, all failed to investigate nurses' confidence calibration systematically and rigorously.

The study

Aim

The aim of this study was to examine the relationship between nurses' clinical experience and calibration of their self-confidence and judgement accuracy for critical event risk assessment judgements.

Design

A confidence calibration (Lichtenstein & Fischhoff 1982) design was used.

Methodology of confidence calibration

Calibration statistics

Three indices describe the association between confidence ratings and judgement correctness: (i) calibration score, (ii) over/under-confidence and (iii) resolution.

Mathematically, subjective probabilities of confidence are analysed as a function of the objective probabilities of the proportion that is correct (Erev *et al.* 1994). The calibration score is the sum of squared deviations from a 45 degree line in a scatter plot of subjective assessments against objectively 'correct'. This sum is weighted by the number of responses in each confidence category and divided by the total number of responses (Soll 1996), as in equation 1 (calibration score; Petrusic & Baranski 1997):

$$\frac{1}{n}\sum_{j=1}^{J} nj(\overline{p}j - \overline{e}j)^2, \qquad (1)$$

where n is the total number of responses, J is the total number of confidence categories, nj is the number of responses in confidence category j, $\overline{p}j$ is the mean confidence level associated with category j, $\overline{e}j$ is the mean proportion correct associated with category j.

The calibration score gives a weighted squared deviation of the mean proportion correct and the mean confidence rating associated with each confidence category (Lichtenstein & Fischhoff 1977, 1982). The calibration score ranges from 0 (perfectly calibrated) to 1 (worst possible calibration), with a lower score indicating better calibration. A score of 1 would be obtained by a judge who always assigned $P = 100$ when the judgement was wrong and $P = 0{\cdot}0$ when the judgement was correct. In contrast, a calibration score of zero (perfect calibration) would occur when a confidence rating assigned by a person is in accordance with the percentage of correct judgements; for example, always assigning a confidence rating $P = 80$ and obtaining 80% correct judgements.

The measure, over-/under-confidence, offers a global index of the confidence/accuracy relationship; it measures the deviation between confidence and proportion correct: $p - e$, where p denotes mean confidence level and e denotes mean proportion of correct judgements. A negative score (mean proportion correct exceeds mean confidence) indicates under-confidence. In contrast, a positive score (mean confidence exceeds mean proportion correct) indicates over-confidence.

The final measure of the confidence/accuracy relationship used in our analysis is resolution. The resolution score gives a weighted squared deviation of the mean proportion correct ($\bar{e}j$) for each confidence category (e.g. 0·50–0·59, 0·60–0·69, and so on) and the overall proportion correct ($\bar{e}$) at the whole group level, as in equation 2 (resolution; Petrusic & Baranski 1997):

$$\frac{1}{n}\sum_{j=1}^{J} nj(\bar{e}j - \bar{e})^2. \tag{2}$$

The resolution score measures how well nurses adjust their confidence ratings in the face of correct and incorrect responses (i.e. resolution). The normalized resolution scores were adjusted by the knowledge index $\bar{e}(1 - \bar{e})$ (Yaniv *et al.* 1991), as in equation 3 (normalized resolution; Yaniv *et al.* 1991):

$$\mathrm{NRI} = \frac{\left[\frac{1}{n}\sum_{j=1}^{j} nj(\bar{e}j - \bar{e})^2\right]}{\bar{e}(1 - \bar{e})}. \tag{3}$$

Normalized resolution scores vary from 0 (no resolution) to 1 (perfect ability to discriminate between correct and incorrect responses).

Calibration curve

A graphical way of analysing the relationship between probability judgements and confidence ratings is the calibration curve. In such curves, the proportion correct is plotted on the ordinate (y axis) for each confidence rating plotted on the abscissa (x axis) (Keren 1991, Soll 1996, Griffin & Brenner 2004). Plotting a calibration curve requires the conversion of confidence levels (continuous data) into ordinal categories, such as 0·50–0·59, 0·60–0·69 and so on. The mean proportion correct for each response group is then plotted against the corresponding mean confidence rating for that category. A 45 degree line thus represents perfect calibration; any deviations from the 45 degree line indicate the extent of miscalibration (e.g. over-confidence or under-confidence). Specifically, the lower the curve below the calibration line the greater the over-confidence, whilst the greater the curve above the line the greater the tendency towards under-confidence.

Participants

Purposive sampling was employed to recruit relevant participants, i.e. experienced nurses and nursing students. Because students represent the extreme end of the distribution of experienced nurses (i.e. they have very little clinical experience as autonomous clinical judges), they are a useful comparator for the experienced nurses. Torgerson and Campbell (2000) have demonstrated that studies can oversample some groups in an experiment without compromising the design principles and assumptions required for comparison, thus making studies more economically efficient. We recruited a convenience sample of experienced ($n = 34$) nurses from the population of ward and critical care nurses in North Yorkshire, United Kingdom, and second and third year nurse students ($n = 103$) from the undergraduate student population in a large university in Northern England.

Vignette scenarios

A clinical vignette questionnaire was constructed to present 25 clinical scenarios to enable the cue values to be systematically altered and to present 'static' background clinical information for an emergency patient (see Appendix 1). Each scenario contained five cues: systolic blood pressure, heart and respiratory rates, temperature and consciousness levels. Cues were presented in units used in clinical practice (e.g. respiratory rate in breaths per minute). Patient consciousness was represented via four levels: alert (A), reacting to voice (V), reacting to pain (P) and unresponsive (U). All the cues are widely used in rapid assessments of consciousness in critical care patients (Independent Healthcare Association 2002). The format and content was also reviewed and approved by a critical care specialist nurse with more than 10 years' specialist experience.

The clinical scenarios were derived from real patient cases by means of a random sample of real patient cases in a large dataset ($n = 673$) of emergency admissions collected in the medical admissions unit of a single NHS District General Hospital during March 2000 by Subbe *et al.* (2001). Judgement criteria (i.e. whether a judgement was 'correct' or not) were derived from case records prospectively collected by Subbe *et al.* (2001). Patients were classed as 'at risk' if they died, were admitted to intensive care or high dependency units, or underwent cardiopulmonary resuscitation.

Data collection

Individual participants were asked to make an independent dichotomous judgement (yes/no) of whether the patient in each scenario was 'at risk of a critical event'. They were then asked to assign confidence ratings (0–100) to their judgements for the risk assessment vignettes. Data were collected between March 2007 and January 2008.

Ethical considerations

The appropriate research ethics committee approved the study. Students were invited to take part by one of the authors (HY) by letter after hearing about the project at the end of a lecture on clinical decision-making in nursing. There was no conflict of interest between the researchers and the students being recruited, as neither researcher was directly involved in the assessment of any of the students. It was made clear to students that their participation in the study was in no way linked to their studies or assessment of performance.

Data analysis

Confidence calibration analyses were performed at the individual (ideographic) level. Individual calibration statistics were calculated to enable hypothesis testing of differences in these scores between experienced and nursing student groups. For data meeting the assumptions required for parametric testing, independent two-sample *t*-tests were used to test for differences between calibration measures in the two groups. Where parametric assumptions were not met, we used the Wilcoxon rank-sum test for the variables of calibration and resolution scores to test the null hypothesis that the median (Mdn) of the differences between groups was zero. A level of $P < 0{\cdot}05$ was used as a cut-off for statistical significance. Calibration analyses were conducted using Stata version 9 (http://www.stata.com).

Calibration curves were obtained by plotting mean confidence ratings against the mean proportion correct using the following confidence categories: 50–59, 60–69, 70–79, 80–89, 90–99 and 100. Confidence categories <50 were associated with very few data points; thus, including these data would significantly bias the calibration curve. To derive more reliable calibration curves, confidence ratings of <50 were therefore excluded.

Results

Participants

Table 1 presents the participant demographics: 98% ($n = 101$) of students were registered for a nursing diploma, whilst 2% ($n = 2$) were registered on a Bachelor degree programme. There were no missing data. The mean age of experienced nurses was 36·6 years, whilst the mean age of nursing students was 29·1 years. The experienced nurses had an average of 12·2 years of clinical experience and 85% of them were women, whilst 89% of nursing students were women.

Under/over-confidence

No statistically significant differences in the proportion of judgements correct were found between the student (mean 77·55%, SD 7·73%) and experienced nurse groups (mean 77·53%, SD 6·82%), $t\ (135) = 0{\cdot}02$, $P = 0{\cdot}99$. However, confidence ratings of experienced nurses (mean 82·22%, SD 9·06%) were statistically significantly higher than those of students (mean 74·37%, SD 11·50%), $t\ (135) = -3{\cdot}62$, $P = 0{\cdot}0004$.

Students were under-confident (mean of over/under-confidence score −3·18%, SD 14·11%) and experienced nurses were over-confident (mean of over/under-confidence score 4·69%, SD 12·74%), $t\ (135) = -2{\cdot}89$, $P = 0{\cdot}005$.

Calibration and resolution

The Wilcoxon rank-sum test revealed no statistically significant difference in calibration between experienced nurses (Mdn 0·046) and students (Mdn 0·053), $z = 0{\cdot}74$, $P = 0{\cdot}46$.

Table 1 Participant demographics

Demographics	Experienced nurses ($n = 34$)	Nursing students ($n = 103$)
Age (years) mean (SD)	36·55 (9·96)	29·05 (9·04)
Clinical experience (years) mean (SD)	12·15 (9·90)	n/a
Gender *n*/(%)		
Male	5 (15)	11 (11)
Female	28 (85)	90 (89)

n/a, not applicable.

Experienced nurses were no better at judgement-confidence resolution (Mdn 0·199) than students (Mdn 0·206), $z = 0{\cdot}30$, $P = 0{\cdot}76$.

Calibration curve analysis

Figure 1 shows the calibration curve used to compare the experienced nurses and students: 6·20% of reported confidence ratings were < 50. Confidence ratings below 50 indicated the participant's awareness (subjective feeling) of an error judgement, with 0 denoting complete certainty of an error. Whilst excluded from the curves generated, the mean confidence for these reports was 34·5% and the mean of proportion correct associated with them was 69·8%. These results imply that participants had insight into subjective errors but were under-confident; they overestimated the degree to which they err.

The calibration curves in Figure 1 are similar for the student and experienced nurse groups. There was a clear cut-off for under-confidence and over-confidence for both groups: with a stated confidence of more than 75% participants tended to be over-confident, whilst confidence ratings of lower than 75 revealed a degree of under-confidence. This phenomenon is depicted by the two curves and is known as over-extremity (Griffin & Brenner 2004) – a tendency to assign subjective probabilities that are consistently too extreme, given the underlying relationship.

Discussion

Study strengths and limitations

Strengths

In this study, we examined nurses' and nursing students' confidence calibration performance by investigating how well their subjective confidence estimates calibrate with objective probabilities of judgement correctness. This offers a careful evaluation of their calibration performance about the appropriateness of confidence on their risk assessment judgements. Given that confidence calibration research (with associated theoretical and methodological foundations) is an important component in the field of nursing decision-making and judgment, our study contributes to current research knowledge of how nurses and nursing students judge and choose from the clinical choices they face in practice.

The idea that nurses make choices and exercise judgement is central to models of contemporary nursing. Almost all models of nursing (McKenna 1997) have at their core a vision of human beings with needs to be met and the nurse as a central component (often *the* central component) in identifying these needs. Benner *et al.* (1999) have long argued that a measure of the expertise of a nurse (and nursing as a professional craft) lies in the ability to identify these needs effectively through skilled 'reflection in action' – or, in the parlance of this paper, clinical judgement. Our study reinforces the hypothesis that clinical experience alone is a necessary but not sufficient basis for claims of expertise, at least in the task of critical event risk assessment.

In constructing the clinical scenarios used, our study used real patient data from a large data set of medical emergency admissions. This approach ensured the face and content validity of the vignettes. The judgment criteria were also established using the real outcomes of records of patient cases.

Limitations

We purposively sampled both experienced nurses and nurse students; this approach helped to identify appropriate participants who possessed relevant characteristics, given our research questions. However, one of the limitations in sampling is that we used volunteers. Because of the time and resource constraints for conducting the study, we also recruited a limited sample of experienced nurses. Randomly sampling participants in sufficient numbers is an important means of fostering generalizability and reducing any selection bias in research with nurse decision-makers. Future researchers might like to replicate our methods with larger numbers of randomly sampled participants.

In real clinical judgement situations, large amounts of clinical information are available to nurses. These include perceptual information cues (e.g. patient's skin colour or smell, etc.); nurses might use these cues in recognizing patients at risk of deterioration. Their confidence may be influenced by these additional clinical information cues, as might their accuracy; we did not give participants in our

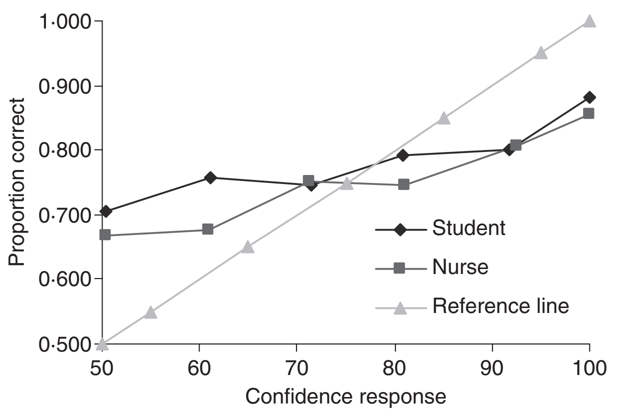

Figure 1 Calibration curves of experienced nurse and nursing student groups.

study the chance to use these cues. Other researchers may like to consider the use and role of clinical simulation environments to develop calibration in controlled environments whilst enabling greater realism in the perceptual cues available.

Miscalibration: under-confidence and over-confidence

We have outlined the miscalibration between nurses' subjective confidence and objective accuracy; specifically, experienced nurses tend towards over-confidence, whilst students were generally under-confident. In doing so, we have revealed that the subjective probability judgements of experienced nurses and students are prone to systematic bias because they over/underestimate their abilities and/or knowledge.

Over-confidence and under-confidence as phenomena are well-documented in the psychological literature (Lichtenstein & Fischhoff 1982, Juslin 1993, 1994, Baranski & Petrusic 1994, Soll 1996, Petrusic & Baranski 1997). We should perhaps not be surprised to find that nurses, like all human beings, are naturally prone to reasoning mistakes involving overestimating or underestimating their knowledge on judgements. Such patterns have been identified in doctors. Hausman *et al.* (1990) report that paediatric residents were over-confident on the correctness of their judgements. Friedman *et al.* (2005) also identifies that physicians' diagnosis was accompanied by inappropriate confidence levels, and that experienced physicians were over-confident than inexperienced physicians. Our study replicates these findings and reveals that even experienced nurses are not immune to such cognitive biases in clinical judgements.

Clinical experience and calibration

Experienced nurses were statistically significantly more confident in their judgements than students, with nurses' confidence increasing in line with their clinical experience. This finding mirrors a number of other studies. Baumann *et al.* (1991) found that experienced nurses were highly confident in their treatment/intervention choices. Hamers *et al.* (1997) found that experienced nurses were highly confident in their judgements of pain intensity. Indeed, accumulating clinical experience has been identified as the element most required for building confidence in nurses making judgements (McMurray 1992, Haffer & Raingruber 1998, Crooks *et al.* 2005). Our findings support the hypothesis that experience statistically significantly increases nurses' confidence in their clinical judgements.

Despite experience being a vital ingredient for building confidence, this does not mean that judgement accuracy is a linear function of experience. Our findings reveal that, whilst confidence increases as a function of the years of clinical experience, the accuracy of judgements was not increased by clinical experience. Calibration was unaffected by the difference in clinical experience between experienced nurses and students.

The tendency towards over-confidence in experienced nurses indicates that they were not aware of the underlying levels of judgemental 'correctness': experienced nurses expressed confidence levels that exceeded their accuracy. Therefore, we found no evidence that better calibration in nurses' critical event risk assessment judgements can be attributed to increased experience. One possible reason for this finding is that experienced nurses may not appreciate (cognitively at least) the uncertainty inherent in clinical practice. Uncertainty is an irreducible part of delivering health care (Eddy 1988), and using unwarranted certainty as a heuristic to reduce the uncertainty in clinical judgements is an unwise strategy. Unwarranted certainty in the prediction of the accuracy on one's judgements has been identified as a major source of over-confidence (Kissinger 1998). Baumann *et al.* (1991) also point out that critical care nurses do not appear to perceive uncertainty in some situations when coping with clinical judgements. Our data reinforce this assertion in that some experienced nurses were quite certain that their judgements were correct and consistently assigned confidence at ceiling (100) levels.

The calibration curves associated with participants in our study reveal over-extremity (Griffin & Brenner 2004): the tendency consistently to assign and rely on extreme (but unwarranted) probabilities in judgement tasks. Not adjusting subjective probabilities leads to an increased likelihood of selecting whatever judgement participants are most predisposed to – regardless of the judgement task information they are faced with. The phenomenon also means that inferring the 'quality' (as correspondence) of judgements from nurses' self-reports of confidence in their judgements may be invalid.

Our results illustrate that experienced nurses had no better resolution than students. Resolution reflects people's ability to sort scenarios into different subcategories where the proportion correct is maximally different from the overall proportion correct (Lichtenstein & Fischhoff 1977), or how well judges use their confidence ratings to differentiate correct from incorrect responses.

Implications for practice

Experienced nurses were biased towards over-confidence. This is problematic for decisions that rely on reflexive awareness of uncertainty in a judgement. However, it

What is already known about this topic

- Whilst miscalibration in the form of over-confidence (and under-confidence) in judgmental ability is a characteristic of human reasoning, it is implicated in delayed profession intervention, unwarranted inaction and healthcare error.
- Critical event risk assessment relies on nurses' judgements and is prone to instances of delayed or absent action and errors.
- It is often assumed that experience is a useful proxy for expertise, but this assumption can be questioned when focusing on judgement as an element of clinical performance.

What this paper adds

- Experienced nurses were statistically significantly more confident than students but no more accurate on risk assessment judgements.
- Whilst students tended to be under-confident, experienced nurses were over-confident.
- Experienced nurses were no more calibrated than students, and experienced nurses were no better at discriminating between correct and incorrect judgements than students.

Implication for practice and/or policy

- Nurses should consider the accuracy of their predictive judgements, whether their confidence in their judgements is warranted, and the means by which they would know this.
- Educators, researchers and practice developers should consider incorporating feedback and debiasing techniques into training for 'judgement rich' tasks (such as critical event risk assessment) and evaluating the impact of such initiatives.

represents something of a dilemma for the conscientious reflective practitioner: appearing confident may increase patients' satisfaction with their treatments and interventions (Johnson *et al.* 1988). Most of us are not in favour of ambiguity being expressed in clinical practice, even where that ambiguity is warranted. Having nurses who are prepared to acknowledge and express the uncertainty in judgements and adjust their confidence accordingly requires patients who are willing to accept that clinicians may not have all the answers and the ability to produce certainty in inherently uncertain situations.

Under-confidence can be equally problematic; it is associated with a lack of initiative, judgemental stasis and diminished ability to take risks (Zakay 1997). Of course, a degree of risk-taking, given the uncertainty in the healthcare environment, is often warranted. Nurses who are consistently under-confident may tend towards reliance on external knowledge resources, even when they are not needed. Whilst being open to advice and seeking to consult resources such as evidence-based synopses and protocols/guidelines can be a cornerstone of good decision-making (Sackett 1997), time, task structure and the quantity of available task information all conspire to mean that not all judgements are descriptively, or indeed normatively, best served by information-seeking in this form (Hammond *et al.* 1997).

On balance, however, it is over-confidence in nurses that has the highest potential for reducing the quality of care and threatening patient safety. Over-confidence may impede both the motivation for – and the information recalled and used in – nurses' reflections on the appropriateness of their judgements; making nurses less likely to learn from their experience. With higher confidence in a given hypothesis in a judgment, the motivation to acquire more knowledge to test the hypothesis is reduced (Kruglanski *et al.* 1991). If experienced judges believe they are better than they really are, there is a little incentive to change or try to reconcile their judgements with external resources (Mumpower & Stewart 1996). Therefore, alongside compromising 'reflection on action', over-confidence also compromises 'reflection in action' (Schön 1987): nurses who are too confident in their preliminary judgements in cases or situations are less likely to gather and appraise clinical information which might change that judgement, prematurely closing the judgement task (Graber *et al.* 2005). Over-confidence may also inhibit clinicians in using decision support systems (Friedman *et al.* 2005).

Importantly, over-confidence has been cited as a contributory factor in healthcare errors (Grager 2005, Berner & Graber 2008). Confidence is implicated in whether a planned task is actually executed (Zakay 1997). Specifically, someone who is confident that their judgement of 'no risk' is correct is more likely to delay, or even take no action, in the face of a patient whose condition is in reality deteriorating. In clinical practice, over-confidence is frequently manifested in delayed or missed preventions/interventions for patients (Berner & Graber 2008). Such missed or delayed actions are at the root of many 'failures to rescue' (Goldhill 2001). As nurses are the professional group most responsible for initiating 'rescue' in clinical settings (Cioffi 2000, Clarke & Aiken 2003), the hypothesis that improving their judgements might improve the identification, treatment and outcomes associated with critical events cannot easily be discounted.

Conclusion

The nurses in this study were systematically biased towards over/under-confidence in their risk assessment judgements. Over-confidence has a negative impact on practice, and combating and minimizing nurses' over-confidence is an important issue for educators, researchers and practice developers alike. In view of the fact that nurses – like all clinicians – must operate in uncertain environments, nurses should consider the accuracy of their clinical judgements and whether their confidence in these judgements is warranted; crucially, they should ask themselves, 'How would I know?'. Strategies (such as debiasing) for improving clinician calibration already exist and should be implemented and evaluated in a nursing context in order that the scale of any potential improvements might be estimated. Educators and researchers should consider incorporating feedback and debiasing techniques into training for other 'judgement rich' tasks (such as, for example, the assessment of labour in midwifery, or the assessment of risk of violence in mental health nursing) and evaluating the impact of such initiatives. Society deserves a confident nursing workforce, but that confidence needs to be warranted; for many areas of nursing the degree to whether this is so is still unclear.

Acknowledgements

We thank Prof. Martin Bland for his advice in medical statistics in this paper. We also thank Dr C. P. Subbe for consenting to us using his dataset in constructing our clinical scenarios.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflict of interest

No conflict of interest has been declared by the authors.

Author contributions

HY and CT were responsible for the study conception and design and data collection. HY performed data analysis. HY and CT drafted the manuscript and provided critical revisions of manuscript for important intellectual content.

References

Baranski J.V. & Petrusic W.M. (1994) The calibration and resolution of confidence in perceptual judgments. *Perception and Psychophysics* **55**, 412–428.

Baumann A.O., Deber R.B. & Thompson G.G. (1991) Over-confidence among physicians and nurses: the micro-certainty, macro-uncertainty phenomenon. *Social Science & Medicine* **32**(2), 167–174.

Benner P., Hooper-Kyriakides P. & Stannard D. (1999) *Clinical Wisdom and Interventions in Critical Care: A Thinking-In-Action Approach*. Saunders, Philadelphia.

Berner E.S. & Graber M.L. (2008) Over-confidence as a cause of diagnostic error in medicine. *American Journal of Medicine* **121**(5 Suppl.), S2–S23.

Cioffi J. (2000) Nurses' experiences of making decisions to call emergency assistance to their patients. *Journal of Advanced Nursing* **32**(1), 108–114.

Clarke S.P. & Aiken L.H. (2003) Failure to rescue. *American Journal of Nursing* **103**(1), 42–47.

Crooks D., Carpio B., Brown B., Black M., O'Mara L. & Noesgaard C. (2005) Development of professional confidence by post diploma baccalaureate nursing students. *Nurse Education in Practice* **5**, 360–367.

DeVita M.A., Bellomo R., Hillman K., Kellum J., Rotondi A., Teres D., Auerbach A., Chen W.J., Duncan K., Kenward G., Bell M., Buist M., Chen J., Bion J., Kirby A., Lighthall G., Ovreveit J., Braithwaite R.S., Gosbee J., Milbrandt E., Peberdy M., Savitz L., Young L., Harvey M. & Galhotra S. (2006) Findings of the first consensus conference on medical emergency teams. *Critical Care Medicine* **34**(9), 2463–2478.

Eddy D.M. (1988) Variation in physician practice: the role of uncertainty. In *Professional Judgement: A Reader in Clinical Decision Making* (Dowie E.J. & Elstein A., eds), Cambridge University Press, Cambridge, pp. 45–59.

Erev I., Wallsten T.S. & Budescu D.V. (1994) Simultaneous over- and under-confidence: the role of error in judgment processes. *Psychological Review* **101**(3), 519–527.

Friedman C.P., Gatti G.G., Franz T.M., Murphy G.C., Wolf F.M., Heckerling P.S., Fine P.L., Miller T.M. & Elstein A.S. (2005) Do physicians know when their diagnoses are correct? Implications for decision support and error reduction. *Journal of General Internal Medicine* **20**, 334–339.

Goldhill D.R. (2001) The critically ill: following your MEWS. *QJM: Monthly Journal of the Association of Physicians* **94**(10), 507–510.

Graber M.L., Franklin N. & Gordon R. (2005) Diagnostic error in internal medicine. *Archives of Internal Medicine* **165**(13), 1493–1499.

Grager M. (2005) Diagnostic errors in medicine: a case of neglect. *Joint Commission Journal on Quality and Patient safety* **31**(2), 106–113.

Griffin D. & Brenner L. (2004) Perspectives on probability judgment calibration. In *Blackwell Handbook of Judgment and Decision Making* (Koehler D.J. & Harvey N., eds), Blackwell Publication, Oxford, pp. 177–199.

Haffer A.G. & Raingruber B.J. (1998) Discovering confidence in clinical reasoning and critical thinking development in baccalaureate nursing students. *Journal of Nursing Education* **37**(2), 61–70.

Hamers J.P., Van den Hout M.A., Halfens R.J., Abu-Saad H.H. & Heijltjes A.E. (1997) Differences in pain assessment and decisions regarding the administration of analgesics between novices, intermediates and experts in paediatric nursing. *International Journal of Nursing Studies* **34**(5), 325–334.

Hammond K.R., Hamm R.M., Grassia J. & Pearson T. (1997) Direct comparison of intuitive and analytical cognition in expert judgement. In *Research on Judgement and Decision Making* (Goldstein W. & Hogarth R.M., eds), Cambridge University Press, New York, pp. 144–180.

Hausman C.L., Weiss J.C., Lawrence J.S. & Zeleznik C. (1990) Confidence weighted answer technique in a group of paediatric residents. *Medical Teacher* **12**(2), 163–168.

Independent Healthcare Association (2002) *Guidance on Comprehensive Critical Care for Adults in Independent Sector Acute Hospitals*. Independent Healthcare Association, London.

Johnson C.G., Levenkron J.C., Suchman A.L. & Manchester R. (1988) Does physicians uncertainty affect patient satisfaction? *Journal of General Internal Medicine* **3**(2), 144–149.

Juslin P. (1993) An explanation of the hard-easy effect in studies of realism of confidence in one's general knowledge. *European Journal of Cognitive Psychology* **5**, 55–71.

Juslin P. (1994) The over-confidence phenomenon as a consequence of informal experimenter-guided selection of almanac items. *Organizational Behaviour and Human Decision Processes* **57**, 226–246.

Juslin P. & Olsson H. (1997) Thurstonian and Brunswikian origins of uncertainty in judgment: a sampling model of confidence in sensory discrimination. *Psychological Review* **104**(2), 344–366.

Keren G. (1991) Calibration and probability judgments: conceptual and methodological issues. *Acta Psychologica* **77**, 217–273.

Kissinger J.A. (1998) Over-confidence: a concept analysis. *Nursing Forum* **33**(2), 18–26.

Kruglanski A.W., Peri N. & Zakay D. (1991) Interactive effects of need for closure and initial confidence on social information seeking. *Social Cognition* **9**, 127–148.

Lichtenstein S. & Fischhoff B. (1977) Do those who know more also know more about how much they know? *Organizational Behaviour and Human Decision Processes* **20**, 159–183.

Lichtenstein S. & Fischhoff B. (1980) Training for calibration. *Organizational Behaviour and Human Performance* **26**, 149–171.

Lichtenstein S. & Fischhoff B. (1982) Calibration of probabilities: the state of the art to 1980. In *Judgment Under Uncertainty: Heuristics and Biases* (Kahneman D., Slovic P. & Tverksy A., eds), Cambridge University Press, Cambridge, pp. 306–334.

McKenna H. (1997) *Nursing Theories and Models*. Routledge, London.

McMurray A. (1992) Expertise in community health nursing. *Journal of Community Health Nursing* **9**(2), 65–75.

Mumpower J.L. & Stewart T.R. (1996) Expert judgment and expert disagreement. *Thinking and reasoning* **2**, 191–211.

Petrusic W.M. & Baranski J.V. (1997) Context, feedback, and the calibration and resolution of confidence in perceptual judgments. *The American Journal of Psychology* **110**(4), 543–575.

Sackett D.L. (1997) *Evidence-Based Medicine: How to Practice and Teach EBM*. Churchill Livingstone, Edinburgh.

Schön D.A. (1987) Teaching artistry through reflection-in-action. In *Educating the Reflective Practitioner* (Schön D.A., ed.), Jossey-Bass Publishers, San Francisco, California, pp. 22–40.

Soll J.B. (1996) Determinants of over-confidence and miscalibration: the roles of random error and ecological structure. *Organizational Behaviour and Human Decision Processes* **65**(2), 117–137.

Subbe C.P., Kruger M., Rutherford P. & Gemmel L. (2001) Validation of a modified Early Warning Score in medical admissions. *QJM: Monthly Journal of the Association of Physicians* **94**(10), 521–526.

Thompson C. (2003) Clinical experience as evidence in evidence-based practice. *Journal of Advanced Nursing* **43**(3), 230–237.

Torgerson D.J. & Campbell M.K. (2000) Use of unequal randomisation to aid the economic efficiency of clinical trials. *British Medical Journal* **321**, 759.

Yaniv I., Yates J.F. & Mith J.E.K. (1991) Measures of discrimination skill in probabilistic judgement. *Psychological Bulletin* **110**, 611–617.

Zakay D. (1997) Post-decisional confidence – can it be trusted? In *Decision Making: Cognitive Models and Explanations* (Ranyard R., ed.), Routledge, London, pp. 233–247.

Appendix 1: Clinical vignette questionnaire

Clinical Information

Mr Robert Wright, 63 years old and 76 kg weight, was presented to the emergency room in your hospital, accompanied by his wife. He was generally feeling unwell, with a tender abdomen and vomited after each meal for past 2 days. He was born in England and he has been married for 38 years. He is a senior engineer in an automotive company. He has no food or medical allergies. There was no report of use of medications. He has no significant past medical history or history of mental illness. The details of family history are unclear. The following sets of information are available to you when you assess Mr Wright on admission. Please make your judgements for each scenario.

The sets of scenarios used the following units:
Systolic blood pressure mmHg;
Heart rate bpm (beats per minute);
Respiratory rate bpm (breaths per minute);
Temperature °C.

An example scenario

Systolic blood pressure 80		Risk (circle)	
Heart rate 118	Yes		No
Respiratory rate 38			
Temperature 36·8			
Conscious level reacting to pain	Confidence (0–100)		