Establishing scales of perceived severity for clinical situations during anaesthesia

T. C. Walsh1* and P. C. W. Beatty2

1Computer Officer (Applications), Manchester Computing Centre, Kilburn Building, Oxford Road, Manchester M13 9PL, UK. 2Imaging Science and Biomedical Engineering, The Stopford Building, The University of Manchester, Oxford Road, Manchester M13 9PL, UK

*Corresponding author. E-mail: Tanya.Walsh@manchester.ac.uk

Background. In many clinical, teaching, and research situations it would be useful to have graded scales of the urgency or other subjective properties for clinical situations that can arise during anaesthesia. Such a scale could serve as a reference point for the appropriate mapping of the urgency in audible alarms or visual warnings, provide a basis for training of graduated difficulty during anaesthesia simulation, provide a benchmark in risk assessment exercises, guide prioritization of decisions in computerized decision support systems for anaesthesia and help in assessing anaesthetist occupational stress.

Methods. A questionnaire-based instrument was developed to assess the perceived severity of a range of anaesthetic clinical situations. Four scales were tested: the severity of the situation for the patient, the urgency of response required by the anaesthetist, attention required by the anaesthetist, and anxiety experienced. Over 300 anaesthetists in three cohorts of 100 were consulted in the selection of the situations to be studied. The final version of the questionnaire, which included 25 situations, was circulated to a further 229 anaesthetists for validation. The pair-wise relationships of the four properties and hence their independence, was examined using Kendall’s τ and correlation analysis.

Results. The subjective assessments of urgency and attention were closely related, as were severity and anxiety. Comparing the mean rank for the severity scale with the subjective risk scores revealed a statistically significant correlation of τ=0.647 (P<0.01).

Conclusion. Subjective assessment of severity by anaesthetists was found to be consistent across the clinical situations studied.

Keywords: anaesthesia; complications; alarms

Accepted for publication: April 25, 2005

In many clinical, teaching, and research situations it would be useful to have graded scales of the urgency for clinical situations that can arise during anaesthesia. For instance such a scale could:

- Provide a benchmark in risk assessment exercises.
- Provide a basis for training situations of graduated difficulty during anaesthesia simulation.
- Help in assessing occupational stress in anaesthetists.
- Serve as a reference point for the appropriate mapping of urgency in audible alarms or visual warnings.
- Guide prioritization of decisions in computerised decision support systems for anaesthesia.

This paper details the development of a subjective scale of severity for a range of clinical anaesthesia situations.

The scale was constructed using a questionnaire reported previously in pilot form by the authors.1

Methods

The questionnaire

The questionnaire in its final form is shown in the appendix (see Supplementary material in British Journal of Anaesthesia online). The rubric to the questionnaire requested that the anaesthetists imagine that they were conducting routine anaesthesia for an ASA II patient when one of the clinical situations listed occurred. This was done to eliminate the effect of modifying the clinical urgency of the situation perceived by the anaesthetists with the underlying condition.
of the patient. Clinical situations were described in simple phrases. The questionnaire was divided into two sections for each clinical situation. In the first section the respondents were asked to grade their response to the clinical situations on a scale of 1 to 10 where 1 corresponded to extremely low and 10 to extremely high.

As urgency is not the only possible attribute of a situation that has an effect on the actions taken by an anaesthetist; responses to three additional subjective properties of the situations, commonly cited in critical incident reports, were included. These were: the perceived severity of the situation, the anxiety induced in the anaesthetist and the level of attention required treating the situation. Thus, the study was capable of generating four independent scales.

The second section of the questionnaire sought to establish a second estimator of severity using a worst-case scenario. Respondents were asked to rate each clinical situation on the basis of their subjective estimate of the worst-case severity of the outcome for the patient on a scale of 1 (no effect) to 6 (death), and the likelihood of that outcome occurring on a similar scale of 1 (never occur) to 6 (always occur). By multiplying the two scores together a new score of severity may be derived, which is more liability-oriented than the general subjective assessment of severity obtained using the first section of the questionnaire. It is related to the assessment of legal liability for negligence propounded in the Hand Formula (i.e. that B, reasonable burden of avoiding a risk=P, the probability of the risk resulting in an injury×L, the severity or loss incurred from the injury caused) and to psychological models of how people deconstruct the assessment of overall risk of an adverse outcome in a complex decision making environment.

The final section of the questionnaire was a free-form section for comments and asked the grade of the responding anaesthetist.

The study reported was performed in 1999 before research governance regulations in the National Health Service (NHS) made it a requirement for Ethical Committee approval for studying NHS staff. After taking advice from the Secretary of the University of Manchester Ethical Committee, it was agreed that the study could proceed without explicit Ethical Committee approval, given that response to the questionnaire was purely voluntary. All data were kept out explicit Ethical Committee approval, given that response to the questionnaire was purely voluntary. All data were kept under the requirements of the Data Protection Act. There was no necessity for responders to return their names and addresses before results were analysed. Where names and addresses were supplied as a way of expressing willingness to participate in future work, these were separated from the questionnaire responses before analysis so that responder responses could not be identified.

Clinical situations

Over 300 anaesthetists were consulted in the selection of the situations for inclusion in the scale, using an iterative method of development and refinement, until no further change to the questionnaire was deemed necessary. This involved circulation to three cohorts of 100 anaesthetists (cycles) selected randomly from a list of responders to a previous questionnaire. An initial questionnaire was constructed by taking advice of local anaesthetists, which gave a basic list of nine situations. This version of the questionnaire was informally piloted on 10 trainee anaesthetists attending the FRCA Primary Course at The University of Manchester, which led to the questionnaire for cycle 1. This questionnaire included 15 situations. In the subsequent cycles, new situations were included, using suggestions made by responders. Removal or inclusion of a situation before the next cycle was influenced by the reliability of scores obtained and comments made in the free-form part of the questionnaire. These comments were particularly useful in clarifying the descriptions of the situation. The order of the clinical situations in the questionnaire and their properties was changed with each cycle to check that order of appearance did not affect responder scores. In all, 32 situations were considered of which 25 were included in the stable, final version of the questionnaire. This stable version of the questionnaire was circulated to a further 229 anaesthetists for validation.

Statistical analysis

Because of the ordinal nature of the data, initially descriptive non-parametric statistics were used to examine the distribution of the data for each of the four situational properties measured. The pair-wise relationships of these properties and hence their independence, was examined using Kendall’s τ and correlation analysis.

A rank order was established using the median values for individual situations. However, this does not give a particularly tractable scale against which to map the properties. As a result, we have followed the recommendations of Edworthy, who, when working in the area of urgency mapping of audible alarm sounds, suggested that the rank order of properties produces the most practical mapping. In the final scales, we have ranked the situations in order using Pearson’s rank to increase usability and sensitivity.

The second estimator of severity derived from the second part of the questionnaire was compared using Kendall’s τ and correlation analysis to the subjective severity rank derived from the first section of the questionnaire. This gave a measure of the internal reliability of the severity scale.

Results

From the final mailing of 229 questionnaires, 99 responses were received, a response rate of 43%. A follow up reminder letter 1 month later elicited a further 22 responses, resulting in an overall response rate of 53% or 121 responders.

Graphical examination of the descriptive statistics for the four scales (Table 1) showed that there was a strong correlation between the severity and anxiety scales, and the urgency and attention scales. With the exception of anticipated apnoea (severity and attention τ=0.125; anxiety and attention τ=0.148), correlations among the properties
were statistically significant (P<0.01) and in a positive direction for all clinical situations. The magnitude of correlation coefficients was greatest for the urgency and attention pairing, ranging from \(t=0.839\) for malignant hyperpyrexia to \(t=0.479\) oxygen supply failure, with the exception of anticipated apnoea and pneumothorax, where the magnitude was greatest for the severity and anxiety pairing \((t=0.647, 0.676, 0.594)\). This was followed by the severity and anxiety pairing for all clinical situations apart from cardiac arrest, kinked/displaced ET tube, malignant hyperpyrexia and profound bronchospasm, where the severity and anxiety pairing followed in magnitude \((t=0.497, 0.455, 0.570, 0.601)\ respectively), and profound hypotension and severe intraoperative haemorrhage where the anxiety and urgency pairing followed in magnitude \((t=0.532, 0.516)\).

Mean rank for each situation on each scale in descending order of rank on the severity scale is shown in Table 2. Comparing the mean rank for the severity scale with the subjective risk scores revealed a statistically significant correlation of \(t=0.647\) \((n=18, P<0.01)\).

### Discussion

**Validity and robustness of the scales**

The experimental design laid emphasis on establishing a good range of situations important to the majority of anaesthetists. These situations were provided by the anaesthetists themselves without prompting or otherwise constraining their choices. The range of mean ranks in the final scales indicates that this was achieved. Though relatively lengthy, the resulting questionnaire took under 15 minutes to complete.
similar scales in other studies. 67

compares favourably with sample sizes used to establish questionnaire in its development phases. This sample size responders who contributed to the framing of the final calibration sample was 121, backed up by the 159 other In this context, our sample size is reasonably large. The truly represents views of the population of all anaesthetists. In this context, our sample size is reasonably large. The calibration sample was 121, backed up by the 159 other responders who contributed to the framing of the final questionnaire in its development phases. This sample size compares favourably with sample sizes used to establish similar scales in other studies.67

Sample bias is a fundamental problem in this type of study. We are only allowed by the Data Protection Act to keep, and therefore contact, people who have replied previously to one of our surveys. Thus, it is reasonable to assume that they are particularly interested in this type of research. However, approaching a random sample of anaesthetists would not eliminate this type of bias, since the interested ones would be most likely to respond. Comparing demography between studies using the same database provides some internal check on consistency of sample but does not eliminate the possibility of bias. Other internal checks on consistency were performed. The inter-quartile ranges for the situations within the scales were narrow indicating that there was little variability in the responses. This is particularly true for those clinical situations that received the highest median ratings. The number of outlying and extreme points, at an average over the four scales of 1.6% of individual responses, was low. We would expect a much higher spread of inter-quartile range if the scales were the result of idiosyncratic choices by responders, rather than the expression of a common view. The severity score from the second part of the questionnaire was significantly correlated with the naïve perceived severity score, indicating acceptable internal consistency. However, the correlation attained accounts for only about 41% of the variation between the scores, suggesting that there might be other factors not captured by the perceived severity scale that affect anaesthetists’ over all perception of the severity of a situation.

The final issue of robustness is the differences shown between the nature and characteristics of the situations in the final questionnaire. The situations generated by the responders fall into two types: those that are observable and objective (e.g. hypotension) and those that are secondary derivations (e.g. myocardial infarction, pneumothorax). We do not believe this dichotomy affected the validity of the scales. The study was strictly contextualized and designed to assess perceived severity of an emerging clinical situation within anaesthesia. Great care was taken in the development of the scales not to impose any limits on the anaesthetists. However, the existence of the two types of situation draws us to consider a fundamental human factors issue about the decision making process, regarding integration of information received from patient monitors with other clinical information. The questionnaire, supported by findings in psychology literature, was predicated on the hypothesis that anaesthetists contextualize all the information they require for a decision and therefore do not see any dichotomy between these two different types of situation. As noted above, our results support this view. There is no indication from the free-form part of the questionnaires that responders wanted to distinguish between the two types of situation. Responders had the opportunity to reject, as well as accept, situations but there is no indication that they did so in a systematic way between the different types of situation. If there was uncertainty between the two sorts of situation then we would have expected the consistency statistics to be poorer. We conclude that the scales represent usably robust measures of their relevant properties and the best estimates of perceived property strength currently available.

Applications of the scales

The scales clearly have a role in the area of assessment of clinical risk. Ideally, clinical risk should be assessed using objective evidence-based assessments of incidence, outcome or possibly economic impact. Unfortunately, the four scales map different aspects of perceived clinical risk. All attempts to derive objective measures of risk associated with the situations to allow comparison of our scales with actual risk proved fruitless. The sources of data from which to derive such objective measures of risk (audit studies, CEPOD, incident reporting studies, etc.) are insufficiently detailed to separate out the specific situations identified by the anaesthetists in this study. There are also conflicts between methodologies in different studies. Thus, to establish the detailed relationship between the perceived strengths of the properties represented by the scales and objective measures of risk for the situations in clinical practice would require a separate tailor-made study. However, in similar circumstances in medicine, where objective evidence is unavailable, the opinions of a panel of experts are often substituted. The scales presented here can be viewed as the opinion of a wide panel of experts.

The use of anaesthesia simulators for critical incident training in anaesthesia is now familiar. Potentially, the scales offer new dimensions to this process. First, by graduating the severity of the simulations, an individual’s training may be made progressive, allowing the objective assessment of performance and exposure to a full range of incident of difficulties. Secondly, the scales allow the simulations to be assessed for urgency, anxiety induced, and attention required, as well as severity.
The scales could be used in the management of occupational stress by providing the starting point for a scoring system of stress exposure based on the anxiety scale. It is logical to assume that occupational stress in anaesthesia increases in relation to the number and nature of critical incidents encountered. By combining an extended anxiety scale of the type presented here with an incident reporting system, a measure of anxiety exposure might be developed giving the potential to study the relationship between stress exposure and performance in normal clinical practice. If a sufficiently sensitive system of stress exposure measurement of this type could be developed it has the potential to enable stress management to be added to other considerations of workload and work pattern already identified as contributing to safety in anaesthesia.

The scales have a particular relevance to audible warnings on patient monitoring systems. The purpose of warnings is to invoke appropriate actions from clinicians by directing attention to the problem and give an indication of the urgency (urgency mapping). To do this, different forms of sounds have been assigned to different aspects of monitoring and their urgency manipulated by varying pitch, repetition rate and frequency content so that the intrinsic urgency of the sounds map to the urgency of the situation. Design recommendations for these aspects of audible warnings have been incorporated into international standards. In designing well urgency mapped warning sounds, it has been argued that it is the perceived urgency of the situation, not its objective urgency, that is most important. The methodology of designing well urgency mapped warning sounds in aviation has shown great promise but is yet to be thoroughly validated in anaesthesia. The scales presented here may be of importance in the design of more effective warnings for use in anaesthesia.

With some qualifications, we have met the objectives of establishing a calibrated scale of severity, urgency, attention, and anxiety for anaesthetic clinical situations, and can conclude that the subjective assessment of the four properties by anaesthetists is reasonably consistent. We conclude that the scales represent usably robust measures of their relevant properties and the best estimates of perceived property strength currently available.

### Supplementary data: appendix

The appendix can be found as supplementary data in British Journal of Anaesthesia online.

### References

2. Judge Learned Hand. United States v. Carrol Towing, 159 F.2d 169 (2d Circ 1947); 1947
9. EN 60601-1-8 Medical electrical equipment Part 1–8: General requirements for safety. CENELEC, Brussels, 2004