

## Use of a new task-relevant test to assess the effects of shift work and drug labelling formats on anesthesia trainees' drug recognition and confirmation

## Nouveau test pertinent à la tâche afin d'évaluer les effets du travail par quarts et des formats d'étiquetage des médicaments sur les capacités d'identification et de confirmation des médicaments par les stagiaires en anesthésie

James F. Cheeseman, PhD · Craig S. Webster, PhD · Matt D. M. Pawley, PhD · Margaret A. Francis, PhD · Guy R. Warman, PhD · Alan F. Merry, FANZCA

Received: 26 April 2010 / Accepted: 12 October 2010 / Published online: 13 November 2010  
© Canadian Anesthesiologists' Society 2010

### Abstract

**Background** Drug administration errors occur in every aspect of clinical practice. Using a novel task-relevant Medication Recognition and Confirmation Test (MRCT), we investigated the effects on performance of working night and day shifts and labelling different drug formats.

**Methods** Anesthesia trainees ( $n = 18$ ) participated in one of two experiments during an 8-12 hr day shift and an 8-12 hr night shift. In Experiment-1 ( $n = 10$ ), we compared standardized colour-coded labels with pictures of ampoules. In Experiment-2 ( $n = 8$ ), we compared colour-coded labels with black and white labels. Sleep was measured with wrist actigraphy during both day and night shift runs over seven to eight days. The MRCT outcome measures were reaction times and drug errors.

**Results** In the two experiments, colour-coded labels were recognized (and therefore selected) more quickly than pictures of conventional ampoules (mean difference 332 msec, 95% confidence interval (CI) 242-422 msec;  $P < 0.0001$ ) and faster than black and white labels (mean difference 96 msec, 95% CI 46-146 msec;  $P < 0.0001$ ). Participants obtained less sleep while working night shifts than while working day shifts (mean difference 57 min, 95% CI 0:15-1:39 hr;  $P = 0.013$ ). Mean confirmation reaction times were slower during night shifts than during day shifts (mean difference 60 msec, 95% CI 1-120 msec;  $P = 0.048$ ). No differences in error rates were observed between shifts or among drug label types.

**Conclusions** Label format influenced recognition and confirmation reaction times to representations of drugs in this study, and we found some evidence to suggest that performance is better during day shifts than during night shifts. The task-relevant test evaluated here may have further application in measuring performance in the wider clinical setting.

J. F. Cheeseman, PhD · C. S. Webster, PhD ·  
M. D. M. Pawley, PhD · G. R. Warman, PhD ·  
A. F. Merry, FANZCA (✉)

Department of Anaesthesiology, Faculty of Medical and  
Health Sciences, University of Auckland, Private Bag 92019,  
Auckland, New Zealand  
e-mail: a.merry@auckland.ac.nz

M. D. M. Pawley, PhD  
Institute of Information and Mathematical Sciences, Massey  
University, Albany, New Zealand

M. A. Francis, PhD  
Department of Psychology, University of Auckland,  
Auckland, New Zealand

A. F. Merry, FANZCA  
Auckland City Hospital, Auckland, New Zealand

### Résumé

**Contexte** Les erreurs d'administration médicamenteuse surviennent dans toutes sortes de situations de la pratique clinique. À l'aide d'un nouveau Test de d'identification et de confirmation des médicaments (MRCT) pertinent à la tâche, nous avons examiné les effets de quarts de travail de nuit et de jour sur les performances, ainsi que les effets de différents formats d'étiquetage de médicaments.

**Méthode** Des stagiaires en anesthésie ( $n = 18$ ) ont pris part à l'une de deux expériences pendant un quart de

travail de jour de 8-12 h et un quart de nuit de 8-12 h. Dans l'expérience 1 ( $n = 10$ ), nous avons comparé des étiquettes standardisées avec codes de couleur à des étiquettes avec des images d'ampoules. Dans l'expérience 2 ( $n = 8$ ), nous avons comparé des étiquettes avec codes de couleur à des étiquettes en noir et blanc. Le sommeil a été mesuré à l'aide d'actigraphie au poignet pendant les quarts de jour et de nuit sur une période de sept ou huit jours. Les résultats mesurés par le MRCT étaient les temps de réaction et les erreurs médicamenteuses.

**Résultats** Dans les deux expériences, les étiquettes avec codes de couleur ont été identifiées (et donc sélectionnées) plus rapidement que les étiquettes présentant des images d'ampoules conventionnelles (différence moyenne 332 msec, intervalle de confiance (IC) 95 % 242-422 msec;  $P < 0,0001$ ) et plus rapidement que les étiquettes en noir et blanc (différence moyenne 96 msec, IC 95 % 46-146 msec;  $P < 0,0001$ ). Les participants ont moins dormi lorsqu'ils travaillaient les quarts de nuit que lorsqu'ils ont travaillé les quarts de jour (différences moyennes 57 min, IC 95 % 0:15-1:39 h;  $P = 0,013$ ). Les temps de réaction de confirmation moyens étaient plus longs pendant les quarts de nuit que pendant les quarts de jour (différence moyenne 60 msec, IC 95 % 1-120 msec;  $P = 0,048$ ). Aucune différence n'a été observée dans les taux d'erreurs entre les quarts ou entre les différents types d'étiquettes.

**Conclusion** Le format de l'étiquette a eu une influence sur les temps d'identification et de réaction de confirmation aux représentations médicamenteuses dans cette étude, et certaines des données observées suggèrent que la performance est meilleure pendant les quarts de jour que pendant les quarts de nuit. Le test pertinent à la tâche évalué ici pourrait également être appliqué pour mesurer la performance dans un contexte clinique plus vaste.

Error in drug administration pervades every aspect of clinical practice, notably in anesthesia.<sup>1-5</sup> There is reasonable agreement about the basic principles of safe administration of medications<sup>6</sup> but relatively little empirical evidence to underpin these principles. Studying medication error in clinical settings is difficult, because errors, although unacceptably frequent, do not occur often enough to lend themselves to study in prospective trials that are moderate in size. In recent years, the effects of fatigue and time of day have received considerable attention as causes of performance impairment.<sup>7-10</sup> Interns in the United States working "traditional" 24-hr shifts suffered from double the number of failures in concentration than those working a 17-hr shift,<sup>11</sup> and during the night shift, they made 36% more medical errors of a serious nature.<sup>12</sup>

This impairment in performance was attributed to the effects of sleep deprivation and the circadian clock, which promotes sleep at night.

As a result of these concerns, the work hours of junior doctors in New Zealand have been restricted since 1985 to a maximum of 17 continuous hours and a total of 72 hr per week. Many of the studies of the effects of fatigue are based on tests, such as the psychomotor vigilance test, that seem to have little to do with clinical medicine in general or anesthesia in particular. Perhaps unsurprisingly, Howard *et al.*<sup>13</sup> demonstrated a significant decrement in the performance of anesthesiologist in the psychomotor vigilance test after 25-30 hr of sleep deprivation, but they demonstrated no change in their performance in a high-fidelity human patient simulator. Conventional cognitive tests (CogState<sup>TM</sup>) have been used to demonstrate reduced performance of anesthesiologists after working several consecutive night shifts,<sup>14</sup> but these tests are relatively time-consuming to complete during work periods and, arguably, do not relate directly to the task of drug administration. Legible drug labels that are colour-coded according to international standards have been adopted widely as an aid to avoiding drug error,<sup>15</sup> and with some overall success, they have been incorporated into a multifaceted system designed to reduce error in anesthesia.<sup>16</sup> However, the influence of drug label format alone on drug administration error has not been definitively elucidated.

Consequently, we have developed a novel task-relevant simulated drug recognition and confirmation test to measure times and error rates of simulated drug recognition and confirmation. In this study, we describe the test and examine two null hypotheses: 1) Speed of drug recognition and accuracy of drug selection, as measured by the new test in terms of reaction time and error rate, does not differ between three different presentations of drug identity; and 2) These measures do not differ between day and night shifts.

## Methods

### Setting, ethics, and consent

Approval was obtained from the Auckland Regional Ethics Committee (ref 97/032) and the University of Auckland Human Participants' Ethics Committee (ref 2005/051). Written consent was obtained from all participants. The study was conducted at Auckland City Hospital (ACH), Auckland, New Zealand from 2005 to 2008. The ACH is a 760-bed tertiary hospital associated with the University of Auckland and credentialed for training by the Australian and New Zealand College of Anaesthetists (ANZCA).

## Participants

Eligible participants were anesthesia trainees allocated to one of four sub-departments at ACH. The ACH participates in a regional training scheme that has slightly more than 90 anesthesia trainees at any one time, all of whom are allocated to this sub-department (15 each time) at some point in their training. These trainees are rostered for three or four consecutive night shifts at a time, interspersed with variable (but longer) periods of day shifts. Data were collected from each participant over a seven-day study period that spanned both day and night shifts. Different participants took part in Experiment-1 (undertaken in the earlier part of the study) and Experiment-2 (undertaken in the later period).

## Sleep data

Participants used actigraphs (Actiwatch-L, Cambridge Neurotechnology Inc, Cambridge, UK) to track their time spent sleeping, which were cross-checked with sleep diaries.

## The medication recognition and confirmation test

The computer-based Medication Recognition and Confirmation Test (MRCT) was designed with reference to established criteria for neurocognitive testing.<sup>17</sup> The user interface is a touch screen to measure the times and accuracy of responses to linked pairs of questions involving (Question One) recognition and selection and (Question Two) confirmation of the identity of pictures of medication labels in any desired format.

In Question One (recognition and selection), each participant is asked to choose a picture out of four different options that corresponds to a medication named in the centre of the screen (Fig. 1A). In Question Two (confirmation), the participant is presented with a single picture selected from the four options with the prompt, “Do you want to administer this drug?” (Fig. 1B). The picture presented in this “confirmation” question is usually the one selected, but one of the other three pictures is substituted in a small percentage of cases. The substitution rate can be varied, and the substitutions can occur randomly, thereby increasing the need for vigilance and the potential for error. The number of questions provided in a single test can also vary.

In this study, the tests were set to 200 iterations of paired pictures consisting of 50 medications used commonly in anesthesia at ACH. Two experiments were conducted (see below). For each experiment, presentation in each pair was

in one of two formats (A and B) specific to that experiment (Figs 1A and 2). Format A and Format B were presented randomly in equal proportions according to a sequence of computer-generated numbers, such that participants could not predict the format that was likely to follow. Each of the 50 medications appeared four times during each test (twice within the 100 Format A question pairs and twice within the 100 Format B question pairs). In the medication confirmation questions, the substitution rate for a false picture was set arbitrarily to 5% overall (60 out of 1,200 questions per participant), and the ratio of these substitutions was 20% intra-class (for example within the class of neuromuscular blockers) and 80% inter-class (for example between neuromuscular blockers and opioids), class being defined by the standard international colour code for user-applied drug labels.<sup>18</sup>

## Participant handling

All participants were familiarized with the MRCT test and practiced the test before commencing the study. During training and before each trial, participants were instructed to strive for maximum speed and accuracy. Testing was undertaken on the third shift in a consecutive series of day or night shifts, respectively. Each participant acted as his/her own control. The start shift run (either nights or days) was determined by the roster, and each test battery was conducted three times (at the start, middle, and end of the shift). To avoid any possible differences in processor speed, the same dedicated desktop computer was used to collect all MRCT data. All testing took place in a quiet room adjacent to the operating rooms. Participants were instructed to use their dominant hand to respond to every question.

## Experiment-1: colour-coded labels vs ampoules

Format A: Pictures of standardized labels colour-coded according to a New Zealand and Australian standard<sup>19</sup> (identical to standards registered in the United States, Canada, and the United Kingdom),<sup>20-22</sup> with both the class and the name of the drug displayed in a large clear font (e.g., “Opioid” and “Fentanyl”). Less salient details, including those required by regulation, were displayed in smaller fonts. These labels incorporated a barcode and were designed in variants for use on prefilled syringes, on ampoules as flag labels, and as user-applied labels for syringes<sup>15</sup> (Safer Sleep, Auckland, NZ).

Format B: Photographs of conventional ampoules with care taken to produce the best possible resolution and clarity (Fig. 2). All ampoules were represented 1:1 with their physical equivalents.

**Fig. 1** Example screen shots of the Medication Recognition and Confirmation Test. (A) The name of a drug (in this case etomidate) is shown in the centre of the screen, and the participant is asked to choose the correct label by touching the appropriate picture. (B) Having chosen the correct drug, the participant is asked to indicate whether he/she wishes to administer the displayed drug (which is usually the drug chosen, but may be a different drug displayed on a random basis) by touching the Yes or No buttons on the screen



Experiment-2: colour-coded labels vs black and white labels

Format A: The same as Experiment-1, Format A.  
 Format B: Similar to Experiment-1, Format A, but with black and white labels instead of colour (Fig. 2).

Data analysis

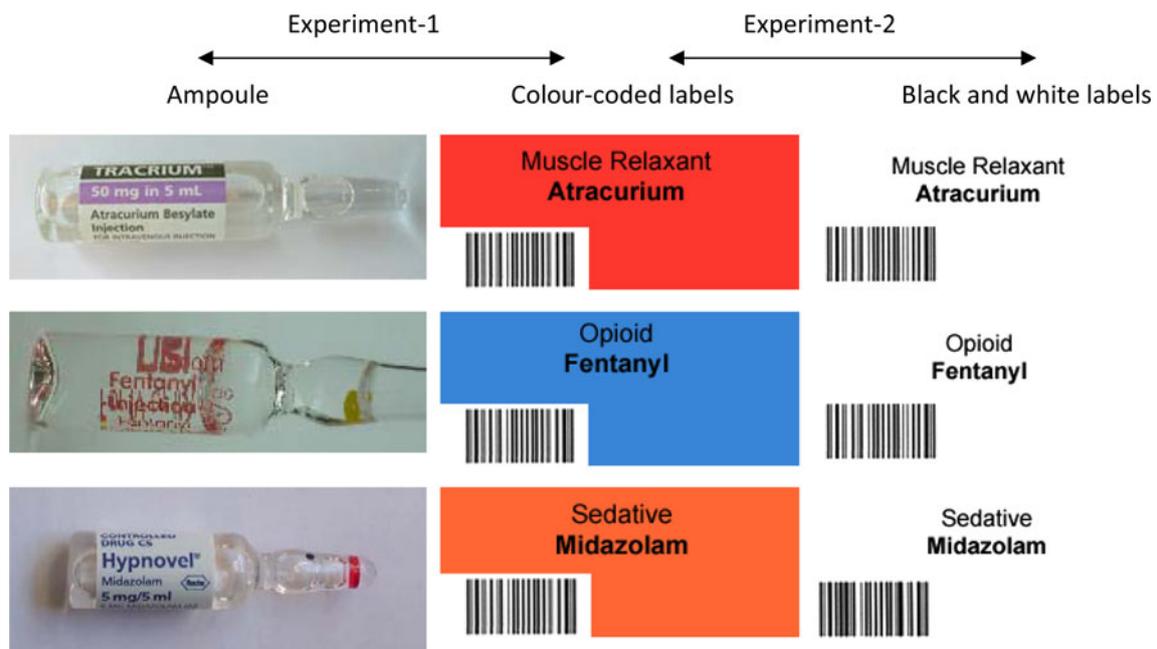
*Software*

Statistical analysis was undertaken with SPSS v.17 (SPSS, Chicago, IL, USA) and the statistical package R, version 2.8.<sup>23</sup> Average daily sleep gained during each shift type

was calculated using Sleep Analysis Software v. 5.0 (Cambridge Neurotechnology Inc, Cambridge, UK).

Data reduction

For each participant, the median reaction time (MRT) in milliseconds (msec) was calculated for the responses to each recognition question in each format (Question One) and each confirmation question in each format (Question Two). The MRT was chosen purposely after examining the distributions of data, because it was a valid measure of central tendency<sup>10</sup> and because transforming the data would make interpretation of the reaction times less intuitive. The total number of errors was calculated for each



**Fig. 2** Three example medications were used in each experiment. In Experiment-1, colour-coded labels were tested against pictures of conventional ampoules. In Experiment-2, colour-coded labels were

matched with black and white labels. Ampoules and labels are not shown to scale here, but they appear 1:1 with physical equivalents in the computer test

participant for each of Questions One and Two for each format.

#### Label type comparisons

When examining differences in label types, separate analyses were performed on each of the following outcome measurements: median recognition MRT, median confirmation MRT, and total errors within each experiment. Outcomes were measured for each participant within a shift-type at a shift-stage, so tests were based on the paired differences between the label types. Participants were modelled as levels of a random factor, as they were considered to have been sampled from the population of anesthetic trainees in the region's training program. Linear mixed-effects models with random intercepts were fitted to the distribution of differences using a normal error structure (identity link-function) and maximum likelihood parameter estimation. The model initially included the following covariates: 1) shift-type (night vs day) 2) shift-stage (start, middle, or end), and 3) sleep duration (minutes) as fixed effect covariates. Covariates were dropped from the model if there was no evidence of an effect at the  $\alpha = 5\%$  level.

When outliers were present, we used the non-parametric Wilcoxon signed-rank test or the sign test (the latter test was used when data had multiple ties) on the paired differences (averaged across shift-stage for each participant), as indicated in the text. We obtained the exact confidence intervals (CI) for the median difference of the population

by the algorithm described by Bauer,<sup>24</sup> and we designated the level of significance, alpha, as 5%.

The sample size estimate was not undertaken because no prior data were available for this purpose. Instead, we recruited as many participants as practical during the time the first author was completing his PhD.

#### Day-night comparisons

When examining differences in shift-type (day vs night), the analysis was carried out on the combined Format A data (i.e., coloured label data) amalgamated from both experiments using a paired Student's *t* test.

### Results

Fourteen trainees consented for Experiment-1 and ten trainees consented for Experiment-2. No trainee who was approached at the outset declined to participate, but four and two trainees, respectively, were then unable to participate because of rostering exigencies. At the end of the study, valuable MRCT data were available from 18 subjects all aged < 40 yr (ten in Experiment-1: five male, five female and eight in Experiment-2: three male, five female); actigraphy data were available from 12 of these participants (five in Experiment-1 and seven in Experiment-2). Missing data were attributable to lack of compliance in completing the tests and equipment failure.

### Sleep data

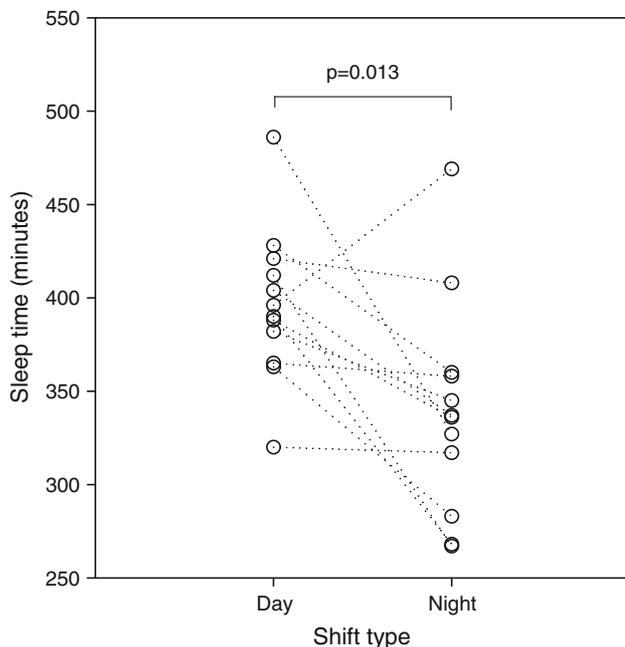
Actigraphy data were in concordance with sleep diaries (sleep diary data not shown). Participants slept an average of 6:36 hr during day shifts compared with 5:39 hr during night shifts. This result corresponds to 57 min less sleep (95% CI 0:15-1:39 hr;  $P = 0.013$ , paired Student's  $t$  test) per 24-hr period while working night shifts than while working day shifts (Fig. 3).

### Experiment-1: colour-coded labels vs ampoules

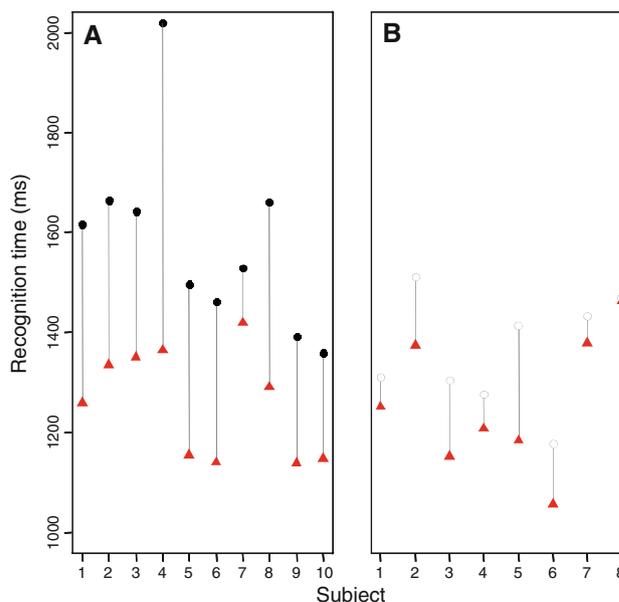
The mean recognition MRT for colour-coded labels was 332 msec faster (95% CI 242-422 msec;  $P < 0.0001$ , mixed-model analysis), shift phase omitted, than that for conventional ampoules (Figs 4A and 5). When shift phase was included in the model, the evidence for a difference in mean recognition MRT remained strong ( $P < 0.001$ , mixed-model analysis), but the magnitude of the difference diminished as shift-stage progressed from early through mid to late (Table). Only four medications were recognized more quickly on average in ampoule format than in label format (metoclopramide, potassium chloride, tenoxicam, and water) (Fig. 5).

Mean confirmation MRT for colour-coded labels was 40 msec faster (95% CI 15-66 msec;  $P = 0.0028$ , mixed-model analysis) than for ampoules (Fig. 6A).

There were 38 errors, 17 errors with the colour-coded label format and 21 errors with the ampoule format (with no



**Fig. 3** Paired comparisons of mean sleep obtained by anesthesiologists during day and night shifts. Mean difference was 57 min (95% confidence interval 0:15-1:39 hr;  $P = 0.013$ , paired Student's  $t$  test)



**Fig. 4** Parallel line plot comparing individuals' median reaction times in drug recognition. Panel (A) Experiment-1 comparison between colour-coded labels ( $\blacktriangle$ ) and conventional ampoules ( $\bullet$ ) ( $n = 10$ ). Panel (B) Experiment-2 comparison between colour-coded labels ( $\blacktriangle$ ) and black and white labels ( $\circ$ ) ( $n = 8$ )

evidence of a difference in proportions;  $P = 0.47$ , mixed-model analysis). There was no evidence of the effects of shift-stage order on confirmation MRT or error rate. No drug stood out as being more error prone than the others.

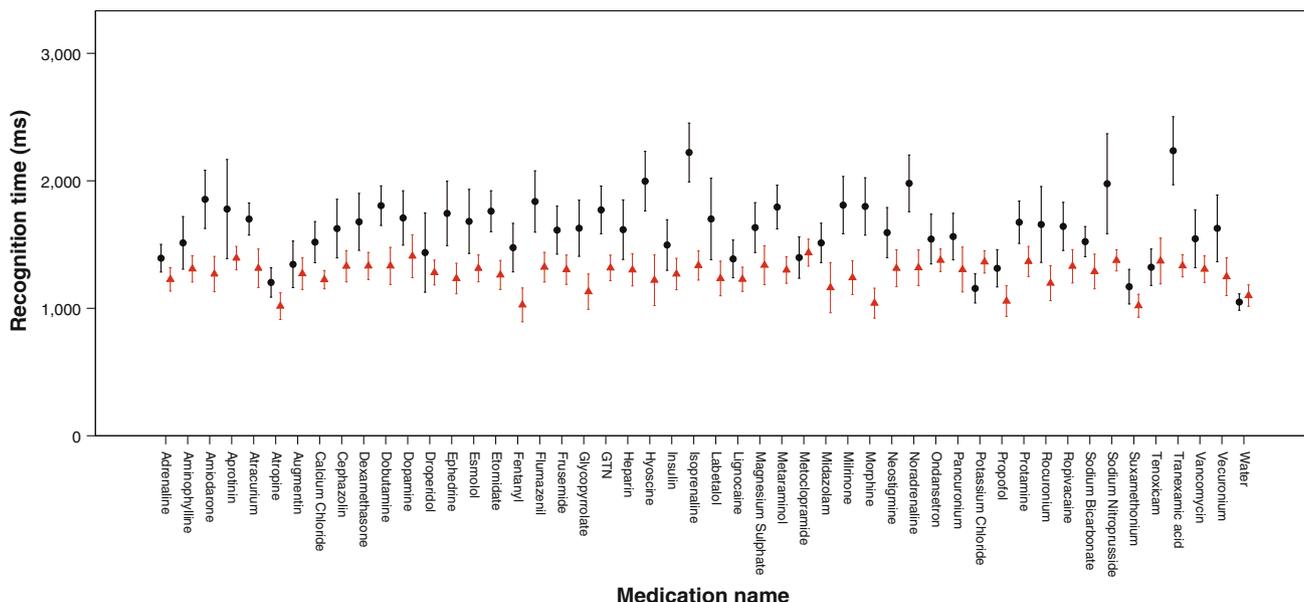
### Experiment-2: colour-coded labels vs black and white labels

The mean recognition MRT for colour-coded labels was 96 msec faster (95% CI 46-146 msec;  $P < 0.0001$ , mixed-model analysis) than that for black and white labels (Fig. 4B). Mean confirmation times for colour-coded labels were 16 msec faster (95% CI 6-26 msec;  $P = 0.0023$ , mixed-model analysis) than that for black and white labels (Fig. 6B). There was no evidence of the effects of shift-stage on recognition or confirmation MRT.

A total of 76 errors were made, 42 with the colour-coded label format and 34 with the black and white label format ( $P = 0.07$ , Wilcoxon signed-rank test).

### Day vs night comparisons: analysis of amalgamated colour-coded label format data (Format A in both experiments)

There were 20,600 responses to the MRCT questions for colour-coded labels (Format A) (10,300 recognition questions and 10,300 confirmation questions). Sleep duration, shift-stage, and order showed no evidence of affecting MRT, so these factors were dropped from the initial model.



**Fig. 5** Experiment-1 comparison of medication recognition times between colour-coded labels (▲) and conventional ampoules (●) (*n* = 10). Each point represents the mean of all participants’ median reaction times to each medication. Error bars indicate 95% confidence intervals

**Table** The magnitude of the difference between colour-coded labels and ampoules was dependent on shift stage

Ampoules – colour-coded labels	Estimate (msec)	(95% CI) (msec)	Contrast
Start shift	419	(321 , 518)	↕ A
Mid shift	326	(227 , 424)	↕ B
End shift	250	(151 , 348)	↕ C

Contrasts between shift stages are denoted by the letters A-C. A: mean -94 msec, 95% confidence interval (CI) -164 to -24; B: mean -170 msec, 95% CI -240 to -100; C: mean -76 msec, 95% CI -146 to -6; *P* < 0.001 in each case. (In Experiment-2, no effects were seen for stage of shift [colour-coded vs black and white] or for confirmation times)

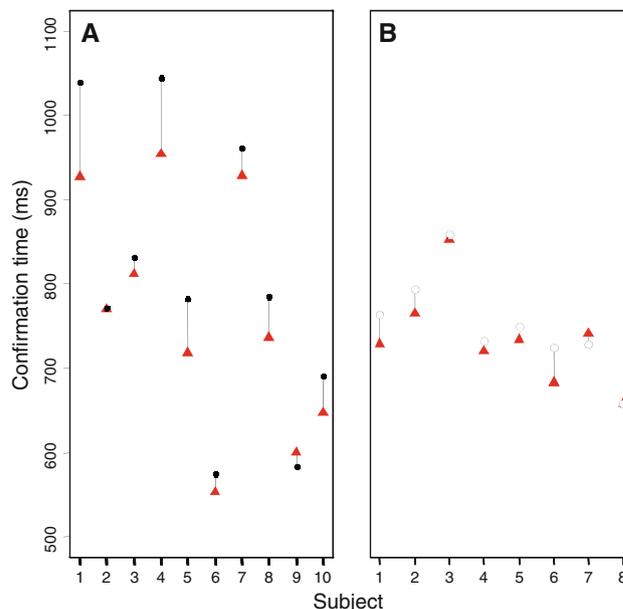
The mean recognition MRT was 56 msec slower (95% CI -2 to 115 msec; *P* = 0.06, paired Student’s *t* test) during night shifts than during day shifts (Fig. 7).

Mean confirmation MRT was 60 msec slower (95 % CI 1-120 msec; *P* = 0.048, paired Student’s *t* test) during the night than during the day (Fig. 7). Analysis of recognition and confirmation errors together revealed no evidence of a difference (*P* = 0.29, mixed-model analysis)

Fifty-nine errors were made with colour-coded labels, 28 during day shifts and 31 during night shifts.

**Discussion**

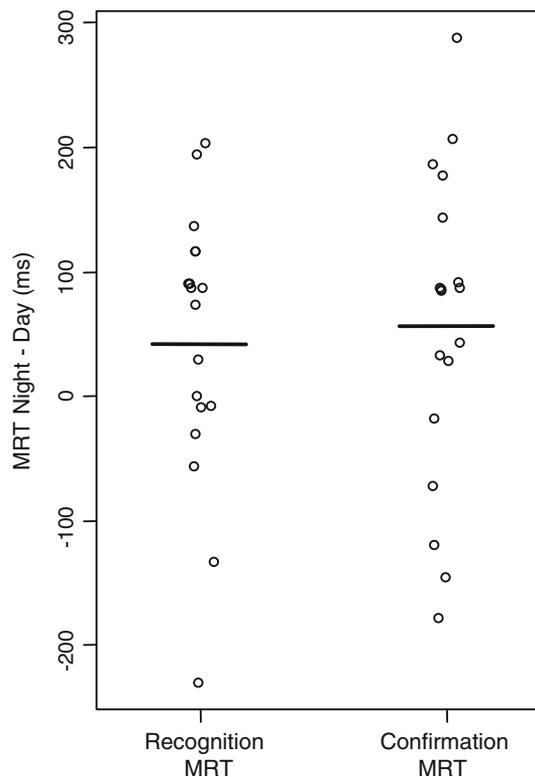
In Experiment-1, photographs of colour-coded drug labels were selected more quickly than photographs of ampoules,



**Fig. 6** Parallel line plot comparing individuals’ median reaction times in drug confirmation. Panel (A) Experiment-1 comparison between colour-coded labels (▲) and conventional ampoules (●) (*n* = 10). Panel (B) Experiment-2 comparison between colour-coded labels (▲) and black and white labels (○) (*n* = 8)

and they were also confirmed more quickly as the required drug when subsequently presented. However, there was no difference in error rates between these two formats.

In Experiment-2, selection and confirmation was also quicker with colour-coded labels than with equivalent black and white labels. Forty-two errors were made with



**Fig. 7** The distribution of differences (night – day) in median reaction times for recognition responses and confirmation responses ( $n = 18$ ). Points are shown perturbed on the x-axis to aid in discrimination. Horizontal lines denote the means of the group sample

the colour-coded labels vs 34 with black and white labels ( $P = 0.07$ ).

In the first Experiment-1, the difference in selection times diminished as the shift progressed from early to mid to late. This result may have been an effect of fatigue, but no other effects from shift phase were identified, so it is difficult to speculate on this theory. We tested for effects of order and found none, suggesting no important learning component in the test.

Participants obtained almost one less hour of sleep while on night shifts than while on day shifts (Fig. 3). Both drug recognition time ( $P = 0.06$ ) and drug confirmation time ( $P = 0.048$ ) were slightly slower during night shifts. However, we found no differences between the night and day shifts with regard to error rates.

The value of these data lies in the clinical relevance of the simulated task. The participants in this study were participating in standard rostered shifts, and the only deviation from normal clinical practice was the requirement to perform our tests. We did not control for caffeine, alcohol consumption, or sleep medication. We were surprised that almost all participants obtained considerably less than the eight hours of sleep considered usual<sup>25</sup> on

both day and night shifts. The relatively small difference between day and night shifts in hours of sleep (approximately one hour) can be attributed to the fact that participants often had the opportunity for sleep while on night duty. Recent data suggest that continuously restricting sleep to six hours each night can have a deleterious effect on vigilance equivalent to 24-hr acute sleep deprivation,<sup>26,27</sup> and self-awareness of reduced performance tends to plateau.<sup>27,28</sup> The trainees who were studied in our hospital had a tendency to fall within this category, at least on night shifts, despite the relatively rigorous restrictions on work hours in New Zealand. Impairment of anesthesiologists' performance due to loss of sleep has been shown previously in a similar New Zealand setting.<sup>10</sup>

Taken collectively, these findings suggest an advantage in favour of the labels that were studied (characterized by a standardized layout, a large clear font, and various other features)<sup>15</sup> over most of the varied presentations manufacturers use on their ampoules. The results also provide limited support for the use of colour-coding, although the advantage conferred by this was restricted to speed of selection and confirmation rather than to increased accuracy.

A limitation of this study lies in its relatively small number of participants (the logistics of collecting data from practicing clinicians during actual periods of work at different times of night and day were challenging) and the fact that they all were a part of a single regional training scheme. Therefore, our findings may not apply beyond the population studied. The strength of this study is its paired design - each individual acted as his/her own control. Also, we were able to collect data from many repetitions of the test.

Questions relating to the presentation of drug information may be easier to study than those relating to time of day. We took great care to make the photographs as clear as possible, but results could be different if the actual ampoules were to be tested against the actual labels. It is also possible that performance in a clinical context could differ from that reported here. On the one hand, performance could be worse if there were distractions; on the other hand, perhaps participants would perform better if they were making decisions with consequences for real patients. The absolute reaction and confirmation times in this study are relatively arbitrary, and one might argue that differences of 330 msec (recognition times) or 40 msec (confirmation times) between colour-coded labels and ampoules are not clinically relevant in themselves. However, these times reflect the mental effort required to complete the tests,<sup>13</sup> and we therefore infer that quicker reaction times to standardized drug representations correspond with easier recognition of the drugs. Over a large number of administrations, quicker reaction times are

likely to facilitate the avoidance of errors. We were surprised not to find a difference in error rates between label types. Despite being instructed to complete the tests as quickly and accurately as possible, the participants in the study knew we were investigating drug error and may have sacrificed speed in favour of accuracy.

The value of colour-coding in promoting safety in drug administration continues to be debated.<sup>29,30</sup> The fact that there is room for improvement in the legibility and distinctiveness of some ampoules is less controversial<sup>2</sup> – but it remains very difficult to persuade regulators or manufacturers to act on the well recognized problem of so-called “look-alike sound-alike”<sup>31</sup> drugs.

The experiments described here employ a simple form of simulation. Complexity in simulation can be varied, from relatively simple task-oriented simulations such as this through to immersion high-fidelity simulation.<sup>32</sup> By capitalizing on the strengths of each level (in this case low cost and ease of multiple tests) the same questions can be studied in different ways. Thus, the methods presented here could be used to obtain the information needed to perfect the presentation of drugs, i.e., investigating such features as font size, tall man lettering, and the use of auditory as well as visual presentation of key information.<sup>15</sup> One could also test for error rates between drugs that look the same rather than using random substitutions as we did in the present trial. Once this has been accomplished, more complex simulation methods could then be used to verify that these findings apply in the wider context of administering a full anesthetic<sup>6</sup> before finally confirming them in a clinical setting.

In conclusion, we have presented a novel task-relevant test of performance in relation to drug administration. We have shown that there are advantages to having legible standardized drug labels that are colour-coded for class of drug, and we have added to the evidence that shift effects (and likely fatigue associated with sleep restriction) influence performance in an anesthetic context.

**Acknowledgements** We are grateful for the participation of the clinical staff from the Department of Anaesthesia, Auckland City Hospital, Auckland District Health Board, New Zealand. We sincerely thank Miss Milica Milovanovic for her assistance with data collection.

**Financial support** J.F.C. received a University of Auckland Doctoral Scholarship and a New Zealand Vice-Chancellors’ Committee William Georgetti Scholarship to conduct this work. Miss Milovanovic received a University of Auckland summer studentship stipend. This project was funded through an equipment grant from the Maurice and Phyllis Paykel Trust. A.F.M. and C.S.W. have a financial interest in Safer Sleep, Auckland, New Zealand.

**Competing interests** None declared.

## References

1. Fernandez CV, Esau R, Hamilton D, Fitzsimmons B, Pritchard S. Intrathecal vincristine: an analysis of reasons for recurrent fatal chemotherapeutic error with recommendations for prevention. *J Pediatr Hematol Oncol* 1998; 20: 587-90.
2. Webster CS, Merry AF, Larsson L, McGrath KA, Weller J. The frequency and nature of drug administration error during anaesthesia. *Anaesth Intensive Care* 2001; 29: 494-500.
3. Abeyssekera A, Bergman IJ, Kluger MT, Short TG. Drug error in anaesthetic practice: a review of 896 reports from the Australian Incident Monitoring Study database. *Anaesthesia* 2005; 60: 220-7.
4. Wheeler SJ, Wheeler DW. Medication errors in anaesthesia and critical care. *Anaesthesia* 2005; 60: 257-73.
5. Merry AF, Webster CS. Medication error in New Zealand - time to act. *N Z Med J* 2008; 121: 6-9. Available from URL: <http://www.nzma.org.nz/journal/121-1272/3015/> (accessed (August, 2010)).
6. Jensen LS, Merry AF, Webster CS, Weller CS, Larsson L. Evidence-based strategies for preventing drug administration errors during anaesthesia. *Anaesthesia* 2004; 59: 493-504.
7. Merry AF, Warman GR. Fatigue and the anaesthetist. *Anaesth Intensive Care* 2006; 34: 577-8.
8. Howard SK, Gaba DM. Trainee fatigue: are new limits on work hours enough? *CMAJ* 2004; 170: 975-6.
9. Dawson D, Reid K. Fatigue, alcohol and performance impairment. *Nature* 1997; 388: 235.
10. Gander P, Millar M, Webster C, Merry A. Sleep loss and performance of anaesthesia trainees and specialists. *Chronobiol Int* 2008; 25: 1077-91.
11. Lockley SW, Cronin JW, Evans EE, et al. Effect of reducing interns’ weekly work hours on sleep and attentional failures. *N Engl J Med* 2004; 351: 1829-37.
12. Landrigan CP, Rothschild JM, Cronin JW, et al. Effect of reducing interns’ work hours on serious medical errors in intensive care units. *N Engl J Med* 2004; 351: 1838-48.
13. Howard SK, Gaba DM, Smith BE, et al. Simulation study of rested versus sleep-deprived anesthesiologists. *Anesthesiology* 2003; 98: 1345-55.
14. Griffiths JD, McCutcheon C, Silbert BS, Maruff P. A prospective observational study of the effect of night duty on the cognitive function of anaesthetic registrars. *Anaesth Intensive Care* 2006; 34: 621-8.
15. Merry AF, Webster CS, Mathew DJ. A new, safety-oriented, integrated drug administration and automated anaesthesia record system. *Anesth Analg* 2001; 93: 385-90.
16. Webster CS, Larsson L, Frampton CM, et al. Clinical assessment of a new anaesthetic drug administration system: a prospective, controlled, longitudinal incident monitoring study. *Anaesthesia* 2010; 65: 490-9.
17. Dorrian J, Rogers NL, Dinges DF. Psychomotor vigilance performance: a neurocognitive assay sensitive to sleep loss. In: Kushida C, editor. *Sleep Deprivation: Clinical Issues, Pharmacology and Sleep Loss Effects*. NY: Marcel Dekker, Inc; 2005. p. 39-70.
18. Anonymous. Anaesthetic and respiratory equipment — User-applied labels for syringes containing drugs used during anaesthesia — Colours, design and performance (ISO 26825:2008): International Organisation for Standardization; 2008. Available from URL: <http://www.iso.org> (accessed (August, 2010)).
19. Anonymous. User-applied labels for use on syringes containing drugs used during anaesthesia (AS/NZS 4375:1996). Wellington: Standards New Zealand; 1996.

20. *Anonymous*. Standard for user-applied drug labels in anaesthesia and critical care (Z264.3-98). Etobicoke: Canadian Standards Association; 1998.
21. *Anonymous*. Standard specification for user applied drug labels in anesthesiology (D4774-94). Philadelphia: American Society for Testing and Materials; 1995.
22. *Royal College of Anaesthetists*. Syringe labelling in critical care areas. RCoA Bulletin 2003; 19: 953.
23. *R Development Core Team*. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2009 .
24. *Bauer DF*. Constructing confidence sets using rank statistics. J Am Stat Assoc 1972; 67: 687-90.
25. *Young T, Palta M, Dempsey J, Skatrud J, Weber S, Badr S*. The occurrence of sleep-disordered breathing among middle-aged adults. N Engl J Med 1993; 328: 1230-5.
26. *Banks S, Dinges DF*. Behavioral and physiological consequences of sleep restriction. J Clin Sleep Med 2007; 3: 519-28.
27. *Olson EJ, Drage LA, Auger RR*. Sleep deprivation, physician performance, and patient safety. Chest 2009; 136: 1389-96.
28. *Van Dongen HP, Maislin G, Mullington JM, Dinges DF*. The cumulative cost of additional wakefulness: dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. Sleep 2003; 26: 117-26.
29. *Fasting S, Gisvold S*. Adverse drug errors in anesthesia, and the impact of coloured syringe labels. Can J Anesth 2000; 47: 1060-7.
30. *Webster CS, Merry AF*. Colour coding, drug administration error and the systems approach to safety. Eur J Anaesthesiol 2007; 24: 385-6.
31. *Filik R, Purdy K, Gale A, Gerrett D*. Labeling of medicines and patient safety: evaluating methods of reducing drug name confusion. Hum Factors 2006; 48: 39-47.
32. *Cumin D, Merry AF*. Simulators for use in anaesthesia. Anaesthesia 2007; 62: 151-62.