Applied Ergonomics 75 (2019) 230–242

Contents lists available at ScienceDirect

Applied Ergonomics

journal homepage: www.elsevier.com/locate/apergo

Eye-tracking reveals *how* observation chart design features affect the detection of patient deterioration: An experimental study

Lillian Cornish[a], Andrew Hill[a,b,*], Mark S. Horswill[a], Stefanie I. Becker[a], Marcus O. Watson[a,b,c]

[a] School of Psychology, The University of Queensland, St Lucia, Brisbane, Queensland, 4072, Australia
[b] Clinical Skills Development Service, Metro North Hospital and Health Service, Herston, Brisbane, Queensland, 4006, Australia
[c] School of Medicine, The University of Queensland, Herston, Brisbane, Queensland, 4006, Australia

ARTICLE INFO

Keywords:
Clinical deterioration
Patient safety
Human factors

ABSTRACT

Particular design features intended to improve usability – including graphically displayed observations and integrated colour-based scoring-systems – have been shown to increase the speed and accuracy with which users of hospital observation charts detect abnormal patient observations. We used eye-tracking to evaluate two potential cognitive mechanisms underlying these effects. Novice chart-users completed a series of experimental trials in which they viewed patient data presented on one of three observation chart designs (varied within-subjects), and indicated which observation was abnormal (or that none were). A chart that incorporated both graphically displayed observations *and* an integrated colour-based scoring-system yielded faster, more accurate responses *and* fewer, shorter fixations than a graphical chart without a colour-based scoring-system. The latter, in turn, yielded the same advantages over a tabular chart (which incorporated neither design feature). These results suggest that both colour-based scoring-systems and graphically displayed observations improve search efficiency *and* reduce the cognitive resources required to process vital sign data.

1. Introduction

1.1. Patient deterioration

Patients who experience clinical deterioration during their hospital stay are at greater risk for a range of serious adverse events, including unplanned intensive care unit (ICU) admission, cardiac or respiratory arrest, and unexpected death (Franklin and Mathew, 1994; Goldhill et al., 1999; Hillman et al., 2001). However, such outcomes are often preventable as they are frequently preceded by derangements in the patient's vital signs that are detectable up to 48 hours beforehand (Endacott et al., 2007; Hillman et al., 2001, 2002).

Paper-based bedside observation charts are a type of clinical form commonly used to support patient monitoring in inpatient settings. These charts, which are typically filled in by nurses, are cognitive artifacts that store and display data for a range of physiological parameters including vital signs such as heart rate, respiratory rate, blood pressure, and temperature. Consequently, they have the potential to play a critical role in assisting nurses and doctors in the early recognition of deteriorating patients, if their designs are optimized to fulfil this objective. Indeed, there is a growing body of empirical evidence demonstrating that observation chart design can have a substantial impact on chart-users' ability to detect abnormal observations that may indicate clinical deterioration (Chatterjee et al., 2005; Preece et al., 2012b; Christofidis et al., 2013, 2014, 2015, 2016).

1.2. Clinician-led approaches to observation chart design

Hospital observation charts have traditionally been designed by local clinical staff perceived to have relevant knowledge or expertise. Consequently, the design (and design quality) of these forms has tended to vary substantially between different local area hospital services, different hospitals, and even different wards within individual facilities (Chatterjee et al., 2005; Preece et al., 2013). Although the importance of regular patient observations has been recognized by clinicians since at least the time of Florence Nightingale (1860), it was not until the 21st century that the first study was published in which an "evidence-based approach" to observation chart design had been employed (i.e. Chatterjee et al., 2005).

Chatterjee et al. (2005) were a team of medical and nursing staff who set out to design a chart with the explicit aim of improving the detection of patient deterioration. They began by conducting an experiment in which they asked clinicians to identify abnormal observations in patient data presented on five different charts used in their

* Corresponding author. Clinical Skills Development Service, Metro North Hospital and Health Service, Herston, Brisbane, Queensland, 4006, Australia.
E-mail address: Andrew.Hill@health.qld.gov.au (A. Hill).

https://doi.org/10.1016/j.apergo.2018.10.005
Received 25 September 2017; Received in revised form 9 October 2018; Accepted 14 October 2018
Available online 12 November 2018

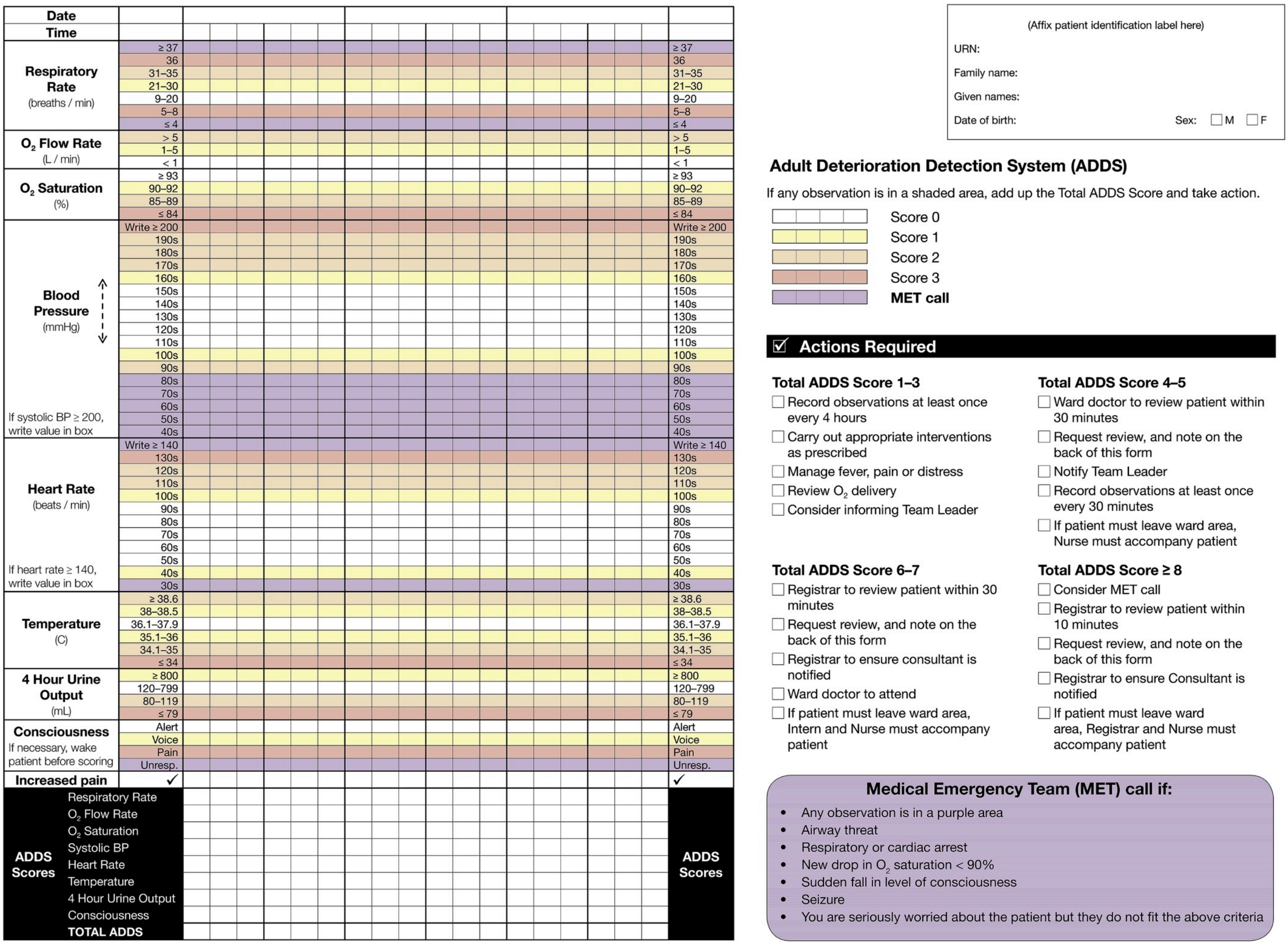


Fig. 1. The version of the ADDS chart upon which the experimental materials were based (inside pages only).

hospital, and measured their detection rates. Subsequently, the team created a new chart, justifying their design choices based on a combination of the empirical data obtained from the experiment and subjective chart preferences gleaned from participant interviews. Finally, they repeated the experimental task with the new chart, demonstrating significantly increased detection rates for abnormal observations across several vital signs.

In subsequent years, other initiatives involving clinician-led, evidence-based approaches to addressing the deteriorating patient problem through chart design were found to be associated with reductions in adverse events (Mitchell et al., 2010) and improved documentation of vital signs (Cahill et al., 2011; Mitchell et al., 2010). However, in each case, it was unclear to what extent the results could be attributed to chart design as opposed to the accompanying training program.

1.3. An interdisciplinary human factors approach to observation chart design

In contrast to prior clinician-led initiatives, Horswill et al. (2010a) conducted the first published program of research in which an observation chart was designed using an explicit human factors approach. This work was conducted for the *Australian Commission on Safety and Quality in Health Care* (ACSQHC) as part of a national initiative to develop a standardized adult form to improve the detection of patient deterioration.

As a precursor to the chart design process itself, an interdisciplinary team of human factors specialists and clinicians (including three individuals with experience using patient observation charts) conducted a heuristic evaluation to systematically compare the usability of a representative sample of 25 existing charts from Australia and New Zealand (Preece et al., 2013). Since there were no previously published usability heuristics for observation charts (or any other paper-based clinical forms), the team began by drafting a set of design rules addressing potential usability challenges associated with these forms. Most of the design rules were adapted from published usability heuristics in the web design and software domains (see Gerhardt-Powals, 1996; Nielsen, 1994; Zhu et al., 2005), with additional rules derived from a task analysis and a heuristic evaluation piloting process. The final set of 64 rules covered eight categories of design considerations, specifically: page layout; information layout; recording observations; integration of track-and-trigger systems; language and labelling; cognitive and memory load; use of fonts; and use of colour (for details, see Preece et al., 2013). Collectively, the rules were intended to improve usability by accommodating for the fundamental limitations of human information processing (e.g. with respect to visual perception, reading, memory, and decision-making). Hence, they embody design choices that: enhance the saliency of important information; eliminate or reduce the need for users to hold information in memory or mentally manipulate data; preference recognition over recall; and exclude redundant, ambiguous or irrelevant information, as well as other sources of extraneous cognitive load. All 25 existing charts were found to have a substantial number of usability problems (i.e. violations of the design rules), with specific issues identified across most or all categories for each chart (Preece et al., 2013).

Subsequently, the team used the output from the heuristic evaluation as a *de facto* usability guide for the design of a new observation chart, the *Adult Deterioration Detection System*, or *ADDS* (Horswill et al., 2010a). Some minor details of the design were also influenced by the preferences of a large sample of health professionals who participated in an online survey (Preece et al., 2012a). After completing an iterative design process incorporating rapid-cycle prototyping, the team subjected the ADDS chart to its own heuristic evaluation. This analysis confirmed, to their satisfaction, that the chart design addressed as many potential usability problems as practicable, with the remaining issues attributable to inevitable and carefully considered trade-offs between competing usability considerations (Preece et al., 2013). Two versions of the chart were produced for the ACSQHC (Horswill et al., 2010a), one of which is presented in Fig. 1. The two versions differed only with respect to the presence or absence of a 'look-up table' for blood pressure – an optional feature that is not relevant to the present study.

To evaluate the effectiveness of their human factors based approach to observation chart design, Horswill et al. (2010a) conducted two empirical experiments comparing the usability of the ADDS charts with that of four charts selected from the heuristic evaluation sample, including both the best design (which was superficially similar to the ADDS) and the worst (which was essentially a table of numerals). In the first experiment, groups of health professionals and novice chart-users were presented with a series of observation charts onto which real vital sign data had been transcribed (Preece et al., 2012b). For each chart, their task was to identify any abnormal observation or to indicate that none were present. The results showed that, when using ADDS charts, both participant groups were significantly faster and more accurate than with any of the other charts (which each yielded around 2.4 to 2.8 times as many errors as the best-performing ADDS, averaging over groups).

In the second experiment, participants completed a series of time-pressured rounds in which they recorded observations for six patients in a simulated hospital ward environment, using the same six charts (Horswill et al., 2010b). Again, the pattern of results was similar for health professionals and novice chart-users. Data recording error rates were low across all charts, and did not differ significantly between the ADDS and the other designs that incorporated some form of 'track-and-trigger' system to assist in recognising and responding to patient deterioration (i.e. by using recorded observations to 'track' the patient's vital signs, and established criteria to 'trigger' standard clinical-care responses to abnormalities). However, participants made fewer scoring errors on the ADDS compared with the best of the pre-existing charts (whose track-and-trigger system incorporated a similarly complex method of scoring multiple vital signs to derive a total "early-warning score" quantifying the patient's overall degree of physiological instability). Of course, the observation chart designs compared in these studies varied unsystematically along multiple design dimensions but, taken together, the results provided initial validation evidence for Preece et al.'s (2013) observation chart usability heuristics as a set, and the design process as a whole (Horswill et al., 2010a).

A multi-site replication of the Preece et al. (2012b) study yielded further empirical support for the overall human factors design of the ADDS charts (Christofidis et al., 2013). The results indicated that even clinicians who were highly-experienced with one of the alternative observation charts used in the experiment (or a similar chart) detected the presence or absence of abnormal observations faster and more accurately when using an ADDS chart, after receiving only a few minutes of video-based training. This suggests that chart design choices that account for human information processing limitations can enhance performance to a greater extent than familiarity and experience. In addition, real-world implementations of versions of the ADDS chart (currently used in 183 public hospitals and primary health care centres in Queensland, Australia,[1] as well as facilities interstate and overseas) have been associated with improved clinical outcomes, including reductions in: the rate of in-hospital cardiac arrest; the overall length-of-stay for hospital patients; the number of patients admitted to the ICU from the ward with severe deterioration (i.e. individuals with > 50% predicted mortality); and the median length-of-stay in ICU (Drower et al., 2013; Joshi et al., 2015). More subjectively, in an independent survey study conducted at 11 sites that trialled versions or derivatives of the ADDS chart, over 75% of nurses agreed that the charts helped manage deteriorating patients, and 80% agreed that the charts' use of colour helped identify patients at risk (Elliott et al., 2016).

Empirical usability studies have also been conducted to test some of Preece et al.'s (2013) specific observation chart design rules by using experimental stimuli that incorporate systematic variations in key design features. For example, contrary to popular belief among clinicians (but consistent with the relevant design rule), one study found that abnormal blood pressure and heart rate observations were detected more quickly and accurately when these two vital signs were plotted in separate graphs (rather than on the same axes), irrespective of whether the participants were experienced clinicians or novice chart-users (Christofidis et al., 2014). Another study, which employed a novice sample, yielded support for design heuristics in favour of using a graphical (vs. numerical) data-recording format, and incorporating an integrated colour-based track-and-trigger scoring-system in which the areas for recording vital sign data include coloured bands representing different gradations of physiological abnormality (see Fig. 1; Christofidis et al., 2016). Similarly, Fung et al. (2014) found that nurses and doctors tested using two different observation chart designs interpreted vital sign data faster and more accurately when it was presented graphically rather than numerically. More broadly, these applied findings are consistent with the outcomes of a review of the effects of data visualization on performance in data exploration tasks, which concluded that graphs can lead to faster decision-making, provided that the interface is congruent with the user's task (Baker et al., 2009). However, despite the mounting evidence that observation chart design features affect the detection of patient deterioration, no study to date has used eye-tracking to empirically investigate the cognitive mechanisms underlying any of these effects in this applied context.

1.4. Proposed cognitive mechanisms for improved user-performance

This paper, by way of example, will focus on two key observation chart design features that Preece et al. (2013) argued would improve the detection of clinical deterioration, and which were subsequently incorporated into the ADDS chart: (1) graphically displayed observations and (2) integrated colour-based scoring-systems. These design features were selected for three main reasons: (1) they were supported by statistically large effects in Christofidis et al.'s (2016) empirical usability study (discussed above); (2) the cognitive mechanisms proposed to underlie them are amenable to investigation via eye-tracking technology (see section 1.5, below); and (3) they could be experimentally manipulated such that the stimuli would conform to the technical limitations of the eye-tracking equipment we intended to use.[2]

For each of these design features, the potential performance benefits to chart-users were explained by Preece et al. (2013) in terms of two mechanisms, both of which were argued to be at play. First, both design features were thought to *improve search efficiency* by making relevant

[1] As at August 9, 2018.

[2] Technical limitations prevent some observation chart design features from being investigated using current eye-tracking technology. For example, in some cases, manipulations that alter the spatial distribution of observations (e.g. separate vs. overlapping graphs, separate vs. grouped scoring rows, closely vs. widely spaced observations) would place some data points too close together for eye-tracking equipment to reliably differentiate between them. This would potentially make it impossible to determine when and for how long the participant fixated on a particular observation or series of observations (e.g., blood pressure observations vs. heart rate observations).

information more visually salient, thus directing attention towards it. Specifically, it was suggested that, by presenting lower-level data as a higher-level summation, graphs make it easier for chart-users to detect important features, such as trends and criterion breaches, which are obscured by tabular data. Similarly, the colour-shaded range-rows found in some track-and-trigger scoring-systems may also draw attention towards abnormal observations. This is consistent with evidence from basic research by Williams and Reingold (2001), who found that colour was the most influential visual feature directing gaze in complex search tasks.

Second, it was suggested that both design features may *reduce cognitive workload* by making vital sign data easier to process. It was argued that, unlike graphs, tabulated observations potentially require chart-users to perform cognitively demanding mental visualizations in order to see trends in the data, including those that signal deterioration. In addition, colour-shaded range-rows were thought to eliminate the need to actively compare numerical vital signs against clinical criteria stored in memory (or listed elsewhere on the chart).

Search efficiency and cognitive workload were also considered to be among the most practically relevant aspects of performance to study in the context of observation charts because identifying critical patient information quickly is crucial in emergency situations, and it is well-known that reducing health practitioners' cognitive load decreases clinical error (Ahmed et al., 2011; Laxmisan et al., 2007).

1.5. Eye-tracking

Search efficiency and *cognitive workload* can be studied through the use of eye-tracking technology. As eye-tracking provides information about where, when and for how long people look, it has proved useful for testing the usability of a wide range of systems, including websites (Bergstrom et al., 2013), surgical interfaces (Barkana and Açık, 2014), cockpit control layouts (Hanson, 2004), and drug labels (Bojko et al., 2005). Its utility in cognitive research is also well established, particularly in the study of reading (e.g. Drieghe et al., 2005) and visual search (e.g. Zelinksy and Sheinberg, 1997). In addition, numerous eye-tracking studies have analysed decision-making in medical domains such as radiology and surgery (e.g. Hermens et al., 2013; Tien et al., 2014; Van der Gijp et al., 2017). However, few such studies have focused on clinical documentation (e.g. Brown et al., 2014; Forsman et al., 2013) and none, to the best of our knowledge, have used eye-tracking to analyse observation charts – one of the most ubiquitous and critical clinical forms.

A person's gaze behaviour or 'scan path' can be described in terms of two main elements: *fixations* and *saccades*. During a fixation, the gaze remains relatively stable, which is associated with the processing of visual information (Duchowski, 2007). Saccades are the fast eye movements that occur between fixations as the gaze moves from one target to the next (Poole and Ball, 2005; Duchowski, 2007). Search efficiency and cognitive workload have a long history of evaluation by way of eye-tracking, and can be measured in terms of *fixation count* (Luria and Strauss, 1975; Duchowski, 2007) and *fixation duration*, respectively (Salthouse et al., 1981; Goldberg and Kotval, 1999).

Regarding the former, fixation count has been established as a global measure of search efficiency in visual search tasks (Duchowski, 2007; Goldberg and Kotval, 1999; Luria and Strauss, 1975; Shen et al., 2007; Williams, 1967; Williams and Reingold, 2001; Zelinsky and Sheinberg, 1997). In the context of identifying abnormal observations on a patient chart, a lower fixation count would indicate that abnormal observations were identified with fewer glances, suggesting that attention – and therefore gaze – was directed more efficiently towards relevant information.

Fixation duration is often used to evaluate cognitive load, having demonstrated its utility in studies of visual search, reading, memory, mental calculations, learning and problem-solving (Duchowski, 2007; Eckstein et al., 2017; Gould, 1973; Moffit, 1978; Qiuzhen et al., 2014; Salthouse et al., 1981). When applied to the interpretation of observation charts, a shorter average fixation duration would indicate that less time was taken to decide whether fixated observations were normal or abnormal, consistent with lower cognitive workload requirements. Although pupil dilation and blink rate have also been validated as measures of mental effort (Chen and Epps, 2014; Eckstein et al., 2017; Zagermann et al., 2016), they are more susceptible to external factors (e.g. differences in luminance between charts could affect pupil dilation) and demonstrate high inter-individual variability (Goldberg and Wichansky, 2003). In addition, blink rate is positively correlated with task duration (independent of effort), introducing a potential confound (Martins and Carvalho, 2015). As a result, these metrics tend to be considered second-line measures of cognitive workload (Poole and Ball, 2005) and were not analysed in the current study.

1.6. Study objectives and hypotheses

The primary aim of the present study was to use eye-tracking metrics to investigate *how* graphically displayed observations and integrated colour-based scoring-systems improve user-performance. Specifically, we investigated whether these two design features: (1) improve search efficiency by directing gaze towards relevant information; and (2) make vital sign data easier to mentally process, reducing cognitive workload, as proposed by Preece et al. (2013). A clear demonstration of these mechanisms would bolster the validity of human-factors approaches to clinical form design, by demonstrating that effective designs are those that cater to ubiquitous limitations on human information processing.

Using a similar computer-based methodology to Christofidis et al. (2016), with the addition of eye-tracking, we evaluated participants' ability to detect abnormal observations on paper-based chart designs in which the presence or absence of the design features under investigation was experimentally manipulated. We hypothesised that, consistent with prior research (Preece et al., 2012b; Christofidis et al., 2013, 2016), graphically displayed observations and an integrated colour-based scoring-system would each yield increases in speed (Hypothesis 1) and accuracy (Hypothesis 2). Given the underlying mechanisms proposed by Preece et al. (2013) to explain these performance benefits, we also hypothesised that the presence of each design feature would reduce fixation counts (Hypothesis 3) and shorten fixations (Hypothesis 4).

2. Method

2.1. Design

The study employed a within-subjects experimental design, with *chart design* as the independent variable. Three designs were compared (see 2.2.2.), and the dependent measures were: (1) response time; (2) error rate; (3) fixation count; and (4) fixation duration.

2.2. Materials and procedure

2.2.1. Patient data

Forty-eight patient cases were used, each containing 18 time-points of data (enough to fill the charts completely) for 8 key parameters: respiratory rate, oxygen flow rate, oxygen saturation, blood pressure, heart rate, temperature, four-hour urine output, and consciousness. Half of the cases contained an 'abnormal observation' for one parameter (3 cases per parameter) outside the normal ranges used in the study (see Table 1; ACT Health, 2011). The remaining cases contained only normal data. The 48 cases were randomly assigned to the 3 chart designs (described in 2.2.2.), with the constraint that each design was allocated 8 'abnormal cases' (one per parameter) and 8 'normal cases'.

Table 1
Normal physiological ranges used in the experiment.

Physiological parameter	Normal range
Respiratory rate	9–20 breaths/min
Oxygen delivery	< 1 L/min (1 or greater is abnormal)
Oxygen saturation	93–100%
Systolic blood pressure	110–159 mmHg
Heart rate	50–99 beats per minute
Temperature	36.1–37.9 degrees Celsius
Four-hour urine output	120–799 mL
Consciousness	Alert

2.2.2. Observation chart designs and experimental stimuli

Three observation chart design extracts were created for the experiment, each derived from a cropped version of the ADDS chart (Fig. 1; Horswill et al., 2010a). They excluded extraneous chart features such as calling criteria, scoring keys and patient identification labels. Additionally, areas for recording dates, times, and early-warning scores were left blank so that direct inferences about the search task itself could be made from the gaze data.

The first design extract was an unedited excerpt from the chart, which therefore incorporated both graphically displayed observations *and* an integrated colour-based scoring-system (Fig. 2a). The second design extract was identical to the first, except it did not include the colour-based scoring-system (the *graphical-only chart*; Fig. 2b). The final design extract lacked both of these design features, instead using a tabulated written-number format with the irrelevant axis scales removed (the *tabular chart*; Fig. 2c). Note that, to prevent a spatial confound, we avoided compressing this design to present the tabulated data more space-efficiently. It should also be noted that the experimentally manipulated design features could not be varied factorially, as it is impossible to add an integrated colour-based scoring-system to a basic tabular chart.

The three chart design extracts were created in Adobe InDesign CS5 (Adobe Systems Incorporated, 2011). The experimental stimuli were produced by adding blue response buttons for ‘normal’ cases (see Fig. 2), and transcribing each set of patient data using a graphics tablet. To confirm that the spatial distribution of target abnormalities present in the experimental stimuli was equivalent across the three chart designs, a series of one-way ANOVAs was conducted. Results indicated that there was no significant difference between designs in the distance of abnormal observations from the top left-hand corner of the chart, nor in their vertical or horizontal distance from the chart edge (all $p \geq .92$).

2.2.3. Apparatus

SMI Experiment Center™ 3.4 software (SensoMotoric Instruments, Teltow, Germany), was used to build the chart task and display the experimental stimuli on a 23-inch 1920 × 1080 LCD screen mounted on a height-adjustable table. An SMI RED™ infrared eye-tracker, positioned beneath the screen, sampled participants' pupil and corneal reflections at 120 Hz, and SMI iView X™ 2.8 recorded the data.

Although the present study focused on the design of paper-based observation charts, we chose to display the experimental stimuli digitally for several reasons: (1) to ensure that stimuli were presented in a consistent manner across experimental trials and conditions; (2) to track gaze more accurately; (3) to allow precise, automated measurement of response times; and (4) to automate coding of incorrect and correct responses, eliminating the potential for coding errors. Since the stimuli were not intended to represent computer-based observation charts, they lacked the additional design features and interactive functionality of such tools (e.g. menu systems, multiple screens to navigate, automated alerts, etc.). Therefore, user-performance on the digitally displayed paper charts was expected to closely reflect performance on the same charts in their original printed form. Notably, other recent studies have also involved participants interpreting paper-based chart designs presented digitally (see Christofidis et al., 2014, 2016). In these studies, observation charts that conformed more closely to human factors design principles yielded faster and more accurate detection of deterioration, consistent with studies in which participants interpreted actual paper charts (Christofidis et al., 2013; Preece et al., 2012b).

2.3. Participants

The present study focused on novice performance for several reasons. First, inexperienced doctors and nurses are often the first to make decisions about deteriorating patients (Endacott et al., 2010). Second, in similar past experiments, novice chart-users from non-clinical backgrounds have consistently produced comparable patterns of results across charts (in terms of speed and accuracy) as clinicians (Preece et al., 2012b; Christofidis et al., 2013, 2014, 2016); hence, testing clinicians would not necessarily add additional value. Third, testing naïve participants avoids potential confounds related to prior chart experience. Fourth, every clinician begins clinical practice as a novice chart-user. Fifth, it is important to study the novice user of any system, because their performance may reveal the impact of usability problems to which experienced users have spent years learning to adapt.

Power analyses (G*Power 3.1.9.2: Faul et al., 2007) were conducted using effect sizes derived from Preece et al.'s (2012b) novice data. Results indicated that, for both response time (assuming $\eta_p^2 = 0.37$) and error rate (assuming $\eta_p^2 = 0.33$), a minimum of 12 participants were required to detect an effect (with 95% power, alpha = .05, and conservatively assuming no correlation between repeated-measures). However, given the lack of prior data regarding the specific set of chart designs used in the present study and their effects on fixation counts and durations, we chose to recruit substantially more.

A convenience sample of 45 novice chart-users was recruited. Fifteen came from a Brisbane university (QLD, Australia) and received course credit, while 30 were recruited by word-of-mouth from the wider Brisbane community and received no remuneration. Two were excluded (see Fig. 3), leaving a final sample of 43 (29 females, 14 males, aged 17–55 years, $M = 23.51$, $SD = 8.30$).

2.4. Data collection

Recruitment and testing were completed between June and August 2014. Participants were tested individually in a quiet room. They completed a demographic questionnaire, then watched training videos that explained the normal ranges for the 8 parameters (see Table 1) and the experiment, including how to identify and respond to normal and abnormal cases on the three charts. Participants then studied a printed summary of the normal ranges. When ready, they were tested for their memory of these ranges via a written examination. Participants who did not score 100% restudied the ranges and re-sat the test. No participant failed on the second attempt.

Next, participants were seated at the table, directly facing the computer screen. The table was lowered or raised, and the chair moved forwards or backwards, to optimise eye-tracking accuracy according to the iView X™ software. Eye-tracker calibration was conducted using the manufacturer's recommended protocol, such that measurement error was < 1° of visual angle in the horizontal and vertical planes. These thresholds were deemed acceptable in prior studies of visual search for numbers (Godwin et al., 2014) and reading (Van Diepen et al., 1995).

Finally, participants completed a series of 48 experimental trials. In each trial, they viewed a single chart displaying a unique patient case. The trials were arranged into 3 blocks, one per chart design. Each block contained 8 “normal” and 8 “abnormal” cases. The order of the blocks, and the order of trials within each block, was randomised for each participant. In each trial, the participant responded (and ended the trial) by clicking on either: (1) an observation thought to be abnormal; or (2) the blue ‘normal’ button at the bottom of the chart, if all observations were thought to be within their normal ranges (see Fig. 2). Eye-tracker calibration was repeated before each block.

(a)

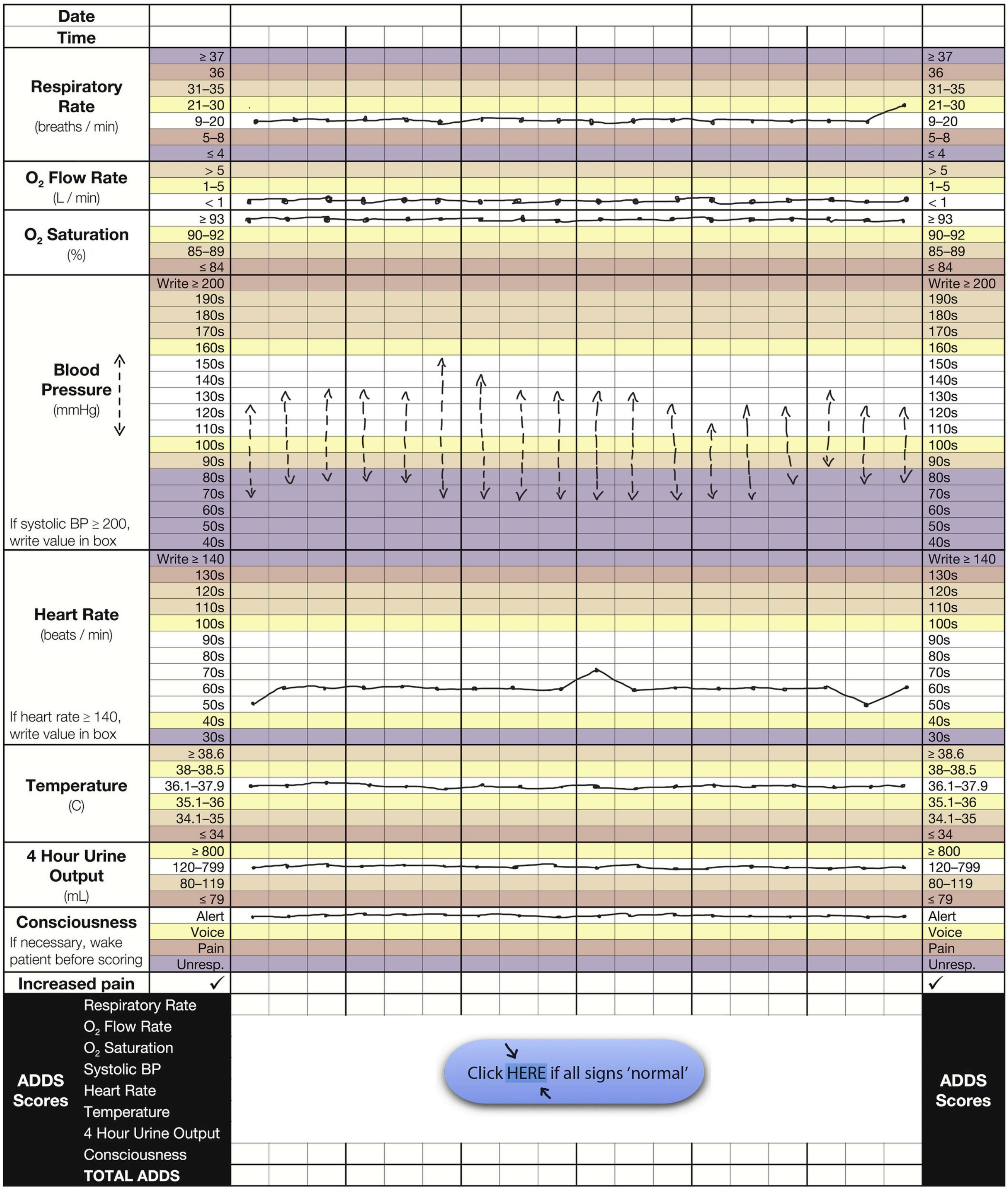


Fig. 2. Examples of the three observation chart design extracts, as seen by participants during the experiment: (a) the 'ADDS chart' (graphically displayed observations *and* an integrated colour-based scoring-system); (b) the 'graphical-only chart' (graphically displayed observations *without* a colour-based scoring-system); and (c) the 'tabular chart' (neither graphically displayed observations nor a colour-based scoring-system). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

(b)

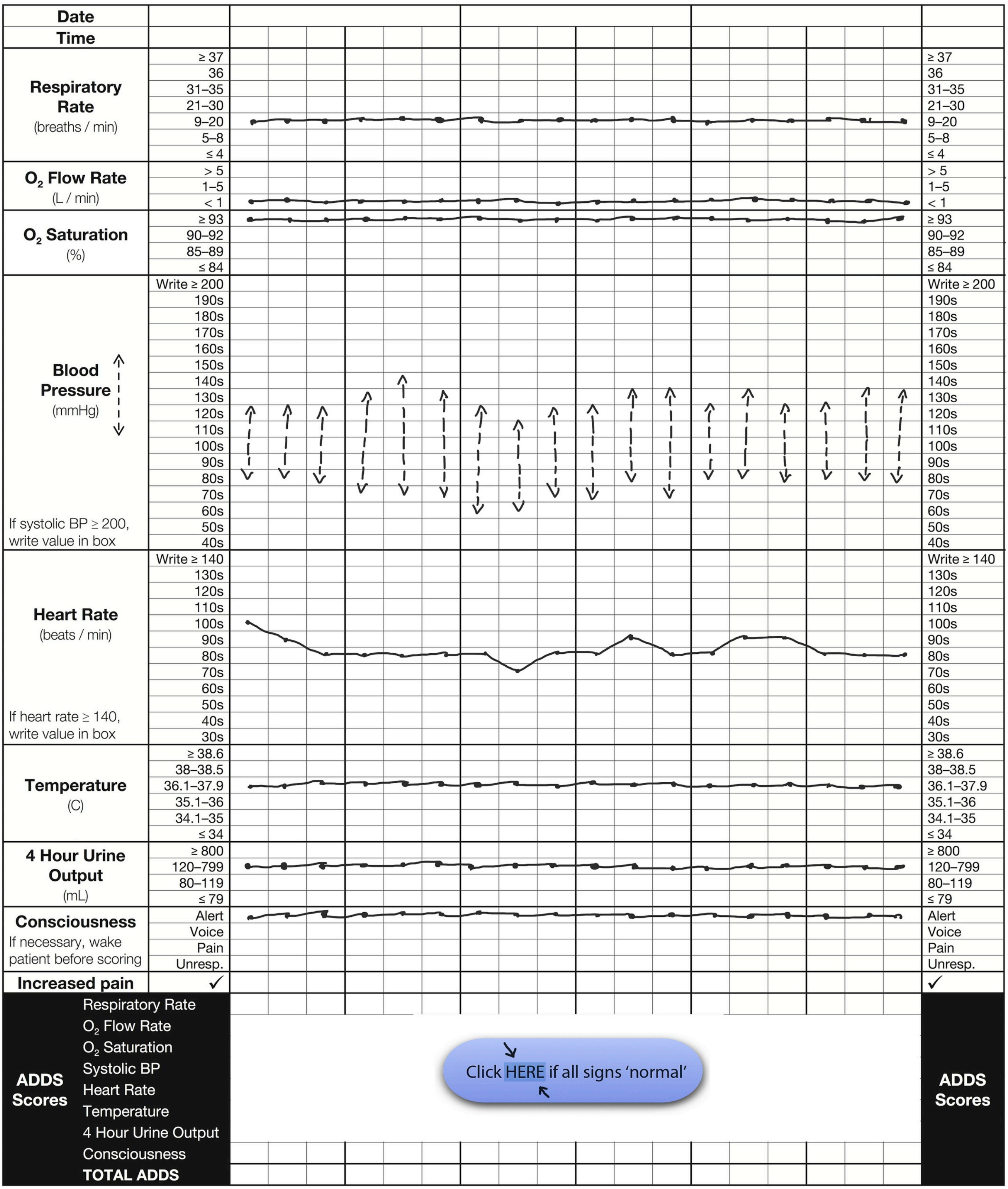


Fig. 2. (*continued*)

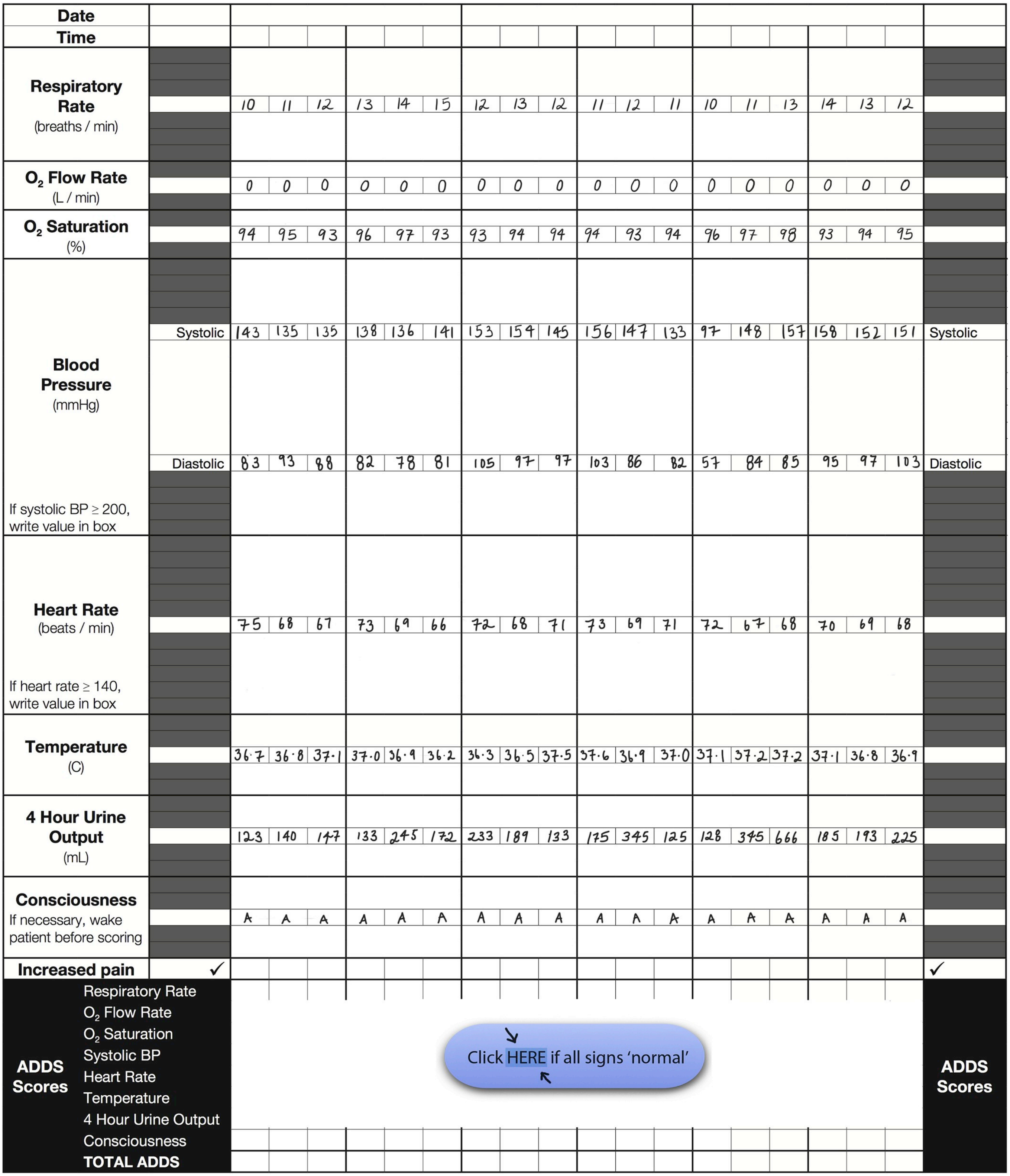
2.5. Ethical considerations

This study was granted ethics approval in accordance with the review processes of the university ethics committees.

2.6. Data analysis

Mouse-click and eye-tracking metrics were extracted from the data recordings using SMI BeGaze™ 3.4 (SensoMotoric Instruments, Teltow,

(c)



Date																		
Time																		
Respiratory Rate (breaths / min)	10	11	12	13	14	15	12	13	12	11	12	11	10	11	13	14	13	12
O_2 Flow Rate (L / min)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
O_2 Saturation (%)	94	95	93	96	97	93	93	94	94	94	93	94	96	97	98	93	94	95
Blood Pressure (mmHg) Systolic	143	135	135	138	136	141	153	154	145	156	147	133	97	148	157	158	152	151
Diastolic	83	93	88	82	78	81	105	97	97	103	86	82	57	84	85	95	97	103
If systolic BP ≥ 200, write value in box																		
Heart Rate (beats / min)	75	68	67	73	69	66	72	68	71	73	69	71	72	67	68	70	69	68
If heart rate ≥ 140, write value in box																		
Temperature (C)	36.7	36.8	37.1	37.0	36.9	36.2	36.3	36.5	37.5	37.6	36.9	37.0	37.1	37.2	37.2	37.1	36.8	36.9
4 Hour Urine Output (mL)	123	140	147	133	245	172	233	189	133	175	345	125	128	345	666	185	193	225
Consciousness If necessary, wake patient before scoring	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Increased pain ✓																		

Fig. 2. (*continued*)

Germany). Mouse-responses were coded as correct if an abnormal observation was clicked or a normal case was classified as normal. Fixations were defined as having a minimum duration of 80 ms and maximum dispersion of 200px (the BeGaze™ defaults). Each participant's mean response time, error rate (%), mean fixation count, and mean fixation duration were calculated for each chart design.

Statistical analyses were performed using IBM SPSS 21.0 (IBM Corp., Armonk, NY: USA). To test Hypotheses 1–4, one-way repeated-measures ANOVAs (comparing chart designs) were conducted on the response time, error rate, fixation count, and fixation duration data,

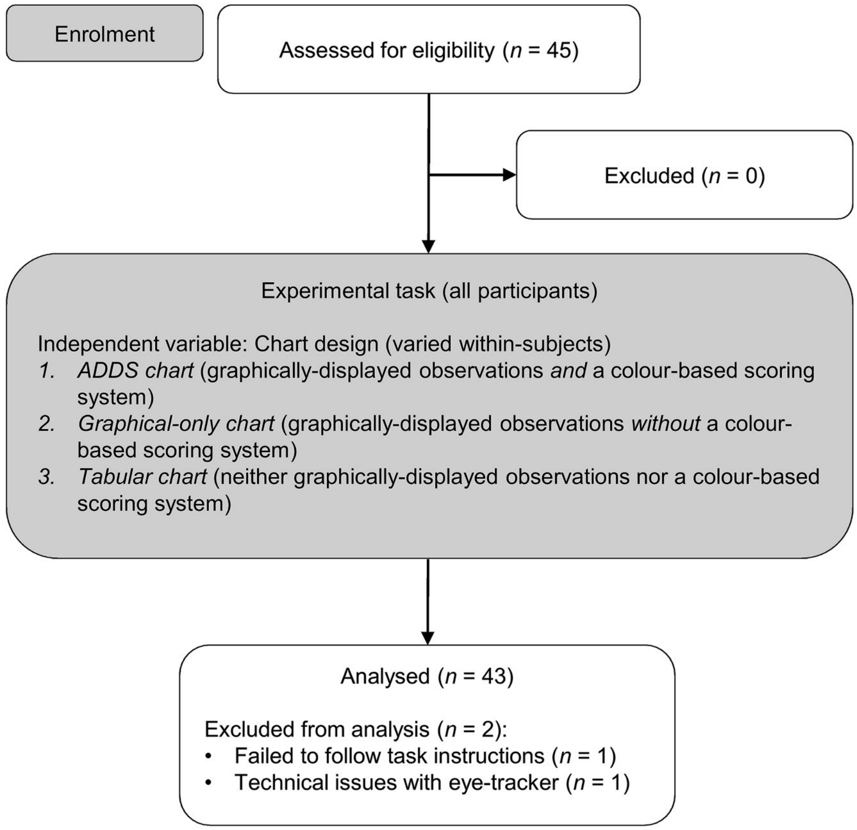


Fig. 3. Flow diagram illustrating the enrolment of participants and analysis of their data.

with alpha set at 0.05 and η^2 as the effect size measure (Howell, 1997). For the response time, fixation count and fixation duration analyses, the Greenhouse-Geisser correction was applied because Mauchly's Test indicated violation of the sphericity assumption. Significant main effects of chart design were followed up using paired-samples t-tests with alpha set at 0.017 (applying the Bonferroni correction) and Cohen's d as the effect size measure (Cohen, 1992).

3. Results

3.1. Response time

Analysis of the response-time data revealed a significant main effect of chart design, $F(1.45, 61.06) = 283.43$, $p < 0.0001$, $\eta^2 = 0.87$ (Fig. 4a). The ADDS produced responses 8.34s faster (CI 6.68–10.00) than the graphical-only chart, $t(42) = -10.17$, $p < 0.0001$, Cohen's $d = -1.68$, which itself yielded responses 21.05s faster (CI 18.32–23.78) than the tabular chart, $t(42) = -15.55$, $p < 0.0001$, Cohen's $d = -2.63$. Responses were 29.39s (CI 26.30–32.49) faster for the ADDS versus the tabular chart, $t(42) = -19.17$, $p < 0.0001$, Cohen's $d = -3.47$.

3.2. Error rate

There was also a significant main effect of chart design on error rate, $F(2, 84) = 70.44$, $p < 0.0001$, $\eta^2 = 0.63$ (Fig. 4b). Participants made 5.09% fewer errors (CI 1.10–9.07) on the ADDS than the graphical-only chart, $t(42) = -2.58$, $p = 0.014$, Cohen's $d = -0.39$, and 18.60% fewer errors (CI 13.85–23.36) on the graphical-only chart versus the tabular chart, $t(42) = -7.89$, $p < 0.0001$, Cohen's $d = -1.20$. The ADDS yielded 23.60% fewer errors (CI 19.76–27.62) than the tabular chart, $t(42) = -12.17$, $p < 0.0001$, Cohen's $d = -1.92$.

3.3. Fixation count

Analysis of the fixation-count data revealed a significant main effect of chart design, $F(1.71, 71.96) = 315.63$, $p < 0.0001$, $\eta^2 = 0.88$ (Fig. 4c). Participants made 14.96 fewer fixations (CI 12.12–17.80) on the ADDS versus the graphical chart, $t(42) = -10.62$, $p < 0.0001$, Cohen's $d = -1.63$, and 30.23 fewer fixations (CI 26.15–34.30) on the graphical-only chart than the tabular chart, $t(42) = -14.97$, $p < 0.0001$, Cohen's $d = -2.38$. The ADDS yielded 45.19 fewer fixations (CI 41.15–49.23) than the tabular chart, $t(42) = -22.56$, $p < 0.0001$, Cohen's $d = -3.70$.

3.4. Fixation duration

There was also a significant main effect of chart design on fixation duration, $F(1.15, 48.29) = 65.33$, $p < 0.0001$, $\eta^2 = 0.61$ (Fig. 4d). The average fixation duration was 66.78 ms shorter (CI 53.28–80.28) for the ADDS versus the graphical-only chart, $t(42) = -9.98$, $p < 0.0001$, Cohen's $d = -1.57$, and 132.57 ms shorter (CI 90.46–174.67) for the graphical-only chart than the tabular chart, $t(42) = -6.35$, $p < 0.0001$, Cohen's $d = -1.13$. Fixation durations were 199.35 ms shorter for the ADDS (CI 155.80–242.90) compared to the tabular chart, $t(42) = -9.24$, $p < 0.0001$, Cohen's $d = -1.66$.

4. Discussion

This study was the first to use eye-tracking to empirically investigate cognitive mechanisms underlying the established effects of observation chart design on the detection of patient deterioration. Participants were tasked with identifying the presence or absence of abnormal observations on three different chart designs while an eye-tracker recorded their gaze. Consistent with predictions and prior research findings (e.g. Preece et al., 2012b, Christofidis et al., 2013,

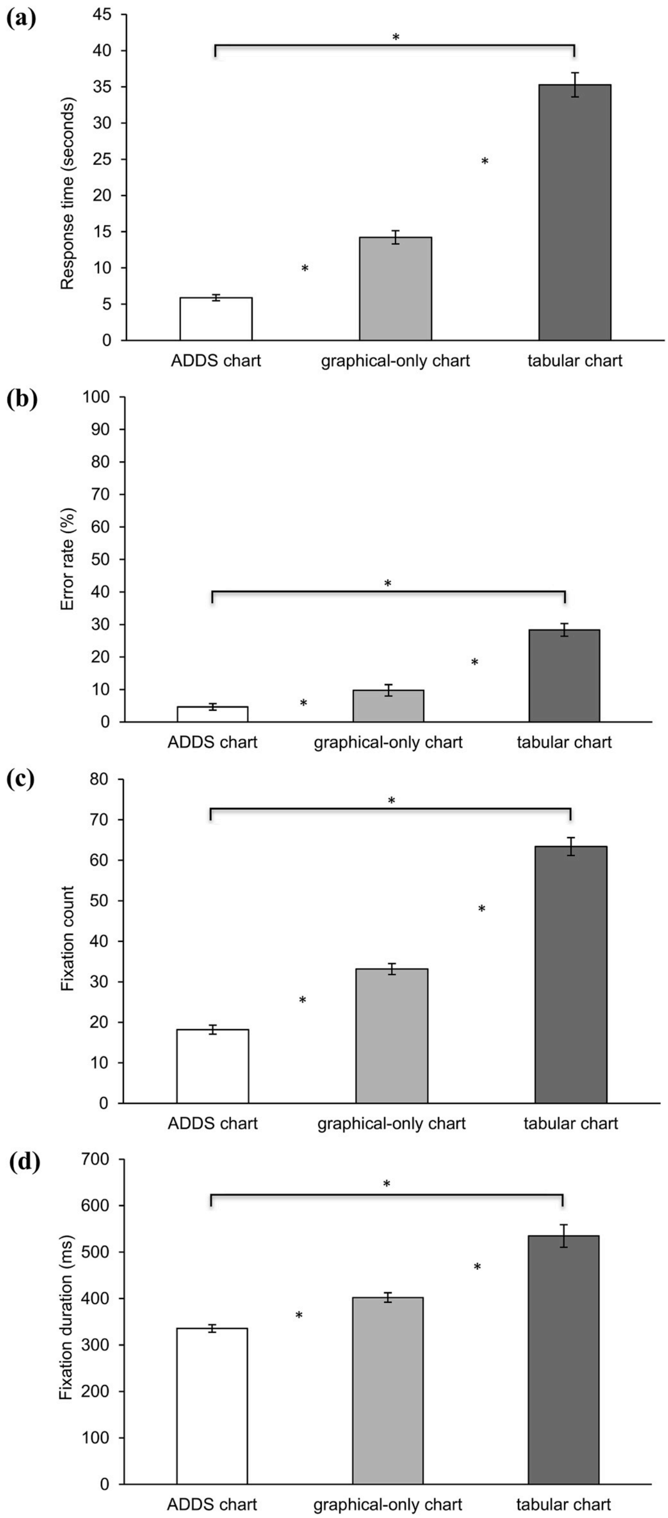


Fig. 4. Mean response time (a), error rate (b), fixation count (c), and fixation duration (d) for interpreting patient data on three different observation chart designs: the 'ADDS chart' (graphically displayed observations *and* a colour-based scoring-system); the 'graphical-only chart' (graphically displayed observations *without* a colour-based scoring-system); and the 'tabular chart' (neither graphically displayed observations nor a colour-based scoring-system). Error bars indicate standard errors of the mean, and asterisks indicate significant differences between charts after applying the Bonferroni correction ($p < 0.017$).

2016), participants were fastest (Hypothesis 1) and made fewer errors (Hypothesis 2) when using a chart that complied with two key usability design recommendations (Preece et al., 2013) – displaying observations graphically, *and* incorporating a colour-based scoring-system (i.e. the ADDS chart). When the colour-coding was unavailable (i.e. when using the *graphical-only chart*), participants' response times and error rates more than doubled. With neither colour-coding nor graphically displayed observations to assist them (i.e. when using the *tabular chart*), participants took six times as long and made over six times as many errors as they did when using the ADDS chart.

Our predictions about *how* these design features improve user-performance were also supported. The eye-tracking findings were consistent with the contention that both graphically displayed observations *and* colour-based scoring-systems can improve search efficiency *and* reduce cognitive workload (Preece et al., 2013). Specifically, the ADDS chart yielded the lowest fixation counts (Hypothesis 3) and the shortest average fixation durations (Hypothesis 4), followed by the *graphical-only chart*, while the *tabular chart* again performed significantly worse than the other designs—a pattern neatly mirroring the error-rate and response-time findings.

The *fixation count* findings suggest that both of the design features manipulated in the present study improved search efficiency by increasing the visual saliency of abnormal observations, directing the gaze towards them. This inference is supported by the findings of basic visual-search studies, which have shown that more visually salient targets reduce fixation count (Williams, 1967; Williams and Reingold, 2001). In addition, the *fixation duration* findings suggest reduced cognitive workload due to enhanced discriminability; that is, fixated information was processed more quickly and easily when presented graphically, and even more so when it was also colour-coded. In particular, the relatively short fixations associated with the ADDS chart likely reflect the fact that its colour-based scoring-system directly signals abnormality, removing the need for chart-users to make mental comparisons with normal ranges held in memory (Preece et al., 2013). While graphical representation alone may also provide visual cues suggestive of deterioration, thus improving processing speed relative to tabulated data, the observation must still be checked against the relevant cut-off level to confirm a criterion breach.

The importance of reducing the cognitive workload associated with observation charts extends beyond improving the speed and accuracy with which deteriorating patients are identified. In the hospital setting, where distractions and multi-tasking are the norm, chart-users are likely to be operating under greater cognitive workload than in a simple laboratory task. Therefore, freeing up finite cognitive resources for higher-order tasks such as clinical reasoning, diagnosis, and prescribing is potentially as critical to patient outcomes as facilitating the detection of deterioration.

Taken together, the findings of the current study not only provide further validation evidence for two key observation chart design rules proposed by the developers of the ADDS chart, but also for the underlying mechanisms that they postulated in formulating their recommendations (Horswill et al., 2010a; Preece et al., 2013). These rules, and the rationales behind them, drew heavily on existing cognitive engineering principles (Gerhardt-Powals, 1996) and usability heuristics (Nielsen, 1994; Zhu et al., 2005) from other domains, and address fundamental limitations on human information processing and their implications for users of a system (such as a website or a clinical chart). Other studies have already shown that, in terms of user-performance, this human factors approach to observation chart design trumps clinicians' design preferences and prior chart experience (Preece et al., 2012a, 2012b, Christofidis et al., 2013, 2014). However, the present study provides the most direct evidence to date that effective observation chart designs are those that cater to ubiquitous human limitations; in this case, visual search and cognitive processing capacities.

4.1. Limitations

The primary limitation of this study is that, as a carefully controlled laboratory-based experiment with novice participants, we cannot be certain to what extent the results would generalise to health professionals in clinical environments. However, there are several important points to consider. First, real-world ADDS chart implementations have led to improved clinical outcomes (Drower et al., 2013; Joshi et al., 2015). Second, it is possible that the distractions, time-pressures and multitasking inherent in clinical environments may actually amplify performance differences between charts by demanding additional cognitive resources from a finite supply (Preece et al., 2012b). This proposal is consistent with the results of an analogous study in which business students interpreted sales data presented in graphs of varied design: Placing participants' working memory under additional strain worsened task accuracy to a greater extent when they were interpreting monochrome graphs rather than graphs that incorporated coloured decision-making aids (Lohse, 1997). Similarly, a basic visual search study by Brouwer et al. (2017) found that increasing participants' cognitive load through multitasking decreased accurate target identification, with a trend towards longer fixation durations. Third, it remains important to study the novice, because inexperienced doctors and nurses are often the first to make decisions about deteriorating patients (Endacott et al., 2010), and every chart-user begins as a novice. Fourth, in previous user-performance experiments, novices and health professionals have repeatedly demonstrated comparable patterns of results across charts, irrespective of prior chart experience (Preece et al., 2012b; Christofidis et al., 2013, 2014, 2016). Furthermore, including participants of varied experience may provide less insight and practical significance than might otherwise be anticipated: first, because nurses are the health professionals overwhelmingly responsible for the recording and initial interpretation of patients' vital signs (Odell et al., 2009); and second, because past research using a similar paradigm to the present study found that doctors and nurses did not differ significantly from one another in the speed and accuracy with which they detected abnormal observations across a range of charts (Preece et al., 2012b). Finally, as discussed above, our eye-tracking metrics address fundamental and ubiquitous aspects of human information processing; hence, the findings are likely to be broadly generalizable.

It could be argued that another limitation of this study was our decision to present the observation chart stimuli on a digital display, rather than on paper. Although this was necessary to maximise the precision of the eye-tracking and performance data, we acknowledge that it may not have been entirely comparable to presenting the charts in their original paper-based form. For example, the digital presentation of stimuli may have differed from a typical paper-based presentation across a range of attributes, such as angle, lighting, and brightness. Nevertheless, in comparable past studies (none of which included eye-tracking), variants of the ADDS chart also yielded fewer errors and shorter response times compared with designs that were less compliant with Preece et al.'s (2013) observation chart usability heuristics, regardless of whether the stimuli were presented digitally (Christofidis et al., 2014, 2016) or on paper (Christofidis et al., 2013; Preece et al., 2012b). It should also be noted that one of the explicit design considerations applied in selecting the colours used for shading on the ADDS chart was that pastels were preferred to maximise legibility under a wide range of lighting conditions (Preece et al., 2013). Furthermore, a study conducted in a simulated hospital ward confirmed that paper ADDS charts could be used successfully by clinicians under low light (Horswill et al., 2010c).

We also acknowledge that the tabular chart design used in this experiment is somewhat removed from those used in real clinical environments, particularly in its lack of space-efficiency (which is arguably the main benefit of a basic tabular chart). For the purposes of this experiment, we reasoned that it was more important to avoid the spatial confound that would have arisen from compressing the chart. Nevertheless, this is unlikely to have affected the fixation counts (because participants had no reason to fixate on the blank space between rows) or fixation durations (because these are independent of saccades).

Finally, a common criticism of eye-tracking is that gaze does not necessarily indicate attention. That is, a person may look at a chart without processing or consciously registering the fixated information. However, the clear and meaningful way in which the eye-tracking findings mapped onto the performance measures suggests otherwise in this case. That is, greater accuracy and quicker response times were clearly linked to lower fixation counts (suggesting more efficient search) and shorter fixation durations (suggesting reduced cognitive workload). While it is possible to attend to visual stimuli without fixating them, it is more common and effective to fixate the objects of one's attention (Eckstein et al., 2017). This connection between gaze and attention is particularly apparent in clearly defined tasks such as visual search (Just and Carpenter, 1976), a link that has been further validated by fMRI and cortical stimulation research (Eckstein et al., 2017).

4.2. Conclusion

It has been established that failure to recognise deterioration, and subsequent delays in intervention, can jeopardise patient safety (Shearer et al., 2012; De Meester et al., 2012; Boniatti et al., 2013). The present study adds to a growing body of evidence that well-designed observation charts can considerably improve detection (Preece et al., 2012b; Christofidis et al., 2013, 2014, 2016), potentially improving patient outcomes by supporting early intervention. This evidence base lends weight to the importance of a human factors approach to the design of clinical charts. Specifically, our eye-tracking findings support the mechanisms postulated by the designers of the ADDS chart underlying two of its key design features (Preece et al., 2013): Both graphically displayed observations and a colour-based scoring-system improved search efficiency *and* reduced the cognitive resources required to process vital sign data. These findings confirm that effective observation chart designs are those which cater to ubiquitous limitations on human information processing.

Although hospitals are increasingly switching to electronic records, the design of paper-based observation charts remains important for two main reasons. First, in developing countries, they are likely to remain the primary tool for storing and interpreting vital sign data for many years to come. Second, it is anticipated that they will always be required as a back-up for electronic systems. The recent *WannaCry* ransomware attack, which affected a large number of National Health System hospitals in the UK, is a timely reminder of the potential vulnerability of electronic records (Collier, 2017). Similarly, there are several reported instances of IT network failures necessitating staff in USA and UK hospitals to revert from electronic to paper-based systems for periods of days (Berinato, 2003; Flinders, 2015). From short interruptions due to routine systems maintenance, through to the rarest and most catastrophic events – such as the prolonged IT and communication systems failures that could result from a solar superstorm (Jonas, 2015) – hospitals cannot rely on electronic systems entirely. However, in the future, even the most experienced clinicians working in digital hospitals will essentially be chart-novices in relation to paper-based clinical forms. Hence, it will become more – not less – important that these charts employ user-friendly designs that do not require extensive training or experience to use effectively.

Funding

This research did not receive any specific grant from funding agencies in the public, commerical, or not-for-profit sectors. SIB was funded by an ARC Future Fellowship [FT130101282].

Conflicts of interest

None.

Acknowledgements

The authors would like to thank Dr Melany J. Christofidis, for her helpful advice and assistance, and the participants who volunteered for this study.

References

ACT Health [Internet], 2011. Compass. Early Recognition of Deteriorating Patient. Available from. http://health.act.gov.au/professionals/general-information/compass/.

Ahmed, A., Chandra, S., Herasevish, V., Gajic, Ognjen, Pickering, B., 2011. The effect of two different electronic health record user interfaces on intensive care provider task load, errors of cognition, and performance. Crit. Care Med. 39 (7), 1626–1634. https://doi.org/10.1097/CCM.0b013e31821858a0.

Baker, J., Jones, D., Burkman, J., 2009. Using visual representations of data to enhance sensemaking in data exploration tasks. J. Assoc. Inf. Syst. Online 10 (7), 553–559.

Barkana, D.E., Açık, A., 2014. Improvement of design of a surgical interface using an eye-tracking device. Theor. Biol. Med. Model. 11, 1–18. https://doi.org/10.1186/1742-4682-11-S1-S4.

Bergstrom, J.C., Olmsted-Hawala, E.L., Jans, M.E., 2013. Age-related differences in eye-tracking and usability performance: website usability for older adults. Int. J. Hum. Comput. Interact. 29 (8), 541–548. https://doi.org/10.1080/10447318.2012.728493.

Berinato, S., 2003. All Systems Down. Retrieved. http://www.computerworld.com/article/2581420/disaster-recovery/all-systems-down.html, Accessed date: 28 August 2017.

Bojko, A., Gaddy, C., Lew, G., Quinn, A., Israelski, E., 2005. Evaluation of drug label designs using eye-tracking. In: Proceedings of the Human Factors and Ergonomics Society, 49th Annual Meeting, vol. 49. pp. 1033–1037. https://doi.org/10.1177/154193120504901107. (11).

Boniatti, M.M., Azzolini, N., Viana, M.V., Ribeiro, B.S., Coelho, R.S., Castilho, R.K., et al., 2013. Delayed medical emergency team calls and associated outcomes. Crit. Care Med. 42, 26–30. https://doi.org/10.1097/CCM.0b013e31829e53b9.

Brouwer, A., Hogervost, M.A., Oudejans, B., Ries, A.J., Touryan, J., 2017. EEG and eye tracking signatures of target encoding during structured visual search. Front. Hum. Neurosci. 11, 1–11. https://doi.org/10.3389/fnhum.2017.00264.

Brown, P.J., Marquard, J.L., Amster, B., Romoser, M., Friderici, J., Goff, S., Fisher, D., 2014. What do physicians read (and ignore) in electronic progress notes? Appl. Clin. Inf. 5, 430–444. https://doi.org/10.4338/ACI-2014-01-RA-0003.

Cahill, H., Jones, A., Herkes, R., Cook, K., Stirling, A., Halbert, T., Yates, A., Lal, S., Gardo, A., Donnelly, R., Gattas, D.J., 2011. Introduction of a new observation chart and education programme is associated with higher rates of vital-sign ascertainment in hospital wards. BMJ Qual. Saf. 20, 932–940.

Chen, S., Epps, J., 2014. Using task-induced pupil diameter and blink rate to infer cognitive load. Hum. Comput. Interact. 29 (4), 390–413. https://doi.org/10.1080/07370024.2014.892428.

Chatterjee, M.T., Moon, J.C., Murphy, R., McCrea, D., 2005. The "OBS" chart: an evidence based approach to re-design of the patient observation chart in a district general hospital setting. Postgrad. Med. 81 (960), 663–666. https://doi.org/10.1136/pgmj.2004.031872.

Christofidis, M.J., Hill, A., Horswill, M.S., Watson, M.O., 2013. A human factors approach to observation design can trump health professionals' prior chart experience. Resuscitation 84, 657–665. https://doi.org/10.1016/j.resuscitation.2012.09.023.

Christofidis, M.J., Hill, A., Horswill, M.S., Watson, M.O., 2014. Observation charts with overlapping blood pressure and heart rate graphs do not yield the performance advantage that health professionals assume: an experimental study. J. Adv. Nurs. 70 (3), 610–624. https://doi.org/10.1111/jan.12223.

Christofidis, M.J., Hill, A., Horswill, M.S., Watson, M.O., 2015. Less is more: the design of early-warning scoring systems affects the speed and accuracy of scoring. J. Adv. Nurs. 71 (7), 1573–1586. https://doi.org/10.1111/jan.12618.

Christofidis, M.J., Hill, A., Horswill, M.S., Watson, M.O., 2016. Observation chart design features affect the detection of patient deterioration: a systematic experimental evaluation. J. Adv. Nurs. 72 (1), 158–172.

Cohen, J., 1992. A power primer. Psychol. Bull. 112, 155–159. https://doi.org/10.1037/0033-2909.112.1.155.

Collier, R., 2017. NHS ransomware attack spreads worldwide. Can. Med. Assoc. J. 189 (22), E786–E787.

De Meester, K., Bogaert, P.V., Clarke, S.P., Bossaert, L., 2012. In-hospital mortality after serious adverse events on medical and surgical nursing units: a mixed methods study. J. Clin. Nurs. 22, 2308–2317. https://doi.org/10.1111/j.1365-2702.2012.04154.x.

Drieghe, D., Rayner, K., Pollatsek, A., 2005. Eye movements and word skipping during reading revisited. J. Exp. Psychol. 31 (5), 954–969. https://doi.org/10.1037/0096-1523.31.5.954.

Drower, D., McKeany, R., Jogia, P., Jull, A., 2013. Evaluating the impact of implementing an early warning score system on incidence of in-hospital cardiac arrest. N. Z. Med. J. 126, 26–34.

Duchowski, A., 2007. Eye Tracking Methodology: Theory and Practice. Springer, London, UK.

Eckstein, M., Guerra-Carrillo, B., Miller Singley, A.T., Bunge, S.A., 2017. Beyond eye gaze: what else can eyetracking reveal about cognition and cognitive development? Developmental Cognitive Neuroscience 25, 69–91. https://doi.org/10.1016/j.dcn.2016.11.001 .

Elliott, D., Allen, E., McKinley, S., Perry, L., Duffield, C., Fry, M., Gallagher, R., Iedema, R., Roche, M., 2016. User acceptance of observation and response charts with a track and trigger system: a multisite staff survey. J. Clin. Nurs. 25 (15–16), 2211–2222.

Endacott, R., Kidd, T., Chaboyer, W., Edington, J., 2007. Recognition and communication of patient deterioration in a regional hospital: a multi-methods study. Aust. Crit. Care 20, 100–105.

Endacott, R., Scholes, J., Buykx, P., Cooper, S., Kinsman, L., McConnell-Henry, T., 2010. Final-year nursing students' ability to assess, detect and act on clinical cues of deterioration in a simulated environment. J. Adv. Nurs. 66 (12), 2722–2731.

Faul, F., Erdfelder, E., Lang, A.-G., Buchner, A., 2007. G*Power 3: a flexible statistical power analysis program for the social, behavioral and biomedical sciences. Behav. Res. Methods 39, 175–191.

Flinders, K., 2015. Network Failure Crashed Frontline Services at London Hospital. Retrieved. http://www.computerweekly.com/news/4500247512/Network-failure-crashed-frontline-services-at-London-hospital, Accessed date: 28 August 2017.

Forsman, J., Anani, N., Eghdam, A., Falkenhav, M., Koch, S., 2013. Integrated information visualization to support decision making for use of antibiotics in intensive care: design and usability evaluation. Inf. Health Soc. Care 38 (4), 330–353. https://doi.org/10.3109/17538157.2013.812649.

Franklin, C., Mathew, J., 1994. Developing strategies to prevent inhospital cardiac arrest: analyzing responses of physicians and nurses in the hours before the event. Crit. Care Med. 22 (2), 244–247. https://doi.org/10.1097/00003246-199402000-00014.

Fung, K., Khan, F., Dawson, J., 2014. The introduction of an integrated early warning score observation chart – a picture paints a thousand words. J. Patient Saf. 10 (1), 13–19. https://doi.org/10.3758/s13423-013-0547-4.

Gerhardt-Powals, J., 1996. Cognitive engineering principles for enhancing human-computer performance. Int. J. Hum. Comput. Interact. 8, 189–211.

Godwin, H.J., Hout, M.C., Menneer, T., 2014. Visual similarity is stronger than semantic similarity in guiding visual search for numbers. Psychonomic Bull. Rev. 21 (3), 689–695. https://doi.org/10.3758/s13423-013-0547-4.

Goldberg, H.J., Kotval, X.P., 1999. Computer interface evaluation using eye movements: methods and constructs. Int. J. Ind. Ergon. 24 (6), 631–645. https://doi.org/10.1016/S0169-8141(98)00068-7.

Goldberg, H.J., Wichansky, A.M., 2003. Eye tracking in usability evaluation: a practitioner's guide. In: Hyönä, J., Radach, R., Deubel, H. (Eds.), The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research. Elsevier, Amsterdam, pp. 493–516.

Goldhill, D.R., White, S.A., Sumner, A., 1999. Physiological values and procedures in the 24 h before ICU admission from the ward. Anaesthesia 54 (6), 529–534. https://doi.org/10.1046/j.1365-2044.1999.00837.x.

Gould, J.D., 1973. Eye movements during visual search and memory search. J. Exp. Psychol. 98 (1), 184–195. https://doi.org/10.1037/h0034280 .

Hanson, E., 2004. Focus of attention and pilot error. In: Proceedings of the Eye Tracking Research and Applications Symposium 2000. ACM Press, New York, NY.

Hermens, F., Flin, R., Ahmed, I., 2013. Eye movements in surgery: a literature review. Journal of Eye Movement Research 6 (4), 1–11. https://doi.org/10.16910/jemr.6.4.4.

Hillman, K.M., Bristow, P.J., Chey, T., Daffurn, K., Jacques, T., Norman, S.L., Bishop, G.F., Simmons, G., 2001. Antecedents to hospital deaths. Intern. Med. J. 31, 343–348. https://doi.org/10.1046/j.1445-5994.2001.00077.

Hillman, K.M., Bristow, P.J., Chey, T., Daffurn, K., Jacques, T., Norman, S.L., Bishop, B.F., Simmons, G., 2002. Duration of life-threatening antecedents prior to intensive care admission. Intensive Care Med. 28, 1629–1634. https://doi.org/10.1007/s00134-002-1496-y.

Horswill, M.S., Preece, M.H.W., Hill, A., Christofidis, M.J., Karamatic, R., Hewett, D., Watson, M.O., 2010a. Human Factors Research Regarding Observation Charts: Research Project Overview. Australian Commission on Safety and Quality in Health Care, Sydney, Australia.

Horswill, M.S., Preece, M.H.W., Hill, A., Christofidis, M.J., Watons, M., 2010c. Recording Patient Data on Six Observation Charts: an Experimental Comparison. Australian Commission on Safety and Quality in Health Care, Sydney, Australia.

Horswill, M.S., Preece, M.H.W., Hill, A., Watson, M.O., 2010b. Detecting Abnormal Vital Signs on Six Observation Charts: an Experimental Comparison. Australian Commission on Safety and Quality in Health Care, Sydney, Australia.

Howell, D.C., 1997. Statistical Methods for Psychology. Duxbury Press, Belmont, CA.

Jonas, F.M., 2015. Solar flares. In: Pelton, J.N., Allahdadi, F. (Eds.), Handbook of Cosmic Hazards and Planetary Defense. Springer International Publishing, Cham, Switzerland, pp. 243–257.

Joshi, K.S., Campbell, V., Landy, M., Anstey, C.M., Gooch, R., 2015. The effect of Rapid Response System revision on standard and specific intensive care unit outcomes in a regional hospital. Anaesth. Intensive Care 45 (3), 369–374.

Just, M.A., Carpenter, P.A., 1976. Eye fixations and cognitive processes. Cognit. Psychol. 8 (4), 441–480. https://doi.org/10.1016/0010-0285(76)90015-3.

Laxmisan, A., Hakimzada, F., Sayan, O.R., Green, R.A., Zhang, J., Patel, V.L., 2007. The multitasking clinician: decision-making and cognitive demand during and after team handoffs in emergency care. Int. J. Med. Inf. 76 (11–12), 801–811. https://doi.org/10.1016/j.ijmedinf.2006.09.019.

Lohse, G.L., 1997. The role of working memory on graphical information processing. Behav. Inf. Technol. 16 (6), 297–308. https://doi.org/10.1080/014492997119707.

Luria, S.M., Strauss, M.S., 1975. Eye movements during search for coded and uncoded targets. Percept. Psychophys. 17 (3), 303–308. https://doi.org/10.3758/BF03203215.

Martins, R., Carvalho, J.M., 2015. Eye blinking as an indicator of fatigue and mental load – a systematic review. In: Conference Paper. Occupational Safety and Hygiene III, https://doi.org/10.1201/b18042-48.

Mitchell, I.A., McKay, H., Van Leuvan, C., Berry, R., McCutcheon, C., Avard, B., Slater, N., Neeman, T., Lamberth, P., 2010. A prospective controlled trial of the effect of a multi-faceted intervention on early recognition and intervention in deteriorating hospital patients. Resuscitation 81, 658–666. https://doi.org/10.1016/j.resuscitation.2010.03.001.

Moffitt, K.W., 1978. Eye Movements as a Correlate of Information Processing in a Visual Search Task (Report No. 78-3). Las Cruces. New Mexico State University, NM.

Nielsen, J., 1994. Heuristic evaluation. In: Nielsen, J., Mack, R.L. (Eds.), Usability Inspection Methods. John Wiley & Sons, New York, pp. 25–62.

Nightingale, F., 1860. Notes on Nursing: What It Is, and What It Is Not. D. Appleton & Company, New York.

Odell, M., Victor, C., Oliver, D., 2009. Nurses' role in detecting deterioration in ward patients: systematic review. J. Adv. Nurs. 65 (10), 1992–2006.

Poole, A., Ball, L.A., 2005. Eye tracking in human-computer interaction and usability research: current status and future prospects. In: Ghaoui, C. (Ed.), Encyclopedia of Human-computer Interaction. Idea Group Inc, Pennsylannia, pp. 211–219.

Preece, M.H.W., Hill, A., Horswill, M.S., Karamatic, R., Hewett, D.G., Watson, M.O., 2013. Applying heuristic evaluation to observation chart design to improve the detection of patient deterioration. Appl. Ergon. 44 (4), 544–556. https://doi.org/10.1016/j.apergo.2012.11.003.

Preece, M.H.W., Hill, A., Horswill, M.S., Karamatic, R., Watson, M.O., 2012a. Designing observation charts to optimize the detection of patient deterioration: reliance on the subjective preferences of healthcare professionals is not enough. Aust. Crit. Care 25, 238–252.

Preece, M.H.W., Hill, A., Horswill, M.S., Watson, M.O., 2012b. Supporting the detection of patient deterioration: observation chart design affects the recognition of abnormal vital signs. Resuscitation 83, 1111–1118. https://doi.org/10.1016/j.resuscitation.2012.02.009.

Qiuzhen, W., Yang, S., Liu, M., Cao, Z., Ma, Q., 2014. An eye-tracking study of website complexity from cognitive load perspective. Decis. Support Syst. 62, 1–10. https://doi.org/10.1016/j.dss.2014.02.007.

Salthouse, T.A., Ellis, C.L., Diener, D.C., Somberg, B.L., 1981. Stimulus processing during eye fixations. J. Exp. Psychol. 7 (3), 611–623. https://doi.org/10.1037/0096-1523.7.3.611.

Shearer, B., Marshall, S., Buist, M.D., Finnigan, M., Kitto, S., Hore, T., Sturgess, T., Wilson, S., Ramsay, W., 2012. What stops hospital clinical staff from following protocols? An analysis of the incidence and factors behind the failure of bedside clinical staff to activate the rapid response system in a multi-campus Australian metropolitan healthcare service. BMJ Qual. Saf. 21 (7), 569–575. https://doi.org/10.1136/bmjqs-2011-000692.

Shen, J., Elahipanah, A., Reingold, E.M., 2007. Effects of context and instruction on the guidance of eye movements during a conjunctive visual search task. In: van Gompel, R.P., Fischer, M.H., Murray, W.S., Hill, R.L. (Eds.), Eye Movements: a Window on Mind and Brain, pp. 598–615. https://doi.org/10.1016/B978-008044980-7/50030-6.

Tien, T., Pucher, P.H., Sodergren, M.H., Sriskandaraja, K., Yang, G., Darzi, A., 2014. Eye tracking for skills assessment and training: a systematic review. J. Surg. Res. 191 (1), 169–178. https://doi.org/10.1016/j.jss.2014.04.032.

Van der Gijp, A., Ravesloot, C.J., Jarodzka, H., van der Schaaf, M.F., van der Schaaf, I.C., van Schaik, J.P.J., Ten Cate, T.,J., 2017. How visual search relates to visual diagnostic performance: a narrative systematic review of eye tracking research in radiology. Adv. Health Sci. Educ. 22 (3), 765–787. https://doi.org/10.1007/s10459-016-9698-1.

Van Diepen, P.M., De Graef, P., d'Ydewalle, G., 1995. Chronometry of foveal information extraction during scene perception. In: Findlay, J.M., Walker, R., Kentridge, R.W. (Eds.), Eye Movement Research: Mechanisms, Processes, and Applications. Elsevier, Amsterdam, The Netherlands, pp. 349–362.

Williams, L.G., 1967. The effects of target specification on objects fixated during visual search. Acta Psychol. 27, 355–360. https://doi.org/10.1016/0001-6918(67)90080-7.

Williams, D.E., Reingold, E.M., 2001. Preattentive guidance of eye movements during triple conjunction search tasks: the effects of feature discriminability and saccadic amplitude. Psychonomic Bull. Rev. 8 (3), 476–488. https://doi.org/10.3758/bf03196182.

Zagermann, J., Pfeil, U., Reiterer, H., 2016. Measuring cognitive load using eye tracking technology in visual computing. In: Lam, Heidi (Ed.), Proceedings of the Sixth Workshop on beyond Time and Errors on Novel Evaluation Methods for Visualization (BELIV '16), Michael Sedlmair, Petra Isenberg, Tobias Isenberg, Narges Mahyar. ACM, New York, NY, USA, pp. 78–85. https://doi.org/10.1145/2993901.2993908.

Zelinsky, G., Sheinberg, D., 1997. Eye movements during parallel-serial visual search. J. Exp. Psychol. Hum. Percept. Perform. 23 (1), 244–262. https://doi.org/10.1037/0096-1523.23.1.244.

Zhu, W., Vu, K.-P.L., Proctor, R.W., 2005. Evaluating web usability. In: Proctor, R.W., Vu, K.-P.L. (Eds.), Handbook of Human Factors in Web Design. Lawrence Erlbaum Associates, Manwah, NJ, pp. 321–337.