

Evaluating the Effectiveness of Clinical Alerts: A Signal Detection Approach

Mei-Sing Ong, MBIomedE¹, Enrico Coiera, MB BS, PhD¹

¹Centre for Health Informatics, Australian Institute of Health Innovation, University of New South Wales, Sydney, Australia

Abstract

Clinical alerts are widely used in healthcare to notify caregivers of critical information. Alerts can be presented through many different modalities, including verbal, paper and electronic. Increasingly, information technology is being used to automate alerts. Most applications, however, fall short in achieving the desired outcome. The objective of this study is twofold. First, we examine the effectiveness of verbal and written alerts in promoting adherence to infection control precautions during inpatient transfers to radiology. Second, we propose a quantitative framework based on Signal Detection Theory (SDT) for evaluating the effectiveness of clinical alerts. Our analysis shows that verbal alerts are much more effective than written alerts. Further, using precaution alerts as a case study, we demonstrate the application of SDT to evaluate the quality of alerts, and human behavior in handling alerts. We hypothesize that such technique can improve our understanding of computerized alert systems, and guide system redesign.

Introduction

Failure to adequately communicate critical information is a major cause of adverse events in hospitalized patients¹. Dissemination of critical information that requires immediate attention to caregivers, who are already burdened with information overload, can be a challenging task².

In recent years, there is a proliferation of interventions implemented to improve alert notification of critical information using various modalities. One such attempt is the application of colored wristband to alert caregivers of certain conditions. In the UK, the National Patient Safety Agency recommends issuing patients with known allergies with a red allergy alert band, and patients suspected of or diagnosed with a disease or condition requiring infection control precautions with a blue precautions alert band³. A more sophisticated approach to alert notification involves utilizing clinical information systems to automate generation of alerts and reminders. Implementation of computerized alert system has been shown to enhance patient care and prevent adverse events⁴⁻⁵.

The effectiveness of these interventions, however, often falls short of expectations. In a study evaluating the effectiveness of allergy wristband, it was found that only 55.2% of patients (n=186) with allergies had an allergy wristband⁶. This intervention was further undermined by the inappropriate use of red (rather than white) wristbands as identification wristbands in some patients, and the use of white wristbands to display allergies. Inconsistency in the use of color-coding was also reported in another study, where red was used to signal at least 10 different statuses or risks³. Poor reliability of the alerts resulted in low compliance with the alerts.

Implementation of computerized alerts faces similar challenges. In a study assessing the effectiveness of computerized test result notification system, designed to minimize breakdowns in critical communication between radiologists and clinicians, it was found that physicians failed to electronically acknowledge over one-third of alerts and were unaware of abnormal imaging results in 4% of cases 4 weeks after reporting⁷. A literature review on drug safety alerts reported that safety alerts were overridden in 49% to 96% of cases, with the exception of serious alerts for overdose, which were overridden in one fourth of all alerts⁸. A common reason for overriding was alert fatigue caused by poor signal-to-noise ratio – the alert was either not serious, or was irrelevant⁷⁻⁹. Several studies further suggested that resistance among physicians to the perceived intrusion of information technology into their practice may have contributed to the high overriding rate of computerized alerts⁹⁻¹¹.

Thus, designing effective alerts for the use of patient care is not a trivial task. The success of an alert notification system is heavily dependent on two major factors: (1) the reliability and detectability of the alerts; and (2) individual user's response bias. The latter is determined by an individual's motivational states, past experiences and knowledge, attitudes and pathological conditions. The ability to quantitatively distinguish these two factors is critical. If the former is found wanting, then improvement strategy should be targeted at improving the reliability

and detectability of the alerts. If human performance is the primary hindrance, resources should be channeled into behavioral change interventions.

To date, the effectiveness of an alert notification system is primarily evaluated through standard measures of sensitivity, specificity and predictive power. These measures provide an indication of how reliable a system is in discriminating true and false alerts. Whilst useful, they do not distinguish between decisional and perceptual contributions to performance¹². Failure to account for this difference confounds our interpretation of the effectiveness of a given system. The lack of reliable tools and framework for evaluating alert notification system is a major barrier to the successful application of such systems.

Study Objectives

The objective of this study is twofold. First, we evaluate the use of alerts in a common clinical process – inpatient transfers to radiology. Specifically, we examine the effectiveness of verbal and written alerts in prompting radiology porters to adhere to infection control precautions during transfer. Second, to theoretically underpin our analysis, a quantitative framework based on *Signal Detection Theory* (SDT) is applied. The framework allows for the perceptibility of the alerts and the internal response of individual caregivers to be analyzed independently. Using precaution alerts during inpatient transfers as a case study, we demonstrate how SDT can be used to understand human behavior in handling alerts. The same technique can be applied to evaluate electronic alerts. We hypothesize that applying such framework can improve our understanding of the alerting process, and can guide system redesign.

Background

What is Signal Detection Theory (SDT)?

SDT is an adaptation of statistical decision theory¹³. The theory models the performance of discrimination task in the presence of uncertainty. An observer is exposed to two types of stimulus, *signal* and *noise*. The task is to distinguish between the two, and to respond only to the signal. An individual's performance is determined by two independent components: (1) a sensory component that determines an individual's *perceptibility* to a stimulus. This is the individual's actual ability to discern true signals from noise; (2) a decision component that determines how an individual responds to a stimulus when confronted with uncertainty. This is the individual's *response bias* or *criterion*, and it reflects the individual's general tendency to respond.

Figure 1 depicts the graphical notation of these concepts. The graph is a theoretical representation of the response curves when only noise is present (*noise distribution*), and when signal is present (*signal distribution*). The height of the curves represents the probability that a subject will respond to the stimuli, for a range of criteria indicated on the x-axis. *Perceptibility* is represented by the degree of overlap between the signal and noise distributions, and *response bias* is indicated by the vertical line. When this criterion is exceeded, the subject will respond to the stimulus. The point where the noise and signal distributions intersect is the neutral point. When the criterion lies to the left of the neutral point, the subject has a bias toward responding to the stimulus, and when it lies to the right of the neutral point, the subject is biased toward not responding to the stimulus. The distance between the criterion and the neutral point is shown as c , and the distance between the means of the two distributions is shown as d' .

The hit rate equals the proportion of the signal distribution that exceeds the criterion, and the false-alarm rate equals the proportion of the noise distribution that exceeds the criterion (Figure 1b). If the criterion is set to a lower value, a higher hit rate can be achieved. However, a more liberal criterion also results in a higher false-alarm rate, as the likelihood of mistaking noise for signal increases. In contrast, a more conservative individual will have a higher criterion value, resulting in a lower false-alarm rate. This is unfortunately accompanied by a drop in hit rate, as the likelihood of rejecting signals as noise increases (Figure 1c).

Thus, performance in discrimination task is highly dependent on the subject's response bias. For a given criterion, the only way to reduce false-alarm rate is to improve signal perceptibility. This is achieved by reducing the overlap between the signal and noise distributions. The higher the degree of overlap, the less detectable the signal is, and the higher the false-alarm rates. As the degree of overlap decreases, signal becomes more distinguishable from its noise counterpart, and false-alarm rate is reduced without also diminishing the hit rate. Perceptibility can be improved by either making the signal "louder", or by diminishing the noise. Graphically, these effects are represented by

increasing the separation between the signal and noise distributions, and reducing the spread of the noise distribution respectively (Figure 1d).

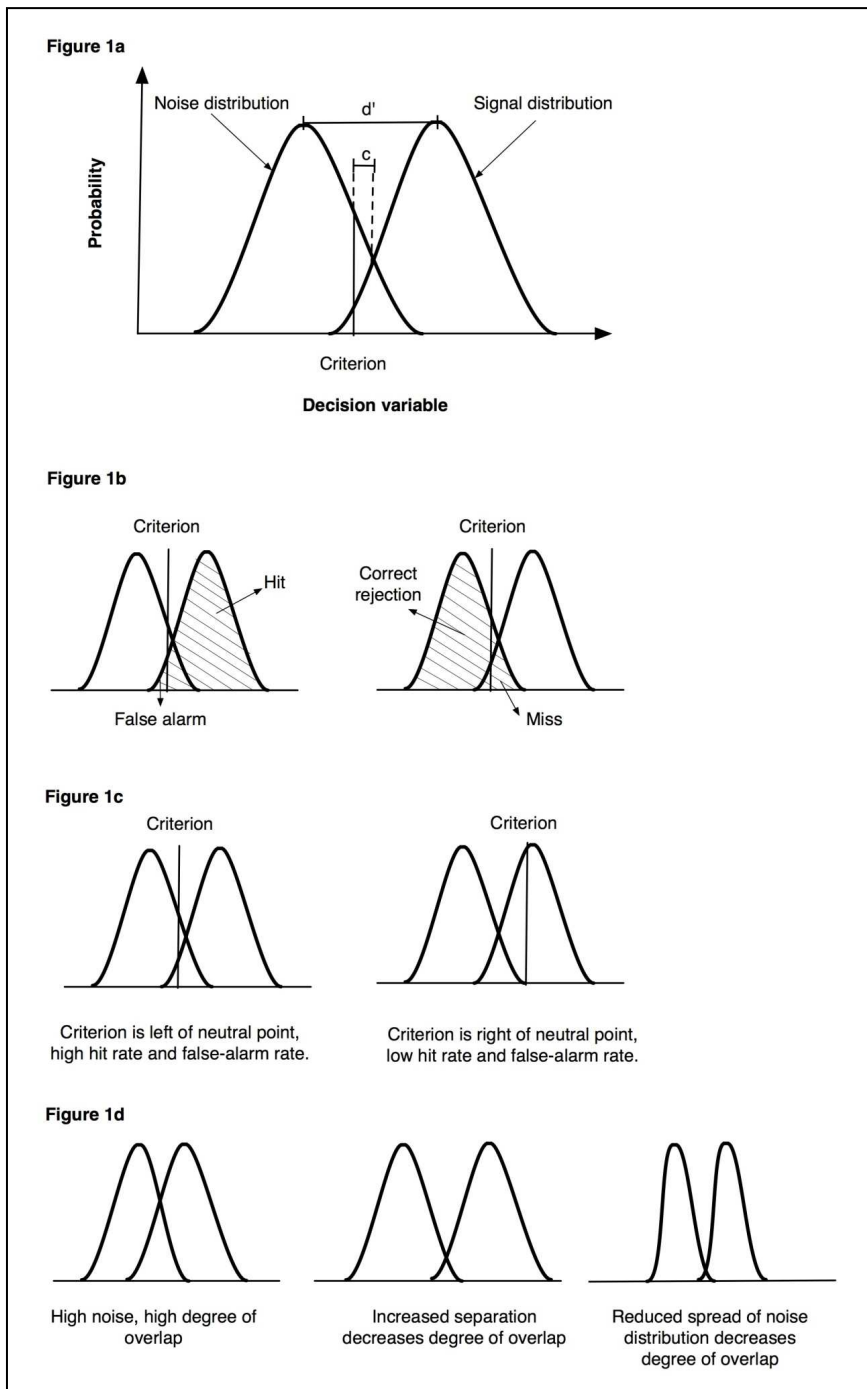


Figure 1. Graphical representation of Signal Detection Theory.

Quantitative Measures of Perceptibility and Response Bias

The effects of sensory and decision components can be estimated independently using parametric or non-parametric statistics. When distribution of data is approximately normal, perceptibility can be quantified with d' , and response bias with β . When distribution of data is unknown, non-parametric measures can be used. Common measures include A' for perceptibility, and c for response bias. Table 1 summarizes these measures.

Table 1. Quantitative measures of perceptibility and response bias¹⁴. In all equations, H represents hit rate, and F the false-alarm rate.

Measure	Description	Mathematical equation
d'	A parametric measure for perceptibility. d' is the distance between the mean of the signal distribution and the mean of the noise distribution. A value of 0 indicates an inability to distinguish signals from noise. A larger value indicates a correspondingly greater ability to distinguish signals from noise.	$d' = \varphi^{-1}(H) - \varphi^{-1}(F)$
β	A parametric measure for response bias, based on the likelihood ratio of a response. When subjects favour neither the <i>yes</i> response nor the <i>no</i> response, β is 1. Values less than 1 signify bias toward responding <i>yes</i> , and values greater than 1 signify a bias toward the <i>no</i> response.	$\beta = e^{\left(\frac{(\varphi^{-1}(F))^2 - (\varphi^{-1}(H))^2}{2}\right)}$
A'	A non-parametric measure for perceptibility. A' typically ranges from 0.5, which indicates that signals cannot be distinguished from noise, to 1, which corresponds to perfect performance. The minimum possible value is 0.	$A' = 0.5 + \text{sign}(H - F) \frac{(H - F)^2 + H - F }{4 \max(H, F) - 4HF}$
c	A non-parametric measure for response bias. c is the distance between the criterion and the neutral point, where neither response is favored. If the criterion is located at this point, c has a value of 0. Negative values of c signify bias toward responding <i>yes</i> , whilst positive values signify a bias toward the <i>no</i> response.	$c = -\frac{\varphi^{-1}(H) + \varphi^{-1}(F)}{2}$

Alert Notification as a Signal Detection Study

In the light of SDT, clinical alerts can be interpreted as signals intended to prompt a caregiver to take the appropriate actions. A *hit* is said to have occurred when the alert is provided, and the correct clinical decision is made as a result. And a *miss* has occurred if the caregiver is alerted, but the appropriate clinical decision is not taken. All decisions made under uncertainty have error rates. At the threshold of uncertainty, an individual can err on the side of making false-positive (risk-averse individual), or false-negative (risk-taking individual) decisions.

Perceptibility can be defined as a caregiver’s ability to correctly discern between relevant alerts and background noise, such as those generated from non-clinical communication and interruptions. And response bias is a caregiver’s personal threshold for rendering positive notification. Thus, the approach isolates the inherent detectability of an alert, a sensory process, from the observer’s decision process.

		Response	
		<i>R</i>	<i>~R</i>
Alert	<i>S</i>	Hit	Miss
	<i>~S</i>	False Alarm	True Negatives

Figure 2. 2x2 contingency table, representing the 4 possible scenarios (*S*=alert is provided, *~S*=alert is not provided, *R*=alert is adhered to, *~R*=alert is not adhered to).

Methods

Data Collection

A prospective observational study was undertaken at a 440-bed metropolitan teaching hospital, with an average occupancy rate of over 90% and over 3000 staff members. Patient transfers from inpatient wards to the radiology were observed over a 6-month period (February to July 2009). The total number of transfers to radiology from the wards over this period was about 9600, a daily average of about 80. Approval to undertake the study was granted by the hospital's ethics committee.

The study was conducted in 2 phases. In an initial pilot, 20 transfers were observed in order to understand the information and process flow of the transfers. All porters working for radiology (*n*=8) were shadowed unobtrusively through the transfer journey by a researcher (MO). Data from this phase guided design of a structured observational instrument, and were not used in further analysis.

In the second phase, the same porters were observed over a convenience sample of 101 transfers, covering transfers from morning to evening, Monday to Friday, using the structured observational tool. Information collected included observed frequency of transfers involving an infectious patient, observed frequency and type of precaution alert notifications, and compliance rate with infection control precautions amongst porters during the transfer process. Inter-rater reliability analysis was performed by a second observer shadowing alongside the first for 12 transfers.

Data Analysis

Using the frequency data collected from observations, SDT measures were calculated to evaluate the perceptibility and response bias of infection control precaution alerts. Both parametric and non-parametric statistics were computed (Table 1). To facilitate comparisons, standard accuracy measures of sensitivity and specificity of the alerts were also computed. Sensitivity refers to the ability of the system to correctly identify infectious cases, and specificity is the ability of the system to correctly identify non-infectious cases.

Results

Compliance rate

One hundred and one patient transfers to radiology were observed. Of these, 27 transfers involved transporting an infectious patient. In 12 cases, infection control precautions were not adhered to by the porters. Of the 27 infectious cases, verbal alerts were provided in 3 cases, and written alerts were given in 19 cases (Table 2). Observers' inter-rater reliability was strong, and calculated as 98% ($\kappa=0.88$).

Other transfer errors were also noted, and these had been reported in a separate report¹⁵.

Effectiveness of Alerts

A porter was alerted of the need to take infection control precautions either verbally by a nurse, or through a written transfer form. Both verbal and written alerts achieved perfect specificity. However, written alerts attained a much higher sensitivity (0.70), compared to verbal alerts (0.11) (Table 3). This indicates that written alerts were more reliable than verbal alerts.

Analysis using SDT, however, shows that the response rates differed markedly between the two modes of communication (Table 3). Verbal alerts achieved a perfect hit rate of 1, indicating that whenever a porter was verbally informed of a patient’s infectious status, appropriate precautions were always taken. Written alerts, on the other hand, attained only a hit rate of 0.44. This implies that when a transfer form was used to communicate a patient’s infectious status, compliance rate was only 44%.

Calculations further revealed that the porters were more perceptible to verbal alerts than written alerts, achieving a near-perfect A' value of 0.98, compared to 0.77 for written alerts. Interestingly, the response bias for the two modes of communication also varied significantly. When precautions were communicated verbally, the porters were highly biased towards taking the appropriate action ($c=-0.83$). In contrast, when precautions were noted in the transfer form only, the porters were more inclined to disregard the alert ($c=0.66$).

Table 2. Observed frequencies, where S=Signal was present, ~S=Signal was absent, R=desired response was observed, ~R=desired response was not observed, \cap =Boolean operator “AND”.

Signal	Response Frequencies			
	$S\cap R$	$S\cap\sim R$	$\sim S\cap R$	$\sim S\cap\sim R$
Verbal alert	3	0	7	91
Written alert	8	11	10	72

Table 3. SDT measures for perceptibility and response bias, and standard accuracy measures for alert sensitivity and specificity.

Signal	SDT Measures						Standard Accuracy Measures	
	Hit rate	False alarm rate	Perceptibility measures		Response bias		Sensitivity	Specificity
			d'	A'	β	c		
Verbal alert	1	0.08	4.52	0.98	0.06	-0.83	0.11	1.00
Written alert	0.44	0.12	1.01	0.77	3.84	0.66	0.70	1.00

Discussion

Effectiveness of Alerts – Verbal versus Written

Whilst the reliability of written alerts was much greater than verbal alerts as indicated by the sensitivity and specificity analysis, evaluation of the alerts using SDT showed that verbal alerts were more likely to result in compliance with infection control precautions compared to written alerts. This is contrary to common belief that an alert system with greater accuracy will generally achieve a better response rate⁸. Clearly, signal reliability is only one of many factors that influence the effectiveness of an alert system.

There is an apparent difference in the compliance behavior when an alert is presented through different modalities. Face-to-face verbal handover has often been advocated as the best handover practice, as it facilitates interactive questioning¹⁶. Our observation confirms that verbal handover is indeed more effective than written handover. The analysis based on SDT further throws light on an unexpected advantage of verbal communication. Independent of signal quality, when a warning was given verbally, compliance rate was significantly higher than when it was given

in written form. In other words, an alert was more likely to be heeded when communicated verbally, than when it was scripted.

Observations of the work culture and informal interviews with clinicians provide several clues to this finding. Due to lack of clinical training, porters were discouraged from making decisions, and were instructed to strictly follow the orders of nurses in charge. Indeed, this work ethos was widespread amongst all allied health workers. When asked about job responsibility, one porter simply summarized “I do what I’m told”. As a result, there was an inflated tendency to comply with verbal instruction given by nurses. Every verbal warning was adhered to without questions.

When a warning was scripted, however, the obligation to comply disappeared. Porters were left to decide based on their own criteria. One factor that could potentially influence their decision was the perceived risk in not taking the appropriate precautions. Informal interviews with the porters revealed that the risk of infection was generally recognized. However, since they were accustomed to being exposed, the need for precautions was given less importance than it deserves. This behavior, known as “normalization of deviance”¹⁷, was observed not just amongst the porters, but permeates through all groups of care workers within the hospital. Non-compliance by superiors and peers further reinforced the behavior as acceptable. Thus, the tendency towards non-compliance perpetuates.

Verbal alert was also found to be more perceptible, despite containing the same message as the written form. A likely explanation was the different level of noise when communicating through different media. Verbal instructions tended to be more directed and pertinent to the tasks of patient transportation. The written form contained more information than was required, since the same transfer form was used for other purposes. Thus written communication was *noisier*. As the signal for infection control precautions had to contend with other information in the form for attention, it was sometimes missed.

Our analysis suggests that verbal handover is a more effective communication channel for the setting under study. However, observations also show that verbal handover often did not occur. Due to the high-paced, interrupt-driven clinical environment, verbal communication is often impractical. Thus, written communication is a necessary process. To address the low compliance rate for written warnings, signal perceptibility can be improved by highlighting critical information to the porters. Noise can potentially be reduced by separating information pertinent to the porters, from other information unrelated to the task of patient transportation. To improve response bias, we believe there must be an increased awareness in the risk of infection hospital-wide. Re-educating the porters alone will not have the desired effects, if non-compliance proliferates amongst other groups of clinical workers.

Other Factors Contributing to Non-compliance

Other factors that could potentially affect compliance rate included stress, workload, or just simple oversight rather than a conscious decision not to do so.

Application

In this study, we have shown how SDT can be applied to assess the effectiveness of verbal and written alerts. The same method can be used to evaluate computerized alert systems. Application of SDT involves the following steps: (1) collecting the response frequency for the signal of interest, when the signal is present and when it is absent; (2) calculate the hit rates and false-alarm rates; (3) deducing the perceptibility and response bias measures.

The most practical benefit of the theory is that it provides a number of useful performance measures. These measures enable us to fine-tune an alert notification system to achieve optimal outcome. A common tool used for such purpose is the *Receiver Operating Characteristic (ROC)* analysis. ROC analysis can be applied to optimize cut-off values with regard to the cost ratio of false-positive and false-negative results. This technique is well-documented and widely used in many applications. In healthcare, SDT in conjunction with ROC analysis have been applied successfully in the evaluation of diagnostic imaging systems¹⁸⁻²⁰.

Limitations

There are several limitations to this study. Firstly, we have reported observations of a single process at one hospital. The patterns of errors observed in this systematic convenience sample may be different for different processes, and at other organizations. The probabilistic assessment used in this study also has its limitations. The sample size of

observations is modest, and the likelihood of some less frequent errors being observed is consequently low. The non-obtrusiveness nature of the observational study also introduced uncertainties in the data. Whilst we were able to observe the existence of infection control measures in the ward such as isolation and verbal communications of the need for infectious precautions, we did not carry out a record review, and so may have systematically underestimated the rate at which infectious state was captured in our observations. And finally, whilst SDT facilitates quantitative differentiation between signal perceptibility and response bias, it does not inform us about the costs versus the benefits of information provision. Studies have shown that as the quantity of information being considered increases, the rate of performing tasks decreases²¹. Thus, the provision of alerts must be balanced with the cost they impose. This issue has been addressed in several studies²¹⁻²².

Conclusion

In this study, we have examined the effectiveness of verbal and written alerts in ensuring adherence to infection control precautions during inpatient transfers to radiology. The application of SDT has shed much light on the effectiveness of these alerts, and areas where improvement is likely to be most beneficial. We believe that the technique can be extended to the evaluation of any alert notification systems, including electronic alerts. With the increased reliance of information technology and computer decision support in healthcare, reliable tools for system evaluation is critical in ensuring that the full potential of technology can be realized.

Acknowledgements

This research is supported by grants from the Australian Research Council (LP0775532), and NHMRC Program Grant 56812.

References

1. Joint Commission on Accreditation of Healthcare Organizations. Sentinel event statistic, June 29. 2004.
2. Hall A, Walton G. Information overload within the health care system: a literature review. *Health Info Libr J*. 2004 Jun;21(2):102-8.
3. Sevdalis N, Norris B, Ranger C, Bothwell S; Wristband Project Team. Designing evidence-based patient safety interventions: the case of the UK's National Health Service hospital wristbands. *J Eval Clin Pract*. 2009 Apr;15(2):316-22.
4. Bates DW, Teich JM, Lee J, Seger D, Kuperman GJ, Ma'Luf N, Boyle D, Leape L. The impact of computerized physician order entry on medication error prevention. *J Am Med Inform Assoc*. 1999 Jul-Aug;6(4):313-21.
5. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *BMJ*. 2005 Apr 2;330(7494):765.
6. Ismail ZF, Ismail, TF, Wilson AJ. Improving safety for patients with allergies: An intervention for improving allergy documentation. *Clinical Governance: An International Journal*. 2008;13(2):86-94.
7. Singh H, Arora HS, Vij MS, Rao R, Khan MM, Petersen LA. Communication outcomes of critical imaging results in a computerized notification system. *J Am Med Inform Assoc*. 2007 Jul-Aug;14(4):459-66.
8. van der Sijs H, Aarts J, Vulto A, Berg M. Overriding of drug safety alerts in computerized physician order entry. *J Am Med Inform Assoc*. 2006 Mar-Apr;13(2):138-47.
9. Glassman PA, Simon B, Belperio P, Lanto A. Improving recognition of drug interactions: benefits and barriers to using automated drug alerts. *Med Care*. 2002 Dec;40(12):1161-71.
10. Magnus D, Rodgers S, Avery AJ. GPs' views on computerized drug interaction alerts: questionnaire survey. *J Clin Pharm Ther*. 2002 Oct;27(5):377-82.
11. Weingart SN, Toth M, Sands DZ, Aronson MD, Davis RB, Phillips RS. Physicians' decisions to override computerized drug alerts in primary care. *Arch Intern Med*. 2003 Nov 24;163(21):2625-31.
12. McFall RM, Treat TA. Quantifying the information value of clinical assessments with signal detection theory. *Annu Rev Psychol*. 1999;50:215-41.
13. Swets J. Signal Detection Theory and ROC Analysis in Psychology and Diagnostics: Collected Papers.
14. Stanislaw H, Todorov N. Calculation of signal detection theory measures. *Behaviour Research Methods, Instruments, & Computers*. 1999;31(1):137-149.
15. Ong MS, Coiera E. Safety through redundancy: a case study of in-hospital patient transfers. *Qual Saf Health Care*. 2010 Oct;19(5):e32.

16. WHO Collaborating Centre for Patient Safety Solutions. Communication during patient handovers. Geneva, Switzerland, World Health Organisation; 2007:1-4.
17. Vaughan D, The Challenger Launch Decision, University of Chicago Press, Chicago and London, 1996.
18. Swets JA. Measuring the accuracy of diagnostic systems. 1998 June; Science. 1988 Jun 3;240(4857):1285-93.
19. Metz CE. Receiver operating characteristic analysis: A tool for the quantitative evaluation of observer performance and imaging systems. J Am Coll Radiol. 2006 Jun;3(6):413-22.
20. Greiner M, Pfeiffer D, Smith RD. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. Prev Vet Med. 2000 May 30;45(1-2):23-41.
21. Horvitz E, Barry M. Display of information for time-critical decision making. Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence, Montreal, August 1995.
22. Horvitz E, Jacobs A, Hovel D. Attention-sensitive alerting. Proceedings of UAI '99, Conference on Uncertainty and Artificial Intelligence, Stockholm, Sweden, July 1999;305-313.