Interpreting Nurses' Responses to Clinical Documentation Alerts

James J. Cimino, MD;[1] Lincoln Farnum, RRT-NPS, CHISP;[2] Kelly Cochran, MS, RN;[2]
Steve D. Moore, BS, EMSE;[2] Patricia P. Sengstack, DNP, RN-BC, CPHIMS;[2]
Jon W. McKeeby, DSc;[2]
[1]Laboratory for Informatics Development
[2]Department of Clinical Research Informatics Informatics Outcomes Research
National Institutes of Health Clinical Center, Bethesda, Maryland, USA

Abstract

Objective: To understand how nurses respond to alerts that detect attempts to enter into electronic health records patient weights that vary significantly from previously recorded weights. Methods: Examination of subsequent patient weights to determine if the alerts were true positive (TP) or false positive (FP), and whether nurses overrode alerts, changed their entry or quit without storing a value. Results: Alerts occurred 2.74%, with 41.9% TP and 58.1% FP. Nurses overrode 30.3% of TP and 97.3% of FP alerts. Conclusions: The alert has an acceptable FP rate and does not appear to cause nurses to change entries to satisfy the alert. The alert improves recording of patient weights.

Introduction

Clinical decision support systems (CDSS), in the form of automated alerts and reminders, are a familiar feature of commercially available clinical information systems. Studies have shown that alerts related to events (such as trends in patient laboratory results) and orders (such as potential medication interactions) can impact clinician decision making and improve patient care, although their effects may be blunted when high false-positive rates induce "alert fatigue". Alerts have been used to improve nursing documentation, generally through the detection of incomplete data recording or entry of out-of-range values, including body temperatures, heights and weights.[1-3] These alerts clearly improve the *consistency* of the data that are entered, but relatively little is known about whether this actually reflects improved data *quality*. There exists the possibility that an alert may cause clinicians to change data such that they satisfy the alert, but don't reflect reality.

We have previously reported the development of an alert for improving the recording of patient heights and weights in a clinical research information system.[4] The alert compares the value being recorded with the average of the two previously recorded values and generates an alert when the deviation between the new and old values exceeds a fixed threshold. That system showed a reduction in discrepant values from 2.4% to 0.9%.

Despite this apparent success, we were concerned about a number of possible problems. The first was that although the threshold for alerting is based on the patients' own data, heights, and especially weights, can change significantly over time, leading to false-positive alerts. For example, the less recent the previous values, the less likely they will reliably reflect the true current state of the patient, especially in children. Thus, some perfectly normal situations may result in false-positive alerts.

A second concern is that the introduction of the alert could lead to unintended negative consequences. For example, one reason for the reduction of discrepant values in our study might have been that, when faced with an alert message, nurses simply chose not to record any value, even if their original value was correct. Another possibility is that nurses might alter the values they record to placate the alert.

We therefore sought to study the impact of the alert on nurse documentation. Observational methods seemed inappropriate for this task since they might alter users' behavior. We therefore chose to examine system records to determine how nurses were responding to alerts. We used patient weights recorded before and after the alert to determine whether the alert was a true- or false-positive and characterized nurses' responses to alerts when the data they entered were correct as well as incorrect.

Methods

The Weight Alert: The weight alert compares the value entered with one or two of the most recent previous weights and issued a warning if the difference was 10% or more. The alert did not consider other factors such as the elapsed time since previous measurements or the patient's age. The alert was created as a medical logic module to run in the Clinical Research Information System (CRIS) at the Clinical Center at the National Institutes of Health. This system is a commercial electronic medical records system (Sunrise Clinical Manager, Eclipsys Corporation, Boca Raton, Florida) that has been implemented to support patient care and clinical research for in-patient nursing stations and out-patient clinics. If a nurse enters a weight that

exceeds the alert's threshold, a pop-up window appears that tells the nurse about the deviation, displays the previous weights, and offers the option of returning to the data entry screen or continuing to record the value. An optional comment field is provided to indicate a reason for their action.

Data Collection: We extracted data for all weight alert events that occurred from August 22, 2008 through November 22, 2009 (15 months). We also extracted the birth date and all recorded weights for each patient that had an alert. Patient and nurse identifiers were automatically removed in accordance with NIH policy and with oversight by the NIH Office of Human Subjects Research (OHSR). Data were obtained from the CRIS. OHSR certification was provided by the NIH Biomedical Translational Research Information System (BTRIS.nih.gov).[5]

Data Preparation: For each patient with an alert, we prepared a data extract (examples shown in Table 1) that consisted of the following:

a) weights recorded for the patient prior to the first alert for that patient, with times and dates normalized to the time of the first weight (the "index weight")

b) weight entered by the nurse that caused the alert (referred to as the "alert weight"), with its date and time, normalized as above

c) nurse's response to the alert (back out without storing the alert weight or proceed and store it)

d) weight the nurse subsequently recorded (if any) related to the alert (the alert weight if the nurse proceeded or some other weight after backing out)

e) all subsequent weights recorded for the patient, with their dates and times normalized as above

f) age of the patient at the time of the index weight

Data Analysis: We used extracts to classify the alerts. If subsequent weights were not available, the alert was coded as inconclusive and excluded from further analysis. We characterized each alert as false-positive (FP) if the alert weight was quantitatively between the previous and subsequent weights (as shown in Patient A in Table 1) or was within 10% of subsequent weights (as shown in Patient B in Table 1). Alert weights that did not meet either of these criteria were characterized as true-positive (TP; as shown in Patient C in Table 1). We examined the TP alert weights to identify systematic types of data entry errors and FP alerts to identify whether patterns in the data might suggest ways to improve alert logic.

Nurse responses to alerts were characterized as either Overridden (if they proceeded with the alert and kept

Table 1: Examples of Weight Alert Extracts					
Patient A	Time	0	370.63	825.67	959.9
	Weight	65	***78.1***	80.6	87.1
Patient B	Time	0	364.19	365.25	
	Weight	17.6	*19.8*	19.3	
Patient C	Time	0	181.98	181.98	400.9
	Weight	27.4	*21.2*	**29.2**	31.8

Times are in days relative to initial weight with time of day as a decimal component. ***Bold italic font*** indicates weight that *triggered* an alert and was subsequently **stored** (alert overridden), *italic font* indicates weight that *triggered* an alert but was not stored (quit), and **bold font** indicates a changed weight that was **stored** after an alert.

the alerting value, as shown in Patient A in Table 1), Quit (if they backed out and did not record a weight, as shown in Patient B in Table 1), or Corrected (if they changed the weight to one that was within the 10% threshold, as shown in Patient C in Table 1).

Results

Alert Performance: A total of 92,723 weights were entered into CRIS during the study period, of which 2,538 caused alerts to fire (2.74%) on 1,874 unique patients. In 684 cases (36.5%), the veracity of the alert was inconclusive. We noted that nurses overrode these alerts 635 times (92.8%), backed out of the alert 49 times (7.2%), changing the weight value 17 times and simply quitting the remaining 32 times. These alerts were excluded from further analysis. Of the remaining 1,190 alerts, 499 were characterized as TP and 691 were characterized as FP, for a true positive rate of 41.9% and a false positive rate of 58.1%. Figure 1 shows how nurses responded to alerts of each type.

Nurse Responses to True Positive Alerts: As shown in Figure 1, nurses overrode alerts later characterized as TPs 151 times (30.3%; see Table 2). No reasons were given for any of these overrides. They backed out of the other remaining 348 alerts, correcting their entries 306 times (87.9% of back-outs, 61.3% of all TP alerts) and simply quit 42 times (12.1% of back-outs, 8.4% of all TP alerts).

Nurse Responses to False Positive Alerts: Nurses overrode alerts later characterized as FPs 672 times (97.3%; see Table 3). Reasons for the overrides were given 23 times (3.4% of overrides, 3.3% of all FPs). They backed out of the other remaining 19 alerts, changing their entries 10 times (53% of back-outs, 1.4% of all FP alerts) and simply quit 9 times (47% of back-outs, 1.3% of all FP alerts).
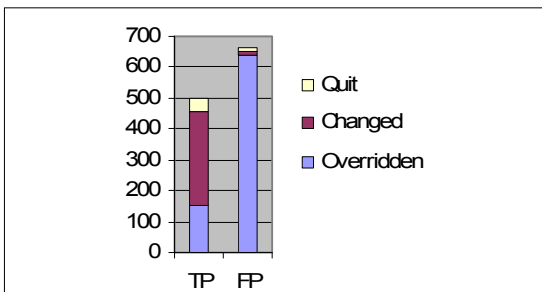
Figure 1: Nurse responses to True Positive (TP) and False Positive (FP) alerts. Note that TP alerts that were changed or quit represent reductions in storage of deviant values. FP alerts that were overridden represent deviant values that were stored despite the alerts.

Data Entry Patterns Related to True Positives: True positive alerts appeared to occur for three reasons. Many cases appeared to be simple transcription errors (e.g., previous weight "46.3", alert weight "15.4", corrected weight "45.4", or previous weight "69.2", alert weight "689", corrected weight "68.9"). In many other cases, the ratio between the alert weight and the corrected weight was 2.2, suggesting that the error was due to measurements being entered in pounds instead of kilograms (e.g., previous weight "62.5", alert weight "139.3", corrected weight "63.3"). In some remaining cases, the amount entered as the weight appeared to be a value for a different body measurement (e.g., previous weight "102.6", alert weight "37.0", subsequent weight "103", suggesting that the value for the alert weight was actually a body temperature).

Data Patterns Related to False Positives: In 244 FPs (34%), the patient's age was less than 18 years and the weight change and time between the index weight and the alert weight were consistent with normal human development. In an additional 228 FPs (32%), the patient was an adult but sufficient time

elapsed between the index weight and the alert weight that the weight change was plausible. In 32 FPs, the weight change between the index and alert weights was consistently noted on subsequent weights, although the elapsed time was short. For example, one patient had an index weight of 38.8, an alert weight of 45.8 two days later, and a subsequent weight of 44.1. In most of these cases, no reason was give to explain the dramatic weight changes reported.

An additional 84 (12%) FPs were actually due to comparison of what appeared to be a correct alert weight (that is, one that was consistent with subsequent weights) to an erroneous value that was stored prior to institution of the alert. The remaining 130 (18%) of FPs fired for unclear reasons – insufficient data were available to tell if the index weight was correct or not, although the alert weight was clearly consistent with subsequent weights.

Discussion

During the study period, nurses entered 92,723 weights and received 2,538 alerts. Of the 1,874 alerts studied, 1,458 were overridden, so that discrepant values were stored 0.61% of the time, instead of the 2.74% rate that would have occurred in the absence of the alert. These rates are consistent with our previous study.[4] The questions we sought to answer in this study relate to whether the alerts are improving or harming the data collection process.

Is the false-positive rate too high? The relative FP rate of 58.1% means that nurses must deal with roughly three false positive alerts for every two true positive ones. However, if the FP rate is extrapolated to all 2,538 alerts (including inconclusive ones), the absolute false positive rate for all 92,723 entered weights is 1.59%, which compares very favorably with rates in other systems that seek to minimize FP alerts to avoid alert fatigue. [2,3,6]

Table 2: Examples of where nurses overrode True Positive alerts.									
Time	0	63	246.96	311.88	319.93	326.98			
Weight	105.6	105.6	102.7	***36.7***	101.6	100.8			
Time	0	127.07	127.07	188.17	189.29	190.18	190.31	191.17	192.18
Weight	52.8	50.2	50.2	***58.9***	48.9	48.9	50	51.3	49.4
Time	0	448.05	448.11	479.09					
Weight	22.4	***60.9***	25.1	25.7					
Time	0	0.22	1.21	7.99	8.04	9.18	336.33	364.31	
Weight	102.2	101.9	101.9	***227.4***	101.4	102	106.7	106.4	
Time	0	1.03	2.05	3	3.13	4.9	6.04	7.04	
Weight	43.5	44.7	46.1	***40.2***	47.6	47.9	48.5	49	
Time	0	2.97	6.78	9.95	16.97	21.03	23.98	27.91	30.97	34.74
Weight	85.4	84.5	83.9	***97***	83.4	82.8	81.9	79.8	80.8	80

Times and weights are depicted as for Table 1 (***bold italic font*** indicates alert weights that were stored)

Time	0	364.01	364.01	734.99					
Weight	95.3	*78.9*	**95**	80.7					
Time	0	0.98	1.97	111.82	153.77	223.79	293.81	376.1	414.86
Weight	95.6	94.8	94.7	*105.9*	**100.8**	106.1	107.3	108.5	107.2
Time	0	0.98	2.03	2.98	2.99	63.01	64.01	98.1	98.85
Weight	49.9	48.1	48.6	*42.2*	**48.2**	41.5	40.4	40.5	40.3
Time	0	35	35	63	88.68	96.81	99.9	266.02	279.96
Weight	162.6	*115.4*	**161.3**	116.8	119.2	118.3	120.5	111.5	111
Time	0	13.76	48.73	90.76	90.76	116.05	116.99	117.93	
Weight	46	45.4	45.4	*50.8*	**47.8**	53.4	52.9	52.7	
Time	0	56	111.6	220.64	227.73	227.73	227.92	248.67	248.93
Weight	122.5	128.5	121.6	*109*	110.6	110.6	110.9	111.9	111.9
Time	0	2.7	10.87	49.8	73.71	77.79	99.75	105.8	134.09
Weight	37.2	37.5	35.4	*32.6*	32.2	33.8	32.3	33.2	33.2
Time	0	181.03	398.06						
Weight	71.8	*87*	76.2						

Table 3: Examples of where nurses changed the value or quit out of False Positive alerts.

Depicted as in Table 1 (*italic font* for weight not stored (quit); **bold font** for changed weight stored after alert)

Are the true positive alerts helpful? Of the 499 alerts identified as TPs, nurses agreed with the alert 70% of the time, a rate that is fairly high in comparison to other studies.[2,3,6] Furthermore, the nurses went on to enter new values 84% of the time when they agreed with the alert. Although a 100% rate would be ideal, this would likely require reweighing patients in some cases. We found that most cases where a nurse did not enter a correct value occurred in the outpatient setting where reweighing may not have been feasible (e.g., the patient went home). We therefore conclude that the alert improved data recording at a rate that is close to ideal, within practical constraints.

The 151 overrides of TPs are of potential concern. Because no reasons were provided in these cases, we cannot know whether the nurses were justified in their insistence on storing apparently discrepant weights or if they simply chose an expedient course of action at the expenses of accurate reporting. We believe that a combination of these explanations is likely. However, when the TP override rate (30.3%), the TP rate (41.9%), and the overall alert rate (2.74%) are considered together, the number of these instances represents less than 0.35% of all recorded weights, which would have been recorded without the alert.

Are false positive alerts harmful? One concern that must always be considered when implementing clinical alerts is whether FP alerts will mislead clinicians. In our study, nurses agreed with FP alerts 2.7% of the time – a much lower rate than with TP alerts, suggesting that they generally recognized these alerts to be false. In 1.3% of the cases, nurses backed out and did not enter new values, thus missing an opportunity to save a correct value. We hypothesize that, because almost all of these occurred in the outpatient setting, they represent instances where confirmation by reweighing was simply not possible or the data actually belonged to another patient. However, when the FP agreement rate (2.7%), the FP rate (58.1%), and the overall alert rate (2.74%) are considered together, the number of these instances represents less than 0.04% of all weights that would have been recorded without the alert. The 10 cases where nurses "corrected" FPs to enter values closer to previous weights (but which were actually discrepant with subsequent weights) are of concern, since there is the possibility that some of these values were deliberately false. However, even if this is occurring occasionally, it is clearly rare.

Does the benefit of the alert outweigh the harm? Taken in total, we conclude that the rate at which the alert is firing inappropriately is acceptably low, that it is improving data by preventing incorrect data storage 348 times per 1,190 alerts, with potential harm by not storing correct values 19 times. We believe that this 18.3:1 ratio indicates that the alert is beneficial.

Can the performance of the alert be improved? A small number of alerts fired due to comparison of the alert weight with previously stored values that were incorrect. Such occurrences should become less frequent as the number of these values decreases. However, the 3:2 ratio of FPs to TPs clearly suggests room for improvement. One possibility is for the logic to take into account factors such as the elapsed

time between the alert weight and the prior values, especially when the patient's age is taken into consideration. The proper "tuning" of the logic to consider age and time is not straightforward and will require careful study to avoid increasing the number of false negatives (that is, situations where the weight entered is deviant but the alert does not fire).

The alert might also be improved by including information in its message about potential reasons for the discrepancy (such as the possibility that the value entered is measured in pounds instead of kilograms, or is missing a decimal point, or is actually a temperature). The usefulness of such suggestions might not outweigh the benefit of the current, briefer message, and is a subject for further study.

Comparison with other studies: In previous studies of nurse responses to patient measurement alerts, Kroth and colleagues showed that an alert related to the recording of implausible body temperatures could induce nurses to repeat their measurements;[2] however, they did not report the rate at which remeasurements were automated or manual. Caraballo and colleagues studied the impact of alerts on height and weight recording and used the health record to verify whether alerts were true or false positive. However, they did not describe the nurses' responses to the alerts, other than to report a decrease in outlier values.[3]

Limitations: Our study deliberately avoided direct observation of nurse behavior. As a result, we can only speculate on the reasons for nurse's responses to the alerts. However, the responses we saw were largely consistent with best practices. Also, because we did not specifically weigh patients ourselves at the time of the alerts, we cannot be sure of their true weights. However, we believe the weight trends are a reasonable surrogate. The differences in nurse responses to TP versus FP alerts (when they could not know the subsequent weights that would be stored in the record) support this contention.

Significance of findings: Our study confirmed that the weight alert has a substantial impact on reducing the recording of discrepant values. More importantly, we showed that the FP rate is acceptably low and that responses to both FP and TP alerts are appropriate. These results translate directly to both improved patient care and improved clinical research, where incorrect weight records could distort study results or lead to improper dosing of medication (thus putting the patient at risk as well as possibly invalidating a treatment protocol). Studies such as ours are important for verifying that specific alerts are improving, rather than harming care. We have

demonstrated that the use of recorded patient data is a valuable approach for studies of selected alerts.

Future studies: This study points the way to improving the weight alert through tuning with age and time lapse parameters, inferring common causes of transcription errors, and including language that warns that the weight change, if true, is cause for concern. Our experience with the weight alert provides valuable insights for development of additional alerts related to patient data collection.

Conclusions

The weight alert implemented in the Clinical Research Information System at the NIH Clinical Center is improving the quality of data being recorded to support high quality patient care and clinical research, with an acceptable impact on user workload. Our study reassures us that negative effects of the alert are minimal and that the few instances where alerts adversely impact data collection are vastly outweighed by the benefits.

Acknowledgments

References
1. Pryor TA. Computerized nurse charting. Int J Clin Monit Comput 1989;6(3):173-9.
2. Kroth PJ, Dexter PR, Overhage JM, Knipe C, Hui SL, Belsito A, McDonald CJ. A computerized decision support system improves the accuracy of temperature capture from nursing personnel at the bedside. AMIA Annu Symp Proc. 2006:444-8.
3. Caraballo PJ, North F, Peters S, et al. Use of decision support to prevent errors when documenting height and weight in the hospital electronic medical record. AMIA Annu Symp Proc. 2008:890.
4. Haerian H, McKeeby J, DiPatrizio G, Cimino JJ. Use of Clinical Alerting to Improve the Collection of Clinical Research Data. AMIA Annu Symp Proc. 2009:218-22.
5. Cimino JJ, Ayres EA. The Clinical Research Data Repository of the US National Institutes of Health. Proc of the 2010 Medinfo (*in press*).
6. Shah NR, Seger AC, Seger DL, Fiskio JM, Kuperman GJ, Blumenfeld B, Recklet EG, Bates DW, Gandhi TK. Improving acceptance of computerized prescribing alerts in ambulatory care. J Am Med Inform Assoc. 2006 Jan-Feb;13(1):5-11.