Using Lean Six Sigma® Tools to Compare INR Measurements from Different Laboratories Within a Community

Brion Hurley, CSSBB; James M. Levett, MD, FACS; Carla Huber, ARNP; Tim L. Taylor, SSBB

Abstract

Whenever measurements are taken on a patient or health care process, the measurement system must be analyzed in order to verify that it is adequate for the application. A Gage Repeatability and Reproducibility (R&R) study was performed to determine if blood samples taken from a patient on warfarin have the same International Normalized Ratio (INR) results when analyzed in different labs (reproducibility) and when analyzed multiple times in the same lab (repeatability). Results showed a statistically significant difference among labs. The therapeutic range for INR is typically 2.0 to 3.0, yet the data showed a difference in INR of 0.4 among labs on a small sample of 10 warfarin patients, almost 50 percent of the range. Root cause analysis revealed that the normalizing number (mean normal prothrombin time, MNPT) utilized in one of the labs varied greatly from other area labs. The team determined a new qualification process for MNPT to keep results consistent across all the labs.

Introduction

Use of anticoagulation within the inpatient and outpatient settings has expanded greatly in recent years, particularly with the aging of the population and the growing number of indications for anticoagulation. Over 2 million patients in the United States are taking warfarin (a blood anticoagulant, also referred to as Coumadin®),[1] and it is estimated that between 3,000 and 4,000 patients in the Cedar Rapids community take warfarin on a regular basis.

Cedar Rapids is a northeastern Iowa community with a population of approximately 150,000, and a primary and secondary service area of around 300,000.[2] Warfarin anticoagulation accounts for approximately 25 percent of adverse drug events (ADEs)[3] (i.e., any unexpected or dangerous reaction to a drug) in the two community hospitals. Cedar Rapids has a high proportion of elderly patients who are at particularly high risk for an ADE.

Several members of the physician and health care provider community believed that establishing an anticoagulation clinic within the Cedar Rapids community would allow standardization of protocols for prescribing warfarin, improved monitoring of patients on warfarin, and improved communication with patients. The features of such a clinic would also improve the culture of safety within the community, reduce errors and adverse drug reactions, and reduce readmissions for complications in patients taking warfarin.

Establishing an anticoagulation clinic specifically addresses two National Quality Forum safe practice items:[4] Item 1. Creation of a health care culture of safety; and Item 18. Utilization of dedicated antithrombotic services that facilitate coordinated care management.[5]

In 2005, the Agency for Healthcare Research and Quality (AHRQ) awarded a 2-year grant (Award No. 5 U18 HS015830-02) to Kirkwood Community College and Physicians' Clinic of Iowa (PCI). The grant supported the creation of the Community Anticoagulation Therapy (CAT) Clinic and the Cedar Rapids Healthcare Alliance (CRHA) to oversee the CAT Clinic. The project represents a broad coalition of stakeholders including provider organizations, hospitals, payers, employers, educational institutions, and patients.

Rockwell Collins, Inc., designer and manufacturer of aircraft avionics and military systems, with several ISO certifications and extensive experience in Lean, Six Sigma[®6] and ISO-9000[7] quality systems, agreed to offer training in Quality, Lean, and Six Sigma concepts. To identify areas for process improvement, Rockwell Collins improvement specialists began work with community health care providers during the analysis and design phase of clinic development. The design and operation of the anticoagulation clinic involved numerous quality concepts used in the avionics industry; one specific experience is discussed in this paper.

Methods

The Rockwell Collins improvement specialists, with a background in Lean and Six Sigma, used the Define-Measure-Analyze-Improve-Control (DMAIC)[8] approach for improving this process. Each phase of the DMAIC model is very structured with respect to both the actions completed in the phase and the improvement tools implemented.

To begin the project, the specialists spent several days evaluating the overall process of anticoagulation management currently being performed at two cardiology clinics in Cedar Rapids. These two practices manage approximately 2,000 patients on long-term warfarin therapy, representing about half of all warfarin patients in the community. The evaluation involved process and value stream mapping, combined with the application of Lean and Six Sigma concepts in the overall analysis.

One of the first steps when making improvements to any metric is to determine if the measurement system is adequate. Early in the experience, it was determined that four laboratories were being used within the community to run tests for the International Normalized Ratio (INR) values—the key measure of warfarin anticoagulation—for the patients. In this process, the question was whether the INR values were comparable among the different labs.

The labs explained that, since normalizing factors are taken into account to adjust for equipment and plasma material differences used in the testing process, INR values are indeed standardized, in contrast to the prothrombin time (PT) values previously used to manage patients. To validate this statement and the measurement system, a Gage Repeatability and Reproducibility (R&R)[9] study was performed at two of the four labs. A Gage R&R study is a structured statistical experiment that is designed to compare the repeatability (within the same lab) and

reproducibility (between labs) and to compare measurement variations to the typical limits needed for the process. The details of the Gage R&R are described below.

To conduct this experiment, it was necessary to make a case that the experiment was needed. Thus, a small study with non-warfarin patients was established to see what preliminary results could be obtained. The patients who participated in the study were informed of the experiment ahead of time, and appropriate consent was obtained. Four volunteers involved in the project were used for the initial study, and six tubes of blood were collected from each volunteer (all from the same blood draw) at the same lab (Lab A). To remove the lab-to-lab blood draw process variation from the study, one lab was chosen as the "Host." Three tubes were sent to a second lab (Lab B), while retaining the other 3 tubes at the first lab (Lab A). Each lab would take PT and INR results at three specific times during the day: 8 a.m., 12 pm, and 4 p.m. Thus, each volunteer would have three tubes tested at each lab, at established times of the day. The experiment would determine if testing time delays or differences between the two labs provided a greater influence on INR determinations (Table 1).

Results

After collecting the data, it wasnoted that there was a consistency to the repeat results within 8 hours of testing. However, there appeared to be a lab-to-lab difference. Table 1 summarizes the data, by patient, to document the observed differences. The Gage R&R calculations showed that the R&R variation accounted for 38.49 percent of the

Table 1. Summary of INR differences between Labs A and B

| Patient | Average INR | | INR Difference |
	Lab A	Lab B	
1	0.953	1.050	0.097
2	0.890	0.963	0.073
3	1.010	1.100	0.090
4	1.000	1.097	0.097
Average	0.963	1.053	0.089

INR = International Normalized Ratio

tolerance, with most of the variation coming from the reproducibility (lab-to-lab differences). Therefore, almost 40 percent of the variation in a patient's therapeutic range (typical therapeutic ranges for INR values are 2.0-3.0; 2.5-3.5; and, 3.0-4.0) could be coming from the measurement system. A good measurement system must be less than 30 percent, and ideally less than 10 percent. The difference between the two labs was statistically significant, even with just the four patients used in the study (P <0.001), which justified using actual warfarin patients. Permission was obtained from the Institutional Review Board, which serves the two community hospitals in Cedar Rapids, to begin work on the followup Gage R&R study.

Since the 8-hour blood testing time delay study did not show any significant differences, the second experiment was revised to focus on the differences in lab results using 10 patients. The 10 patients who participated in the study were informed of the experiment ahead of time, and appropriate consent was obtained. They were selected on a "first-come, first-served" basis. Each patient had his/her blood drawn from Lab A, and the blood was then separated into six tubes.

Three tubes remained at Lab A for testing, and three tubes were immediately sent to Lab B for testing. The labs were located near one another, so after some discussion with lab personnel, it was agreed that there was minimal concern about travel and handling effects.

The initial findings with the non-warfarin patient Gage R&R study were confirmed by a significant difference between INR results from Lab A and Lab B ($P <0.001$) (Table 2). The Gage R&R study results determined a percent of tolerance of 116 percent, which is far worse than the initial study. This variation required attention first, before trying to address the root cause of the patient's INR result variation, since the measurement system can be a significant factor in these differences.

Table 2. Summary data from warfarin patient Gage R&R study

Patient	Average INR		INR difference
	Lab A	Lab B	
1	2.777	3.170	-0.393
2	2.100	2.320	-0.220
3	2.887	3.110	-0.223
4	1.693	1.830	-0.137
5	2.920	3.160	-0.240
6	1.267	1.413	-0.147
7	3.877	4.320	-0.443
8	2.090	2.240	-0.150
9	2.993	3.300	-0.307
10	3.300	3.553	-0.253
Average	2.590	2.842	-0.251[a]

a $P <0.001$

R&R = repeatability and reproducibility

The next step was to determine the root cause of the differences. A better understanding of the INR formula was needed. The INR value is a normalized value to measure the coagulation time of a patient's blood. If the results are too high or too low, compared to the recommended therapeutic range, then the patient's dosage is adjusted accordingly. Since dosage changes are made based on this value, it is critical that the INR values be precise and accurate. A defect of the anticoagulation management process occurs when dosage is changed unnecessarily (Type I error) and when dosage is not changed when it should be (Type II error). Reducing the variation in the INR actual values would reduce the chances of Type I and Type II errors.

The INR value is calculated from a blood test, using the prothrombin time (often called pro time, which measures how many seconds it takes the blood to coagulate) and the International Sensitivity Index (ISI) rating of the thromboplastin (reagent) being used at the lab to test the blood. The purpose of pro time/INR ratio is to offset differences in thromboplastins and instruments being used at various labs, so a patient can get the same results, regardless of where the blood test is performed.

The third variable in the INR equation is a normalizing factor, called the mean normal prothrombin time (MNPT), which is the geometric mean of the PT of 20 or more non-warfarin

patients, conducted annually at each lab. This is the factor used to standardize labs to one another:

$$INR = (Patient\ pro\ time/MNPT)^{ISI}$$

For Lab A, MNPT = 11.4 and ISI = 1.93; for Lab B, MNPT = 11.9 and ISI = 1.97. It was initially thought that the differences in ISI values were the cause of the variations, since they were directly attributable to the manufacturer of the thromboplastin. However, looking at the INR equation, the impact on INR results seemed to be more highly affected by MNPT differences than ISI differences. To clarify this, a hypothetical simulation was conducted using patient data but keeping either the ISI or MNPT value constant for both labs.

After running the simulation with both the ISI and MNPT values held constant, the difference between labs was still significant if the ISI values were equal (Table 3). They became insignificant if the MNPT values were equal (average difference = 0.019).

This indicated that the lab differences were a result of differences in MNPT value. The next step was to determine why the MNPT values were different.

First, the process used to calculate MNPT had to be understood. Each lab used 20 non-warfarin patients, both males and females, to determine the MNPT on an annual basis. Each lab took a minimum of 20 individuals,

Table 3. **Simulation results: Constant ISI value (1.97), different MNPT values**

Patient	Average INR		INR difference
	Lab A	Lab B	
1	3.167	2.855	-0.313
2	2.318	2.138	-0.180
3	3.106	2.967	-0.139
4	1.832	1.719	-0.113
5	3.157	3.005	-0.152
6	1.409	1.276	-0.153
7	4.323	4.016	-0.307
8	2.240	2.130	-0.110
9	3.301	3.081	-0.220
10	3.554	3.406	-0.148
Average	2.841	2.659	-0.181[a]

a $P < 0.001$

ISI = International Sensitivity Index; MNPT = mean normal prothrombin time; INR – International Normalized Ratio

tested their blood with the current lot of thromboplastin, and calculated the geometric mean of the pro time determinations from those patients. To better understand how the labs came up with vastly different MNPT values, the data from those 20 patients were needed. Since there was no accepted standard for the optimal MNPT value, the data from all four labs were obtained and used to help understand which lab MNPT value was more accurate.

The normal patient pro time results from each lab are shown in Figure 1. The results from Lab A differed significantly from the other labs. The difference between the pro times of Lab A and Lab B was statistically significant ($P = 0.0003$). Compared to Lab C and D, Lab B appeared to

be closer in average. Thus, there was reason to suspect that Lab A was doing something different from the other labs, which could have led to lower pro time values. This would change the MNPT value, which consequently changed patients' INR results. However, the possibility could not be excluded that Lab A had the correct MNPT value, and the other lab results were less accurate.
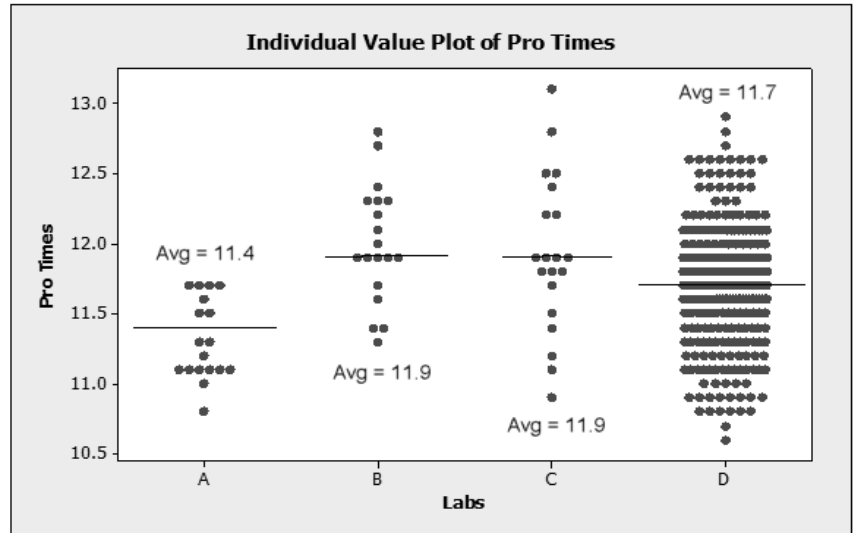
Figure 1. Individual values plot of pro time results from non-warfarin patients by lab.

To review the results and brainstorm the potential causes for the discrepancy, we convened a meeting that included representatives from the four labs and the test equipment manufacturer. All data from the study were presented and reviewed with the team. A Cause-and-Effect (Fishbone) Diagram was then created to organize the causes.[8] The Rockwell Collins support team facilitated the session.

The group identified numerous possible causes. Under "Measurement," some labs performed statistical analysis on the results to identify outlier MNPT values (results that would fall outside 3 standard deviations from the group mean of the sample), while other labs used all patient determinations without checking for outliers. Under "Methods," the criteria for selecting a "non-warfarin normal" patient seemed to be rather ambiguous, where some labs used staff and personnel in the lab, and others used patients coming in for blood testing who were not on warfarin. As a result of the brainstorming session, the group came to the consensus that a new MNPT calculation process was needed, one that would be consistent among all four labs in Cedar Rapids.

Under the new MNPT process, blood samples would be collected within the same group of "normal" patients using new criteria provided by the equipment manufacturer. The samples would be divided evenly and provided to each lab to collectively set up the equipment and MNPT values at their facility. Labs with MNPT results that differed from the others could be quickly identified and corrected before test results were provided to actual patients.

The group also decided to transition over to Innovin®, a manufactured thromboplastin that has a consistent ISI value. This change minimizes the amount of adjustment that needs to be made with every new lot of material shipped and avoids the need for coordination of thromboplastin lots within a community. Previously, all labs within a city, region, or community had to have the same thromboplastin lot number, which required a great deal of coordination from both the labs and the thromboplastin supplier. This change became official in Cedar Rapids in March 2007.

Discussion

Once the Cedar Rapids community had successfully addressed its measurement system issues, the focus could move to individual patients and how they could better control their INR results. Luckily, the groundwork had already been established on how to capture and analyze patient INR data. In practical terms, most patients in our community typically use the same lab for all their testing, so lab differences would not be likely to directly affect the variation of their results, although it could consistently offset the true result for each patient. To adjust for that lab difference, some patients might get more or less medication than they need.

On a national scale, many patients on warfarin are "snowbirds," who travel south for the winter. Blood testing still occurs in their winter destination, so this variation in lab values might be even greater than that in their local community. If the labs are not in agreement, then some patients are getting results higher or lower than their true INR value and are unknowingly receiving a different amount of medication due to the difference in testing methods.

One of the goals for improving the process is to reduce the communication time from completion of the lab testing until the patient is notified of the INR result. Lean tools, useful for addressing process speed, are employed in this process to address the lag time. We observed that each lab had different names for these lag times, and some labs measured the lag times at different points in the process. Consequently, the time data needed from the labs had to be standardized first, before the overall cycle time could be properly managed and tracked in the system.

The new system also integrates a first in - first out (FIFO) approach, showing the order of patients needing to be called, based on how long their results have been in the system. Specific time goals can also be established within the new system to highlight specific parts of the process that take longer than expected (e.g., time to enter data into the system when sent from the labs) and to help identify areas for improvement.

Another observation is that some of the labs batch INR results and fax them to the clinic all at one time or when some predetermined number deemed sufficient are available to be faxed. This is commonly done when processes are handed off to another area because it is an efficient way to minimize the total amount of time for the labs to fax results. However, batching often extends the overall length of time that patients have to wait to receive their results. Batching might be optimal for the labs, but the customer (patient) does not get timely results with this approach. "One-piece flow" is a key teaching point of Lean methodology, since it seems counterintuitive to those working in the process. After explaining the impact on the patient, the labs have been encouraged to fax the results as soon as they are completed to reduce the overall time it takes to communicate back to the patient.

With improved INR data coming from the labs, another goal of the project was to identify the reasons why so many dosage changes had been made in the past. Since the INR value was the key indicator for dosage changes, the team agreed that having the ability to monitor and track real-time patient INR data was critical. In the process of designing the CAT Clinic processes and procedures, it became clear that a software system and database were needed. No software products available on the market fit the needs of the project, so a new software system was

designed by an outside software development company and programmed specifically for the CAT Clinic.

One of the key features needed in the new database system was Statistical Process Control (SPC).[10] A control chart of each patient's historical INR values was integrated into the system, along with a histogram, enabling the nurse practitioner to make a statistically based decision as to whether a patient required a dosage change. Typically, this decision was based solely on the most recent results (last INR value compared to current INR value), background knowledge of the patient, and the experience level of the nurse in an anticoagulation clinic. If the patient's INR result was within the control limits, then the patient was deemed "under control," and the standard dosing guidelines were applied to that patient. If the patient was deemed "out of control," then a more detailed evaluation of the patient's compliance to their dosing, diet, and medication usage was necessary before making any permanent dosage changes.

With SPC in place, the patients' results were entered into the system and the chart displayed, along with any "out-of-control" conditions. The control chart helped identify these conditions through the use of trend analysis and statistical control limits. These limits helped identify the region where a high probability of INR results should fall, specific to each patient. Results outside that region were unlikely (low probability of occurring), based on that patient's historical results, and had to be investigated.

Trend analysis looks for patterns in the data, such as 6 increasing or decreasing points in a row, or 9 data points above or below the average line. These may still fall within the statistical limits, but they would be indicators that something abnormal is happening, prompting an investigation. Not all identified "out-of-control conditions" are valid or indicate a problem, but they all require evaluation.

If an out-of-control condition were deemed relevant, and an assignable or identifiable cause were determined for the out-of-control condition, then either the cause would be noted and the data point essentially ignored, or the process would be changed (dosage change or change in testing frequency) to correct the issue. An example of an out-of-control condition is a situation where a patient admits to recently ingesting a large amount of alcohol. This would explain the higher INR result, so there would be no evidence showing the dosage amounts to be incorrect. In such a case, the data point would be isolated and reviewed with the patient, who would be asked to return for another INR test at a later date after the effects of that issue have been minimized.

Repeat INR determinations would be made until the patient was back in therapeutic range. An incorrect conclusion from that situation might result in a permanent dosage change that could inappropriately lower the patient's INR, especially if the patient avoided alcohol ingestion. This dosage adjustment, to account for outliers and special causes, could lead to serious complications. This new approach requires the clinic staff to ask a lot more questions of the patient to determine whether a special cause exists. To help increase understanding of this concept, clinic and lab personnel are now provided with SPC training to explain the details behind outliers and out-of-control conditions.

Figure 2 shows an actual patient's control chart (the patient's name has been changed). The bar in the middle of the figure denotes the patient's therapeutic range; the changing line indicates the patient's actual INR results; the areas directly above and below the bar in the middle of the figure are the patient's statistical process control limits (three
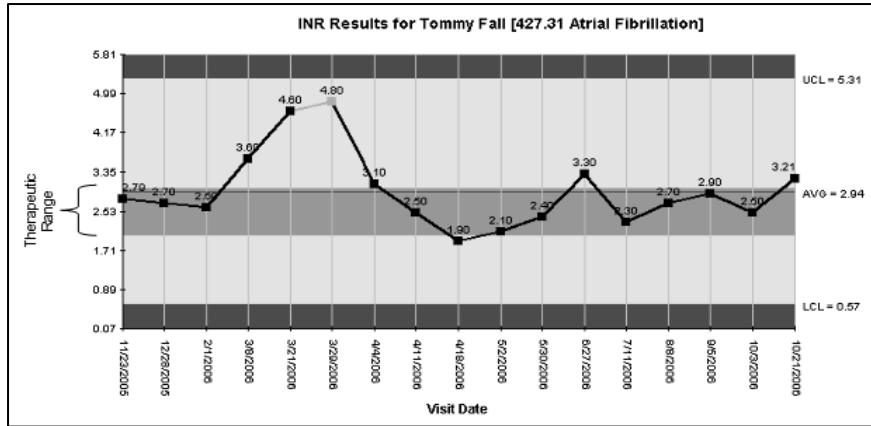

Figure 2. Warfarin patient SPC chart, from INR Pro© software.

standard deviations above and below the average); and the very top and bottom bars are considered an unlikely INR value for the patient that falls outside individual control limits.

Based on historical data, we have learned that this patient does not stay within the therapeutic range very often. The range is 2.0 to 3.0, yet his average result is 2.94, on the high end of the range. On March 29, 2006, "Tommy" had an INR determination of 4.80, which violated one of the out-of-control condition criteria (2 out of 3 consecutive data points beyond the 2-sigma warning limits) and therefore indicated that an investigation was needed. His INR results appeared to be steadily increasing, possibly due to the medication. Later on June 27, 2006, he had a result of 3.30, which was outside his range but well within his control limits, so an unnecessary dosage change was avoided. By now, there were enough data to support a dosage change to lower the INR result, so that the results would be closer to 2.5. The key difference in this dosing approach is that more than one data point was used to determine the need for a dosage change.

The use of SPC differs from the current method used by many providers because with SPC, the nurse practitioner and physician no longer make dosage change decisions based on "gut-feel" or intuition, which often leads to an effect called "tweaking,"[11] commonly referred to as a Type I error. Tweaking is intuitive to most people, since it attempts to make minor process changes to achieve a better end result. For example, if the patient had an INR of 2.65, and the appropriate range was 2.0 to 3.0, with 2.5 being the target value, it would be intuitive to want to slightly reduce the patient's dosage in order to get closer to the ideal value of 2.5. However, the difference between 2.65 and 2.5 is most likely due to "noise," or random variation in the process. Noise can result from a patient's eating habits, lab variation in testing, blood draw techniques, and time of day, among many other possible factors.

In attempting to make fine adjustments to the patient's dosage, tweaking may erroneously make things far worse for the patient. Every time the dosage is changed, the risk of hospitalization increases or a bleeding episode might occur due to excess or insufficient medication. The patient also might become confused about the new medication dosage and when it is to be taken. Warfarin patients are often elderly, which makes it harder to implement dosage changes.

Another concern is that some patients have more variability in their INR results compared to other patients. If a patient has a wider range or variation in his/her INR results, then the provider must be more cautious when making dosage changes for that patient, since historical data suggest that large changes in INR results would be expected for that patient.

Control charts must be understood in the clinical setting. Traditional control charts use an estimate of 3 standard deviations to determine the width of the control limits. Due to the lack of frequent results (lab tests taken weekly or monthly, not daily) and the fact that the random variation in INR readings has not been quantified (How much does an INR result vary throughout a given day?), modifications to the traditional control chart for application in anticoagulation management might be needed. As the use of SPC and control charting increases for managing anticoagulation patients, additional analysis must be performed to better understand whether 3 standard deviations is an appropriate calculation for control limits.

When the statistical upper control limit for a patient exceeds a value of 5.0—considered to be "high risk" for a complication—there is another concern with this new approach. If a patient has a therapeutic range of 2.5 to 3.5 and his/her INR is 4.7 (less than the upper control limit), then some intervention may be need in order to reduce the immediate risk of a complication. This would be the case even though an INR result of 4.7 was not considered "out of control." Additional guidelines and decision rules need to be developed to deal with these unique situations in order to avoid additional patient harm. Regardless of these concerns, it is imperative that INR results be viewed in a control chart format before making any decisions about a patient's dosage amounts.

Conclusion

In summary, the use of statistical tools and techniques, such as Gage Repeatability and Reproducibility studies and Statistical Process Control (key tools utilized in Lean Six Sigma programs), have practical application in health care, specifically in an anticoagulation management process. The benefits expected from these improvements include reduced dosage changes (causing fewer adverse drug events), shorter response times in communicating results to the patient, better patient access to historical results (Web-based access to their data to help them take more accountability for managing their results), and less variation in INR results among the labs (causing fewer dosage changes and more time in range).

Acknowledgments

Author Affiliations

Rockwell Collins, Inc. (Mr. Hurley, Mr. Taylor); Physicians' Clinic of Iowa (Dr. Levett); CAT Clinic (Ms. Huber).

Address correspondence to: Brion Hurley, 1425 Highway A1A #20, Satellite Beach, FL 32937; e-mail: brionhurley@hotmail.com.

References

1. Gitter MJ, Jaeger TM, Petterson TM, et al. Bleeding and thromboembolism during anticoagulant therapy: A population-based study in Rochester, Minnesota. Mayo Clinic Proc. 1995; 70: 725-733.

2. Cedar Rapids Chamber of Commerce. Demographic information for the Greater Cedar Rapids area: 2005-2006. Washington, DC: U.S. Department of Commerce, Bureau of Census; 2005.

3. Fihn SD, Callahan CM, Martin DC, et al. The risk for and severity of bleeding complication in elderly patients treated with warfarin. Ann Intern Med 1996; 124: 970-979.

4. Safe Practices for Better Healthcare 2003. Washington, DC: The National Quality Forum. Available at: 216.122.138.39/pdf/reports/safe_practices.pdf. Accessed March 7, 2008.

5. Feldstein AC, Smith DH, Perrin N, et al. Reducing warfarin medication interactions: An interrupted time series evaluation. Arch Intern Med 2006; 166: 1009-1015.

6. George ML. What is Lean Six Sigma? New York, NY: McGraw-Hill; 2003.

7. Hoyle D. ISO 9000 quality systems handbook. 5th ed. Burlington, MA: Elsevier; 2006.

8. Breyfogle FW III. Implementing Six Sigma: Smarter solutions using statistical methods, 2nd ed. Hoboken, NJ: Wiley and Sons; 2003.

9. Measurement Systems Analysis Work Group. Measurement systems analysis reference manual, 3rd ed. Automotive Industry Action Group. Daimler Chrysler Corporation, Ford Motor Company, and General Motors Corporation; 2002.

10. Western Electric. Statistical quality control handbook. Indianapolis, IN: Western Electric Corporation; 1956.

11. Deming WE. Out of the crisis. Cambridge, MA: Massachusetts Institute of Technology; 1986.