Accred Qual Assur (2002) 7:461–467
DOI 10.1007/s00769-002-0527-2

GENERAL PAPER

Jan Krouwer

Using a learning curve approach to reduce laboratory errors

Received: 3 February 2002
Accepted: 17 July 2002

Presented at the European Conference on Quality in the Spotlight in Medical Laboratories, 7–9 October 2001, Antwerp, Belgium

Jan Krouwer (✉)
Krouwer Consulting, 26 Parks Drive, Sherborn, MA 01770, USA
e-mail: info@krouwerconsulting.com
Tel.: +1-508-653-2379
Fax: +1-508-647-9380

Abstract Hospital laboratories have error rates that are too high and in some cases may be responsible for adverse patient treatment. This paper introduces reliability growth management (RGM), which is based on learning curve theory, as a method to improve laboratory error rates. RGM is widely used in the defense and automotive industry to solve problems when resources are limited and knowledge about the product and/or process is incomplete. An example of RGM, which was used to improve the reliability of instrument assay systems in the medical diagnostics industry is presented. RGM is a closed-loop process that entails creating a goal and event model, classifying events with failure review and corrective action system (FRACAS), tracking progress and predicting completion with Duane analysis. Results achieved by RGM were far better than those obtained by previously used methods. RGM techniques can be transferred to hospital laboratories to reduce laboratory error rates. The advantages of RGM compared to other quality initiatives such as ISO 9000 and Six Sigma are discussed.

Keywords Reliability growth management · Quality · ISO 9000 · FRACAS · Six Sigma

Abbreviations *RGM* Reliability growth management · *FRACAS* failure review and corrective action system · *CAP* College of American Pathologists · *LIS* laboratory information system · *HIS* hospital information system

Introduction

Recently, much attention has been devoted to preventable medical errors [1, 2] . Not surprisingly, this attention has in turn focused on errors in hospital labs. For example, a recent case involved a woman who underwent surgery to remove malignant tumors, based largely on a lab result. It turned out that the lab result was wrong and that she did not have cancer [3]. Of course, lab errors do not have to be as serious as incorrect assay results. If a lab test result is late (e.g., does not meet an agreed upon turn-around-time), then this too is a failure.

These failures may be regarded as quality problems. Improving quality in labs has been an ongoing activity for many years both by manufacturers of assay instrument systems and by the hospital labs themselves. In some countries, quality is also monitored by regulatory agencies. Whereas these efforts have been fruitful, recent studies show that failures still occur in hospital labs at rates that are much higher than those of other industries [4, 5].

The purpose of this article is to illustrate a new method to improve lab quality. Based on learning curve theory, this method has been applied for years in the defense and automotive industry and recently in the medical device industry. Its principles will be illustrated with an example of how it was applied to improve the reliability of instrument assay systems. The steps to transfer this technology to hospital labs will be discussed. Finally, the advantages of learning curve theory over other popular quality programs such as ISO 9000 and Six Sigma will be discussed.

Reliability growth management – a technique based on learning curve theory

Reliability growth management is credited to Duane, who while at General Electric observed that reliability improvement for a variety of different systems followed a similar pattern. Improvement was proportional to the cumulative time that the systems were under test (Fig. 1). This observation is based on learning theory [6]. Thus, as a system was under test, failures were observed. A fix for the observed failure was proposed and applied. The system was then retested. This cycle of operating the device, observing failures, fixing failures, and retesting continued until the reliability goal was reached.

In reliability engineering, reliability growth management is different than demonstration testing since in demonstration testing, reliability is assessed after a product has been designed and manufactured. In reliability growth management, reliability is assessed and improvements are made during the design and development phase.

Duane's initial observations were developed over several years into a codified practice known as reliability growth management that is widely practiced in the defense and other industries [7]. The practice was adapted to the medical diagnostic industry at Ciba Corning Diagnostics [8] to improve the reliability of instrument assay systems.

Reliability growth management entails the following steps:

1. Creation of a high level quantitative quality goal
2. Creation of a model to ensure that all relevant events are counted
3. Classification of errors as to their frequency and severity
4. Ranking of errors by Pareto analysis
5. Formation of corrective action teams to solve the top problems
6. Tracking the top errors
7. Prediction of when the goal will be reached

Figure 2 shows this closed-loop process.

Designing in quality vs. testing in quality

Many quality initiatives emphasize "designing in quality", or doing it right the first time and cast a dim view of "testing in quality". Designing in quality is seen as preferable because there is an implicit assumption that it is more efficient than testing a product or process to determine what needs to be changed. In fact, "designing in quality" and "testing in quality" are not mutually exclusive. Each should be used where appropriate even on the same program. The choice of the right quality program depends on the state of knowledge of the technology.

If one has a complete state of knowledge (e.g., where exact equations describe knowledge), where one knows all of the failures that might occur and knows precisely how to change the design to mitigate these failures, then one can and should design in reliability and avoid testing in quality. However, for complex systems and processes (e.g., the clinical laboratory) the number of potential failure modes is quite large. There are often hundreds, if not, thousands of ways a complex process or system can fail. Furthermore, the means of mitigating these failures are often not known with certainty. For example, changing software to mitigate one failure may result in a new, unintended failure. If the state of knowledge is not com-

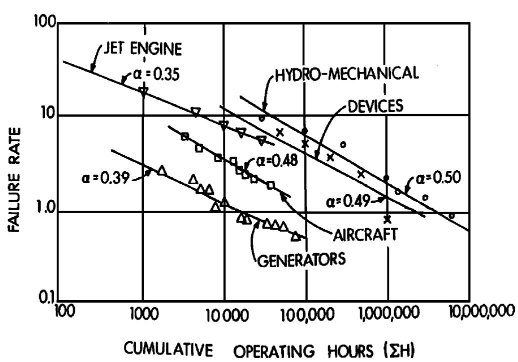

Fig. 1 The original Duane plot. The log failure rate (*Y*-axis) plotted vs. the log cumulative test time (*X*-axis) gives straight lines for different devices. The numbers on the plot are the growth rates (the slope of the lines). Higher growth rates (steeper slopes) mean faster reliability improvement © 1964 IEEE

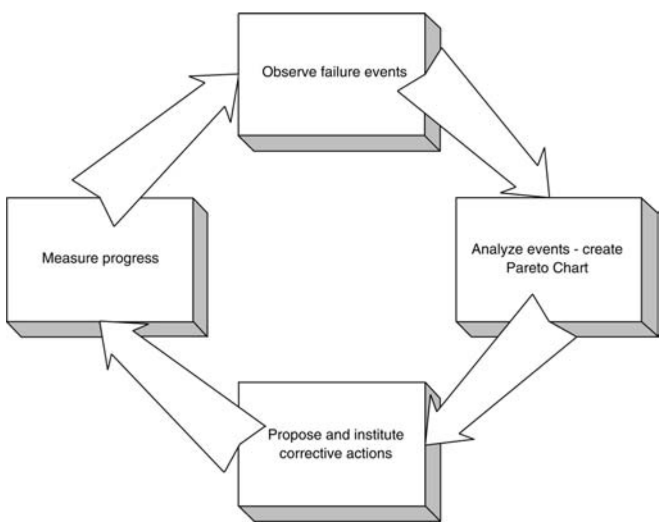

Fig. 2 Reliability growth as a closed loop process

plete, then reliability growth methods are often the fastest way to develop or improve products or processes by testing to expose problems, proposing solutions, and retesting. In this sense, "testing in quality" can be a more productive means of generating and confirming our knowledge. Diagnostic instrument assay systems and clinical laboratories are examples of complex products and processes where the state of knowledge is not complete because of the following:

– Integration of disparate technologies
– Unintended modes of operation, particularly with respect to the operator and patient samples
– Off-the-shelf technologies are often used that are not optimized for the function of the system or process

If one were to insist on designing in quality where the state of knowledge is incomplete, then one would have to fund more research until the state of knowledge was adequate. This option is not commercially viable.

While perhaps not apparent, testing in quality is used in many fields. For example, in the pharmaceutical industry, one proposes a drug, synthesizes it, tests its efficacy in animals and humans and if approved, provides it to the public. Often, the drug is not completely effective and has side effects, hence an improved drug is proposed and the cycle is re-entered.

Reliability growth management – an example

The metric and goal setting

The area chosen for quality improvement was the reliability of assay instrument systems. An existing metric, the unscheduled service call rate was chosen. Shortened to the "call rate", this metric measured the annualized unscheduled service call rate for instrument systems that were under warranty. An advantage of using an existing metric was that it was accepted by management. Additionally, it encompassed most customer quality concerns and issues. Moreover, it had a direct financial value to the company. That is, since each unscheduled service call was expensive, a high call rate reduced profit and potentially reduced customer acceptance of the product.

Setting an appropriate goal was an important element of the program. Overly optimistic goals were often previously proposed by managers. These goals were unsupported by either benchmarking or data analysis and were typically ignored since they were unrealistic. Goals based on benchmarking and data analysis as part of reliability growth management provided incremental improvement targets. These realistic goals did play a role in design strategies.

The model

The model is a key element of the program. Since the objective was to measure and improve reliability before the product was released, a model was needed to estimate what the future reliability would be in the hands of the customer. The conventional wisdom at the time equated "hard" failures with unscheduled service calls. Hard failures were events that a customer could not resolve without a service call such as a failed power supply. However, summing the actual hard failures did not add up to the observed number of unscheduled service calls.

Examination of service call records led us to the notion of "soft" failures as an additional cause for unscheduled service calls. Soft failures were problems that could be resolved by customers, such as clearing a cuvette jam. By conducting data analysis, it was determined that multiple occurrences of the same soft failure often led to an unscheduled service call, even though the system issue was resolved by the customer. Adding (weighted) soft and hard failures accounted for most of the unscheduled service calls. The model's fit to observed data was further improved by adding a factor to account for training issues. These were repeat service calls caused by the service department's inability to fix certain problems in one service visit, especially for a new product.

The equation for the model is:

$$\text{unscheduled service calls} = C \times (A \times \text{hard failures} + B \times \text{soft failures}) \quad (1)$$

where A, B, and C are coefficients optimized by examining actual data.

Data collection

The model required new efforts in data collection. This was especially needed for the soft failures. If an event occurred such as a cuvette jam that was cleared by the customer, it would be impractical for the manufacturer to find out about it by asking the customer. Fortunately, the instrument system had an extensive diagnostic database that captured all events, including soft failures like cuvette jams. Moreover, the system also had a modem, which permitted downloading this data into a central database for analysis.

The framework for data analysis and reports

The framework used for data analysis was a process known as failure reporting analysis and corrective action system (FRACAS). A database was used with several key tables [9, 10].

Table 1 An example of a Pareto analysis report. The percentage contribution to the service call metric is shown for each failure mode. Only the top failure mode contributors are shown

Failure mode	Contribution (%)
Unscheduled recalibration	22
Reagent level sense failure	18
Cuvette jam in incubator	15
Software locked up during sampling	12
Reagent storage thermal error	10

Events. This table contained all failure events, each of which was assigned a failure mode selected from the failure mode table. Each event also had the frequency of the event's occurrence, the classification of the event, and a variety of other demographic data related to the event.

Failure modes. This table contained a list of failure modes and an associated corrective action ID.

Corrective actions. This table contained demographic data for each corrective action.

Classification. This table contained weighting factors according to the severity of an event.

Data analysis

The first step in data analysis was to review the records for the latest time period. Each failure event was assigned to either an existing or new failure mode. The list of failure modes was developed over time as knowledge grew about how the system failed. It was important to determine the possible root cause of each failure. The root cause could be different from the observed symptom. For example, a symptom might be the repeated calibration failure of an analyte, whereas the root cause might be traced to a reagent. If the specific reason were known for the fault in the reagent, this too would become part of the failure mode text. Reclassification of events was possible as more information was acquired.

All failure modes had their severity multiplied by their frequency of occurrence. Pareto analysis over a recent time period was performed which ranked those failure modes highest with the highest contribution to the overall failure rate. A Pareto analysis in the form of a table (an example follows, see Table 1) would be issued regularly to provide management with progress reports.

Corrective action

The corrective action process relied on teams to fix each of the problems at the top of the Pareto table. Since these problems had the highest impact on the failure rate and since there were limited resources, it was imperative to marshal resources effectively. Without this type of Pareto analysis, some other selection method might be used to assign people to work on problems such as fixing problems in the order in which they were first reported or perhaps working on a problem that was interesting. These selection methods were almost always slower in achieving results in reliability improvement.

On a superficial level, an observer would not notice any difference between a traditional reliability program and that of reliability growth management. In each case, the observer would see people working very hard to solve problems.

It was also important to distinguish between a recovery and a corrective action. Recoveries were actions that repaired the system, but without any certainty that the problem would not reoccur. For example, a customer that cleared a cuvette jam would have performed a recovery. A corrective action would be a design change to eliminate the cuvette jam from occurring.

Measuring progress

In addition to the Pareto analysis, a central feature of reliability growth was estimation of the speed of fixing problems. This was performed by graphing the log of the cumulative failure rate (*Y*-axis) vs. the log of the cumulative time that the system was under test (*X*-axis). This graph was often a straight line, whose slope is the growth rate (Fig. 3).

The growth rate is an essential feature of the learning curve phenomenon. Thus, for a system under test, there were a variety of failures occurring which summed to the overall failure rate. During this testing, teams worked on solving the top problems. If the teams fixed problems at a faster rate than new ones occurred, a growth rate would be observed. The failure rate, known as the instantaneous failure, was estimated from the data by Eq. 2.

$$\lambda(t) = 1- \alpha)KT^{-\alpha} \quad (2)$$

where λ is the failure rate at time t (the end of the cumulative test time T), α is the growth rate estimated as the regression slope of the Duane plot, K is the intercept estimated as the regression intercept of the Duane plot, T is the cumulative time that the system or process is under test.

Equation (2) could also be used to predict failure rates for future time periods. This was of key importance to management, who always wanted to know when reliability would reach an acceptable level.

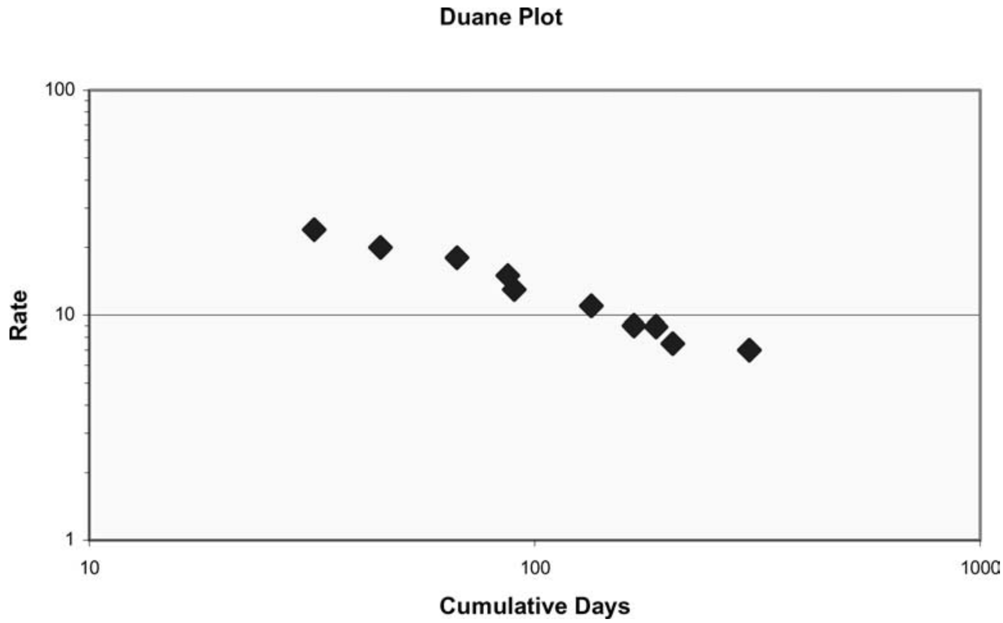

Fig. 3 An example of a Duane plot for a medical device. The log failure rate (*Y*-axis) plotted vs. the log cumulative test time (*X*-axis) gives a straight line.

There was an important difference between predictions based on reliability growth compared to traditional predictions. To see this, it would be helpful to generalize the nature of failures, which could be classified into three general categories:

A. Failures that had been previously observed and that still occurred for which no fix had yet been found
B. Failures for which a fix had been found, but had not been applied to the system
C. New failures, which might be altogether new or caused by fixes from category B

In traditional predictions, managers counted failures in category A, and set to zero failures in categories B and C. That is, all fixes were assumed to be 100% effective without causing any new problems. Moreover, the existence of other new problems was ignored. One factor to cause this optimism was the tremendous pressure by management to speed up the development process. What manager would care to admit that his or her fixes might cause additional problems, or that altogether unknown problems were lurking in the system.

In reliability growth predictions, there was no distinction among categories A–C. The failure rate predictions were data driven, based solely on the demonstrated rate of fixing problems. Putting things another way, reliability growth predictions were based on past data, while traditional predictions were based on assumptions about future events.

Results achieved with reliability growth management

The simplest way to evaluate the results of the reliability growth program was to compare the achieved reliability of programs that used reliability growth with those that didn't. This was facilitated by an assumption that the failure rate of a system was proportional to its complexity. Complexity, in turn was approximated by standard cost. Figure 4 shows the reliability results for the programs that used or did not use reliability growth management.

The improvement in reliability could be equated to profitability since each unscheduled service call cost about $1000. If there were 1000 units in the field under warranty, each lowering of the unscheduled service call rate by one saved the company one million dollars.

The key benefit of using reliability growth management over traditional programs was simple – it led to the most rapid improvement in reliability in the shortest development time. This translated directly into improved profitability. Yet the reasons for this improvement are important to understand as they have implications beyond programs in improving instrument systems.

1. By developing a metric and model that counted all failures – coupled with a process to classify and rank problems, the limited resources available could be used most efficiently to work only on problems most affecting the speed of reliability improvement.
2. By using a data driven prediction process, management had an accurate forecast of when the reliability goals would be reached. This allowed management to plan for contingencies.

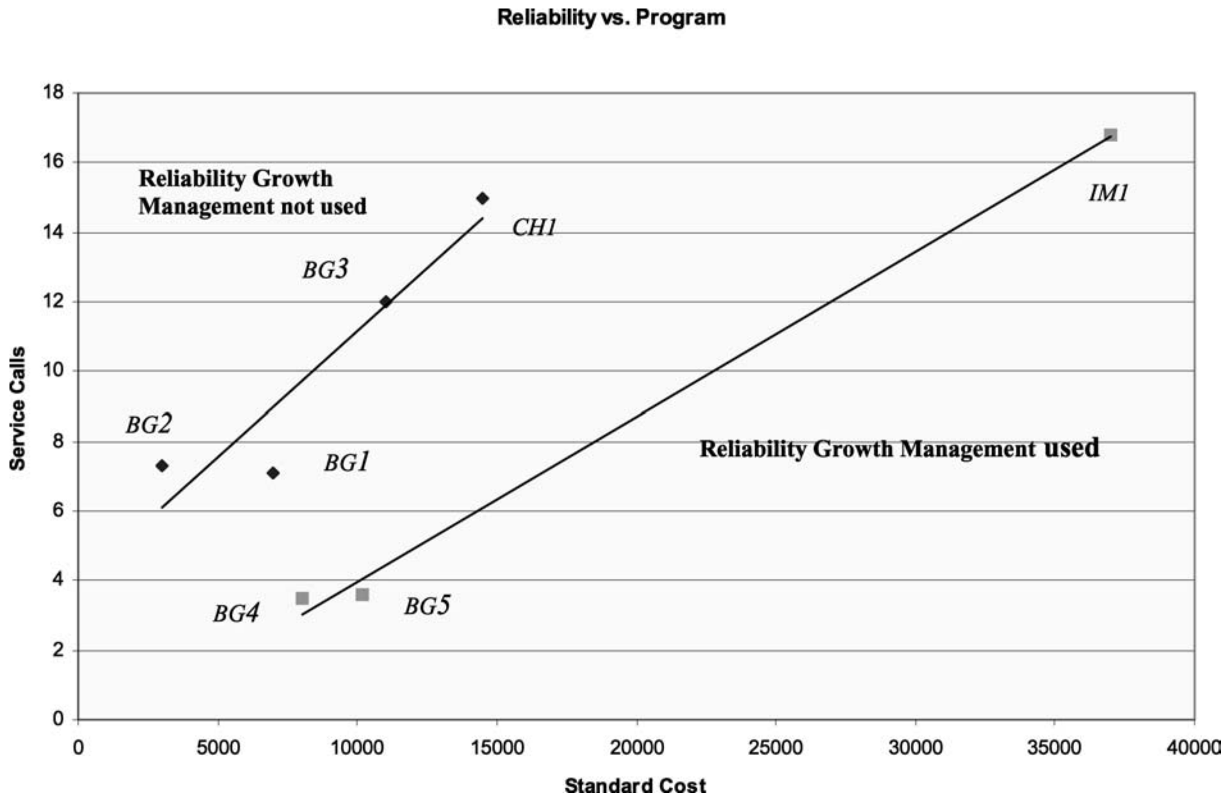

Fig. 4 Reliability achieved for different systems as a function of whether reliability growth management was used. The service call rate (*Y*-axis) was plotted against the standard cost (*X*-axis), with the standard cost acting as a surrogate for complexity. Analyzer abbreviations are *BG*, blood gas; *CH*, chemistry; *IM*, immunoassay

Applying learning curve theory to the reduction of lab errors

One can envision lab errors as a variety of external failures observed by the lab's customers. Examples of these failures include:

– Incorrect assay results that may or may not lead to clinician complaints, patient injury or even death
– Test orders that are filled incorrectly
– Test results that are late
– Lab reports that are not understood

A recent study of lab errors [4] compared error rates measured in three hospitals with CAP Q probe data. In the study there were 14 error categories. The highest hospital error rate for a category was estimated at over 100,000 errors per million tests. The Q probe data had similar rates.

To apply learning curve theory to the lab testing process requires establishing a metric and its desired target value. One could envision incidents per million tests as such a metric. One would also have to create weighting factors for failure mode types so that all failures would contribute to the metric, weighted by their severity. Clearly, a wrong assay result would have the highest weighting. Actual weights should be based on reviewing existing failures.

As much of the failure information as possible should be obtained automatically (e.g., electronically), since it is unlikely that resources will be otherwise available to collect this data manually. Electronic data collection could be used to assess incorrect results by analyzing quality control data, proficiency test results, delta checks, multivariate profiles, and patient means. The laboratory information system (LIS) or hospital information system (HIS) could be used to assess test order fulfillment errors and turn-around-time events. Clinician and other user complaints would also help clarify errors.

With data collection and classification, one could then classify errors as to their severity and frequency and perform a Pareto analysis. A Duane graph would be prepared to track incident rates vs. test volume so that one could calculate a growth rate and be able to predict when improvement goals would be reached.

Learning curve theory vs. ISO 9000 and Six Sigma type quality programs

ISO 9000 provides a quality framework for all aspects of an organization (e.g., including purchasing and personnel, in addition to functions more traditionally thought of as quality related, such as manufacturing). The trend towards covering all aspects of how an organization achieves quality is appealing, as no area within the organization is overlooked. Yet, simply covering all of the bases can provide a false sense of security. How do we know that the processes and procedures are optimal? ISO 9000 does not prescribe standards for these functions, other than to require that something exist in each functional area and the requirement that there is documentation to prove that an organization "does what it says it does."

Another popular quality initiative is Six Sigma. This program has its roots in the notion that the frequency of errors can be measured using a standard deviation. The idea is to reduce the probability of these errors to a level of 3.4 failures (or less) in a million opportunities, or equivalently, six standard deviations (6 sigma). Many clinical processes are 3 sigma or 4 sigma processes. Six Sigma is a process for identifying sources of variation, and methodically working towards a system or process design that meets the Six Sigma criteria. Again, the reliability growth methodology does not conflict with Six Sigma. Reliability growth tools were designed specifically to improve the reliability of complex systems and includes a set of tools tailored to this end (e.g. FRACAS, Duane model, Pareto analysis). In the end, the real challenge is how quickly an organization can acquire the knowledge that will allow them to reduce failures.

Conclusions

Given the large number of problems reported in Nevalainen's study [4] and the limited resources available, it is difficult to imagine how problems in so many areas could be addressed using existing quality improvement programs without a large commitment of resources. One can imagine that if the problems were easy to solve, they would have already been fixed.

The learning curve theory approach described here is a specific tool that can be applied to improve quality by concentrating the limited resources on the most important problems.

Acknowledgements The majority of the work of transferring reliability growth management from the defense industry to the medical device industry was performed by Keith K. McLain, while he was at Ciba Corning Diagnostics. Keith is now at Ortho-Clinical Diagnostics (email: KMcLain@ocdus.jnj.com).

References

1. Leape LL (1994) J Am Med Assoc 272:1851
2. Committee on Quality of Health Care in America, Institute of Medicine (2001) Crossing the quality chasm. A new health system for the 21st century, National Academy Press, Washington DC. Available on line at http://www.nap.edu/books/0309072808/html/, accessed Oct. 25, 2001
3. Sainato D (2001) Clin Chem News 27:1
4. Nevalainen D, Berte L, Kraft C, Leigh E, Picaso L, Morgan T (2000) Arch Pathol Lab Med 124:516
5. CAP website: http://www.cap.org/html/lip/benchmarks/
6. Duane JT (1964) IEEE Trans Aerospace 2:563
7. US Army Communications Research and Development Command (1981) Mil-Hdbk 189 Reliability growth management. US Army Communications Research and Development Command, Fort Monmouth, NJ
8. KcLain KK (1994) Medical Device Diagnostic Industry 166
9. US Army Communications Research and Development Command (1995) MIL-STD-2155 Failure reporting, analysis and corrective action system (FRACAS). US Army Communications Research and Development Command, Fort Monmouth, NJ
10. Villacourt M, Govil P (1994) http://www.sematech.org/public/docubase/abstracts/2332agen.htm