Quality Management in Health Care

Experimental design in health care

Colletta H. (K.K.) Moore

Design of experiments (DOE) is a collection of statistically-based methods for testing multiple process improvement ideas after a quality improvement team has already made initial improvements to remove defects and stabilize the process. Although experimental design techniques are not new, their use in improving administrative processes has not been fully exploited. The power of this tool is illustrated through an actual emergency room case study. Anderson Area Medical Center located in Anderson, South Carolina used traditional quality improvement methods and DOE to significantly decrease patient dissatisfaction from an average of .27 to .06.

To be published in the December 1993 issue of *Quality Management in Health Care*

A 250-bed hospital in the Northwest was concerned about the turnaround time for radiology reports, which averaged about 48 hours. A team was formed to discover ways of reducing this average. Over several months, the team thoroughly and correctly applied several familiar techniques such as flow-charting the process, brainstorming causes of system failure, organizing these causes on a fishbone chart, and prioritizing them on a Pareto chart. Nine months into the project, changes were implemented and positive results were obtained. Turnaround time was reduced from an average of 48 hours to 22 hours. While the improvement was considerable, it was not enough. Certainly, the physicians in this hospital did not think that an average of 22 hours to turn around a radiology report was acceptable. In this case, the team could simply go through another improvement cycle by either doing another project with the same team or starting over with fresh team members; but, substantial further gains will probably not be realized since the same techniques would be used. More often, however, the team will terminate its quality improvement efforts, rejoicing in its success even though the process still doesn't meet customer requirements. A significant amount of time and resources are spent on the initial success, yet the alignment of the process with needs remains elusive.

With the current health care crisis facing the United States, most health care organizations recognize the need for quality improvement. However, many of the methods by which processes can be improved are overlooked and still not understood.

Colletta H. (K.K.) Moore, *M.S., is a consultant specializing in the health care industry with Qual-Pro, Inc. in Knoxville, Tennessee. She has authored manual texts for many of the QualPro seminars, including* Basic Statistical Quality Improvement Techniques for the Health Care Industry *and* Experimental Design Techniques for Service Processes.

The author gives special thanks to Bill Manson, Bob McMillan, Nicole Narciso, and Paul Plsek for their help in preparing this paper.

Fisher's work focused on process innovation after stability is achieved. Shewhart developed control chart techniques, which are used to determine whether a process is stable. Stability means that successive measurements are representative of the same statistical distribution. Limits are placed on the chart such that any measurement outside the limits indicates that the distribution has changed. By finding why the distribution changed and eliminating the cause, the variation of results can be reduced and quality improved.

Fisher's work is complementary to Shewhart's work and fills a distinct need in process improvement. Sometimes a process is not good enough even if stability is attained, and all points on the chart are within limits. Then the reasonable approach would be to change the process. This is when Fisher's work is applicable.

Shewhart was concerned primarily with manufacturing applications. His methods are described in the monumental book, *The Economic Control of Quality of Manufactured Product*, published in 1931.[3] Since that time, the techniques of Shewhart have been refined, extended, and applied in many industries. In contrast, much of the early work on DOE concerned agricultural applications, guided by Fisher and others at the Rothamstead Experiment Station, an agricultural research facility in England.[4] The variability of field trials necessitated a disciplined respect for the uncertainties of experimental data. The experiments are run in the "noisy" and imperfectly controlled conditions of the real world.[5] Two of Fisher's books, *Statistics for Research Workers* and *The Design of Experiments*, have been through many editions and are regarded as classics in this field.[6,7] Other significant contributors to experimental design theory and practice include F. Yates, G. Box, R. Plackett, J. Burman, D. Finney and W. Cochran.[8-16] In the United States, much of experimental design application was also done at schools of agriculture, which is a major reason for the United States' dramatic increases in crop yields since the 1920s. Its use in manufacturing processes has been just as dramatic. For example,

- COBE Cardiovascular, Inc., a manufacturer of specialty medical products, produces a membrane lung used during open heart and other complicated surgeries. Data indicated that 5 percent to 15 percent of the total membrane lung production was leaking. Experimental design techniques helped reduce leakers to only 1.8 percent of total production, saving approximately $250,000 per year in scrap reduction, but most importantly, significantly reducing the risk of field failures and customer complaints.

- Beaulieu Carpets of Dalton, Georgia used DOE to virtually eliminate a color streaking problem called banding in several of its carpet styles. Seconds were reduced from 53% to 0%, which resulted in $200,000 in direct labor savings and the reassignment of inspection and repair personnel to more productive activity.

- Boise Cascade Corporation's Southern Operations in DeRidder, Louisiana, used experimental design techniques to reduce raw material costs by more than $2 million per year after learning that more expensive wood did not make better paper. These landmark findings led to the team being awarded the 1993 RIT/*USA Today* Quality Cup for manufacturing improvements.[17]

Like the rest of quality improvement theory first used in industry, experimental design technology, too, needs to be brought into health care. Although the medical literature is filled with hundreds of articles describing "experiments," the concept of experimental design presented in this paper differs. Typical experiments in medical care involve, say, a proposed new drug or treatment. Unknown to the patients, the researchers separate them into two groups, experimental and control. One group gets the new drug or treatment; the other does not. After tracking the health of the two groups over an appropriate, often long, period of time, the researchers test to see if there is a statistically significant difference between the two groups. In contrast, the experimental design techniques discussed here involve the testing of multiple ideas simultaneously with minimal experimentation costs. It is these designs that are relatively unknown in the health care industry. The mathematical complexity of most texts process improvement consultant. Members included the Director of ER, the Nurse Manager/Trauma Coordinator, the Outpatient/ER Registration Coordinator, the Manager of Admitting, a staff nurse from day shift, a staff nurse from evening shift, a surgical technician, and a management engineer. The team used the following procedure for experimental design:

- Determine that meaningful experimentation is possible
- Generate suggestions for system changes
- Design and run the experiment
- Analyze the results and verify conclusions
- Make decisions and/or process changes; plan further experimentation

Notice that this general procedure parallels the familiar Shewhart cycle of Plan-Do-Check-Act (PDCA).[18]

Determine That Meaningful Experimentation Is Possible

While there are advanced statistical methods which allow one to experiment meaningfully with an unstable process, commonly used experimental design techniques are built on the assumption of process stability. This means that a team must first work to remove special causes of variation (i.e., achieve stability on a control chart). Without stability, the special causes will increase the experimental error and make it difficult to obtain useful results. Once the process is stable, experimental design is used to shift and/or reduce common cause variation. In other words, DOE is used to change the system and/or improve consistency.

The response, or outcome of interest, for the ER team was patient dissatisfaction. Therefore, the team began a measurement system before the planned experiment which assessed patient dissatisfaction for discharged ER patients. Each day, 40 randomly selected patients were surveyed by telephone. The p chart in Figure 1 illustrates the proportion of dissatisfied patients each day. The first significant shift resulted from an overall awareness of the project. Pursuing continuous improvement by using traditional QI techniques, the team organized survey comments on a Pareto chart. The single most common complaint was waiting time. As a result, the team began measuring actual wait time for a random sample of patients each day. However, a scatter plot revealed that *actual* wait times were not related to overall dissatisfaction. Thus, even if the team had continued to spend energy reducing wait times, one would not see any effect on the ultimate key measure, patient dissatisfaction. The pa-

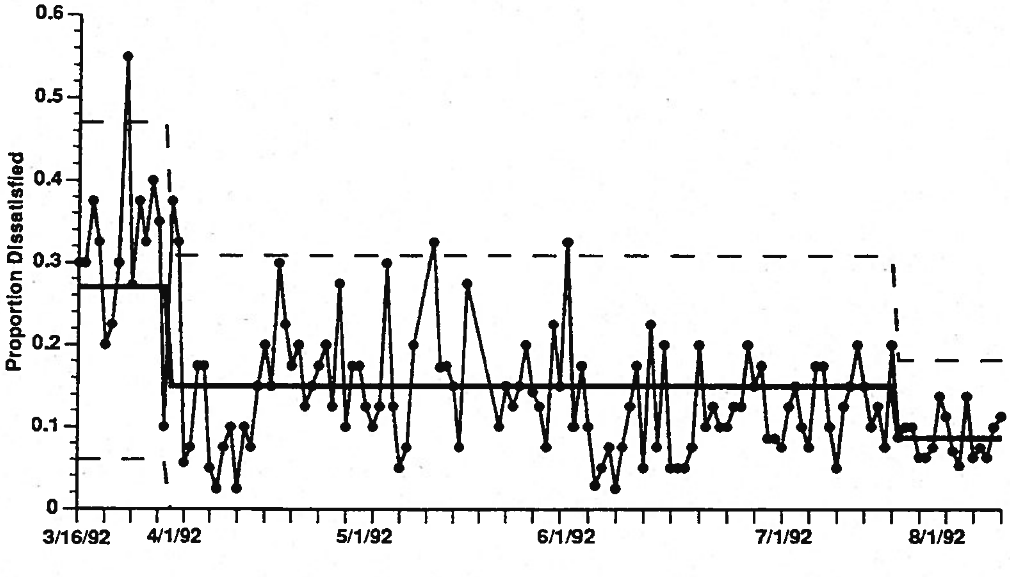

Figure 1. Using traditional QI methods, average proportion of patient dissatisfaction was reduced from 27% to 8%.

Factor		**Level**
A: Fast Track	–	*Current condition*
	+	Non-emergent patients will receive service in a separate, clinic area
B: Internal Operations	–	*Current condition*
	+	Follow-up call next day at home to patient by R.N.; communicate to patient at time of discharge that an attempt will be made to see the original doctor on follow-up visit; send stable patients upstairs to Radiology for quicker x-ray response
C: Triage Area Changes	–	*Current condition*
	+	Handout pamphlet explaining ER process and triage; place a solid partition between the two triage desks to allow for more privacy; hand out a "welcome" card with names of triage nurse, charge nurse, and nurse manager as well as who to call if there are any problems
D: Outside Waiting Room	–	*Current condition*
	+	Student volunteer to staff ER waiting area to help families and maintain coffee pot; place television in waiting room
E: Pediatric Changes	–	*Current condition*
	+	Give coloring books and crayons to children; play children's movies in pediatric treatment room; apply large Disney characters on wall

Table 1

Design and Run the Experiment

Once the factors and levels are determined, the combinations of factors varied in the initial experiment need to be selected. This initial experiment is usually run just to narrow the list of factors in preparation for subsequent, more focused experiments. This is comparable to the motivation used in Pareto analysis, a traditional quality improvement method. The initial designs that are used, called screening designs, are used to provide an efficient way to evaluate a large number of factors and select from them a smaller set of important factors for subsequent experimentation.[19] One type of screening design was introduced by Plackett and Burman (1946).[20]

Plackett-Burman designs have several properties. The first property relates to the number of *runs* in the experiment. A *run* is a determination of response for a particular combination of factor levels. The choice of Plackett-Burman experiments depends on the number of factors tested. Plackett-Burman designs can hold up to n-1 factors in n different runs, where n is a multiple of four. For example, if there are 4-7 factors one can use an 8-run design, if there are 8-11 factors one can use a 12-run design, etc. The ER team used an 8-run design since it experimented with 5 factors. A second property of a Plackett-Burman design is that all *main factor effects*, or the changes in the response that occurs when a factor is changed from one level to another, can be calculated in such a way that they are unaffected by play children's movies, and apply large Disney characters on the wall for cosmetic purposes (E+). The other treatment combinations are decoded similarly.

Now the team had to decide the order in which to run each of the eight different treatment combinations. As mentioned previously, randomization is crucial. However, in this case, the staff did not feel that factor A, fast track, could be turned off and on weekly. Therefore, restricted randomization was present. Fast track was not offered for the first four weeks of the experiment (runs 5, 3, 2, 8), but then was offered for the last four weeks of the experiment (runs 4, 6, 1, 7).

The ER team had to deal with other logistics before the experiment began. Assignments were given to specific team members to coordinate the events associated with each factor change. For example, one team member was responsible for printing the pamphlets while a different team member was responsible for obtaining the children's movies and VCR for the pediatric treatment room. Additionally, experimentation costs had to be considered. With the exception of fast track, costs associated with making the factor changes were negligible. Staffing costs for fast track were between $5,800-$6,700/week. A schedule of weekly changes was published and posted in the staff lounge. Furthermore, weekly ER staff meetings served as just-in-time inservices on factor changes for each specific week. All changes began on a Monday, with the nurse manager overseeing the changes. While the changes were considered somewhat disruptive, timely communication of information was crucial to obtaining staff buy-in.

Another consideration of the ER team involved the ability to assess a Hawthorne effect. With so much attention being paid to patient satisfaction in the ER, perhaps there could be an upward trend that makes runs in later weeks appear to have better results. This is another reason why it is important to stabilize the process before experimenting. Furthermore, we were able to assess process stability during the experimental course by maintaining the same control chart. Incidentally, one special cause was detected during the experiment, but it could not be explained. The data from that day were not used in the analysis.

Each week, the various factor changes were made while the same measurement system was continued. That is, the identical survey instrument was used throughout the DOE as was used to make the initial improvement using traditional QI methods. For each run, the proportion of dissatisfied patients in each daily sample of 40 discharged patients was recorded. We then calculated the average and range for each week. The combined results are provided in Table 3. Y-bar represents the average proportion of dissatisfied patients for each treatment combination, or week. R represents the range of the 7 daily fractions of dissatisfied patients for each treatment combination. The averages are used to estimate the effect of each factor while the range is used to assess the variability of the response at a particular test condition. It is this latter information that is used to determine the experimental error which is needed to test the statistical significance of each of the effects.

Run	A	B	C	D	E	F	G	$\bar{y}$	R
1	+	+	+	–	+	–	–	.032	.075
2	–	+	+	+	–	+	–	.054	.075
3	–	–	+	+	+	–	+	.075	.100
4	+	–	–	+	+	+	–	.039	.100
5	–	+	–	–	+	+	+	.071	.150
6	+	–	+	–	–	+	+	.039	.075
7	+	+	–	+	–	–	+	.075	.150
8	–	–	–	–	–	–	–	.050	.100
	–EG	–DE	–EF	–FG	–CF	–DG	–DF		
	–CD	–CG	–BG	–BE	–BD	–CE	–BC		
	–BF	–AF	–AD	–AC	–AG	–AB	–AE		

Table 3

Analyze the Results and Verify Conclusions

The estimated effect of fast track (A) on patient dissatisfaction can be found by calculating the average proportion of dissatisfied patients when fast track was offered (A+) minus the average proportion of dissatisfied patients when fast track was not offered (A–). Notice that A+ occurred in runs 1, 4, 6, and 7 and A– occurred in runs 2, 3, 5, and 8. The numbers needed in

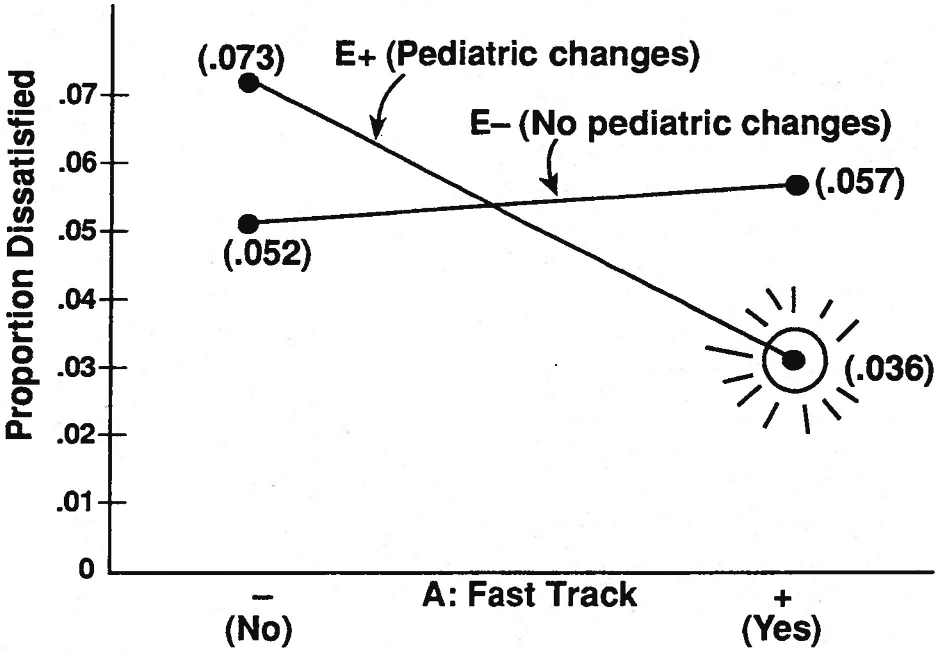

Figure 3. The combination of fast track and pediatric changes significantly reduced patient dissatisfaction.

focusing on only factors A and E, there are only four possible treatment combinations. These four different conditions occurred twice during the 8-run Plackett-Burman experiment. For example, referring to Table 3, Runs 2 and 8 involved both of these factors being "off," which resulted in the average proportion of dissatisfied patients of .054 and .050, respectively. Therefore, the average proportion of dissatisfied patients during this specific condition was .052. The results can be reorganized as shown below:

Runs	A	E	$\bar{y}_1$	$\bar{y}_2$	$\bar{\bar{y}}$
2, 8	–	–	.054	.050	**.052**
6, 7	+	–	.039	.075	**.057**
3, 5	–	+	.075	.071	**.073**
1, 4	+	+	.032	.039	**.036**

Figure 3 helps illustrate the AE interaction. If an interaction is not significant, the two lines would appear parallel, indicating that the change produced by each factor is independent of the level of the other. Although this graphical illustration provides a crude estimate of the AE interaction, it works well in practice. As shown, the combination of fast track and the pediatric changes significantly reduced patient dissatisfaction. The team's process-oriented explanation of this interaction was the increased satisfaction with the adult patient population as a direct result of increased satisfaction with the pediatric patient population. Many of the fast track patients were children; therefore, the parents were pleased with the special service. If the pediatric patients did not use fast track, then at least they were provided service under more pleasant conditions in the pediatric treatment room. Again, the parents were pleased. The combination of these services effectively filtered the patient population base into two natural groups so that the staff could better service each one individually.

LESSONS LEARNED

The ER case study resulted in the following *lessons learned*:

- **"DOE is not exclusively a statistician's tool."**
  The managers and staff at Anderson Area Medical Center played a very active role in the application of DOE in the ER. They successfully followed appropriate methodology, and they managed the DOE by coordinating the weekly changes and seeing that the design was adhered to throughout the 8-week period. This experience resulted in the transfer of skills to the managers of this area who are now able to execute an experiment successfully. Even the quantitative skills necessary for analyzing and interpreting the results have been transferred, albeit to a smaller group. Certainly, Shewhart's vision is alive at Anderson Area Medical Center:

  > *"The long-range contribution of statistics depends not so much upon getting a lot of highly trained statisticians into industry as it does on creating a statistically minded generation of . . . others who will in any way have a hand in developing and directing processes of tomorrow."* [22]
  > *- Walter A. Shewhart*

  The managers at Anderson Area Medical Center recognize that DOE is a *manager's* tool for promoting innovation in the organization. In addition, the use of experimental design also taps the knowledge of the local work force by seeking their input for DOE factors. Therefore, all managers should understand the use-potential of the technique.

- **"We can apply DOE in health care processes."**
  After applying DOE to the ER process, the diversity of the applications are evident. For example, one might want to reduce bed sores by experimenting with different types of beds, different types of dressings, etc. Other experimentation could include the use of different communication devices, different types of information provided, etc. to reduce patient transportation time. In a clinical sense, one could experiment with different types of medication, doses, timings of medication, etc. to improve clinical outcomes.[23] Many opportunities exist for efficient evaluation of several factors simultaneously in such studies. Note, however, that many health care processes are not yet in statistical control and are not, therefore, immediate candidates for DOE work. Understanding of the prerequisites is important.

- **"DOE should play an integral role in our typical CQI efforts."**
  Recognizing that stability is a prerequisite to experimental design, traditional QI techniques should be used initially. However, these techniques are often not sufficient in achieving true breakthroughs in performance. Continuously improving processes to meet customer needs and reducing variation should be the desired outcomes, for these lead to improved competitive position. Process innovation and design changes are required. Therefore, DOE should be built into the overall strategy for any process improvement effort.

- **"To understand the effect of any process change, you do not have to experiment with only one factor at a time."**
  Experimental design is based on proven theory which refutes the OFAT myth. Not only is OFAT inefficient, but it also could lead to the wrong conclusions. Even though multifactor experiments such as Plackett-Burman designs do not allow explicit estimation of most two-factor interactions, there is often a warning of interaction effects which would be missed using OFAT. In the ER case study, OFAT would have resulted in the erroneous conclusion that patient dissatisfaction could not be significantly reduced with the five factors that were tested. It was the presence of the interaction of two factors which held the key to the breakthrough.

SUGGESTED READINGS

Box, G. E. P. and Bisgaard, S. "The Scientific Context of Quality Improvement." *Quality Progress*, June, (1987): 54-61.

Box, G. E. P., Hunter, W. G. and Hunter, J. S. *Statistics for Experimenters*. New York: John Wiley & Sons, 1978.

Box, Joan Fisher. R.A. *Fisher: The Life of a Scientist*. New York: John Wiley & Sons, 1978.

Cochran, William G. and Cox, Gertrude M. *Experimental Designs*. New York: John Wiley & Sons, 1950.