Q Manage Health Care
Vol. 14, No. 4, pp. 206–216

The One-Person Randomized Controlled Trial

Jesper Olsson, PhD; Darcey Terris, MBA; Matthias Elg, PhD; Jonas Lundberg, MSc; Staffan Lindblad, MD, PhD

Currently, the gold standard for collection of clinical evidence is the randomized controlled trial (RCT), preferably with large, multicenter samples of subjects. Although this approach provides valuable information, many clinicians find it difficult to translate RCT results to the individual patient level. In this report, a statistical approach called Design of Experiments (DOE) is described as a method of applying the principles of RCTs one person at a time. An overview of the method, with a simple clinical example, is presented. As shown, DOE is a more efficient method than the sequential approach often taken by clinicians and their patients when evaluating various treatment choices. Further, the effect of multiple interventions can be assessed, alone or in combination with each other. In this way, DOE can be an important addition to the field of evidence-based medicine, although further studies are needed.

Key words: *evidence-based medicine, randomized controlled trials, research design*

Currently the prevailing design for developing evidence in medicine is the randomized controlled trial (RCT). RCTs are characterized by strict adherence to inclusion and exclusion criteria and the randomization of study subjects to avoid bias and confounding in the observation of an intervention's effect. In general, results from multicenter trials with large samples of subjects are given greater weight as evidence in constructing guidelines for clinical decision making. The remaining question, however, is outside the trial, "To whom do the results apply?"[1] Individual variation in patients' health states and their responses to interventions requires clinicians, in daily practice, to extrapolate. Although careful control of the experimental process is a key attribute of RCTs, it often results in significant reductions in the external validity of the study's results. This lack of external validity has been suggested as one explanation for the observed underuse of clinical guidelines.[1]

From the Swedish Association of Local Authorities and Regions, Stockholm, Sweden, affiliated with Medical Management Centre, LIME, Karolinska Institutet, Stockholm, Sweden, and Quality Sciences, Chalmers University of Technology, Gothenburg, Sweden (Dr Olsson); the Department of Tropical Hygiene and Public Health, University of Heidelberg, Heidelberg, Germany, and the Division of Health Services Research, Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, Ohio (Ms Terris); the Department of Quality Technology and Management, Linköping University, Linköping, Sweden (Dr Elg); the Centre for Medical Innovations and Medical Management Centre, LIME, Karolinska Institutet, Stockholm, Sweden (Mr Lundberg); and the Medical Management Centre, LIME, Karolinska Institutet, Stockholm, Sweden (Dr Lindblad).

Corresponding author: Jesper Olsson, PhD, The Swedish Association of Local Authorities and Regions, Hornsgatan 20, S-118 82 Stockholm, Sweden (e-mail: jesper.olsson@skl.se).

We thank Tom Nolan, Mats Brommels and John Bridges for their valuable ideas.

Traditional methods of constructing evidence-based clinical guidelines, however, should not be abandoned. Systematic collection and dissemination of evolving medical knowledge is a critical starting point for safe and effective medical intervention at the individual patient level. What is needed, however, are tools that can be readily implemented by a clinician to facilitate translation of general research findings to the unique patients who present each day in their office. The aim of this article is to suggest one such tool, a statistical method referred to as Design of Experiments (also called DOE, DOX, or factorial design), that has been used for several decades in industry to efficiently determine which factors are critical to a process and how these factors might be adjusted to optimize the process under consideration. An introduction to DOE is provided, with a simple clinical example, to demonstrate the power of the method to bridge the gap between population-level observations and context-specific applications.

PRIOR USE OF DOE IN THE CLINICAL MANAGEMENT OF INDIVIDUAL PATIENTS

To identify previous examples of the use of DOE in the clinical management of individual patients, a literature search and review was performed. As a first step, PubMed (www.pubmed.org) was used to identify articles containing any of the following phrases—"design of experiments," "factorial experiment," or "factorial design"—in their title or abstract. This search yielded a total of 2680 articles. RefViz software (Thomson ISI Researchsoft) was then used to identify subtopics and organize related articles into clusters. A dummy entry was included for the authors' current article. In this way, related articles could be selected by reviewing the cluster that contained the authors' current work.

In the end, 2643 of the 2680 references (99%) fell into 51 clusters. The authors' current article was included in a group of 218 references that were associated with the key words "patient," "therapy," and "intervention." All articles from all groups that included the key word "patient" were then combined with this initial group, yielding 414 articles from 20 groups. Duplicates were removed, resulting in a final list of 400 articles related to the potential use of DOE in the clinical management of individual patients.

A second method was subsequently used to identify additional related articles. Here, we identified publications that referenced an article by McKinlay and colleagues, where a version of DOE was used in a health care setting.[2] In this manner, a total of 38 additional publications were identified.

Through review of the 438 identified articles, only 3 appeared to address the use of DOE in a health care delivery setting. The perspectives of the articles, however, were focused on optimization of the health care production process. In one article, for example, DOE was used to understand what nonmedical factors impact medical decision making among physicians. Among the 438 articles reviewed, no article addressed the use of DOE for the clinical management of individual patients.

But clinicians are often faced with individual patients for whom the optimal process of care is not clear. Several treatment options may be available, which alone, or in conjunction with others, may provide the best outcome for this individual. The physician is faced with trying to understand patient-level processes in order to achieve the best possible outcome and DOE is a well-known and accepted method for optimizing processes, no matter the nature or context of those processes. From the literature search and review, the use of DOE for the clinical management of patients appears to be a novel approach to bridging the gap between knowledge gained through large, randomized, controlled studies and the needs of individual patients.

ADVANTAGES TO USING DOE FOR THE CLINICAL MANAGEMENT OF INDIVIDUAL PATIENTS

The set of statistical methods referred to as "Design of Experiments" provides a systematic approach to understanding and optimizing processes in which several factors may influence the observed outcome and in which the various factors may interact with

each other. Most importantly, DOE provides theory-grounded justification for shortcutting the trial-and-error method of investigation, significantly reducing the number of iterations required to understand and/or optimize a patient's course of treatment. The foundations of DOE arose in the 1920s from the work of Sir Ronald Fisher, a British statistician and geneticist, working on the effects of nutrition and soil type on plant fertility.[3] Today, DOE is used in a wide range of industries to address problem areas and to increase productivity and efficiency.

In clinical care, the traditional approach taken is one of trial and error—the patient is asked to sequentially moderate one aspect of his or her behavior or treatment at a time and then report his or her observations to the clinician. This approach can be time-consuming, especially if there are multiple factors being investigated. The patient may lose interest and/or diligence in carrying out a lengthy series of iterations. The resulting data may provide little insight into how the final course of treatment should optimally proceed.

In DOE, a series of carefully planned tests is performed, with controlled changes made in each test to the input variables that have been identified as potentially critical to the process under consideration. In contrast to the traditional sequential testing approach described above, input factors are varied simultaneously. Through analysis of the outcomes under various input scenarios, the input variables that are critical to a process outcome and the optimal settings for the input variables can be identified. Depending on the DOE method used, both the impact of individual factors (main effects—as defined below) and the impact of multiple factors together (interactions—as defined below) can be identified. For patients, both of these features are important. In the traditional sequential testing approach, interaction patterns between input factors typically remain undetected.

In this report, DOE is proposed as a clinical tool to identify factors critical to the management of individual patients. Our focus is on a simple DOE model that can be easily implemented and analyzed—requiring nothing more sophisticated than a calculator. Step-by-step instructions are provided, with reference to an actual DOE experiment employed by one of the authors (S.L.) to identify which, among 4 behavioral changes, could be implemented to improve the quantity and quality of his sleep. For additional background on the use of DOE, we recommend Del Vecchio[4] or Moen et al.[5]

COMMON TERMS USED IN DOE

Before we begin, several definitions are required.[4]

A *factor* is an event or phenomenon that influences the observed outcome for a process under consideration. Factors are also called independent or control variables.

A factor's *main effect* refers to the effect of each factor on the observed outcome *without regard to the other factors under consideration*.

An *interaction* is said to be present when the observed effect of more than 1 factor *acting together* is more (or less) than the sum of the main effects of the factors.

A *full-factorial design* is a set of experiments that includes every combination of factor levels for all of the factors under consideration.

A *fractional-factorial design* is a systematically chosen subset of the full-factorial design experiments. The experiments are typically chosen so that there is balance between the factors and factor levels represented, but not all possible combinations of factor levels are included. This type of design reduces the number of experiments required, but also the amount of information possible from the analysis. Information about higher order interactions (interactions between more than 2 factors) is lost, but the main effects and interactions between pairs of factors can be identified.

A *half-factorial design* is the most basic fractional-factorial design and is the design illustrated in our example. Only one half of the full-factorial design experiments are included.

A *run* refers to a test of a specific combination of factor levels used within a DOE experiment. The number of runs needed for the planned experiment

depends on the number of factors, the factorial design, and how many replications of each run that is made.

AN EXAMPLE OF THE USE OF DOE FOR THE CLINICAL MANAGEMENT OF INDIVIDUAL PATIENTS

To illustrate the use of DOE at the individual patient level, an empirical example is described, with step-by-step instructions. The reader is cautioned, however, that the description is not a "one-size fits all" approach for every complex treatment decision faced by a clinician and his or her patient. It is rather a short and useful introduction to the DOE methodology.

Step 1: Define the problem

As with every experimental process, the first step is to clearly define the problem to be investigated. For our example, the participant had a long-standing problem with insomnia, characterized by sleep of short duration and poor quality. He previously tried short periods of various behavioral changes, but had experimented in an unsystematic manner. As a result, he did not find any factor that appeared to influence the quantity or quality of his sleep. Upon the suggestion of his fellow author, the participant agreed to undergo a systematic investigation of the effects of the various behavioral interventions, based on a simple DOE design.

The problem is thus defined as: Can behavioral changes be made which will improve the quantity and quality of the participant's sleep?

Step 2: Decide how the process outcome will be measured

The accuracy and precision of the conclusions drawn from any experiment are dependent on strict control of the experimental process and careful measurement of the outcome variable(s). This is also true for DOE. For the insomnia example, 2 outcome variables were chosen to encompass both the quantity and quality of sleep received. It is possible to have a larger number of outcome variables if relevant to assessment of the problem, including those that would capture unwanted side effects.

The first outcome measure, the length of sleep received, was measured by direct observation of the time asleep, given by the difference in time from when the subject laid down in bed at night and the time the subject woke up. The times were measured by a luminescent clock at the bedside and confirmed by the spouse.

The second outcome measure, the perceived quality of sleep received, was measured upon wakening through the use of a visual analogue scale (VAS) with 10 cm between the endpoints—"Totally worthless" sleep and "Couldn't have been better" sleep. The result was measured in centimeters from the origin and recorded on a separate form each morning.

Step 3: Decide which factors will be included in the experiments

There are 2 basic approaches to designing a set of DOE experiments. The first is to focus on identification of factors, among a large number of potential candidates, which are critical for the process under investigation. In this type of DOE design, referred to as a "screening" or "saturated" design, only the main factor effects are considered. Hence, further tailoring the process to focus at the individual patient level is often needed.

A saturated DOE is often followed by a DOE aimed at determining the optimal factor levels required to maximize the output of the process under consideration. In this type of design, the interactions between factors are also considered. Regardless the choice of DOE design, the potential critical factors to be incorporated may be identified from clinical guidelines, the literature and patients' preferences. In our example, a DOE design was chosen to investigate both the main effects and interactions among 4 possible interventions (factors) to address the participant's insomnia.

1. Do not eat or drink after 8:00 PM (Food & Drink)
2. Do a 20-minute period of a specific yoga exercise before going to bed (Yoga)
3. Go to bed before 10:00 PM (Bedtime)

4. Stay in bed, asleep or awake, until 6:00 AM (Rise Time)

As explained in the next section, the more factors (and factor levels) included in a series of DOE experiments, the more runs required.

Step 4: Decide how the factors will be varied during the experiments

In DOE, the factors may be varied continuously or in discrete increments and each factor may vary differently. For our insomnia example, we use a commonly used DOE design based on 2 discrete levels for each factor. The 2 different levels for the 4 factors in the example are shown in Table 1.

To ensure reasonable separation between the levels specified for the Food & Drink and Bedtime factors, an hour break was imposed between the high and low factor levels. For example, the high level for Food & Drink was specified as "no food or drink after 8:00 PM." The low level for the same factor was specified as "food and drink after 9:00 PM." By imposing this hour separation (essentially no food or drink between 8:00 PM and 9:00 PM), the participant was prevented from eating at 7:59 one night, qualifying under the high factor level, and then at 8:01 PM the next night, for the low factor level.

For a full-factorial DOE design in which all of the factors have the same number of levels, the following formula can be used to calculate the number of experiments needed:

$$X^k = \text{number of experiments required for a full-factorial design,}$$

where X is the number of factor levels (must be the same number for all factors) and k is the number of factors.

With the 4 factors identified in Table 1, a total of 16 ($=2^4$) experiments are needed for a full-factorial design. If 5 factors had been included, the number of experiments would increase to 32 ($=2^5$) and with 6 factors it would increase to 64 experiments. As shown, the number of experiments needed to conduct a full-factorial design quickly increases when the number of factors or factor levels increases. It is in these cases that fractional designs become useful. For example, the 6-factor design, with 64 experiments required for the full-factorial design, may be reduced to 32 experiments for a half-factorial design.

Table 2 shows the 16 experimental runs that would be included in a full-factorial design for the 4 factors

Table 1

THE FACTORS AND FACTOR LEVELS CHOSEN FOR THE INSOMNIA EXPERIMENTS

Factor	High level	Low level
Food & Drink	No food or drink after 8:00 PM	Food and drink after 9:00 PM
Yoga	20 min of a specific yoga exercise before bedtime	No yoga exercise before bedtime
Bedtime	Go to bed before 10:00 PM	Go to bed after 11:00 PM
Rise Time	Stay in bed, asleep or awake, until 6:00 AM	Get out of bed within 1 h of waking up

Table 2

FULL-FACTORIAL DESIGN FOR THE 4-FACTOR – 2-LEVEL INSOMNIA EXPERIMENTS

Run scenario	Factors			
	Food & Drink	Yoga	Bedtime	Rise Time
1	−	−	−	−
2	+	−	−	−
3	−	+	−	−
4	+	+	−	−
5	−	−	+	−
6	+	−	+	−
7	−	+	+	−
8	+	+	+	−
9	−	−	−	+
10	+	−	−	+
11	−	+	−	+
12	+	+	−	+
13	−	−	+	+
14	+	−	+	+
15	−	+	+	+
16	+	+	+	+

identified in Table 1. The common DOE shorthand is employed of using "+" to denote when the high level (increased level) is used for a given factor in a given run and "−" to denote when the low level is used for a given factor for a given run.

Step 5: Decide whether a full or fractional-factorial design will be used

As previously stated, one of the main advantages of using DOE is that the number of experimental runs can be dramatically decreased by systematically choosing a subset of which experiments to include. Although the complete set of possible information concerning the process under consideration will not be obtained in a fractional-factorial design, much can be learned from a relatively small number of iterations. The cost (ie, loss of information concerning factor interactions) is often outweighed by the benefits (ie, quicker results with less burden on the patient). There are standard guides for constructing different factorial designs depending on the number of factors under investigation and how many experiments one wants to perform. Design templates are given in the recommended textbooks.[4,5]

As illustrated in Table 3, a half-factorial design was chosen for the insomnia example. In this design, the number of experimental runs is reduced from 16 to 8.

Table 3

HALF-FACTORIAL DESIGN FOR THE 4-FACTOR – 2-LEVEL INSOMNIA EXPERIMENTS

Run scenario	Factors			
	Food & Drink	Yoga	Bedtime	Rise Time
1	−	−	−	−
2	+	−	−	+
3	−	+	−	+
4	+	+	−	−
5	−	−	+	+
6	+	−	+	−
7	−	+	+	−
8	+	+	+	+

The results from the reduced number of experiments will allow us to investigate the main effects of each factor and all interactions between any 2 factors. However, the design will not allow the separation of interactions between any set of 3 factors.

Step 6: Decide whether to replicate the design and how to replicate the design

One of the main advantages of DOE is that the number of experimental runs can be reduced without significant loss of information. However, the ultimate goal of DOE is to efficiently obtain high-quality information concerning a process under consideration. As discussed in Step 2, the accuracy and precision of the conclusions drawn from a DOE analysis are dependent on careful measurement of the outcome variable(s). The outcome variables chosen in the insomnia example, however, are both dependent on patient self-reports. This is an example of a risk of error in measurement that could lead to excessive scatter in the data. As in all research, scatter in the data can make it more difficult to identify the differences that result from true factor effects.

Replication, specifically including multiples of each run scenario in the set of experimental runs, provides an insight into the scatter associated with similar runs and makes it easier to identify the differences due to true factor effects.[6] In the end, the expected benefits of including replication in the design must be balanced against the additional burden placed on the patient. If at all practical, replication of experimental runs should be included.

Table 4 lists the experiments performed in the insomnia example. As shown, each experimental run scenario was included twice.

Step 7: Implement randomization

Although systematic choices are made as to which experimental runs are included in a given DOE design, the experimental runs do not need to be ordered, when implemented, in the same pattern with which they were selected. In fact, the exact opposite is true—the order of the experimental runs should be randomized to prevent sequencing effects from clouding the detection of the true factor effects.

Table 4

RANDOMIZED AND REPLICATED HALF-FACTORIAL DESIGN FOR THE 4-FACTOR – 2-LEVEL INSOMNIA EXPERIMENTS

Day	Run scenario	Factors			
		Food & Drink	Yoga	Bedtime	Rise Time
1	3	−	+	−	+
2	3	−	+	−	+
3	7	−	+	+	−
4	7	−	+	+	−
5	8	+	+	+	+
6	8	+	+	+	+
7	4	+	+	−	−
8	4	+	+	−	−
9	1	−	−	−	−
10	1	−	−	−	−
11	5	−	−	+	+
12	5	−	−	+	+
13	2	+	−	−	+
14	2	+	−	−	+
15	6	+	−	+	−
16	6	+	−	+	−

For the insomnia example, a decision was made to randomize the order of experimental run scenarios, but to replicate each run after randomization, as shown in Table 4. Although it would be best, experimentally, to randomize all of the runs including the replications, a compromise was made to simplify the experimental process for the participant. Randomization can be performed through the random number function present on most scientific calculators or by simply drawing from a hat slips labeled with each run scenario.

Step 8: Implement additional experimental controls as needed

Often the best way to reduce the scatter of the data collected for any experimental process is to consider, beforehand, what may go wrong at each step and then implement appropriate controls. Here the responsibility lies with the clinician to instruct his or her patients to rigorously follow the prescribed order of the experimental runs and at the factor levels specified for each. Additional aids/reminders can be provided. In the insomnia example, a diary was constructed with clearly written instructions as to what behaviors should be implemented each night. A sample of a diary entry is shown in Figure 1.

Day 1—the night April 5 to 6
- Food & Drink after 9 PM (none between 8 and 9 PM)
- Do 20 min of yoga exercise
- Go to bed after 11 PM
- Stay in bed, asleep or awake, until 6 AM

Figure 1. An example of a daily diary entry for the insomnia experiments.

Step 9: Run the experiments

With careful instructions, patients are then sent off to run their experiment, record their data, and report back to the clinician. Intermittent checks-ins between the patient and clinician may be helpful. In this way, unanticipated challenges in carrying out the experimental runs can be identified. It may be more efficient to stop a poorly implemented experimental process, retool the design, and then begin again. Otherwise, the subject continues to collect what may be useless data and may become unwilling to participate in such experiments in the future.

As shown in Table 5, the following data were collected for the insomnia experiment.

Step 10: Analyze the data and draw your conclusions

The main effects of each factor can be calculated using the following formula[6]:

$$\text{Factor Effect} = \text{absolute value of } \frac{\Sigma Y_+}{n_+} - \frac{\Sigma Y_-}{n_-}$$

where ΣY_+ is the sum of all of the outcomes observed when the given factor was at the + level, ΣY_- is the sum of all of the outcomes observed when the given factor was at the − level, n_+ is the number of times when the given factor was at the + level, and n_- is the number of times when the given factor was at the − level.

Table 5

DATA COLLECTED FOR THE 4-FACTOR – 2-LEVEL INSOMNIA EXPERIMENTS

Day	Run scenario	Food & Drink	Yoga	Bedtime	Rise Time	VAS score	Hours slept
1	3	−	+	−	+	4.0	4.0
2	3	−	+	−	+	5.4	5.5
3	7	−	+	+	−	6.2	6.5
4	7	−	+	+	−	8.4	7.0
5	8	+	+	+	+	7.0	7.0
6	8	+	+	+	+	7.0	6.0
7	4	+	+	−	−	5.7	5.5
8	4	+	+	−	−	3.8	4.0
9	1	−	−	−	−	4.4	4.5
10	1	−	−	−	−	4.7	5.0
11	5	−	−	+	+	6.3	5.0
12	5	−	−	+	+	4.4	6.5
13	2	+	−	−	+	6.0	6.5
14	2	+	−	−	+	6.6	6.0
15	6	+	−	+	−	4.3	5.0
16	6	+	−	+	−	4.9	5.0

What is important when judging the relative importance of factor effects is the absolute value of the effect—not whether the observed effect is positive or negative. Whether the effect is positive or negative comes into consideration only after the analysis, when planning the optimal intervention for a given patient.

In the insomnia example, the Food & Drink factor's main effect on the quality of sleep, as measured by the VAS, is calculated as shown below.

$$\Sigma Y_{+} = 7.0 + 7.0 + 5.7 + 3.8 + 6.0 + 6.6 + 4.3 + 4.9 = 45.3$$

$$\Sigma Y_{-} = 4.0 + 5.4 + 6.2 + 8.4 + 4.4 + 4.7 + 6.3 + 4.4 = 43.8$$

Food & Drink VAS Factor Effect

$$= \Sigma Y_{+} - \Sigma Y_{-} = (45.3/8) - (43.8/8) = 0.1875$$

The main effects for all of the factors in the insomnia example, for both the quantity of sleep measured in hours, and the perceived quality of sleep as measured by the VAS, are given in Table 6. As shown, the Bedtime factor had the largest effect on both the quantity and perceived quality of sleep obtained by the participant and the Food & Drink factor had the least effect on these 2 outcome measures.

Table 6

THE MAIN EFFECTS FOR ALL FACTORS INCLUDED IN THE INSOMNIA EXPERIMENTS*

	Food & Drink	Yoga	Bedtime	Rise Time
Hours slept	0.13	0.25	0.88	0.50
VAS score	0.19	0.74	0.99	0.54

*VAS indicates visual analogue scale.

To calculate interaction effects, one must first identify when the interaction between pairs of factors is considered to be "+" and when it is considered to be "−." This is accomplished by multiplying the signs of the relevant factors pairs for each experimental run. For example, as shown in Table 5, in the first day (run scenario 3) the Food & Drink factor is "−," the Yoga factor is "+," the Bedtime factor is "−," and the Rise Time factor is "+." The interactions between (*a*) Food & Drink and Yoga and (*b*) Food & Drink and Rise Time would be characterized as "−" (the result of a negative being multiplied by a positive) and the interactions between (*c*) Food & Drink and Bedtime and (*d*) Yoga and Rise Time would be characterized as "+" (the result of multiplying 2 negatives or 2 positives).

Beyond calculating the factor effects and ranking them in order of absolute magnitude, it may be helpful to draw a scree plot, as shown in Figure 2 for the main factor effects on hours slept. A scree plot is constructed by first creating a *y*-axis with a range slightly larger than the range of effects observed for the factors. The factor effects are then plotted, from left to right, in descending order of the size of the effect, regardless of whether the observed effect is positive or negative. Although a subjective opinion as to which effects are critical is still required, often a breakpoint, where the slope of the plotted line changes, is observed. The breakpoint, if present, provides some guidance as to which factors may be considered to provide an important effect on the process under consideration.

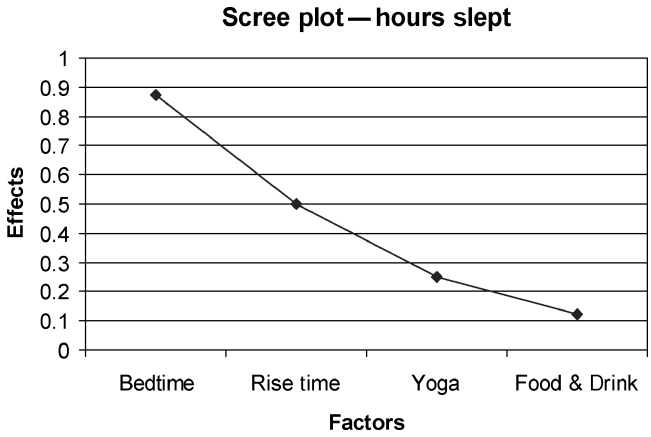

Figure 2. A scree plot of the main effects for the insomnia experiments.

This procedure, however, does not identify whether any of the factor effects, or interactions between pairs of factors, rises to the level of significance. This can easily be tested by standardizing the effects and plotting them on normal probability paper (as shown in Fig 3 for hours slept), or by using one of several DOE software packages (for further information, see recommended texts). In Figure 3, the *x*-axis labeled represents the number of standard deviations from the mean number of hours of slept. The main and interaction effects that cross the dotted line are significant at the $\alpha = .10$ level. The Bedtime factor's main effect and the interaction of the Food & Drink and Rise Time factors were found to be significant for both the number of hours slept and the quality of sleep obtained (figure not shown).

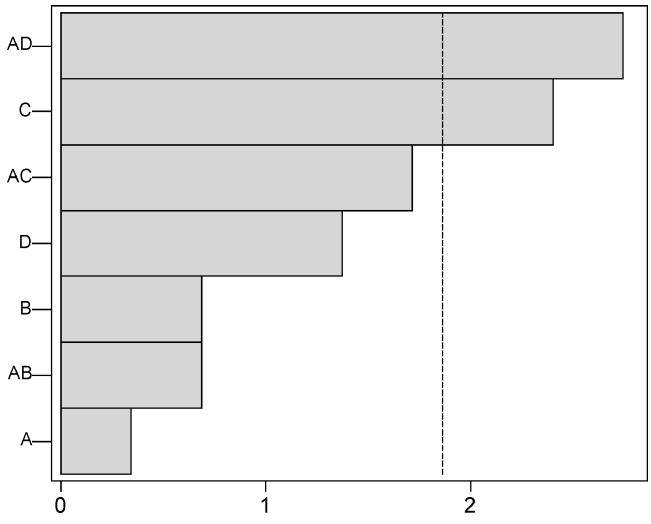

Figure 3. A pareto chart of the main and interaction effects for the insomnia experiments. Response is hours of sleep, $\alpha = .10$. A = Food & Drink, B = Yoga, C = Bedtime, and D = Rise time.

The final stage of analysis is optimization of the process through determination of the factor levels required to maximize the outcome for the patient. In what direction (high + or low −) should each factor be adjusted to produce the best response? A third type of graph, shown in Figures 4 and 5, can be used. Here the mean change in the response variable (hours of sleep in this example) is plotted for when each experimental factor and pair of factors move from low (−) to high (+). As shown in Figure 4, the Bedtime factor is associated with a greater number of hours of sleep when set at the high (+) level, for example, going to bed before 10:00 PM. To take advantage of the significant interaction between the Food & Drink and Rise Time factors, the participant should refrain from eating and drinking after 8:00 PM (the "+" factor level) and should stay in bed, sleepless or not, until 6:00 AM (the "+" factor level), as shown in the interaction plots given in Figure 5.

DISCUSSION

In the previous example, the power of using a basic DOE method to identify which behavioral changes could be implemented to optimize the number of hours slept and quality of sleep for an individual with long-standing insomnia is illustrated. Although implementation of DOE for the clinical management of individual patients can be relatively simple, there are important issues that remain unaddressed in our brief overview. The most obvious challenges are that human beings, in general, are much more complex than industrial process lines and, having free will, are more difficult to control.

In addition, although we have been able to identify which behavioral changes, and in which

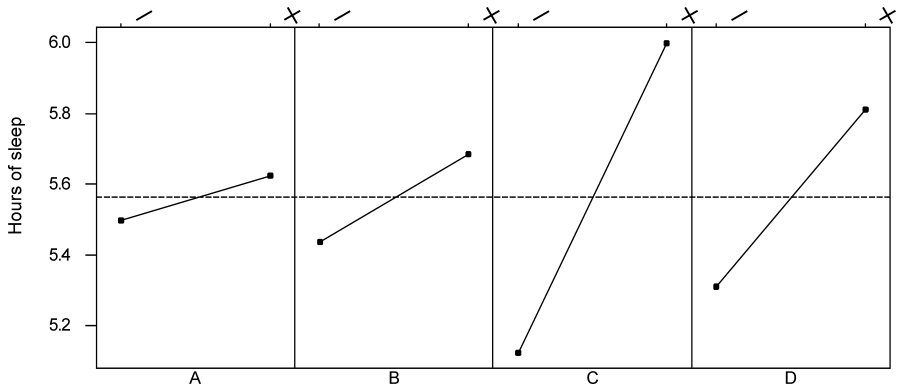

Figure 4. A main effects plot of data means for hours of sleep for the insomnia experiments.

combinations, could be made to optimize the participant's sleep, we have done so within a fairly narrow context during the period of experimental runs. Human beings are affected by various cycles, including daily cycles (eg, blood glucose levels are typically low in the morning and high in the evening), weekly cycles (eg, days of work and days of leisure), monthly cycles (eg, menstrual period), yearly cycles (eg, variation in activity patterns according to seasons), and life cycles (eg, changes in sleep needs according to age). Variation in cycle times could be included as an additional factor in the DOE design, but it may be more efficient to repeat the DOE experiments at different times within these cycles.

The overview presented scratches only the surface of the field of DOE and overlooks many considerations that must be made for more complex designs and analyses. More advanced DOE texts should be consulted and the support of a knowledgeable statistician is recommended if venturing far beyond full-factorial or half-factorial designs with more than 2 factor levels.[7,8] It should be noted, however, that

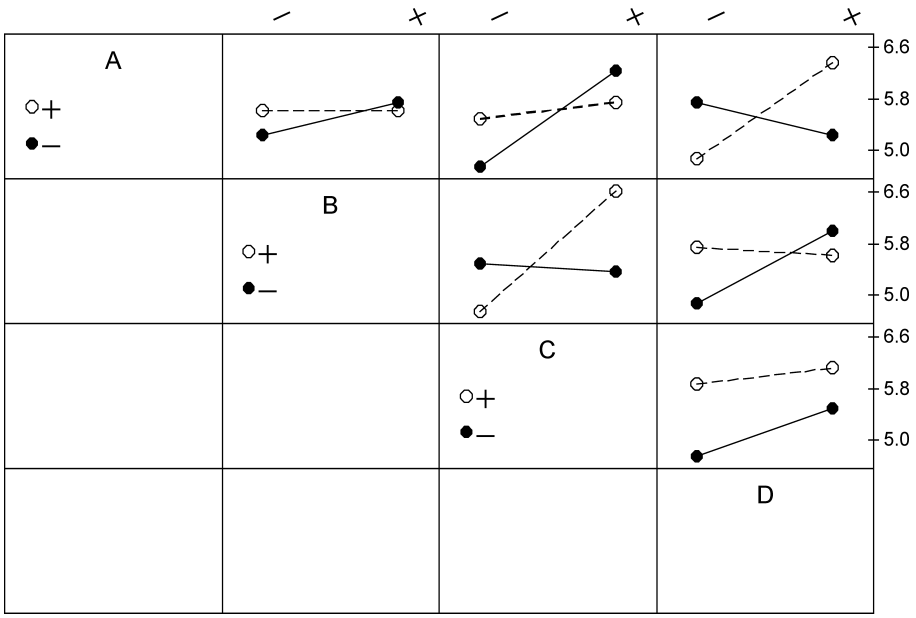

Figure 5. An interactions plot of data means for hours of sleep for the insomnia experiments.

with the more complex DOE designs the analysis may require more than a calculator, but implementation of the experimental runs, once the run scenarios and order have been determined, can be as simple as was introduced in the insomnia example. As such, DOE can be a valuable tool to bridge the gap between clinical guidelines and RCT results and the efficient clinical management of individual patients.

Finally, the value derived from conducting DOE experiments for the clinical management of individual patients does not need to be restricted to the final results. Benefits can be derived from the clinician discussing potential factors and factor levels with the patient and directly including patient preferences in the design. The patient then becomes an active participant in the experimental process and may, as a result, have a stronger commitment to compliance with the optimal therapy identified.

REFERENCES

1. Peter R. External validity of randomized controlled trials: "To whom do the results of this trial apply?" *Lancet*. 2005;365:82–93.
2. McKinlay JB, Potter DA, Feldman HA. Non-medical influences on medical decision-making. *Soc Sci Med.* 1996;42(5):769–776.
3. Box JF. *R.A. Fisher, the Life of a Scientist*. New York: John Wiley & Sons Inc; 1978.
4. Del Vecchio RJ. *Understanding Design of Experiments*. Cincinnati, Ohio: Hanser/Gardner Publications Inc; 1997.
5. Moen R, Nolan T, Provost L. *Improving Quality Through Planned Experimentation*. 1st ed. New York: McGraw-Hill; 1991.
6. Anderson MJ, Whitcomb PJ. *DOE Simplified: Practical Tools for Effective Experimentation*. New York: Productivity Press Inc; 2000.
7. Box GEP, Hunter WG, Hunter JS. *Statistics for Experimenters*. New York: John Wiley & Sons; 1978.
8. Montgomery D. *Design and Analysis of Experiments*. 4th ed. New York: John Wiley & Sons; 1997.