BETTER THAN TAGUCHI ORTHOGONAL TABLES

DORIAN SHAININ AND PETER SHAININ
Shainin Consultants Inc., 35 Lakewood Circle South, Manchester, CT 06040, U.S.A.

SUMMARY

There has been a great amount of publicity about Taguchi orthogonal tables. This paper will evaluate the pros and cons of that approach. In addition an American approach, having the same initial goals of the Taguchi approach, will be presented in detail, representing a significant improvement in meeting those goals without confounding interactions with any main effect or with other interactions. In addition, this constructive alternative generally requires a much smaller number of tests.

KEY WORDS Taguchi method Orthogonal tables

INTRODUCTION

Since Sir Ronald A. Fisher described statistically designed experiments related to his work in the fields of genetics, biology and agriculture[1] his ideas have been applied to several other fields. When two or more input variables (x factors) are studied simultaneously to evaluate their influence upon an output of interest y, *three* concepts permit the separation of statistical estimates of what each x alone does to y, called main effects; and of what each product of two or more xs do to y, called interaction effects, each independent of the other interactions and independent of each main effect. The experimenter *plans* a *balanced*, symmetrical test plan called a full factorial design, a form of an orthogonal array; and he will deliberately plan to *replicate* or repeat every test combination of the two or more levels of each variable, testing them in a preplanned, *randomized* sequence.

An increasing amount of effort has been expended during the past 45 years to adapt the powerful concepts of balance, replication and randomization to the field of industrial manufacture of product. Engineers and managers appreciate strategies that can confirm or deny the validity of cause and effect relationships; especially if those approaches can operate economically at desired values of statistical risk and confidence. In time, significant contributions were made to the improvement of product quality and reliability (Q & R), as well as to other management activities. Unique ideas have now simplified, as well as made more rigorous, the statistical analyses of test results.[2, 3]

Some professionals have improved the cost-effectiveness of statistical designed experiments (SDE) by conducting *extremely* objective clue-generating strategies that narrow down the kinds of variation of the quality characteristic y to precede the designing of the experiment. Multi-vari charts, paired comparisons and concentration diagrams represent examples. This preliminary research practice virtually always reveals unsuspected x variables, thus avoiding the use of variables which would be the usual subjectively selected ones based on 'tenable hypotheses'. Such subjectivity often leads to insignificant results, representing an unfortunate loss of time and effort.

After the SDE, particularly when interactions of practical significance have been revealed, other pioneers have developed approaches that find the combination of new levels of the xs which will optimize a single y output; and even an optimum compromise for multiple y requirements. Note the use here of 'practical' significance (magnitude of improvement) in contrast to 'statistical' significance (not just chance results).

Understandably the more variables used in a single SDE, the more opportunities exist to discover dominant main effects and particularly to reveal controlling interactions. But then the test plan would require a rapid increasingly large number of test combinations. Accordingly it became popular to consider the use of only a fraction of the number of test combinations needed for a full factorial SDE. That interest spread because many practitioners did not take the time to find out, or for other reasons never realized the 'price' paid* when one uses fractional factorial SDEs. Some Western promoters of a moderate amount of fractionating even claimed that interactions (three-factor or higher order) would be too difficult to explain; nor could they be important! The gas law $PV = MRT$ (the British

* Certain interaction effects lose their contrast, evaluation of the magnitude of their influence level to level, and so knowledge of their existence is gone! Significant main effects and important interactions have aliases — other 'confounding' interaction names. Thus *wrong* answers can, and often do come from the time, money and effort of the experiment.

0748-8017/88/020143-07$05.00

Received February 1988

scientist Robert Boyle ennuciated it in 1662) plots as a simple graph. It depicts a three-factor interaction affecting *y* as pressure, or as volume. These promoters may have forgotten their course in physics? Then six years ago Americans began to hear publicity about Taguchi orthogonal tables, described as Japan's 'secret super weapon' the real reason for developing an international reputation for quality. Claim: a large number of variables could now be handled with practical efficiency in a single SDE! As later details became available, many professionals realized that these arrays were fractional factorials, and that Taguchi went to greater extremes than the Western statisticians in the degree of fractionating. He often fills the selected design with as many single factors for which it has room. The design becomes 'saturated'. The growing interest in Taguchi arrays during the past five years in the U.S.† attests to the fact that industrial people either are not aware of the now still greater 'price' paid, or know of no other way to handle more and more variables at one time.

This paper will first describe the alternative approach, and then explain the logic of the disadvantages of employing fractional factorials.

VARIABLE SEARCH PATTERNS

Dorian Shainin's work in exploring many variables together started in 1952, close to the time when it is understood Dr. Genichi Taguchi was getting started, having joined a Japanese communication system research laboratory in 1949. Presumably, some time later, he got into thinking about the advantages of statistically balanced experiments. In Boston, Massachusetts, Shainin was with the management consultant firm of Rath & Strong, Inc. Early attempts with his colleagues developed random balance, and later multiple balance SDEs.[4] Only the several variables that fit the clues revealed by prior multi-vari charts[5] were used. With just 30 to 50 test combinations more variables could be accommodated simultaneously than with Western fractional factorial practices. Admittedly, they avoided saturated fractionals because of the horrendous confounding involved.

Clients of Rath & Strong nicely isolated controlling main effects, and some two-factor interactions when they were lucky with the random sample of combinations of levels used. Multiple balance discovered more two-factor interactions, and an occasional three-factor one, but it still was not all that was desired as an efficient *x*-searching procedure. Then in the early 1960s, working at Chevrolet (GMC) and at Vickers (Sperry Rand), problems with assemblies of a great many parts led to a new idea: component search patterns. It used a rapidly converging, strategic process of elimination. The problems previously seemed insoluble, probably because of mysterious higher-order interactions involved. The conversion, of the rather amazing power of this new idea, from parts or components to *variables* became quite straightforward.

Start with two assemblies that, for some unknown reasons, perform with a different magnitude of *y*. Interchange a pair of common parts between the assemblies. If the original performance figures do or do not follow the part, the conclusion is obvious. If, instead, only one or both change an unexpected, but statistically significant amount, then something is different between the parts, but there is also a contribution coming from some other part in the rest of the assembly. As one might expect, in hundreds of such applications of component search patterns (CSP), the original performance difference practically never came from a difference between a single pair of parts. By sequentially swapping a second, third, etc., suspected pair, in order of expected decreasing importance, a great number of new combinations are simultaneously tested — and two-, three- and higher-factor interactions reveal themselves!

But before any interchanging of common components is done, check for repeatable outputs following disassembly and rebuilding; since swapping would involve some degree of such physical actions. Tear down and rebuild the two assemblies, two times each. That is Stage I. If they both repeat in accordance with two simple criteria, it can be statistically shown with 95 per cent confidence that the observed small changes in *y* are trivial in comparison to the original difference in outputs between the lower-*y* and the higher-*y* assemblies. Swapping, called Stage II, can then commence.

If the units fail to meet the criteria, the specific source of the lack of repeatability has to be isolated by using a progressive tear down and rebuild procedure. First one part, then two together, then three, etc., until the step sensitive to the change in *y* reveals itself. Often it is a gasket thickness, torque on a nut, a spring contacting a different spot or the like. After that source is modified and the statistical criteria can be met, start Stage II.

Directly following the detection of the second component which makes a significant difference, a capping run checks whether any of the remaining parts not yet interchanged could have a significant influence. That group, the rest of the parts, are treated as a single subassembly. It is interchanged with respect, or in relation to the two previously isolated components. If one or both of the new *y*s are disturbed significantly, the capping run is not successful. Continue the sequential swapping until a third important part discloses itself. Now run a three-part capping run test; and so on until a successful capping run, where neither *y* changes, terminates Stage II.

† The authors were surprised to learn, during a recent trip to Japan, using persistent and *intensive*, repeated questioning (through interpreters) to get beyond the party-line stories, that his [Taguchi] methods had only seldom been used in that country.

For the next, and last stage, enter all the numerical y results in a full factorial matrix for the two or more pairs of parts previously isolated Thus Stage III determines the magnitudes of each main effect and of each interaction effect. A detailed physical inspection of those parts then finds the x causes of the y difference.

When working instead with *variables*, two important 'universals' make a similar searching technique readily operable. Among many, even hundreds of candidate (usually unsuspected) factors, a single one or a single group of interacting factors will be the root cause of variation of y. Secondly, its independent influence is not a straightforward portion of the sum of all the independent influences. In statistical language, the standard deviation σ of y is not equal to the sum of the separate, independent σ effects. Hence the second 'universal': the σ of y equals the square root of the sum of the separate, independent *variance* influences. A variance is the standard deviation squared, σ^2. Suppose only two xs were involved, one causing five units and the other 1 unit of y variation. The observed total output variation would not be 6 units. Instead it would be 5·1 units, the square root of 26. That root cause x, causing 5 units of y variation, has been called the Red X by these authors. It contributes 25 times as much as the 1 does to the observed 5·1. If one succeeds in removing the 1·0 influence, the output would then be 5·0, reduced by only 1/10th of the change of 1·0. If one reduces the 5·0 influence by the same amount, to 4·0, the output would be the square root of 17, or 4·1 units. Attempting to control a non-Red X buys almost no improvement, certainly much less than one would expect from a change to that one smaller main effect.

The top section of Figure 1 illustrates the planning for the Stage I test for variable search patterns. One needs to have the factors arranged in two sets, *equivalent* to the two assemblies described above that performed differently; one set expected to have a higher y output than the other. Say the multi-vari chart clues derived from the variation of y point to 11 'x' variables. A full factorial matrix, with two levels of each variable denoted as the + and the − levels, would consist of 2048 test combinations (cells). It could evaluate 2047 independent main effects and interactions as candidates, in all of the combinations of the 11, for the Red X. But one would not reasonably have enough time to complete such an extensive experiment. So plan to start with two of those test combinations that may exhibit a large repeatable difference. Select cells with a different sign (level) of each factor, in order to permit the later swapping of Stage II. One cell is the HIGH column set of signs of the factors, the other LOW, shown in Figure 1.

To anticipate effective implementation of their findings, plan a team of three experimenters assigned from the engineering, manufacturing and quality departments. They arrange the 11 factors in

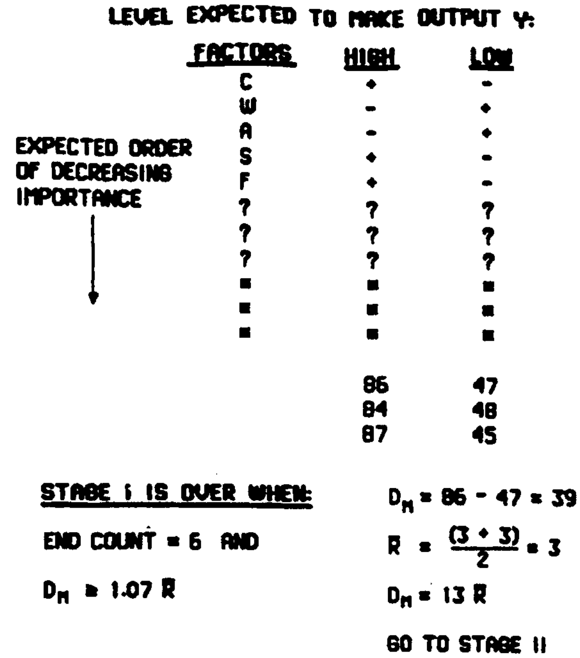

Figure 1. Variable search patterns: Stage I

a suspected consensus order of decreasing importance. Figure 1 shows that they may have an opinion about five of them. The remaining less important, trivial six can thus be arranged in any order. They assign at random, opposite signs for each such trivial variable.

The lower section of Figure 1 depicts the results of empirical testing of all these previous hypotheses. Material was fabricated at the simultaneous sign levels for all 11 factors, as shown in the HIGH cell, and the output y was measured as 86 units. Similarly, the LOW cell result was 47. The team gets early congratulations for their prediction. But if the 47 comes on the left and the 86 on the right, under LOW, just change the headings!

If those first two test results come out close to each other, say, 51 and 50, then try reversing one or more pairs of signs to see if a cancellation of influence was taking place.

With a now larger starting difference between ys, replicate. Should the agreement in the cells turn out to be poor, look for a clue as to one or more factors to be added to the 11, at the top of the list of suspected variables — something important was causing that lack of repeatability. Qualifying for Stage II may take some time. Usually, in hundreds of applications, it has been from a few hours to, say, a week.

'END COUNT = 6' means that the three results in one cell all rank above the three results in the other cell, as they do in Figure 1. D_M means the 'difference between the *medians*'. $\bar{R}$ means 'average range'. For readers interested in the supporting statistics, Figure 2 shows the calculated source of the 1·07 coefficient. For the end count of 6, there are exactly 20 *different* ways to arrange three HIGH cell results and three LOW cell results in rank order. The number of combinations of six items taken three

END COUNT OF 6

$D_M \geq 1.07\ R$

$$F = \frac{n\,\sigma_{\bar{x}}^2}{\sigma_x^2} \qquad \text{USING MEDIAN, } M, \text{ INSTEAD OF } \bar{x} \qquad F = \frac{n\,\sigma_M^2}{\sigma_x^2}$$

$n = 3$

$$F_{(1,4)_{.95}} = 7.71 = \frac{3\,\sigma_M^2}{\sigma_x^2} \qquad \text{Taking the square root of both sides}$$

$$2.78 = \frac{1.73\,\sigma_M}{\sigma_x} \qquad \frac{\sigma_M}{\sigma_x} = \frac{2.78}{1.73} = 1.61$$

$$\sigma = \frac{R}{d_2}$$

SUBGROUP SIZE	d_2
2	1.128
3	1.693
4	2.059
5	2.326

$$\sigma_M = \frac{R}{d_2} = \frac{D_M}{1.128}$$

$$\sigma_x = \frac{R}{d_2} = \frac{R}{1.693}$$

$$\frac{\sigma_M}{\sigma_x} = \frac{D_M/1.128}{R/1.693} = 1.61 \qquad D_M = \frac{1.61(R)1.128}{1.693} = 1.07\ R$$

Figure 2. Two criteria for significant Stage I result

at a time is 6!/3!(6−3)! = 20. Only one of those 20 can show an end count of 6 deceptively, therefore risk = 0·05, and confidence = 0·95.

Fig. 3 shows the results, one by one, of the interchanging of one variable at a time, in comparison to the rest of the factors unchanged, denoted here by R, the 'Rest'. The subscripts, L and H, indicate the cell from which they came. This interchanging directly corresponds to the above component search pattern strategy. The empirical test results are now compared with the medians, M, from Stage I. When small variations occur, within the control limits, we cross out, disregard the influence of that variable in the ANALYSIS column, either as a main effect or as a member of an interaction involving that factor. But, if the results from one or both of the swaps goes beyond a control limit, call that influence statistically significant with a preselected level of 95 per cent confidence. The value for t of 2·776 comes from the 'Student's t' table, extensively published, for 95 per cent confidence. $\bar{R}/d_2$ is an estimate of σ.

If either one, or both of the interchanged test results departs from the previous median value, tabulated in column 'M', by an amount exceeding either control limit of $M \pm 4{\cdot}92$, the variable identification letter cannot be crossed out. 'W+' means that W plus some other variable not yet identified is having a significant influence. Tests 9 and 10 are the equivalent CSP 'capping run'. If R, the Rest of the yet unswapped variables, had not been eliminated in Figure 3, that unsuccessful capping run would be followed by an interchange of the next factor F, etc., until a third variable shows significance and a three-factor capping run would be tested.

Following a successful capping run, Stage III shown in Figure 4 shows all the previous data entered in a full factorial matrix and depicted by a graph. The non-parallel lines represent an interaction influence beyond the average spacing and average slope denoting the strengths of the two main effects, respectively.

FOR SIGNIFICANT STAGE II RESULTS: ONE OR BOTH OUTSIDE THE "CONTROL LIMIT" RANGE MEDIAN, $M \pm t_{.95}\ (\bar{R}/d_2)$

FOR 4 DEGREES OF FREEDOM:
$M \pm 2.776\ (3/1.693) = 4.92$

TEST	COMB.	RESULT	M	ANALYSIS
1	$C_L R_H$	85	86	
2	$C_H R_L$	48	47	✗
3	$W_L R_H$	81	86	
4	$W_H R_L$	70	47	W+
5	$A_L R_H$	84	86	
6	$A_H R_L$	49	47	✗
7	$S_L R_H$	58	86	
8	$S_H R_L$	86	47	S+
9	$W_H S_H R_L$	89	86	
10	$W_L S_L R_H$	43	47	✗

STAGE II IS OVER WHEN: ✗

Figure 3. Variable search patterns: Stage II

FRACTIONAL FACTORIALS

Start with a full factorial, two-level, three variable SDE. The array table for the two levels, + and − would be as given in Table I. This 2^3 U.S., U.K. *orthogonal array* has three inputs, factors A, B and C, providing the signs for the interaction columns by multiplication of the signs.

When the lower case letter is present for a given cell, the corresponding *single* capital letter input is at the + level; when it is absent, that input is at the − level. For (1), A, B and C are all −. That notation makes it easy, particularly for larger arrays, to apply

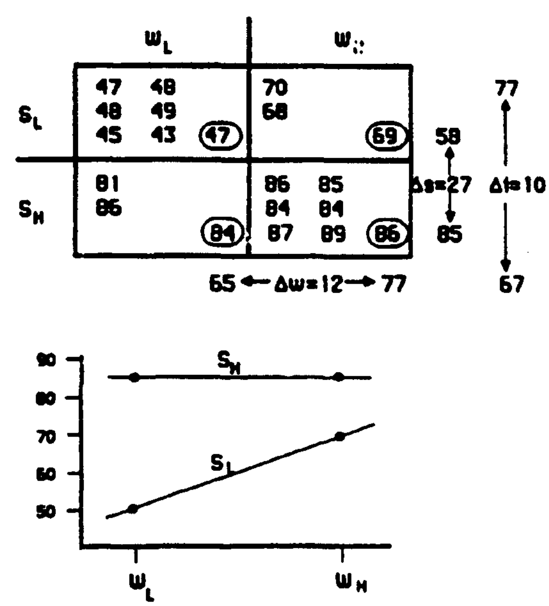

Figure 4. Variable search patterns: Stage III, factorial matrix

Table I

Factor column Cell No.		A 1	B 2	AB 3	C 4	AC 5	BC 6	ABC 7	y
1	(1)	−	−	+	−	+	+	−	#
2	*a*	+	−	−	−	−	+	+	#
3	*b*	−	+	−	−	+	−	+	#
4	*ab*	+	+	+	−	−	−	−	#
5	*c*	−	−	+	+	−	−	+	#
6	*ac*	+	−	−	+	+	−	−	#
7	*bc*	−	+	−	+	−	+	−	#
8	*abc*	+	+	+	+	+	+	+	#

the Yates' algorithm to estimate the separate independent magnitudes of the influence of each interaction and of each main effect. The # symbol represents the numerical average of the two or more replication results in each test combination cell.

A four-factor, two-level, full factorial test plan would have 16 cells, and the corresponding table would have 15 columns. Each of the 15 permits the calculation of a 'constrast' between the $\bar{y}$ (average of the averages) of the +s and the $\bar{y}$ of the −s, representing an estimate of the magnitude of the influence upon y of that interaction or that main effect. One could use eight cells with four factors, resulting in 'a half replicate' fractional factorial design. In the eight-cell table above, one would plan the fourth factor D to use the signs of any one of the four *interaction* columns. Say the experimenter elects to use the ABC column. Now that contrast column has two names. If its influence comes out strong, one will not know from such 'planned confounding', whether it was caused by the D main effect or by its alias, the ABC interaction. Unfortunately, if its influence comes out not to be significant, that could be the result of both having an important, not necessarily equal, influence but in opposite directions, reducing the net effect. No one would know.

A further 'price' paid comes from the fact that with these four factors, 15 independent influences have to be accounted for. Consider two axioms of what might be considered the algebra of ± signs: (1) either sign squared = +; (2) + times either sign does not change that sign. Accordingly, since the sign of D was made equal to the sign of the ABC interaction, one could write, in sign algebra, $D = ABC$; now multiply both sides by BC: $BCD = A$, since B squared and C squared will always be + and thus not change the sign of A. Check it out. Put D in column 7, the alias of ABC. Now multiply the signs of columns 2, 4 and 7. In each cell you will get the sign of A. So the alias of A is the BCD interaction. Notice that one can move letters from one side of the sign equation to the other. No division as in the algebra of Euclidean space. Hence by putting D in column 7, one automatically creates seven aliases:

$$A = BCD,\quad B = ACD,\quad C = ABD,\quad D = ABC,$$
$$AB = CD,\ AC = BD,\ AD = BC$$

and all the troubles described above for $D = ABC$ unfortunately apply to the other six aliases. But two names for each of seven columns account for only 14 contrasts. The 15th is the $ABCD$ four-factor interaction. Since $D = ABC$, $ABCD$ will always be + for all of the eight cells. The contrast between the + average and the − average has disappeared. All evidence of its influence is lost!

Some Western statisticians have used five factors with eight test cells, a quarter replicate fractional factorial. Now each of the seven columns has four names, and three interactions have no contrast.

TAGUCHI ORTHOGONAL ARRAY EXAMPLE

This case history is often used to introduce the Taguchi arrays[6] (see Table II).

This actual example of a tile manufacturing experiment was performed by Ina-Seito Co., in 1953. The letters A to G represent seven factors, different raw material amounts and ingredients tested at two levels, 1 and 2 shown in each column. In order to see the equivalence of this array to the previous Western, unconfounded full factorial, substitute + for each level 1, and − for level 2. You will then see how Taguchi inverted columns and rearranged the column positions.

Pros

(a) Taguchi's tables retain the balance, or symmetry feature of the Western SDEs. When the average of the 1s is compared with the average of another level, 2s or 3s, all the other contrasts have contributed the same combination of their levels to the single level of the factor being evaluated, thus becoming neutralized.

(b) The test work environment in this example was 'actual large scale manufacturing', rather than the laboratory. The presence of more variables, including unsuspected ones, permits the data (when replicated) to show more scatter. That would warn the experimenter that he would be neglecting at

Table II

Factor Column Test No.	A 1	B 2	C 3	D 4	E 5	F 6	G 7	No. of defectives in 100 tiles
1	1	1	1	1	1	1	1	16
2	1	1	1	2	2	2	2	17
3	1	2	2	1	1	2	2	12
4	1	2	2	2	2	1	1	6
5	2	1	2	1	2	1	2	6
6	2	1	2	2	1	2	1	68
7	2	2	1	1	2	2	1	42
8	2	2	1	2	1	1	2	26

least one strong variable in the real world of the shop.

(c) When a single strong variable is suspected, these balanced arrays could confirm it as an influential main effect. Examples would be the change of state of a material with temperature, or with pressure; the strength of an adhesive with curing time, or with temperature.

(d) The Taguchi examples claim that one additional actual test set under the best conditions indicated by the earlier experiment confirms conclusions. Sounds good. But unfortunately a confidence interval checks that conclusion; and a higher confidence number provides a larger interval!

Cons

(a) This example depicts a saturated fractional factorial, quite representative of many Taguchi applications. Each of the seven columns has 16 names, 15 interaction causes of variation in addition to the main effect used as a title. Information about 15 additional interactions has been lost. None of this potentially positive knowledge disappears when one uses the constructive alternative, variable search patterns.

Taguchi (Reference 6, p. 171) states (not so clearly?):

> 'One often hears that ever since experiments by orthogonal arrays have been performed, it has become possible to apply the results of small-scale experiments from the laboratory directly on-site. This is because a factorial effect that is consistent even when the conditions of other factors change has a good chance of being reproduced even if the single condition of scale changes.
>
> 'However, this *does not guarantee that the experiment will succeed* with the use of the orthogonal array. If there is additivity to the main effects, either an experiment by an orthogonal array or an experiment by changing one factor each time will work well, but if the interactions are great neither will go well. That a certain factor influences the performance characteristic means that this factor imparts effect to the performance characteristic, or in other words that *it performs work*. Therefore, fundamentally, there should be additivity to the quantities of work. But for various reasons, additivity fails, that is interactions exist. *In nearly all cases, interactions is* (sic) *not considered in this book. This is not because there is no interaction. It is because, since there can be an interaction, we perform experiments only on the main effects, having cancelled interactions*. If the interactions are great, no assignment works well except experiments on a certain specific combination. Good results are obtained neither by an experiment on one factor at time nor by an experiment using an orthogonal array if interactions have been omitted.'

We have heard it expressed that all interactions disappear if the variables are plotted on logarithmic scales, or if the output is plotted against an abscissa of the product of the input variables. Having worked with Pratt & Whitney Aircraft for several years in the development of the RL–10 (cryogenic liquid hydrogen–oxygen second stage rocket engine used in Centaur and Saturn), we wonder what the U.S. Air Force reaction would be to: 'we could remove the synergistic creation of that engine's power, the hydrogen–oxygen interaction, by changing the scale on which we plot the data'!

(b) Conducting the test sequence in systematic, rather than random order, permits spurious associations. Simply eye-balling these tests results clearly shows something other than the seven factors selected is controlling the number of defectives: with time, going from test no. 1 to no. 8, the 16, 17 level descended to 6, then jumped to 68 to start the next cycle down. If factor A were the strong one, as the Taguchi report claimed, that Red X would not have permitted the second 6 to occur with A at level 2!

(c) Not replicating, or repeating the test results in random sequence prevents obtaining a valid statistical estimate of error or noise. In this example the authors applied the 'Shainin test': use random numbers in place of the reported test data, to check the discriminating ability of any promoted new plan. Taguchi reported from his (very questionable) statistical analysis of his numbers:

16, 17, 12, 06, 06, 68, 42 and 25

that factors A, D, E, F and G were significant at a risk level of 0.01.

The authors used eight pairs of random numbers, 68 and under, three times. They came out:

60, 32, 03, 11, 04, 61, 66 and 08
61, 24, 12, 26, 65, 14, 54 and 66
01, 16, 60, 36, 59, 46, 53 and 42

and with exactly the same Taguchi statistical analysis:

(i) the first set showed that factors B, C, F and G were significant at a risk level of 0·01, with factor A at 0·05.
(ii) the second set showed that factors A, C, D and F were significant at a risk level of 0·01.
(iii) the third set showed that factors A, B, C and G were significant at a risk level of 0·01, with factor F at 0·05.