# Why Small Samples Can Increase Accuracy

**Robert Hamlin, University of Otago**

## Why Small Samples Can Increase Accuracy

### Abstract

*This paper examines the impact of size on experimental accuracy. Many designs used within marketing were originally developed within agriculture. These include the instruments of parallel comparison, which include; full factorials, fractional factorials, latin squares and derivatives such as conjoint analysis. It is well understood within agricultural research that the sample size used within these experiments should be kept to a minimum if maximum reliability is to be achieved. This understanding does not appear to have been transferred to Marketing. This article explains the reasons behind this counterintuitive claim that the smallest possible dataset is technically as well as adminstrationally desirable when these instruments are deployed.*

### Introduction

Marketing utilises a variety of experimental designs that were initially developed for use in agricultural research, either directly by transfer (latin square, full factorial) or by subsequent development of these initial designs (conjoint analysis). With one or two caveats, the transfer has been a happy one (Hamlin 2005). However there are one or two insights derived from their development and application within Agriculture that do not seem to have been widely disseminated within the marketing research literature. This article examines one of these insights: That it is a key requirement of experimental reliability that the samples from which the results are derived is as *small* as possible. This is achieved by using an experimental pattern that has the highest efficiency and by using the smallest possible sample size within each treatment condition. For a discipline that, typically takes the attitude of: 'The bigger the better' for its research samples, this may seem a surprising statement. This article demonstrates why it is as applicable to marketing as it is to agricultural research.

### The economic and practical advantages of small sample size

High efficiency in an experimental design has the obvious attraction that a result can be obtained after a much lower expenditure of time, money and other research resources. The same comments can be made with regard to a small individual sample for each treatment condition within any such design. A further benefit of both of these features is that any experiment that possesses them may be administered with a very much lower degree of disruption of the environment in which it is undertaken. This is important as much of the research work using agricultural designs in Marketing, since their introduction by Brunk and Federer in 1953, has been administered in a commercial environment, such as a store or supermarket (Brunk and Federer, 1953 a&b; Cox, 1964; Dodds, Monroe and Grewall, 1991; Kennedy, 1970; Rui and Meyers-Levy, 2009). It can be difficult to access these retail environments for the purposes of research. Under such circumstances, where the cooperation of a commercial partner is required, the efficiency of the experimental design may determine if research is undertaken at all.

### The technical advantages of small sample size

Beyond these advantages there is a much more subtle, yet highly important benefit endowed by high efficiency. Nearly all the experimental designs sourced from agriculture are instruments of *parallel* comparison, which rely on the controlled application of the independent variables to *equivalent* experimental units. The means of these individual units are then compared to the mean of the entire experimental population or to a single 'control' condition if a partially confounded design is being used.

The larger the experiment becomes in terms of the number and/or size of the experimental units deployed, the harder it becomes to ensure that they are all equivalent for the purposes of these comparisons as micro environmental changes are incorporated into the research sample. The less certain the equivalence of the treatments is; then the less reliable the overall experiment will be. It is for this reason that Sir Ronald Fisher, the developer of nearly all these experimental designs, made the following comment in one of the books in which he developed these techniques:

"...the problem of designing economical and effective field experiments is reduced to two main principles (*i*) the division of the experimental area into the plots as small as possible...; (*ii*) the use of [experimental] arrangements which eliminate a maximum fraction of soil heterogeneity, and yet provide a valid estimate of residual errors."

**[Fisher, 1950: 510]**

The first of Fisher's requirements for effectiveness is clear enough. The second, when applied to Marketing, may be restated as: 'Use the most efficient design (with regard to required sample size) that may be applied to the research objectives.' In order to explain why this is so important, it is useful to turn to the 'classic' illustration of this concept as it is used in agricultural training (Figure 1).

**Figure 1: 'Classical' example of the role of efficiency in agricultural research latin square v. equivalent full factorial**

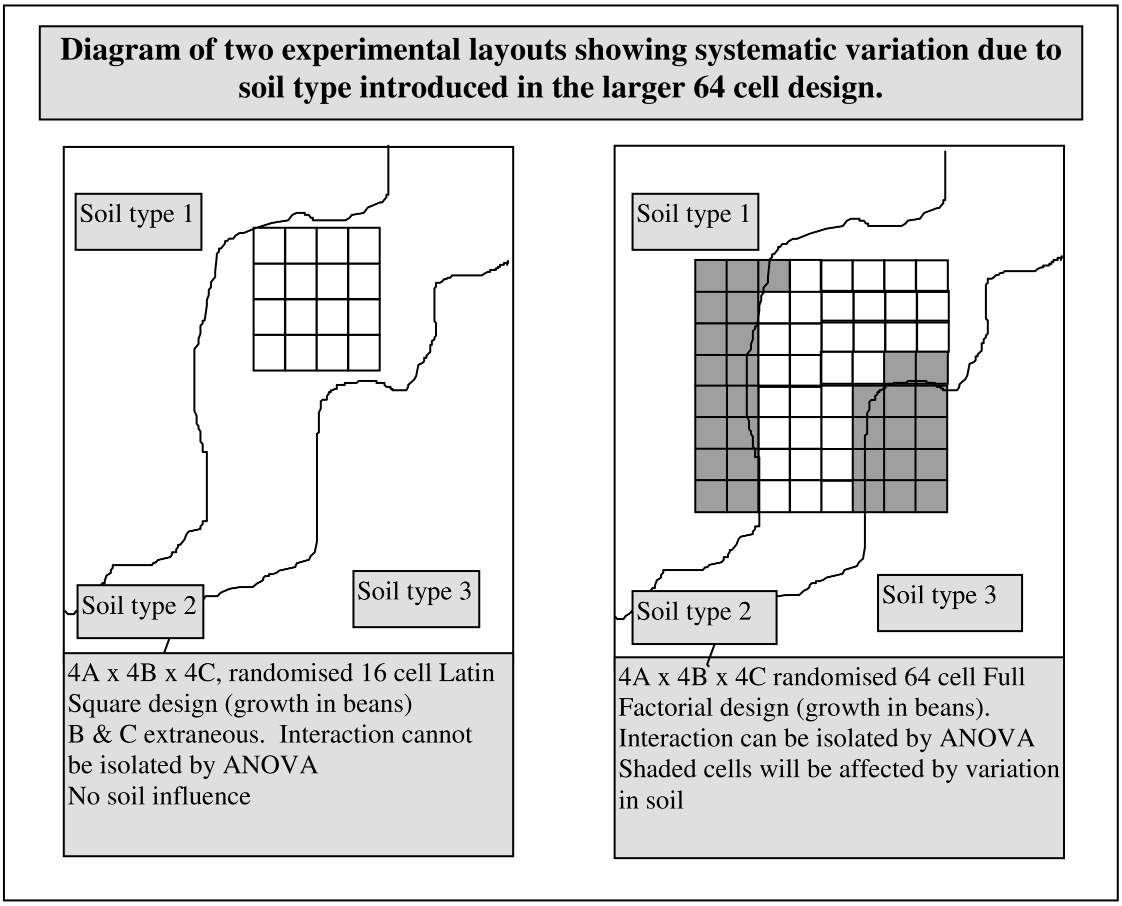


Figure 1 compares two equivalent designs, a 4x4 Latin Square and a 4x4x4 Full Factorial laid out in the same field - which like all fields has variations in soil type across it. The factorial design, although *appearing* to offer more information, suffers from having a range of soil types included in it as a consequence of its larger size. If the treatments were arranged systematically within the layout, this will lead to a biasing of results. If they were arranged randomly, the result is an increase in the error term and a reduction in the power of the experiment.

While neither of these situations is desirable, the latter option is obviously the better of the two if an incorrect conclusion is to be avoided. As a consequence, randomised versions of Full Factorial and the larger Latin Square designs are published in tabular format for the use of agricultural researchers. The same comments can be made with regard to a latin square experiment in which the individual plots are four times the size. The outcome is likely to be less rather than more reliable, due to the variations in environmental conditions within individual treatment conditions that would then contribute to the error term.

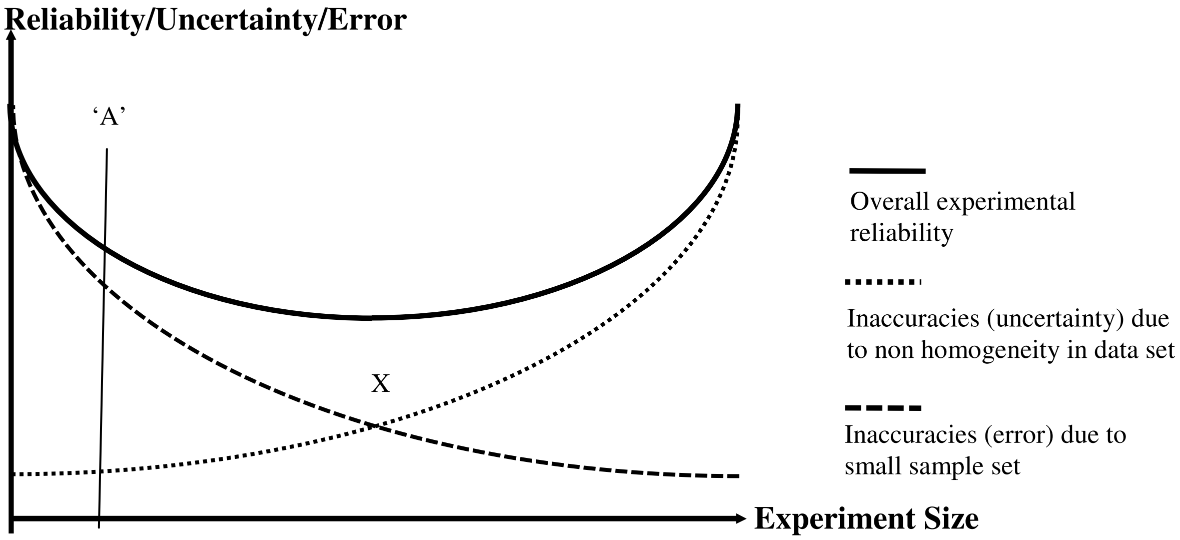
Similar situations can be demonstrated in Marketing. For example, 4x4 Latin Square using one store at a specific of the week would take four weeks to conduct. An equivalent 4x4x4 factorial experiment would have taken four months, or would require the use of four different locations. Such a situation could lead to considerable problems with history effects, especially if the treatments were not randomised. For example, a very large experiment examining the effect of country of origin on consumer evaluations of products in New Zealand was completely destroyed by the impact on consumer attitudes stemming from the incident involving French nuclear testing at Muroroa Atoll that occurred during the course of the experiment. France was one of the independent variable set, and consumer attitudes towards it changed to the degree that the error term overwhelmed the significance of any results. Had the researcher used a partially replicated latin square design, rather than its less efficient full factorial equivalent, then the work would have been concluded before this unfortunate event occurred!

## Technical implications

A trade-off therefore exists (Figure 2). Increasing the sample size of an experiment can lead to an increase in accuracy/power that is statistically predictable via decreases in the mean squared error term (MSE). This is achieved, either by decreasing the numerator, the total sum squares for error, by increasing the stability of individual ‘cell’ observations by increasing the number of data points per cell, or by increasing the denominator term, the numbers of degrees of freedom associated with the sum of squares for error, by increasing the number of ‘cells’ in the experiment.

However as the size of an experiment relying on parallel comparison increases in size it becomes increasingly subject to external sources of uncertainty due to a non homogeneity in the experimental environment. Unlike the potential sources of error associated with a smaller sample size, the extent and nature of the distortions introduced by these external factors of uncertainty cannot be statistically estimated. It is therefore a true uncertainty rather than a statistically quantifiable risk. The fact that many of these uncertainties may exist within, or enter into, an experimental population without the knowledge of the researcher makes any aspect of the research design that increases the likelihood of such an event even less desirable.

**Figure 2: The 'optimal' sample size for an instrument of parallel comparison.**




The fact that this second source of error is an unquantifiable uncertainty rather than a chance of error that may be estimated statistically means that researchers should make their experiments as small as possible within the constraints of the relationships that they are investigating. This may be done by decreasing the number of observations per cell to a minimum. Adopting the most efficient experimental design that satisfies their requirements also allows them to reduce the required overall sample size, and thereby reduce the chance of an unpredictable distortion occurring due to a non homogeneity of the experimental environment. Such a smaller experiment may be statistically less powerful, but it is predictably so.

In light of this, Fisher's recommendations become more understandable. If the line 'A' in Figure 2 represents the maximum acceptable risk statistical risk for a particular research objective, then the researcher should design their experiment to be as close to this line as possible. They should not, and indeed cannot, aim for the theoretical low point of risk represented by point 'X', as the uncertainty curve cannot be defined or used for the purposes of accurately establishing position of this point. It is for this reason that when one looks at a series of agricultural research plots on a typical field research station, they appear to be extremely small – Typically they are not much larger than a room in a house (Figure 3).

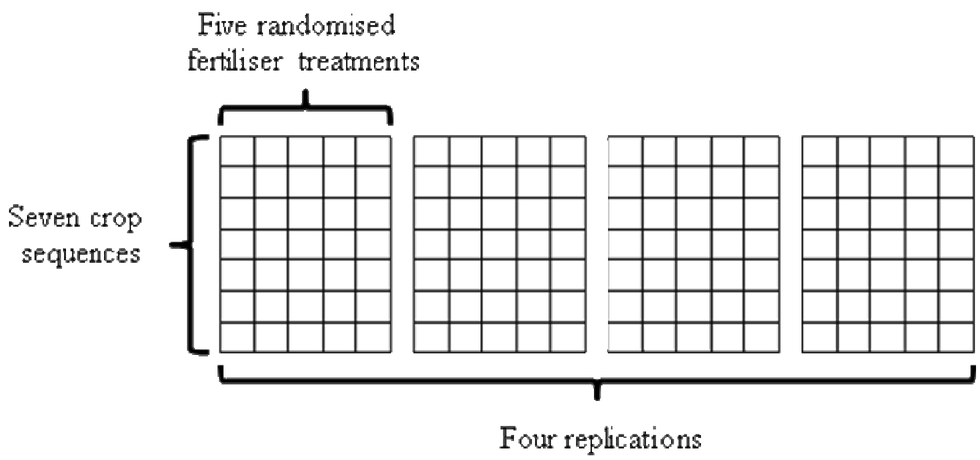
They are this size because this is the minimum that can be established without 'edge effects' becoming an issue (An edge effect is where plants on the edge of a sample plot do significantly better, or worse, than their neighbours in the middle.) If this was not an issue, they would be yet smaller! The researchers do not do this to save money, but to increase accuracy for the reasons stated above. As in the last sixty years commercial agriculture has achieved the unlikely feat of feeding the World in the face of a nearly five-fold increase in its population, largely via the application of the results of such research, the practical aspects of their approach to the application of experimental design are to be respected!

## Conclusions

As Marketing appears to consistently value very large sample sizes in its research; the concept that a small sample size may be technically as well practically desirable when certain experimental patterns are used is an important point, While this position may be justified for survey research, and for some experimental designs, it is potentially undesirable for 'agricultural type' experimental designs that rely upon parallel comparison if the internal validity of the experiment is to be maximised. As these approaches represent a significant proportion of all published research in some areas of our discipline (e.g. consumer behaviour), the idea that the optimal sample size may not be the largest that resources allow, but is in fact the smallest that the most efficient appropriate methodology permits, is worthy of further dissemination.

One final observation is that if circumstances within agricultural research situations do require a larger dataset, then this increased size is usually achieved by replication of small sample patterns, which allows any environmental irregularities impacting upon the larger design to be partitioned as sum of squares for replication in the analysis of variance table. While this well-proven practice is common in agricultural research (e.g. Figure 3), it is rare in published academic marketing research.

**Figure 3:** **'Standard' field research plots of 7x7m.**
Total experiment with all replications occupies 0.6 hectares
(Berzsenyi, Gyrffy & Lap, 2000)



## References


Berzsenyi, Z., Gyrffy, B. & Lap, D. 2000, Effect of crop rotation and fertilisation on maize and wheat yields and yield stability in a long-term experiment, European Journal of Agronomy, 13/2-3, 225-244

Brunk, M.E. & Federer, W.T., 1953, How Marketing Problems of the Apple Industry Were Attacked, and the Research Results Applied, Methods of Research in Marketing No. 4, Department of Agricultural Economics, Cornell University, Agricultural Experiment Station, New York State College of Agriculture, Ithaca, New York, January 1953, Pamphlet, Ithaca, NY

Brunk, M.E., & Federer, W.T., 1953, Experimental Designs in Probability Sampling in Marketing Research, American Statistical Association Journal, September 1953, 440-452

Cox K. 1964, The Responsiveness of Food Sales to Shelf Space Changes in Supermarkets, Journal of Marketing Research, 30, 63-67

Dodds, W.B., Monroe, K.B. & Grewal, D., 1991, Effects of Price Brand and Store Information on Buyers' Product Evaluations, Journal of Marketing Research, 28, 307-319

Fisher, Sir R.A. 1950, Contributions to Mathematical Statistics, John Wiley & Sons Ltd., New York, NY

Hamlin, R.P. 2005 The Rise & Fall of the Latin Square in Marketing, A Cautionary Tale European Journal of Marketing, 39, (3/4), 328-350

Kennedy, J.R., 1970, The Effect of Display Location on the Sales and Pilferage of Cigarettes, Journal of Marketing Research, VII, 210-215

Rui, J., Meyers-Levy, J., 2009, The Influence of Self-View on Context Effects: How Display Fixtures Can Affect Product Evaluations., Journal of Marketing Research, 46 (1), 37-45

Brodie, R.J., Danaher, P.J., 2000. Building models for marketing decisions: Improving empirical procedures. International Journal of Research in Marketing 17 (2-3), 135-139.