DESIGN OF EXPERIMENTS FOR CONTROLLED EXPOSURE STUDIES

John O. Rawlings

North Carolina State University

Raleigh, N.C. 27695-8203

Presented at

Workshop on Controlled Exposure Techniques and the

Evaluation of Tree Responses to Airborne Chemicals

Atlanta, Georgia

February 4, 1986

INTRODUCTION

I want to thank the organizers for this opportunity to discuss experimental design for the controlled exposure studies of the effects of air pollutants on forests. A common complaint of statisticians is that they are called upon for their expertise only after the "damage has been done" so it is indeed a pleasure to participate in this early planning of the exposure studies in forest species. The air pollution studies are extremely important, of major interest to the population as a whole, not just the scientists, and they tend to be very expensive studies.

Every effort must be made to insure that the results from these studies are as accurate and precise as resources will permit and that they provide a valid basis for the intended inferences.

There are three major points I would like to address as an introduction to the discussion which follows this paper. First, I want to discuss the importance of clearly specifying in some detail the objectives of the studies. The purpose is <u>not</u> to constrain in any way the objectives of the study but, rather, to encourage a clear definition of the objectives whatever they might be. In particular, I want to concentrate on the importance of specifying the reference population of environments to which the results of the study are to be applied. Secondly, I want to discuss several key considerations in experimental design. Finally, I wish to make a suggestion on the use of dose response modeling and make an appeal for an early committment to a coordinated analysis of all forestry exposure data.

SPECIFICATION OF OBJECTIVES:

On the surface, it seems unnecessary to even mention the need to specify objectives. We all have in mind general objectives when we set out to run an experiment. However,

very often the objectives are not stated with sufficient detail to distinguish among several different objectives. In many cases, the optimum design for one objective will not be the optimum design for another. It is possible that the optimum design for one objective will not even provide relevant information for the other objective.

This conflict of objectives is illustrated with the recent National Lakes Survey. It is my understanding that the objectives of the study included both assessment of the number of lakes and the surface area of the lakes that were acidified. The optimum sampling design for estimation of "number of lakes" is not the optimum sampling design for "surface area". Therefore, the design used in the survey was a compromise which provided reasonably good information on both objectives but not maximum information, for the resources, on either objective.

We can illustrate the need to be specific in the statement of the objectives using the air pollution studies. The general objective, to study air pollution effects of ozone, for example, does not tell the whole story. Are we interested in simply determining if ozone has an effect on forest productivity or are we interested in assessing the magnitude of the ozone effect on the forestry industry? Is

it the quantity or the quality of forest products which is important? Are all species to be considered or only loblolly pine? Is it the Southeast Coastal Plain environment, the Northeast, or the Northwest environments that are of interest? Even the two objectives of detecting an effect versus assessing the magnitude of the effect require very different kinds of studies if the final inference is to be statistically defensible.

The key point I wish to make is that the statement of objectives <u>must</u> include specification of the set of environments, or the population of environments, to which the final inference is to apply. Regardless of how much we know about the interaction of our biological material with various environmental factors, there will always be a major, random, environmentally induced component of variance in biological studies; both because we cannot identify and understand all of the environmental factors which affect plant growth and because we cannot control many of the environmental factors we do understand. Therefore, if one is to make an inference over a set of environmental conditions, for example the environments of the SE Coastal plain for the next 20 years, the experiments must be designed so that a measure of how the plants will respond <u>in</u> <u>those environments</u> is available.

Every factor that contributes to one environment being different from another is involved in every experiment regardless of whether that factor has been given any special attention in the experimental design. There are undoubtedly many environmental variables that have not been identified but which affect plant response. These unidentified variables, of necessity, are simply allowed to take whatever values they happen to have in the particular test environments. They are responsible, in part, for the "random" environmental variation we observe both within and between experiments. The environmental variables we can identify are handled, basically, in three ways:

(1) The environmental factor is recognized as being important and a decision is made to control that variable at some constant level _for all studies_. For example, supplemental irrigation might be used to ensure that the plants never suffer drought stress.

(2) The environmental factor is recognized as being important and the research is designed to include the study of the effects of this variable and its interactions. The information obtained is then included in any predictions of the behavior of the system; e.g., of the pollutant effects. For example, levels of moisture stress might be included as a treatment factor so that the effects of moisture stress on

the pollutant effects are understood.

(3) The environmental variable is ignored, perhaps because it cannot be controlled experimentally. It plays the same role in the experiment as the unidentified environmental variables. Its effects become part of the random environmental "noise" in the system. If moisture level, for example, is simply ignored, the effects of the "random" plot-to-plot differences in moisture level within an experiment become part of experimental error. The "random" differences in moisture stress over the test environments contribute to the "environment by pollutant" interaction effects and become part of the error of prediction of the effects of the pollutant.

When the intent of the study is to make an inference over a population of environments that include variation in that factor, the first course of action can be recommended only if the environmental factor is known to have no influence on the effects of the pollutant; that is, if there is no interaction between the factor and the pollutant. This will be a rare situation and, in general, should not be the chosen course of action. An inference to the population of environments implies that the level of this "controlled" environmental factor will be allowed to vary according to its rules of behavior in that population and, consequently,

any inference based on data from a "controlled" experiment
will be in error to the extent that there is interaction
between the factors.

Notice that this does not preclude controlling certain
environmental factors. It is perfectly legitimate to
specify that certain factors are to be held constant if, in
fact, that is compatible with the intended population of
inference. For example, if a crop is produced only under
green house conditions of adequate moisture and nutrients
under a fixed temperature regime, it would be appropriate to
"fix" the levels of these factors. Likewise, if the
objective is only to determine if there is an effect of the
pollutant, it may be desirable to hold certain factors
constant in order to gain in precision for that specific
objective.

The second course of action is to be preferred to the
extent that resources permit. It is the intent of science
to understand as thoroughly as possible the systems we
choose to study. This includes how these systems are
influenced by their surroundings. Certainly, it is
desirable to identify the key environmental variables, those
having major impact on the process, and incorporate their
effects in the dose response models. This is an expensive

option and there are obviously limits to the number of such
factors that can be studied in depth. From the perspective
of predicting the impact of pollutants, the gain from
understanding the environmental variables is greater
precision since the environmental variation caused by these
key variables can now be taken into account in the
prediction.

Regardless of the magnitude of the resources, however,
there will always be many environmental factors that remain
as part of the random environmental background, that fall in
category three. To the extent that these "ignored"
environmental variables affect the plants' response to the
pollutant, an inadequate (nonrepresentative) sampling of the
environmental conditions for the test environments will
produce biased estimates of the effects of the pollutant.
Further, the variations in response attributable to these
"ignored" factors contribute to the error of prediction.
Consequently, any statistically valid inference of the
effects of air pollution, for example, must be based on test
data in which the "ignored" environmental factors were
represented much the same as they will be in the population
of environments to which the inference is intended. This is
accomplished, conceptually, by using as test environments a
<u>random</u> sample of the reference set of environments.

To emphasize the role of these environmental factors in research, I would like to show a diagram used by George Box in a recent seminar at NC State University. It takes the form of a slight elaboration on our usual statistical models:

$$Y = f(X; \theta) + \epsilon(W)$$

where $f(X; \theta)$ is the part of the model that incorporates the factors being studied X, and their parameters θ. The change from the usual model is in expressing the residuals, ϵ, as functions of another set of variables represented by W. This emphasizes the fact that what we usually regard as random error, something mysterious and beyond our control, is in fact a result of the ignored environmental variables. They play a role in the research regardless of whether or not they are explicitly considered. Of course, the purpose of research is to move as many of the variables as possible from W to X and, as a result, become more and more precise about what can be said about the process being studied.

Let me reiterate using moisture availability as an environmental variable for illustration. All research could be conducted only under well-watered conditions. The

10

predictions of the effect of the pollutant would be biased

unless moisture level had no effect on the impact of the

pollutant or the population of environments of interest

involved only "well-watered" conditions. Alternatively, the

effects of moisture stress on pollutant effects could be

studied, incorporated into the model, and then any

predictions to specific environments would take into account

the levels of moisture stress. Finally, moisture level

could be simply ignored in the experiments; that is,

moisture variability would be one of the variables in W, a

component of random error. Variability in moisture stress

over the set of environments to which the predictions are

applied would become part of the random errors of

prediction.

Of course, it is very difficult and expensive to do an

adequate sampling of environments. This is particularly

difficult for the studies which are labor intensive and

require fairly sophisticated technology such as the ozone

studies on plants. Nevertheless, it is important to

recognize this problem from the very beginning so that

either (1) an reasonably adequate sampling of environments

can be planned, or (2) the limitations of the data can be

appropriately recognized when the regional inferences are

attempted or (3) the objective can be recognized as being

unreasonable with the resources in hand.

EXPERIMENTAL DESIGN

Experimental design includes the sampling of
environments discussed above. It also includes the choice
of treatment factors and "levels" of each treatment factor,
the numbers of replications within each environment sampled,
the choice of field plot design, size and shape of plots and
blocks which will minimize the impact of local environmental
(soil) variations, and consideration of potential covariates
that might provide statistical control of some local
variations. Of course, we cannot cover the entire field of
experimental design today. I have chosen just a few points
I consider to be particularly relevant.

1. Factorial treatment combinations:

The advantages of factorial experiments should not be
overlooked. This point is particularly relevant in view of
my previous comments on identifying key environmental
variables and building the dose response model to include
their effects. Investigation of treatment factors one-at-a-
time is inefficient and provides no measure of interaction
between treatment factors. The only way interaction effects

can be estimated is by using combinations of levels of
relevant factors in the same experiments. There is no loss
of information in factorial experiments if a factor is
included which turns out to have no effect; the levels of
the factor simply serve as additional replicates of the
other treatments. This extra replication is referred to as
"hidden replication".

The factorial studies should be considered in two
distinct stages. An early objective in a relatively new
area would be to identify those key environmental factors
that influence the response of the plant to air pollution.
A preliminary study for this objective would involve many
environmental factors, all that the researcher thought to be
of primary importance, in a high level 2^n factorial study.
With limited resources, it might be necessary to use partial
replicates of 2^n factorials in these initial studies. That
is, a subset of all-possible treatment combinations is
chosen such that the interaction effects of interest can be
estimated. The _purpose_ of these initial studies is simply
to identify key variables which then will be included in
further research. The second series of studies would be
devoted to developing dose response functions which take
into account the effects of these key variables as well as
the pollutant. These again would be factorial experiments

but with only the important factors involved and with enough
detail to characterize the dose response models.

A compromise design which allows several environmental
factors to be considered and at the same time provides some
information on the air pollutant dose response function
might use, for example, a half-replicate of a 2^4 factorial
in all combinations with four levels of the pollutant. The
half-replicate of the four environmental factors, eight
treatment combinations, would be chosen so as to sacrifice
the four factor interaction. In combination with the four
levels of the pollutant, this design would require 32
experimental units. Five levels of the pollutant would
require 40 exposure chambers. Such a design provides
estimates of the main effects of the environmental factors,
aliased with their three-factor interactions, and estimates
of the interaction effects of the air pollutant with the
environmental factors (Table 1). The latter are aliased
with the four-factor interactions between the pollutant and
three environmental factors. This design provides no direct
estimate of error from true replication. The three-factor
interactions between the pollutant and two environmental
factors would be used as the estimate of experimental error.

Table 1. Analysis of variance for a half-replicate of a 2^4 factorial of four environmental factors with four levels of the pollutant. The half-replicate is chosen so that the four-factor interaction among the environmental factors is sacrificed.

SOURCE OF VARIATION	D.F.	ALIASED WITH:
Total	31	
Pollutant (P)	3	
Factor A	1	B x C x D
Factor B	1	A x C x D
Factor C	1	A x B x D
Factor D	1	A x B x C
A x B	1	C x D
A x C	1	B x D
A x D	1	B x C
Pollutant x A	3	Pollutant x B x C x D
Pollutant x B	3	Pollutant x A x C x D
Pollutant x C	3	Pollutant x A x B x D
Pollutant x D	3	Pollutant x A x B x C
Pollutant x A x B	3*	Pollutant x C x D
Pollutant x A x C	3*	Pollutant x B x D
Pollutant x A x D	3*	Pollutant x B x C

* Three factor interactions pooled for an estimate of experimental error.

Different levels of some treatment factors can be compared _within_ exposure chambers. Different cultivars, for example can be grown in the same chamber. The possibility of including such a factor within the exposure chambers in a conventional split-plot design, with very little additional cost, should not be overlooked.

2. Numbers of levels of air pollutant treatment.

There is a common misconception that the more distinct

levels of a quantitative factor the better. In general,
this is not true. Given fixed resources, a fixed number of
exposure chambers, greater efficiency is obtained from a
relatively small number of treatment levels with a
corresponding higher number of replications at each level.
The optimum number of treatment levels depends on the
complexity of the response curve and the degree of faith one
has in the assumed response function. For example, the
optimum strategy if the response curve is known to be linear
is to split the replicates equally between two extreme
exposures. If the response curve is known to be quadratic,
the optimum number of distinct levels is three. This
optimum number of levels, and their placement, depends on
the true dose response curve, which of course is not known.
To allow for a test of the adequacy of the response model
and to provide protection against misguessing the nature of
the response curve, a few additional points are needed. The
extra design points become increasingly important for
protection as the response to the pollutant becomes more and
more concentrated in a narrow dose range. Thus, one would
certainly use more levels than the optimum number for a
specific dose function but the philosophy of using as many
levels as possible should not be pursued. The general rule
would be to use a few more points than required to
characterize the curve; a few points more than the number of

parameters in the most complex model anticipated. This, of course, assumes that the remaining chambers are dedicated to replication or the investigation of other factors.

3. Placement of design points.

One of my students, Karen Dassel, and I have spent considerable time investigating the optimum design for the Weibull response curve. This is a three parameter nonlinear model used in many of the NCLAN (National Crop Loss Assessment Network) studies. The optimum number of distinct points for estimation of the model parameters is three if the response is known to be Weibull. The optimum placement of these points can also be specified but, again, only if we know the values of the parameters of the response curve. This is, of course, an unrealistic situation for the experimenter for if he knew the curve the experiment is no longer needed. Nevertheless, the study does point out some general rules:

(1) When the "region of interest" covers only a part of the dose response curve, there are sizeable gains in precision to be realized from having the experimental dose range go well beyond the region of interest. With ozone, for example, the region of primary interest is from

approximately 0.025 to 0.07ppm. For the Weibull response model, the optimum minimum dose is as near zero as practical and the optimum maximum dose is the dose that generates approximately an 80% yield reduction; approximately 0.15 to 0.20 ppm depending on species. Precision of predicted losses drops rapidly as the minimum dose level rises or as the maximum dose level falls. THERE IS A STATISTICAL ADVANTAGE OF HAVING DOSES ABOVE THE RANGE OF DOSES EXPECTED IN NATURE.

(2) Given that the end points of the dose range have been determined as above, it is best to concentrate the remaining dose levels in the "region of activity" of the response curve. This, again, depends on some prior idea of where the response will occur and can only be applied realistically after some information is obtained. With poor prior information on the nature of the response curve, uniform distribution of the design points over the dose scale appears to provide protection against misguessing. Uniform distribution of the design points results in some loss in precision, compared to concentrating the points in the region of activity, if the dose response curve has been guessed correctly. But, on the other hand, the precision with uniform distribution can be much better if the curve has not been guessed correctly.

4. <u>Placement of experimental plots and supplemental</u> <u>information</u>.

Exposure studies with perennial crops are long-term commitments. Errors at the beginning of the study will be around to cause trouble for a long time. Therefore, any extra effort in choice of experimental units is likely to pay good dividends over the course of the study. Blocking of experimental units should be used to remove as much of the test site variability as possible. The experimental units (exposure chambers) should be located on uniform material within blocks. A uniformity trial on the experimental area before the experiment is started may be worthwhile. Exposure studies have the advantage in that growth of plants (at least initial growth, stand, etc.) can be observed before the chambers are placed on the plots and before treatments are assigned to plots. Thus, plots can be chosen for uniformity on the basis of the growth of the same plants on which the study will be conducted. In addition, supplemental information can be obtained on the experimental field (soil measurements, etc.) which can also be used to define the experimental units and classify them according to their similarities and differences.

5. <u>Covariates and companion plots</u>.

Use of initial plant and soil data for defining plots
and blocks has been mentioned. The use of this kind of
information to remove field variability through covariance
analysis may be particularly useful in forestry studies
where selection of uniform field conditions may be
difficult. Initial plant size, for example, has been shown
repeatedly to be very effective in improving the precision
of experiments.

NCLAN has also attempted to use supplemental information
in the form of "companion plots", plots immediately adjacent
to the exposure chambers. This is essentially an entire
network of ambient air plots (control plots) interspersed
among the chambers. For the annual plants, this has only
occasionally proven useful for improving the precision of
dose response models. Companion plots may, however, be much
more useful in experiments with trees where effects are
allowed to accumulate over more time and where the
experimental areas are likely to be more heterogeneous
(because of larger plots and, therefore, a larger
experimental area and because of less mixing of the soils by
tillage). Serious consideration should be given to the use
of tree growth data from companion plots to improve the
precision of the studies.

6. Power of test.

The objectives should include a statement of the desired precision of the estimates and the desired power for critical tests of significance. These considerations dictate the size of the experiment -- the number of replications, the number of test environments and the number of times the studies are repeated. Frequently, precision and power are not explicitly considered until after the study has been completed, if then, and the data are being analyzed. Relative precision and power of different experimental designs can be compared before the experiment is run, and absolute power can be computed if an estimate of the coefficient of variation or the experimental error is available. With this information, experiments can be designed with sufficient power to meet the objectives. Much of the controversy over scientific findings arises because of an inadequate power in the designed experiments to consistently detect real effects, an underdesigning of the experiments. As a consequence, the significant results in some studies and insignificant results in others are interpreted (incorrectly) as inconsistencies in the process; not as inadequacies in the research.

Computation of power or, equivalently, the required number of replicates, will quickly reveal whether the objectives are reasonable with the available resources. A seemingly logical and economical study to assess the impact of air pollution would use only ambient air and charcoal filtered plots with several replicates of each. How many replicates are needed with this design to detect a reasonably high 5% change in productivity between these two levels of pollutant? If we assume the coefficient of variation is 10% and that any change in productivity will be a decline, one needs approximately 22 replications to have just a 50:50 chance of detecting the difference; fifty replicates are needed to have a probability of .80 of detecting the change. If the change to be detected is dropped to 2.5%, 200 replications are needed to have an 80% chance of detecting the change. Clearly, such a study should not be undertaken unless the expected change is considerably greater than 5% or the coefficient of variation is lower than 10%.

ANALYSIS OF DATA.

There are two points I wish to emphasize concerning the analysis of data. The first point is to encourage the use of response models and regression analyses to characterize

the response of quantitative variables to pollutants rather
than using simple means comparison procedures. A good
response model, one which adequately characterizes the plant
response, greatly increases the precision of the predicted
yield losses over using simple comparison of two treatment
means. Again, I use the Weibull model from NCLAN for
illustration of this concept. Suppose one had the
objective of predicting yield loss for one specific dose
interval, say between 0.025 and 0.06 ppm. One simple
experimental design would place half of the experimental
units at each end of the dose range of interest, at .025 and
.06. The predicted relative yield loss would be estimated
as the difference between the two means divided by the mean
for the lower dose. Note that all of the experimental
effort is concentrated on this one interval; no information
is available on any other interval. Dassel's study of
design of experiments for the Weibull response model has
shown, however, that much higher precision is obtained, for
this same interval, from fitting the Weibull model and
predicting the relative yield loss from the fitted Weibull.
In addition, the fitted Weibull dose response model provides
information for all intervals within the limits of the
study, not just the .025 to .06 interval. Use of
appropriate dose response models can greatly increase the
amount of information realized from an experiment.

Secondly, and finally, I want to put in a plug for statistics. Some of you probably think that most of this talk has been a plug. Here, however, I want to push for the direct involvement of statisticians in the conduct of the experiments as team members, not simply as the statistician to whom the graduate student or technician is sent when a problem arises. Obviously, this point is a little out sequence since the statistician as a team member is going to be involved in all of the points we have discussed. I introduce it at this point because I want to encourage an activity that falls primarily under the purview of the statistician.

The program should be structured so that one individual (a statistician directly involved in the program) is given the responsibility of thoroughly and critically analyzing all data. If started early, this has the tremendous advantage of feed-back to the researchers of additional information needed for modeling, of inadequacies in the experimental designs or conduct of the experiments, or of inadequate power of experiments as designed. It also has the advantage that one individual has the responsibility of asking questions which are frequently ignored; are the models adequate, are the basic assumptions of the analysis

satisfied, do the residuals behave properly, etc. Such questions frequently can be better answered from a coordinated analysis of several data sets than from the individual experiments. For illustration of this point, I have been concerned from the beginning of the NCLAN studies that the variance associated with the yield responses might not be homogeneous. This concern arose from the often observed phenonmena of the variance increasing with the level of performance. However, no solid indication of heterogeneous variance was ever obtained from any one study. Recent analysis by Virginia Lesser of eight sets of soybean data from North Carolina (Heagle), with 74 degrees of freedom for the pooled residuals, provides fairly clear indication that there is a tendency for the variance to increase with the mean and that a log transformation would be helpful. At the same time, this larger data set provided no indication that the Weibull model was inadequate. Thus, analyses of the combined data sets has both highlighted a heretofore undetected problem and provided support for the response model being used.

SUMMARY

There are many aspects of experimental design which could be discussed today. I have tried to mention a few

which I have found to be particularly important and often
not given the attention they deserve. I would like to
summarize this presentation by listing these key points.

1. Clear specification of the objectives is a requisite
   for good experimental design. This must include an
   understanding of the population of environments to
   which the inferences are to be made. Decisions must
   be made, often with limited information, as to which
   environmental factors are most likely to influence
   the response to the pollutant so that they might be
   studied and their effects included in the response
   models.

2. The greater efficiency of factorial experiments
   should not be overlooked. Partial replicates of 2^n
   factorials serve as useful screening studies to
   identfy the important interacting environmental
   factors. The key environmental factors must then be
   incorporated in the dose response studies if their
   effects on the dose response model are to be
   understood and taken into account.

3. There is no need to have an excessively large number
   of distinct levels of the pollutant in order to
   characterize the dose response curve. More and
   better information is obtained if a minimal number of
   levels is used with a corresponding increase in

numbers of replications or numbers of environmental factors included in the study.

4. The response curve is estimated best if the design points cover a wide enough dose range to allow a major part of the dose response curve to be realized. This range is usually well beyond the levels of the pollutant expected in nature. The optimum placement of these points is dependent on the nature of the response curve. A few additional points and equal distribution of the design points over the interval seem to protect against misguessing the response curve.

5. Preliminary information on the soils and the plants at the test sites can improve the precision of the experiments by facilitating the placement of plots and blocks and as covariate information in the analyses of data. This may be even more useful for perennial crops than for annual crops.

6. The power of a study for specific hypotheses should be considered before the study is run. Computations of power will frequently alert one to inadequate experiments or overly ambitious objectives.

7. Dose response models and regression analyses should be used to characterize the response to the pollutant whenever possible. In general, more precise

predictions of effects can be made from appropriate response equations than from simple comparison of means.

8. The planning of a major research effort should include early and coordinated analyses of all results. The combined analyses provide opportunities for finding inadequacies in the research and the models which might otherwise go undetected.