PLANNING A SAFETY STUDY OF AN AGRICULTURAL PRODUCT: EFFECTS OF LAND APPLICATION OF PHOSPHOGYPSUM ON RADON FLUX

Ramon C. Littell

Sudeep Kundu

Recommended Citation

PLANNING A SAFETY STUDY OF AN AGRICULTURAL PRODUCT: EFFECTS OF LAND APPLICATION OF PHOSPHOGYPSUM ON RADON FLUX

Ramon C. Littell and Sudeep Kundu
Department of Statistics
Institute of Food and Agricultural Sciences
University of Florida
Gainesville, FL 32611

Abstract

Traditional agricultural research has been concerned largely with demonstrating that new products or new practices increase yield from plants or animals; i.e. that a change has occurred. Concepts of experimental design have been effectively employed in production-agriculture research planning to control extraneous variation and thereby reduce experimental error. Good data analysis practices have been employed to control type 1 error rate and to correctly compute errors of estimation. In recent years, increased emphasis has been placed on food safety and environmental impact of agricultural products. Studies of these issues are concerned with measuring small effects with required precision or establishing upper bounds on the effects. Statistical emphasis is on limiting the margin of error and the type 2 error rates. This paper discusses these concepts in the context of an environmental study of effect of phosphogypsum (PG) on radon flux. An experiment in progress revealed essentially no statistically significant effect of the phosphogypsum. Two statistical questions were then raised: 1) How large of an effect would have been detected in the study? and 2) How should a future study be conducted that would produce measurements of the effect with specified precision? A retrospective power analysis was performed to estimate the minimum detectable effect (MDE) in the existing study in response to question 1. In response to question 2, a new study was designed with required numbers of plots and measurements to meet precision and power objectives, using variance component estimates from existing data.

1. Introduction.

Phosphogypsum is a by-product of the phosphate fertilizer and phosphoric acid manufacturing industries. Large quantities of phosphogypsum have accumulated in Florida. The material has been considered as a potential source of sulphur and calcium for soils deficient in these minerals. In particular, it is being considered as a soil amendment for Florida pastures (Rechcigl, Roessler, and Alcordo, 1991). However, uranium and its decay series are associated with phosphate rock, resulting in the presence of radium-226 in phosphogypsum and the exhalation of radon-222 (called radon flux). The concentration of radium-226 in phosphogypsum derived from Central Florida rock is approximately 20 to 30 $pCi\ g^{-1}$, and the radon flux is on the order of 20 to 30 $pCi\ m^{-2}\ s^{-1}$ (Roessler, 1987).

Experiments are in progress at the UF/IFAS Agricultural Research and Education Center at Ona, Florida to evaluate the beneficial effects of land application of phosphogypsum (PG) on bahiagrass and ryegrass pasture lands, and other experiments are in place to assess environmental effects (Rechcigl et al, 1991). The present paper deals with an environmental study of effects of phosphogypsum application on a bahiagrass pasture. The basic experiment was a randomized complete blocks design, with two blocks each containing four quarter-acre plots. Phosphogypsum rates of 0.0, 0.4, 2.0 and 4.0 $Mg\ ha^{-1}$ were randomly assigned to the plots within each block. Numerous radiation measurements were made of airborne radon, radon flux, and radionuclides in soil, water, and vegetation. Generally, levels of increased radiation were not detected statistically. This raised two questions: 1) What level of increased radiation would have been detectable with the experiment? and 2) How should a new study be conducted that would ensure estimation of increased radiation with desired precision? These questions will be addressed in relation to radon flux.

2. Results from the bahiagrass study.

Radon flux measurements were obtained using Large-area activated charcoal canisters (Hartley and Freeman, 1985). Two samples per plot were obtained by exposing canisters for 24 hours and shipping collections for laboratory analysis. Results from five collections were available, yielding 80 measurements in all (2 blocks x 4 PG levels x 5 collections x 2 samples/plot/collection). A model for the radon flux data is

$$Y_{ijkl} = \mu + b_i + \tau_j + e_{ij} + c_k + (cb)_{ij} + (c\tau)_{ik} + (cb\tau)_{ijk} + \epsilon_{ijkl},$$

where

b_i = random effect of ith block,

τ_j = fixed effect of jth treatment (PG rate)

e_{ij} = random error associated with plots within blocks

c_k = random effect of kth collection time

$(cb)_{ik}$ = random interaction of coll k, blk i

$(c\tau)_{jk}$ = random interaction of coll k, trt j

$(cb\tau)_{ijk}$ = random interaction of coll k, blk i, trt j

ϵ_{ijkl} = random effect of lth sample in blk i, trt j,
              at coll k.

All random effects are assumed independent with mean zero and respective variances σ_B^2, σ_P^2, σ_C^2, σ_{CB}^2, σ_{CT}^2, σ_{PC}^2 and σ_S^2. Since a Blk*Trt combination refers to a particular plot, the variance components for Trt*Blk and Coll*Trt*Blk are subscripted with P and PC, respectively. Notice that "collection" is treated as a random effect, although in standard repeated measures experiments it would normally be considered a fixed effect. In this experiment, collections were taken at arbitrary, if not random, points in time. The objective of sampling repeatedly over time was to provide time replication rather than to estimate time trends. These considerations are more consistent with objectives of random effects than fixed effects. An analysis of variance of the radon flux data based upon the model is shown in Table 1.

Notice from the expected mean squares in Table 1 that treating collection as a random effect results in the CT variance component σ_{CT}^2 appearing in the expected mean square for Treatments. However, since MS(Coll*Trt)>MS(Coll*Trt*Blk), the ANOVA estimate of σ_{CT}^2 is negative, so its presence henceforth will be ignored. Thus MS(Trt*Blks) becomes an appropriate error term for testing Treatment effects averaged over collections.

Radon flux means for the four PG rates and five collections are presented in Table 2 and Figure 1. Standard error estimates for comparing PG means in Table 2 were computed from variance component estimates derived from the analysis of variance shown in Table 1.

No clear evidence of increased radon flux due to PG application rate is present in Tables 1 and 2. The pertinent issue, however, is not whether there was an increase in radon flux due to PG application, but rather the amount of increase. Due to the magnitude of experimental error and the limited

number of replications in the experiment, it is not possible
to make a meaningful assessment of the amount of increase in
radon flux. An alternative is to place an upper limit on the
amount of flux increase that could have occurred with
reasonable probability. A confidence interval seems to be a
logical approach to constructing such a limit. This requires
identification of a parameter estimate about which to
construct the confidence interval. One possible parameter
estimate is the slope of a regression of flux on PG rate.
Other possibilities include differences between mean flux at
the non-zero PG rates and mean flux at the zero PG rate.
Still another possibility is the difference between the
averages of the "high" PG rates (2.0 and 4.0) and the "low" PG
rates (0.0 and 0.4). These estimates, with standard errors
and upper 95% confidence limits are presented in Table 3.
Clearly, all of these estimates are small and none is
"significantly different from zero."   This conclusion,
coupled with the fact that there is only a suggestion of a
numerical trend, detracts from any of these estimates as the
basis for a confidence interval. Consequently, the use of
confidence intervals as a means of establishing an "upper
limit" for the PG effect is more or less meaningless. It can
only be concluded from the data that the PG effect was too
small to be detected in this experiment. We therefore return
to the analysis of variance as a means to estimate components
of random variation in the data.

If PG treatment effects are negligible, then the sources of
variation labeled Blocks, Treatments, and Trt*Blk in Table 1
can be combined into a single source which measures "Plot"
variation.   Likewise, Coll*Blk, Coll*Trt, and Coll*Trt*Blk
sources can be combined into "Plot*Collection" variation.
These are all random effects, measuring variation in the
sampling scheme. The revised ANOVA is given in Table 4, along
with estimates of the variance components.

3. Minimum detectable PG effect in present study.

Failure to detect or measure the PG effect raises the question
"How large of an effect would have been detected with this
experiment?"   This question can be addressed by a
retrospective power analysis, as described by O'Brien and
Muller (1992).   It also is closely related to the minimum
detectable effect (MDE) of Lynch (1992).

We shall consider the difference between the mean flux for a
given nonzero PG rate (which will generically be labelled
"Trt") and the flux for a zero PG rate (labelled "Con") as the
parameter of interest, and refer to it as "δ." This leads to
a one-sided hypothesis testing situation of $H_o: \delta = 0$ versus
$H_a: \delta > 0$. An appropriate test statistic is

$$T = (\overline{Y}_{Trt} - \overline{Y}_{Con}) / (2MS(P)/20)^{1/2}$$

Quantitatively stated, the problem is to find the minimum value of δ that would produce a specified power (say, 0.8) for an α-level t-test of H_o: $\delta=0$ versus H_a: $\delta>0$. That is, we want δ such that

$$Power(\delta) = P\{T > t(\alpha,df) \mid \delta\} = 0.8,$$

where $t(\alpha,df)=2.353$ is the upper $\alpha=0.05$ critical value of the central t distribution with df=3 degrees of freedom.

The power of the test is actually a function of the non-centrality parameter

$$\lambda = \delta / \sqrt{Var (\overline{Y}_{Trt} - \overline{Y}_{Con})},$$

where $Var(\overline{Y}_{Trt} - \overline{Y}_{Con}) = 2(\sigma_s^2 + 2\sigma_{PC}^2 + 10\sigma_P^2)/10$ is the "true" variance of $\overline{Y}_{Trt} - \overline{Y}_{Con}$, rather than δ. This "true" variance is unknown, but it is common practice to use an estimate in its place. An unbiased estimate of the "true" variance is $(2/20)*MS(Plot) = 0.000565/10$ from the ANOVA in Table 4. This gives $1/2(0.000565/10) = 0.007517$ as an estimate of the "true" standard error in the denominator of lambda. The power, as a function of δ, and using the estimated variance, can now be computed and plotted as shown in Figure 2. The MDE is then obtained by solving the equation Power(δ) = 0.8, giving MDE = 0.024801. We conclude that, if the "true" difference between the treatment and control mean were 0.024801 or larger, then the statistical test would have rejected H_o: $\delta = 0$ with probability 0.8. In this sense, the calculated MDE = 0.024801 is an estimate of the smallest difference that would have been detected.

A simpler calculation can be used as an approximation to MDE. An approximate sample size formula (in this case, sample size refers to the number of plots per mean) for a t test to detect a difference of δ is given by

$$r = ((t(\alpha,df) + t(\beta,df))\sigma_{Diff}/\delta)^2,$$

where $\sigma_{Diff}^2 = 2(\sigma_s^2 + 2\sigma_{PC}^2 + 10\sigma_P^2)/10$ is the variance of a difference between two plot means (averaged over five collections and two samples per plot). In the present bahiagrass study, r = 2. Using $(2/10)*MS(Plot) = 0.000565/5$ from the ANOVA in Table 4 as an estimate of σ_{Diff}^2 and solving for $\delta = MDE$ yields

$$MDE = (t(\alpha,df)+t(\beta,df))*(2/10)*MS(Plot)/2)^{1/2}$$

$$= (2.353 + 0.978)*0.007517$$

= 0.025039.

This is a very good approximation to MDE = 0.024801 calculated from the power function. Also, it is notable that the approximate MDE can be easily calculated from the usual LSD for comparing two means:

MDE = (t(α,df)+t(β,df))/t(a/2,df)*LSD.

These calculations have used an estimate of experimental error variance, MS(Plot) as if it were the "true" experimental error variance, without accounting for uncertainty in estimation of the variance. The result will be underestimation or overestimation of MDE depending on whether MS(Plot) is an underestimate or overestimate of EMS(Plot). A conservative estimate of MDE can be obtained by using an upper confidence limit on EMS(Plot) instead of MS(Plot) in place of the "true" EMS(Plot). The standard upper 95% confidence limit for EMS(Plot) yields the corresponding lower 95% confidence bound for the power function, as shown in Figure 2.

4. Strategy for new study.

A new study has been proposed which will focus on safety aspects of land application of phosphogypsum, particularly regarding radon flux. It will be planned to achieve certain statistical objectives in terms of powers of tests and precision of estimation.

There were four essential stages in planning the new study:

A: Select basic experimental design.

B: Select PG rates to be studied.

C: Define statistical criteria in terms of power and precision.

D: Determine feasible sampling schemes and optimize.

These stages will be discussed in turn.

A. <u>Selection of basic experimental design</u>. The original study was run in a randomized complete blocks design. Analysis of variance of the radon flux data revealed no statistically significant evidence of block variance in the original study. However, there were only two blocks, so this cannot be interpreted as evidence that, indeed, there were no block differences. In addition, there is evidence of significant variation between plots within blocks. Smaller plots are envisioned for the new study than for the original study, meaning that blocks in the new study would be more nearly the

size of plots in the original study. Thus a randomized blocks
design was chosen for the new study to control for this
potential source of variation. Whether the blocks will be
complete (contain all treatment levels) or incomplete will
depend on the number of PG application rates, i.e. the
treatment levels.

B. Selection of levels of PG to be studied. This stage was
concerned with deciding what actual rates of PG should be used
in the new study; that is, with choosing treatment levels.

The maximum PG rate of 4.0 Mg ha^{-1} used in the initial study
was too low to be detected without very large sample sizes, so
a higher maximum rate must be used in order to be
statistically measurable with meaningful precision in an
experiment that is financially feasible. Also, PG rates as
high as 20 Mg ha^{-1} were discovered in use in other agricultural
applications. Moreover, in safety studies it is often
advisable to investigate rates of the material in question
greater than those anticipated for actual use. These
considerations led to the decision to use an upper rate of 20
Mg ha^{-1}.

Control rates of 0.0 Mg ha^{-1} were considered necessary in the
experiment. The decision remained of what rates to use
between 0.0 and 20 Mg ha^{-1}. Linear response to PG application
rate is a reasonable assumption for radon flux, which would
imply using only the highest (20 Mg ha^{-1}) and the lowest (0.0
Mg ha^{-1}) PG rates. Nonetheless, it was decided to run an
intermediate rate of 10 Mg ha^{-1} to reveal nonlinear trends.
Thus the experiment will have three treatments; PG application
rates of 0.0, 10.0, and 20.0 Mg ha^{-1}.

C. Specify statistical criteria in terms of power and
precision. Two general criteria were developed concerning the
two non-zero PG application rates. First, the experiment
should provide sufficient information to make the effect of
the 10 Mg ha^{-1} application rate statistically detectable. That
is, the minimum detectable effect (MDE) should be less than or
equal to 10 Mg ha^{-1}. Equivalently, the power of a statistical
test comparing the mean radon flux at 10 Mg ha^{-1} with the mean
radon flux at 0.0 Mg ha^{-1} should be greater than 0.80. Second,
following EPA guidelines suggested in Hartley and Freeman
(1985), the experiment should provide sufficient information
to make the margin of error (MOE) of estimating the linear
rate effect less than 25% of the effect itself. This is
equivalent to requiring that the MOE of the difference between
the mean flux at 20.0 Mg ha^{-1} and the mean flux at 0.0 Mg ha^{-1}
be less than 25% of the expected difference between the means.

Engineering calculations produce anticipated radon flux
increases of 0.019 pCi m^{-2} s^{-1} for a PG application rate of 10

Mg ha^{-1} and 0.038 pCi m^{-2} s^{-1} for a PG rate of 20 Mg ha^{-1}. Thus
the criteria become

MDE < 0.019

and

MOE < 0.25*0.038 = 0.0095.

D. <u>Determine feasible combinations and optimize</u>.   Either
criterion (test power or estimation precision) will entail
sampling sufficiently intensively and extensively to attain a
bound on the variance of the difference between two treatment
means. As in section 3, we generically denote the variance of
the difference as $\mathrm{Var}(\overline{Y}_{Trt}-\overline{Y}_{Con})$.

The (approximate) minimum detectable effect (MDE) and margin
of error (MOE) are given by the expressions

$$\mathrm{MDE} = (t(\alpha,df)+t(\beta,df))*(\mathrm{Var}(\overline{Y}_{Trt}-\overline{Y}_{Con}))^{1/2}$$

and

$$\mathrm{MOE} = t(\alpha/2,df)*(\mathrm{Var}(\overline{Y}_{Trt}-\overline{Y}_{Con}))^{1/2}.$$

For $\alpha=0.05$ and $\beta=0.20$, the ratio MDE/MOE is approximately 1.2
for a wide range of df values. Thus the criterion MOE<0.0095
is more stringent than the criterion MDE<0.019, so we need
only be concerned with the bound on MOE.

The variance $\mathrm{Var}(\overline{Y}_{Trt}-\overline{Y}_{Con})$ will be estimated using variance
component estimates from the existing study.

In terms of the variance components,

$$\mathrm{Var}(\overline{Y}_{Trt}-\overline{Y}_{Con}) = 2[\sigma_S^2/src+\sigma_{PC}^2/rc+\sigma_P^2/r],$$

where r = number of plots per treatment (blocks), c = number
of collection times, and s = number of samples per plot at a
given collection time.

Initially, for sake of illustration, we take s = 2, as in the
existing study. The variance expression then becomes

$$\mathrm{Var}(\overline{Y}_{Trt}-\overline{Y}_{Con}) = 2[(\sigma_S^2+2\sigma_{PC}^2)/2rc+\sigma^2/r].$$

In terms of expected mean squares,

$$\mathrm{Var}(\overline{Y}_{\mathrm{Trt}} - \overline{Y}_{\mathrm{Con}}) = 2[(\mathrm{EMS(P)} - \mathrm{EMS(P*C))}/10r - \mathrm{EMS(P*C)}/2rc]$$

$$= 2[\mathrm{EMS(P)} + (5\text{-}c)\mathrm{EMS(P*C)}/c]/10r$$

$$= [\mathrm{EMS(P)} + (5\text{-}c)\mathrm{EMS(P*C)}/c]/5r,$$

where EMS(P) and EMS(P*C) are the expected mean squares from Table 4 for Plot and Collection, respectively. Thus an estimate of the variance for a specific r and c is obtained by inserting the mean squares in place of the expected mean squares. For example, with r = 12 and c = 6, one obtains the estimated variance

$$\mathrm{Var}(\overline{Y}_{\mathrm{Trt}} - \overline{Y}_{\mathrm{Con}}) = [0.000565 + (5\text{-}6)0.000239/6]/60$$

$$= 0.000009.$$

Our problem is to determine the number of blocks r and collections c to meet the stated objective MOE<0.0095. The combination of r=12, c=6 and s=2 satisfies the above inequality, with an MOE of 0.00614. But the combination of r=12, c=6 and s=1 is almost as good with an MOE of 0.00650, while it would reduce the cost of the radon flux measurements by almost half. It is also notable that r=6, c=12 and s=1 yields MOE=0.00708. This reveals the fact that increases in space replication (numbers of plots per treatment) are more effective than increases in time replication (numbers of collections) and increased subsampling (number of samples per plot per collection.)

At this point, other considerations take over. The scientists believed that combinations requiring approximately 70 samples were cost feasible. Also, they believed that the maximum number of space replications (blocks) that could be managed was 12. Calculation of MDE and MOE for a variety of numbers of blocks r and collections c with rc approximately equal to 70 shows that the best choice is r=12 and c=6.

5. Summary and Conclusions.

It is often the case that an initial attempt to assess effects of a product released in the environment are done in studies that are not sufficiently large to measure the effect. Results from these initial studies are nonetheless useful to determine the magnitude of effect that would have been detected, and also to determine sample sizes needed in future studies that will provide meaningful measures of the effects. Determination of the sample sizes requires specification of the degree of precision desired and estimates of variation from the various sources. These principles have been

illustrated with results from a phosphogypsum land application study of effects on radon flux and other radiological measurements. The illustration has been in terms of radon flux. Final plans for a proposed study would consider cost and objectives regarding the other radiological measurements.

References

Hartley and Freeman (1985). "Radon flux measurements on Gardinier and Royster phosphogypsum piles near Tampa and Mulberry, Florida,"

Lynch, R. O. (1993). "Minimum Detectable Effects for 2^{k-p} Experimental Plans," Journal of Quality Technology, in press.

O'Brien, R. G., and Muller, K. E. (1992). "Unified Power Analysis for t Tests through Multivariate Hypotheses," to appear in Applied Analysis of Variance in the Behavioral Sciences, New York : Marcel Dekker.

Rechcigl, J.E., Roessler, C.E. and Alcordo, I. S. (1991). "Influence of Phosphogypsum on forage yield and quality, and on the environment in a typical Florida spodosoil," progress report to Florida Institute of Phosphate Research, March 1991.

Roessler, C. E. (1990). "Control of Radium in phosphate mining, beneficiation and chemical processing," Technical Reports Series No. 310, The Environmental Behavior of Radium, International Atomic Energy Agency, Vienna, 1990.

Table 1. Analysis of variance of radon flux data.

Source of Variation	df	Mean Squares	Expected Mean Squares
Blocks	1	0.000673	$\sigma_S^2+2\sigma_{PC}^2+10\sigma_P^2+8\sigma_{CB}^2+40\sigma_B^2$
Treatments	3	0.000548	$\sigma_S^2+2\sigma_{PC}^2+10\sigma_P^2+4\sigma_{CT}^2+20\sigma_T^2$
Trt*Blk	3	0.000546	$\sigma_S^2+2\sigma_{PC}^2+10\sigma_P^2$
Collections	4	0.003179	$\sigma_S^2+2\sigma_{PC}^2+8\sigma_{CB}^2+4\sigma_{CT}^2+16\sigma_C^2$
Coll*Blk	4	0.000339	$\sigma_S^2+2\sigma_{PC}^2+8\sigma_{CB}^2$
Coll*Trt	12	0.000219	$\sigma_S^2+2\sigma_{PC}^2+4\sigma_{CT}^2$
Coll*Trt*Blk	12	0.000227	$\sigma_S^2+2\sigma_{PC}^2$
Sampling	40	0.000076	σ_S^2

Table 2. Radon flux means (pCi $m^{-2}s^{-1}$) for four rates of PG application (Mg ha^{-1}) and five collection times (October 1990-January 1992). Entries are based on four measurements from two blocks and two samples per plot per collection.

PG Rate	Collection 1	2	3	4	5	Mean
0.0	0.034	0.043	0.028	0.014	0.047	0.033
0.4	0.033	0.036	0.024	0.024	0.052	0.034
2.0	0.037	0.045	0.041	0.020	0.080	0.045
4.0	0.041	0.030	0.035	0.027	0.059	0.038
Mean	0.036	0.039	0.032	0.021	0.060	0.038

	Std. Error
Comparing PG means averaged over collections:	0.007517
Comparing Collection means averaged over PG rates:	0.019934
Comparing PG means at same collection:	0.012333
Comparing Collection means at same PG rate:	0.022068

Table 3. Estimates of PG treatment effects and standard errors.

Effect	Effect Estimate	Standard* Error	Upper 95% Confidence Limit
Slope	0.002	0.0017	0.0060
0.4 - 0.0	0.001	0.0075	0.0186
2.0 - 0.0	0.012	0.0075	0.0296
4.0 - 0.0	0.005	0.0075	0.0226
Trt - Con	0.006	0.0061	0.0201
"Hi" - "Lo"	0.008	0.0053	0.0285

*Note computation of variance ($\hat{\sigma}^2_{CT} < 0$, so set $\hat{\sigma}^2_{CT} = 0$).

Table 4. Condensed analysis of variance for radon flux.

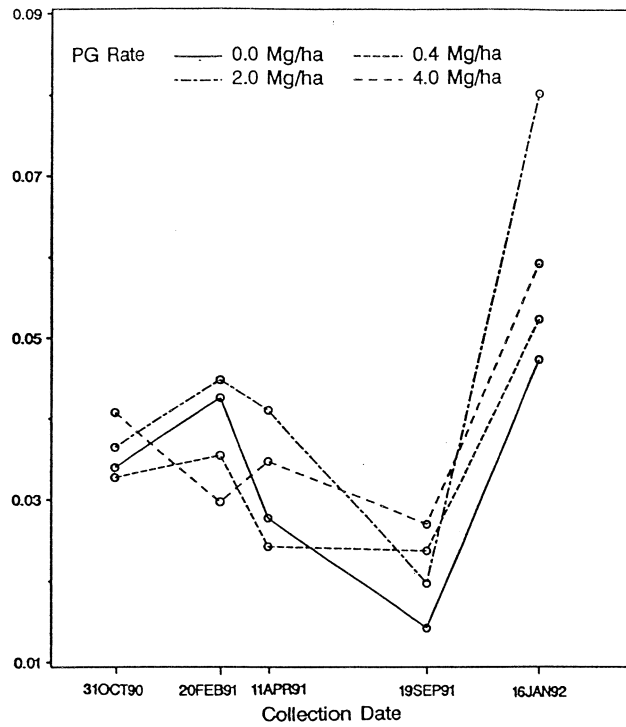| Source of Variation	df	Mean Squares	Expected Mean Squares	Variance Estimate
Plot	7	0.000565	$\sigma^2_S + 2\sigma^2_{PC} + 10\sigma^2_P$	0.000033
Coll	4	0.003179	$\sigma^2_S + 2\sigma^2_{PC} + 16\sigma^2_C$	0.000184
Plot*Coll	28	0.000239	$\sigma^2_S + 2\sigma^2_{PC}$	0.000081
Sampling	40	0.000076	σ^2_S	

Figure 1. Radon Flux Means (pCi m^{-2} s^{-1}) for
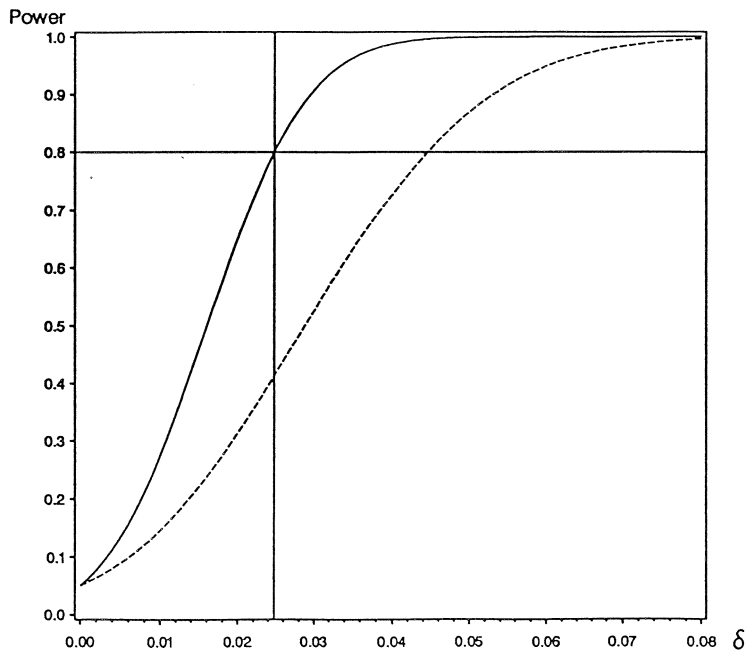Four PG Rates at Five Sampling Times

Figure 2. Power of t-test for Comparing Two PG
Rate Means when True Difference is δ