

Whole-Profile Soil Carbon Stocks: The Danger of Assuming Too Much from Analyses of Too Little

A. N. Kravchenko*

Dep. of Crop and Soil Sciences
Michigan State Univ.
East Lansing, MI 48824-1325

G. P. Robertson

W.K. Kellogg Biological Station
and
Dep. of Crop and Soil Sciences
Michigan State Univ.
East Lansing, MI 40024

Whole soil profile assessments of soil C stocks are enormously important for evaluating a soil's potential for sequestering C. Carbon stock measurements are often highly variable, which makes the detection of statistically significant differences among different land uses and management systems difficult. A common mistake is to interpret a lack of statistical significance as evidence for the absence of differences. This mistake is costly: failure to identify practices and habitats that accumulate C can affect our understanding of soil C cycling and hide possible sequestration potentials. In this study, we investigated inadequate replication as a major cause of a lack of statistical significance. Power analysis is a well-established statistical tool that allows researchers to infer whether the lack of statistical significance is due to the absence of meaningful differences or due to insufficient replication. We used data from two published studies to illustrate the value of power analysis. In particular, we showed that in these studies the chances of detecting even substantial (e.g., 50%) differences among experimental treatments can be as low as 20 to 60% due to the natural variability in soil C concentrations and stocks. Carbon variability in different soil horizons suggests that statistical analyses of soil C stocks should be conducted separately for each sampled horizon and that conclusions about the profile as a whole should be made based on the individual horizon results. Devising an optimal sampling strategy that identifies the appropriate number of replicates needed to detect significant C change by horizon is essential.

Abbreviations: CRD, completely randomized design; CT, conventional tillage; LTER, Long-Term Ecological Research; NT, no-till; RCBD, randomized complete block design.

Whole-profile assessments of soil C change are increasingly important for estimating contributions of various management practices and land uses to soil's potential for supporting crops, sequestering C, and contributing in general to a sustainable biosphere (VandenBygaart and Angers, 2006; Baker et al., 2007; Grandy and Robertson, 2007; Richter and Mobley, 2009). Soil C is, however, especially variable at depth, and high variability in deeper soil horizons complicates the task of detecting statistically significant differences even where such differences exist.

The danger of missing significant changes in soil C at depth is one thing; the potential for this variability to obscure changes in other, less variable parts of the soil profile is another. Both can lead to errors in our understanding of terrestrial C cycling and to missed opportunities for—and misappropriation of—management approaches to soil C storage.

A number of recent studies reported no statistically significant differences in whole-profile soil C stocks due to no-till and other management practices (e.g., a survey by VandenBygaart et al., 2003). When a statistical test demonstrates a lack of statistical significance, however, it cannot be taken as sufficient evidence for the absence of differences. In such a case, there is a possibility of committing a Type II error, that is, of declaring that there are no differences when in fact they are present. Researchers must evaluate how high the probability of a Type II error is, given a hypothetical difference of a certain size. Only when Type II error is sufficiently low can they conclude that the difference of the hypothesized size is indeed absent.

Soil Sci. Soc. Am. J. 75:2010

Posted online 8 Nov. 2010

doi:10.2136/sssaj2010.0076

Received 13 Feb. 2010.

*Corresponding author (kravche1@msu.edu).

© Soil Science Society of America, 5585 Guilford Rd., Madison WI 53711 USA

All rights reserved. No part of this periodical may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Permission for printing and for reprinting the material contained herein has been obtained by the publisher.

Low Type II error means high power of the statistical test. This evaluation forms the basis of the post hoc power analysis. It allows a cause for the lack of statistical significance to be inferred: the absence of biogeochemically significant differences or simply an inadequate sample replication. Such assessment becomes especially important for soil attributes with high intrinsic variability. A better approach commonly recommended by statisticians is an ad hoc power analysis conducted before experiment layout or sampling. The ad hoc power analysis provides the optimal number of replications or samples that must be taken to detect as statistically significant a hypothesized biogeochemically significant difference of a certain size with acceptably low probabilities of both Type I (e.g., $\alpha = 0.05$) and Type II errors. This will ensure that replication in the planned experiment or sampling effort will be adequate for detecting acceptable differences. Unfortunately, either ad hoc or post hoc power analyses are rare for studies dealing with whole-profile C concentrations and stocks.

A combination of high intrinsic variability and low sample replication can lead to low statistical power, rendering experimental results of limited value. In fact, it is alarming to consider that many studies of soil C have been performed under scenarios where the design of the experiment made it almost impossible to detect even large differences in C. For example, see comments by VandenBygaart (2009) and sample number calculations reported in Yang et al. (2008).

Our objectives in this study were twofold. First, we investigated the value of post hoc power analysis using two published studies of soil C comparisons, thereby highlighting the difficulty of detecting whole-profile soil C changes where variability is high and statistical power concomitantly low. Second, we developed specific strategies for planning, sampling, and analyzing experiments for detecting soil C change.

MATERIALS AND METHODS

Data

The data for our illustration came from two sources. The main source was the study by Syswerda et al. (2011), who collected the data at the W.K. Kellogg Biological Station Long-Term Ecological Research (LTER) site in southwest Michigan, established in 1988. The experiment was a randomized complete block (RCB) design with six replicated plots for each treatment and five soil subsampling stations within each replicated plot. The experiment consisted of seven treatments representing a variety of agronomic management practices and land uses. The details on the experimental layout and management can be found at www.kbs.msu.edu/lter (verified 15 Sept. 2010). Measurements of soil C concentrations and bulk densities were conducted in 2003 by horizon to 100 cm, referred to here as surface (Ap horizon), middle (B/Bt horizon), and deep (Bt2/C) layers. For this illustration, we used C concentration and C stock data from only two of the LTER corn (*Zea mays* L.)–soybean [*Glycine max* (L.) Merr.]–wheat (*Triticum aestivum* L.) rotations, the chisel-plowed (CT) and no-till (NT) treatments, on account of general interest in conservation tillage effects on soil C sequestration. To ensure the most accurate assessment

of variability, however, the variability due to experimental blocks, plots, and subsampling stations was characterized using the entire data set.

The second source of data was the study by Blanco-Canqui and Lal (2008), conducted at 11 sites representing Major Land Resource Areas of the eastern United States. At each site, three samples were collected from adjacent conventionally plowed, no-till, and woodlot areas. Three samples were collected from each treatment at each site. Our data analyses for each site were conducted separately as a completely randomized design (CRD), as done in the original study, although a more preferable approach would have been to combine the data from all 11 sites and use site as a blocking criterion in a randomized complete block design (RCBD) analysis (Franzluebbers, 2009). For post hoc power analyses, we followed the same separate-site data analysis as the one reported in the study. Blanco-Canqui and Lal (2008) measured soil C concentrations and bulk densities at multiple depths and reported both C concentration and C stocks results. Because measures of variability for soil C stocks were not reported, however, we examined only the soil C concentration data for which variability information was available. We used data from depths of 5 to 10, 30 to 50, and 50 to 60 cm. For these depths, we recorded the mean soil C concentrations (in g kg^{-1}) and the LSD values from Blanco-Canqui and Lal (2008, Fig. 2). We only considered comparisons between two treatments, CT and NT; however, estimates of variability were obtained based on data from all three treatments.

Although our analysis was based on comparisons between only two treatments, the results can be easily extended to multiple-treatment and multiple-factor experiments.

Power Analysis

Detailed descriptions of the Type I and Type II errors, concepts of power, and power analysis can be found in most texts on statistical principles and experimental design (e.g., Ott and Longnecker, 2001). Quinn and Keough (2002) provided a comprehensive list of useful suggestions for power analysis applications in the biological sciences. Barker Bausell and Li (2002) offered a detailed description of the procedures and tables for power analysis application in a variety of experimental designs. Here we only briefly mention the key points relevant to the power analysis conducted in this study.

The questions that can be answered by an ad hoc or post hoc power analysis can be formulated in two different ways. Either we can ask:

Given the variability of the studied data and given the experimental settings, e.g., the numbers of replications and subsamples, blocking, multiple sites, etc., what is the probability (the power) that we will detect as statistically significant a certain change in C, if such change indeed has taken place?

Or we can ask:

Given the variability of the studied data and experimental settings, how many replications or subsamples should be or have been taken to be able to detect as statistically significant a certain change in C, if such change indeed has taken place, with a certain probability (power) (say, 90%)?

Both questions are basically two different ways of addressing the same point—was or will be the number of samples sufficient?

Figure 1 outlines the main steps of power analysis with considerations for each step. Steps 1 and 2 are to specify the hypothesized difference between the treatment means and to estimate variance components

relevant to the particular experimental design. The hypothesized differences are those judged to be of practical significance. Estimates of variance components should be conservative representations of the researcher's knowledge of sources of variability in the planned or conducted experiment. Once those two parameters are specified, the power (probability) of detecting the hypothesized difference, if it is indeed present, during a statistical test, can be calculated. Alternatively, we can determine the number of samples (replications or subsamples) needed in order for the experiment to have a specified level of power.

For the Syswerda et al. (2011) study, we used the measured data to estimate the three variance components describing the three random sources of variability present in a RCBD with subsamples: block variance, plot variance, and residual (subsample) variance (Table 1). For the Blanco-Canqui and Lal (2008) study, we followed the CRD

analyses conducted by the researchers separately for each site. The estimates of the residual variance were calculated for each site from the LSD values reported in the study. All power calculations were based on a 0.05 probability of Type I error. Power analysis was conducted using the PROC MIXED procedure of SAS (SAS Institute, 2001) based on the approach outlined by Stroup (2002).

RESULTS AND DISCUSSION

Power Analysis of Total Carbon Concentration Data by Depth

To identify the size of the change in the studied variable, e.g., the concentration of soil C that we believe would be of practical significance, we must take into account the natural distribution of soil C with depth. Of course it is impossible to identify a single

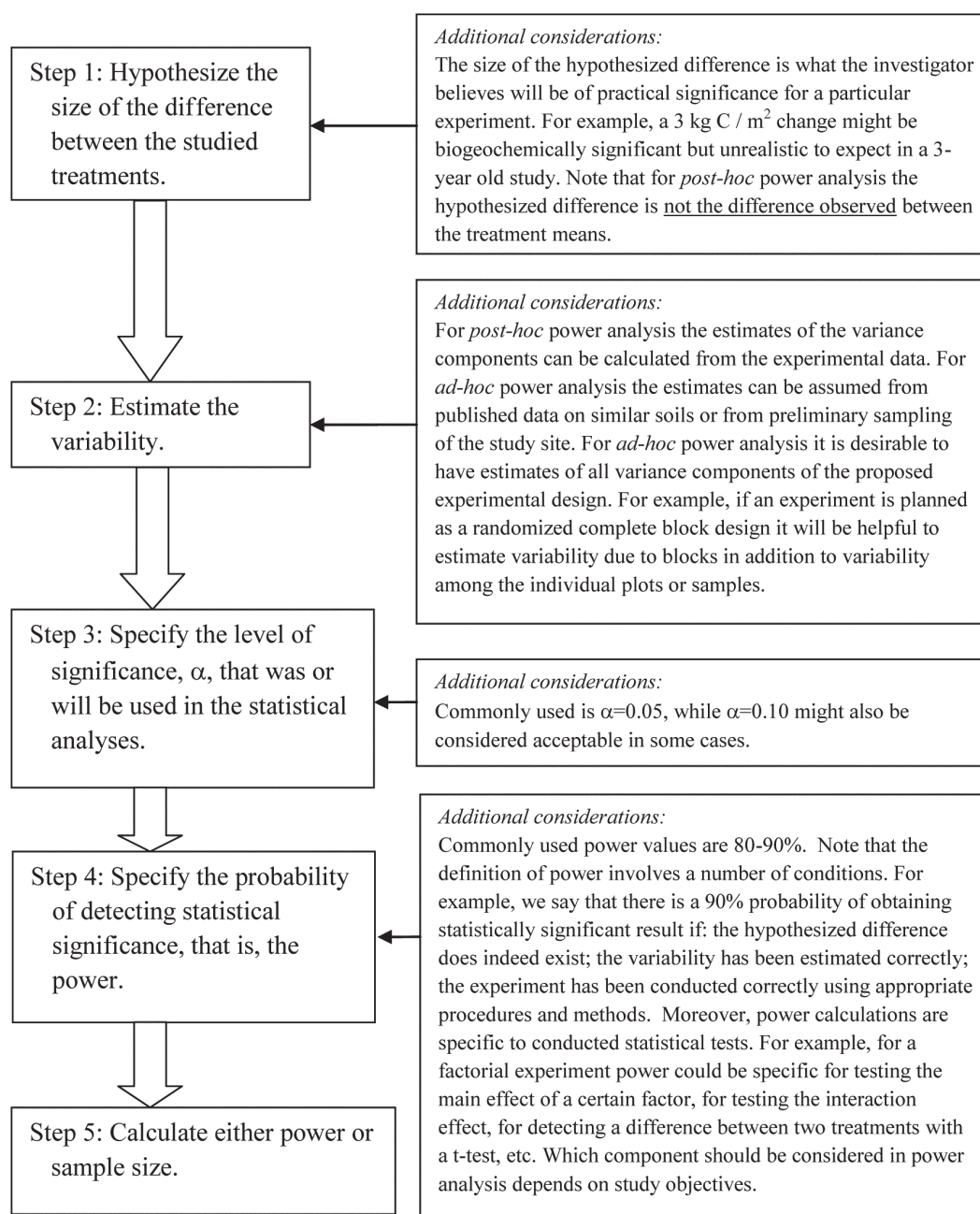


Fig. 1. Main steps of power analysis.

Table 1. Grand average values and estimates of variance components for total C concentrations and total C stocks used in the power analyses (from Syswerda et al., 2011).

Horizon	C concentration				C stock			
	Mean	Variances			Mean	Variances		
		Block	Plot	Residual		Block	Plot	Residual
	g kg ⁻¹				kg m ⁻²			
Surface (Ap), 0–20 cm	12.1	0.77	0.01	3.35	3.4	0.61	0.13	0.66
Middle (B/Bt), 30–40 cm	4.4	1.56	0.10	4.10	2.5	0.77	0.30	2.24
Deep (2Bt/C), 40–100 cm	1.9	0.09	0.01	1.19	1.4	0.19	0.30	1.58
Whole profile, 1 m					7.4	0.10	0.30	4.86

absolute value for a change in soil C concentration that would be suitable for every soil layer. For example, a change of 5 g kg⁻¹ in the total C concentration can be expected to take place in a soil surface horizon where average total C concentrations are around 10 to 15 g kg⁻¹. Such a change cannot be expected in the deeper soil layers, however, where average soil C concentrations are around 1 to 3 g kg⁻¹. Thus, it makes sense to specify changes in C on a relative basis.

We considered several sizes of change in the total C concentration, expressed as a percentage of the overall averages, e.g., changes of 10, 50, 100, and 200%. For example, in the Syswerda et al. (2011) study, if the average total C concentrations were equal to 12.1, 4.4, and 1.9 g kg⁻¹ in the surface, middle, and deep horizons, respectively (Table 1), then a 50% change relative to the average C in the three horizons would be 6.1, 2.2, and 0.9 g kg⁻¹, respectively.

In further calculations, we considered a hypothesized scenario that in the course of a long-term experiment, the total C concentration under CT would become lower than that under NT. Because the size, not the direction, of the difference in C matters in a comparison between the two treatments, however, an identical conclusion would hold regarding an experiment's power and sample sizes if soil C under CT were hypothesized to become greater than under NT.

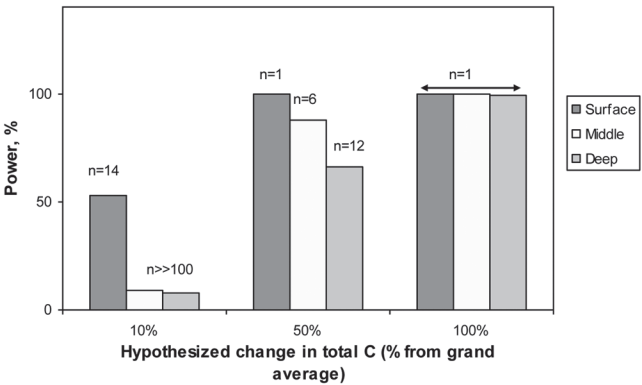


Fig. 2. Probability (power) of detecting a statistically significant ($\alpha = 0.05$) treatment effect in total C concentration from the Long-Term Ecological Research experiment (randomized complete block design [RCBD] with 30 samples: six replicate plots and five subsamples per plot) as a function of relative differences between no-till (NT) and conventional tillage (CT) treatments for surface, middle, and deep soil layers (based on data from Syswerda et al. 2011). The numbers above each bar represent the number of subsamples that need to be taken in each of six replicated plots of this RCBD experiment to detect the hypothesized differences between CT and NT treatments with 90% probability (power) at $\alpha = 0.05$.

The power values calculated for Syswerda et al. (2011) for three potential differences in total C concentration for the three studied layers are shown in Fig. 2. Note that the power values reported in Fig. 2 are applicable for the specific case of a RCBD experiment with six replications and five subsamples per experimental unit and under the assumption that the observed variability estimates (Table 1) correctly reflect the real variability. The probability (power) of detecting statistically significant differences is highest in the surface layer, followed by the middle and then deep layers. For example for the surface layer, if a difference of 10% in the total C concentration truly exists between CT and NT (far left bar in Fig. 2), there is a 52% chance of detecting this 10% change at $\alpha = 0.05$ with a total of 30 soil samples (6 replicate plots \times 5 subsamples per plot). Likewise for the middle layer, if a difference of 50% in the total C concentration truly exists between CT and NT, then there is an 88% chance of detecting a 50% difference in total C at $\alpha = 0.05$ with this replication scheme.

Alternatively, it is also possible to use this analysis for planning purposes. Because it is not possible to change the number of blocks or plots in this long-term field experiment, only the number of subsamples per plot can be adjusted. Power analysis shows that for a 90% chance (power) of detecting a truly existing 50% difference in middle horizon soil C, we would need to collect at least six subsamples from each replicated plot (Fig. 2). For the same probability of detecting a true 50% difference in deep horizon soil C, we would need to collect twice as many subsamples.

From a biogeochemical standpoint, however, it is extremely unlikely that a 50% deep-profile soil C change would occur after only 15 to 20 yr. Were we to wish instead to detect a more modest 10% change, the number of samples that must be taken unfortunately becomes substantially larger: $>> 100$ (Fig. 2).

The power values calculated for Blanco-Canqui and Lal (2008) for three potential levels of change in total C concentrations are shown in Fig. 3. The mean power values were based on the power values calculated individually for the 11 experimental sites. Our analysis showed that in a CRD experiment of this sort, with only three replications, the chances of detecting a 10% difference in C concentration between CT and NT are extremely low for all soil layers. The largest power observed in any of the sites hardly exceeded 20% even for the 5- to 10-cm depth. For the 5- to 10-cm depth, the chances of detecting a 100% difference in C with three replicates per site increased to $>80\%$. In the 50- to 60-cm depth, however, the chances of detecting a 100% difference remained $<50\%$ (Fig. 3). That is, even if true C concentrations in one treatment at the 50- to 60-cm depth were twice that of the other treatments, statistical analyses would probably have missed it and we would have reported the absence of statistically significant differences.

Power calculations for each individual site are shown in Fig. 4. We considered the very conservative scenario of differences in management (CT vs. NT) leading to a factor of 4× change (200%) in the total C concentration at the 50- to 60-cm depth (Fig. 4). Although such change is extremely optimistic given the relatively short duration of the experiments at most sites, it should be easier to detect—yet in only six of the 11 studied sites would it have been detected with >80% certainty. In the other five sites, the probability of getting statistically significant results in a sampling with only three replications under even such substantial differences in soil C concentration would be as low as 50% (Fig. 4). Thus, based on the results of this study, the hypothesis of no difference between CT and NT can hardly be accepted for half of the studied sites. The only remaining conclusion would have been that an insufficient number of replicates were sampled.

Power Analysis of Total (Whole-Profile) Carbon Stocks

Our analysis of the variability in the soil C stock data from Syswerda et al. (2011) showed that the contribution of individual layers to the overall variability of whole-profile (0–1-m) C stock within individual plots was approximately 20% for the surface layer, 40% for the middle layer, and 40% for the deep layer. That is, middle and deep layers contributed >80% of the variability in the whole-profile C stock (Table 1).

These differences in the variability in soil C stocks by soil profile layer suggest that for optimal detection of treatment differences, analyses should be conducted separately for each soil layer. In other words, if differences between treatments occurred just in a certain layer, the differences will be more likely to be detected if analyzed by layer. If we consider stocks in the soil profile as a whole, statistically detectable differences in specific layers might otherwise be obscured by the greater overall variability of C stocks in the entire profile.

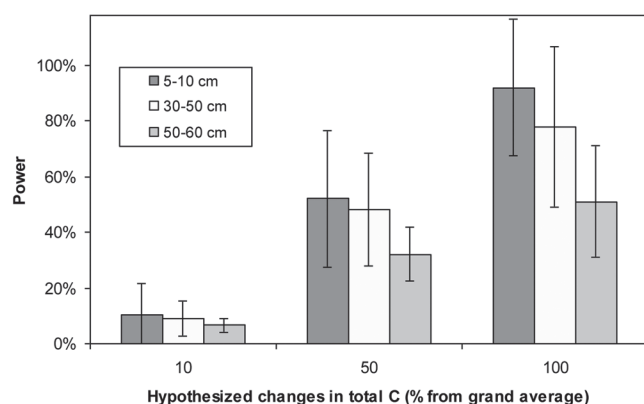


Fig. 3. Probability (power) of detecting statistically significant ($\alpha = 0.05$) treatment effects from the 11 Major Land Resources Area sites sampled by Blanco-Canqui and Lal (2008), a set of 11 completely randomized design experiments with three replications, as a function of relative differences between conventional tillage (CT) and no-till (NT) treatments for three soil depths. The graph shows the average power values calculated based on the analyses conducted for each of the 11 experiments. Vertical lines within bars represent standard deviations.

Scenario analyses of the Syswerda et al. (2011) data illustrate this point. In a first scenario, we hypothesized that the difference between CT and NT treatments in total C stock occurred in the surface layer only, while there were no differences in the rest of the profile. We considered a hypothesized difference of 50% or 1.7 kg C m^{-2} . In a second scenario, the 50% difference existed in the middle layer (1.25 kg C m^{-2}), while there were no treatment differences in the surface and deep layers. In a third scenario, the 50% difference existed in the deep layer only (0.7 kg C m^{-2}). We conducted power analyses both for the case where the C data of each layer were analyzed separately and for the case when C stocks from the three layers were added together to represent the whole profile.

Our results showed that in the first scenario, when the surface layer data were analyzed separately, the power of detecting statistically significant difference was quite high (>90%). In fact, only two subsamples per plot was sufficient to detect the difference in the surface layer with a power of 90% (Fig. 5). When the stocks for the entire profile were added and the resulting data were subjected to statistical analysis, however, the power of detecting differences was just above 55%. To raise the power to 90%, at least 23 subsamples per plot would need to be analyzed (Fig. 5, fourth bar from left).

We observed similar trends of lower power for the second and third scenarios, where true 50% differences existed hypothetically for middle and deep layers, respectively. As would be predicted based on the greater C stock variability in the subsurface horizons, the power was very low even when each layer was analyzed separately—about 50 and 22%, respectively, for the middle and deep layers. But whole-profile analysis made it even more unlikely that the true 50% differences would be revealed—about 38 and 18%, respectively. More than 100 subsamples per plot would need to be analyzed to detect the truly existing 50% change with a 90% probability (power) (Fig. 5).

These low probabilities were lower than those for soil C concentrations alone (Fig. 2) because of the additional

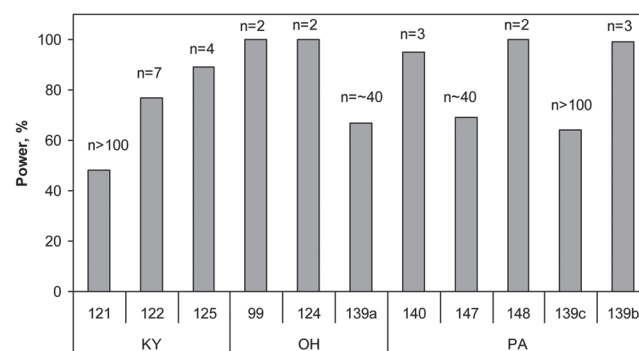


Fig. 4. Probability of getting statistically significant ($\alpha = 0.05$) results in analyses of soil C concentrations at the 50- to 60-cm depth if there were a 200% (4×) difference in soil C concentrations between the conventional tillage (CT) and no-till (NT) treatments of Blanco-Canqui and Lal (2008). The data are from the 11 Major Land Resources Area sites reported (completely randomized design [CRD] with three replications). Numbers above the bars represent the number of replications needed in a CRD analysis conducted separately at each site to detect the hypothesized 200% (4×) difference with 90% probability (power) ($\alpha = 0.05$).

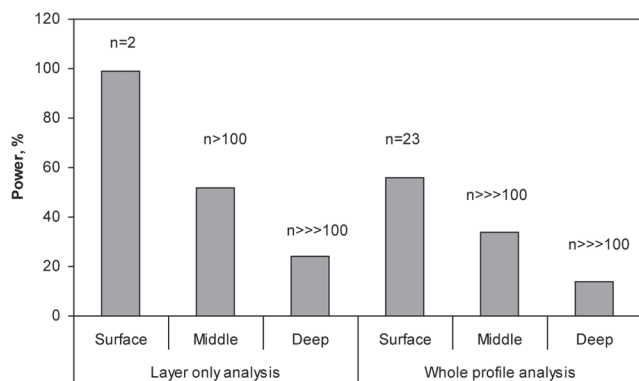


Fig. 5. Probability of detecting a statistically significant ($\alpha = 0.05$) treatment effect in total C stocks from the Long-Term Ecological Research experiment (randomized complete block design [RCBD] with six replications and five subsamples per replicated plot) if the difference between no-till (NT) and conventional tillage (CT) treatments was 50% of the grand average in each layer (data from Syswerda et al., 2011). Layer-only analysis refers to the statistical analyses conducted separately for each layer in which the 50% difference occurred. Whole-profile analysis refers to statistical analyses conducted using calculated whole-profile C stocks. The numbers above the bars represent the number of subsamples needed in each of six replicated plots of this RCBD experiment to detect the hypothesized difference in C stocks with 90% probability ($\alpha = 0.05$).

variability added by bulk density measurements. Total C stock measurements require the combination of C concentration and bulk density to yield an areal-based estimate of soil C (e.g., kg C m⁻²). Where bulk density is measured only once per soil profile position within a treatment, or—worse—assumed, this variability is inappropriately hidden and inappropriately inflates the likelihood of detecting a C stock change.

These scenarios illustrate the deficiency of whole-profile C analysis for detecting real C change in individual layers, and the concomitant danger that real change—whether positive or negative—will be overlooked by insufficient sample numbers or data aggregation. In such cases, the conclusion of no difference between treatments will reflect only the fact that sampling in more variable layers is inadequate to detect even a substantial C change. Policy decisions based on such analyses would be misguided at best and entirely inappropriate.

SUMMARY AND RECOMMENDATIONS

The absence of statistically significant differences among experimental treatments—so-called negative results—can be properly judged only with knowledge of the probability of having overlooked real differences had they been present. Post hoc power analysis provides this knowledge but is too-seldom applied. One of the reasons for its being seldom applied is a lack of emphasis on the importance of controlling Type II error and overemphasizing Type I error in statistical courses taught to agronomists and soil scientists. The absence of post hoc power analysis becomes especially pernicious when important management or policy decisions are based on otherwise inappropriate conclusions.

Power analysis allows the probability of detecting C change with a given sampling design and effort to be readily calculated.

Only then can one evaluate whether the lack of a significant treatment difference is likely to be the result of inadequate sampling vs. the absence of a true effect.

Power analysis is especially important for assessments of soil C stock changes because the variability of C and bulk density usually increases with depth. Consequently, the sampling effort needed to detect a significant change in subsurface horizons will often be substantially greater than the effort needed to detect change in the surface (e.g., A or Ap) horizons. Ad hoc power analysis can identify the amount of sampling effort required to achieve a specified level of certainty before sampling and should be included in the design of all soil C sampling strategies. Post hoc power analysis should be required for all reports that purport to conclude insignificant change.

Whole-profile C analysis should be regarded with particular care. It is important to recognize that conclusions about whole-profile C change can be compromised by higher variability in one horizon than another. In such cases—which are probably the rule rather than the exception—horizons should be analyzed by layer. The construction of whole-profile change should then be based on results only from the layers where statistically significant differences have been detected. Alternatively, researchers should be prepared to state the power of their sampling design—the probability of having detected change of some meaningful magnitude with the number of samples collected. And we should always remember that the absence of evidence is not evidence of absence.

REFERENCES

- Baker, J.M., T.E. Ochsner, R.T. Venterea, and T.J. Griffis. 2007. Tillage and soil carbon sequestration: What do we really know? *Agric. Ecosyst. Environ.* 118:1–5.
- Barker Bausell, R., and Y.-F. Li. 2002. *Power analysis for experimental research*. Cambridge Univ. Press, Cambridge, UK.
- Blanco-Canqui, H., and R. Lal. 2008. No-tillage and soil-profile carbon sequestration: An on-farm assessment. *Soil Sci. Soc. Am. J.* 72:693–701.
- Franzluebbers, A.J. 2009. Comments on “No-tillage and soil-profile carbon sequestration: An on-farm assessment.” *Soil Sci. Soc. Am. J.* 73:686–687.
- Grandy, A.S., and G.P. Robertson. 2007. Land use intensity effects on soil C accumulation rates and mechanisms. *Ecosystems* 10:59–74.
- Ott, R.L., and M. Longnecker. 2001. *An introduction to statistical methods and data analysis*. 5th ed. Duxbury Press, North Scituate, MA.
- Quinn, G.P., and M.R. Keough. 2002. *Experimental design and data analysis for biologists*. Cambridge Univ. Press, Cambridge, UK.
- Richter, D.D., and M.L. Mobley. 2009. Monitoring Earth's critical zone. *Science* 326:1067–1068.
- SAS Institute. 2001. *SAS user's guide*. Version 9.1. SAS Inst., Cary, NC.
- Stroup, W.W. 2002. Power analysis based on spatial effects mixed models: A tool for comparing design and analysis strategies in the presence of spatial variability. *J. Agric. Biol. Environ. Stat.* 7:491–511.
- Syswerda, S.P., A.T. Corbin, D. Mokma, A.N. Kravchenko, and G.P. Robertson. 2011. Agricultural management and soil carbon storage in surface vs. deep layers. *Soil Sci. Soc. Am. J.* (in press).
- VandenBygaart, A.J. 2009. Comments on “Regional study of no-till effects on carbon sequestration in the midwestern United States.” *Soil Sci. Soc. Am. J.* 73:1435.
- VandenBygaart, A.J., and D.A. Angers. 2006. Towards accurate measurements of soil organic carbon stock change in agroecosystems. *Can. J. Soil Sci.* 86:451–463.
- VandenBygaart, A.J., E.G. Gregorich, and D.A. Angers. 2003. Influence of agricultural management on soil organic carbon: A compendium and assessment of Canadian studies. *Can. J. Soil Sci.* 83:363–380.
- Yang, X.M., C.F. Drury, M.M. Wander, and B.D. Kay. 2008. Evaluating the effect of tillage on carbon sequestration using the minimum detectable difference concept. *Pedosphere* 18:421–430.