Using Field Experiments to Evaluate the Impact of Financial Planning and Counseling Interventions

J. Michael Collins[a]

Field experiments, which are a powerful research technique, are common in some fields, but they have not been widely used in studying the effect of financial and counseling planning interventions. Financial services can benefit from the expanded use of field experiments to explore potential causal mechanisms for the effects of financial planning and counseling interventions. This article describes the value of field experiments as well as the potential problems with the approach, in this context. Researchers and practitioners in financial planning and counseling should explore opportunities to conduct field experiments, especially in situations where studies can be carefully designed and implemented in a standardized way with a sufficient number of people and where valid measures are available.

Keywords: financial planning, financial counseling, field experiments

Financial planning and counseling studies is a broad field covering a range of disciplines. Research in this arena includes consumer decision making, microeconomics, consumer finance, financial capability and literacy, household finance, behavioral finance, and even social work and psychology. Each of these fields and subfields has its own research traditions and legacies, but the study of personal finance is beginning to develop its own identity, building on these and other disciplinary backgrounds (Remund, 2010; Schuchardt et al., 2007; Tufano, 2009; Xiao, 2016; Xiao, Ford, & Kim, 2011; Xiao, Lawrence, & Francis, 2014).

Research fields evolve based on the evidence developed through studies. Some studies develop theories; others are descriptive, illustrating how theories are reflected in behaviors or markets. Because fields begin to propose diagnostic procedures and test interventions—or "treatments," borrowing from the medical field—research studies focus more on causal inference. For example, since the 1940s, hundreds of thousands of randomized controlled trials have been published in health care research. Randomized designs can overcome selection bias, allowing researchers to estimate causal effects in a more robust way than other study designs (Jadad & Rennie, 1998).

A number of recent literature reviews in the field of financial literacy conclude that more experimental designs are needed to support such robust causal inferences (Collins & O'Rourke, 2010; Lusardi & Mitchell, 2014). An expanded repertoire of experimental techniques can inform the practice of financial planning in new ways, resulting in more efficient and effective approaches. Of particular interest are field experiments in which programs or interventions are tested with people making decisions in real-world settings (as opposed to hypothetical decisions in laboratory settings). As in medicine, financial planning is often designed to diagnose issues and then offer appropriate treatments to people who need assistance in making financial decisions or using financial tools. A growing number of field experiments have been performed in household finance and financial education (e.g., see Collins, 2013; Duflo & Saez, 2003; Ludwig, Kling, & Mullainathan, 2011), but field experiments remain relatively rare in financial planning and counseling research.

This article provides a brief overview of field experiments as a research strategy, followed by some important considerations for personal finance researchers considering using a field experiment. The article concludes with a number of cautions and best practices for designing field experiments.

[a]*Associate Professor, School of Human Ecology and La Follette School of Public Affairs, University of Wisconsin-Madison, 1300 Linden Drive, Madison WI 53706. E-mail: jmcollins@wisc.edu*

http://dx.doi.org/10.1891/1052-3073.28.2.247

Designing Field Experiments

Even a cursory review of recent articles in journals focused on financial planning and counseling will show that the majority of studies use descriptive or correlational designs, often relying on survey data, administrative data, or primary data collection for summary statistics to describe the central tendency of a behavior or set of behaviors for various financial attributes. For example, a study might estimate savings or borrowing behaviors by race, gender, or another demographic characteristic.

Studies may use more complicated techniques to estimate conditional behaviors, controlling for a wide range of observed factors. Especially with large datasets and rich data, these types of analyses can be quite powerful, describing stark differences for different types of individuals. Yet, these studies often include a pointed caveat that any relationships cannot be described as causal, but only observational, because of the study design. There may be a statistical association between, for example, the level of emergency savings and experiencing a hardship, but hardships and savings levels are both endogenous with unobserved factors, such as motivation level and social networks.

Thus, although studies that show these types of associations are much needed, they cannot conclude that an intervention will cause people to change behaviors. Even well-designed matched-comparison studies, which offer estimates that are less biased by selection effects, can only support causal inferences to the extent that observable factors explain differences between groups. Matched comparisons are perhaps more robust than correlational designs, but they may not be convincing enough to support the prescription of new policies or programs.

Experiments offer another way to gather data regarding the likely impact of an intervention. In an experiment, the researcher assigns each participant to a treatment or control group, using an observable but random mechanism so that any differences between the treatment and control groups at the beginning of the study are also random. Ideally, this also means that any differences observed after the treatment can be ascribed to the randomly assigned treatment. With this design, the impact of the intervention can be directly measured and tested statistically by simply comparing the treatment and control groups. If assignment is truly random, there is no need to control for any other factors.

There are many types of experiments (Levitt & List, 2009). One common design is a laboratory experiment. In this design, "subjects" are asked to perform tasks under various treatment or control conditions. Lab experiments may take place outside of an actual laboratory—for instance, in a classroom or even online—but the context of the tasks and the decisions people make in response to that context are hypothetical. Participants usually know they are subjects and are making choices that may not have a strong impact on them personally. Lab experiments are a powerful and important research tool, especially to test different approaches or aspects of theories. They are also often efficient—they require small sample sizes relative to descriptive or correlational studies, are faster to complete, and can be run in an incremental series to triangulate the effects of various treatments. But lab experiments are not usually directly relevant in the real world. Even when they have a high degree of internal validity—that is, the treatment and control are designed with great care to rule out endogeneity—they may lack external validity. In other words, the results of a lab experiment may not be directly generalizable to actual people making decisions about their own financial situations (Harrison & List, 2004).

A classic field-based randomized controlled trial assigns individuals or households to treatment or control groups, delivers a program or service, then compares the two groups to estimate effects. Random assignment may happen at the individual, household, or other level. In some cases, assignment to the treatment or control group is based on school, office, or other kind of site. This method is often used when it is impossible to isolate groups of people—social interactions within a given site may mean that participants will likely know who received or did not receive a treatment. To avoid this, everyone at a given site is placed in the same group, and the random assignment is made by site.

Another common variation is to randomize participants by cohort: one set of clients is randomly assigned to receive a treatment in one period, and the control group clients receive the treatment in a later period. This overcomes the potential inequity created by some clients being denied what could be a valuable service. In some experimental designs, randomized waitlists or queues are used to define cohorts.

There are many other variations on field experiments that researchers might consider. If a goal of the study is to

determine the effects of a treatment on certain subgroups, stratified or cluster-level randomization designs might be relevant. Likewise, as with cohort- or site-based designs, it may be necessary to randomize participants within particular regions or time periods.

Pitfalls of Field Experiments

Although field experiments are powerful, a randomized controlled trial is not appropriate for every study. Randomization-based approaches take each person's potential outcome as fixed and consider the assignment to treatment as random. As a result, there is no opportunity to observe what would have happened to someone in the treatment group if they had been in the control group. We infer that treatment group participants are similar to the control group. This may be a strong assumption, and it can mean that the results of these studies are not automatically generalizable to other settings.

There are at least five other threats to the validity of even a well-designed study. First, the treatment in a randomized controlled trial needs to be based on a well-designed, well-run, highly standardized program. In the field, programs and processes can be implemented with a high degree of variation. Each client in the treatment group may not actually receive the same treatment. Worse, variation in treatment may not be random. For example, program administrators may informally promote certain aspects of a treatment for some groups of clients. Unless researchers can observe and document these kinds of variations, the estimates resulting from a field trial, especially estimates of effects for subgroups may be biased.

A second threat to the validity of field experiments is the requirement that all participants consent to be in the study to comply with the human subjects provisions of the institutional review board (IRB). When interventions are targeted to economically vulnerable people or are related to financial behaviors, IRB review committees often have heightened levels of concern. People who are willing to consent to be part of a study and cooperate in data collection, especially before they know if they are in the treatment group, may be measurably different from those not willing to consent and cooperate. This presents a serious problem for external validity (Barrett & Carter, 2010; Deaton, 2010). Study participants may not reflect the general client population. To compensate for this possibility, the consent process must be documented in a way that allows researchers to measure whether the requirement that participants make an affirmative decision to participate changes who the composition of the study group. It is also crucial that random assignment occurs after the consent process so that any differences resulting from that process are randomized across groups.

A third threat to validity is noncompliance. It is extremely rare to have every participant who is offered a service actually complete that service; sometimes, only a fraction of people in the treatment group actually complete the full treatment. The best practice is to estimate the effects of the random assignment to treatment regardless of whether clients actually take part. Called intent to treat (ITT), this is the best, least biased measure of the effects of an experiment. The assumption is that assignment to the program or treatment is random and exogenous. The choice to take part in the program is not exogenous—it is a choice. The effects of treatment on the treated (TOT) are biased by those who select into cooperating; thus, TOT is not a valid estimate of the average effect of the program. Yet, ITT average treatment effects are dampened by nonparticipants, which means that the overall estimates are lower than they would be if all participants took part. There are a range of strategies for dealing with the lack of take-up, including using random assignment to exogenously predict TOT (see Angrist & Pischke, 2008, for a discussion). But when take-up rates are very low, it may not be possible to arrive at an unbiased estimation of effects.

A fourth threat to the validity of field experiments is problems or failures in the fidelity of implementation and attrition over time. These situations are most common when the program being evaluated has not been fully standardized and tested in advance (see first threat), but they can occur even with well-designed interventions. One common problem is that some clients in the control group will cross over to the treatment group—when they become aware of the treatment they have not received, they find an alternative way to get that service or program. Flaws in the administration of the program or collection of data may also create problems, for instance, by not allowing enough time between participation and data collection for impacts to show or allowing too much time so that impacts dissipate. The most serious problem is program attrition. The longer a program lasts and the longer the interval between participation and

data collection, the more clients will move, die, or drop out, making the study sample biased in both observable and unobservable ways.

The fifth threat to field experiments is a lack of a valid outcome measures. The financial planning field has a wide range of measures of personal financial condition, including measures of subjective perceptions or attitudes as well as objective knowledge and behaviors. There are also a range of health, mental health, and well-being measures that might be relevant. Many measures used in financial planning and counseling, such as account balances, have a high degree of variance, making finding statistically results less likely, especially in smaller samples. Other data, especially self-reported survey items, may be biased and need to be validated with other measures or administrative data. Moreover, researchers must be attentive to avoid assigning qualitative values to certain outcomes or behaviors, marking them as "good" or "bad." For example, for some clients, an appropriate outcome might be to pay off debt, whereas for others, it could be to take on more debt, depending on the situation. Clients will reflect different preferences, cultures, and contexts. Researchers should use outcome measures that reflect the goals and priorities of clients' financial situations. Well-designed studies use a small number of well-validated, reliable measures that are also not a burden for clients to provide.

Best Practices in Field Experiments for Financial Planning Studies

The simplest field experiment is one in which observably independent subjects all receive the same treatment and the analysis is a comparison of the sample means for the treatment and control group. Unlike observational studies, no elaborate controls or methods are needed. If people are not observably independent—for example, if some sets of clients share the same employer—clustered or stratified designs may be appropriate. In these cases, the researcher ideally plans ahead to account for these shared attributes and uses the experimental design to adjust for covariates. Any analysis beyond the comparison of average treatment effects is secondary and mainly intended to examine channels of influence, not to "find" effects among subgroups in the absence of evidence for overall effects.

Many experiments are set up as pre- and post-designs in which the change in treatment groups is compared to the change in control groups—the difference in differences. This allows the researcher to affirm that there are no differences between groups pretreatment (Figure 1). But if the study has a valid assignment process, a post-treatment only design, which assumes similar starting points and compare outcomes, is an option (Figure 2).

Figure 1. Experimental design: pre-post.

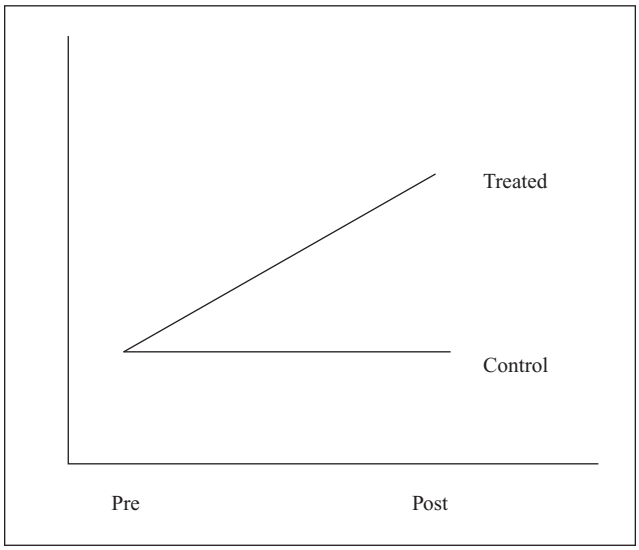
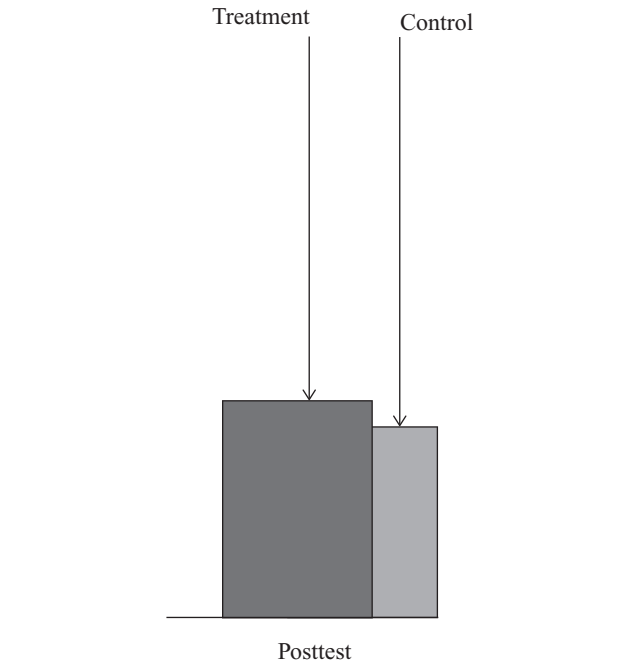


Figure 2. Experimental post-only design.



Good experimental designs have no more than two or three key outcomes measures that are pretested and clearly defined. The analysis plan is designed in advance, including making sure the sample size has enough statistical power to allow detection of effects. A carefully designed study will have a sufficient sample size to allow estimates even assuming the worst cases for consent, take-up, and attrition. An underpowered study will have undetectable effects or very large bounds. Performing a minimum detectable effects (MDE) calculation given a realistic sample size and power assumptions is essential to assure the study will yield detectible effects.

There are a number of accessible texts researchers new to experimental methods may benefit from consulting. Angrist and Pischke (2009) provide an applied explanation of most common causal inference models, often using the randomized experiment as the benchmark. Other books are written in the context of development economics but are equally useful to design studies in personal finance. For example, Taylor and Lynch (2016) offer an easy to use toolkit for consumer finance research, including methods for collecting and analyzing data. Likewise, Banerjee and Duflo (2017) have edited a handbook volume covering a range of best practices and emerging methods. Other texts in experimental economics and behavioral finance may also be useful resources.

Conclusions

A number of potential field experiments could be conducted in financial planning and counseling research. For example, existing programs could test varying timing, modalities, or intensities or combinations of different approaches to study the mechanisms people use to make and execute financial plans. Studies could isolate activities such as goal setting, reminders, and attention to follow-through. Studies using field experiments do not need to be complicated. In fact, keeping the experiment simple makes it possible to explain mechanisms that more complex designs cannot identify.

Field experiments are a powerful tool, but they should not replace other study designs. Table 1 illustrates a range of study designs; the choice of design depends on the researcher's objectives. As Deaton (2010) concludes, "Randomized experiments cannot automatically trump other evidence, they do not occupy any special place in some hierarchy of evidence" (p. 426). A randomized controlled trial can be combined with other approaches, using a carefully designed program and quality measures, to show causal effects other methods cannot. But the design and implementation of the trial may actually be more important than the statistical analysis work. Any researcher considering a field experiment should make sure the program is well documented. If it is not, a better first step is to conduct a process study and then a descriptive or correlational study.

TABLE 1. Comparing Designs

Study Method	Objective	Example
Observational study	Show patterns/ trends Illustrate theory in data	Examine correlation between gender and risk taking
Matched comparison	Model of selection Show conceptual validity	Match men and women on observable traits
Natural experiment	Policy evaluation with external validity	Identify behavior changes after a change in policy or regulation
Lab experiment	Test or develop theory Fast and low cost	Test priming mechanism for risk attitude
Field experiment	Evaluate mechanisms with high internal validity	Evaluate curriculum/ online program

There is the potential for field experiments to be more widely used in financial planning research. Expanded experimental techniques should offer insights for policy and practice and ultimately help the field better serve individuals and families.

References

Angrist, J. D., & Pischke, J. S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton, NJ: Princeton University Press.

Angrist, J. D., & Pischke, J. S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton, NJ: Princeton University Press.

Banerjee, A. V., & Duflo, E. (2017). An introduction to the "Handbook of Field Experiments." In A. Banerjee & E. Duflo (Eds.), *Handbook of economic field experiments* (pp. 1–24). Amsterdam, Netherlands: Elsevier.

Barrett, C. B., & Carter, M. R. (2010). The power and pitfalls of experiments in development economics: Some non-random reflections. *Applied Economic Perspectives and Policy*, *32*(4), 515–548.

Collins, J. M. (2013). The impacts of mandatory financial education: Evidence from a randomized field study. *Journal of Economic Behavior & Organization*, *95*, 146–158.

Collins, J. M., & O'Rourke, C. M. (2010). Financial education and counseling: Still holding promise. *Journal of Consumer Affairs*, *44*(3), 483–498.

Deaton, A. (2010). Instruments, randomization, and learning about development. *Journal of Economic Literature*, *48*(2), 424–455.

Duflo, E., & Saez, E. (2003). The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment. *The Quarterly Journal of Economics*, *118*(3), 815–842.

Harrison, G. W., & List, J. A. (2004). Field experiments. *Journal of Economic Literature*, *42*(4), 1009–1055.

Jadad, A. R., & Rennie, D. (1998). The randomized controlled trial gets a middle-aged checkup. *JAMA*, *279*(4), 319–320.

Levitt, S. D., & List, J. A. (2009). Field experiments in economics: The past, the present, and the future. *European Economic Review*, *53*(1), 1–18.

Ludwig, J., Kling, J. R., & Mullainathan, S. (2011). Mechanism experiments and policy evaluations. *Journal of Economic Perspectives*, *25*(3), 17–38.

Lusardi, A., & Mitchell, O. S. (2014). The economic importance of financial literacy: Theory and evidence. *Journal of Economic Literature*, *52*(1), 5–44.

Remund, D. L. (2010). Financial literacy explicated: The case for a clearer definition in an increasingly complex economy. *The Journal of Consumer Affairs*, *44*(2), 276–295.

Schuchardt, J., Durband, D. B., Bailey, W. C., DeVaney, S. A., Grable, J. E., Leech, I. E., . . . Xiao, J. J. (2007). Personal finance: An interdisciplinary profession. *Journal of Financial Counseling & Planning*, *18*(1), 61–68.

Taylor, E. B., & Lynch, G. (2016). *Consumer finance research methods toolkit*. Retrieved from http://www.imtfi.uci.edu/cfrmp

Tufano, P. (2009). Consumer finance. *Annual Review Financial Economics*, *1*(1), 227–247.

Xiao, J. J. (Ed.). (2016). *Handbook of consumer finance research* (2nd ed.). New York, NY: Springer Publishing.

Xiao, J. J., Ford, M. W., & Kim, J. (2011). Consumer financial behavior: An interdisciplinary review of selected theories and research. *Family and Consumer Sciences Research Journal*, *39*(4), 399–414.

Xiao, J. J., Lawrence, F., & Francis, A. (2014). Financial counseling: Categorizing research papers published in Journal of Financial Counseling and Planning. In V. J. Mason (Ed.), *Proceedings of the Association for Financial Counseling and Planning Education* (pp. 4–7). Retrieved from https://www.researchgate.net/publication/271271126_Financial_Counseling_Categorizing_Research_Papers_Published_in_Journal_of_Financial_Counseling_and_Planning