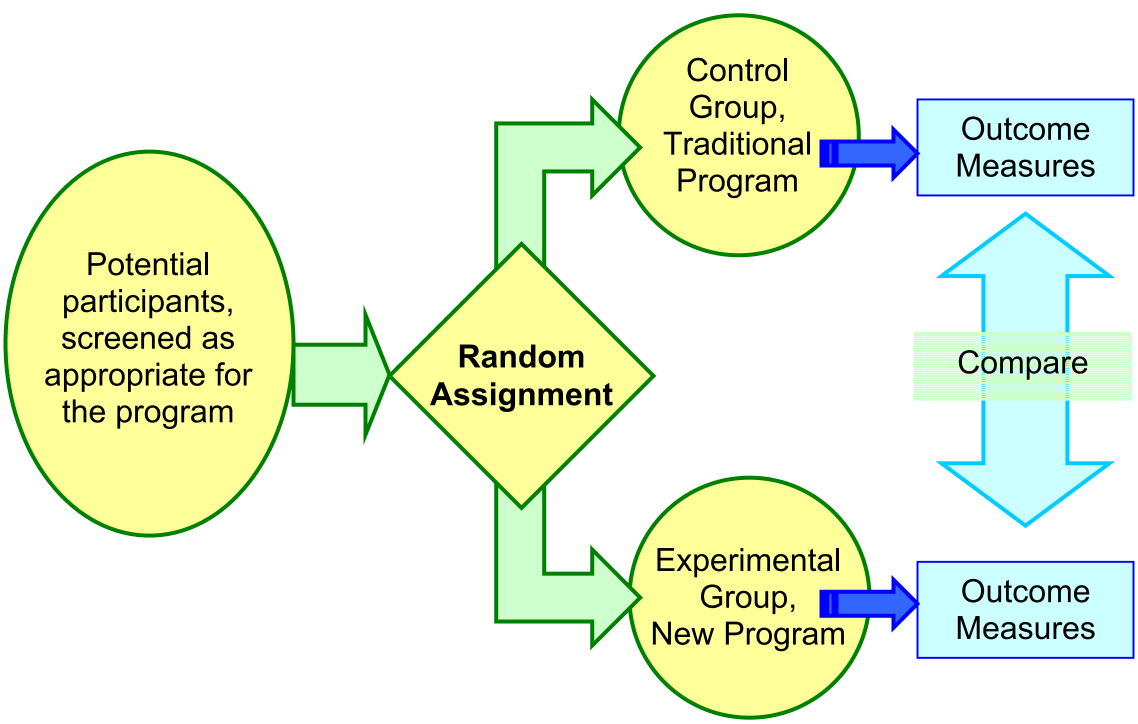
Matching Procedures in Field Experiments

To achieve the perfect experimental comparison, the field experimentalist would like to replay history. In this fantasy, a new approach or a new program would be tried. Then, the history rewind button would be pressed to replay events under the traditional or existing program with the same human beings to see whether things turned out differently. No one could argue that the two groups were dissimilar because they were the same people.

In the real world, we have to look elsewhere for comparisons, but the key idea underlying their selection is still similarity between those undergoing the new approach (the experimental group) and those experiencing the traditional approach (the control group). We would like the control group to be so similar to the experimental group that the two can be regarded as essentially the same. High similarity or close equivalence makes it easier to argue that differences in outcomes are the result of the way participants were treated rather than of preexisting differences between the groups.

Random Assignment. Random assignment is considered the ideal method of selecting a control group in impact evaluations of social programs. In the simplest design, potential program participants are assigned to either an experimental group, usually the group in which some new method or service is being tried, or to a control group, usually the traditional program. "Random" means that each individual has an equal probability of ending up in either group. The assignment of each participant is like the flip of a coin. This can be diagrammed in the following manner:



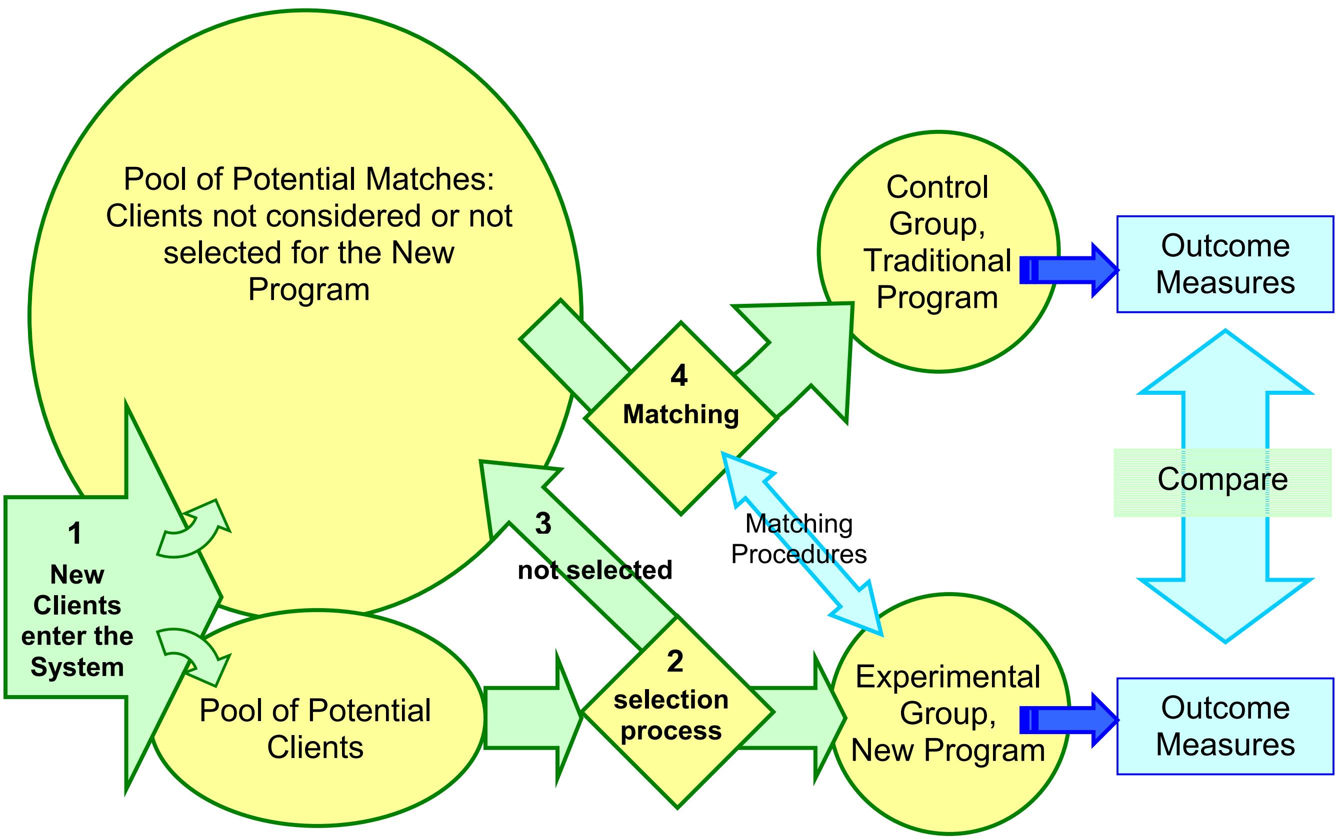
This method is considered ideal because, with only the effort of shuffling people around, essentially equivalent experimental and control groups can miraculously be formed. Operationally, similarity of averages or proportions between groups is the proof of equivalency under random assignment. For example, the two groups will have approximately the same percentage of males, similar age distributions, and the like, and most importantly, similar averages and proportions for variables that are directly related to the new program or services being tested. Because they are practically equivalent, we say we have "controlled for" those variables. And, under random assignment dogma, we believe that even unknown or hidden variables are controlled.

Reasons Why Random Assignment is not used. Random assignment is sometimes resisted because it is mindless. The fate of human beings, often individuals about whom we are very concerned, appears to be left to the role of the dice. Indeed, it can be highly unethical to assign individuals or families to services on a random basis when we know with some certainty the consequences or the value of the services and how they ought to be matched to clients. However, when we do not know or when the number of clients outstrips the capacity or the service system, random assignment may be justified.

Other reasons are more practical. The program may have been put into operation before the evaluation design was considered. Perhaps, all clients who apply or are found to be appropriate for the new program are automatically admitted. No one is left for simultaneous assignment to a control group. It may be that admission was designed to be based on decisions of practitioners and/or clients with no allowance for input from researchers. People may be told all the rational reasons why they may or may not be admitted with no allowance for other considerations, let alone the irrationality of chance selection. In some instances, program intake procedures would have to be scrapped and redesigned to permit random assignment to be introduced.

The Matching Alternative. When random assignment is not possible, matching may be a viable alternative. Matching refers to selection of control group cases based on specific criteria of similarity. In pair matching, we try to find similar individuals or families, one at a time. Some research designs may be based on comparisons of the pairs that are created through this process. The goal of matching in most field experiments of social programs, however, is to achieve comparable groups that are similar in the way that randomly assigned groups are similar, as described above.

Using this method, cases are assigned to the new program in the ordinary fashion. Cases that are not considered (or could not have been considered) for the program represents a primary pool of clients from which matched cases may be selected. The pool may be made up of clients who entered the system before the new program began. Alternatively, it may be made up of clients who entered the system contemporaneously with the new program but were not considered for it. Still others may have been considered for the new program but rejected. If these cases were specifically rejected because they were not appropriate for the program, they will usually not be appropriate cases for matching. In some instances, however, clients are considered acceptable but are not admitted through happenstance or lack of funding, and may, therefore, be among the most appropriate cases for matching. The general process is shown in the following diagram. The important steps are numbered 1 through 4:



This kind of matching involves the following process.

1. Determine the common information (variables) available both for the cases in the experimental group and the pool of potential matches. Information that is available for only one side is not useful for matching purposes. Part of this process may involve determination of the number of holes in the data (missing values) and its reliability.

2. Select the most relevant demographic, psychosocial, and program-related variables from among the common set. This is partly a function of common sense and partly a function of theoretical considerations. One set of critical variables consists of the screening criteria used in the selection process (2 in the diagram).

3. Determine the hierarchy of importance of the variables. This is a process of ranking or weighting variables to determine which are the most important. Which

characteristics are most relevant to participation in the program and program outcomes and which are less relevant?

4. Set up a method for collecting and updating these variables for the cases in the pool and for members of the experimental group. The usual method is periodic extraction of data from official sources (e.g., agency information systems) and maintenance in a research database available to program evaluators and administrators.

5. Produce a method of selecting the best matches using these variables. This involves a computer program that searches the pool for the best possible matches with new cases entering the experimental group.

6. Simultaneous with 1, determine whether the outcome measures anticipated for experimental cases can also be measured for control cases. Set up procedures to collect these data. In some instances, the outcome measure may be available through official sources, but if important observational and interview data are being collected for experimental cases they must also be collected for control cases. It is possible, however, that impact comparisons will be limited to a subset of known experimental outcomes. The type of data collection required for outcome measures will be an important consideration in the timing of the matching process. For example, a design in which data must be collected directly from individuals or families at the start of their program experiences will require contemporaneous and closely synchronized matching—control cases must be immediately identified and approached. On the other hand, outcome measures that are based on routine data stored in official sources or follow-up data after cases close permit a more relaxed approach to matching.

7. Set the process in motion. Monitor the experimental and control group for similarity based on information stored in the research database. In most cases the selection process will occur on a regular basis, e.g., monthly, as data become available on new experimental and pool cases. Regular data updates are avoided when the matching is retrospective. Full retrospective matching is employed when all the potential matches are cases from an earlier period before the new program began. Under this method, the entire pool for matching is obtained at one point in time and never updated.

Limitations. The fundamental limitation of pair matching is that it is based on variables common to experimental and pool cases—a finite set. Complete matching (in which pairs are exactly matched) is seldom possible on more than four or five variables and then only with large pools of potential control cases. A larger number of variables can be utilized but latitude in matching must be tolerated and weighting must be employed to determine those variables for which the closest matches are required. As it turns out, this often works out well when the goal is overall similarity of the experimental and control groups and no particular importance is placed on the matched pairs themselves.

Random assignment theoretically protects against unknown dissimilarities between experimental and control groups, particularly when the groups are large.

Matching does not. The method only controls for known variables used in matching and any unknown variables that are highly correlated with them.

Can Matching by Used in Our Program? There are certain questions to answer when matching is being considered. Here the most important ones:

1. Does a pool of cases of sufficient size exist from which cases similar to cases in the program (experimental) group can be selected?

2. Can common information be assembled for both the program and pool cases?

3. Does this information include the critical characteristics that govern screening and admittance to the experimental group as well as other relevant variables?

4. Can this information be obtained reliably and economically? (Are the relevant organizations or agencies responsive? Will there be ongoing costs associated with data extraction?)

5. Do we have the equipment and software necessary to receive, convert and store data files?

6. Is it possible to measure important outcomes for control cases? (Do we have access to control cases, if necessary? Are there differences in confidentiality of information on control versus experimental cases? Must permission be sought to protect the human rights of control cases? Do we have sufficient resources to engage in data collection from control sources, if necessary?)

7. Once the process of obtaining data and selecting matches is in place, are there resources and personnel sufficient to conduct routine activities and monitor the process?

Tony Loman, Ph.D.
tloman@iarstl.org