APPENDIX V-A

EVALUATION DESIGN

An evaluation design defines the structure for the evaluation. It should be determined by the nature of the evaluation questions asked and, therefore, the kind of evidence that is needed to answer the questions.

The design specifies what people or units to study, how they will be selected and at what time intervals; what methods of data collection will be used and how the information will be collected, including issues of consent and confidentiality; and the kinds of comparisons that are planned. We usually associate evaluation design with outcome and economic evaluations. However, process and monitoring evaluations also benefit from a design. Although process and performance monitoring evaluation designs may be considerably less formal (often referred to as a plan rather than a design), they do address important decisions about the evaluation, for example, what data will be collected and how confidentiality will be protected.

TYPES OF EVALUATION DESIGN

The general types of evaluation designs are:

1. Non scientific designs. These include:
    - Internal self-evaluation (interviews, focus groups, diaries, surveys of staff and participants) usually conducted for the purpose of program improvement or insight.
    - Formative evaluation and Process and Outcome monitoring. The systematic documentation of aspects of program performance that are indicative of whether the program is functioning as intended or according to some appropriate standard.

2. Scientific designs, requiring a formal outcome evaluation. These designs 1) compare program participants with an equivalent group of people who did not receive the program in order to see whether participants' gains exceed those made by non participants or 2) compare program participants before and after they receive the program intervention in order to see if they have made gains on key outcomes. They use systematic inquiry to study or analyze the causes of change or to compare the efficacy of interventions. The scientific method designs include:
    - Experimental
    - Quasi-experimental
    - Non-experimental

DESIGN ELEMENTS

All designs should address these elements:

- Groups to be studied
    a. Are you using a design with one group where outcomes will be measured against a baseline value or comparison with a standard or one where results from an intervention group will be compared to those from a comparison/control group?
    b. Will the design require developing protocols for selection and sampling and what are the issues that have been raised during stakeholders meetings that may impact the content of these protocols?
- Source of data / information
- Evaluation measures or data collection / method tool
- Frequency at which information is gathered or measurements made
- Method(s) of information/data analysis

OUTCOME EVALUATION DESIGN

In deciding on the design type, it is important to determine first, whether there is a need for a scientific method design and secondly, the feasibility of using a particular approach. Where there is evidence in the peer review literature or from formal evaluations done by others that a particular intervention is effective, one can assume that the planned intervention is effective and that there is no need to prove it again. In this case a performance monitoring evaluation can be employed that focuses on whether the elements of the intervention have been carried out as planned (process measures) and whether the outcomes of the intervention are consistent with those of programs that have proven efficacy.

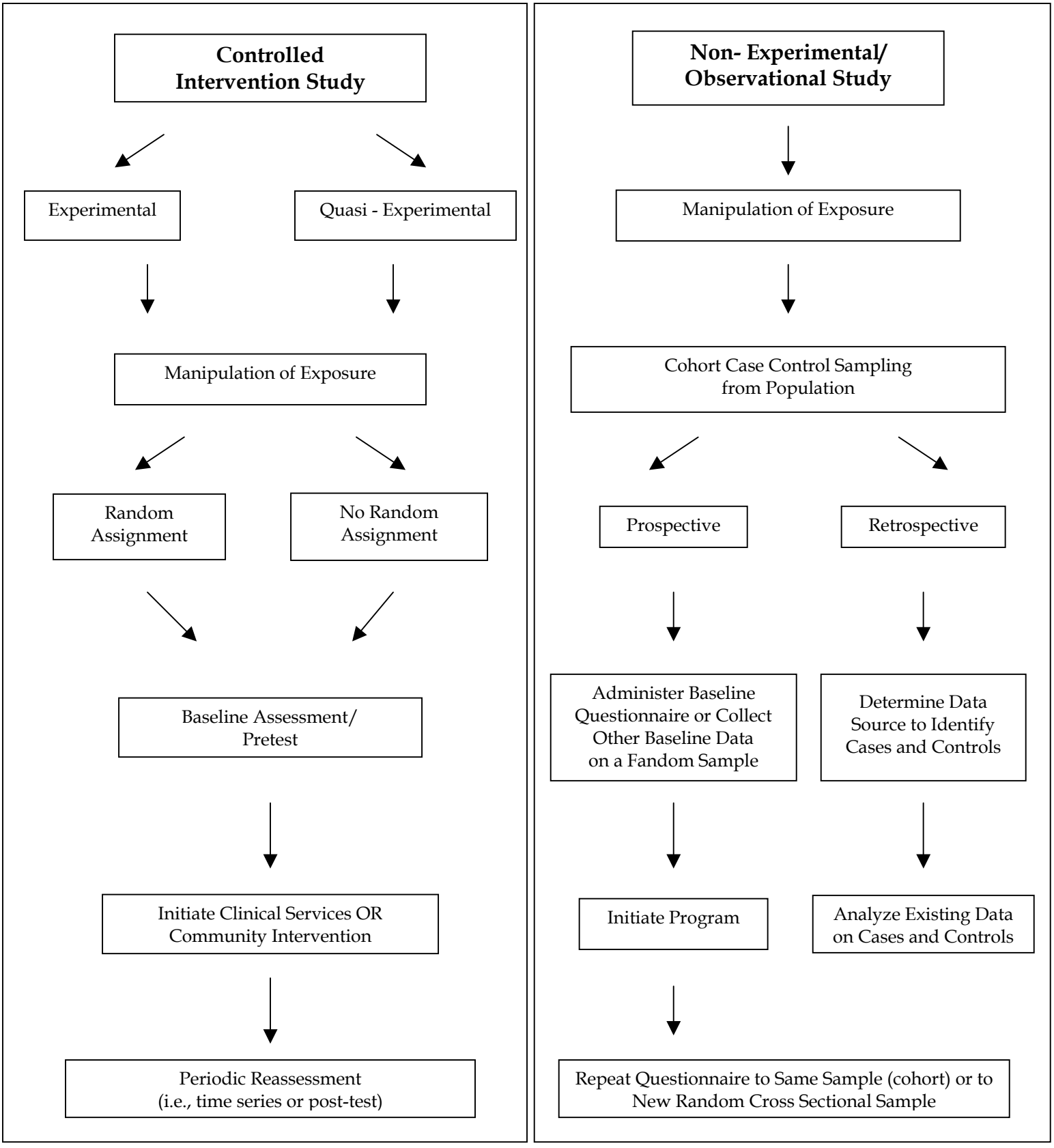
In a demonstration project, a pilot of a new model or a formal research intervention, it is preferable to conduct a scientific evaluation using an experimental, quasi-experimental or non-experimental/ observational design. The diagram "Research Evaluation Designs" provides a decision tree for the selection of an evaluation research design. It is a modified version of a diagram used by Dr. Arden Handler, the Director of the MCH training program at the University of Illinois in Chicago.

The designs outlined on the left are most suitable for use when a new service or program is in the planning stages. The experimental model can be used in a situation in which the exposure to an intervention and selection of subjects can be controlled. The quasi-experimental design is appropriate in a situation in which the exposure to an intervention, the components of the intervention or the choice of subjects is not under the control of the program/evaluator or there are logistic or ethical problems in randomizing participants. This is very frequently the case in public health departments. The term "exposure" for public health programs generally means having received a clinical, educational or behavioral intervention. Random selection assures that participation can be determined in a way that would not bias the selection towards any particular group. Researchers use special random number tables to select subjects. Most public health agencies do something less formal like recruiting every patient that comes in on a certain day of the week or recruiting every third patient that comes in for a service.

The right hand side of the diagram outlines a non-experimental or observational model where it is not possible to control the exposure to the intervention. Thus, we can only study the program's effects on those who use it and compare them with those who do not use it. This is frequently the case in public health. For example, a new prenatal clinic is opened in an underserved area or there is a new state policy that expands the percentage of the population eligible for health insurance. In these examples it would be unethical or politically impractical to limit access to a service that people perceive as a need. If there is advance notice that a new program is coming and its potential impacts can be anticipated, data can be collected across the population that will be eligible for the service before and after the program has been initiated (i.e. prospective). This is baseline data that will later be compared to follow-up data. Where the timeline prevents a baseline evaluation, we can recruit a sample of people who happened to use a service and a comparable (e.g., demographically) group who did not happen to use it and compare them. Another option would be to conduct retrospective analysis of data (e.g. birth certificate data collected prior to the program initiation) that can be analyzed for outcomes of interest in a population comparable to the one that now uses the new service.

This is a very simple presentation of a much more complex process. However, the use of this simple schema can facilitate public health program evaluation. Wherever possible, consultation should be obtained from an epidemiologist or and evaluator.

RESEARCH EVALUATION DESIGNS



Experimental Evaluations

There are three primary types of experimental evaluations. They include pretest/post-test, time series and cross-over designs.

Pretest Posttest – Control Group Design

An intervention and comparison/control group are randomly chosen and assessed or measured for the risk factor or outcome of interest. The intervention group receives the services, while the control group receives some other service or no services. At the conclusion of the intervention, the two groups are retested or reassessed. For community wide interventions aimed at increasing awareness, improving knowledge level, changing attitudes or behaviors, these are pre and post exposure cross-sectional surveys.

Time Series Design

This design uses the same procedures as the above pretest-posttest design, but continues to follow the intervention and control groups at defined intervals over a longer period of time. This design is used to determine if the outcome of interest is sustained over time.

Cross-Over Design

This design allows for all study participants to receive the intervention. This is important in public health agencies when it is expected that an intervention will be beneficial and it would be ethically and politically suspect to deny these benefits to a group that had been used as a comparison. Like the pretest-posttest design, randomization occurs and two groups are formed and tested, one group receives the intervention and both groups are followed and measured repeatedly over time. At a certain point the second group is exposed to the intervention and the testing continues for both groups until the scheduled conclusion of the intervention/evaluation activities. The program would be considered effective if after exposure, both groups show improvements or increases in test results.

Quasi-Experimental Evaluations

Quasi-experimental evaluations are used when it is not possible to randomly assign participants to intervention and comparison groups. In this situation, the two primary alternatives are nonequivalent or nonrandom control group design or time series design.

Nonequivalent Control Group Design

This design is comparable to the pretest-posttest – control group design, however there is no formal randomization of participants. In these situations, control groups can be selected with less rigorous techniques such as recruiting at a specific site on certain days of the week, or in different neighborhoods with similar population characteristics. For those quasi-experimental designs using comparison groups, it is essential to have groups with similar characteristics.

Time Series Design

A time series design uses one group, and follows that group for a specified number of observations/measurements.

Using non-random techniques introduces issues of bias and other outside influences. Identifying a comparable group(s) is one way of controlling for the limitations of the design. If a comparison group is not available, testing your participants over a specified period of time, before you introduce your intervention, allows you to control for knowledge or behavior that exists in both groups and is not affected by the intervention.

Non-Experimental Design

The use of a case (study) and control group, or the pre-testing of participants may be too time or resource consuming. Given that possibility, one may test a single group of participants conducting pretest and posttests to measure differences in attitudes, knowledge, etc. The lack of comparison is the most distinguishing characteristic of a non-experimental design. With this type of design you should still try to account for and explain possible outside influences or bias.

Non-Experimental/Observational Evaluation

This approach can be done using either a prospective or retrospective study.

Pretest-Posttest Design

This design is the most common approach used in prospective studies. A group of individuals in a target population or community are tested or surveyed for the knowledge, attitudes, behaviors or health characteristics of interest. An intervention such as the opening of a new clinic or the expansion of insurance coverage occurs. The community can then be reassessed with another cross-sectional survey. Survey questions can include whether or not the person received insurance coverage or new services.

Retrospective Record Review

This is the most common approach used in retrospective studies. A data source must be identified that contains information on those who were exposed to an intervention or condition and those who were not, for example those who received a new type of health insurance coverage and those who did not. The data source must also contain information on the outcomes of interest in both groups. Comparisons can then be made on the association of particular outcomes with exposure to the intervention of interest. Every attempt is made to match the two groups on demographic or risk factors that are known or suspect to influence the occurrence of a targeted outcome.