The Design of Outcome Studies in Social Work: Quasi-experimental and Experimental Approaches

Bruce A. Thyer, Ph.D.
Florida State University - USA

A paper presented at the Symposium on Social Work Practice Research,
Held at Peking University - Hong Kong Poly U
China Social Work Research Centre, Beijing, China

There are many legitimate ways to evaluate the outcomes of social work services

The Major Categories of Evaluation Studies are

•Quantitative studies, using large groups of clients.
•Qualitative studies, using small numbers of clients
•Mixed-Methods studies, using both quantitative and qualitative approaches
•Single-Subject Research Designs

The Focus of Today's Talk will be on the first approach -

Quantitative approaches known as quasi-experimental and experimental group research designs.

Group Research Designs May be Used to Answer both Simple and More Complex Questions

Simple Questions Include:

•How do the clients fare, immediately after receiving a social work service?
•How do clients fare, after receiving a social work service, later on? (e.g., months or years later)
•Do clients IMPROVE after receiving a social work service?
•Do any improvements MAINTAIN, later on?

More Complex Questions Include:

- Do treated clients improve more than untreated clients?
- Do treated clients receiving a new or experimental intervention improve more than clients who received existing or standard services?
- Did the experimental treatment CAUSE any improvements?
- Are any advantages of one social work intervention maintained over time?
- Apart from focused problems, are there any generalized benefits from the intervention?
- Does the intervention have any unintended negative side effects?

Simple questions only require a simple research design, in order to be answered.

More Complex questions require more complex research designs, in order to be satisfactorily answered.

Evaluating outcomes is relatively easy.

Determining the CAUSES of outcomes, or of differences between social work intervention outcomes, is much more difficult.

There are several pre-requisites needed to conduct a proper quantitative group outcome study:

- A large number of clients/participants
- The use of one or more valid outcome measures (e.g., dependent variables)
- One or more conditions or treatments, such as the *experimental treatment* which is the focus of the study, and *control or comparison conditions* such as standard services, an alternative treatment, no-treatment or placebo treatment (the independent variables)
- Proper statistical analysis of changes or differences

Research designs used in social work outcome studies can be roughly grouped into three categories*:

- Pre-experimental Designs (involve one group of clients)
- Quasi-experimental Designs (involve two or more groups of clients, created naturalistically or by convenience)
- Experimental Designs (involving two or more groups of clients, created using random assignment methods)

*for today's purposes we will omit discussing time series designs and their variations

There are Several Conventional Symbols Used to Outline the Design of Group Outcome Studies

- X is used to refer to the clients in one group receiving the experimental treatment or intervention
- Y is used to refer to the clients in another group receiving a different treatment
- Z to yet another treatment
- And so forth

There are Several Conventional Symbols Used to Outline the Design of Experiments

- O is used to refer to a *observational period*, when all clients are assessed using one or more outcome measures.
- O_1 refers to the first observation or assessment of clients.
- O_2 the second assessment, O_3 the third assessment, etc.

Some **Pre-Experimental Designs** May be diagrammed as

X – O

Known as the post-test only design.

This involves **one group** of clients, who are assessed on some valid outcome measure, AFTER they have received a social work intervention.

This can be used to empirically answer the simple question:

What is the status of clients who received our agency's services?

By sampling all clients, or a representative sample of them, one can greatly improve an agency's evaluation efforts beyond the level of the impressions of the social workers, or by using only anecdotal reports.

The X – O Design can be used to answer question such as:

- What percentage of Hong Kong Poly U bachelors students who apply to Peking University, are admitted to graduate school?
- What percentage of women seeking help for being battered, report being battered one year later, after receiving social work services?
- What percentage of teenagers who received a school-based sex education program report reliably using condoms, six months later?

The diagram

$O_1 – X – O_2$

Means

- A group of clients was assessed at one point in time, O_1 and then they received the intervention labeled X. After treatment, they were assessed a second time, O_2

By assessing clients' function **before** receipt of services, and then a second time, **after** receipt of services, you can answer the question:

"Do clients **improve**, immediately after receipt of social work services?

What does this diagram indicate?

$O_1 – X – O_2 – O_3$

This indicates that a group of clients was assessed at one point in time, O_1 and then they received the intervention labeled X. After treatment, they were assessed a second time, O_2 and sometime later they were assessed a third time, perhaps a follow-up period, O_3

This can help to answer the question:

Do any improvements MAINTAIN, later on?

This Pre-test Post-test Single Group Design can be used to answer questions such as:

- Do client's average scores on a valid measure of depression improve, after participating in a group treatment for depressed persons.
- Do student's self-esteem scores improve after participating in a positive youth development program in Hong Kong schools?
- Does the percentage of youth attending a particular school report a reduction in school bullying, after an anti-bullying program was implemented at the school?

And so forth

- By adding additional periods of assessment you can help determine if any improvements are **maintained over time**.

Yes, but……..

Assume that you have conducted such studies, and obtained favorable results. It is certainly good that your clients improved, on average. But can you really assert that your treatment worked? That is, that it ***caused them to get better?***

Not really, because the improvements may have been caused for some other reason(s) besides your treatment.

Can you come up with any alternative explanations apart from the treatment, which could account for this apparent improvement?

Of course you can, for example…

- Some problems naturally vary over time.
- People tend to seek treatment when their problems are at their worse. If they are assessed then, and reassessed sometime later, they may display lower scores but this is not because of treatment, merely a general movement to a more *normal* level of depression.
- Sometimes problems vary by the time of the year.

Of course you can, for example…

- Clients who you treated may minimize the severity of their problem, so that your feelings are not hurt, or to please you.
- Many psychosocial problems respond favorably to *placebo factors*, nonspecific positive effects related to receiving care.

- Maybe something good (or bad) happened at the local, community, state or national level, e.g., 9-11, the "Obama Effect", a Supreme Court decision, etc.

All these rival explanations complicate our ability to make a causal inference.

Collectively these alternative accounts are called *threats to internal validity*

An internally valid study is one which permits a high degree of causal inference.

Quasi-experiments and true experiments are intended to increase internal validity by reducing sources of bias

By eliminating rival explanations, one narrows in on the real explanation for observed changes

- Eventually, it is hoped that by reducing biases, one will be able to determine what the real effects of treatment really are.

Quasi-experiments

- These involve a comparison group. Some designs include:

$O_1 – X – O_2$ $O_1 – X – O_2 – O_3$

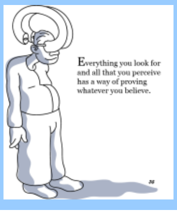
- - - - - - - - - and the - - - - - - - - - - - - -

O_1 O_2 O_1 $O_2 – O_3$

The pretest-posttest no-treatment control group design, and with a further follow-up period. The dashed line separating the groups means they were NOT created using random assignment.

This design adds a no-treatment control group, clients assessed at roughly the same points in time as those in the treatment group.

Having a no-treatment control group helps control for extraneous influences such as the passage of time, regression to the mean, concurrent history, and so forth, all possible sources of bias.



Suppose you have a social work program intended to help unemployed persons find work. Some applicants get enrolled in the program right away, and others are placed on a waiting list, due to a lack of resources to help everyone at the same time. This could result in a naturalistically occurring pretest-posttest no treatment control group design which could be diagrammed as follows:

$O_1 – X – O_2$
- - - - - - - -
O_1 O_2

Lets use numbers, reflecting percent employed, after 3 months, and that the numbers looked like this:

	Pretest	Posttest
Got social work help	0%	80%
	- - - - - -	- - - - -
On waiting list. No help	0%	30%

If all we had was information from the top group, it would look like 80% of our clients obtained jobs, because of the social work program. Sounds good!

But by having the no-treatment control group data, we now also know that about 30% of the unemployed found a job WITHOUT the social work program.

We can more conservatively conclude that the REAL effects of the social work program helped about 50% of the unemployed, not 80% , since 30% found jobs on their own.

Other quasi-experiments can look like:

$O_1 – X – O_2$
- - - - - - - - - - - -
$O_1 – Y – O_2$

Used to compare one program (X) with another (Y)

Or $O_1 – X – O_2$
- - - - - - - - - - - -
$O_1 – Y – O_2$
- - - - - - - - -
O_1 O_2

Used to compare X and Y against each other, and against no services at all. There are many variations of these designs.

The *Logic* of Experiments

”Because the resulting experimental and control groups differ from one another only because of chance, whatever influences may be competing with an intervention to produce outcomes are present in both groups to the same extent, except for chance fluctuations…Any given difference in outcome among randomized experimental and control groups, therefore, can be compared to what is expected on the basis of chance.” (p. 283)

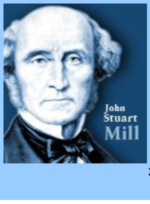
cf. Rossi, Freeman & Lipsey (1999). *Evaluation: A systematic Approach.* Thousand Oaks, CA: Sage)

From *A System of Logic*, by John Stuart Mill (1843)

- **The Method of Difference**

“If an instance in which the phenomenon under investigation occurs, and an instance in which it does not occur, have every circumstance save one in common, that one occurring only in the former, the circumstance in which alone the two instances differ, is the effect, or cause, or an necessary part of the cause, of the phenomenon.” (p. 455)

- See also his Method of Residue, Direct Method of Agreement, Joint Method of Agreement and Difference, & Method of Concomitant Variations



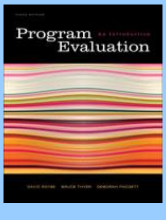
The field of social work contains 100s of published examples of quasi-experiments!

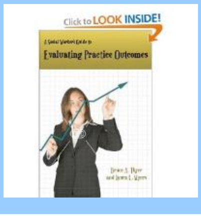
Many of these are described in my book, *Quasi-experimental Research Designs* (2012), Oxford University Press.

And in several of my other books:

Royse, Thyer & Padgett (2010). *Program evaluation: An Introduction* (5th edition). Belmont, CA: Cengage.

Thyer, B. A. & Myers, L. L. (2007). *A social worker’s guide to evaluation practice outcomes*. Alexandria, VA: CSWE.





Problems with Quasi-experiments?

There may be blunt or subtle differences between groups created naturally, which can give rise to apparent differences post-treatment. Differences NOT due to the various effects of treatments received.

To control for this possible source of bias, TRUE EXPERIMENTS create comparison groups via the process of random assignment. This can be used to ensure (usually) that the different groups of clients are essentially the same on all relevant dimensions at the beginning of the study. Thus any differences post-treatment, can be more reliably ascribed to the differing treatments they received.

True experiments are diagrammed just like quasi-experiments, but without the dashed line separating groups.

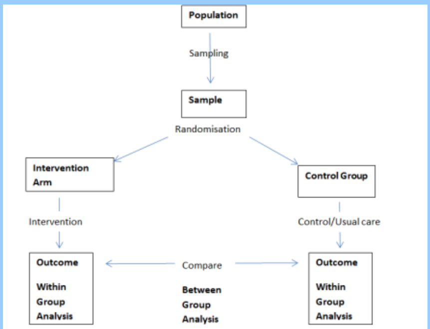
This diagram depicts a QUASI-experiment:

$$\begin{array}{ccc} O_1 - X - O_2 \\ \text{- - - - - - - -} \\ O_1 \qquad\quad O_2 \end{array}$$

This diagram depicts a TRUE experiment:

$$\begin{array}{c} O_1 - X - O_2 \\ O_1 \qquad\quad O_2 \end{array}$$

Here is another way to diagram experiments:



Contrary to popular belief, the use of experimental methods to evaluate social work outcomes is NOT rare.

The earliest randomized controlled trial in social work began in 1937, and was first published in 1949!

See Powers, E. (1949). An experiment in prevention of delinquency. *Annals of the American Academy of Political & Social Science, 261,* 77-88.

This use of experimental methods to evaluate social work programs historically rivals their use in medicine and psychology's evaluation of psychotherapies!

I am completing a comprehensive bibliography listing English-language publications reporting data-based studies using experimental methods and randomized controlled trials, all involving random assignment. This far I have located over 450, through the end of 2013. When finished I anticipate locating over 500 such studies!

Thyer, B. A. & Massie, K. M. (2014). *A Bibliography of Randomized Controlled Experiments in Social Work (1949 – 2013): Solvitur Ambulando.* Unpublished manuscript.

There are very many examples of using quasi-experimental and experimental research designs to evaluation social work and health care programs in China! For example:

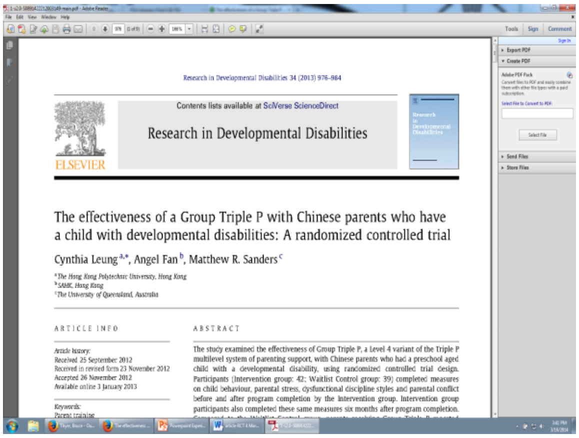
Pang, X. et al. (2013). Does women's knowledge of voting rights affect their voting behavior in village elections? Evidence from a randomized controlled trial in China. *The China Quarterly.*

Wong, K K. et al. (2008). A randomized controlled trial of a supported employment program for persons with long-term mental illness in Hong Kong. *Psychiatric Services, 58,* 84-90.

Chien, W-T. et al. (2004). One-year follow-up of a multiple-family-group intervention for Chinese families of patients with schizophrenia. *Psychiatric Services, 55,* 1276-1284.

Smith, H. L. (2005). Introducing new contraceptives in rural China: A field experiment. *Annals of the American Academy of Political and Social Science, 5999,* 246-271.

Hser, Y-I. et al. (20110). Effects of a randomized contingency management interventio on opiate abstinence and retention n methadone maintenance treatment in China. *Addiction, 106,* 1801-1809.



Research in Developmental Disabilities 34 (2013) 976–984

Contents lists available at SciVerse ScienceDirect

Research in Developmental Disabilities

The effectiveness of a Group Triple P with Chinese parents who have a child with developmental disabilities: A randomized controlled trial

Cynthia Leung [a,*], Angel Fan [b], Matthew R. Sanders [c]

[a] The Hong Kong Polytechnic University, Hong Kong
[b] SAHK, Hong Kong
[c] The University of Queensland, Australia

ARTICLE INFO

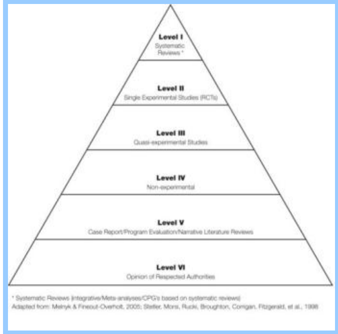
Article history:
Received 25 September 2012
Received in revised form 23 November 2012
Accepted 26 November 2012
Available online 3 January 2013

Keywords:
Parent training

ABSTRACT

The study examined the effectiveness of Group Triple P, a Level 4 variant of the Triple P multilevel system of parenting support, with Chinese parents who had a preschool aged child with a developmental disability, using randomized controlled trial design. Participants (Intervention group: 42; Waitlist Control group: 39) completed measures on child behaviour, parental stress, dysfunctional discipline styles and parental conflict before and after program completion by the intervention group. Intervention group participants also completed these same measures six months after program completion.

When it comes to answering **causal questions**, randomized controlled trials are considered near the top of the evidence pyramid.



* Systematic Reviews (integrative/Meta-analyses/CPG's based on systematic reviews)
Adapted from: Melnyk & Fineout-Overholt, 2005; Stetler, Morsi, Rucki, Broughton, Corrigan, Fitzgerald, et al., 1998

Of course not all important questions in social work deal with 'causality'.

Experimental evidence, including Randomized Controlled Trials of Clinical Interventions, are important building blocks for developing credible, useful knowledge for social work.

In China, you have an opportunity to infuse a more empirical orientation to the development, delivery, and evaluation of social work services than we have adopted in the West, including America.

Much money is wasted on ineffective and sometimes harmful social work services, and people are often not served well.

As professional social work education expands in China, graduate training should include:

- Training in the design and conduct of evaluation research, including quasi-experimental and true experimental methods.
- Including advanced quantitative skills needed for the proper statistical analysis of such data.
- A thorough grounding in the foundations of evidence-based practice, as exemplified through the work of the *Peking University Center for Evidence-based Medicine and Clinical Research*.
  http://english.bjmu.edu.cn/art/2010/10/14/art_6145_49662.html
- And the Hong Kong Branch of the Chinese Cochrane Center
  http://www.hkcochrane.cuhk.edu.hk/hkcochrane/about.html

The Hong Kong Branch of the Chinese Cochrane Centre
中國考科藍中心香港分中心
THE COCHRANE COLLABORATION

This is an exciting time to be involved in promoting high quality social work in China. I wish you well in your endeavors.

Thank you for your kind attention.

Are there any questions?

If you would like a copy of this Powerpoint, email me at Bthyer@fsu.edu