

URR-8105

THE WORLD BANK
DEVELOPMENT ECONOMICS DEPARTMENT
URBAN AND REGIONAL ECONOMICS DIVISION

Urban and Regional Report No. 81-5

QUASI-EXPERIMENTATION IN AN URBAN CONTEXT

A review of experience in the evaluation
of World Bank urban shelter programs.

Michael Bamberger

July, 1981

This report was prepared as part of a program of Monitoring and Evaluation of urban shelter programs which is being conducted by DEDRB. The views reported here are those of the author and they should not be interpreted as reflecting the views of the World Bank or its affiliated organizations.

ISSN 2460

CONTENTS

	<u>Page</u>
Introduction:	1
SECTION I - The Main Types of Urban Shelter Program being studied	3
SECTION II- The Quasi-Experimental Designs used in the Urban Shelter Evaluations	6
SECTION III- Operational Problems in the Implementation of the Research Designs	11
SECTION IV - Evaluation of the Research Designs in terms of their Ability to resolve the Main Threats to Validity of the Research Findings	19
SECTION V - The Practical Utility of Longitudinal Impact Studies in the Urban Context	43
SECTION VI - Conclusion	49
References	55

QUASI-EXPERIMENTATION IN AN URBAN CONTEXT

A review of experience in the evaluation
of World Bank urban shelter programs.

Introduction

In 1975 the World Bank began, in cooperation with the International Development Research Center of Canada, a research program to evaluate a number of the newly developing low-income shelter programs in urban areas. The research began in 3 countries (El Salvador, Zambia and Senegal) and has subsequently expanded to include the Philippines, Indonesia, Colombia and Kenya. The original 5 year research program had as one of its objectives the development of a standardized evaluation methodology which could be applied in other urban projects. 1/ One objective of all of these evaluations has been to estimate the impact of the shelter program on beneficiaries, and in some cases the urban poor in general. To achieve this all of the research designs have included longitudinal surveys which have either used or have considered using some type of quasi-experimental design. The purpose of this paper is to review the experience with the application of quasi-experimental designs in complex urban settings in developing cities, and to discuss the implications for the design of future evaluations of this kind.

1/ A series of 7 manuals are being prepared by the Urban and Regional Economics Division of the World Bank on the methodology of evaluation of urban shelter programs. The manuals cover overall evaluation design, longitudinal impact evaluation, questionnaire design, basic evaluation systems, non-survey techniques, statistical analysis and computer issues in data analysis. The draft version of all manuals will be completed during 1981.

It is important to state at the outset that longitudinal impact analysis is only one of the many types of evaluation studies which have been conducted in these projects. 1/ From the point of view of program management, other types of studies such as efficiency analysis and short-term monitoring are usually of higher priority as they provide more immediate feedback into project design and implementation.

The paper is divided into six sections. Section One describes the main features of the two main types of urban shelter programs being studied. Section Two describes the designs which have been used to evaluate each of the 11 World Bank shelter programs where evaluations have been completed or are in progress. Section Three discusses the operational problems associated with the application of experimental (or quasi-experimental) designs to the evaluation of each of these types of project. Section Four evaluates the research designs in terms of their ability to resolve the main threats to the validity of the research findings. Section Five evaluates the practical utility of longitudinal studies and describes some of the main findings on project impact which have been generated; and Section Six presents conclusions and recommendations for future impact evaluations of this kind.

1/ The main types of study are usually classified as follows: evaluation of project progress, efficiency, effectiveness in achieving its objectives (including longitudinal impact studies) and generation of planning information for future projects.

SECTION ONE: THE TWO MAIN TYPES OF URBAN SHELTER PROGRAM BEING STUDIED

As many features of the evaluation designs discussed in this paper derive from the special characteristics of the programs being studied, it will be helpful to begin with a brief description of each of the two main types of urban low-income shelter program financed by the World Bank and which have been the subject of the evaluation studies.

The first is the sites and services program, in which plots of land are sold, or in some cases leased to low-income families. The plots are provided with certain basic services, such as water, sanitation, roads and possibly community services. Some programs may include a partially built house, while others involve no construction at all, the purchaser building the entire house, either through his own efforts or through organized mutual help.

Sites and services projects are often quite small, offering in some sites as few as 250 to 500 dwelling units, although others include as many as 5,000 units. Normally the total number of units will provide housing for not more than about ten per cent of a city's low-income population. Participating families are required to make a monthly payment which, although designed to be as low as possible, will normally be at a level to exclude at least the poorest twenty per cent of urban families. Given the limited supply of shelter units and the need to ensure that families can meet the monthly

payments, sites and services projects always involve a selection process. In some cases, there is the requirement that stated income must be confirmed by an employer, or that monthly mortgage payments be deducted from a participant's paycheck. Both of these requirements effectively exclude families working in the informal sector. The requirement of a certain time living in the city may also be included to discourage rural-urban migration.

A sites and services project involves the movement of families from their existing place of residence to a new community built on previously unoccupied land. In some cases the distances involved are relatively small, but where projects are located on the outskirts of large cities, families may have to move fifteen miles or more from their former place of residence and employment.

The second type of project involves the upgrading of existing communities or areas of a city. Some projects concentrate on improving the housing stock through dedensification and construction materials loans, while others focus on upgrading public services, such as roads, water, and drainage. In the latter case, it is hoped that families will have an incentive to make their own investments in housing when they see improvements being made in the community infrastructure.

Upgrading projects are often very large and in many cases are designed to cover most of the low-income population of a city. Some upgrading projects directly affect all families in a given area. This is the case when reblocking or dedensification entails the re-alignment of dwellings, or the demolition of a certain number of units and the moving of their occupants to overspill areas. Other projects that concentrate on roads and other basic

services will affect some families directly and others not at all. The difference in impacts can be even greater in projects that involve a number of agencies providing various services. An example would be a case where day-care centers, job-training programs, community centers, health clinics, and paved roads are provided to different areas of a city. Many families will have no access to some services while being directly affected by others.

Unlike a sites and services project, upgrading seeks to improve the quality of existing housing, rather than add to the total housing stock. The cost to families of participating in upgrading projects is generally much lower than for sites and services. The goal is often to ensure that almost no families are excluded by an incapacity to pay, in contrast to the selection procedures typical of sites and services programs.

SECTION TWO: THE QUASI-EXPERIMENTAL DESIGNS USED IN THE
URBAN SHELTER EVALUATIONS

Defining the Quasi Experimental Design

A quasi-experimental design is one which uses a true experimental design as the paradigm but where it is not possible to satisfy all of the experimental conditions required by this design. The researcher is aware of the ways in which his design departs from the experimental paradigm and tries to compensate for this in the statistical analysis or the interpretation of the findings. There are a wide range of quasi-experimental research designs, some of which approximate the true experimental design sufficiently closely for strong inferences to be made about causation, whilst others fail to achieve so many of the experimental requirements that it is impossible to make any but the most general inferences about possible project impact. The most comprehensive discussion of the threats to validity of different research designs is contained in Cook and Campbell. ^{1/} Their scheme will be used in later sections to evaluate the urban shelter evaluation designs described in this section.

The research designs used for impact evaluation

Table 1 summarizes the research designs used for impact evaluation in 9 World Bank shelter projects in 7 countries. Again it is important to recall that the longitudinal impact evaluation is only one of several types of evaluation conducted in each of these projects.

3 of the reported studies were of sites and services projects whilst 6 covered upgrading programs. The following are the main features of the sample designs:

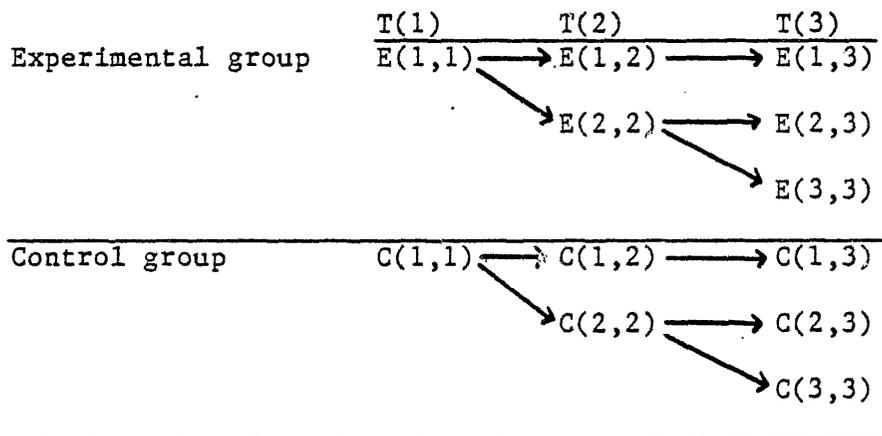
^{1/} Thomas D. Cook and Donald T. Campbell "Quasi-experimentation Design and Analysis issues for field settings" Rand McNally 1979

1. In 8 out of the 9 studies a Mixed Panel Sample design was used. 1/ With this design a sample of dwelling units is selected and the occupants of each dwelling are interviewed in Time 1. In the subsequent interviews the same dwellings are revisited. If the same family is still living in the structure they will be reinterviewed, if they have moved the new family will be interviewed.

In those projects which involve moving the family will be followed to the new site and the interview conducted with them (or if they subsequently move) with the new family who occupies their new house.

The mixed panel design can be represented as follows:

Fig. 1: The Mixed Panel Sample Design



where: E = experimental group
C = control group
First number in parenthesis =
time period in which respondent
first interviewed.
Second number in parenthesis =
present time period of interview

1/ In the final study it has not yet been decided whether to use the Mixed Panel Sample or to select independent random samples for each subsequent interview.

Table 1: RESEARCH DESIGNS FOR IMPACT EVALUATION OF WORLD BANK SHELTER PROGRAMS

Country	Project	Type of sample	Type of control	No. of appli- cations	Sample Size	Main problems
Zambia	Urban 1. Upgrading and sites and services	Mixed panel sample covering 2 experimental communities	Families who stay compared with movers to over-spill areas. No external control due to flooding	3		<ol style="list-style-type: none"> 1. Control group lost in floods. 2. Difficult to match cases between surveys 3. Lack of Computer access
El Salvador	Urban 1. Sites and services with 7000 plots in 5 cities	Mixed panel sample covering 2 experimental communities	Stratified sample from 3 main types of low-income settlements	3	200 300 con- trol	<ol style="list-style-type: none"> 1. Difficult to obtain access to large local computer. 2. Delays in project initiation reduced time over which changes could be observed. 3. Relatively high turnover in control group
Senegal	Urban 1. Sites and Services. One site for 14,000 families	Mixed panel sample selected in phases to reflect fact that project occupancy occurred in phases over several years.	3 control areas selected as typical of low-income areas	*		Extremely slow rate of plot occupation made it impossible to conduct original impact design
Philippines	Urban 1. 1 upgrading project for 180,000 population. 1 sites and services site	Mixed panel sample. Smaller panels selected for special studies.	Internal control provided by areas of project not yet reblocked. 3 external control areas typical of low-income settlements.	4		<ol style="list-style-type: none"> 1. Slow rate of project implementation means longer time period required to measure changes. 2. Lack of access to local computers

(Table 1 continued)

Country	Project	Type of sample	Type of control	No. of appli- cations	Sample Size	Main problems
Philip- pines	Urban 3. 13 up- grading sites and 3 sites and ser- vices	Stratified non- proportional sample covering all 13 upgrad- ing sites. Mixed panel sample.	No external con- trol group. Control provided internally through multiple regression analy- sis	3	1300	Too early to know.
Colombia	Urban 1 Upgrading project in 23 cities	Random sample of families in project areas in 3 cities. Not yet decided whether to use mixed panel or independent random sample designs.	External control in some cities. Internal controls developed through multiple regres- sion	Not de- cided		<ol style="list-style-type: none"> 1. Not possible to cover all project cities. 2. Slow development of project and low in- tervention level mean only small number of families may be iden- tified who have been affected by the project.
Indo- nesia	Urban 3 Jakarta Up- grading	Revisit in 1981 to 750 of sam- ple of 5000 families inter- viewed in 1976. Interview con- ducted with main occupant of dwelling. Mixed panel sample.	Sample strati- fied according to whether in- cluded in up- grading program and period when upgraded	2	5000	<ol style="list-style-type: none"> 1. Not possible to cor- rect certain weak- nesses in original sample and question- naire. 2. Difficult to obtain information on fami- lies renting rooms in a house occupied by owner.
	Urban 3. Upgrad- ing pro- ject in 4 other cities	Mixed panel sample of families in 3-4 project sites	Representative sample from sites not affected by the project	2	500 per city	Too early to know
Kenya	Urban 1 1 sites and ser- vices project		No external control			<ol style="list-style-type: none"> 1. Difficult to keep record of who living in house as many families illegally sub-let

The diagram shows that in T(2) the sample will be composed of: re-interviews with families who are still living in the same house: E(1,2) for the experimental group and C(1,2) for the control group and, first interviews with new families who replace families who have left: E(2,2) and C(2,2). Assuming no new structures have been built the complete sample in T(2) can be assumed to be a representative sample of all families. Thus E(1,2) + E(2,2) provides a representative sample of all families living in the project in T(2). At the same time E(1,2) is a representative sample of original families still living in the project and E(2,2) is a representative sample of new families who have moved into the project since T(1). The sample thus has considerable flexibility. Similar arguments can be made for the sample in T(3) although it is now composed of 3 sub-groups.

2. In only one case (El Salvador) was the control sample designed to provide a representative sample of the total low-income population of the city. In a further 4 cases typical control areas were chosen although no claim was made as to their representativity. In the other 4 cases there was no external control group. However, in 2 of these 4 cases a control group is developed within the sample (this question of internal controls is discussed in more detail later).
3. Although in one case (Zambia) the selected control area was completely eliminated by flooding, in most cases it was simply not possible to use the traditional type of control, either because the project was so large as to cover nearly all of the low-income population (Jakarta) or because the way in which

services were to be distributed made it impossible to make any clear distinction between control and experimental areas (Philippines Urban 3 and Colombia).

4. In four of the seven cases where the information is available, it was decided to conduct 3 or more surveys at different points of time, rather than the 2 surveys required for the simple before and after design. The reason for this is that changes occur at different rates for different types of impact so that different time periods may be appropriate. In many cases trends are also not linear so that it is useful to have more than 2 observation points (for example investment in housing is likely to rise and then fall, whereas income from sub-letting part of the house will often fall - during the period of reconstruction - and then rise).

SECTION THREE: OPERATIONAL PROBLEMS IN THE IMPLEMENTATION OF THE RESEARCH DESIGNS

Although each study encountered a number of unique problems [see Table 1] it is possible to identify a number of common difficulties which were faced in the implementation of most of the research designs. These problems can be classified into the following groups:

- a) Problems related to the schedule of project implementation of the research designs. Most studies faced the problem of delays in project implementation. As the research contract was normally for a specified number of years, any delays in the project implementation schedule meant a shorter period of years over which project impact could be observed. In many

cases the delays in plot occupation were as long as two years. As the process of house construction and consolidation often takes up to two years this meant that many of the 5 year research programs were almost ending by the time it was possible to begin measuring some of the main project impacts. In one case the delays were so long that it was never possible to measure impact, even with a one year extension.

The implementation delays were often compounded by the numerous changes in project design and implementation schedules. In upgrading projects which involve movement of houses, it was usually not possible to know in advance which families would be moved (affected) so that often baseline samples had to be made excessively large in the hope of catching a sufficiently large number of families who would subsequently be affected. Even with large base samples it was often necessary to make some hurried modifications in the sample once it was known which families would move. Often this meant that there would be no "before movement" measures for many of the moving families.

b) Problems related to the identification and classification of families

The mixed panel study design used in most studies requires that the original sample be revisited and that it is established in each case whether the same or a different family is interviewed. In many studies it proved extremely difficult to achieve these two seemingly simple administrative tasks. In many of the control areas there is often no precise address for a structure, and even with fairly careful mapping it may be difficult on a return visit to establish which is the original structure. Houses can be merged, numbers can change, the entrance can be changed from one street to

another and even roads can vanish or appear. It is also more difficult than might be thought to determine whether the family is the same as during the previous interview. In many cultures a family can have several different family names and the name given to the interviewer can be different on each occasion. In Indonesia if a family has a run of bad luck they may decide to change their name to one which is more auspicious. In one interview the woman may be classified as family head and her name be noted, whilst in the next interview the name of her companion may be given. For all of these reasons the process of determining whether the family has moved or stayed in the same house is complex and entails a margin of error.

The identification problem can be further complicated in some project areas. Many upgrading projects involve the physical movement of up to 25 per cent of the structures, usually with new numbers being given to structures. For example, in Zambia families who were required to move to overspill areas were first moved to a temporary unregistered plot. This was later given a permanent number unrelated to the earlier temporary number. This made it almost impossible to follow-up on individual families.

In theory there are a number of consistency checks which can be used during the analysis to determine whether the family occupying a house has changed. In fact none of these tests have proved to be foolproof. 1/

1/ For a discussion of the problems involved in applying these consistency checks see David Lindauer "Longitudinal Analysis and Project Turnover. Lessons from El Salvador" Urban and Regional Economics Division. September 1979.

One important finding of the research is that many projects have differential impacts on owners and renters. Whilst house owners usually benefit from the increased value of their properties, renters may either be forced to leave due to reconstruction or to their inability to pay the increased rent. For these reasons it is important to be able to identify and study both owner and renter families. In practice it has proved to be very difficult to identify families who are renting a room or part of a house. The first problem is that many projects forbid or discourage subletting so that the owner is likely to conceal the fact that he/she is subletting by simply stating that the rooms are occupied by friends or relatives. The situation is often complicated by the fact that many rental arrangements are informal. Rent may be paid partly in the form of services or only on an irregular basis. A further complication is that it is often difficult to define clearly the limits of a structure. In squatting areas where no-one has clear tenure rights it will often not be clear how many rooms or structures "belong" to the person being interviewed. For all of these reasons it is often extremely difficult to obtain an accurate estimate of the proportion of renters or the income from rent.

c) Sample attrition

One of the arguments often used against panel study designs is that the rate of sample attrition will be so high as to drastically reduce the size of the panel and make the remaining cases too unrepresentative to be usable. To date the studies have shown that attrition rates, whilst significant, have been lower than expected, particularly among project participants. In all cases a sufficiently large proportion of the original families have been available for reinterview to make the panel design useful.

d) Problems of access to computer facilities

In all projects, with the possible exception of Senegal, the research teams have suffered from lack of access to adequate data processing facilities. In addition to slowing the process of report preparation, the delays have often meant that problems such as difficulties of matching cases have not been recognized until it is too late to rectify them. In several cases the problem has been partially resolved by conducting most of the data analysis in Washington, although the process of data transference produces its own problems as well as defeating the purpose of developing a local research capacity.

A comparison of evaluation problems with sites and services and upgrading projects

From the point of view of the research design, sites and services projects have a number of advantages making them easier to evaluate [see Table 2]. In the first place, there are no special problems in defining which families will be affected by the project. The houses are constructed on previously empty land and there is no ambiguity as to whether a family is a participant. Families tend either to receive the same package of basic services, or to receive clearly distinguished sets of services at specified times. Sites and services also tend to be limited to a relatively small proportion of a city's population, so it is usually feasible to select a control group of families who will not participate but will have characteristics fairly similar to those of participants.

Upgrading projects are more difficult to evaluate for a number of reasons. In practice, different families in an upgraded area will be differentially affected by a project, a fact complicated by the aim of most

upgrading projects to cover a large population. At one extreme, a family may gain direct access to a paved road, may be located close to a public water supply, or be able to pay for the installation of private water connection. At the other extreme, one can often find families in an upgraded area who do not seem to have been affected by the project at all. Such families may live in a sector where roads have not been paved or drainage not installed, and they may not be able to afford a private water connection. In other projects, involving the coordination of a number of government agencies, a large diversity of services may be provided over widely scattered areas. One family may live near a day-care center, but too far from a job-training program to be able to enroll. Another family may have access to a health clinic, but not to community center. Such circumstances make it very difficult to use a simple random sample, as the number of families who receive a particular service may be too low for a significant number to be captured by the sample.

In the case of many upgrading projects, it is extremely difficult, perhaps even impossible, to determine precisely which families lie within the project area and can be considered part of the experimental group. The scale of many projects can also make it difficult to locate a control group of unaffected families with similar characteristics.

Table 2. A COMPARISON OF SOME OF THE FEATURES OF SITES AND SERVICES AND UPGRADING PROJECTS WHICH AFFECT THE DESIGN OF AN IMPACT EVALUATION

Feature	Sites and Services	Upgrading	Impact on evaluation design
Proportion of low income families affected	Normally less than 10%	Can cover up to 75% of the urban poor	Much more difficult to select a comparable control group for upgrading.
Standardization of the package of services	Families receive same package, or one of a small number of clearly defined options	Wide differences in services actually received. Difficult to measure what each family received	Difficult to define and standardize the "treatment" for upgrading
Uniqueness of participants	Many similar families	Projects often cover almost all poor, or only affect a very special group not easily matched	Difficult to find equivalent control group for upgrading
Definition of project area and participants	Area consists of new housing and is clearly defined	Limits of project effective area may be ambiguous	More difficult to define experimental group for upgrading
Selection of participants	All participants must satisfy criteria in terms of income, family size etc. This will often eliminate up to 25% of the poorest families in the city.	All families in upgraded areas are automatically included	In neither cases are participants randomly selected so there are problems of locating an equivalent control group. In the case of sites and services the problem is to control for variables such as motivation which is impossible to match. With upgrading the problem is to find similar socioeconomic areas when the project may include all squatter areas.

(Table 2 continued)

Feature	Sites and Services	Upgrading	Impact on evaluation design
Speed of project implementation	May be a gap of 3 years or more between time when first and last families occupy their project houses	In large areas the project progresses in stages, with several years between first and last stages receiving the services	In both cases the delays complicate the design. Easier to control in upgrading as there is a phased sequence whereas in sites and services the occupation is unpredictable. In upgrading the later areas can be used as control group for earlier areas.

SECTION FOUR: EVALUATION OF THE RESEARCH DESIGNS IN TERMS OF THEIR ABILITY TO RESOLVE THE MAIN THREATS TO VALIDITY OF THE RESEARCH FINDINGS.

In the previous section we reviewed some of the operational difficulties faced in the implementation of the research designs. The purpose of the present section is to review the extent to which the research designs permit valid inferences to be made about the impacts produced by the projects. To do this we follow the scheme proposed by Cook and Campbell in which four main types of threat to validity are defined. ^{1/} For each type of threat a number of indicators are used to evaluate the design. Each project was reviewed in terms of each indicator. If no problem was found a score of 1 is given. If the design was found to be very weak in terms of the indicator a score of 5 is given. Intermediate scores indicate the relative degree of problem. The scores were then averaged over the 9 designs. This is of course only an approximate ordinal scale. If the average score is below 2 it is assumed that this indicator is not a major problem. If the score is above 4 then the design is very weak in this respect. Scores between 2 and 4 indicate that the problem affects the research design and the validity of the interpretation but may not be a source of major weakness. It should be stressed that the following tables are only intended as a way to provide an overall summary and that when reviewing a particular design in detail much more specific attention would be paid to each individual design component.

The first group of threats to validity refer to Statistical Conclusion Validity, or the extent to which the research design would permit the

^{1/} Cook and Campbell (op.cit) Chapter 2. The list of indicators has been modified for the evaluation of urban shelter programs. See Bamberger "Statistical evaluation of project impact through longitudinal studies" Third draft. June 1980. Urban and Regional Economics Division, World Bank.

identification of project impact if it really did occur. Six indicators are used and summarized in Table 3:

1. Statistical power of the test:

The main issue is whether the sample sizes are large enough to permit statistically significant results to be found. No general answer can be given as the required sample size varies for each dependent variable being studied. In general, however, estimates were made of required sample size during the research design and the analysis has proved the sizes to have been adequate for most purposes. The only major problem occurred when a project developed so slowly that less families entered the experimental group than planned. The average score of 1.9 shows this was not a very frequent problem.

2. Reliability of measures of impact:

Problems were found in several research designs due to difficulties of quantifying some types of impact. Some of the biggest problems related to the estimation of additional income from renting, and the evaluation of project impact on renters. Problems such as this reduced the accuracy of the measurement of outputs sufficiently to give an average score of 2.6.

3. Uniformity of treatment implementation:

In all projects there were unplanned variations in the implementation process. This affected both timing and the package of services received by different families. The problem was normally more severe for upgrading projects than for sites and services.

Table 3. Evaluation of the 9 research designs in terms of indicators of their resolution of the 4 main threats to validity of the inferences which can be made about project impact
(Score of 1 = no problem. Score of 5 = serious problem)

<u>STATISTICAL CONCLUSION VALIDITY</u>	<u>Mean score for 9 studies*</u>
1. Statistical power of test	1.9
2. Reliability of measures of impact	2.6
3. Uniformity of treatment implementation	2.3
4. Random interferences in experimental setting	2.2
5. Control for intervening variables	2.8
6. Matching of cases over time	2.0
AVERAGE FOR ALL INDICATORS	2.3
<u>INTERNAL VALIDITY</u>	
1. History	3.6
2. Maturation	2.1
3. Testing	2.0
4. Instrumentation	2.0
5. Regression	1.0
6. Selection bias	1.7
7. Experimental mortality	2.3
8. Effect of projects on control areas	2.4
AVERAGE FOR ALL INDICATORS	2.1
<u>CONSTRUCT VALIDITY OF CAUSAL RELATIONS</u>	
1. Lack of clearly defined causal model	3.2
2. Unclear operational definition of treatments	1.6
3. Mono operation bias	1.8
4. Interaction between different treatments	2.8
AVERAGE FOR ALL INDICATORS	2.4
<u>EXTERNAL VALIDITY</u>	
1. Representativeness of experimental group	3.2

* Note that in some cases information was available on less than 9 research designs as in some cases the study is not sufficiently advanced to provide the required information.

Once the nature of the problem had been realized, some of the more recent research designs have developed methodologies and systems of analysis which permit compensation for the variations and in fact use them to increase the power of the analysis. ^{1/} Overall this problem was rated at 2.3 due to the high ratings for some upgrading programs.

4. Random interference in experimental setting:

Even when control groups were used, it was found that many unique historical events occurred in one or more of the study areas.

Although some of these events could seriously affect impacts they are almost impossible to control for in the research design. Events of this kind included major interventions by the government or a political party in the project area, projects by other donor agencies, internal political activity etc. Although the effects of many of these events can be evaluated to some extent on common sense grounds, they produce substantial interference with the research design and hence produced an average score of 2.6. These events tend to be less serious in upgrading programs which include a large number of sites in different parts of the city. They tend to be most serious when the project is concentrated in one large site.

^{1/} The new designs include detailed measurement of "exposure to project inputs" so that it is possible to quantify for each family the "amount" of each input they have received. These inputs are then used as independent variables in a multiple regression analysis with project impact as the dependent variable. The partial coefficient of each project input in the regression analysis indicates its contribution to overall project impact. This analysis technique can potentially resolve many of the problems encountered in earlier studies.

5. Control for intervening variables:

In none of the projects was it possible to achieve random assignment of families to the experimental and control groups, so that in even the strongest research designs there is a non-equivalent control group. This raises the fundamental problem of determining whether observed differences between the project and control groups are really due to the impact of the project or to initial differences between the two groups. The average score on this indicator is 2.9. In those designs which do not include a control group, it is almost impossible to control for the effect of intervening variables. When a non-equivalent control group is used it is usually possible to make some adjustment for the effects of intervening variables. The way in which this is done will be illustrated with the example of the El Salvador evaluation.

Table 4 shows the results of a comparison of some of the socio-economic characteristics of experimental and control families in one of the El Salvador projects in 1977, the time at which the first application of the survey was made. It was found that when the two groups were compared using the T-Test or Chi-Square there were found to be statistically significant differences between the two groups on 4 of the 8 variables. The differences referred to the years of education of the household head, family size, household income last month and the age of the head of household. Although there was considerable overlap between the distributions of the two groups, the differences are sufficiently large to establish that there is a non-equivalent control group design.

Table 5 shows the effect of ignoring the differences between the two groups. When the T-Test is applied it is found that there is a statistically greater increase of income among participating than among control families. This would normally be taken to mean that the project had produced an impact on income. The problem with the use of the T-test is that it assumes equivalency in T(1) of the two groups being tested, and has no way to compensate for the initial differences between the two groups. 1/ Table 6 shows the way in which multiple regression analysis can be used to control for these differences. In this case the dependent variable is defined as family income in T(2). The variables where initial differences existed in T(1) are included in the analysis as independent variables (income in T(1), family size, education of head and age of head). We also include Participant Status. This latter is a dummy variable with the experimental group being assigned a value of 1, and the control group 0.

The regression coefficients in Table 6 (column B) indicate the amount of change in income in T(2) produced by a unit change of the independent variable when we hold constant the other independent variables. Thus a difference of 1 colon of income in T(1) produces an increase of 0.83 colons of income in T(2). The probability column also shows that this coefficient is statistically significant. In fact it can be seen that all 4 of the variables identified as having initial differences in T(1) are shown to have an effect on income in T(2).

1/ In fact the T-Test is still valid in that it shows there was a statistically significant difference between the two samples. What we are not able to do is to generalize from the sample to any wider population. In other words we cannot state that participation in the project will produce the observed difference in income for other groups who do not have the special characteristics of the families included in this test.

Table 4. The problem of non-equivalent control groups. Initial differences between the control and experimental groups in Sonsonate, El Salvador at the start of the evaluation (1977)

Variable	Participants	Control	Different	Test	Score	Probability
Years of education	4.52	3.05	+1.47	T	3.96	0.0001*
Weeks worked last month	3.9	3.78	+ .12	T	1.39	0.168
Months in present job	112	126	-14	T	-1.01	0.339
Family size	5.67	4.82	+ .85	T	2.89	0.004*
Household income last month	385	301	+84	T	3.52	0.001*
Age of head	37.2	43.5	-6.3	T	-3.7	0.0001*
Sex of head (% male)	60.3	66.7	-6.4	T	.725	0.39
Branch of economic activity				2 X	4.04	0.67

Note: For all calculations N = 238

* = statistical difference between two groups at 0.05 level or beyond

Source: Michael Bamberger "Statistical evaluation of project impact through longitudinal studies" Table 62. Urban and Regional Economics Division. World Bank. June 1980 (third draft).

Table 5. T-Test for the difference of means for income change of families in the experimental and control groups. Sonsonate. El Salvador. 1977-1980 (figures in colones)

	No. of cases	Mean change	Standard deviation	T Score	D.F.	2 tail probability
Participants	156	199	254	2.86	209	0.005
Control group	82	116	189			

Note: The F Test showed that the variances of the two distributions were significantly different so the version of the T-Test was applied for separate variances.

Source: FSDVM. Socio-Economic study of Sonsonate.

Table 6. The use of multiple regression analysis to control for initial differences between the experimental and control groups. The example of income changes in Sonsonate, El Salvador.

Dependent variable = Income in T(2)

Independent variables	BETA	B	F	Probability
Income in T(1)	0.49	0.83	67.9	0.01
Family size	0.14	21.9	5.9	0.05
Education of head	0.17	17.8	6.9	0.01
Age of head	0.13	3.58	4.1	0.05
Participant status	0.01	68.2	3.31	0.10

Note: Other non-significant variables not included in the table. These include: weeks worked last month, months in job, sex of head, type of residence.

Conclusion: When we control for income in T(1), family size and education of head, there is no difference between participants and control group for change in income.

Source: Michael Bamberger "Statistical evaluation of project impact through longitudinal studies" (third draft) June 1980.

For the present discussion the most important finding is that when we hold constant each of these 4 independent variables, the coefficient for participant status is no longer statistically significant. This means that the apparent project impact on income was in fact produced by the initial differences between the two groups.

Although this regression approach can be very valuable in reducing the effect of initial differences between the two groups it is important to appreciate that it is never a perfect solution. 1/ Although we can control for differences such as income, age, education and family size, we cannot as easily control for differences in motivation. It may be for example that families who chose to apply for projects are more ambitious and this produces the difference in outcome. Although this is an important issue, it is likely to be less serious than in many of the social science research studies often quoted to illustrate the point. Typical examples cited in the literature refer to drug control programs which are only given to those with the worst drug problems, or reading programs only given to those with the worst reading disabilities; or at the other extreme training programs are only offered to subjects most likely to succeed. In these cases the severe drug users are likely to regress and be worse after the program than the control group (creating the incorrect impression that the program had negative effects).

1/ See Cook and Campbell (op.cit) for a detailed discussion of the limitations of regression analysis. See especially chapters 4 and 7.

Although there is self-selection in sites and services, the problem is not likely to be so severe for the following reasons:

- a. Only a relatively small proportion of the population are selected. This means that there are probably many motivated families who do not get into the project.
- b. There are normally not very dramatic differences between participants and control groups on socio-economic characteristics and that there is considerable overlap between the two groups. Thus the groups are not too dissimilar.
- c. As project selection is usually either randomized 1/ or based on criteria of capacity to pay, it is unlikely that only highly motivated families will be selected.
- d. In some cases (for example El Salvador) the control group is stratified into different settlement types, some of which are quite similar to the projects in terms of the type of ownership patterns and housing quality. Some of the families motivated to improve their housing and general welfare conditions will probably have entered one of these other projects, with the result that not all motivated families will be found in the project.

6. Matching of cases over time.

As discussed earlier there are a number of administrative problems involved in ensuring that cases can be matched over time. Although a high

1/ In some cases selection is based on the weighting of attributes such as overcrowding, quality of present housing and services, etc.

degree of accuracy in matching cases has been achieved in most of the well planned evaluations, [although often at a high administrative cost], 1/ in those cases where the research program ran into organizational problems (change of staff, financial problems etc.) it sometimes became impossible to match cases.

The second group of threats relates to the Internal Validity of the research design and its ability to control for factors which might distort the process of measurement and interpretation. The original formulation of Cook and Campbell was modified to give a list of 8 internal threats. The severity of each of these is shown in Table 3.

1. History:

The fact that most projects occupy one, or a very small number of large areas of the city, means that there are likely to be many historical events which occur there and not in some of the control areas (or vice versa). During the course of the research these historical events have included: severe flooding, civil war, epidemics, Presidential visits, etc. Although one can take account of the possible influence of these events in a common sense way, there is no statistical procedure which can control for this. It can be seen from Table 3 that this is on average the most serious of all the threats to validity.

1/ In some cases this involved re-interviewing some families, in others the use of complicated and expensive consistency checks on the computer.

2. Maturation:

This problem is not as severe in housing as with research on children or psychiatric research, where one can expect to find strong maturation effects. In the present studies maturation largely refers to the rate at which families improve their houses. Where a control group exists, it is possible to control for this quite well, but it can of course be a problem where there is no control.

3. Testing:

In general the measurement techniques seem to have been non-reactive in that there is no evidence that they affect the behaviour of the respondent. An exception could be the panel studies in which families were requested to keep a daily record of all sources of income and all expenditures. The process of keeping the record might have some initial effect on expenditure patterns but as the study continued over 3 years it is likely that families would have returned to their normal expenditure patterns.

4. Instrumentation:

As the studies progressed mistakes were sometimes detected in a few questions and in some cases improvements were made. In most cases this took the form of adding rather than changing the questions, but sometimes changes were made in the way the questions were asked. The area where the biggest changes could be expected are in the estimation of income. Better ways were discovered to identify sources of earnings from informal employment and income received in the form of transfers from relatives and friends. This additional information was normally obtained by asking additional questions about the

new sources of income so that it is possible to make adjustments when comparisons are made over time thus avoiding most of the instrumentation problems.

. 5. Regression:

This problem arises when the experimental and control groups are selected from the extremes of the distributions of two populations. In these situations the members of the two groups will tend to regress towards the population mean of their group in T(2). This can produce statistical differences which could be confused with treatment effects. Due to the considerable overlap between the control and experimental groups, this problem is apparently not very severe in the present research context.

6. Selection bias:

This is a serious problem and we have discussed at length the approaches used to overcome it.

7. Experimental mortality.

A high dropout rate can produce serious problems with a simple panel study design. However, with the Mixed Panel Study designs which are used in most of the studies we are discussing, the problem is largely overcome. Two types of problem remain. The first occurs, as in the case of Senegal, where the rate of progress of the project is so slow that a very high proportion of the originally selected families never moves to the project. The second problem is when there is a high turnover rate, but it is difficult or impossible to detect because families are illegally selling or sub-letting their house. If these changes cannot be detected, what are inferred to be projects impacts (for example higher income) may be caused by the fact

that poorer families have been replaced by richer families. This only appears to have been a serious problem in one study.

8. Effect of project on control areas.

Because of the dynamics of a city, there are many ways in which the development of large scale housing programs will affect other areas of the city. This is a type of interaction not found in more traditional experimental designs and hence often not mentioned in the research literature. If a prestigious housing program is developed with external funding, and high visibility (foreign visitors, etc.) the government may decide to divert schools, clinics, etc., which had been planned for other areas of the city, to the project. In this way other areas may become worse off as a result of the project. In other cases, if the project creates resentment within certain government agencies, these agencies may try to sabotage the project by giving priority to other areas. The project is also likely to produce externalities such as changing land values or an increase in the housing stock which lowers rents in other parts of the city. Although it is extremely difficult to evaluate statistically these types of interaction, they are likely to reduce the validity of the inferences which can be made about project impact.

The third threat relates to construct validity of causal relations. This relates to the danger that through inadequate theoretical models or interpretation, a wrong interpretation may be placed upon the findings of the analysis. Of the 4 groups of threats, this group is the second most problematic with an average score of 2.4. The four indicators are discussed below:

1. Lack of clearly defined causal models:

All of the earliest research designs suffered from this weakness to some extent because not enough was known about the way in which the housing programs would operate to permit the development of sophisticated causal models. Although, this has the second highest score of all indicators, The better research designs have been able to develop reasonably adequate general explanatory models.

2. Operational definition of treatments:

This problem arises when the research design is not able to define or quantify the project inputs. This has not proved to be a serious problem as most inputs are relatively easy to quantify.

3. Mono-operational bias:

This is a problem when only one indicator is used to measure project impact. Most of the research designs have made at least some attempt to use "triangulation" and to develop a number of independent indicators of project impact, particularly in the areas of greatest interest to management. ^{1/} In general this has not proved too serious a problem.

4. Interaction between different treatments:

All of the projects involve a number of different treatments which are received by different families in different combinations. A reasonably adequate methodology has been developed for resolving this problem and

^{1/} For example at least 5 independent estimators have been used to measure increase in the value of the house.

for resolving this problem and is being applied in the newer research designs.

1/ This is based on an extension of the multiple regression approach used to control for the effect of intervening variables. In this case a measure is taken of the exposure of each family to each project input. The project components are then included as independent variables and the partial regression coefficients are used to obtain a rough estimate of the relative importance of each input in contributing to the observed project impact.

The logic of this approach can be illustrated with the simplified hypothetical example given in Fig. 2. In this example a project includes 4 components: housing material loans, piped water, paved roads and tenure. We wish to evaluate the relative impact of each of these components, and their combined impact, on the value of housing. We also believe the impact may vary according to the income of the family and the family size. Table 7 shows how each of these variables can be quantified so as to measure the exposure of families to each program input. The amount of loan received can easily be measured as a continuous variable. Access to piped water is defined in binary form, although in some cases it is possible to measure the quantity of water consumed. Distance from the paved road is used as a proxy for the effect of the road on the family, and tenure is defined as another binary variable. Increase in housing value is defined as the increase in the imputed rent. Using these variables an equation can be specified as follows:

1/ The most recent example is the evaluation of the Zonal Improvement Program in Manila.

$$Y = a + b_1 X_1 + b_2 X_2 + b_3 X_3 + b_4 X_4 + b_5 X_5 + b_6 X_6$$

The coefficients of each X can be interpreted as indicating the effect which a one unit change in this variable would have on imputed rent, when all other inputs are held constant. ^{1/}

Although this technique provides a relatively effective way to resolve the problems presented by interaction between different inputs, not all of the research designs contained sufficiently precise measurement of exposure to project impact so this issue has presented problems in several designs and hence has a relatively high score.

The fourth threat to validity derives from problems of External Validity. In the present analysis this is only evaluated in terms of the representativeness of the experimental group and the extent to which it is possible to generalize to wider population groups. This factor which has the highest score (3.2) of the four groups of threats, is closely related to the adequacy of the control groups used in the research design. Two of the research designs have no control group and in only 2 of the designs is the control group considered adequate to permit generalization to the whole low-income population.

^{1/} This is the simplest additive form of the model. It will often be necessary to introduce interaction terms.

Fig. 2. A simple causal model to evaluate the relative impact of different project inputs.

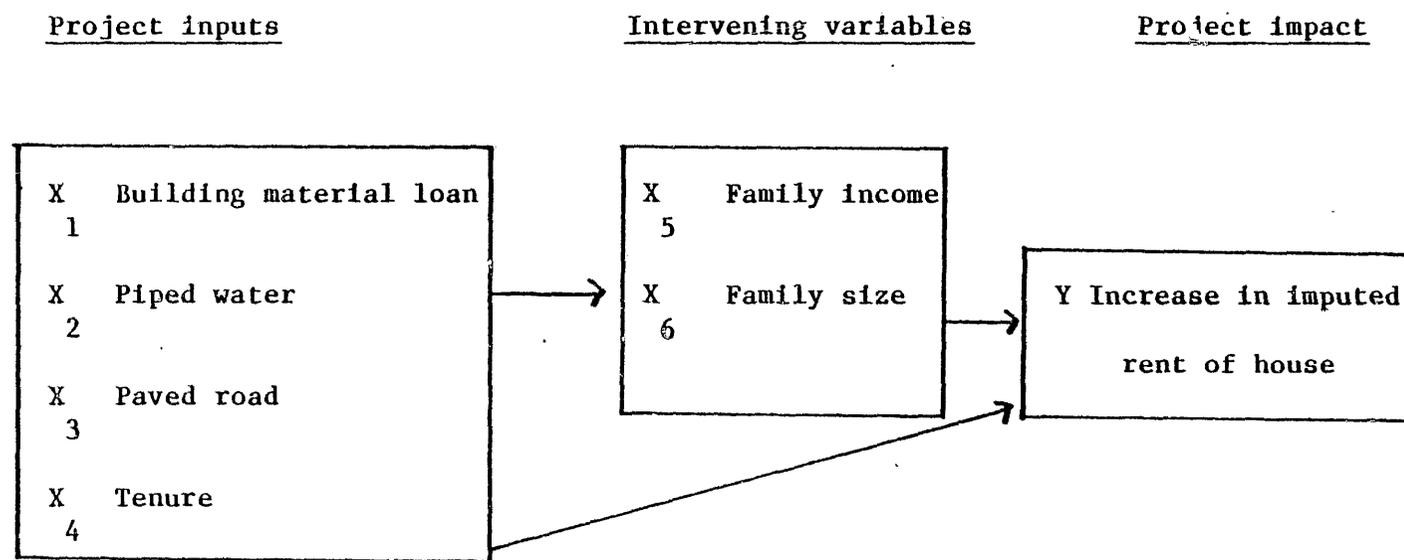


Table 7. Specification and measurement of the variables used in the causal model presented in Fig. 2.

Type of variable	Variable	How quantified	Symbol
Program input (independent variable)	Housing material loan	Continuous	X ₁
	Access to piped water	Binary	X ₂
	Distance from paved road	Continuous	X ₃
	Provision of tenure	Binary	X ₄
Intervening variables	Household income	Continuous	X ₅
	Family size	Continuous	X ₆
Project impact (Dependent variable)	Change in imputed rent	Continuous	Y

Summary of the validity of the research designs

Up to this point we have been discussing the average validity of all 9 research designs. Whilst this is useful for pointing out the many administrative problems encountered in field settings, it is perhaps not a good way to evaluate the theoretical validity of the research models as the average includes a number of designs which proved to be very weak, often for operational rather than theoretical reasons (for example delays in project start-up, changes of staff, loss of control groups through flooding, etc.). Table 6 shows the wide variations between projects.

In an attempt to evaluate how well the designs can operate in reasonably favorable circumstances, table 9 summarizes the average scores for the 4 strongest designs. It can be seen that these 4 designs have average scores of less than 2 for each of the first 3 groups of indicators. This means that the designs are able to resolve reasonably well problems of statistical conclusion validity, internal validity and construct validity of causal relations. The main problems relate to random interferences with the research design. Given the large size and small number of project sites, it is very likely that particular historical events such as natural disasters or political interventions will have different impacts on the project and control areas. There is almost no way in which the research designs can overcome these problems. In upgrading projects covering a large number of sites this is less of a problem but even there it still exists.

Table 8. Comparison of the 9 research designs in terms of their resolution of the 4 main threats to validity

(Score of 1 = no problem. Score of 5 = serious problem)

	Statistical conclusion validity	Internal validity	Construct validity of causal relations	External validity
Zambia	3.0	2.5	3.5	5.0
El Salvador	1.8	2.0	1.5	2.0
Senegal	3.2	2.6	3.5	3.0
Philippines				
Urban 1 (Tondo)	2.0	1.8	1.5	3.0
Urban 3 (ZIP)	1.7	1.7	1.3	2.0
Colombia	2.2	2.2	2.3	3.0
Indonesia				
Urban 3 (Jakarta)	2.3	2.0	2.3	3.0
Urban 3 (other)	2.2	1.7	2.5	3.0
Kenya	2.2	2.5	2.8	5.0
AVERAGE FOR ALL PROJECTS	2.3	2.1	2.4	3.2

Note: Some of the newer designs could not be evaluated on all indicators as the information is not yet available.

Table 9. Summary of the average validity of the strongest research designs (average of the best 4)

<u>STATISTICAL CONCLUSION VALIDITY</u>	<u>Mean score for best 4 designs</u>
1. Statistical power of test	1.5
2. Reliability of measure of impact	2.0
3. Uniformity of treatment implementation	2.2
4. Random interference in experimental setting	2.5
5. Control for intervening variables	2.0
6. Matching of cases over time	1.5
AVERAGE FOR ALL INDICATORS	1.9
 <u>INTERNAL VALIDITY</u>	
1. History	3.2
2. Maturation	1.7
3. Testing	2.0
4. Instrumentation	1.5
5. Regression	1.0
6. Selection bias	1.5
7. Experimental mortality	2.0
8. Effect of projects on control areas	2.2
AVERAGE FOR ALL INDICATORS	1.9
 <u>CONSTRUCT VALIDITY OF CAUSAL RELATIONS</u>	
1. Lack of clearly defined causal model	2.2
2. Unclear operational definition of treatment	1.2
3. Mono operation bias	1.2
4. Interaction between different treatments	1.7
AVERAGE FOR ALL INDICATORS	1.9
 <u>EXTERNAL VALIDITY</u>	
1. Representativeness of experimental group	2.5

Note: Projects compared: El Salvador, Philippines (1) and (3) and Colombia.

Score of 1 = No problem
Score of 5 = Very serious problem

The designs have usually not been able to adequately resolve the threats to external validity. In most cases the project population is different from the control groups (either in terms of the attributes of families or in terms of the characteristics of the project area) and it is difficult to generalize from the results of the project to the potential impact on the whole low-income population. This problem is perhaps less important than it may seem. Even if there was a perfectly controlled experiment with random assignment to the experimental and control groups, it would still be very difficult to make generalizations about what would happen if the project were repeated. The reason for this is that all projects use scarce resources such as land and materials. This means that subsequent projects will have different cost structures and this will affect location and probably design in quite significant ways. Consequently, it will not be possible to strictly replicate the project so that some departure from strict generalizability of the results for the first evaluation may not be too serious.

SECTION FIVE. THE PRACTICAL UTILITY OF LONGITUDINAL IMPACT STUDIES IN
THE URBAN CONTEXT

Section 4 argued that with good planning, and good luck, it is possible to design a methodologically sound evaluation of an urban shelter program. The question must now be asked as to whether these longitudinal designs, with all of the expense they entail, have proved themselves to be useful. In many evaluation programs at least half of the budget is devoted to the longitudinal impact study, and as there are usually more requests for studies than can be covered within the budget and available human resources the question must be asked as to the utility of the longitudinal studies.

If program managers are asked their criticisms of longitudinal studies they will usually cite one or more of the following list of complaints:

- a) The studies take so long to produce results that they are of no practical utility to the project manager.
- b) The studies are too theoretical, producing large volumes of statistical analysis which cannot be understood by program staff.
- c) The studies are too expensive.
- d) The studies are not responsive to changes in project timetables and design so they are not able to evaluate new aspects of the program as they develop.
- e) Some of the more research minded of the managers also question whether it is possible to establish causal relations in the complex urban milieu, and whether even a well designed evaluation can tell much about project impacts.

Some of the dissatisfaction of management with longitudinal studies derives from a misunderstanding as to what this type of study can be expected

to produce. In most cases the evaluation is supervised by the manager responsible for project implementation and as longitudinal studies are designed to help with long-term planning, rather than project execution, it is not surprising there is often some dissatisfaction with the "lack of usefulness" of the studies. For this reason we begin with a discussion of the objectives of longitudinal studies and what they have been able to achieve. In the final section of the paper we will try to draw conclusions as to the role of the longitudinal study in an evaluation program.

The functions of longitudinal studies and what they have been able to achieve

As longitudinal studies are designed to measure changes over time, almost by definition their main utility will not be felt until the second or third measurement has been taken. In most cases this will mean that a period of at least 3 years will pass before the main benefits of the study are seen. However, a number of short and medium range outputs will be generated as by-products.

Short-run outputs

The longitudinal study involves a comprehensive socioeconomic survey of the target population and usually of a control sample of families from the main types of low income housing. This survey information has proved very useful in evaluating the affordability and accessibility of the project to the target population. The data on the economic characteristics of the population has helped in planning employment and income generating programs such as small businesses, consumer and production cooperatives, and savings and loans programs. It has also been possible to apply hedonic price analysis to data

on the control group to obtain indicators as to the different types of housing and services likely to be demanded by participants and the relative priority of each of them. 1/

Medium-run outputs

In most studies the interview is repeated after one or at the most two years. The second survey is usually conducted at the time when families are beginning to settle into their new houses, or when they have completed the initial investment in upgrading their existing house. At this point the studies are able to make a number of very useful contributions which derive from the fact that the same families are being reinterviewed.

A first important contribution has been to throw some light on the question of drop-out rates. It is believed by many people that large numbers of the poorer families will be forced to leave the project areas due to the high costs. The panel studies have proved valuable in providing empirical data on this issue. In most cases it has been found that the turnover rate is much lower than had been expected 2/ and there is little evidence to show that more poor families are leaving. 3/

The only case where high turnover rates have been found is in Kenya, where a high proportion of plot owners

1/ See for example: John Quigley "The distributional consequences of stylized housing programs" Urban and Regional Report No. 80-18. DEDRB. August 1980 and Emmanuel Jimenez "The value of squatter settlements in developing countries". Urban and Regional Report No. 80-17. DEDRB. Nov. 1980.

2/ This has been found in Jakarta, Manila and El Salvador.

3/ This is the case in El Salvador. There is some tentative evidence in Manila that turnover may be higher for poorer families but the findings are not conclusive.

in the Dandora sites and services project have opted to continue living in a nearby squatter settlement and are subletting the house.

The medium range studies have also provided important information on the costs to the families of the process of house consolidation. By the nature of progressive development, each family makes its own decision as to how much to invest in completing the house so the total cost can only be obtained by studying a sample of families over a period of years.

Long-run outputs

Although the above mentioned short and medium-run outputs are useful by-products, the main claim to utility of the longitudinal studies is the information they can provide on the long-run impacts of the projects on participating families. How well have the studies been able to evaluate project impact and what type of results are being produced?

There are several limitations on the longitudinal studies which sometimes reduce their utility. Often there are considerable delays in project implementation, so that by the time a five year evaluation is being completed, it may still be too early to provide conclusive results. In many cases the participants being studied may only have been living in the project or may only have been affected by upgrading, for about 2 years. A second limitation is that many of the changes in which we are interested may take much longer than five years to appear. Health and demographic changes are slow to develop, and improved productivity through higher school attendance may also take many years before the benefits are felt. Despite these constraints a number of useful results are beginning to be produced, the most important of which are the following:

- a) Preliminary evidence suggests that most projects do not have a significant impact on earned income, but there is a very significant impact on rental income. For poorer families the increased rental income can account for 25% or more of total family income.
- b) Turnover rates among house owners appear to be very low in most cases. It does not seem to be the case that the poorer families prefer to sell their house and realize a capital gain. The information is not yet so clear on the impact of the project on renters.
- c) It is possible for families to complete a good quality house for less than half (and often only about 25%) of the cost of the cheapest type of public or private housing available.
- d) The evidence seems to suggest that the housing projects do not directly act to stimulate migration to the city. Most project participants have been living in the city for many years before the project began.

As the shelter programs expand in size, it becomes necessary to obtain much wider information on program impacts, the type of housing constructed and community preferences with respect to services. It also becomes important to have a better understanding of the dynamics of the urban housing market, and this is where the longitudinal studies can be extremely valuable. In addition to providing information on whether project benefits are received

by the target population, the panel studies can also describe changes in the composition of the population, the development of rental accommodation, the mechanisms of housing finance and preferences for different types of housing. One of the big advantages of the panel study is that it can follow particular families over time, and can thus provide a deeper understanding of the processes of change than could be obtained from cross-sectional studies.

SECTION SIX. CONCLUSIONS

World Bank urban shelter programs are intended to produce a significant impact on the social and economic conditions of the urban poor. Given these ambitious objectives, and the increasing scope of the programs, a need has been felt for systematic research which can evaluate the extent to which the programs objectives have been achieved and which can provide guidelines for the development of future programs. The most appropriate research design suggested by the social science literature for this purpose is one of the variants of the quasi-experimental design. In the present context this means the interviewing of a sample of project participants on two or three occasions at least a year apart. At the same time a group of families, selected to match participants, and who live in low income housing, is also interviewed and forms the control group.

The application of the quasi-experimental design faces a number of practical and theoretical problems not encountered when the method is applied to a smaller scale projects. Much of the paper is devoted to a discussion of the theoretical and operational validity of the research designs in terms of 4 main groups of threats to the validity of the inferences which can be made about project impact. Although several of the designs were found to be very weak, often due to operational problems beyond the control of the researchers (delays in project implementation, natural disasters, etc), the strongest research designs were found to be relatively robust and able to cope to a reasonable degree with the main threats to validity. The two main weaknesses were found to be the impossibility of eliminating the effect of particular events (natural, political or economic) which had different impacts

on the participant and control groups; and the virtual impossibility of designing a completely equivalent control group. This latter factor places some limitations on what can be said about project impact, and limits the extent to which generalizations can be made from the results.

Although these limitations are important, it is argued that many of the problems related to the non-equivalent control group can be resolved (or reduced) through the use of multiple regression analysis. It is also argued that the constraints upon the generalizability of the results are less crucial than might be thought as it will almost never be possible to completely replicate a project. The finite size of the city and the scarcity of resources such as land, materials and financing, mean that the design of subsequent projects will be significantly different from the first project studied. Consequently the evaluation results will be used as general guidelines for planning future projects rather than as a strict blue-print, so a certain margin of uncertainty as to the causal linkages may not be too important.

A review of the research designs showed that to date they have all been quite similar. All have used mixed panel designs in which the second and third surveys are repeated with the families living in the same structures where the original interview was conducted. If the same family still lives there they will be reinterviewed. If the family has moved the interview will normally be conducted with the new occupant. In this way the sample can be divided into a panel sub-sample, and a sub-sample of new families. These two can be analyzed separately or they can be combined to provide an approximately random sample of all families living in the project at a given point in

time. This design provides a strong analytical base. The only problem arises if a large number of new structures are being built. In this case, unless the replacement strategy is carefully designed there is a danger of a bias being introduced through the exclusion of families living in new structures.

The mixed panel design has been preferred to the selection of new random samples each time the interview is repeated. Although the independent sample design is easier to administer and ensures representativity, it suffers from the disadvantage that certain powerful types of statistical analysis used in the analysis of panel studies cannot be used. In particular it is not possible to regress the value of the dependent variable in T(2) on its value in T(1). This regression technique is very important in adjusting for the effects on non-equivalent control groups.

Despite the analytical benefits of the mixed panel design, there are many situations in which the matching of cases over time can be very difficult or unreliable. In these cases it may be better to opt for independent random samples at each point of time.

A significant modification is being introduced into the design of newer evaluations, particularly those referring to upgrading projects. Instead of having an external control group, the design covers the total universe potentially affected by the program. Through the use of regression analysis and the inclusion of measures of "exposure to project impact" it is possible to generate an internal control group. Each family which has not received a particular service will form part of the control group with respect to that particular component. The same family, if it has received other

services, might at the same time be part of the experimental group for the latter components. This design can produce powerful analytical models and is particularly useful in those cases where it is not possible to locate an external control group. This research design also makes it possible to evaluate the relative contribution of each project input to the total project effect. Although these new approaches, based upon multiple regression analysis, are potentially very attractive, further work needs to be done to evaluate their theoretical validity and practical difficulties.

Having established that it is possible to design relatively robust evaluation models which provide a satisfactory degree of validity in the interpretation of the findings, the final question to be asked relates to the function of longitudinal impact studies in an evaluation program. Longitudinal studies are expensive, time consuming and do not produce most of their main findings until it is too late for them to be built into the planning of the program they are studying. Given these circumstances, many program managers who have found shorter-term studies to be very useful, continue to have serious reservations about the longitudinal studies.

What then should be the function of the longitudinal study? Should it always be included in the evaluation design, or only under special circumstances? Must it always be so expensive and time consuming or are there simpler approaches?

It must be stated at the outset that one of the main functions of the longitudinal study is to contribute to the planning of future projects. Although a certain amount of short and medium run feedback can be generated, the main findings will not be produced until project implementation is

is virtually completed. As the results are mainly of use in the design of future projects, when the studies are being designed it should be asked whether it is expected there will be a future project. If there is not likely to be a future project there may be no justification for conducting the longitudinal studies.

If it is expected that there will be subsequent projects then the longitudinal studies can provide very valuable planning information. Some of the topics on which the longitudinal study can provide information better than other types of study, are the following:

- a) Modifications in the design of similar projects in the future. This can refer to physical design, affordability and selection, employment and income generation and the relative treatments of owners and renters.
- b) Structure of demand for housing and the identification of new housing designs and programs. By observing the reaction of different population groups to the project, and the changes in ownership and rental patterns, it is possible to understand which groups the project most appeals to and which types of demand are not being satisfied.

It can be seen from the above discussion that the longitudinal impact study is mainly providing information for a different group than the short-term evaluation studies. Given this fact the decision as to whether the long-term studies should be conducted should often be taken by a different group than those reviewing the short and medium range studies. In the case of

longitudinal studies the appropriate group to make the decision might be the Ministry of Planning or the higher levels of the Ministry of Housing, rather than the sites and services or upgrading project units. It also follows from this that the longitudinal studies may often be done by a different institution than the one doing the shorter term studies. Often the longitudinal studies will be sub-constructed to a university or research institution whilst the shorter term studies will be done in-house by the executing agency. This means that many of the criteria used by the project executing agency to evaluate the usefulness of the longitudinal studies are inappropriate. Longitudinal studies should be evaluated in terms of their role in long-term planning, rather than in terms of how well they can help resolve the immediate problems of the project implementer.

REFERENCES

- Bamberger, M., "Planning the Evaluation of an Urban Shelter Program: Key Issues for Program Managers", Series on the Evaluation of Urban Shelter Programs. No. 1, Urban and Regional Economics Division, Development Economics Department, World Bank. Jan. 1981 (draft)
- " " "A Basic Methodology for Impact Evaluation in Urban Shelter Programs", Series on the Evaluation of Urban Shelter Programs. No. 2, Urban and Regional Economics Division, Development Economics Department, World Bank. Feb. 1981 (draft).
- " " "Statistical Evaluation of Project Impact through Longitudinal Surveys", Series on the Evaluation of Urban Shelter Programs. No. 6, Urban and Regional Economics Division, Development Economics Department, World Bank, June 1980 (draft).
- Bamberger, M., Gonzalez-Polio, E. and Sae-Hau, U., "Evaluation of the First El Salvador Sites and Services Project", Urban and Regional Report No. 80-12, Urban and Regional Economics Division, Development Economics Department, World Bank. November 1980.
- Cohen, J. and Cohen C., "Applied Multiple Regression/Correlation Analysis for the Behavioural Sciences", Lawrence Erlbaum Associates Publishers, New Jersey. 1975.
- Cook, T.D., Campbell, D.T., "Quasi-experimentation. Design and Analysis Issues for Field Settings", Rand McNally. 1979.
- Jimenez, E. "The Value of Squatter Settlements in Developing Countries", Urban and Regional Report No. 80-17, Urban and Regional Economics Division, World Bank. November 1980.
- Lindauer, D., "Longitudinal Analysis and Project Turnover. Lessons from El Salvador", Urban and Regional Economics Division, World Bank. September 1979.
- LP3ES, "Jakarta Monograph Report", Series on the Evaluation of the Kampung Improvement Program. Prepared for the Government of the Republic of Indonesia, Directorate General of Cipta Karya, Jakarta. March 1981.
- P.T. Resources Jaya Teknik Management Indonesia, "Analysis and Evaluation of Impacts of KIP Implementation in Jakarta", Jakarta. 1979.

- Quigley, J., "The Distributional Consequences of Stylized Housing Programs", Urban and Regional Report No. 80-18, Urban and Regional Economics Division, World Bank. August 1980.
- Research and Analysis Division (RAD), "Interim Report. Tondo Foreshore Dagat-Dagatan Development Project", Report Series 80-1, National Housing Authority. Philippines. 1980.
- "House Consolidation Study", Report Series 80-2, National Housing Authority. Philippines. 1980.
- "Preliminary Estimates of Project Turnover", Report Series 80-4, National Housing Authority. Philippines. 1980.
- Sae-Hau, U. "The Use of Computers for the Cleaning and Analysis of Evaluation Survey Data: Some Practical Issues", Series on the Evaluation of Urban Shelter Programs. No 7, Urban and Regional Economics Division, Development Economics Department, World Bank. March 1981 (draft).
- Sanyal, B., Valverde, N. and Michael Bamberger, "Final Report on the Evaluation of the First Lusaka Upgrading and Sites and Services Project", Urban and Regional Economics Division, Development Economics Department, World Bank. February 1981 (draft).
- Senga, Ndeti and Associates, "Medis 8", Monitoring and Evaluation Study of Dandora Community Development Project for the Government of Kenya. Nairobi. February 1980.