Several studies have evaluated the experimental housing allowance program authorized by Congress in 1970 and implemented by the Department of Housing and Urban Development. The studies that are relied upon most have nevertheless been considered inadequate because they were essentially case studies. An evaluation design has recently been suggested, the random-comparison-group design, that makes the data from the Administrative Agency Experiment useful for evaluation. The outcomes in the AAE program sites are compared here to projected outcomes based on an analysis of Annual Housing Survey data. The evaluation finds that rent burdens were reduced and housing quality was improved for many who lived in substandard units. These impacts were achieved without producing inflation. Substantial confidence in the findings is permitted not only by the logic of the method itself, but by corroboration from other research. The design employed is found to be practical and powerful for certain evaluations, and its use in connection with other public programs can be recommended. The random-comparison-group design could substitute for randomized experiments in situations in which such controlled experimentation is not possible or was not carried out.

RENT SUBSIDIES

An Impact Evaluation and an Application of the Random-Comparison-Group Design

BYRAN O. JACKSON
Washington University

LAWRENCE B. MOHR
University of Michigan

BACKGROUND AND PURPOSE OF THE EXPERIMENTAL HOUSING ALLOWANCE PROGRAM

The experimental housing allowance program (EHAP) was authorized by Congress in the Housing and Urban Development Act of 1970.

AUTHORS' NOTE: *This research was supported in part by a grant from the Sloan Foundation to the Institute of Public Policy Studies, The University of Michigan, for studies in public management.*

EVALUATION REVIEW, Vol. 10 No. 4, August 1986 483-517

Three experiments, the demand experiment, the supply experiment, and the administrative agency experiment, were designed to test the concept of providing direct cash payments to families in need to assist them in obtaining adequate housing. The experimental program was established to answer a number of important questions concerning housing allowances including the following:

(1) Who would participate in a housing allowance program? What types of households (husband-wife, single parent)? Could both whites and minorities secure adequate housing and participate?

(2) How would participating households use their allowance payments?

(3) Would the quality of housing improve for participating households?

(4) Would a housing allowance program cause participants to change the location of their housing?

(5) Would landlords and homeowners rehabilitate substandard properties and increase maintenance?

(6) What would happen to the price of housing? Would there be significant market responses to a housing allowance program?

(7) What alternatives exist for administering the program?

(8) What are the probable costs of a nationwide housing allowance program? (Struyk and Bendick, 1981)

Even though each of the three experiments conducted under EHAP separately addressed many of the questions raised above, they all had specialized functions. The demand experiment examined the questions of how housing-allowance recipients used their housing allowances, residential mobility generated by the experiment, and the fulfillment of housing quality requirements. Conducted in Phoenix, Arizona, and Pittsburgh, Pennsylvania, the demand experiment was the only true experiment in terms of research design. Participants were randomly assigned to treatment and control groups in both cities. The cities were selected on the basis of availability of housing and the quality of the housing stock (U.S. Department of Housing and Urban Development, 1980: 20-29).

The supply experiment examined the impact of housing allowances on housing markets. The major issues addressed were the impact of the program on housing price inflation in markets where vouchers were being used, and the extent to which landlords and homeowners would rehabilitate substandard properties and increase maintenance. Two counties were selected for the experiment—Brown County, Wisconsin, and St. Joseph County, Indiana (U.S. Department of Housing and Urban Development, 1980). Unlike the demand experiment, the supply experiment's design permitted open enrollment to all income-eligible citizens within each of the jurisdictions as opposed to random selection of participants from an eligible applicant pool. Lacking a control group, the experiment relied more heavily on time-series data collected over the 10-year period, 1974 to 1984.

The administrative agency experiment (AAE), which will serve as the focus of this analysis, was designed to provide information on management issues in an allowance program. The experiment was implemented in eight different housing markets across the country by eight public agencies. Information was collected on different administrative procedures, the costs of the procedures. and the experiences of agencies and participants in the program. The criteria for selecting the agencies included interest of local agencies and government bodies in participating; market size and vacancy rate; diversity of population and housing market characteristics; and an assessment of the overall feasibility of running the AAE in the site. The sites and agencies selected were Jacksonville Department of Housing and Urban Development, Jacksonville, Florida; San Bernardino County Board of Supervisors, San Bernardino, California; Department of Community Affairs, Springfield, Massachusetts; State of Illinois Department of Local Government Affairs (Office of Housing and Buildings), Peoria, Illinois; Social Services Board of North Dakota, Bismarck, North Dakota; Durham County Department of Social Services, Durham, North Carolina; Tulsa Housing Authority, Tulsa, Oklahoma; and the Housing Authority of the City of Salem, Salem, Oregon.

The administrative agency experiment's design was the most controversial and considered the weakest among the three experiments. The major problem stems from the lack of controlled variation in the administrative functions administered by the agencies (Kershaw and Williams, 1981). Each agency was given a broad set of program guidelines and was allowed to develop its own procedures for eight required administrative functions (outreach, income certification, enrollment, counseling, services, housing inspections, payments, and termination). The rationale for this approach was to select the most successful procedure for each of the administrative functions in designing the national housing allowance program. Even though the agencies differed in terms of how they designed and implemented their administrative procedures, the eligibility requirements for participation in the program were consistent across the eight cities.

In addition to case study material collected on the implementation of the administrative functions, termed as "soft data" by HUD officials, hard data were collected on participant attrition at each stage of the program as well as on the demographic characteristics of participants. In addition, panel data on housing conditions of a random sample of the participants (n = 747) were collected as part of the AAE data base. The original intent in collecting these data was to allow for comparisons of the impact of the housing allowance program across the AAE, demand, and supply experiments. This cross-program comparison, called the "integrated analysis," never evolved according to plan. Nevertheless, a final report synthesizing results from the three separate experiments was produced by the Urban Institute (see Struyk and Bendick, 1981).

The data from the administrative agency experiment are an undervalued resource for scholars and policymakers designing and evaluating social welfare programs. It is true that the exact content of "the program" remains undefined because of variations in implementation across the individual sites. Nevertheless, the essential core of any such program is the granting of subsidies to eligible families, and this was common to the sites in the administrative agency experiment. The purpose of this article is to use the hard data collected in the administrative agency experiment both to learn something about the impact of the program and to illustrate how data collected from demonstration projects such as the AAE can be combined with more general purpose data (such as those collected through the decennial census or the Annual Housing Survey) to evaluate the impact of government programs on various social problems. The potential value of this demonstration is significant for a number of reasons. First, real-world experimentation such as the EHAP is extremely costly (see Table 1).

Second, experiments like the supply and demand experiments are limited in their generalizability given the nature and small number of cities used in each. For example, the two cities in the supply experiment—Green Bay and South Bend—have been criticized for "having market conditions qualitatively different from those in large, old, decaying, urban centers such as Detroit, Cleveland or Newark" (Kain, 1981: 358). Given the inability to generalize from the EHAP results, the Department of Housing and Urban Development has in fact authorized a new Housing Demonstration Project costing almost $1.5 million. This

TABLE 1
Experimental Housing Allowance Program Costs
(in millions of dollars)

Experiment	*Payments to Households*	*Administration and Program Operation*	*Research and Monitoring*	*Total*
Demand	$ 4	$ 2	$25	$ 31
Supply	40	18	41	99
Administrative agency	10	3	9	22
Integrated analysis	0	0	7	7
Total	$54	$23	$82	$159
Percentage	34	15	51	100

SOURCE: Struyk and Bendick (1981: 297); estimates as of April 1980; copyright 1981 by The Urban Institute.

demonstration, according to Secretary of Housing and Urban Development Samuel Pierce, "will show that vouchers are the most efficient vehicle for providing housing assistance to low-income families, regardless of where they live—in big cities, small towns, or rural areas (U.S. Department of Housing and Urban Development, 1985).

Third, personnel used in these two experiments were affiliated with the experiment as opposed to being drawn from the local personnel population that would normally run a housing allowance program. Their involvement serves as a possible contaminating factor for the results obtained. Finally, the use of experimental and control groups such as were employed in the demand experiment is politically infeasible in most studies given that every group has a desire to get some of the treatment. Therefore, the potential for alternative designs to yield comparably reliable results is an important issue.

For these reasons we give special attention to data collected in the AAE, which represents the typical demonstration project used to test the effects of government subsidies in that the design involves no random assignment. The AAE represents an opportunity for the employment of the "random-comparison-group design" (Mohr, 1982) in testing the effects of a housing allowance program on low-income families. The design differs from most others in that it purports to accomplish much of what is accomplished by a randomized experiment, but without the random assignment of subjects to treatments (another example is the regression-discontinuity design; see Trochim, 1984). No application of the random-comparison-group design has previously been published.

RESEARCH DESIGN AND RESULTS

TREATMENT AND COMPARISON GROUP SAMPLES

This article demonstrates the use of the random-comparison-group design in cases where researchers are either unable to or fail to assign clients to treatment and control groups at random. In the Experimental Housing Allowance Program's administrative agency experiment, evaluation data were collected only on clients who received the treatment. The data collection effort involved taking a random sample of clients who enrolled in the program and collecting extensive information on their housing conditions before and after the program. The data were collected in three waves (consisting of the same program clients) immediately before the treatment, immediately after the treatment, and near the termination of the program. The initial sample had a total of 795 enrollees, but not all of them had complete records. Eliminating the latter left a sample size of 747 for the purposes of this analysis. In addition to demographic information, such as race, sex, and household size, data on housing conditions, rent, and mobility (before and after the program) were collected.

The comparison group to be employed and discussed in the next section was constructed for the analysis using similar data from the national file of the Annual Housing Survey for the years 1974 and 1975, administered by the Department of Housing and Urban Development. This time period closely approximates the time period in which the experiment was run. In determining the comparison group, only households that satisfied the criteria used in certifying clients for the EHAP program were chosen. The unweighted sample size for the Annual Housing Survey comparison group was 3040. The Appendix outlines in detail the criteria used in selecting households for the comparison group.

MEASURING HOUSING DEPRIVATION

The term housing deprivation is used to describe areas of housing need that are most prevalent among low-income families needing housing assistance. The three most common indicators used in determining the level of a family's housing deprivation are the percentage of income a family pays for rent, the number of persons per room in a housing unit, and the quality of a family's housing unit. Neighborhood quality is also used, but less frequently, as a measure of housing deprivation.

When measuring a family's rent relative to its income, most analysts use 25% as a crude demarcator of rent burden. Families paying more than 25% of their income for rent are considered as having an undesirable rent burden. In terms of crowding, the total number of rooms in a housing unit is divided by the number of residents living in the given unit to determine persons per room. Families with more than 1.5 persons per room are considered to be living in overcrowded housing.

The indicators for measuring housing quality are less straightforward than those for rent burden and crowding. The U.S. Bureau of the Census at one point used lack of plumbing as the only measure of housing quality. Census enumerators later began classifying units as "dilapidated" and "not dilapidated." Today, the question of what indicators to use in measuring housing quality remains a topic of much debate. The Appendix outlines in detail the operationalization of housing quality, as well as the other measures of housing deprivation used in this analysis.

For the purposes of this analysis attention will be devoted to four variables: the magnitude of reductions in out-of-pocket rent itself (controlled for income), rent burden, number of persons per room, and the incidence of substandard housing. The major questions raised are: What was the impact of the program in reducing rent, rent burden, crowding, and substandard housing for participating clients in the eight sites under investigation? Second, we ask: How was the subsidy used? That is, how much of it was spent to ameliorate each of the types of deprivation targeted? These goals resemble the original goals of the demand and supply experiments much more than those of the AAE; the purpose is in part to investigate the value of AAE-type data as a substitute for or supplement to the other, more unusual kinds of experiment.

These questions do comport with the design and goals of the housing allowance program. The program operated on the notion that families could find decent, safe, and sanitary housing at affordable prices if the federal government could provide them with a housing voucher reflecting the difference between 25% of their income and the cost of a *standard unit* in the market in which they lived. By examining the average reductions in the areas of deprivation, one would hope to be able to estimate the extent to which these goals were fulfilled.

TABLE 2
Results Based on the Before-After Design

	Before	*After*	*Impact*
Rent	$110.35	$63.64	−$46.71
Rent burden	.43	.23	−.20
Quality	.12	.06	−.06
Crowding	.70	.65	−.05

RESULTS

The work presented here will measure program impacts on housing deprivation employing two different quasi-experimental designs. First, we employ the "before-after design" or Campbell and Stanley's "Design 2" (Campbell and Stanley, 1966: 8). Design 2 is applied to the treatment group alone. Next, the random-comparison-group design (Mohr, 1982) will be employed using the treatment group and the Annual Housing Survey comparison group developed for this analysis.

Design 2. The before-after or pretest-posttest design rests on the assumption that if not for the program, housing deprivation for the treatment group would have remained constant at preprogram levels. Therefore, the program's impact may be obtained by subtracting the "before" measures from the "after." Table 2 presents these results for rent, rent burden, housing quality, and crowding.

If one were to base one's conclusions on an analysis of the Design-2 data shown in Table 2, the following observations would seem to be indicated: On average, the subsidy produced a rent saving of $46.71 per month, which means that nearly half (42%) of the original or "before" rent was freed for other uses. Rent burden, similarly, went from 43% of income to 23%, so that the burden was approximately halved by the subsidy. The subsidy does appear to have brought the rent burden down below the level of the 25% mark, which was a goal. The housing quality data indicate that the failure rate went from 12% of the items failed to 6%. Because quality was measured on a six-item scale (see Appendix), these figures mean an improvement from an average of 0.72 items failed per household before to 0.36 items afterwards. In other words, the housing of these individuals was only slightly substandard to begin with and the program improved quality by halving the average number of items failed. This is a modest impact in terms of raw improvement, but improvement generally becomes quite difficult to achieve as one approaches the very bottom or top of a measurement scale, as here. Crowding appears truly not to have been a problem to begin with. Given a maximum of 1.5 persons per room as a standard, the before-program average of 0.7 persons per room represents adequate conditions; it compares well with the commonly held middle-class model of a family of four in a six-room house (0.67 persons per room). The change from before to after is negligible, but little change in the average figures was, under the circumstances, to be expected. To summarize the above, it appears from the before-after data that quality and crowding were not major problems for these groups and that only a small portion of the subsidy was spent on improvement in those categories. The bulk of the impact of the program apparently went to bring the rent burden down to a reasonable level.

These results may be trusted, however, only insofar as one may safely assume that, on average, nothing else besides the subsidies occurred to change housing status on these dimensions over the time period covered. That is too strong an assumption. From general observation and experience alone, one would guess that rents would creep upward over time, often at a greater rate than income. Crowding might well become worse through both increase in household size and the choice of some to move to smaller quarters rather than accept higher rent. Housing quality might also tend to get worse, both as an alternative to higher rent and as the result of unchecked deterioration over time. Using the unmodified "before" data risks distorting the true effects of the subsidies by ignoring these possible hidden movements and assuming that, in the "null case" (that is, without the treatment, or with a zero treatment effect) housing status after the term of the program would not have changed at all from before. (We will see when the random-comparison-group analysis is reported that some of the conclusions based on the before-after data are in fact quite reasonable; if we had *only* these data, however, and so needed to be concerned about events other than the subsidy program that might have affected the posttest scores, it would be quite difficult if not impossible to decide whether they were reasonable or not.) As an alternative to undertaking a more elaborate and sophisticated analysis, we might simply assume that the differences or impacts in Table 2 are conservative, as the true null-case posttest conditions are probably worse (higher scores) than the before data indicate, but (a) we cannot really be sure of that; they may in fact be about right or may even err on the liberal side, (b) if the before-after results are indeed conservative, we do not know how conservative, and (c) we would also be essentially uncertain whether the differences are conservative in all of the categories of deprivation or just some of them, so that there would be substantial ambiguity regarding the distribution of the effects of the subsidies across the categories of rent, quality, and crowding.

Before leaving Design 2, it should be noted that the results on all of these before-after differences are statistically significant. That is rather a meaningless and irrelevant bit of information, however, because it tells us only that it is unlikely that the before-after differences were caused by random forces. It does not tell us what we are really concerned about, namely, the extent to which the differences are owing to *non*random forces other than the treatment, such as the kind of normally expected change suggested above or, perhaps, the occurrence of unusual events.

In short, it would be highly desirable to have better indicators of the null-case outcomes than we have in these "before" measures. When before measures are available but cannot be trusted completely, the logical place to turn is in the direction of a comparison group. Optimally, the treatment and comparison groups should be composed by a process of randomization—beginning with one large group of cities or individuals and subdividing it at random into experimental and control groups. In this case, that simply was not done; the evaluators must make do with the treatment group that was actually used and whatever comparison groups they can find. Furthermore, it is doubtful that political and bureaucratic realities would have permitted a rigorous randomization even if it had been pressed by those thinking ahead to the needs for evaluation. The design that generally results in such circumstances is "Design 10" (Campbell and Stanley, 1966: 40), in which the comparison group is essentially an arbitrary one, usually judgmentally selected to be as much like the treatment groups as can be obtained under the circumstances.

We do not have data on such a group to present here. Moreover, it is important to recognize that the exercise would be essentially uninformative. We intend to present data on a different sort of comparison group—one that gives an accurate idea of the null-case outcomes. Had we a regular Design-10 comparison group as well, we might have found either that it gave exactly the same accurate information as the group we actually used or that it was off by a little, or perhaps by a lot in one direction or the other, leading toward a certain distortion of the program effects (such distortion is generally called "selection bias"). But that would be just this one case. It would say nothing about Design-10 comparison groups in general. We could reach no conclusions about such groups beyond the single case at hand. Moreover, if we had *only* the Design-10 comparison group, there would be no way of knowing how much it distorted outcomes by selection bias in the *present* case, or in which direction. One could compare it to the treatment group on certain available characteristics such as income, age, sex, and racial composition, but these particular "selection" controls are certainly not all of the variables that matter for rent, crowding, and housing quality. In principle, one can never know in such a situation whether one has taken account of all of the variables that do matter in a substantial way or not. The null-case outcomes would therefore remain quite uncertain.

The random-comparison-group design. Instead of an arbitrary or judgmentally "close" comparison group, we have opted for what has been called a "random comparison group" (Mohr, 1982). In ordinary Design 10, potential selection bias refers to the possibility that the treatment and comparison groups differ *from one another* on variables that could, in whole or in part, determine outcomes. The critical feature of the random-comparison-group design is that it introduces a "criterion population" and shifts the concern to possible differences of the treatment and comparison groups, not from one another, but *from that population.* Each group is thus a separate and delimited source of potential bias. The difference in the two approaches would seem to be negligible, but a further characteristic of the random-comparison-group design is that it stipulates the employment of the criterion population itself, or a random sample from it (hence the label of the design), as the comparison group for the evaluation study. This eliminates the threat of selection bias from one of the two possible sources (the comparison group), thus "minimaxing" beforehand the total possible bias. But further, it provides a basis for using statistics to derive an interval estimate of program effect in which one can have a substantial degree of confidence (Mohr, 1982: 64, 66-71).

The criterion population should (1) include the treatment group, (2) be very much larger than the treatment group, and (3) ideally be the population to which one would like to generalize the results of the evaluation. In this case, we chose the whole country as the criterion population—guided by these three desiderata and by the knowledge that the Annual Housing Survey of a national sample existed—and selected the study sample as described in the Appendix. We note in passing that the criterion population need not always be a national sample of individuals or households. Depending on guideline 3, it might be the organizations of a certain type in a state or region, the children of a school district, the professional employees of major accounting firms, and so forth. Finally, it should be noted that populations that include the treatment group have no doubt been used from time to time for comparison groups in program evaluation. In fact, some attempt was made to construct a comparison group through the use of Annual Housing Survey data in the 1981 evaluation of the Section 8 Housing Assistance Program (U.S. Department of Housing and Urban Development, 1981). However, the design employed lacked rigor and made no attempt to adjust for self-selection into the program. In general, the gain will never be great if the analysis based on such a design is a straightforward regression (or comparable) analysis, for then the statistical validity of the causal inference depends on treating the data as though they were from a randomized experiment. That is, the results are only valid if the treatment group can be considered tantamount to a random sample from the population, which, of course, it generally cannot be. A modified statistical approach is needed to extract the true benefits of the comparison group's being an inclusive population.

The random-comparison-group design and Design 10 are similar in structure—both designs feature primarily a pretest and a posttest on a treatment group and a comparison group (other controls or predictors may be used in place of or in addition to the pretest). Analytic procedures for the two are fundamentally different, however. The first difference is that, in Design 10, all subjects in both groups are thrown together—in a regression analysis, for example—to derive estimates of the parameters in question, whereas in the random-comparison-group Design only the comparison group is used for most parameter estimation. It makes sense to use both groups in Design 10 because the question of bias involves differences from one another and neither group can therefore be the more valid or preferred source of information about parameters. It makes sense to use only one group in the random-comparison-group design, however, because the bias question involves differences from a criterion population and one group *is* that population, or a random sample from it. When one needs the regression parameter of posttest on pretest, for example, the true population magnitude is to be found directly in the comparison group (within random-sampling error); mixing the treatment group into the analysis as well yields a bigger "N," but that in no way begins to compensate for the fact that it also unnecessarily adds uncertainty and confusion about the magnitude concerned. The treatment group's contribution to parameter estimation may be flawed, both because the treatment itself has muddied the waters and because the sample is not a random one.

Before proceeding, we take this opportunity to alert the reader that the analysis must necessarily be somewhat involved. The random-comparison-group approach is new and, unlike the fine points in an application of ordinary statistical inference, its various aspects, offshoots, and subparts are not second nature to the average reader.

The first step in the analysis, then, is to estimate within the random comparison group alone the relevant relations of posttest, or deprivation scales, to all predictors to be employed except the treatment. These data are shown in Table 3.

The regression equation on which these results are based has the following form for each dwelling unit in the Annual Housing Survey sample:

$$Y = a + b_1X_1 + \ldots + b_kX_k + e \qquad [1]$$

where Y is an outcome—a housing deprivation score; a is the constant or Y-intercept; X_1 through X_k are predictors of the outcome, including the before measure; b_1 through b_k are the associated slopes or regression weights; and e is the error term. Geometrically, this equation denotes a (k + 1)-dimensional hyperplane. Because a picture becomes essential in the following analysis, however, we present for illustrative purposes in Figure 1 a plot of postprogram rent on preprogram rent alone, omitting all other predictor variables in order to get down to graphable dimensions. (This particular relationship is selected simply because it represents the best prediction of an outcome by a single predictor in the study.)

Next, let us symbolize the program effect for a given outcome measure, such as rent, as b_T'. This is obtained by simple subtraction of what would have been from what was, that is, subtraction of the expected *null-case* mean outcome for the treatment group (predicted on the basis of the comparison group experience) from the *actual* mean treatment-group outcome:

$$b_T' = \overline{Y}_E - \hat{Y}_{OE} + e_{\hat{Y}} \qquad [2]$$

TABLE 3
Comparison Group: Parameter Estimates and t-Values

Predictor Variable	*Rent 75 Par. Est.*	*t-Value*	*Predictor Variable*	*Rent Burden Par. Est.*	*t-Value*
	Rent 75			*Rent Burden*	
Intercept	14.97	6.01	Intercept	56.81	41.29
Rent 74	.87	80.99	Rent burden 74	.36	25.79
Black	−0.80	−0.70	Black	−1.90	−3.02
Midwest	−3.86	−2.67	Midwest	−3.23	−4.07
South	−5.64	−4.02	South	−8.77	−11.71
Mover	.27	.25	East	2.49	3.44
Age	−0.19	−0.55	Mover	.78	1.31
Female	1.92	2.05	Age	−1.48	−7.65
Income 75	.002	5.50	Female	−0.85	−1.60
Household size	1.46	4.69	Income 75	−0.007	−40.66
			Household size	1.42	8.29
	Housing Quality			*Crowding*	
Intercept	.05	4.03	Intercept	7.57	3.12
Quality 74	.65	46.90	Crowding 74	.67	51.29
Income 75	-7×10^{-6}	−4.52	Income 75	.004	12.54
Household size	.0095	5.46	Black	1.96	1.64
Black	.04	5.82	Age	−1.14	−3.35
Female	−0.02	−3.87	South	−1.61	−1.12
Age	.0002	−0.09	East	−4.75	−3.52
South	.04	5.20	Midwest	−2.92	−1.96
East	.01	1.37	Female	2.05	2.12
Mover	.006	1.03			
Midwest	−.009	−1.16			

where E refers to the experimental or treatment group; $\bar{Y}$ is a mean outcome or posttest score, such as mean rent; $\hat{Y}$ is a predicted mean; O stands for the null case (so that $\hat{Y}_{OE}$ symbolizes the predicted or expected null-case outcome for the treatment group), and $e_{\hat{Y}}$ is the conservative half of a 95% confidence interval around the predicted value $\hat{Y}_{OE}$ (see Mohr, 1982: 68-69).

In equation 2, $\bar{Y}_E$ is obtained from the observed data for the treatment group, given previously in Table 2 and repeated in Table 4 for convenience. $\hat{Y}_{OE}$ is easily obtained by plugging values into equation 1, as follows: The regression coefficients in that equation ($a, b_1, \ldots, b_k$), are the estimates of the population parameters for equation 1, obtained from the Annual Housing Survey sample (comparison group) and given

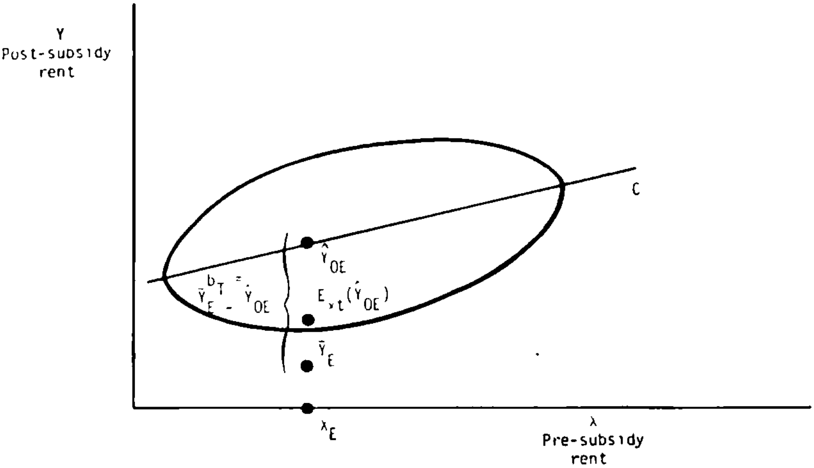

Figure 1: Predictions of Null-Case Outcome from the Random-Comparison-Group Design

in Table 3. The values used for the variables X_1 through X_k are the treatment-group means for the respective predictors. The error term is, of course, omitted in obtaining a predicted, or expected score.

In sum, the predictor-variable data (using all predictors) for the treatment group are weighted by the respective population parameter estimates from the comparison group to give the outcome expected if there had been no program. Or, to put it another way, we consider the treatment group to be members of the referent population with certain characteristics, and we use sound estimates of the population parameters to project outcomes for subjects with precisely those combinations of characteristics. That is the strength of the random-comparison-group design; at best, it uses a correct model of the relevant experience of the correct population to infer what would have happened to the treatment group in the absence of the program.

Table 4 shows $\hat{Y}_{OE}$ and b_T', as well as $\bar{Y}_E$, for all of the housing deprivation scales considered. A better view of the estimation technique may be obtained by examining the plot in Figure 1. There, the predicted treatment group mean for the null case, $\hat{Y}_{OE}$, is seen to fall on the regression line above $\bar{X}_E$, the pretest mean (which, in this simplified, two-dimensional version stands for the centroid of the space defined by all of the predictor variables actually employed), and b_T', the program impact, is readily seen to be the distance between this predicted point

TABLE 4
Results Based on the Random-Comparison-Group Design

	Rent	*Rent Burden*	*Housing Quality*	*Crowding*
$\overline{Y}_E - \overline{X}_E$	−46.71	−0.2	−0.06	−0.05
$\overline{Y}_E$	63.64	0.23	0.06	0.65
$\hat{Y}_{OE}$	118.42	0.42	0.13	0.65
$e_{\hat{Y}}$	1.23	0.007	0.005	0.001
b_T'	−53.55	−0.18	−0.07	0
s_{uC}	24.66			
$1.96s_{uC}$	48.33	0.27	0.27	0.49
Ext ($\hat{Y}_{OE}$)	70.09	0.15	−0.14	0.16
b_T	−5.22	0.09	0.21	0.49
s_{uE}	29.66	0.14	0.1	0.18
$1.28s_{uE}$	37.96	0.18	0.13	0.23
Ext*($\hat{Y}_{OE}$)	108.05	0.33	−0.01	0.39
bT*	−43.18	−0.09	0.08	0.26
R_E^2	0.43	0.3	0.09	0.68
R_C^2	0.78	0.48	0.57.	0.75

and the observed outcome, $\overline{Y}_E$ (in the simplified plot, we disregard the small quantity $e_{\hat{Y}}$). That distance, −$53.55 (see Table 4), is fairly substantial, indicating a moderate to strong effect of the subsidies on rent.

It is of some interest in passing to compare the $\hat{Y}_{OE}$ row of Table 4 with the before measures. For example, the effect of the program on rent (controlled for income as well as other factors) is apparently greater than use of the before measure suggested; the before measure ($110.35) does indeed underestimate what rents would have been on average without the program ($118.42).

The analysis in terms of $\hat{Y}_{OE}$ and b_T' makes a strong assumption, however, that is generally not warranted—namely, the assumption that the treatment group is typical of the population, as though it were a random sample. In principle, the predicted value $\hat{Y}$ for any given X should be exactly on the regression line *only* if the subjects were randomly selected from among those with that X score, or were a random sample of the whole population whose mean happened to be X. Because our treatment group was not randomly selected, however, perhaps its mean would, without the subsidy program, have fallen above or below the line rather than right on it. Unfortunately, we now have no way of knowing exactly where it would have fallen. In other words, although there is no selection bias in the comparison group, there may still be some in the treatment group; given that it is nonrandom, we have no basis for assuming that it is "typical." To put it another way, one considers the population parameters to govern the treatment group, but considers the treatment-group *disturbance* terms to be a nonrandom subset. What then should be our estimate of the null-case outcome?

In this case, which is the standard situation in the random-comparison-group design, it is well both to be conservative and to look to the within-group error variances for guidance. Mohr (1982) suggests two "extreme" or conservative predicted values—Ext($\hat{Y}_{OE}$) and Ext*($\hat{Y}_{OE}$) —depending upon the proportion of the treatment group one feels or fears might possibly, in the null case, have been posttest outliers on the population regression plot. Before selecting one of these estimators, let us take a moment to clarify the concept of "outlier." An outlier is defined here to be an individual who falls, say, in the 2½% of the population that would—without the program—have been furthest from the regression surface on the favorable or low-rent side. In other words, an outlier is someone in the 2½% of the population that would have been "best off" without the program. In a normal distribution, which these error terms approximate rather well, that 2½% would fall beyond 1.96 error standard deviations below the surface.

How much of the treatment group would have been included in this 2½% of the population—that is, how many would have been low-deprivation outliers without the program? One extremely conservative assumption is about 50%. That is, *even without the program,* approximately 50% of the treatment group would have become so well off in housing that they would have fallen into the best 2½% of eligibles in the country by the time of the posttest. That would place the *mean* outcome score for the treatment group at the 2½% point, or Ext($\hat{Y}_{OE}$) in Figure 1. To show just what sort of null-case assumption this is and the relation of the subsidized families to their counterparts in the rest of the country, a simplified version of the treatment-group plot is superimposed on a similar version of the comparison-group plot in Figure 2, with the treatment-group mean at the "extreme" point, Ext($\hat{Y}_{OE}$) (the true plots are shown in Figures 3a and 3b for orientation).

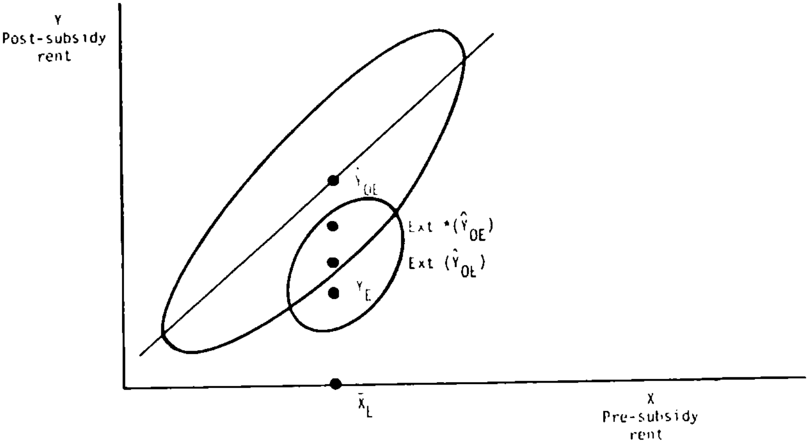

Figure 2: Conservativeness of the Prediction Ext($\hat{Y}_{OE}$) When There Is Variance in Y_T

This technique gives us an *interval estimate* of the magnitude of the program effect, b_T, such that the magnitude is equal to or more favorable than the distance between the observed outcome and this conservative estimate or extreme point, that is,

$$\mathrm{Ext}(\hat{Y}_{OE}) = \hat{Y}_{OE} - 1.96(S_{uC}) \quad [3]$$

and

$$b_T \leq \overline{Y}_E - \mathrm{Ext}(\hat{Y}_{OE}) + e_{\hat{Y}} \quad [4]$$

or, from equation 3,

$$b_T \leq \overline{Y}_E - \hat{Y}_{OE} + e_{\hat{Y}} + 1.96(S_{uC}) \quad [4a]$$

where s_{uC} is the error standard deviation for the comparison group and $\hat{Y}_{OE}$ and $e_{\hat{Y}}$ are as defined in equation 2 (one must be careful to make the sign on $1.96(s_{uC})$ such that the new estimate is more conservative than that given by equation 2—from which inequality 4a differs essentially only in its last term).

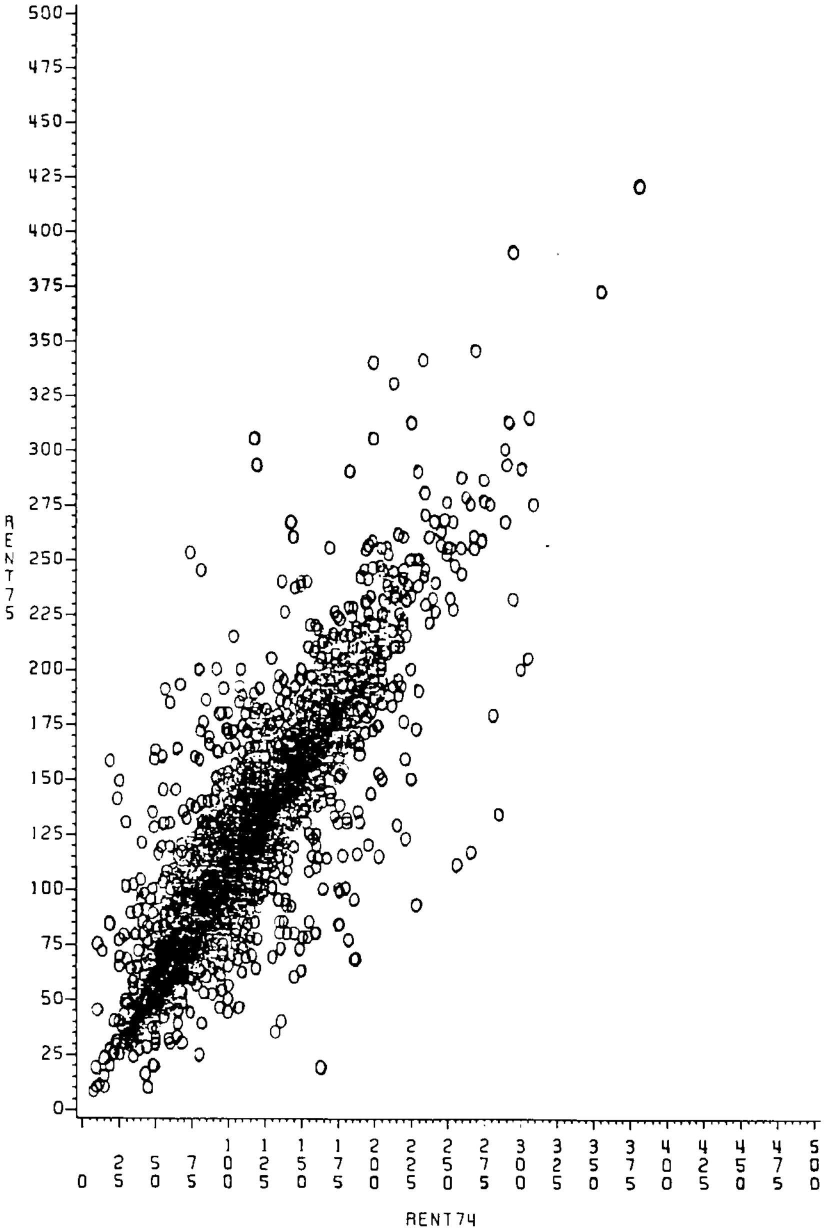

Figure 3a: Comparison Group Plot: Post-Rent * Pre-Rent

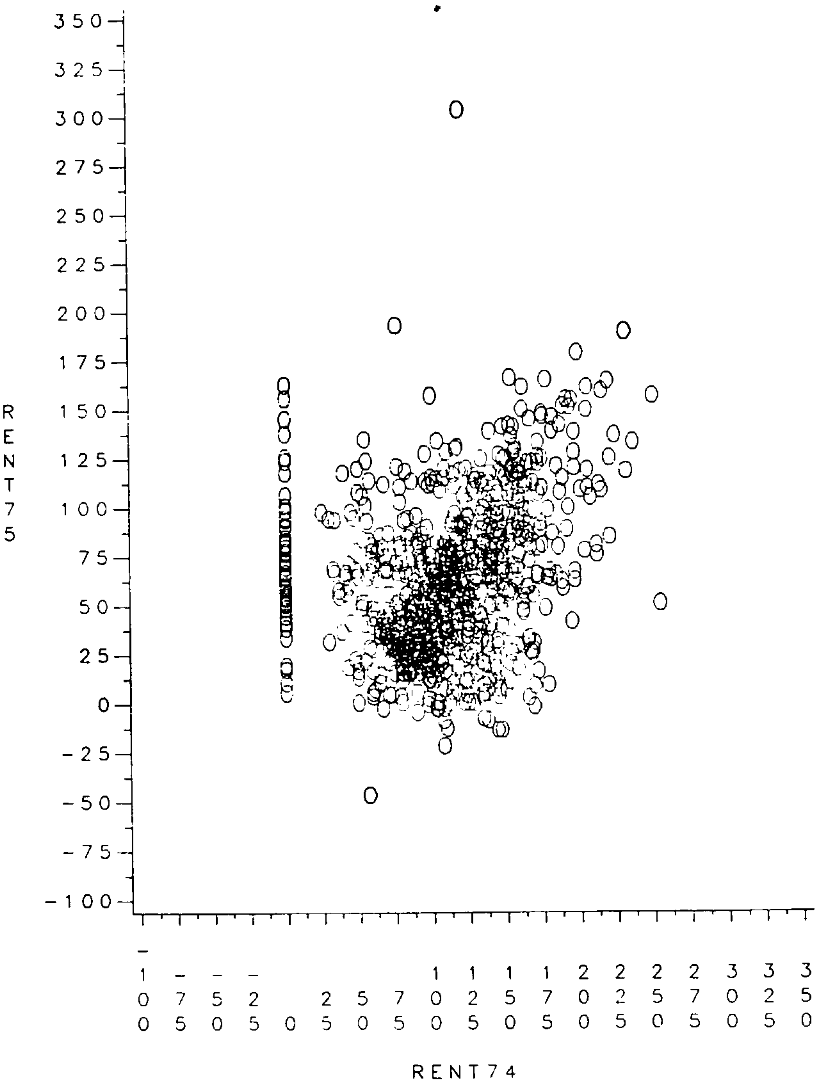

Figure 3b: Treatment Group Plot: Post-Rent * Pre-Rent

Before continuing, it is well to note that the interval estimate rendered in this analysis is not a conventional confidence interval for a regression coefficient. The role played by such a confidence interval in making an estimate is, in this analysis, essentially taken by the prediction-error term, $e_{\hat{Y}}$. Indeed, as is clear from equation 2, b_T does not estimate a parameter in a regression model at all, but rather is a "free standing" symbol indicating the effect of a treatment.

It is clear from Figure 2 that the average observed or posttest rent—$\overline{Y}_E$—is barely more favorable than it would have been without the program (–$5.22—see Table 4), given this very conservative estimate of what would have been, namely, Ext($\dot{Y}_{OE}$). From the analysis, it also becomes clear what conditions are necessary to be able to infer a favorable effect with so conservative a technique: either (a) the true program effect must be extremely large, so that the mean outcome for the treatment group is far more favorable than the mean outcome for individuals in the population with the same beginning characteristics, or (b) the error variance must be extremely small, that is, individuals in the population must be clustered tightly around the regression surface, so that even with a moderate program effect, $\overline{Y}_E$ will lie far beyond the extremes. The latter condition may in fact occur frequently in social programs because an outcome or posttest is often highly predictable from the *pretest*. Thus, it should be noted in passing, the pretest can assume substantial importance in evaluation design and analysis (Mohr, 1982: 71-72)—much more so than in theoretical research. In the present case, however, it appears that there is so much natural instability in income and housing conditions among the subject population that, except for the category of rent, the prediction accuracy of "after" from "before" and other measures, although substantial, is still too low to make a moderate program effect detectable with so conservative an approach (see Table 3 and 4).

There will doubtless be general agreement, however, that the approach is unwarrantedly conservative in the present instance if one wishes to make a reasonably accurate estimate of the effect of the program rather than simply avoid error. The treatment group is a broadly based selection of eligibles from all major regions of the country, not chosen on the basis of any apparent peculiarities. Its members are probably fairly typical of the population as a whole. We may be unwilling to assume that they are in essence a random sample, but we certainly do not need to assume that they are so extreme that 50% of them would have been outliers, as in Figure 2. One way to look at the possible conservativeness of the assumption in the general case is the following: The estimate b_T as defined in inequality 4 is ultraconservative to the extent that the treatment group is heterogeneous, that is, to the extent that the original variance in its Y scores is substantial. The extreme opposite—a homogeneous treatment group—would be a group with one common outcome score, as though it were one individual. In the null case, that one score could reasonably have fallen anywhere in the plot. Estimating that it would have fallen at the 2½% mark is not terribly conservative because that estimate does not make the score into a hardly believable outlier. Substantial variance in the outcome scores, however, makes outliers out of 50% and *very distant* outliers out of as much as 30%-40% of the group when its midpoint is placed at the extreme point, as in inequality 4. To see this visually, note in Figure 2 how far the treatment-group plot in this assumed position sticks out below the population plot of which it is supposed to be a subgroup.

An excellent way to moderate the ultraconservatism is by assuming a still-conservative but more believable proportion of outliers (after defining the outlier point as, say, 2½%). One might allow 10% outliers for the null case, for example, instead of either 2½%, as in equation 2, or 50%, as in inequality 4. To do this, one would move the expected null-case mean for the treatment group from the outlier point toward the center by $1.28(s_{uE})$, which covers 40% of the cases in a normal distribution, leaving 10% rather than 50% as outliers. We would then have a new prediction, $Ext^*(\hat{Y}_{OE})$, as follows (see Figure 2):

$$Ext^*(\hat{Y}_{OE}) = Ext(\hat{Y}_{OE}) + 1.28(S_{uE}) \qquad [5]$$

and a new program effect parameter, b_T^*, as follows:

$$b_T^* \leq \overline{Y}_E - Ext^*(\hat{Y}_{OE}) + e\hat{Y} \qquad [6]$$

or, from equations 6, 5, and 3,

$$b_T^* \leq \overline{Y}_E - \hat{Y}_{OE} + e_{\hat{Y}} + 1.96(S_{uC}) - 1.28(S_{uE}) \qquad [6a]$$

It is still rather difficult to believe that 10% of the treatment group are outliers given the similarity of original means and variances and other characteristics of the treatment and comparison groups in the present case, but that is a reasonable penalty to incur for an estimate of minimum program effect in which both analyst and audience may have substantial confidence.

Table 4 shows that for rent $b_T^* = -\$43.18$. The connection between the treatment group and the population employed as the comparison group permits this result to have a valid statistical interpretation (not possible in Design 10), as follows (Mohr, 1982: 71): Assuming that as many as 10% of the treatment group might have been population outliers on the favorable side—that is, beyond the -.025 level without the treatment—one may say with 97.5% confidence that one of the effects of the program was to lower average rent paid by \$43.18 or more. Similarly, we find that, at the same level of confidence, rent burdens were reduced by an average of at least 0.09. Because the final outcome $\overline{Y}_E$ for rent burden was observed to be 0.23, the program may be said here, as with the before-after design, to have brought average rent burden below the 25% mark, as desired. The before-after design suggested a modest effect of the program on housing quality (-.06). If the treatment group were a *random* sample of the eligible population, the random-comparison-group method would also find a modest effect (in Table 4, $b_T' = -.07$). Given the hedging necessitated by the nonrandomness of the treatment group, however, no favorable impact of the program on housing quality may be inferred with confidence. Similarly, no favorable program impact on crowding may be inferred, although in that case, as noted previously, none was to be expected; all relevant measures indicate that crowding would not, on average, have been a problem without the program.

It will be useful at this point to say a brief, clarifying word about the relation of the random-comparison-group design to selection bias. There is essentially no way to overcome the possibility of selection bias in a design featuring the arbitrary selection of treatment and comparison groups. Methods that approach the problem by using additional data to model the selection decision—model the differences between the two groups—depend on getting those differences just right (see Blumstein and Cohen, 1979; Heckman, 1976). However, one never knows whether one has included all of the important variables. Amemiya's (1973) method for overcoming the selection bias involved in truncated samples does not apply to the ordinary case of program evaluation because of the absence of a truncation point: All relevant scores on the dependent variable may fall anywhere along its entire range. The present method is different from these others in two respects. It operates in part by means of the selection process itself, which is how the bias is neutralized in true experiments and in the regression-discontinuity design (Campbell and Stanley, 1966). In this case, there is the selection of a certain sort of comparison group. For the rest, it does not try to neutralize selection bias so much as to apply a practical and persuasive hedge against it.

Use of the subsidy. Having derived solidly based, if perhaps conservative, estimates of the various impacts of the program, it remains to speculate on how the subsidy was used by the recipients. That is not an easy task. To accomplish it precisely, one would want to know the exact conditions of housing, rent, and income that would have prevailed in the absence of the program—what we have called the null-case outcomes—and then the true conditions at the posttest time selected. Part of the subsidy could then be deduced to have been spent by that time for specific improvements in the housing standard and the balance for reducing the rent expenditure, thus freeing funds for other expenses of living. We may do this only roughly for three reasons.

First, we do not have firm data on the null-case conditions. We have only the before measures, whose accuracy must be questioned, and the estimates based on the comparison group, which must be hedged. The latter are clearly the preferred *type* of indicator; we will therefore use the estimates based on the comparison group and resort to providing conclusions in a range that is somewhat wider than would be desired.

Second, the housing conditions on which we have data are only quality, as measured, and crowding. These leave out many possible expenditures that are housing related and to which the subsidies of some households may have been devoted such as quality of neighborhood, draftiness, paint condition, privacy, location, esthetics, and elements of total space such as porches, bathrooms, basement, yard, closets, and size of rooms. The key to reaching reasonably accurate conclusions about expenditures for improved housing without these details lies in the variable "rent." The use of the subsidy may conveniently be divided into just two parts—the amount that went into higher rent and the amount that went into relief of the rent burden (that is, smaller out-of-pocket expenditure for the same rent). Both parts may readily be calculated from the available data. The portion of the subsidy spent for improved conditions of housing is contained in the first part—the amount spent for higher rent.

Third, once it is determined how much of the subsidy was devoted to higher rent, it is still unclear how much of that amount simply covered a price increase and how much went toward better housing. As noted, we do not have detailed measures of housing conditions in the AAE data for the treatment group, nor do we have them in the Annual Housing Survey data for the comparison group. With a few reasonable assumptions, however, we can go a substantial distance toward separating price increase from quality increase. We know that some families in the program moved and some did not. We will assume that those who moved did so to change their housing conditions in one or more desirable ways (indeed, moving in the context of this program was done in order to get out of unacceptable housing and qualify for the subsidy) and further assume that those who stayed in the same housing either did not get improved conditions or, at any rate, did not pay for them. The average additional rent paid by the *movers* then represents the total amount that recipients paid for desirable changes in housing conditions (whether at inflated prices or not). The balance of all increased rent covered price increases on housing in which recipients remained over the whole time period covered. Thus, we will observe the allocation of the subsidy to three functions: relief of rent burden, improvement in housing conditions, and price increases for current housing.

The average subsidy in the program was \$80.88 per month and the average posttest out-of-pocket rent paid was \$63.64. Thus, the sum of these, or \$144.52, was the average dollar rent actually paid to the landlord after the program. The amount that *would have been* paid, given as $\hat{Y}_{OE}$ in Table 4, was \$118.42, so that the difference, \$144.52 – \$118.42 = \$26.10, is the amount of the average subsidy that was put into higher rent. This represents 32% of the total subsidy, a portion to which we will return momentarily for further analysis. The balance, \$54.78 or 68%, went to relieve the rent burden; that is, it was freed for other uses (we neglect the interval "$\pm e_{\hat{Y}}$" because it will become part of a larger interval momentarily; otherwise, the quantity \$54.78 would be the same as that indicated by b_T' in equation 2 and Table 4). Knowing that 68% of the subsidy went toward (a) relief of the rent burden, it remains to divide the 32% between (b) better quality housing, and (c) a price increase for the same dwelling unit. This we attempt by examining the record of the movers (N = 356, or 45% of the total AAE sample—see Table 5).

The average dollar rent paid at posttest time by those families that actually moved was \$160.92 (including \$90.30 in average subsidy), and the null-case rent (based on $\hat{Y}_{OE}$) for that group averaged \$102.43 (note that the random-comparison-group design enables the analyst to derive a null-case estimate for any subgroup, or indeed any individual, and not

TABLE 5
Comparison Group, Treatment Group, and Treatment-Group Movers

	Comparison Group	*Treatment Group*	*Treatment-Group Movers*
Sample size	3040	795	356
Average subsidy	–	\$80.88	\$90.30
Average rent:			
After ($\bar{Y}$)	\$119.21	63.64	\$ 70.62
Before ($\bar{X}$)	110.12	110.35	90.83
Predicted ($\hat{Y}_{OE}$)	–	118.42	102.58
Average rent burden			
After	.395	.232	.270
Before	.393	.426	.348
Average quality			
After	.157	.062	.048
Before	.153	.116	.158
Average crowding			
After	.63	.65	.70
Before	.64	.70	.81

just the treatment group as a whole). The difference, \$58.49 per family or 65% of the subsidy, is the amount paid in higher rent, which in this subgroup means the amount paid for improved conditions of housing. Inasmuch as these subjects made up 45% of the recipients and their rent increase represents the total of improved housing conditions for the treatment group as a whole, then (.45)(\$58.49) = \$26.32 is the average amount put into better housing within the treatment group as a whole. This yields a result that may appear surprising. The quantity \$26.32 is 32% of the average subsidy (\$80.88) for the treatment group. It will be recalled that 32% of the subsidy over the whole group was paid in higher rent. Thus, we have found that *all* of the average subsidy going into higher rent was spent for better conditions of housing; none, on the average, went toward increased price for the same housing. Although somewhat surprising, the result is solidly based in null-case projections, for both the treatment group and the "mover" subgroup, within the random-comparison-group framework. As it happens, the finding accords well with the results of the supply experiment described previously, so that both are supported by the corroboration. It appears that there was little price inflation (above whatever may be reflected in the comparison-group data) for the current housing of these individuals during the period of the study.

These estimates, however, assume that the treatment group is average in posttest rent for similar households in the population (where "similar" means having the same mean scores as the treatment group on income, prior rent, and the other causal variables in Table 3). We have found that such projections need to be converted to interval estimates to provide a safe conservatism. In the present analysis, however, conservative cannot be taken to mean a smaller impact of the program on rent relief, for that would simply imply *liberal* inferences of impact on improvement in conditions of housing. What is needed, then, is a *symmetric,* or two-sided, interval around the rent impact, based on the same method as was used in the previous section, that is, equations 5 and 6. This yields an interval estimate of 68% ± 14% for the proportion of the subsidy going toward rent relief and 32% ± 14% for the part going toward higher rent. Furthermore, detailed, comparative random-comparison-group analyses for movers and "stayers" show that when the latter interval is at its high end, or 46%, nearly all (85%) of this rent increase is found to be attributable to improved quality, and when it is at its low end, or 18%, so much is attributable to quality that substantial price *de*flation is implied for those who did not move to better housing (an unlikely result that therefore suggests, as suspected, that our interval of ±14% is unnecessarily large because the assumption of 10% outliers is still unnecessarily conservative).

In brief, we find that 68% ± 14% of the subsidy went to relieve the rent burden and the balance to improvements in quality. The interval is large, but (a) its method, which allows 10% outliers on each end, is apparently too conservative, so that the 68-32 split is undoubtedly more accurate than the breadth of the interval suggests, and (b) it still permits informative conclusions about the impact of the program to be drawn. It appears that the bulk of the subsidy went to liberate funds in these low-income families for nonrent uses. The remainder was spent on improvements in housing quality obtained by families that had to move to better housing in order to qualify for the program. This was nearly half (45%) of the treatment group (whereas 30% moved during this time in the comparison group, and it is unclear whether they moved to better or worse housing). On average, the group of movers started out paying much lower rent than the treatment group as a whole ($90.83 versus $110.35), received a larger subsidy ($90.30 versus $80.88), and ended by paying higher amounts for rent out of pocket ($70.62 versus $63.64).

The movers improved their conditions of housing substantially, represented by at least 18% and probably about 32% of the subsidy for the group as a whole. This amount primarily bought quality that is *not* detailed in our measures. It is reflected in part, but only in part, by a larger-than-average increase in quality as measured here—moving from nearly one item failed in the null case to essentially zero after the subsidy—and a larger-than-average, though still very small, improvement in crowding, going from 0.8 persons per room in the null case to 0.7 after the program. The evaluation could have been even more informative if more detailed observations had been recorded on conditions of housing in both the treatment and comparison groups, but the latter is perhaps too troublesome to be expected.

The precision provided by the random-comparison-group design could have been improved by accepting less confidence in the results—the usual trade-off in classical statistical inference—and by better prediction in the comparison-group model. On the latter point, the present evaluation suffers because there is substantial unpredictable fluctuation in rent and income over a year's time in the poverty class. That is why the R^2 in the national sample is as low as it is (although very high in comparison with theoretical social research). There would surely be evaluations in which the posttest were more predictable from the pretest and other variables, so that even conservative estimates from the random-comparison-group design would yield readily satisfying precision in results.

CONCLUSIONS AND POLICY IMPLICATIONS

In this analysis we have utilized data from the EHAP Administrative Agency Experiment and the Annual Housing Survey to test the effects of housing vouchers on the housing conditions of low-income families. With a full understanding of both the design and data limitations of the AAE, we proceeded to show how demonstration projects of this sort can be used in social program evaluation with a reasonable degree of confidence in the results through employing the random-comparison-group design. We examined two types of program outcome in the analysis: (1) the impact of the program on reducing housing deprivation scores for participants, and (2) the manner in which the subsidy was used by participating households. These are key issues concerning housing vouchers and our findings tend to support as well as augment some of the findings from the supply and demand experiments. (We remind the reader that the supply and demand experiments have the drawback that they represent only two locations each, whereas the AAE represents eight widely scattered and diverse markets, with null-case projections based here on a national sample.)

In examining program impacts on reducing housing deprivation, our results showed that, overall, the program had its greatest impact in reducing rents and rent burdens. In terms of housing quality as measured directly for this analysis, the impact was slight for the group as a whole. For movers, however, it is clear that a substantial shift from substandard to standard housing did occur. One indicator is that this group paid about 65% of its subsidy in higher rent. Another (see Table 5) is that the mean failure score for movers on the quality index was one full item before the program (16% of the six-item index) and one-third of an item afterwards (5%—see the Appendix for the discrepancy between this index and the criteria for substandard housing actually used in the AAE).

Our findings on allocation of the subsidy were strikingly similar to findings from the demand experiment. In the demand experiment approximately 75% of the housing allowance went toward reduction in rent burden. We found approximately 68% of the subsidy going to address this problem in the AAE. The other 25% of the subsidy in the demand experiment went toward increased housing expenditures, which could have and probably did prominently include improvements in housing quality. We found that 32% of the subsidy went toward higher rents in the AAE and all of this amount seemed to have been devoted to improvements in housing quality.

One of the major findings from the supply experiment was that a full-scale open-enrollment allowance program did not have perceptible effects on rents or property values in either a tight housing market (Green Bay) or a loose market (South Bend). Lowry offers the following explanation:

> One reason was that the program increased aggregate housing demand by less than 2 percent. Another was that it proved relatively easy and inexpensive to transform substandard to standard dwelling units. When a renter joined the program without moving, his rent typically increased by less than 2 percent, even though his landlord may have made minor repairs to bring the dwellings up to program standards [Lowry, 1983: 26].

The present findings buttress these conclusions. In examining the 32% of the subsidy not going to reduce the rent burden, we asked how much went toward improved housing quality and how much toward inflated rents. Using outcomes for movers as the basis for our analysis, we conclude that all of the subsidy went for improved quality and none for inflated rent. In fact, for recipients who did not move, we found that price deflation might have occurred. The on-site observer case study data from the AAE tends to support this finding. Generally, landlords as well as recipients in the AAE found housing assistance with "no strings attached" an attractive idea and may very well have made rent adjustments to meet HUD's recommended fair market rents (see Jackson, 1982: 175 for a full listing of these reports).

Clearly, it is problematic to generalize to a national program of subsidies from a study that included only pockets of the population and that was evaluated after only one year. With that caveat, our findings support the idea that by providing housing vouchers to low-income families, substantial reductions in their out-of-pocket income for rent will occur. For those families having to move to satisfy the minimum standard requirement, there is less relief of the rent burden, but substantial improvements in housing quality are made. Keeping in mind the reservations that opened this paragraph, all of these results may apparently be obtained without creating rent inflation in the housing market.

We cannot conclude, however, without taking note of an important caveat based on previous work. Budding (1978) and Jackson (1982, chap. 4) show that a large percentage of preprogram enrollees were dropped from the program because their housing units were substandard and they could not find adequate housing in the housing market. In other words, the minimum housing-quality standard lowered the participation rates of needy families. As Frieden (1980) concluded some years ago, housing allowances do work. It would be an error, however, to assume that they leave no major pockets of need unmet.

APPENDIX

A. DEVELOPMENT OF THE COMPARISON GROUP FOR THE ADMINISTRATIVE AGENCY EXPERIMENT USING ANNUAL HOUSING SURVEY DATA

Data from the 1974 and 1975 national files of the Annual Housing Survey were utilized in constructing a comparison group for the administrative agency experiment. The purpose of this appendix is to describe the selection technique used in constructing the comparison group.

First, the years 1974 and 1975 were chosen as the two years that most closely approximated the time period in which the experiment was implemented. Second, a procedure was developed comparable to the one utilized by the Department of Housing and Urban Development for selecting families who would be eligible for the housing allowance program. The key to the procedure involved first determining a market rent that the Department of Housing and Urban Development would be willing to pay for a housing unit of standard quality in a given housing market. This rent is frequently referred to as a fair market rent and is normally determined by the regional area office in conjunction with the local housing authority.

The procedure for determining this rent is imprecise and subject to variation from one locale to the next. Recently, hedonic price indices have been employed in determining fair market rents (Follain and Malpezzi, 1981). Instead of using an elaborate technique such as a hedonic price index to determine the fair market rents for this analysis, average rents for units with 0 to 9 bedrooms were used. An average rent for each number of bedrooms was taken for each region. Substandard units and units subsidized by the government were eliminated from the sample prior to taking the averages.

In a memo issued to the public agencies administering the administrative agency experiment, the Department of Housing and Urban Development specified the number of bedrooms for which a family would be subsidized based on the number of persons in the family. The next step was to assign the appropriate fair market rent (the average rent) to each household in the AHS based on the household size, the number of bedrooms HUD was willing to subsidize in that instance, and the region in which the household was located.

Finally, to determine whether or not a family would be eligible for a housing allowance, the assigned fair market rent was divided by the family's total income. Those families with rent-income ratios above .25 were considered eligible for the program. The procedure was employed on renters in 1974 and 1975. Only families qualifying for the program in both years were used in the comparison group. In this fashion, the selection procedure employed here approximated closely the one used by the housing allowance agencies in selecting AAE participants.

B. MEASURES OF HOUSING DEPRIVATION USED IN THE TREATMENT AND COMPARISON GROUPS

The national housing goal of the United States adopted by Congress in the Housing Act of 1949 called for a "decent home and suitable living environment for every American family." The definition of a decent home and suitable living environment has varied over time and it has been left to policymakers to decipher congressional intent.

Three major aspects of housing have evolved as the principal components of congressional concern: deficient or substandard housing, crowding, and affordability. Measures used in this analysis for these three forms of housing deprivation come from a general set of guidelines issued to all experimental housing allowance agencies by the Division of Housing Assistance Research within the Office of Policy Development and Research, HUD. The three categories of deprivation were operationalized as follows.

CROWDING

Families having more than 1.5 persons per room were considered as living in overcrowded conditions. For this analysis, the total number of persons in a household was divided by the total number of rooms in the housing unit. Instead of using the dichotomy, "crowded-not crowded" based on the demarcator of 1.5, the exact ratio of persons to rooms was employed as a continuous variable. This allowed us to examine the extent to which persons per room changed as an impact of the program. Sample data from the administrative agencies' enrollment and payment operating forms were used to operationalize the pre- and postprogram measures of crowding for the treatment group. Data from the 1974 and 1975 AHS tapes were used to measure persons per room for the comparison group.

RENT BURDEN

The proportion of a family's income devoted to rent had been the principal measure of housing affordability in housing policy analysis. Most studies, as well as the guidelines issued to EHAP agencies, use .25 as an appropriate fraction of income to be paid in rent. Families paying more than 25% of their income for rent are considered as having a deprivation-level rent burden.

For this analysis, rent burden was operationalized by dividing a family's gross rent (rent plus utilities) by the total family income. Data from AAE enrollment and payment operating forms on rent and income were used in developing before-and-after measures for the treatment group. Instead of using a dichotomy based on .25, the full rent/income ratio was employed to indicate rent burden before and after the program. For the comparison group, similar data were lifted from the AHS tapes for 1974 and 1975.

HOUSING QUALITY

The guidelines issued by the Department of Housing and Urban Development to EHAP program agencies provided for a minimum housing quality standard. Although this standard was not rigorously enforced in the administrative agency experiment it did serve as a guideline for agencies in assessing the quality of units in their respective jurisdictions. The minimum standard consisted of 14 components falling into three general categories: *basic housing services* (core room presence, complete plumbing, complete kitchen facilities, light fixtures, electrical services); *safety features* (adequate exits, presence and safety of heating equipment); *structure and surface condition* (room structure, room surface, floor structure, floor surface, roof structure, exterior walls). For a complete description of these minimum standard requirements see the Minimum Standard Requirement report produced by ABT Associates (Baker et al., 1980). A fourth category, the *presence of rats,* was added for use in this analysis.

Given that the Annual Housing Survey data do not contain enough detail to allow full operationalization of the minimum standard requirement, a modified housing-quality index was constructed based on comparable items from the AAE housing-quality surveys and the AHS housing-quality section. The housing-quality index used in this analysis for both the treatment and control groups was based on the following components taken primarily from the minimum standard requirements:

1. complete kitchen facilities,
2. complete plumbing facilities,
3. adequate heat,
4. roof condition,
5. wall condition, and
6. presence of rats.

Instead of classifying units as substandard for a failure on any one of these items, a ratio measure was developed measuring the percentage of items failed by each unit in the treatment and comparison groups at the two relevant time periods.

REFERENCES

AMEMIYA, T. (1973) "Regression analysis when the dependent variable is truncated normal." Econometrica 42: 999-1012.

BAER, W. C. (1976) "The evolution of housing indicators and standards: some lessons for the future." Public Policy (Summer): 361-393.

BAKEMAN, H., C. A. DALTO, and C. WHITE (1980) Minimum Standards Requirements in the Housing Allowance Demand Experiment. Cambridge, MA: ABT.

BLUMSTEIN, A. and J. COHEN (1979) "Control of selection effects in the evaluation of social problems." Evaluation Q. 3,3: 583-608.

BRADBURY, K. and A. DOWNS [eds.] (1981) Do Housing Allowances Work? Washington, DC: The Brookings Institution.

BUDDING, D. W. (1978) Report on Housing Deprivation Among Enrollees in the Housing Allowance Demand Experiment. Cambridge, MA: Abt Associates.

CAMPBELL, D. T. and J. C. STANLEY (1966) Experimental and Quasi-Experimental Designs for Research. Chicago: Rand McNally.

Congressional Budget Office (1978) Federal Housing Policy: Current Programs and Recurrent Issues. Washington, DC: The Brookings Institution.

DELEEUW, F. (1972) The Design of a Housing Allowance. Working Paper, Urban Institute. Washington, DC: Superintendent of Government Documents.

FRIEDEN, B. (1980) "Housing allowances: an experiment that worked." The Public Interest 59 (Spring): 15-35.

FREIDMAN, J. and D. WEINBERG (1983) The Great Housing Experiment. Beverly Hills, CA: Sage.

HARTMAN, C. W. (1983) "Housing allowances: a critical look." J. of Urban Affairs 5, 1: 41-55.

HECKMAN, J. (1976) "The common structure of statistical models of truncation, sample estimation for such models." Annals of Economic and Social Measurement 5: 475-492.

JACKSON, B. O. (1982) "The linkage between implementation processes and policy outcomes: an analysis of HUD's administrative agency experiment." Ph.D. dissertation, University of Michigan.

———(1983) "The impact of federal housing policy on low and moderate income black Americans: an analysis of a dream deferred." Presented at the Annual Meeting of the National Conference of Black Political Scientists, Houston, Texas, April 27-30.

KAIN, J. F. (1981) "A universal housing allowance program," pp. 339-373 in K. Bradbury and A. Downs (eds.) Do Housing Allowances Work? Washington, DC: The Brookings Institution.

KERSHAW, D. N. and R. C. WILLIAMS (1981) "Administrative lessons of the experimental housing allowance program," pp. 285-338 in K. Bradbury and A. Downs (eds.) Do Housing Allowances Work? Washington, DC: The Brookings Institution.

LOWRY, I. S. (1983) Experimenting with Housing Allowances. Cambridge, MA: Oelgeschlager, Gunn and Hain Press.

McKENNA, W. (1982) The Report of the President's Commission on Housing. Washington, DC: Government Printing Office.

MOHR, L. B. (1982) "On rescuing the nonequivalent-control-group design: the random-comparison-group approach." Sociological Methods and Research 11, 1: 53-80.

OZANNE, L., S. MALPEZZI, and T. THIBODEAU (1980) Characteristic Prices of Housing in Fifty-Nine Metropolitan Areas. Washington, DC: The Urban Institute.

STRUYK, R. and M. BENDICK (1981) Housing Vouchers for the Poor: Lessons from a National Experiment. Washington, DC: Urban Institute Press.

TROCHIM, W.M.K. (1984) Research Designs for Program Evaluation: The Regression-Discontinuity Approach. Beverly Hills, CA: Sage.

U. S. Department of Housing and Urban Development (1980) Experimental Housing Allowance Program Conclusions: The 1980 Report. Washington, DC: Government Printing Office.

———(1981) Participation and Benefit Rates in the Section 8 Program. Washington, DC: Government Printing Office.

———(1985) News release, May 16. Washington, DC.

WEICHER, J. C. (1979) National Housing Needs and Quality Changes During the 1980's: A Forecast. Working Paper 1345-2. Washington, DC: The Urban Institute.

———(1980) Housing: Federal Policies and Programs. Washington, DC: The American Enterprise Institute for Public Policy Research.

Byran O. Jackson is Assistant Professor of Political Science and an Adjunct Assistant Professor in the Urban Affairs Program at Washington University, St. Louis. He has also served as a Social Science Research Analyst for the Office of Policy Development and Research in the Department of Housing and Urban Development in Washington, D.C.

Lawrence B. Mohr is Professor of Political Science and Public Policy at the University of Michigan. His primary research areas are organization theory, program evaluation, and the philosophy of social research.