



**BOLD
THINKERS
DRIVING
REAL-WORLD
IMPACT**

Impact of Housing and Services Interventions for Homeless Families (Family Options Study)

Supporting Statement for Paperwork Reduction Act Submission, Part B

**Contract No.
GS-10F-0086K, Order No.
DU206SF-13-T-00005**

36-Month Follow-up Data Collection

Final

November 20, 2013

Prepared for:
Anne Fletcher
U.S. Department of HUD
Office of Policy Development and
Research
451 Seventh Street SW Room 8140
Washington, DC 20410

Prepared by:
Abt Associates Inc.
4550 Montgomery Avenue
Suite 800 North
Bethesda, MD 20814-3343

Family Options Study 36-Month Follow-Up Data Collection – Draft

Table of Contents

Part B: Collection of Information Employing Statistical Methods.....1

- B.1 Identification of Appropriate Respondents.....1
 - B.1.1 Sample Recruitment and Random Assignment.....1
 - B.1.2 Universe of Households and Survey Samples.....3
- B.2 Administration of the Survey.....5
 - B.2.1 Sample Design.....5
 - B.2.2 Estimation Procedures.....6
 - B.2.3 Degree of Accuracy Required.....9
 - B.2.4 Procedures with Special Populations.....14
- B.3 Maximizing the Response Rate and Minimizing Non-Response Risk.....14
- B.4 Test of Procedures.....15
- B.5 Individuals Consulted on Statistical Aspects of the Design.....16

References.....17

Part B: Collection of Information Employing Statistical Methods

B.1 Identification of Appropriate Respondents

B.1.1 Sample Recruitment and Random Assignment

The Family Options Study is conducted as a randomized experiment. From September 2010 through January 2012, the research team enrolled 2,282 homeless families into the study in 12 sites.¹ Each family was randomly assigned to one of four interventions:

1. Permanent Housing Subsidy (SUB)
2. Project-Based Transitional Housing (PBTH)
3. Community-Based Rapid Re-housing (CBRR)
4. Usual Care (UC)

In an effort to maximize the likelihood that families randomly assigned to the study interventions could actually receive the assigned intervention, the research team established conditions that had to be met for random assignment to proceed:

1. Families had to meet the general eligibility requirements—reside in shelter at least 7 days prior to enrollment and have at least one child age 15 or under with them—for participation in the study and had to consent to enroll in the study.
2. Intervention slots² had to be available at the time of random assignment or anticipated within 30 days.
3. Families had to meet more particular intervention-specific program eligibility criteria for at least two interventions for which slots are available.³ These more particular requirements include such

¹ The final enrollment count for the study was 2,307 families. However, upon reviewing baseline data collected, the research team determined that 25 families had been enrolled in error and did not satisfy the family eligibility requirement of having at least one child age 15 or younger. These 25 families, scattered at random among the four random assignment arms in the study design, have been removed from the research sample without skewing the statistical equivalence of the arms, leaving 2,282 families.

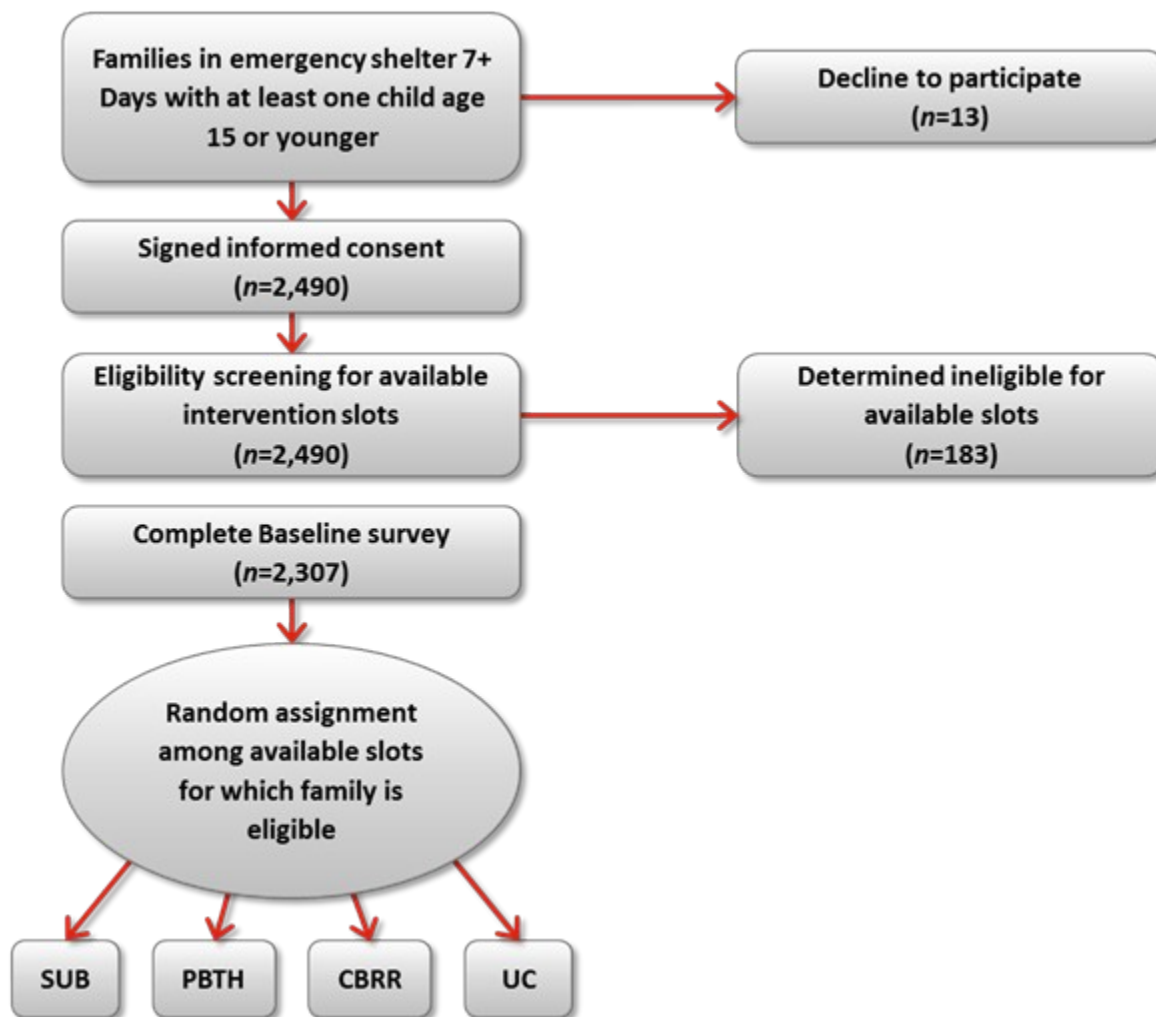
² The term *slot* refers to opportunities for placement in a study intervention. For SUB intervention, a slot refers to a housing choice voucher or a unit in a public housing or project-based assisted development. For the PBTH intervention, a slot refers to a family’s housing unit or space at a transitional housing facility. For the CBRR intervention, a slot refers to rental assistance provided to the family to subsidize the rent of a housing unit in the community. When we refer to an intervention slot as available, we mean that there was an open space for a family to be placed in that study intervention.

³ Initially, random assignment was contingent upon family eligibility for available slots in at least three of the four interventions, in order to maximize the number of experimental comparisons that each family could be included in. In August 2011, this condition was relaxed to two of the four interventions in order to take full advantage of service slots that had been reserved for families in the study. Sometimes it took longer to utilize all of the slots for one intervention type than others. In those instances, as one intervention type—CBRR for example—was fully utilized, but another intervention type had slots remaining—SUB for example—we relaxed

things as having no criminal background, being clean and sober, having some form of income, not owing back rent to a housing authority, etc.

Enrollment in the study and conducting random assignment was a multi-step process, as shown in Exhibit B-1. The right side of the exhibit summarizes the conditions under which some families would be ineligible or decline to participate.

Exhibit B-1. Random Assignment Process



The study was not designed to capture the experiences of families who seek assistance directly from transitional housing programs without first entering emergency shelters. The design relied on emergency shelters as the point of intake for families in the study.

The design excluded families who left shelter in less than 7 days because the more intensive interventions considered in this study are not considered appropriate for families with such transitory needs. We expected shelters to continue to provide all services and referrals they ordinarily provided to help families

the requirement in order to take advantage of the slots reserved for the study.

leave shelter up until the point of random assignment. As also shown at the top of the exhibit, the population of interest for this study is all families who had been in an emergency shelter for at least 7 days and who had at least one child 15 or younger at baseline. This restriction was included because child outcomes are important to the study, and we will not have a large enough sample to consider outcomes for youth who become young adults in the course of the follow-up period. Families were then assigned, as close to the 7-day mark as was feasible, to the Subsidy Only (SUB), Community-Based Rapid Re-housing (CBRR), Project-Based Transitional Housing (PBTH), or Usual Care (UC) interventions.

The study design also recognized that not all families were eligible for all interventions. Consistent with this consideration, families were screened as to their eligibility for each specific service provider in their site, prior to random assignment. Families were randomly assigned among only interventions for which they appeared eligible, based on their responses to a set of intake screening questions. As long as one provider within each experimental intervention at a given site would accept a family with a particular profile, that family was considered eligible for that intervention.

As we describe below, this design assures that comparisons of interventions will involve well-matched groups in each intervention. To achieve this, a family will be included in the pairwise impact comparisons only for pairs of interventions to which it *could have been assigned* (i.e., interventions that were available at the time of that family’s randomization and for which the family was eligible). The design thus assures that any observed differences in outcomes are caused by the differential treatment families receive following random assignment, and not by any pre-existing differences among the families.

Although assignment to interventions was conducted at random, within interventions families were not assigned at random to specific service providers that provided the intervention. Allocation among providers was made instead on the basis of family characteristics, as is customary practice in the housing assistance system. Thus, for example, if one or more of the transitional housing programs in a site specialized in families with a particular profile (only families with domestic violence issues, or only families where the mother has been clean and sober for some period), then among families randomly assigned to PBTH, only those that fit that program were assigned to that service provider. This preserves and studies programs as they currently operate.

B.1.2 Universe of Households and Survey Samples

This supporting statement seeks clearance for a data collection effort to collect four types of data:

1. An adult survey for heads of households;
2. Ages and Stages Questionnaire for heads of households reporting on up to two children ages 20 to 66 months;
3. Child assessments administered directly to up to two focal children ages 3 years 6 months to 7 years 11 months; and
4. A child survey for up to two children in the household (ages 8 years to 17 years 11 months).

Adult Survey

The universe for the adult survey for heads of households is the entire sample of families who were enrolled in the study (2,282 families). To be eligible for the study, the families—defined as at least one

adult and one child—had to experience homelessness, receive assistance at an emergency shelter, and remain in the shelter for at least seven days. Exhibit B-2 summarizes the definition and sample sizes for all of the random assignment groups.

Exhibit B-2. Definition and Size of Randomly Assigned Groups in the Family Options Study

Group	Intervention Definition	# Assigned per Site	Total # Assigned
SUB	Subsidy only; defined as deep, permanent housing subsidy that may include housing related services but no supportive services.	0-77	599
CBRR	Community-Based Rapid Rehousing: Time-limited housing subsidy that may also include housing-related services and limited supportive services	8-83	569
PBTH	Project-Based Transitional Housing: Time-limited housing subsidy coupled with supportive services	0-66	368
UC	Usual Care: Other assistance available in the community	21-81	746
	Total, all Intervention Groups	58-281	2,282

Child Survey and Assessments

During the study’s 18-month follow-up data collection effort, the study randomly selected up to two focal children from each respondent family for child data collection. The 36-month follow-up data collection will attempt to collect data from the already-selected focal children (in 18-month respondent families) and from newly selected children in those families who did not respond at 18 months.

The focal children sampling plan used in the 18-month follow-up gave every child a known probability of being sampled (allowing us to project to the universe of all children of the study families), and oversampled those children who could be assessed directly. The focal child selection process to be used in the 36-month follow-up (for families who still need focal children to be selected) will be identical except that the age criteria will shift upward to reflect the time elapsed between data collection waves.

The focal child selection process considers a) whether the child was with the parent at baseline; b) whether the child is with the parent at the follow up or the child’s recent activities are sufficiently known by the parent; and c) the age of the child. Specifically, in the 18-month follow-up, the study randomly selected up to two focal children from children age 12 months to 17 years 11 months (as of 18-month follow-up) who were living with family or apart from the family at follow-up.⁴ Children who were age 3 years 6 months to 17 years 11 months, living with the family at study entry in the shelter, and living with the family at follow-up were oversampled for inclusion into the focal child sample. As noted above, in the 36-month follow-up, for families who still need focal children to be selected, the study will follow the same selection criteria used at 18-months to select focal children for the 36 month follow-up child data collection, with age ranges increased by 18 months.

⁴ For children not living with the family at follow-up, the parent needed to have knowledge of a child's activities in the last 30 days in order for that child to eligible to be selected as a focal child.

The type of data collected from and about a focal child depends on the age of the focal child at the time of data collection. Focal children who are ages 20 months to 5 years 6 months will be subjects of the Ages and Stages Questionnaire administered to their parents as part of the adult interview. Focal children who are ages 3 years 6 months to 7 years 11 months will be administered a small number of direct, in-person assessments. Focal children who are ages 8 -17 years will be administered a child survey questionnaire.⁵

We use information from the 18-month data collection to estimate the child sample sizes by age. For purposes of estimating burden for the 36-month data collection effort, we estimate a focal child sample of 3,010. Allowing for the sample to age as time elapses between the 18 and 36 month follow-up data collection, the projected focal child sample is distributed across age categories as shown in Exhibit B-3:

⁵ The Ages and Stages Questionnaire, child assessments, and child survey were used to collect data about and from focal children during the 18-month data collection. In that collection effort, a grant from the NICHD funded the child data collection.

Exhibit B-3. Focal Child Sample Size^{6,7}

	Ages and Stages (Children 20 months to 5 years 6 months)	Child Assessments (Children 3 years 6 months to 7 years 11 months)	Child Survey (Children 8-17 years)
Sample Size	300	1,330	1,380

B.2 Administration of the Survey**B.2.1 Sample Design**

The study sample for the 36-month data collection is 2,282 families. For the 36-month follow-up survey, the research team will attempt interviews with all members of the research sample, except for those who requested no further contact or who were found to be deceased at the time of the 18-month interview.⁸ Therefore, no sampling is required for the 36-month follow-up surveys with adult respondents.

There are an estimated 3,010 children in the focal child sample for the 36-month follow-up. As noted above, the research team may need to select new focal children if we complete an adult survey with someone who did not complete the 18-month interview. We may also need to select a focal child in cases where we completed an 18-month interview with the adult but did not select a focal child because no children met the focal child sampling criteria at the time of the 18-month interview (e.g., the parent could not report on the child’s activities or no child under 18 was available).

B.2.2 Estimation Procedures

The study will generate separate impact estimates for different comparisons among interventions. This will include six pairwise comparisons of a single assignment arm to another single assignment arm, plus four additional comparisons of a pooled sample from multiple assignment arms to a single assignment

⁶ There will be an estimated 3,010 children in our focal child sample. Of these, 300 are eligible for Ages and Stages only; 1,130 are eligible for child assessments only. There are an estimated additional 200 focal children who were assessed with Ages and Stages at 18 months, and will also be old enough for the child assessments at 36 months. These 200 children will be part of the child assessment sample but their parent will also be asked to complete the ages and stages questionnaire. Thus there are 200 focal children who are subject to two of the three child data collection protocols. To avoid duplicate counting in the overall focal child sample size, these 200 focal children are reflected in the child assessment sample.

⁷ The focal child sample sizes for the child assessments and child surveys were estimated by assuming that the families who did not respond to the 18-month data collection would have had the same number of focal children identified in each age category as the families who did respond to the data collection. It was additionally assumed for the two older age groups that the number of children aging into the ranges would be equivalent to the number aging out of the ranges. The sample size for the child assessments is calculated by dividing the number of identified focal children for these assessments at 18 months (1,079 children) by the percent of families who had focal children identified (81 percent). Similarly, the sample size for the child survey is calculated by dividing the number of identified focal children for the child survey at 18 months (1,122 children) by the percent of families who had focal children identified (81 percent).

⁸ During the 18-month interview we found that 6 adult respondents were deceased and 2 refused to be interviewed and asked that the research team not contact them again. These families will not be included in the 36-month survey.

arm. The particular configurations to be contrasted in this way—and the impact questions each will address—are shown in Exhibit B-4. The comparisons will be analyzed separately using the same estimation model.

Exhibit B-4. Policy Questions Addressed By Six Pairwise Comparisons and Four Pooled Comparisons

Pairwise/Pooled Comparison	Policy Question Addressed
1. PBTH versus UC	Impact of a deep, temporary housing subsidy with heavy services
2. SUB versus UC	Impact of a deep, permanent housing subsidy
3. CBRR versus UC	Impact of a deep, temporary housing subsidy with light services
4. PBTH versus SUB	Impact of a deep, temporary housing subsidy with heavy services compared to a deep, permanent subsidy without services
5. PBTH versus CBRR	Impact of a deep, temporary housing subsidy with heavy services compared to a deep, temporary housing subsidy with light services
6. SUB versus CBRR	Impact of a deep, permanent housing subsidy with no services compared to a deep, temporary housing subsidy with light services
7. PBTH + SUB + CBRR versus UC	Impact of any kind of housing subsidy for homeless families compared to the usual services offered in the community
8. PBTH + SUB versus CBRR	Impact of interventions that are more costly compared to a less costly intervention
9. PBTH versus CBRR + SUB	Impact of a housing subsidy with heavy services compared to a housing subsidy with light or no services
10. SUB versus PBTH + CBRR	Impact of a housing subsidy with no time limit compared to a time-limited housing subsidy

Implementation of Random Assignment

To conduct these impact analyses, we will first determine the “randomization set” and the “assessed ineligible set” for each family randomized:

1. Randomization set: the set of interventions to which the family might have been assigned based on availability of interventions and the assessed eligibility of the family for the available interventions. In total, seven different randomization sets occurred: {PBTH, SUB, CBRR, UC}, {PBTH, SUB, UC}, {PBTH, CBRR, UC}, {SUB, CBRR, UC}, {PBTH, UC}, {SUB, UC}, and {CBRR, UC}.
2. Assessed ineligible set: the set of interventions for which a family was assessed as ineligible for assignment. In order to remain in the study, the maximum number of interventions for which a family could have been assessed ineligible was two.⁹ Therefore, each family has one of seven possible “assessed ineligible sets.” These sets are {none}, {PBTH only}, {SUB only}, {CBRR only}, {PBTH and SUB}, {PBTH and CBRR}, and {SUB and CBRR}.

Each of these sets contains important information for the impact analysis. The randomization set for each family determines the impact comparisons in which the family will be included. A family will only be included in the pairwise comparisons of its assigned intervention with other interventions in its randomization set, in keeping with the experimental design of the study. The assessed ineligible set, in turn, captures some characteristics of the family that may be correlated with outcomes. We therefore plan to include indicators for the family’s assessed ineligible set as covariates in the impact regressions

⁹ Families who, after responding to provider-specific eligibility questions, were assessed as eligible for only one of the interventions available to them were not enrolled into the study. There were 183 families who were not enrolled in the study because of this reason.

(described below).¹⁰ Families within the same assessed ineligible set may have responded to different eligibility questions based on the interventions and providers available to them. The random assignment design allows us to expect that the sets are symmetric across intervention groups, however.

Method of Estimation

For each pairwise comparison, we will estimate impacts using a sample of families who have both interventions in their randomization set and were assigned to one of the two interventions, so that the actual interventions experienced by the two groups represent the policy contrast whose impact we want to examine. Following standard practice, we will use regression adjustment to increase the precision of our impact estimates (Orr, 1999). Consider two interventions q and r (e.g., PBTH vs. SUB), where we treat the second option (r) as the base case. Then, we would estimate the impact on an outcome Y (e.g., at least one night homeless or doubled up during past 6 months, working for pay in week before survey, adult psychological distress) of intervention q relative to intervention r by estimating equation (1) for those families who had both options q and r as possible assignments, and were assigned to one of them. The estimation equation is

$$(1) \quad Y_i = \alpha_{q,r} + T_{q,i} \delta_{q,r} + X_i \beta_{q,r} + \sum_{j=1}^6 I_{j,i} \theta_{q,r,j} + \sum_{k=1}^{13} I_{k,i} \varphi_{q,r,k} + e_i$$

where

Y_i = outcome Y for family i ,

$T_{q,i}$ = indicator variable that equals one if family i was assigned to intervention q ,

$\delta_{q,r}$ = impact of being assigned to intervention q relative to being assigned to intervention r ,

X_i = a vector of background characteristics of family i ,

$I_{j,i}$ = indicator variable for “assessed ineligible set” j for family i (omitted set is {none}),

$I_{k,i}$ = indicator variable for “site-RA regime”¹¹ k for family i ,

e_i = residual for family i (assumed mean-zero and i.i.d.),

$\alpha_{q,r}$ = a constant term, and

$\beta_{q,r}$, $\theta_{q,r,j}$, $\varphi_{q,r,k}$ = other regression coefficients.

In this model, we make the assumption that the true impact of intervention q relative to intervention r is homogeneous across sites. The impact parameter $\delta_{q,r}$ will be a weighted average of the point estimates of site-level impacts, with each site-level impact weighted by the number of families in the site.

¹⁰ While the assessment of *eligibility* for interventions reflects on characteristics of the families themselves, the assessment of *availability* of interventions just prior to random assignment does not tell us about family characteristics. Thus, we propose to include indicators for assessed ineligible set rather than randomization set (which combines availability and eligibility information) as covariates in the analysis.

¹¹ Ten of the 12 sites had a single random assignment regime during the 15-month study enrollment period. The remaining two sites changed random assignment probabilities a single time each. Therefore, there are a total of 14 site \times RA regime groups. Thirteen indicator variables are included in the equation and one is omitted. These indicator variables are included so that the impact estimate will be based on within-site comparisons.

Standard Errors

We plan to estimate the model above using ordinary least squares (OLS), which assumes that the outcome data have a normal distribution (i.e., form a bell-shaped curve) with a common variance (i.e., are homoscedastic). We have no reason a priori to expect homoscedasticity, however, since some types of families could have higher variability in their outcomes than other families and the different interventions could themselves influence this variability. Furthermore, many of our outcomes are binary; applying OLS to such binary outcomes (i.e., using the “linear probability model”) induces heteroscedasticity.¹²

To address the potential of heteroscedasticity, we will compute robust standard errors (i.e., Huber-Eicker-White robust standard errors; Huber, 1967; Greene, 2003; White 1980, 1984). To address concerns about the linear probability model, we will also estimate some of the specifications using logistic regression models specifically designed for binary outcomes as sensitivity checks. Previous experience suggests that inferences will be quite similar. If we find divergent results, we will present the impact estimates from logistic models for binary outcomes in the main impact tables.

Missing Covariate Data

A small amount of baseline covariate data is missing because some heads of households did not provide responses to certain items on the baseline survey. As the baseline survey was administered prior to random assignment, missing baseline data cannot be correlated with assignment status. Given the small amount of missing covariate data, a number of approaches are available to us to handle the missing data. We plan to use “single stochastic imputation” to impute the missing data based on the values of non-missing covariates. This procedure adds random perturbations (randomly drawn from estimated distributions of residual variance) to the predicted values of missing covariates.¹³ Single stochastic imputation has the virtue of superior statistical power (through preservation of degrees of freedom) over the alternative method of imputation of artificial values and addition of dummy variables to indicate the presence of missing data. Single stochastic imputation also has the virtue of simplicity compared to the alternative method of multiple imputation (which involves the creation of multiple sets of data for analysis).¹⁴

Strategy for Addressing the Multiple Comparisons Problem

The impact analysis will involve a large number of hypothesis tests due to the inclusion of six pairwise impact comparisons, four pooled comparisons, and many outcome measures. Testing such a large number of hypotheses heightens the danger of “false positives” arising in the analysis, i.e., of obtaining statistically significant impact findings where true impact is zero. This danger is called the “multiple

¹² Joshua D. Angrist and Jörn-Steffen Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press. 2008. p. 47.

¹³ Single stochastic imputation may be used for binary variables as well as continuous variables. For binary variables, a random draw is made from the binary distribution using the probabilities derived from the prediction model.

¹⁴ Strengths and weaknesses of various methods of handling missing data are described in Paul D. Allison, (2002). *Missing Data*. Thousand Oaks, CA: Sage University Paper No. 136 and in Michael J. Puma, Robert B. Olsen, Stephen H. Bell, and Cristofer Price (2009). *What to Do When Data Are Missing in Group Randomized Controlled Trials* (NCEE 2009-0049). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

comparisons problem”; the risk of false positives rises above the desired 5 or 10 percent chance as the number of hypothesis tests performed rises above one. To address the multiple comparisons problem we will separate the hypothesis tests into “confirmatory” tests and “exploratory” subsets. Only the most important outcomes will be included in the confirmatory group. These hypothesis tests will be conducted for:

- The 6 pairwise policy comparisons and 1 pooled comparison (PBTH + SUB + CBRR versus UC), and
- A single composite outcome constructed from 2 binary outcomes within the housing stability domain
 - “At least one night spent homeless or doubled up during past 6 months” at the time of the 36-month follow-up survey
 - “Any return to emergency shelter in the 36 months following random assignment”¹⁵ as measured from HMIS administrative data

All other impact estimates will be considered exploratory. For the confirmatory set, we will perform a formal multiple comparisons adjustment that controls the “family-wise error rate” (the probability that at least one of these tests will be found statistically significant by chance if the null hypotheses for all seven tests are true) at the 0.10 level. We will characterize findings of statistical significance for confirmatory outcomes as the *proven* impacts of the policies being compared, and findings of statistical significance for exploratory outcomes as merely suggestive of the impacts that *may have* occurred.

B.2.3 Degree of Accuracy Required

The research team has estimated the minimum detectable effects for this evaluation that will be available through the impact analysis. The analysis of statistical power is presented here.

Power Calculations for Binary Outcomes

In this section, we consider statistical power to estimate impacts of interest. Specifically, we report minimum detectable effects (MDEs). MDEs are the smallest true effects of an intervention that researchers can be confident of detecting as statistically significant when analyzing samples of a given size. The power analyses are computed based on actual numbers of families assigned to the interventions and available for each pairwise comparison. These sample sizes differ somewhat from the planned design due to constraints on families’ eligibility and availability of slots by site.

Our analysis indicates that the proposed design will have sufficient statistical power to detect impacts of the magnitude we might expect to occur for two of the central outcomes of the study—housing stability and child separation from the family for some pairwise comparisons. As discussed below, we will be able to detect effects on these outcomes as small as 8.3 percentage points for the CBRR vs. UC and SUB vs. UC pairwise comparisons and as small as 10.6 percentage points for the PBTH vs. UC comparison..

Exhibit B-5 shows the MDEs by pairwise comparison for the pooled study sample of 1,729 which is 75 percent response of the full sample of 2,282 families. The MDEs presented are the minimum detectable

¹⁵ A stay in emergency shelter after random assignment will be considered a “return to emergency shelter” if HMIS records show that (1) previous program exit had destination of permanent housing or (2) there have been 30 days of non-enrollment since previous program exit.

differences in outcomes (in percentage points) between two randomly assigned groups with 80 percent power when we perform a two-sided¹⁶ statistical test at 10 percent level of significance, assuming a regression R^2 of 0.04¹⁷ and no finite population correction.¹⁸ The differences are shown for various average outcome levels for second assignment group in each comparison.

The last column of the first row of Exhibit B-3 shows that for a mean group outcome of 0.5, the MDE for the CBRR vs. UC comparison is 8.3 percentage points. This means that if the true effect of CBRR compared to UC is to change the prevalence rate of an outcome measure—such as return to shelter housing, or percent of families whose head is a leaseholder at 36-month follow-up—from 50 percent to under 41.7 percent (for return to shelter) or above 58.3 percent (for lease holding), we would have an 80 percent likelihood of obtaining an impact estimate that is statistically significant. If the true effect is less than 8.3 percentage points, there is a lower likelihood that differences between these assignment groups will be statistically significant, though many might still be detected.

The hypothesis is that the interventions to be tested in relation to the Usual Care intervention—all involving housing assistance or subsidy of some sort—will have fairly large effects on housing stability. Drawing on the longitudinal HMIS analysis of shelter utilization (AHAR, 2008; Culhane et al., 2007), we estimate that of families who remain in shelter for at least seven days without any special assistance, approximately 50-60 percent find housing that keeps them from returning within a multi-year follow-up period. There is substantial potential for the proposed interventions to expand this percentage, by using subsidies to eliminate the risk of shelter return for many families in the other 40-50 percent of the population. Housing subsidies remain available to families many months after first receipt, during which time they should provide a sufficiently stable and improved housing option compared to shelters that, for most families, precludes the need for returns to shelter. Research in St. Louis, Philadelphia, and New York City (Stretch & Krueger, 1993; Culhane 1992; Shinn et al., 1998) tends to support this projection. For example, in St. Louis just 6 percent of families who left shelter with a housing voucher returned, compared to 33 percent of those without subsidized housing.¹⁹ Housing stability differed by more than 60 percent between those who received a subsidy (80 percent in stable housing at five years) and those who

¹⁶ While one-sided tests would decrease MDE's, we believe one-sided tests are inappropriate because we care about negative impacts; i.e., they are in a substantive sense not equivalent to a finding of no impact. To see this consider comparing Transitional Housing to Subsidy Only. There a negative point estimate implies that one of the interventions is worse than the other. We care about that, above and beyond the idea that the other intervention is not better.

¹⁷ Since we will estimate regression-adjusted impact estimates, we assume an amount of explanatory power for the regressions. An R^2 of 0.04 is assumed. The regression R^2 was chosen to be identical to the calculated R^2 of the impact regression on the outcome “Did not have a place of one’s own to stay or living with others at some point during the past year” in the Effects of Housing Vouchers on Welfare Families Study, Mills et al., 2006, Exhibit 5.3. Outcomes with higher regression R^2 's will have smaller MDE's.

¹⁸ Applying the finite population correction (FPC) would reduce the MDE's. However, we believe not applying the FPC more accurately represents our uncertainty as to results holding true in future similar applications of the intervention approaches.

¹⁹ Note that this observational pattern is not a direct measure of the impact of subsidized housing on shelter return. Likely the families who exited shelter with subsidies differed from the without-subsidy group on other factors that led to their better outcomes. But even if the difference in unadjusted shelter return rates exaggerates the true impact of a subsidy by an extreme amount—say, 2 or 3 times—the observed 27 percentage point difference would mean an impact of 9 to 13 percentage points.

did not (18 percent stable at five years) in the New York study. Thus, we conclude that an MDE of 8.3 to 10.6 percentage points assures confident detection of the type of impact on housing stability we would expect from the tested interventions (CBRR, SUB, and PBTH) when compared to the Usual Care group.

Exhibit B-5. Minimum Detectable Effects for Prevalence Estimates by Pairwise Comparison

Sample	Expected Number of Completed Follow-up Survey Interviews		MDE if Mean Outcome for the Second Assignment Group is:		
	First Assignment Group	Second Assignment Group	0.1 (or 0.9)	0.3 (or 0.7)	0.5
CBRR vs. UC	430	432	5.0 pp	7.6 pp	8.3 pp
SUB vs. UC	450	408	5.0 pp	7.6 pp	8.3 pp
PBTH vs. UC	270	256	6.4 pp	9.7 pp	10.6 pp
CBRR vs. SUB	288	327	5.9 pp	9.0 pp	9.8 pp
CBRR vs. PBTH	176	174	7.8 pp	11.9 pp	13.0 pp
SUB vs. PBTH	179	192	7.6 pp	11.6 pp	12.6 pp

Notes: (1) The MDE's are based on calculations which assume that two-sided tests are used at the 10 percent significance level, the desired power is 80 percent, and the regression R^2 is 0.04. (2) All MDE's assume a 75% survey response rate, with no finite population correction.

A similar conclusion holds for the prevalence of child separation from the family during the follow-up period. This is likely to be a less common occurrence, making the column of Exhibit B-5 labeled “MDE if Mean Control Group Outcome is: 0.3” likely the most relevant one.²⁰ Here, a slightly smaller true impact can be detected with 80 percent assurance.

Power Calculations for Earnings

Exhibit B-6 shows the MDEs for earnings impacts by pairwise comparison. These MDE's are based on the adult earnings outcomes from the Moving To Opportunity (MTO) Demonstration (Orr, et al., 2003), a study of families who were living in distressed (i.e., barely better than emergency shelters) public housing or private assisted housing projects in high poverty neighborhoods at baseline. The first row of the exhibit shows that the analysis will be able to detect a difference between mean annual earnings of the CBRR and UC groups of \$1,170 with 80% likelihood. Given that only two of the interventions tested have a partial focus on the labor market—though better, more stable housing may enable steadier employment and resulting greater earnings—the study design is weaker for detecting these effects. On the one hand, it is by no means assured that even an intervention directly focused on employment and training could produce an earnings impact of over \$1,200 per year. On the other hand, a true impact substantially smaller than this amount—say, an impact on annual earnings of \$600—would have little potential to move families out of poverty and hence may not be important to detect with high confidence.

²⁰ We note that Cowal, et al, (2002), finds 44 percent. In as much as that estimate applies here, we will have slightly lower power.

Exhibit B-6. Minimum Detectable Effects for Annual Earnings Impacts by Pairwise Comparison

Sample	Expected Number of Completed Follow-up Survey Interviews		MDE (dollars)
	First Assignment Group	Second Assignment Group	
CBRR vs. UC	433	435	1,170
SUB vs. UC	453	411	1,172
PBTH vs. UC	272	257	1,498
CBRR vs. SUB	290	329	1,385
CBRR vs. PBTH	177	175	1,837
SUB vs. PBTH	180	194	1,783

Notes: (1) The MDE's are based on calculations which assume that two-sided tests are used at the 10 percent significance level, the desired power is 80 percent, and the regression R^2 is identical to the MTO adult annual earnings impact regression. (2) All MDE's assume a 75% survey response rate, with no finite population correction. (3) The variance of earnings is derived from the standard error of the ITT impact estimate for the experimental group (n=1,729) vs. the treatment group (n=1,310) in the MTO Demonstration: \$254.

Power Calculations for Child Well-Being Outcomes

Exhibit B-7 shows the MDEs for the child well-being outcomes that are measured in the child assessments. The expected sample size for children in the age range to be administered the child assessments is 1,330 children.²¹ The MDEs shown in Exhibit B-7 are based on a 61 percent completion rate, which is the expected completion rate for these assessments in the 36-month data collection.²² The table shows MDEs for binary outcomes and for standardized continuous outcomes (rightmost column). The minimum detectable effect size column shows the MDEs in standard deviation units. Exhibit B-8 shows the MDEs for those child well-being outcomes that are measured in the child survey. These MDEs assume a 64 percent completion rate on a sample size of 1,380 children.²³ We expect that the child well-being outcomes that are measured in the adult survey will have somewhat larger sample sizes (we assume a 75 percent completion rate for the adult survey) and so will have MDEs somewhat smaller than those shown in Exhibits B-7 and B-8.

²¹ See Exhibit B-3.

²² The child assessment data collection at 18 months had an overall completion rate of 66 percent. This rate was the product of the 81 percent response rate of the family heads and the 81 percent completion rate for focal children identified in interviews with family heads. At 36 months, we assume a 75 percent response rate for family heads and the same 81 percent completion rate for identified focal children in the child assessment age range. The product of these two assumptions is an overall child assessment completion rate of 61 percent.

²³ The child survey data collection at 18 months had an overall completion rate of 69 percent. This rate was the product of the 81 percent response rate of the family heads and the 85 percent completion rate for focal children identified in interviews with family heads. At 36 months, we assume a 75 percent response rate for family heads and the same 85 percent completion rate for identified focal children in the child survey age range. The product of these two assumptions is an overall child assessment completion rate of 64 percent.

Exhibit B-7. Minimum Detectable Effects for Child Well-Being Outcomes Measured in Child Assessments

Sample	Expected Number of Completed Child Assessments		MDE if Mean Outcome for the Second Assignment Group is:			Minimum Detectable Effect Size
	First Assignment Group	Second Assignment Group	0.1 (or 0.9)	0.3 (or 0.7)	0.5	
CBRR vs. UC	204	252	6.9 pp	10.5 pp	11.5 pp	0.23 sd
SUB vs. UC	213	194	7.3 pp	11.1 pp	12.1 pp	0.24 sd
PBTH vs. UC	128	121	9.3 pp	14.1 pp	15.4pp	0.31 sd
CBRR vs. SUB	136	155	8.6 pp	13.1 pp	14.3 pp	0.29 sd
CBRR vs. PBTH	83	82	11.4 pp	17.3 pp	18.9 pp	0.38 sd
SUB vs. PBTH	85	91	11.0 pp	16.8 pp	18.4 pp	0.37 sd

Notes: (1) The MDE's are based on calculations which assume that two-sided tests are used at the 10 percent significance level, the desired power is 80 percent, and the regression R^2 is 0.04. (2) All MDE's assume a 61% assessment completion rate, with no finite population correction. (3) Minimum detectable effect sizes are measured in standard deviations.

Exhibit B-8. Minimum Detectable Effects for Child Well-Being Outcomes Measured in Child Survey

Sample	Expected Number of Completed Child Survey Interviews		MDE if Mean Outcome for the Second Assignment Group is:			Minimum Detectable Effect Size
	First Assignment Group	Second Assignment Group	0.1 (or 0.9)	0.3 (or 0.7)	0.5	
CBRR vs. UC	222	223	6.9 pp	10.6 pp	11.6 pp	0.23 sd
SUB vs. UC	232	211	7.0 pp	10.6 pp	11.6 pp	0.23 sd
PBTH vs. UC	139	132	8.9 pp	13.6 pp	14.8 pp	0.30 sd
CBRR vs. SUB	148	169	8.2 pp	12.6 pp	13.7 pp	0.27 sd
CBRR vs. PBTH	91	90	10.9 pp	16.6 pp	18.2 pp	0.36 sd
SUB vs. PBTH	92	99	10.6 pp	16.2 pp	17.6 pp	0.35 sd

Notes: (1) The MDE's are based on calculations which assume that two-sided tests are used at the 10 percent significance level, the desired power is 80 percent, and the regression R^2 is 0.04. (2) All MDE's assume a 64% survey response rate, with no finite population correction. (3) Minimum detectable effect sizes are measured in standard deviations.

B.2.4 Procedures with Special Populations

In this study we may encounter interview respondents whose first language is Spanish. As we did in prior data collection efforts, we will translate the 36-month adult follow-up survey instrument into Spanish, for administration in the language most comfortable for the respondent. The consent forms also will be made available in Spanish. All baseline interviews were conducted in either English or Spanish, with no need for other languages, thus we do not anticipate other language needs for the 36-month follow-up.

Our data collection efforts to date have shown that a small number of adult respondents were incarcerated after enrollment. We will work with the contractors' IRB to determine whether or not we can pursue special populations. In addition, during the 18-month data collection some adults reported that selected focal children were not able to participate in the data collection because of medical conditions. We have not pursued data collection with children in these circumstances.

B.3 Maximizing the Response Rate and Minimizing Non-Response Risk

The target response rate for the 36-month data collection is 75 percent of all eligible adult and child sample members. The research team will strive to exceed this target if possible to increase the precision of estimates and minimize non-response bias.

For the 18-month survey, the contractor has exceeded the target 75 percent response rate for the head of household survey with an overall completion rate of 81 percent

During the 36 month data collection period, non-response levels and the risk of non-response bias will be minimized in the following ways:

- Local site interviewers. The Contractor will rely on the local site interviewers to lead the local data collection activities for the 36-month follow-up survey. These local interviewers are established in the study communities and are well known to the families participating in the study. The Contractor will also recruit additional interviewers who are skilled at working with this population, to support the data collection.
- The contractor team will use a field data collection management system that permits interactive sample management and electronic searches of historical tracking and locating data. This will allow the contractor to monitor response rates closely and to work the sample groups for each of the study interventions evenly.

The contractor team has conducted intensive tracking with study families since study enrollment. The tracking approach was designed to maintain contact with the study sample every three months over the follow-up period which is essential to achieving the highest possible response rate for the 36-month follow-up survey. The success of this approach is reflected in the high response rates achieved in the 18-month follow up survey.

Prior to the start of data collection for the 36 month survey, the data collection team will review the study sample to identify cases that do not have tracking updates from any point of tracking. These cases will be classified as “high priority cases” and will be assigned to a Senior Field Interviewer. In addition to any information obtained through the tracking process (e.g. returned letters, information from secondary contacts, interwave tracking interviews), the research team will also contact the program providers in the study sites to which the respondents were referred for assistance, to request any information available about the family’s location, to ensure that all possible sources of locating information are available to the team. High priority cases will be reviewed by the Field Manager regularly to make sure all leads are followed. Field interviewers will receive a comprehensive document for all released cases containing respondent’s information history collected through the baseline, all interwave tracking components and 18-month follow-up efforts (all address, home/cell phone numbers and emails), secondary contacts and any relevant notes collected during the tracking efforts.

As noted in Part A.9, we will again use incentives to thank participants for their time responding to the 36-month data collection effort. Adult respondents will receive \$60 for completing the 36-month adult survey; child respondents will receive \$25 incentives for their participation. Incentive payments will be provided in the form of a money order. These amounts represent a modest increase of \$10 over the amount provided for participation in the 18-month follow-up effort.

With continuity of data collection staff, ongoing tracking and increased incentives, we are well-positioned to achieve the target 75 response rate for the 36-month data collection. However, if we find variation in response rates across intervention groups or sites, the Department may offer differential incentive payments as done in the MTO Final Evaluation (Gebler, 2012). This would involve providing a slightly higher incentive payment (\$75) for adults who are in the harder to complete groups.

B.4 Test of Procedures

The instruments used in the 36-month follow-up data collection effort are virtually identical to the ones currently in use for the 18-month follow-up data collection effort. The interview lengths used to calculate the burden estimates in Section A12 are based on the actual experience with the 18-month survey administration efforts for both the adult and child respondents. The small number of new items that are added to the adult interview were used in other prior studies of similar populations. We therefore do not anticipate any challenges for respondents to answer them.

B.5 Individuals Consulted on Statistical Aspects of the Design

The individuals shown in Exhibit B-9 assisted HUD in the statistical design of the evaluation.

Exhibit B-9. Individuals Consulted on the Study Design

Name	Telephone Number	Role in Study
Dr. Stephen Bell Abt Associates Inc.	301-634-1721	Co-Principal Investigator
Dr. Marybeth Shinn Vanderbilt University	615-322-8735	Co-Principal Investigator
Dr. Jill Khadduri Abt Associates Inc.	301-634-1745	Project Quality Advisor
Mr. Jacob Klerman Abt Associates Inc.	617-520-2613	Project Quality Advisor
Dr. Martha Burt Consultant to Abt Associates Inc.	202-261-5551	Project Advisor
Dr. Dennis Culhane University of Pennsylvania	215-746-3245	Project Advisor
Dr. Ellen Bassuk, Center for Social Innovation and National Center on Family Homelessness	617-467-6014	Project Advisor
Dr. Beth Weitzman New York University	212-998-7446	Project Advisor
Dr. Larry Orr Consultant to Abt Associates Inc.	301-467-1234	Project Advisor

Inquiries regarding the statistical aspects of the study's planned analysis should be directed to:

Dr. Stephen Bell	Co-Principal Investigator	Telephone: 301-634-1721
Dr. Marybeth Shinn	Co-Principal Investigator	Telephone: 615-322-8735

References

- Allison, Paul D., (2002). Missing Data. Thousand Oaks, CA: Sage University Paper No. 136
- Angrist, J. D., & Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton, NJ: Princeton University Press.
- Burt, M.R. (2006). Characteristics of transitional housing for homeless families: Final report. Prepared for the US Department of Housing and Urban Development.
- Burt, M. R., The Urban Institute, Pearson, C. L., Montgomery, A. E., & Walter R. McDonald & Associates, Inc. (2005). *Strategies for preventing homelessness*. Washington, D.C and Rockville, MD: Urban Institute and Walter R. McDonald & Associates, Inc
- CDC Youth Risk Behavior Surveillance substance use questions. (2011). http://www.cdc.gov/healthyyouth/yrbs/pdf/questionnaire/2011_hs_questionnaire.pdf
- Cowal, K., Shinn, M., Weitzman, B.C., Stojanovic, D., & Labay, L. (2002). Mother-child separations among homeless and housed families receiving public assistance in New York City. *American Journal of Community Psychology*, 30, 711-730.
- Culhane, D., Metraux, S., & Byrne, T. (2011). A prevention-centered approach to homelessness assistance: A paradigm shift? *Housing Policy Debate*, 21, 295-315
- Culhane, D. P., Metraux, S., Park, J.M., Schretzman, M. & Valente, J. (2007). Testing a typology of family homelessness based on patterns of public shelter utilization in four U.S. jurisdictions: Implications for policy and program planning. *Housing Policy Debate*, 18(1), 1-28.
- Culhane, D. P. (1992). The quandaries of shelter reform: An appraisal of efforts to “manage” homelessness. *Social Service Review*, 66, 428–440.
- Dahl, R.E. & Harvey, A. G. (2007). Sleep in children and adolescents with behavioral and emotional disorders. *Sleep Medicine Clinics*, 2, 501-511.
- Duffer, Allen P. et al., "Effects of Incentive Payments on Response Rates and Field Costs in a Pretest of a National CAPI Survey" (Research Triangle Institute, May 1994).
- Fisher, B. W., Shinn, M., Mayberry, L., & Khadduri, J. (in press). Leaving homelessness behind: Housing decisions among families exiting shelter. *Housing Policy Debate*.
- Gebler, Nancy and Margaret L. Hudson, Matthew Sciandra, Lisa A. Gennetian and Barbara Ward. (2012). Achieving MTO's High Effective Response Rates: Strategies and Tradeoffs. *Cityscape*, 14(2), 57-86.
- Goodman R (1997). The Strengths and Difficulties Questionnaire: A Research Note. *Journal of Child Psychology and Psychiatry*, 38, 581-586.
- Greene, W. H. (2003). *Econometric analysis: Fifth edition*. Upper Saddle River, NJ: Prentice-Hall, Inc.

- Gubits, Daniel et al., *Interim Report, Family Options Study*, Washington, DC: U.S. Department of Housing and Urban Development, March 2013.
http://www.huduser.org/portal/publications/homeless/hud_503_FOS_interim_report.html
- Herbers, J. E., Cutuli, J. J., Supkoff, L. M., Heistad, D., Chan, C.-K., Hinz, E., & Masten, A. S. (2012). Early reading skills and academic achievement trajectories of students facing poverty, homelessness and high residential mobility. *Educational Researcher*, 41(9), 366-374.
- Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. Lucien M. Le Cam & Jerzy Neyman & Elizabeth L. Scott (editors). Berkeley, CA: University of California Press, Vol. 1, 221-233.
- Locke, G., Khadduri, J. & O'Hara, A. (2007). Housing models (Draft).
- Masten, A. S., Neemann, J., & Andenas, S. (1994). Life events and adjustment in adolescents: The significance of event independence desirability, and chronicity. *Journal of Research on Adolescence*, 4(1), 71-97.
- Mayberry, L. S., Shinn, M., Benton, J. G., & Wise, J. (in press). Families experiencing housing instability: The effects of housing programs on family routines and rituals. *American Journal of Orthopsychiatry*.
- Mills, G., Gubits, D., Orr, L., Long, D., Feins, J., Kaul, B., Wood, M., Amy Jones and Associates, Cloudburst Consulting, & the QED Group. (2006). Effects of Housing Vouchers on Welfare Families: Final Report. Prepared for the U.S. Department of Housing and Urban Development, Office of Policy Development and Research. Cambridge, MA: Abt Associates Inc.
- Mistry RS, Lowe E, Benner AD, Chien N. (2008). Expanding the family economic stress model: Insights from a mixed methods approach. *Journal of Marriage and Family*, 70,196–209.
- Orr, L.L., Feins, J., Jacob, R., Beecroft, E., Sanbonmatsu, L., Katz, L., Liebman, J. & Kling, J. (2003). *Moving to Opportunity interim impacts evaluation: Final report*. Cambridge, MA: Abt Associates Inc. and National Bureau of Economic Research.
- Ponitz, C. C., McClelland, M. M., Jewkes, A. M., Connor, C. M., Farris, C. L., & Morrison, F. J. (2008). Touch your toes! Developing a direct measure of behavioral regulation in early childhood. *Early Childhood Research Quarterly*, 23, 141–158. doi: 10.1016/j.ecresq.2007.01.004.
- Puma, Michael, J., Robert B. Olsen, Stephen H. Bell, and Cristofer Price (2009). What to Do When Data Are Missing in Group Randomized Controlled Trials (NCEE 2009-0049). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
- Rafferty, Y., Shinn, M., & Weitzman, B. C. (2004). Academic achievement among formerly homeless adolescents and their continuously housed peers. *Journal of School Psychology*, 42(3), 179-199. doi: 10.1016/j.jsp.2004.02.002
- Ramirez, M., Masten, A., Samsa, D. (1991, April). *Fears in homeless children*. Paper presented at the biennial meeting for the Society for Research in Child Development, Seattle, WA.

- Rog, D.J. & Randolph, F.L. (2002). A multisite evaluation of supported housing: Lessons learned from cross-site collaboration. *New Directions for Evaluation*, 94, 61-72.
- "National Adult Literacy Survey Addendum to Clearance Package, Volume II: Analyses of the NALS Field Test" (Educational Testing Service, September 1991), pp. 2-3.
- Shinn, M., Weitzman, B. C., Stojanovic, D., Knickman, J. R., Jiminez, L., Duchon, L., James, S., & Krantz, D.H. (1998). Predictors of homelessness from shelter request to housing stability among families in New York City. *American Journal of Public Health*, 88 (10), 1651-1657.
- Shinn, M., Schteingart, J. S., Williams, N. P., Carlin-Mathis, J., Bialo-Karagis, N., Becker-Klein, R., & Weitzman, B. C. (2008). Long-term associations of homelessness with children's well-being. *American Behavioral Scientist*, 51, 789-810.
- Snyder, C. R., Hoza, B., Pelham, W. E., Rapoff, M., Ware, L., Danovsky, M., . . . Stahl, K. J. (1997). The development and validation of the Children's Hope Scale. *Journal of Pediatric Psychology*, 22(3), 399.
- Squires, J., Bricker, D. D., Potter, L., & Twombly, E. (2009). *Ages & stages questionnaires : a parent-completed child monitoring system* (3rd ed.). Baltimore: Paul H. Brookes Pub. Co.
- Stretch, J. J. & Kreuger, L. W. (1993). A social-epidemiological five year cohort study of homeless families: A public/private venture policy analysis using applied computer technology. *Computers in Human Services*, 9(3-4), 209-230.
- U.S. Department of Housing and Urban Development (2012). Volume I of the 2012 Annual Homeless Assessment Report (AHAR). <https://www.onecpd.info/resource/2753/2012-pit-estimates-of-homelessness-volume-1-2012-ahar>
- U.S. Department of Housing and Urban Development (2013). Volume II of the 2012 Annual Homeless Assessment Report (AHAR).
- Voight, A., Shinn, M., & Nation, M. (2012). The longitudinal effects of residential mobility on the academic achievement of urban elementary and middle school students. *Educational Researcher*, 41, 385-392.
- White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48, 817–830.
- White, H. (1984). *Asymptotic theory for econometricians*. Orlando, FL: Academic Press.
- Woodcock, R. W., McGrew, K. S., Mather, N., & Schrank, F. A. (2001). *Woodcock-Johnson III tests of cognitive abilities*. Itasca, IL: Riverside Pub