Data Collection and Analysis Plan

FAMILY OPTIONS STUDY

U.S. Department of Housing and Urban Development | Office of Policy Development and Research

Visit PD&R's website

www.huduser.org

to find this report and others sponsored by HUD's Office of Policy Development and Research (PD&R). Other services of HUD USER, PD&R's research information service, include listservs, special interest reports, bimonthly publications (best practices, significant studies from other sources), access to public use databases, and a hotline (800-245-2691) for help accessing the information you need.

Data Collection and Analysis Plan

FAMILY OPTIONS STUDY

Prepared for:
U.S. Department of Housing
and Urban Development
Washington, D.C.

Prepared by:
Abt Associates Inc.
Daniel Gubits
Michelle Wood
Debi McInnis
Scott Brown
Brooke Spellman
Stephen Bell

In partnership with :
Vanderbilt University
Marybeth Shinn

March 2013

Disclaimer

Family Options Study Revised Data Collection and Analysis Plan

Table of Contents

Chapter 1. Introduction and Evaluation Design

1.1 Overview and Objectives

The U.S. Department of Housing and Urban Development (HUD) has invested considerable time and energy in strategies to address family homelessness. In addition, in response to HUD Continuum of Care (CoC) funding requirements and in an effort to maximize the impact of limited resources, communities are systematically examining their homeless assistance systems and deciding which housing and service interventions should be funded. Unfortunately, past research is inadequate to guide federal policy and local practice. While there is a significant amount of research on the characteristics and needs of homeless families and an emergent body of descriptive research on intervention programs and outcomes for families who use them, there is almost no information about the relative effectiveness of different interventions. In response to this need for information, HUD is sponsoring the Family Options Study[1], a multi-year evaluation led by Abt Associates.

The objective of the Family Options Study is to provide research evidence to help federal policymakers, community planners, and local practitioners make sound decisions about the best ways to address homelessness among families. The study will compare four combinations of housing and service interventions for homeless families who have been in emergency shelters for at least seven days. The study is conducted as a rigorous, multi-site experiment, to determine what interventions work best to promote family stability and well-being. Within the limits of statistical power, the study will also analyze what types of families benefit most from each intervention. The four interventions studied are:

- ***Permanent Housing Subsidy (SUB).*** The permanent subsidy is typically in the form of a Housing Choice Voucher, *without* dedicated supportive services.
- ***Project-Based Transitional Housing (PBTH).*** This intervention features temporary housing assistance offered for up to 24 months (with average expected length of stay of 6 to 12 months) in transitional housing facilities combined with supportive services.
- ***Community-Based Rapid Re-housing (CBRR).*** CBRR provides temporary rental assistance for 2 to 6 months (potentially renewable for periods up to 18 months) in conventional, private-market housing, with limited, housing-focused services.
- ***Usual Care (UC).*** UC includes any additional time spent in emergency shelters and the services that people would normally access on their own from shelter in the absence of these other interventions.

This document is the deliverable under Task 4, the revised Data Collection and Analysis Plan (DCAP). It updates the conceptual framework for the study developed in the 2009 Research Design document developed under Task Order 1 of this contract. This document revises and expands the plans for data collection documented in the 2009 Research Design deliverable. The current document also adds details about plans for conducting the impact analysis and provides an updated assessment of statistical power for the study, taking into account final enrollment across the four interventions and target response rates for the follow-up survey.

[1] This study was originally titled *The Impact of Housing and Services Interventions on Homeless Families*.

The document is organized as follows. This chapter provides an overview of the evaluation design, including the interventions studied, research questions, outcome measures, and random assignment process. Chapter 2 discusses the follow-up survey and administrative data sources to be used in the study and the plans for data collection. Chapter 3 presents the detailed analysis plan, including the hypotheses to be tested, impact estimation model, strategy for addressing multiple comparisons, subgroup analysis, addressing no-shows and crossovers, and sample sizes and statistical power.

In this first chapter, we discuss the key features of the experimental design. We begin by describing the four interventions to be tested (Section 1.2). Next, we describe the design of the study, including how families were assigned to interventions in order to permit experimental contrasts (Section 1.3). Finally, Section 1.4 lists the study's research questions and specifies the outcomes to be measured to assess the effects of the tested interventions.

1.2 Interventions Studied

This section describes in detail the four interventions to be studied and the characteristics of each, in terms of type and duration of housing, and types, duration, and intensity of services. The section also discusses selection of programs in each community that met the intervention definitions developed for the study. Exhibit 1-1 displays the requirements for the interventions that the team developed for the study, organized by these dimensions. Families in the study were randomly assigned to one of these four interventions, and the study will compare the effectiveness of interventions by comparing the outcomes of families assigned to these interventions.

Permanent Housing Subsidy (SUB): Shelter followed by a dedicated permanent housing subsidy with limited housing placement assistance, but no other targeted services after placement.

This intervention is a deep, permanent housing subsidy, in the form of a tenant-based or project-based voucher or a public housing unit. The subsidy is deep in that participants pay approximately 30 percent of their income towards rent, with the remainder of rent or housing cost subsidized by a public program (this is the formula for tenant-based voucher assistance and public housing). Altogether, 10 of the 12 sites offered this intervention, and in all but two cases the subsidies were in the form of tenant-based housing vouchers. The two exceptions were Honolulu, HI, where the PHA provided public housing units, and Bridgeport, CT, where the subsidy intervention is in the form of units subsidized by project-based vouchers.

This intervention may also include housing placement assistance to help participants lease a unit and make use of the voucher and move-in services such as furniture, transportation, or help with getting utilities turned on. No other services beyond what is typically offered to all voucher program participants were offered to families in the study.

Project-Based Transitional Housing (PBTH): Shelter followed by transitional housing with intensive services, without a guaranteed housing subsidy upon program exit.

This intervention is the current transitional housing model of time-limited housing assistance coupled with a wide array of services that includes, at a minimum, assessment of family needs, case management, and provision of, or referral for, services to meet identified needs. Service domains for assessment and referral may include employment; child care; housing; transportation; entitlements; medical; behavioral health; trauma; safety; emotional, cognitive, and developmental needs of children; child welfare; and family preservation and reunification, as well as other services.

Exhibit 1-1: Requirements of Each Intervention

	SUB	CBRR	PBTH	UC
Housing Subsidy Type	Deep rent subsidy	Deep rent subsidy	Deep subsidy (in some cases, zero rent)	No active referral to deep housing subsidies or programs by the shelter system beyond what is ordinarily done; families may apply for such housing assistance on their own.
Housing Duration	Permanent	Temporary: expected average of 6 months, maximum of 18 months	Temporary: expected average of 6 months, maximum of 2 years	Shelter stays as usual
Service Type	Housing search, placement and move-in only All mainstream services allowed, but without case management and referrals as part of the program	Limited to housing search, lease up assistance, and basic service coordination All mainstream services allowed, but without case management and referrals as part of the program	Comprehensive assessment of all family members; case management; referrals and/or direct service provision for identified needs. Domains for assessment and referral will likely include but not be limited to employment; child care; housing search and placement; transportation; entitlements; medical; behavioral health; trauma; safety; emotional, cognitive, and developmental needs of children; child welfare: family preservation and reunification.	Services ordinarily provided in shelter, which may include referrals to other programs including transitional housing programs and to services available in the community
Service Duration	Maximum of 1 month post move-in	Same as limit for housing	Same as limit for housing. Some programs provide some follow-on services after families move from the housing.	Same as what is available in any of the UC assistance received
Service Intensity	Any, for period and purposes allowed	Any, for period and purposes allowed	Services should be available as needed and requested	Same as what is available in any of the UC assistance received

For this study, the transitional housing provided is in project-based facilities and may or may not require rent contributions from the family. Any time limit on tenure up to and including two years would satisfy our definition. Referrals to permanent housing assistance at the end of the transitional housing period are permitted, but permanent assistance cannot be guaranteed. Thus, a transition-in-place model that allows families to stay if they can afford it or can find a subsidy is not included in this intervention.

We expected to find—and did find—variation across programs and sites regarding the philosophies of transitional housing programs and the restrictions imposed on participants with respect to motivation, sobriety, and requirements for participation in treatment or activities designed to foster self-sufficiency. In selecting sites, we attempted to locate communities with some transitional housing programs with very few restrictions, so that most families would be eligible for at least one of the transitional housing options, but at the same time we both wanted and needed to study transitional housing as it exists. For the PBTH intervention, there is considerable variability among services in PBTH program both within and across sites in structure (for example, delivered on-site or off-site), emphases (for example, the degree of focus on parenting, building adult human capital and employment, managing behavioral health issues, child care, or children's programming), intensity, and duration.

In terms of housing support, a transitional housing program with a two-year housing subsidy may not be substantially different in the short term from the permanent subsidy offered in the SUB intervention. But we hypothesize that the expectation that the SUB housing subsidy will be permanent will affect the behavior of families, compared to behavior under a temporary subsidy in the PBTH intervention. Nevertheless, the fact that the PBTH and SUB interventions will look very similar, up to the time limit of the transitional housing programs, suggests that it would be necessary to follow families beyond the 18-month survey currently planned in order to detect differences. It should be noted, however, that previous research has shown that actual stays in PBTH average 6-12 months, even though programs allow families to remain for up to 24 months.[2]

Community-Based Rapid Re-housing (CBRR): Shelter followed by temporary rental assistance without services.

Families assigned to this intervention receive short-term rental assistance, lasting from 2 to 18 months, with limited services focused on housing and on "coordination" or referral to services available in the community. In the study sites, CBRR is funded through the federal Homelessness Prevention and Rapid Re-housing Program (HPRP), which permits rental assistance for up to 18 months, with eligibility re-determination required every three months. However, in this study, to allow the clearest possible test of the impact of services, CBRR is intended to be similar in duration to PBTH (six months), ensuring that the distinction between the two interventions is in the location of housing and in the service models. Thus, the goal was for the CBRR intervention tested in this evaluation—like the typical duration of PBTH—to consist of ***temporary rental assistance*** available to each family for approximately six months. Rental assistance limited to two months or less was not included in the CBRR intervention. To provide a contrast with PBTH, the CBRR intervention includes only services focused on housing placement and service coordination, rather than the comprehensive array of services provided on-site and by other agencies as an integral part of a PBTH program. In addition, CBRR and PBTH are distinguished by the location of the housing. PBTH provides housing in project-based settings while the CBRR intervention involves conventional housing in the community.

---

2 Within a 12-month reporting period, the 2010 AHAR reported an average length of stay in transitional housing of 176 nights.

Usual Care (UC) Group: Shelter with whatever services are normally provided in shelter and any other housing or services interventions that may be available in the community.

The key policy question to be answered by this study is the effect of the more resource-intensive interventions (SUB, CBRR, and PBTH) compared to what would happen in the absence of these interventions. Families in this intervention receive “usual care,” which means whatever programs, assistance, and services are normally available in the community, excluding only the resources made available by the study to the families participating in the three “treatment” arms of the study. For example, the housing vouchers set aside for the study may not be made available the Usual Care group (or to the PBTH or CBRR group), but families assigned to UC may still join or remain on any waiting lists for vouchers or other forms of housing assistance available in the community.

Families randomly assigned to Usual Care will not be prevented from obtaining any other type of assistance they might find on their own in their communities, including transitional housing (including transitional housing programs that are part of the study) if they find it on their own or receive information about it from shelter staff. The UC group also is not precluded from receiving one-time assistance funded by the federal HPRP program, such as rapid re-housing. They may also be referred to other, non-housing services, but not to services that are packaged with one of the “treatment” arms of the study. In each community selected for the study, we are documenting what UC assistance involves and will document the services that each family assigned to the UC group receives.

The UC intervention includes all family emergency shelters participating in the study in a research site. The study placed no limits on further lengths of stay in shelter beyond the 7 or more days that the family had already been there at the time of random assignment, although shelters may impose their own, pre-existing, time limits. The study also placed no limits on the services the shelter may offer, including services after shelter such as a link to a case manager if they are ordinarily offered to the shelter’s clients.

We originally considered defining this intervention as Shelter Only and preventing direct referrals to other housing supports for families assigned to this group. This would have meant that those assigned to UC would have been precluded from any referrals to transitional housing by shelter staff, even if this was the shelter’s usual practice. Our conclusion was that defining the comparison intervention as UC was necessary for ensuring that families who agreed to participate in the study were not made worse off relative to families who opted out of the study.

Implications for program and site selection. The study aims to estimate the impact of the three different approaches—relative to each other and to UC. To do so, the study team needed to begin by defining the approaches. Although practitioners and researchers use shorthand terms such as “transitional housing” or “supportive housing,” these labels do not reflect uniform approaches. In reality, as Rog and Randolph (2002) note,[3] even when programs of a particular “type” are specifically chosen for study, their characteristics can overlap considerably with other programs that nominally use a “different” model. Therefore, the study team first specified a consistent definition of each intervention (shown in Exhibit 1-1). Then, during initial site selection, the team visited potential study programs, collected data on their operations, and completed an assessment for each candidate program. We selected programs that fit our

[3] 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.

definitions of the interventions based on our assessments, rather than based on programs' self-descriptions. Team members also visited all programs in which at least five study families participated in the middle of the study to collect additional information about program operations (see Section 2.4). Some programs place restrictions on the types of families that they will accept. Although we considered asking programs to waive restrictions so as to accept all families to all program models, we concluded that this would require changing the program models, so would not fairly represent the types of programs currently offered to homeless families. Rather, we sought sites where, across all participating programs that fit into a particular intervention type, most or all families would be accepted. Similarly, we looked for sites where at least three of the interventions were available. The design allowed us to include sites that do not have all interventions, or where some families are not eligible for any programs in a particular intervention, as described in the next section. However, this reduces statistical power to examine some experimental contrasts, so we tried to minimize this.

1.3 Study Design

The study is being carried out under a randomized experimental design. From September 2010 through January 2012, the research team recruited 2,307 homeless families who had been in emergency shelter for at least 7 days across 12 sites. The study is not designed to capture the experiences of families who seek assistance directly from transitional housing programs without first entering emergency shelters.

We excluded families who left shelter in fewer than 7 days because the more intensive interventions considered in this study may not be necessary for families who can resolve a housing crisis quickly. During those first 7 days, we expected shelters to continue to provide all services and referrals they ordinarily provide to help families leave shelter up until the point of random assignment. Families were then assigned, as close to the 7-day mark as was feasible, to the SUB, PBTH, CBRR, or UC interventions. We also excluded families who, in the judgment of emergency shelter staff, would be eligible for assistance from permanent supportive housing (PSH) programs if such assistance was available in the community. Permanent supportive housing programs funded by HUD require participants to have severe and persistent chronic disabilities to be eligible. These exclusions of families with disabilities, who would be eligible for PSH programs where such programs were available, were applied by shelter staff. If no PSH programs were available in a community all families who met the other study criteria and who agreed to participate in the study were included.

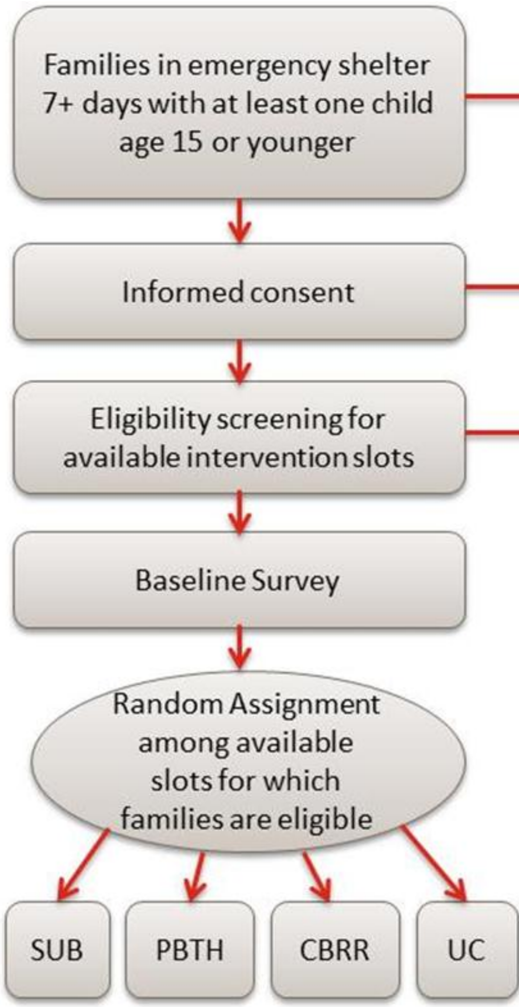
Implementing a random assignment design presented several challenges. In the original design of the study, all families enrolled in the study were to be randomly assigned to one of the four assignment groups (PBTH, SUB, CBRR, or UC). A number of factors prevented the study from being implemented exactly as planned, with each family having a chance of being assigned to all four groups. First, three of the 12 sites were able to provide only three of the four assignment groups.[4] Second, families only had assignment groups available to them where at least one provider of the service type had an available slot. Third, some service providers had their own unique eligibility requirements for families. Prior to random assignment, families were screened for eligibility for the providers that had available slots. For an intervention option to be available to a family undergoing random assignment, there needed to be at least one slot available at an intervention provider for which the family met provider-specific eligibility

[4] Atlanta and Baltimore did not have subsidies (SUB) available for families in the study and Boston did not have providers of PBTH.

requirements. Cumulatively, these factors resulted in most study families not having all four assignment options available to them at the time of random assignment. Of the 2,307 families enrolled in the study, 469 had all four assignment options available to them at random assignment, 1,555 families had three assignment options, and 283 families had two assignment options.

Exhibit 1-2 illustrates the random assignment model we used to use to allocate families to interventions.

Exhibit 1-2: Random Assignment Design



As shown at the top of the exhibit, the study population is all families who have been in an emergency shelter for at least seven days and who have at least one child 15 or younger. The latter restriction was imposed because child outcomes are important to the study, so we wanted to make sure that the family would still have a child under the age of 18 at the time of follow-up data collection. (The study will not attempt to measure outcomes for young adults over the age of 18 who were children at the time of random assignment, because the sample will not be large enough.)

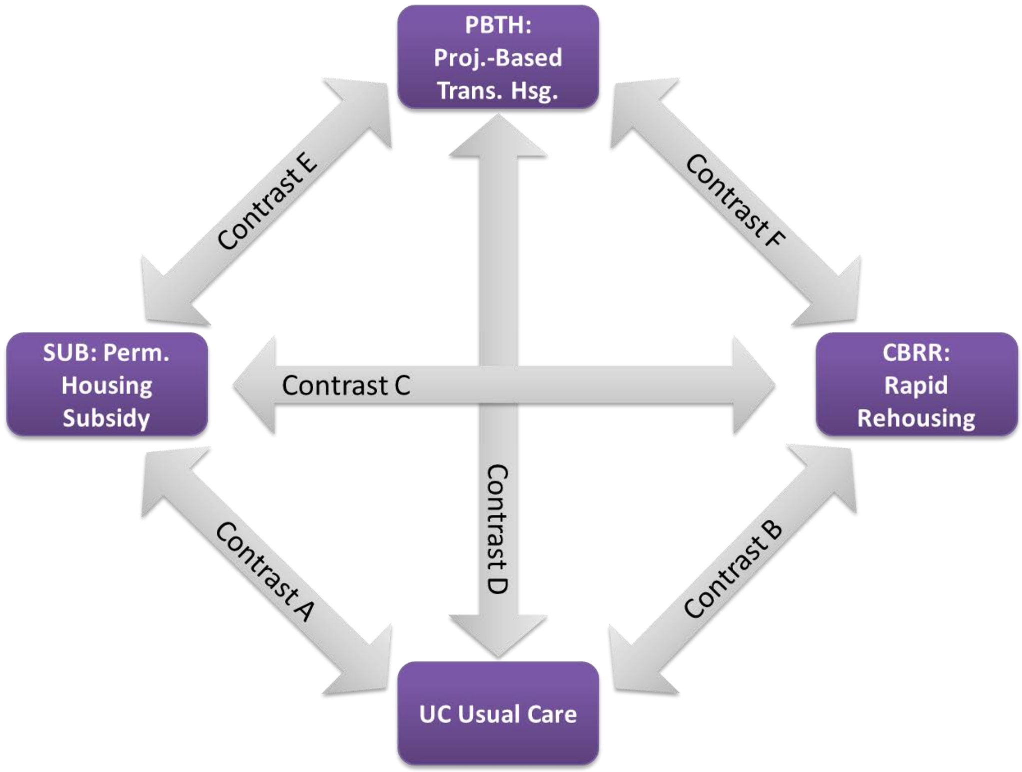
In each site, families who met the basic study eligibility requirements (presence in shelter for 7+ days and at least one child age 15 or younger) and who provided informed consent were randomly assigned to one of the interventions or to UC. The intervention options available to a family were determined by: (1) the availability of the intervention at the time of random assignment; and (2) eligibility of the family for the available interventions. Families were only considered for random assignment to interventions to which they appeared eligible and where slots were available at the time of random assignment. Eligibility for available slots was determined by families' responses to screening questions selected for each program. These questions were designed to screen out families who would not meet the program's core eligibility requirements. In all sites, UC was always available, and we assumed all families were eligible for UC.

This design assures that comparisons of interventions will involve well-matched groups across interventions and that any observed differences in outcomes are caused by the differential treatment families receive and not by any pre-existing differences among the families.

Although assignment to interventions was at random, within interventions families did not need to be assigned to service providers at random. Instead, assignment sometimes was made on the basis of family characteristics, following current program practice. For example, if a transitional housing program specializes in families with a particular profile (only families with young children, or only families where the mother has been clean and sober for some period), then among families randomly assigned to Project-based Transitional Housing at that site, only those that fit that profile were assigned to that service provider.

The research team is tracking families at three-month intervals and administering very short tracking surveys at six-month intervals in preparation for a follow-up survey to be administered at 18 months. This 18-month follow-up survey will allow us to address questions about the effectiveness of the interventions for different outcomes. Comparison among the four interventions allows six experimental contrasts, as shown in Exhibit 1-3.

Exhibit 1-3: Contrasts Among Experimental Interventions



1.4 Research Questions and Outcome Measures

The study seeks to answer six major questions:

1. What is the relative effectiveness of homeless interventions in ensuring housing stability of homeless families?
2. Are the same interventions that are effective for short-term stability of homeless families effective for longer-term stability as well?
3. What is the relative effectiveness of homeless interventions in ensuring the well-being of homeless parents and self-sufficiency of homeless families?
4. Do some interventions promote family preservation and benefit children's well-being more than other interventions?
5. Are different homeless interventions more effective for some categories of homeless families than for others?
6. What features of housing and services explain the effectiveness (or lack thereof) of various homelessness interventions?

These primary research questions ask about five broad classes or domains of outcomes:

- Housing Stability
- Family Preservation
- Self-Sufficiency
- Adult Well-Being
- Child Well-Being

These family-level outcomes will be measured through the 6- and 12-month tracking interviews, the 18-month follow-up interviews, and administrative records from Homeless Management Information Systems (HMIS). Measurement of child well-being will include direct observation of younger children and interviews with older children during the 18-month follow-up interviews.[5]

Within each of these broad classes of outcomes, we will measure several different constructs in order to capture adequately the broad range of potential outcomes that these multi-faceted housing and service interventions may produce. These constructs are discussed in detail in Chapter 3 and displayed in Exhibit 3-3.

[5] This data collection has been funded by the National Institute of Child Health and Human Development (NICHD).

Chapter 2. Data Collection

The analysis of impacts for the Family Options Study will rely on four types of data:

- Surveys of participating families at several points;
- Administrative records from homeless assistance providers and HUD;
- Random assignment records established at enrollment;
- Information about the intervention programs (structure, rules, characteristics of the housing and services interventions provided to families, and program costs)

Surveys of participating families were conducted at baseline, prior to random assignment, and will be conducted at 6, 12, and 18 months after random assignment. The interviews conducted 6 and 12 months after random assignment are brief tracking interviews that collect updated contact information and information about family composition and housing situation. The study will also use two primary sources of administrative data: (1) Homeless Management Information System (HMIS) records maintained by homeless assistance service providers; and (2) HUD records from the Public Housing Information Center (PIC) system and the Tenant Rental Certification System (TRACS). The research team used a secure website to collect identifying information and to conduct random assignment. The random assignment database contains a unique random assignment record for each participating family. The study is also conducting on-site interviews with program staff in order to describe the interventions in detail. These interviews focus on program structures and rules and on the characteristics of the housing assistance and services provided by the programs that are included in each treatment arm at each site. The team will also collect data from program staff to measure the costs of serving families.

The research team will use these data to address the research questions identified in Chapter 1 and to test the hypotheses discussed later in Chapter 3. This chapter provides an overview of each of these data sources, including a description of the key analytic variables derived from each and the timeline and procedures for data collection. Chapter 3 includes examples of analytic tables based on the data elements from these data sources.

2.1 Participant Surveys

The evaluation team will administer surveys to study subjects at several points. The purpose of each survey, its timing, and the data collection process are summarized in Exhibit 2-1. Exhibit 2-2 provides information about the content of the survey instruments and the measures to be obtained from it.

Exhibit 2-1: Role of Survey Data

Data Source	Purpose	Collection Process
Baseline Survey	• Characteristics of the study sample • Covariates to use in impact analysis to improve precision of estimates • Subgroups for impact analysis • Contact information to maintain locating information for sample	• In-person 40-minute interview prior to random assignment • Completed for the full sample randomly assigned
6- and 12-Month Tracking Interviews	• Updated contact information • Information about changes to family composition • Information on housing location and type • Data from the tracking interviews are used to: – Measure outcomes (family preservation and housing stability) – Assess take-up of assigned intervention – Assess crossover	Telephone interview (10 minutes) conducted 6 and 12 months after random assignment
18-Month Follow-up Survey	• Information from adult respondent about housing history and stability; housing quality and affordability; employment, self-sufficiency, and hardship; family composition; adult health and well-being; child well-being; and receipt of services • Data from the 18-month follow-up survey are used to: – Measure outcomes in the five domains (housing stability; family preservation; self-sufficiency; adult well-being; child well-being)	In–person interview lasting 60 minutes, conducted 18 months after random assignment*

*Observation of younger children and interviews with older children will be conducted at the same time as the in-person interview with the adult respondent.

Exhibit 2-2: Content of Participant Surveys

Domain/Construct	Timing			Use in Analysis		
	Baseline	Tracking	Follow-up	Covariate	Outcome	Describe the Sample
Housing History and Stability						
Pre-shelter housing	✓			✓		✓
History of homelessness	✓			✓		✓
Historical indicators of vulnerability (foster care as child, criminal justice history, exposure to trauma, social supports)	✓			✓		✓
Barriers to obtaining housing	✓			✓		✓
Homelessness since Random Assignment			✓		✓	
Number of moves		✓	✓		✓	
Receipt of housing assistance		✓	✓			
Move in to assigned intervention		✓	✓			
Housing Quality, Affordability						
Housing quality			✓		✓	
Housing crowding			✓		✓	
Housing affordability			✓		✓	
Employment, Self-sufficiency, Hardship						
Current status (type, hours, earnings)	✓		✓	✓	✓	✓
Work history (number of jobs since random assignment)			✓		✓	
Education and training	✓		✓	✓	✓	✓
Household income/sources	✓		✓	✓	✓	✓
Economic stressors			✓		✓	
Food security			✓		✓	

Domain/Construct	Timing			Use in Analysis		
	Baseline	Tracking	Follow-up	Covariate	Outcome	Describe the Sample
Family Preservation						
Household/family composition	✓	✓	✓	✓		✓
Family separation and reunification			✓		✓	
Adult Well-being						
Health status (mother only)	✓		✓	✓	✓	✓
Mental health	✓		✓	✓	✓	✓
PTSD symptoms	✓		✓	✓	✓	✓
Maternal depression	✓			✓		✓
Substance use	✓		✓	✓	✓	✓
Domestic violence	✓		✓			✓
Criminal justice involvement	✓				✓	
Child Well-being						
School attendance			✓		✓	
School mobility			✓		✓	
Grade completion			✓		✓	
Health status			✓		✓	
Source of primary health care			✓		✓	
Family routines and home environment			✓		✓	
Behavior			✓		✓	
Parenting			✓		✓	

Domain/Construct	Timing			Use in Analysis		
	Baseline	Tracking	Follow-up	Covariate	Outcome	Describe the Sample
Demographic Characteristics						
Gender (adults and children)	✓	✓*		✓		✓
Age (adults and children)	✓	✓*		✓		✓
Race/ethnicity	✓			✓		✓
Contact information						
Contact information for up to three people who will always know how to locate respondent; confirmation of respondent's current location.	✓	✓	✓			
Service Receipt						
Service use			✓			
Relationship with service providers			✓			

*New household members

Baseline Survey

The baseline survey data are used for information on background characteristics of study participants. The baseline survey variables include housing history, homelessness, barriers to obtaining housing, employment status at baseline, family composition, income and income sources, adult physical health, adult behavioral health, substance use, demographic characteristics, and contact information. These detailed data will be used to describe the families in the study sample, form subgroups for separate analysis, provide covariates in the impact analyses, help to adjust for non-response on the follow-up surveys, and provide contact information for participant tracking and locating for the follow-up survey. All family heads who agreed to participate in the study completed a baseline survey during the enrollment session conducted at the emergency shelter, prior to random assignment. The baseline survey was conducted in person using Computer-Assisted Personal Interviewing (CAPI) software.

Tracking Surveys

The tracking surveys are administered 6 and 12 months after random assignment. The primary purpose of these short, 10-minute interviews is to update contact information for the respondent, including secondary contacts. In addition, the surveys collect information on current housing status, type of housing, and family composition. Information on housing status will be used to construct measures of housing stability, and information on the type of housing will provide additional information beyond what is obtained from program providers to assess take-up of the assigned intervention and crossover. Information on household composition will be used to construct measures of family preservation. The tracking interviews are conducted by phone.

18-Month Survey

The 18-month survey will collect information on a range of outcomes to be measured over the follow-up period (for many outcomes the focus is six months prior to the interview). Outcome measures constructed from data collected in the follow-up survey will permit comparisons of the outcomes of the four interventions included in the study. The instrument contains items related to housing history and stability; housing quality and affordability of the current housing unit; employment income, self-sufficiency, and economic hardship; family preservation; adult well-being; child well-being; and receipt of services. The follow-up survey includes a parent-on-child module in which the parent reports on characteristics of a focal child who was present at baseline with the parent to measure child well-being outcomes. At the same time as the interview with the adult informant, survey staff will observe younger children and interview older children.[6] The 18-month survey is designed to take 60 minutes to administer and will be administered in person. The next section discusses the process and timing of follow-up survey data collection in more detail and steps to ensure the highest possible response rate.

Follow-up Survey Data Collection

This section describes the process for collecting follow-up survey data in the 18-month interview. The section includes a description of CAPI; interviewer recruitment and training; child data collection activities; monitoring and verification; and the data collection period.

---

6 These direct child observations and parent-on-child modules for additional children have been funded by NICHD.

Computer-Assisted Personal Interviewing (CAPI). The project team will administer the survey using CAPI software. Using CAPI reduces errors in survey administration because skip patterns are automated so the interviewers do not need to follow complex skip patterns manually. Automated interview technology also prevents entry of invalid answers and prompts interviewers to validate or confirm critical data items such as income, earnings amounts, and receipt of housing assistance.

After receiving Office of Management and Budget (OMB) approval, the survey team will program the 18-month follow-up survey instrument in CAPI. Trained CAPI testers will rigorously test the CAPI instrument to ensure that response codes are programmed correctly and that skip patterns are followed as intended. The final programming will be tested by the project staff who will check that all text substitutions are in place, all ranges established, and all skip logic and question routing have been checked. Finally, using simulated data, we will create a set of frequencies that will be thoroughly examined for any remaining errors in the programming logic. This rigorous testing of the CAPI programming ensures that the survey data will be collected consistently. The CAPI software is also integrated with participant tracking data to easily facilitate sharing of updated contacting and locating data in and out of the CAPI system.

Interviewer Recruitment and Training. The follow-up survey will be conducted in person. Leading our team of highly skilled interviewers are the site interviewers who conducted the baseline and tracking efforts for the study. The current site interviewers will be supported by additional skilled interviewers trained in tracking and locating procedures for highly transient populations.

To ensure that all of the participant-level data are collected consistently, interviewers will undergo extensive training. Training will cover the following topics:

- Study background
- Purpose of the follow-up data collection
- Renewing the informed consent and obtaining parental permission and child assent
- Survey administration protocols
  - Adult interview
  - Child interview (ages 8–17)
- Tracking and locating procedures and documentation
- Data security

Further, interviewers will receive comprehensive training on direct child assessments for the child data collection discussed below. Each assessment will have a separate training protocol. Interviewers will be trained on the objective and protocols for each individual assessment. They will observe a mock assessment in progress. Finally, they will be asked to conduct a mock assessment as part of the certification process. Those who conduct the mock assessments proficiently will be certified.

Data Collection. We will conduct the follow-up survey over a 16-month field period beginning in June 2012 and continuing through the end of September 2013. Because the random assignment period occurred between September 2010 and January 2012, interviews will take place 18-20 months after random assignment. The data collection period is long, and that allows opportunity for study procedures to lapse a bit for interviewers. To control for drift lapses, we will conduct quality control monitoring periodically over the 16-month field period to ensure that the study protocols remain fresh in the minds of all interviewers.

Exhibit 2-3: Duration of Period Between Random Assignment and Follow-Up Survey Interview, By Intake Cohort

Intake Cohort/Date of Random Assignment	Survey Field Date/Duration of Follow-Up Period
September 2010	June 1, 2012/ 20+ months
October 2010	June 1, 2012/ 19+ months
November 2010	June 1, 2012/ 18+ months
…	18+ months
November 2011	June 1, 2013/ 18+ months
December 2011	July 1, 2013/ 18+ months
January 2012	August 1, 2013/ 18+ months

Each month, up until the final two months of the survey field period, we will release one monthly enrollment cohort for interviewing. The approach allows for a minimum of eight weeks to work the cases in the field, so each month new cases will be added as old ones are still being worked. As shown in Exhibits 2-3 and 2-4, all enrollment cohorts will be released for interviewing no earlier than the 18th month after random assignment. This ensures that interviews will first be attempted with all families in the study at least 18 months after their random assignment.

In considering the potential duration of survey follow-up for each of the study's 17 monthly intake cohorts, only six of these cohorts are displayed in Exhibit 2-3—the first three and the last three—since the other 11 cohorts all have "18+ months" as their potential duration of follow-up. Note that "18+ months" means that interviewing will take place at the earliest on the first day following the 18-month anniversary of random assignment (e.g., a June 1, 2012 interview for someone randomly assigned on November 30, 2010). That is because participants in each monthly interview cohort are released for interviewing after the end of the 18th month following random assignment.

Exhibit 2-3 displays the cohorts whose duration of follow-up, if interviewed immediately after being released for follow-up interviewing, is a full 19 or more months past random assignment (top two rows). Because these special cases are limited to the first two monthly cohorts of families randomly assigned, for the vast majority of the evaluation sample the potential duration of follow-up is the same. Exhibit 2-4 displays the interview period for each intake cohort graphically.

Given the study's experimental design, it is imperative to track production (the results of efforts to administer surveys) by intervention group (i.e., treatment vs. control) over time in order to ensure that all random assignment sample group members respond to the survey data collection at comparable rates. Varying response rates across the groups could bias the results and jeopardize the ability to compare outcomes across those groups.

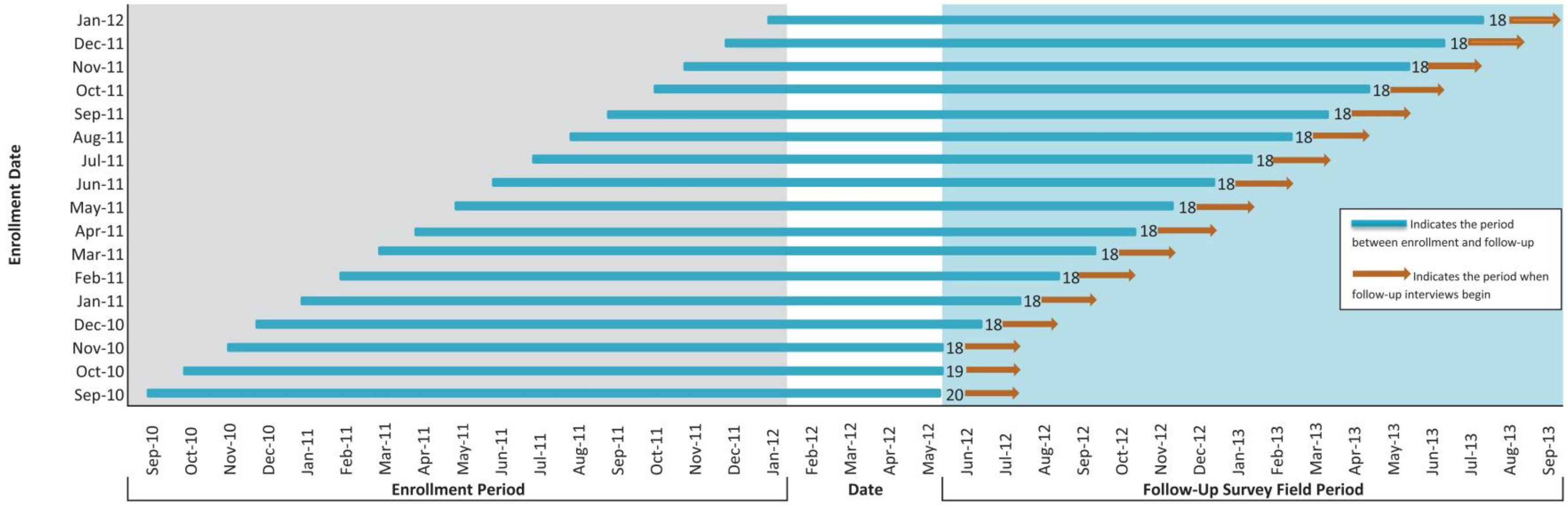
To ensure that response rates do not vary too much across groups, we will monitor production reports carefully. We will provide HUD with monthly reports to show the progress of the follow-up survey and the tracking efforts, by random assignment cohort and by site. Our survey budget is based on an overall response rate of 75 percent. If it is possible to exceed this target within the resources available, we will do so.

Information on families who were lost to follow-up, including their site, assigned intervention, and number of attempts to contact the family will be recorded. This information will be used to analyze patterns of non-response.

Child Data Collection. NICHD awarded a grant to the research team to expand the analysis of child outcomes. In addition to completing the 18-month survey with the adult sample member, interviewers will also attempt to collect additional data from and about up to two children in the family. First, the parent-on-child module in the adult survey will be expanded to collect information about a second child. Interviewers will also conduct a series of assessments with a sample of children ages 3–7. Interviewers will also conduct assessments and a child interview with a sample of children ages 8–17. The child data collection components will be conducted in person at the same time as the adult interview.
Wherever possible, we will conduct the child assessment and/or interview data collection during the same appointment as the adult/parent. It is likely that in some instances we may need to schedule follow-up visits for the child components due to scheduling conflicts. Regardless of when we collect the child data, we will follow the same protocol.

1. *Informed Consent & Parental Permission.* First, interviewers will explain the purpose of the adult interview and seek renewed consent from the adult sample member to continue his/her participation in the study. The interviewer will then explain the child data collection components to the adult sample member and seek the parental permission to proceed with the child data collection.
2. *Completion of the Adult 18-Month Survey.* Once all required permission and consents are obtained, the interviewer will administer the adult survey module, and the parent-on-child module.
3. *Child Assent.* Once the adult survey is completed, the interviewer will sit down with the child(ren) to explain the study and review what it is that they want the child to do. Children will be allowed to ask any questions and then asked if they would like to participate.
4. *Child Assessment and/or Interviews.* Once a child gives his/her assent to participate, the interviewer will prepare to do the child data collection.

Exhibit 2-4.
Enrollment Dates and Corresponding Follow-Up Survey Field Periods



Notes

Enrollment Period

Enrollment began in September 2010 with six [6] families in this cohort

The last family is expected to be enrolled in January, 2012

Follow-Up Survey Field Period

The design assumes a 16-month field period beginning in June, 2012 and ending in September, 2013

The period between random assignment and release (for follow-up) is at 18+ month for all cohorts, with the exception of the first two--September, 2010 and October, 2010--where the spread is 20+ and 19+ months, respectively

Abt SRBI has determined through prior experience that it will be necessary to allow eight (8) weeks to work the sample to completion

In order to maintain a 16-month follow-up field period (and to allow sufficient time to work the sample to completion), the first three (3) cohorts (September, October, and November, 2010) will be released in the same month--June,2012; the durations from enrollment to release for these three cohorts will be 20+, 19+, and 18+ months, respectively; on the other end, the time period to complete the interview for the final cohort to be enrolled (January,2012), will be compressed from the normal eight weeks to four weeks

Child data collection is different than survey interviewing with an adult. Children can be more cautious about the process or they may be uncomfortable with a stranger. Further, parents may be uneasy with an interviewer alone with their child. Thus, our protocol is designed to ensure that the parent, the child, and the interviewer can all be at ease with the following precautions:

- No child data collection will be conducted without a parent or guardian being home at the same time.
- Interviewers will ask a parent to remain present in the room with children—but out of the child's sight line--during the child assessments with children ages 3–7. We do not want our interviewers to be alone with children at any time during the data collection. Having a parent present, but out of the line of vision, ensures that the parent and child are both comfortable with the interviewer's presence, but the parent cannot influence the child's behavior.
- During the child interview with 8–17 year olds, we will ask parents to be out of the room and out of earshot so that the youth feel comfortable answering the questions honestly. However, we will make every effort to set up the interview in an area that allows the parents to visually see the child and interviewer, even though they are out of earshot.
- In order to ensure confidentiality:
  - Parents will be informed that they will not receive copies of their child's responses
  - Children will be reminded that their parents will not see or hear their responses to the survey

2.2 Administrative Data

To supplement the outcome measures that will be derived from the survey data, we will collect three types of administrative data: data from the HMIS of each study site and data from HUD's PIC and TRACS systems on receipt of housing assistance, as shown in Exhibit 2-5.

Homeless Management Information System (HMIS) Data

An HMIS[7] is an electronic data collection system that stores longitudinal, person-level information about persons who access the homeless services system in a Continuum of Care (CoC). HMIS is a valuable resource because of its capacity to integrate and unduplicate data from all homeless assistance and homelessness prevention programs in a CoC. While HMIS are locally administered systems, all use universal data standards for variable definition and file content, making the data extremely well-suited for analysis of programs across a range of communities. Local homeless assistance providers collect data from program participants and submit the data to the HMIS system in the CoC.

[7] See www.HMIS.info,

Exhibit 2-5: Role of Administrative Data

Data Source	Purpose	Collection process
Homeless Management Information System (HMIS)	• Measure participation of study sample in homeless assistance programs covered in HMIS • Used in the study to: – Measure return to shelter, one component of housing stability outcome – Assess take-up of assigned intervention – Assess crossover – Verify start/end dates for participation in assigned intervention	• Individual-level data collected from community-level administrators of HMIS – Negotiate data sharing agreements (one or more per site) – Ask local HMIS administrators to match study sample against system and return matches • Variables: – Program ID – Entry date – Exit date – Person ID – Household ID
HUD Public Housing Information Center (PIC)	• Measure receipt of housing assistance through the Housing Choice Voucher program or Public Housing • Used in the study to: – Assess take-up of assigned intervention – Assess crossover	• Individual-level data collected from HUD (PD&R). HUD will match the study sample identifiers against extracts from PIC and provide data from HUD 50058 for matches – Establish a timeframe for extracting PIC data and specifications • Variables: – Receipt of subsidy – Dates of subsidy receipt – Other variables from 50058 form
HUD Tenant Rental Certification System (TRACS)	• Measure receipt of housing assistance through project-based assistance • Used in the study to: – Assess take-up of assigned intervention – Assess crossover	• Individual-level data collected from HUD (PD&R). HUD will match the study sample identifiers against extracts from PIC and provide data from HUD 50059 for matches – Establish a timeframe for extracting TRACS data and specifications • Key variables: – Receipt of subsidy – Dates of subsidy receipt

Given the universal data standards in place for HMIS and what appears to be quite complete use of HMIS in the study sites, we believe that HMIS is a valuable source of data for the Family Options Study. HMIS will serve three main purposes for the study:

- Using program entry and exit dates recorded in HMIS, we can measure the length of time study families spend in the assigned PBTH or CBRR intervention and the time that families assigned to UC remain in shelter. We will cross-check this with information collected in the follow-up survey about time spent in the assigned intervention.
- Using information on entry into emergency shelter, HMIS provides an additional source of data (in addition to the follow-up survey) to ***measure housing stability, the key outcome domain for the study***.
- By matching the study participants with HMIS, we ***can measure crossover—****which occurs if a study participant receives an intervention to which she was not assigned***.** HMIS will facilitate this measurement because it will identify participants assigned to other interventions who receive assistance from PBTH or CBRR.

Using HMIS data will require negotiating agreements with local HMIS administrators in the 12 study sites. While collection of HMIS data from all sites would be ideal, we believe that HMIS data will still be valuable to the study, even if we are only successful in completing data-sharing agreements and accessing data in a subset of study sites. For other study sites, lengths of stay, return to homelessness, and crossover into PBTH and CBRR will be measured based only on survey data. We can use sites for which we have both types of data to assess the quality of the survey data relative to the HMIS data.

The study will require the examination of individual-level HMIS records found by matching identifiers for the study sample to the HMIS in the communities. Abt Associates will provide a file with identifiers of the family heads in the study sample, and the local HMIS administrator will match to the system and return data for all individuals found in the system.

An important factor affecting the viability of HMIS for the study in any particular site is the coverage of HMIS—that is, what proportion of the community's residential programs for homeless people participates in the HMIS and the extent to which those programs consistently enter data into the HMIS. Based on information reported by communities to HUD in spring 2011, downloaded from the Homelessness Data Exchange, we believe HMIS coverage in the study sites is quite high. Exhibit 2-6 shows the HMIS participation rates for the CoC as a whole, which would support the crossover and housing stability analysis. The exhibit also shows participation for the emergency shelter and transitional programs in the study, which would support the length of participation analysis. HMIS bed participation refers to the percentage of beds in the communities that are covered in the HMIS. Thus, for example, data on clients staying in 83 percent of the beds in emergency shelters in Alameda County that are participating in the study are included in HMIS.

The Abt team will contact the HMIS administrator at each site to confirm the level of participation of local programs. We will also assess the process for gaining local approval to secure individual-level data. We believe that the process for securing approval will be relatively straightforward, since we have informed consent from study participants. However, the process may be time-consuming, and we may need to work with the HMIS administrator to develop approaches for extracting relevant client data from the community's system. If securing the HMIS data appears valuable for a given site, then we will enter into a data-sharing agreement with the site and provide technical specifications and support to the site to extract the data. We will extract data from HMIS once, at the end of the follow-up period prior to analysis of the follow-up survey data. Once extracted, we will append the HMIS data to the other study data available for each participant.

Exhibit 2-6. HMIS Participation in the Study Sites

	HMIS Bed Participation Rates for All Providers in the CoC		HMIS Bed Participation Rates for Providers in the Study	
Study Site	***ES-FAM***	***TH-FAM***	***ES-FAM***	***TH-FAM***
Alameda County	53%	93%	83%	93%
Atlanta	86%	87%	100%	85%
Boston	91%	96%	86%	NA
Connecticut[a]	94–100%	78–100%	100%	100%
Denver	100%	100%	100%	100%
Honolulu	96%	92%	100%	100%
Kansas City	100%	100%	100%	100%
Minneapolis	84%	81%	100%	68%
Phoenix	96%	95%	100%	89%
Salt Lake City	100%	96%	100%	96%
Baltimore	88%	98%	100%	98%
Louisville	80%	100%	100%	100%

[a] This study site is comprised of four CoCs in the New Haven/Bridgeport, CT, area; therefore, the figures reported for CoC coverage represents the range of coverage levels in these four CoCs.

HUD PIC Data

Administrative data from HUD's PIC system provides an important source of information on receipt of housing assistance through HUD's Housing Choice Voucher and Public Housing programs. These PIC data will be used to measure take-up of assigned intervention for those assigned to SUB, and to measure crossover by families assigned to other interventions who receive housing assistance. We will work closely with HUD PD&R to obtain extracts from the ***HUD PIH Information Center (PIC)*** during the post-random assignment follow-up period to measure receipt of housing assistance for a one- or two-year period preceding random assignment through at least the 18-month follow-up survey. We will coordinate closely with HUD staff to determine the precise timing of the data requests and the number of extracts that will be necessary to provide data covering the period desired. As has been the practice with other studies, we have planned that HUD staff will assist us in obtaining PIC extracts for the families in the study by matching PIC extracts to the identifiers of the families who enroll in the study. Abt's analysts and programmers are familiar with PIC data and have used these data extensively to measure receipt of housing assistance for many studies, including the Moving to Opportunity Interim Evaluation and the Effects of Housing Vouchers on Welfare Families.

We recommend collecting a test file after the follow-up survey begins. The first extract will be used to review PIC file contents, layout, and match rate for the study sample and to test the approach for constructing the analytic variables to measure receipt of housing assistance. Depending on guidance from

HUD regarding timing of data availability in PIC extracts, we will develop a plan for subsequent data collection extracts to provide PIC data for all participants through the 18-month follow-up period.

HUD TRACS Data

HUD also provides assistance to families through contracts with property owners under the project-based Section 8 programs administered by the Office of Housing. Information regarding households living in Section 8 projects is reported to HUD by property owners on HUD form 50059 and is maintained by HUD in TRACS. Because some members of the study may have received assistance in those programs during the follow-up period, we also will collect (approximately three) extracts of TRACS data from HUD and use it to identify families who had received project-based assistance in order to measure receipt of housing assistance. We propose the same schedule for TRACS as described above for PIC but will discuss the schedule with HUD experts to ensure the timing of extracts meets the needs of the analysis in the most efficient way.

2.3 Random Assignment Data

The research team developed a secure website to collect identifying information and to conduct random assignment. The website contains these random assignment records as well as information about the extent to which families enrolled in the assigned intervention program. The random assignment record includes three key identifiers for the family—name of the adult, SSN, and date of birth. The random assignment record also includes date of random assignment, random assignment result, and name of program to which the family was referred. The random assignment process also generates a unique family identifier for each enrolled family and this is used to link the random assignment record to the baseline survey data. The random assignment website is also used to record information about whether the family enrolled in the assigned intervention and the date of enrollment.

Data collected in the random assignment record will be linked with participant surveys and administrative data. The primary purpose of the random assignment record is to provide unique identifiers for each participant, and information on date and result of random assignment and date of take-up of the assigned intervention.

2.4 Program Data

Interviews with Program Staff

The research team is also conducting interviews with program staff in the emergency shelters in the study sites and the housing and service providers that serve study participants, in order to collect information about the intervention programs in the study sites. These program data will be used to determine the extent to which programs represent the intended treatment conditions and to understand the nature, quality, and costs of services provided to participants in the study. The data will be used for descriptive assessments of the interventions and used to interpret impact findings—providing information about what the interventions actually provided to study participants.

The research team is collecting program-level information during site visits to each participating site to interview program managers and staff. The team also reviews program background documents and materials to obtain this information. We collect program-level data during the latter months of enrollment and after enrollment ends but while study participants are receiving the designated housing and services interventions.

The program data collection includes general information about each program such as program budget, staffing (FTE), type of services offered, and arrangements used to provide services. We also collect information about the ***housing assistance*** and the ***services*** provided in the study interventions. For each of the services offered, the research team is documenting sources of funding, costs of providing the service, description of the service, and measures of quality. The team is examining the following program dimensions:

- **Housing Duration.** Is the housing provided time-limited, and if so to what duration? What are average stays in housing?
- **Housing Structure.** Is the housing provided primarily congregate, scattered-site, a mixture?
- **Participant Selectivity.** What are the program eligibility requirements? What is the population that the program targets? How strict are the requirements?
- **Nature of Services.** What approach/model does the program use? What does case management include? Are services provided on site or through referrals to providers in the community?
- **Service Intensity.** How intense are the services provided? How many clients do social service staff work with? What is the range of services that are available?
- **Service Quality.** Dimensions of quality include the use of assessments, family involvement in identifying needed services, and qualifications of staff.
- **Staff Attitudes.** To what extent do staff see families as having rights to make their own choices or as needing guidance to make wise decisions?
- **Program Restrictiveness.** To what extent does failure to cooperate with treatment plans or deviant behavior lead to exclusion from the program?

For each program included in the interventions in each site, we will also collect cost and funding data. We will use this cost and funding data to assess the cost-effectiveness of the study interventions. During initial site visits, the research team will document the sources of cost data for each program and gather basic information about the manner in which financial records are maintained and where key cost items are documented. In later site visits, the study team will collect the cost data.

2.5 Cost Data

Chapter 3 discusses the impact analysis that will provide estimation of the causal impact of each of the four interventions, relative to each other and to UC. Estimates of causal impacts are only one input into decisions about homeless policy. The interventions are likely to vary in cost. Housing assistance provided in PBTH and CBRR is temporary; subsidies in the SUB intervention will be indefinite (as long as regular eligibility requirements for the subsidy are maintained). Interventions without services will usually require fewer resources than those offering social services. Thus, if we were to conclude that there is little difference in outcomes across a pair of interventions, this might suggest choosing the lower-cost intervention. Conversely, if we were to find that an intervention had both better outcomes and larger costs, then we would need to consider whether the better outcomes were commensurate with the higher costs.

The study design does not include a full cost-benefit analysis with monetized benefits. However, the study will analyze the costs to permit a comparison of costs per family across the different interventions. This will allow us to evaluate the intervention impacts in the context of cost differences.

To collect cost data, the research team will review financial records and meet with program staff to document program budget information on: (1) housing subsidy amounts (including operating costs of facility or project-based programs, where relevant); (2) supportive service provision (if relevant to the intervention); and (3) program administration. To derive appropriate daily unit cost estimates, we will also collect general program information, such as program unit capacity, typical occupancy/enrollment rates, service types, and definition of a service unit.

The research team will collect cost data from each of the programs operating each study intervention in each of the sites, including the shelters providing UC, the PHAs providing the SUB intervention, the agencies administering CBRR, and the providers operating PBTH. Cost data will include the costs of providing the housing assistance or shelter and any non-housing services provided to each family in the study.

Unit costs such as costs per family per month or costs per family per day will be multiplied by the length of time the family received the housing assistance or the services.

Measuring Costs of Housing Assistance per Family

Exhibit 2-7 shows the data to be used for measuring the costs of housing assistance in the SUB, CBRR, and PBTH interventions. The monthly cost per family of the voucher will be derived from the HUD Form 50058 for each family and from PHA data collected during the site visits on the average administrative fee per family. The operating and capital costs of the average public housing unit (if possible, the average family unit) will come from the records of the Honolulu PHA.

The cost of CBRR will either be the actual cost per family of the rental assistance provided or the average monthly cost of CBRR assistance provided by the HPRP agency, depending on how the agency keeps its records.

For PBTH, we will gather information from the financial records of the sponsor of the transitional housing on the costs of program administration; housing operations staff (e.g., property managers or in-house maintenance staff); facility operating costs (e.g., staff costs and contracted services costs associated with debt service, utilities, management, maintenance, insurance); and operating reserves. Cost data will also include capital costs of properties owned and managed by the program and leasing costs for properties that the program does not own. For capital costs, the team will collect information on the property (address, year built, type and size of facility, type of construction, unit size, property tax assessment appraised value) to calculate a daily unit capital value for the property. The sources of cost data will include program budgets, audited cost reports, or other financial records the program typically uses to report the unit costs associated with providing housing assistance to participating families and amounts of household rents. We transform the estimates of both operating and capital costs into a cost per day, since stays in transitional housing will be measurable on a daily basis.

Exhibit 2-7: Sources of Cost Data

	Data for Unit Cost	Data for Duration for Each Sample Family	Cost Calculation for Each Sample Family
Housing Assistance or Shelter Cost			
SUB Housing Choice Voucher†	Rent (subsidy plus tenant rent) per family per month for sample families from PIC data, form 50058 Average administrative fee per month from PHA data	PIC data Follow-up survey is back-up if 50058 is missing	(Rent per month + admin fee per month) * number of months using voucher = cost per family
SUB Public Housing††	Operating cost per unit per month from PHA data Capital cost per unit per year from PHA data	PIC data Follow-up survey is back-up if 50058 is missing	(Operating cost per month + capital cost per unit per year/12)*number of months in public housing = cost per family
CBRR	Average housing assistance per month per family or actual assistance for each sample family from HPRP administering agency	Program records of HPRP administering agency—collected from HMIS Follow-up survey is back-up if HMIS records for the agency are not available	Either average housing assistance per month * number of months assisted = cost per family or Actual cost of housing assistance per family
PBTH	Operating cost per year from transitional housing provider financial records Capital costs per year from provider financial records	Program records of TH administering agency—collected from HMIS Follow-up survey is back-up if HMIS records for the agency are not available	[(Operating cost per year +capital cost per year)/365]*number of days in program = cost per family
UC	Operating cost per year from transitional housing provider financial records Capital costs per year from provider financial records	Program records of TH administering agency—collected from HMIS Follow-up survey is back-up if HMIS records for the agency are not available	[(Operating cost per year +capital cost per year)/365]*number of days in program = cost per family
Service Cost			
Services associated with SUB	Data from PHA on unit cost of any services not covered by the administrative fee	One-time costs, duration not relevant	Average cost per family
Services associated with CBRR	Data from the HPRP agency on the unit cost of service coordination or other services	Data from the HPRP agency on average number of units per family	Average unit cost per family *average number of units
Services associated with PBTH	Data on service costs from financial records of providers, turned into a unit cost	Data from the PBTH provider on average number of units per family	Average unit cost per family *average number of units
Services associated with UC	Data on service costs from financial records of providers, turned into a unit cost	Data from the shelter provider on average number of units per family	Average unit cost per family *average number of units

†The cost for the project-based vouchers used in Bridgeport will be derived the same way as Housing Choice Vouchers in other sites.

††Will discuss with Honolulu PHA whether can adjust for larger average unit size of family housing.

We anticipate that the most common housing assistance received by families assigned to UC will be continued residence in emergency shelter. For the emergency shelters in each site, we will calculate the daily cost per family of serving study participants. For emergency shelters, the cost components will be similar to those in PBTH (program administration; housing operations staff; leasing costs (if the emergency shelter property is not owned and managed by the program); facility operating costs; and capital costs (if the emergency shelter is owned and managed by the program). As with PBTH, the anticipated sources of these cost data will include program budgets, audited cost reports, or other financial records the program typically uses to report the unit costs associated with providing housing assistance to participating families. As with PBTH, we will turn these estimates into a cost per day.

Duration of receipt of SUB (the voucher or the public housing unit) will be taken from the HUD Form 50058 for each family. Duration of receipt of CBRR will come from the records of the agency administering the HPRP program as recorded in HMIS in each community. Duration of stays in PBTH will come from the program records of the participating programs, collected from data reported to the community's HMIS, which will show entry and exit dates. Duration of continued shelter stays for the UC treatment arm will also come from agency records from data entered into the HMIS. For all four interventions, we will have back-up information on duration of the use of the program from the 18-month follow-up survey if administrative records are missing for the family.

Measuring Costs of Services per Family

Based on information gathered to date from the participating programs, we anticipate that participants assigned to each intervention will receive some services, although the PBTH intervention is intended to provide the most intensive supportive services to families in the study.

In each study site, we will collect service cost data for each program associated with each intervention. For the SUB intervention, we will be careful to distinguish this cost from any housing-related services covered by the Housing Choice Voucher administrative fee.

We have started this process by documenting which services each program provides to participants in the study, using the program data collection underway in the study sites.

Service cost data will include staff costs and program operating expenses such as facility rent for program staff, staff transportation, and supplies. We will also collect information on costs of direct client support, if any, associated with each of the services (e.g., food, clothing, furniture, transportation assistance, and other client assistance). For each service, we will work with program staff to develop an appropriate unit of service that can be used to calculate unit costs of providing each type of service to study participants.

Cost data will be collected using a standard protocol that we will adapt from the cost data collection tools used in previous studies like the Costs of Homelessness Study. The cost data will be collected for any services received that are directly related to a program to which the family was randomly assigned or to programs in which the families assigned to UC participate in during the period between random assignment and the end of the follow-up period.

Receipt of non-housing services will be measured from program participants in the follow-up survey, but we will not ask families about the duration of service receipt, which may or may not be the same as the duration of the housing or residential assistance. Instead, we will ask program staff for the typical durations of each service if they are not the same as the durations of the housing or residential assistance and apply these quantities to the unit costs of those services.

The period of receipt of services may or may not coincide with the period of the receipt of the housing assistance provided by a program. When direct measures of the duration of services are not available from program records, information on typical exposure to services collected from interviews with program staff will be used to approximate duration of service receipt. Data collection will occur close to the end of the study follow-up period, when financial records for the time period of interest are complete.

2.6 Database Development

Construction of the Evaluation Data Base

In preparation for data analysis, the study team will construct an analysis data file from the random assignment, tracking, survey, and administrative data collected for study families. The first step will be to build family composition histories from the detailed household rosters collected at baseline, tracking, and follow-up interviews. This will involve a large number of variables and up to four observations for each family in the study, in order to track changes in family composition as individuals leave and join the family over time. Once families are defined longitudinally, we will need to construct their housing/homelessness history throughout the follow-up period. Other outcomes from the survey data concerning adult and child well-being will be aligned across respondents and over time.
We will follow a similar process for data file development to that used on prior HUD random assignment evaluations such as the Effects of Housing Vouchers on Welfare Families. This includes identifier matching across multiple sources (the family ID established at random assignment), variable construction along a continuous timeline as well as for fixed reference periods or point-in-time "snapshot" measures, and cross-validation of measures available from multiple sources.

Chapter 3. Analysis Plan

This chapter describes the analysis plan for the impact evaluation. The chapter begins with a discussion of the hypotheses that will be tested. Then we present a full list of outcomes that will be examined for impacts. In Section 3.3, we discuss the impact estimation model and various details of the estimation including our plans for weighting observations and for handling missing data. In Section 3.4, we discuss our planned approach to addressing the multiple comparisons problem. Then (Section 3.5) we describe how no-shows and crossovers (noncompliance with assignment) will affect our impact estimates and our plans for addressing noncompliance with assignment in the analysis. Sections 3.6 and 3.7 discuss plans for examining how baseline family characteristics and program design features moderate the size of intervention impacts. Section 3.8 describes our plans for analyzing intervention costs and Section 3.9 shows sample sizes of enrolled study participants and the minimum detectable effects that these samples imply.

3.1 Hypotheses

The research questions stated in Chapter 1 focus on the relative effectiveness of the study interventions in the five outcome domains of housing stability, self-sufficiency, family preservation, adult well-being, and child well-being. The design of the Family Options Study allows for six pairwise comparisons among the four assignment groups in the study. As shown in Chapter 1, Exhibit 1-3, the six pairwise comparisons are:

1. What is the impact of PBTH assistance for homeless families compared to the Usual Care offered in the community?
2. What is the impact of SUB assistance for homeless families compared to the Usual Care offered in the community?
3. What is the impact of CBRR assistance for homeless families compared to the Usual Care offered in the community?
4. What is the impact of PBTH assistance for homeless families compared to SUB assistance?
5. What is the impact of PBTH assistance for homeless families compared to CBRR assistance?
6. What is the impact of SUB assistance for homeless families compared to CBRR assistance?

Formally, each pairwise comparison will test the hypothesis that assignment to the first group produces outcomes equivalent to assignment to the second group. The alternative hypothesis for each comparison, that assignment to the first group produces different outcomes from assignment to the second group, is based on differences in the assistance provided by each intervention. At a basic level, the four interventions can be characterized as combinations of types of housing subsidies and types of services, as shown in Exhibit 3-1.

Exhibit 3-1: Cross-Classification of Subsidy and Service Options for the Family Options Interventions and Usual Care Group

Housing Subsidy	Homeless Service System Services Provided		
	Some*		None
Some	Heavy	Light	
Deep, permanent	----	----	**SUB**
Deep, temporary	**PBTH**	**CBRR**	----
None	**----**	**----**	**UC****

Notes:
* Besides the heavy/light distinction, the *nature* of services also may differ between PBTH and CBRR. CBRR focuses on services to help with locating housing, lease up, and settling in. In contrast, PBTH provides more intensive social services such as assessments, provision of and referral to job-related services, counseling, substance abuse treatment, and family- and child-oriented services.

** UC in many instances may involve continued stay in emergency shelter. In many emergency shelters some services are provided. We are documenting the type and intensity of such services, and at this point would characterize them as light to very light.

The six pairwise comparisons may then be interpreted as addressing a particular policy question about the effects of bundles of housing subsidies and services. The comparisons are listed again in Exhibit 3-2, with their corresponding policy questions.

Exhibit 3-2: Policy Questions Addressed By Six Pairwise Comparisons

Pairwise 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

While families who are assigned to UC are likely to have limited access to PBTH, SUB, and CBRR due to PBTH and CBRR slots being filled by other study families and long waiting lists for SUB, they are not prohibited from accessing PBTH, SUB, or CBRR (as described in Chapter 1, Section 1.2). Therefore, the contrasts with UC may be thought of as the impact of providing direct, immediate access to an intervention versus the usual services available in a community and lower-than-normal access to transitional housing, housing vouchers, and rapid re-housing. Special analyses described later in this chapter will attempt to measure the effects of PBTH assistance versus *none* of these three, CBRR versus none of its contrasting interventions, and SUB versus none of its contrasting interventions.

In addition to the six pairwise comparisons above, we are also able to combine assignment groups in various ways to examine additional comparisons. The following three comparisons have been identified as of interest to HUD:

1. What is the impact of any kind of housing subsidy for homeless families (PBTH + SUB + CBRR) compared to the usual services offered in the community (UC)?
2. What is the impact of interventions that are more costly (PBTH + SUB) compared to a less costly intervention (CBRR)?[8]
3. What is the impact of a housing subsidy with heavy services on homeless families (PBTH) compared to a housing subsidy with light or no services (SUB + CBRR)?
4. What is the impact of a housing subsidy with no time limit (SUB) compared to a time-limited housing subsidy (PBTH + CBRR)?

Other pooled comparisons are possible but have been deemed not of interest because of their ambiguity of interpretation or lack of policy significance.

3.2 Outcomes

The study will collect outcome data in the five outcome domains of housing stability, self-sufficiency, family preservation, adult well-being, and child well-being. In Exhibit 3-3, we list the outcomes in each domain for which the study will estimate impacts. Except where noted, the 18-month follow-up survey will be the source of data for these outcomes. (Chapter 2 describes all of the sources of data that are being collected for the study.) Examining a large number of outcomes for impacts, as we propose to do here, increases the likelihood of finding significant impacts simply by chance. We describe our approach to addressing this issue (the "multiple comparisons problem") in Section 3.4.

Exhibit 3-3: Outcomes in the Family Options Study

Domain ▪ General Outcome – Specific Outcome for Which Impact Will Be Estimated
Housing Stability
▪ Homelessness during follow-up period – At least one night homeless or doubled up during past 6 months – At least one night homeless during past 6 months – At least one night doubled up during past 6 months – Any return to emergency shelter since RA [from HMIS data] – Number of days homeless or doubled up during past 6 months – Number of days homeless during past 6 months – Number of days doubled up during past 6 months

[8] As discussed in Chapter 2, we will collect data on intervention costs for the study's final report. Total cost per family is dependent on both cost per time period and the average duration that families receive an intervention. Our expectation is that cost per family will be higher in PBTH and SUB than in CBRR.

Domain ▪ General Outcome – Specific Outcome for Which Impact Will Be Estimated
▪ Housing independence – Living in own house or apartment at time of survey – Living in own house or apartment at time of survey with no housing assistance – Living in own house or apartment at time of survey with housing assistance
▪ Number of moves – Number of places lived/stayed during past 6 months
▪ Housing affordability – Rent as percentage of family income
▪ Housing quality – Persons per room – Housing problems scale
Self-Sufficiency
▪ Employment status – Work for pay in week before survey – Any work for pay since RA – Calendar months worked for pay since RA (includes partial months) – Hours of work at main job
▪ Income sources/amounts – Annualized current earnings – Total family income – Anyone in family had earnings in the last month – Anyone in family received TANF in last month – Anyone in family received SSDI in last month – Anyone in family received SSI in last month – Anyone in family received SNAP/Food Stamps in last month – Anyone in family received WIC in last month
▪ Education and training – Participated in any school or training lasting two weeks or more since RA – Number of weeks in training programs since RA – Participated in school/academic training lasting two weeks or more since RA – Participated in basic education lasting two weeks or more since RA – Participated in vocational education/training lasting two weeks or more since RA
▪ Food security/hunger – Food security scale
▪ Economic stressors – Economic stress scale
Adult Well-Being
▪ Maternal physical health – Health in past 30 days
▪ Maternal mental health – HOPE Scale – Kessler-6 Psychological Distress Scale

Domain ▪ General Outcome – Specific Outcome for Which Impact Will Be Estimated
▪ Maternal trauma symptoms – Post-traumatic Stress Diagnostic Scale (PDS)
▪ Maternal substance use – Rapid Alcohol Problems Screen – Drug Abuse Screening Test
▪ Experience of domestic violence – Experienced physical abuse or threat of violence from relationship partner in last six months
Child Well-Being (for focal child)
▪ School attendance – Child is currently enrolled in school – Days in the past month child has missed school – Number of child care arrangements since RA – Number of schools since RA
▪ Grade completion – Highest grade that child has completed – Held back in a grade since RA – School grades from last term
▪ Health status – Child's general physical health – Child received physical examination or check-up in past 12 months – Birth weight for child born since RA
▪ Behavioral strengths and challenges – Received request for school/child care conference in last 6 months because of child's behavior problems – Suspended or expelled from school/child care in last 6 months – Strengths and Difficulties Questionnaire (SDQ) scales ▪ Emotional symptoms ▪ Conduct problems ▪ Hyperactivity/inattention ▪ Peer relationship problems ▪ Pro-social behavior ▪ Total problems (sum of emotional symptoms, conduct problems, hyperactivity/inattention, and peer relationship problems) – Arrested in the last six months – Had any problems in the last six months that involved police contacting parent
▪ Preschool/Head Start – Child is currently enrolled in Early Head Start – Child is currently enrolled in Head Start

Domain ▪ **General Outcome** – **Specific Outcome for Which Impact Will Be Estimated**
▪ Home environment – Family routines scale – Sleep scale – HOME Parental warmth subscale – HOME Parental lack of hostility subscale – HOME Learning/literacy subscale (ages 0–2 yrs) – HOME Developmental stimulation subscale (ages 3 and above) – HOME Access to reading subscale (ages 3 and above) – Confusion, Hubbub, and Order Scale (CHAOS)
▪ Parenting – Parenting stress scale – Challenging environment for effective parenting scale
Family Preservation
▪ Child separations – Any children with parent at baseline but not at follow-up – Any temporary separations (now reunited) from children since baseline
▪ Placements in foster care and informal placements – Any children with parent at baseline but not at follow-up, currently with relatives not in foster care arrangement – Any children with parent at baseline but not at follow-up, currently with relatives in foster care arrangement – Any children with parent at baseline but not at follow-up, currently with non-relatives in foster care arrangement
▪ Reunifications of children with parents – Any children not with parent at baseline now with parent at follow-up
▪ Whether housing contributed to other family separations or unifications – Any other adults in household at baseline but not at follow-up – Any other adults not in household at baseline but in household at follow-up – Any temporary separations (now reunited) from other adults who were in household at baseline – Any time when respondent was in residential treatment program in past six months – Any time when respondent was in hospital in past six months – Any time when respondent was in jail or prison in past six months

3.3 Impact Estimation Model

The study will generate separate impact estimates for each pairwise comparison, including the six pairwise comparisons of a single assignment arm compared to another assignment arm plus additional comparisons of pooled assignment arms compared to a single assignment arm. The pairwise comparisons will be analyzed separately using the same estimation model. We begin the explanation of the estimation model by first introducing some terminology to describe how random assignment was implemented in the Family Options Study.

As described in Chapter 1, and shown in Exhibit 1-2, families passed the first eligibility screen for the study by staying in an emergency shelter for seven days and having at least one child age 15 or younger. Families who fit this profile were approached about participating in the study. If the family agreed to participate, then the household head was administered a baseline survey, and the family was randomly assigned to one of the four interventions in the study. The PBTH, CBRR, and (in some sites) the SUB interventions had multiple service providers in each site. Prior to random assignment, the number of slots currently available at all providers of the interventions was assessed. An intervention was deemed "available" if at least one slot at one provider of that intervention in the site was currently available to provide the intervention. Then a series of questions to assess provider-specific eligibility for the available interventions and programs was administered to the household head. Each provider had a unique set of eligibility requirements, with some providers having more stringent requirements than others. The household head was only asked the eligibility questions relevant to the providers that currently had slots available. Thus, eligibility for interventions and programs that did not have currently available slots was not assessed. A family was considered "eligible" for a particular intervention if the household head's responses to the eligibility questions showed that the family met the eligibility requirements for at least one provider of that intervention that currently had an available slot.

Based on this approach to random assignment, we will determine for each family its "randomization set" and its "assessed ineligible set."

- Randomization set: the set of interventions to which it was possible for a family to be assigned based on availability of the intervention and on assessed eligibility of the family. In the study, each family has one of seven randomization sets. These sets are {PBTH, SUB, CBRR, UC}, {PBTH, SUB, UC}, {PBTH, CBRR, UC}, {SUB, CBRR, UC}, {PBTH, UC}, {SUB, UC}, and {CBRR, UC}.
- Assessed ineligible set: the set of interventions for which a family has been 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 analysis. The randomization set of each family determines the pairwise 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. 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 control variables in the analysis.[10] Families within the same assessed ineligible set may have responded to

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.

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 control variables in the analysis.

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 six 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

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

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

$\beta_{q,r}$, $\theta_{q,r,j}$, $\phi_{q,r,k}$ = 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.

Standard Errors

We plan to estimate the model above using ordinary least squares (OLSs), 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

[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 × 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.

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 such 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.

Statistical tests for impacts

We will use estimated standard errors to perform tests of statistical significance for the impact estimates. As noted in Section 3.1, we will test and seek to reject the null hypothesis that assignment to intervention q produces equivalent outcomes as assignment to intervention r. We propose to use 0.10 as the level of statistical significance for hypothesis testing. This level of significance has been commonly used in many recent social policy experimental evaluations.[13]

Covariates

The rich data collected in the baseline survey allows for the inclusion of a broad range of domain covariates in the estimation model. This adjusts for any chance imbalances between assignment groups on family background characteristics and can decrease the standard error of the impact estimate (thereby increasing statistical power) by explaining more of the family-to-family variance in outcomes. Our goal in choosing covariates is to pick variables that capture substantial variation within the sample that may be related to outcomes. In Exhibit 3-4, we present a list of potential covariates to include in the model. This list will be revised after the evaluation team completes preliminary analyses of the final baseline survey data (June 2012) and assesses the distributions of responses to the various survey items.

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

---

12 Angrist, J. D., & Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton, NJ: Princeton University Press, 47.

13 Recent evaluations that have used a 0.10 level of significance include the Moving to Opportunity Fair Housing Demonstration (Sanbonmatsu et al., 2011), Employment, Retention, and Advancement Demonstration (Hendra et al., 2011), the U.K. Employment, Retention, and Advancement Demonstration (Hendra et al., 2011), the Sectoral Employment Impact Study (Maguire et al., 2010), the Accelerated Benefits Demonstration and Evaluation (Michalopoulos et al., 2011), the Head Start Impact Study (Puma et al., 2010), the Opening Doors Demonstration (Scrivener and Weiss, 2009), and the Building Strong Families Evaluation (Wood et al., 2010).

residual variance) to the predicted values of missing covariates.[14] 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).[15]

Exhibit 3-4: Covariates to Be Considered for Inclusion in the Impact Estimation Model

Covariate
For all outcomes
Previously experienced homelessness
Previous homelessness duration (total of all spells of homelessness)
Earlier stay in emergency shelter in year prior to baseline (from HMIS)
Previously stayed with family or friends because of economic necessity
Total time stayed with family of friends because of economic necessity in last 5 years (or since age 18)
Number of months since family had a regular place to stay (linear and quadratic terms)
Age (linear, quadratic, and cubed terms)
Gender (male, female, missing)
Marital status
Race/ethnicity
Highest level of education (less than HS diploma, HS diploma, GED, more than HS/GED)
Family income (linear and quadratic terms)
Receipt of various types of public assistance at baseline - TANF - Food Stamps/SNAP - SSI/SSDI - Medicaid - WIC
Number of children (One child, two children, three children, four children, five or more children)
Currently employed
Past eviction/lease violations
Criminal record
Have disability that limits or prevents working for pay
Family member with disability that limits or prevents household head working for pay
Ever worked for pay
Years since last worked for pay (linear and quadratic terms)

14 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.

15 Strengths and weaknesses of various methods of handling missing data are described in Allison, P.D. (2002.), *Missing data,* Thousand Oaks, CA: Sage University Paper No. 136, and in Puma, M. J., Olsen, R. B., Bell, S. H., & Price, C. (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.

Covariate
Child under age six in family
Two or more adults in family
Mental health/PTSD symptoms
Substance abuse problem
Site dummies
For child outcomes only
Age of focal child at baseline
Gender of focal child

Addressing Non-response to Follow-up Survey

The study has a 75 percent target response rate for the 18-month follow-up survey. The presence of non-response to the follow-up survey concerns us in two ways. First, non-response to the follow-up survey presents a challenge to the internal validity of the study if the intervention groups (i.e., PBTH, SUB, CBRR, and UC) have different patterns of non-response. Second, follow-up survey non-response can threaten the generalizability of results to the entire enrolled sample if survey nonrespondents differ from respondents, even if they do so *symmetrically* across randomization arms. To address both of these issues, we will prepare a set of weights that adjust for survey non-response for each pairwise comparison.[16] The weights will be used in all impact regressions. These weights will be constructed as follows: (1) we will regress a dummy variable for survey response on a large number of baseline characteristics and use the results to generate a propensity to respond for each family; (2) we will divide each intervention group into quintiles based on this propensity; (3) within each intervention group-quintile, the weight for the respondents will be the total number of sample families in the quintile divided by the number of respondent families in the quintile. This last step raises the representation of respondents to the level of the full sample in the weighted data, thereby restoring the composition of the analysis data to that of the full sample on the factors used to estimate propensities to respond. However, the weighted outcome data may still depart from the full sample on other unadjusted background factors related to subsequent outcome levels; if so, some amount of non-response bias will remain in the impact estimates—but less than without the reweighting.

3.4 Strategy for Addressing the Multiple Comparisons Problem

Statement of the Problem

The number of hypothesis tests that could be included in the Family Options Study when testing for non-zero impacts is quite large. First, the study has four random assignment arms:

- Project-Based Transitional Housing (PBTH)
- Community-Based Rapid Re-housing (CBRR)
- Subsidy Voucher (SUB)

[16] The construction of weights to address survey non-response is discussed in Little, R. J. A. (August 1986), Survey nonresponse adjustments for estimates of means, *International Statistical Review, 54* (2), 139-157.

- Usual Care (UC)

This implies a total of six pairwise comparisons in determining the relative effectiveness of the different interventions (shown in Chapter 1, Exhibit 1-3).

Second, the study has five outcome domains (housing stability, self-sufficiency, adult well-being, child well-being, and family preservation), with each domain containing several different outcome variables for each of which an impact hypothesis test could be run.

Cumulatively, the multiple arms, multiple domains and outcomes, and multiple subgroups have the potential to generate an extremely large number of hypothesis tests—easily over one hundred (six pairwise comparisons × five outcome domains × multiple outcomes within each domain).

Simply stated, the multiple comparisons problem is that as the number of hypothesis tests conducted grows, the likelihood of finding a statistically significant impact somewhere among the tested outcomes simply by chance increases far above the desired risk level for producing "false positive" results. In particular, the probability of finding at least one significant impact at the 0.10 level in *k* independent tests when all true impacts are 0 is given by equation (2):

$$(2) \qquad Prob(\min \mathrm{p} \leq 0.10 \mid all\ true\ impacts = 0) = 1\text{-}0.90^{k}$$

If 10 independent tests are performed, then the probability of finding at least one significant impact at the 0.10 level—often taken as the litmus test for a "successful" intervention—when all true impacts are equal to zero is $1 - 0.90^{10} = 0.65$; i.e., about two-thirds of the time one would conclude an unsuccessful intervention is successful. When 20 independent tests are performed, the probability is 0.88; i.e., nearly nine times out of ten. The probability of finding at least one significant impact (or more generally, rejecting at least one null hypothesis) when all true impacts equal 0 (or more generally, when all null hypotheses are true) in a "family" of *k* tests is called the family-wise error rate (FWER). In general, the FWER decreases as the *k* test statistics used become more correlated (i.e., the outcome measure tested become more closely related), leading to somewhat less risk of false positive conclusions than indicated in the previous numerical estimates). Many multiple comparison adjustment procedures have been devised to keep the FWER at or below the desired level (such as 0.05 or 0.10), some of which take account of correlation among outcomes.

Overview of Our Response to the Problem

Faced with the multiple comparisons problem as a statistical challenge, evaluators typically choose one of three approaches:

1. **Ignore the problem.** This yields incorrect statistical inferences, inferences that can conclude that the intervention "worked" to a high level of statistical certainty when it did not. This seems unacceptable.
2. **Adjust the standard of evidence used to declare all individual impact estimates statistically significant.** This procedure, due to Bonferroni, requires a *p*-value for any individual impact hypothesis test to be .10/N (or .05/N) or below—where N is the number of hypothesis tests conducted—rather than .10 (or .05) as is customarily done. Given the very large number, N, of

statistical tests that might be conducted, this procedure would raise the standard of evidence so sharply that the statistical power of the evaluation—i.e., its ability to detect non-0 impacts when they do occur—would sharply decline. Again, this is an unacceptable option.

3. **Adjust the standard of evidence used to declare a subset of individual impact estimates statistically significant.** Divide hypothesis tests into a small set of "confirmatory" tests and a much larger set of "exploratory" tests and make statistical adjustments to the former to maintain the integrity of the statistical inferences made at the confirmatory level.

This last strategy, which we propose to adopt here, hinges on the definition and implications of "confirmatory" hypothesis tests. Following Schochet (2009), we would define confirmatory hypothesis tests as those tests that "assess how strongly the study's prespecified central hypotheses are supported by the data."[17] Statistically significant findings from confirmatory hypothesis tests are considered definitive evidence of a non-zero intervention impact, effectively ending debate on whether the intervention achieved an impact in the study sites. All other hypothesis test results will be deemed "exploratory." For these tests, statistically significant impacts constitute suggestive evidence of *possible* intervention effects and will be presented heavily emphasizing that point—i.e., that unlike the confirmatory test findings, they do not identify proven impacts.

This approach is consistent with what we see as the emerging consensus in the evaluation community.[18] Staff at HUD/PD&R has indicated that the housing stability domain is the most important outcome domain of the study. Therefore, we propose to designate a small set of impact comparisons/hypothesis tests related to housing stability as the "confirmatory set." These hypothesis tests will be conducted for:

- The six pairwise policy comparisons and one pooled comparison (PBTH + SUB + CBRR versus UC), and
- A single composite outcome constructed from two binary outcomes within the housing stability domain
  - "At least one night spent homeless or doubled up during past six months" at the time of the 18-month follow-up survey
  - "Any return to emergency shelter in the 18 months following random assignment"[19] as measured from HMIS administrative data

The six pairwise comparisons need to be included in order to assess the relative effectiveness of the interventions in contributing to housing stability (thereby addressing the study's first research question stated in Section 1.4). We also include the pooled comparison of PBTH + SUB + CBRR versus UC because it will provide evidence on whether a housing subsidy (of any type) improves housing stability.

---

17 Schochet, P. (2009), An approach for addressing the multiple testing problem in social policy impact evaluations, *Evaluation Review, 33*(6), 549.

18 Schochet (2009) is the most prominent recommendation of this approach in the recent evaluation literature.

19 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.

We propose to examine this composite housing stability outcome from different sources of data to examine the general construct of housing stability. Using two sources of data to construct this outcome allows us to measure housing stability as robustly as possible.[20] This strategy will give us a total of seven confirmatory tests. Below we describe the hypothesis test we propose to use for each pairwise comparison, and the pooled comparison and the multiple comparison adjustment we will use to measure the significance of the seven tests.

Implementing the Multiple Comparison Procedure for the Confirmatory Impact Analysis

In order to structure a single statistical test for each pairwise or combined policy comparison, we propose to construct a single composite housing stability outcome variable from the two source measures described above to use in the confirmatory analysis. As noted above, the composite outcome will be equal to one if either of the component outcomes is equal to one.[21] The impact regression for the composite outcome will be as follows:

$$(3) \quad Y_{Ci} \quad = \quad \alpha_{q,r} + T_{q,i}\delta_{q,r} + X_i\beta_{q,r} + \sum_{j=1}^{6} I_{j,i}\theta_{q,r,j} + \sum_{k=1}^{13} I_{k,i}\phi_{q,r,k} + e_i \ ,$$

where

Y_{Ci} = composite outcome Y for family i,

$$Y_{Ci} = \begin{cases} 1, & \text{if either } Y_{1i} = 1 \text{ or } Y_{2i} = 1 \\ 0, & \text{if both } Y_{1i} = 0 \text{ and } Y_{2i} = 0 \end{cases}$$

Y_{1i} = "At least one night homeless or doubled up during past six months" (survey),

Y_{2i} = "Any return to emergency shelter since RA" (HMIS data), and

all other variables and coefficients have same definitions as in Equation 1.

The p-value on the impact coefficient $\delta_{q,r}$ from each pairwise comparison will be adjusted to account for the fact that there are seven confirmatory tests. The adjustment procedure will control the FWER at a 0.10 level for the seven tests. The Bonferroni-Holm step-down method is one procedure available to us to control the FWER and will be used unless we find a "resampling" approach to multiple comparison adjustments that is statistically more powerful (i.e., has greater ability to detect non-zero impacts when they do occur) than the Bonferroni-Holm method while still controlling the FWER at the 0.10 level across

[20] As explained later in this chapter, special procedures will be used to code this variable when data are missing for one of the two source variables (which can only happen for the first measure, and will happen for families that are non-respondents to the 18-month follow-up survey).

[21] While data for the second source measure, from the HMIS administrative data system, will be available for all families, the first survey-based measure will be available only for families that respond to the 18-month follow-up survey. This creates the potential for bias in impacts if survey response rates differ between assignment arms. To avoid this potential bias, we will use multiple imputation to impute the survey component of the composite for survey nonrespondents using an imputation model that includes baseline characteristics and the HMIS administrative outcome as predictors. See Puma et al. (2009) for a description and justification of this method.

the seven tests.[22] The Bonferroni-Holm method begins by ordering conventional p-values from the seven independent hypothesis tests from smallest to largest. The impact estimate from the hypothesis test with the smallest *p*-value is deemed statistically significant if the *p*-value is less than 0.10/7, approximately 0.014. If this test was deemed significant, then the second smallest *p*-value would be deemed significant if less than 0.10/6, approximately 0.017. The procedure continues with the divisor increasing by one at each "step." If at any step the *p*-value does not meet the required standard, that pairwise comparison and all remaining pairwise comparisons (i.e., those with larger *p*-values) are deemed to not have statistically significant impacts on housing stability.

3.5 Addressing No-Shows and Crossovers

This section addresses the issues of "no-shows" and "crossovers" (or more generally, the issue of noncompliance with group assignment) in the impact analysis of the Family Options Study. These issues, which are common in social experiments with two assignment groups, become more complex in the context of this study's four assignment groups. There does not appear to be a literature on the multi-armed intervention case.

In this section, we discuss the analytical challenge presented by noncompliance with random assignment and the information we will collect to detect the existence of no-shows and crossovers. At the current time, the incorporation of this information into the impact analysis remains an open challenge and we are actively working to develop an analytical approach to do so.

The ITT-TOT Distinction in Two-Way Random Assignment Studies

It is helpful to begin by outlining the basic problem of noncompliance, and analytic responses to it, in the simpler case of two-way random assignment to a treatment group that gets an intervention and a control group that does not. Two policy questions concerning program impacts could be of interest in this type of two-arm (one treatment, one control) random assignment evaluation:

- What is the effect of offering the intervention to a pool of eligible individuals or families? This is the "intention to treat" (ITT) impact measure.
- What is the effect of receiving the intervention for the subset of the individuals or families receiving the offer who take advantage of it by actually participating in intervention services? This is the "treatment on the treated" (TOT) impact measure.

For example, in the context of housing assistance to low-income families, one could ask:

- How much does the pool of families given housing subsidy vouchers benefit on average from this step, inclusive of families that both do and do not use the vouchers to lease up an apartment or other rental property?
- How much does having vouchers benefit the subset of families that actually *use* their vouchers to lease up on average?

22 Our work in this area, as well as the statistical tool development in the literature, is still evolving.

The answer to the first, ITT question tells us the effect of what the intervening entity (e.g., HUD) can actually change—by placing a voucher in the hands of an eligible family. The answer to the second, TOT question tells us the value of voucher availability when used—the step that could fundamentally change the family's future housing stability and overall well-being. Both perspectives are important to policy.

Hence, noncompliance with the intent of the intervention—that all families given vouchers use them—turns a simple question of "Is there any impact?" into two more subtle questions that both need to be addressed by the evaluation. Moreover, the number of new research questions introduced by noncompliance in four-armed experiments is larger than in this two-armed case. This adds considerable complexity to the analytic methods needed to answer all of the TOT questions of policy interest in the current study, as discussed below. To set the stage, however, we first briefly describe the two-armed version of the noncompliance phenomena and the conventional analytic response from the literature in the next section before returning to the challenges faced when four randomization arms are present. The latter takes us outside the existing evaluation literature into new methods development that is still underway.

Standard Methods of Addressing No-Shows and Crossovers in Impact Analysis

For social experiments with a single treatment group and a single control group, methods have been developed to address both possible types of noncompliance with the assigned status of a research case: no-shows (people assigned to the *treatment* group who *do not* receive the treatment) and crossovers (people assigned to the *control* group who *do* receive the treatment). Angrist, Imbens, and Rubin (1996) define the four possible types of people in this straightforward social experiment:

- "Never-takers"**—**those people who would not have taken up the treatment whether they were assigned to the treatment or control group.
- "Always-takers"**—**those people who would have taken up the treatment had they been assigned to the treatment group or would have found a way to obtain the treatment had they been assigned to the control group.
- "Compliers"**—**those people who would have taken up the treatment had they been assigned to the treatment group and would not have received the treatment had they been assigned to the control group.
- "Defiers"**—**those people who would have refused the treatment had they been assigned to the treatment group but would have found a way to obtain the treatment had they been assigned to the control group.

The standard approach in analyzing experimental data that contain non-compliers is to adopt what Angrist, Imbens, and Rubin term the "monotonicity assumption," and assume that "defiers" do not exist in the sample.[23] This reduces the number of possible types from four to three, allowing for an analytical solution that produces both ITT and TOT impact estimates in the two-arm case. Essentially, this methodology attributes all of the treatment group-control group difference in outcomes to the "compliers" group to get a TOT estimate, making the assumptions that the intervention does not affect the "never-takers" in either the treatment or control group and equally affects the "always-takers" in both groups.

[23] Angrist, J. D., Imbens, G. W., & Rubin, D. B. (June 1996), Identification of causal effects using instrumental variables, *Journal of the American Statistical Association, 91*, 448.

Complexity of Crossovers in the Family Options Study

Addressing no-shows and crossovers in the Family Options study is considerably more complex than in the two-arm social experiment example given above. In the Family Options study, there are 3 possible ways for each family not to comply with the intent of its random assignment arm. This occurs when the family participates in an intervention intended not for its research group but for one of the other three arms. For example, a family assigned to CBRR could possibly find a way to receive a housing choice voucher, transitional housing, or UC (not utilizing any program would be equivalent to "crossing over" to UC). Likewise, a family assigned to UC could possibly find a way to receive rapid re-housing, a housing choice voucher, or transitional housing. In recognition of these possibilities, many more than just four subtypes of families need consideration. These types are defined by how a family would respond to being assigned to each of the four random assignment arms (i.e., how a family would respond should they be assigned to CBRR, how they would respond should they be assigned to SUB, to PBTH, and to UC). Since possible responses to being assigned to any given arm include complying with that assignment and three kinds of crossover (i.e., participation in the services intended for other arms), the total possible number of family subtypes is $4 \times 4 \times 4 \times 4 = 256$.

It is possible to reduce the number of family types that need to be addressed in the analysis by focusing on individual two-way comparisons and making assumptions similar to the "there are no defiers" assumption in the simpler case. Considering only two interventions at a time, as will be done when making pairwise impact comparisons in the Family Options Study, shrinks the number of relevant subtypes to $4 \times 4 = 16$. We can reduce further the number of types by making additional assumptions. In the spirit of Angrist, Imbens, and Rubin's "monotonicity assumption," it seems reasonable to reduce the number of possible types from 16 to 11.[24] Exhibit 3-5 shows the 11 remaining possible types of families needing separate consideration in the TOT impact analysis in the illustrative pairwise comparison of CBRR versus SUB. Type 1 in the exhibit is the "complier" type, complying with assignment if assigned to CBRR or to SUB. Type 2 complies with assignment when assigned to CBRR, but crosses over to CBRR when assigned to SUB. Type 3 also complies when assigned to CBRR but crosses over to PBTH when assigned to SUB, and so on.

Exhibit 3-5: Possible Family Types in the CBRR versus SUB Comparison

Type Number	Possible responses to being assigned to:	
	CBRR	***SUB***
1	CBRR	SUB
2	CBRR	CBRR
3	CBRR	PBTH
4	CBRR	UC
5	SUB	SUB
6	PBTH	SUB

24 We use two assumptions to reduce the number of types from 16 to 11: (1) if families respond to being assigned to group *x* by crossing over to group *y*, then they will not cross over to group *x* in the situations they are assigned to any of the other three groups, and (2) if families respond to being assigned to group *x* by crossing over to group *y*, then they will respond to being assigned to group *y* by complying with the assignment.

Type Number	Possible responses to being assigned to:	
	CBRR	*SUB*
7	PBTH	PBTH
8	PBTH	UC
9	UC	SUB
10	UC	PBTH
11	UC	UC

Note: Type 1 is a "complier." Type 2 complies when assigned to CBRR, but finds their way into CBRR if assigned to SUB. Type 3 complies when assigned to CBRR, but finds their way into PBTH if assigned to SUB, and so on.…

Even with the simplification of the noncompliance problem to 11 possible family types, the problem remains quite complex. We are actively working on devising additional simplifying assumptions. If we are unable to make progress on this problem, we will simply report the extent and patterns of crossovers among arms that occur in the sample and discuss how this extent of noncompliance might cause TOT impacts of intervention *use* to differ from the ITT estimates of the impact of intervention *availability*, publishing only the latter. In pairwise comparisons of CBRR, PBTH, or SUB with UC, the existence of crossovers will make impact estimates conservative (i.e., it will attenuate them toward 0) and thus will not abrogate findings showing impact to be statistically significantly different from 0 even when subject to this attenuation. In pairwise comparisons that do not include UC, the effects of crossover on the magnitude of impacts will be ambiguous.

Information on No-shows and Crossovers in the Family Options Study

The Family Options Study will have high-quality information on which families do not receive the treatment to which they were assigned (no-shows), and mostly complete information on whether families received treatments to which they were not assigned (crossovers). The information on no-shows is being collected by study team members who make follow-up phone calls to service providers about every family who has been assigned to one of the CBRR, SUB, or PBTH groups. Information is being collected about whether families contact the provider to which they have been referred, whether families have moved in (received the treatment), and various reasons why families have not moved in.

The study will collect information used to identify crossovers (families who receive a different treatment than the one the experiment assigned them to receive) from HUD administrative data, the tracking and follow-up surveys, and HMIS data. Both HUD administrative data and tracking and follow-up survey data will be available to identify crossovers to SUB. Local HMIS databases and tracking and follow-up surveys will provide information on crossover to PBTH and CBRR. We expect HUD administrative data and HMIS data to provide more reliable information than the survey data. As shown in Chapter 2, Exhibit 2-7, the sites included in the study have relatively high HMIS coverage of local providers. The quality and variety of these data sources should lead to mostly complete information on crossovers.

3.6 Examining Family Characteristics that Moderate Impacts (Subgroup Analysis)

The fifth of six research questions for the Family Options Study asks, “Are various homeless interventions more effective for some categories of homeless families than for others?” A rationale for asking this question is given in the design report: “Providers often embody their hunches about the characteristics of people for whom their programs will work in restrictions on entry, but these restrictions vary considerably within programs of particular types. Thus, it is important not only to compare programs directly, but, within the limits of statistical power, to examine how programs work for different subgroups of families.” (p. 16). This section discusses the plan for subgroup analyses, which are often referred to as analyses of the moderating effect of family characteristics on the magnitude of impacts. A major consideration in planning these moderating analyses is the amount of statistical power we will have to detect impacts within subgroups and to detect differences in impacts across subgroups.

As explained above in connection with the multiple comparisons problem, we plan to consider all moderator analyses exploratory, and any findings of statistical significance that emerge will be characterized as suggestive evidence rather than as definitive proof. Even with this lower standard of evidence (see Section 3.4 above for a detailed discussion of our approach to the multiple comparisons problem), the power of the study to detect significant effects in the moderator analysis will be smaller than it is in the main impact analysis, as shown below in Section 3.9.

In order to look at impact variability across types of families, we plan to construct a “challenge” index and an “instability” index for each family in the study. We will first examine whether the indices moderate the effects of the interventions using interactions between the continuous indices and the interventions in impact regression models. This approach has the potential to provide greater power to detect differences in impacts by family types than examining individual subgroups, depending on the distribution of index values in the full sample.[25] Next, if we find evidence of moderating effects of these indices, we will use the indices to form subgroups with distinct impacts (if possible).[26] The index cutpoints that define the subgroups with different impacts may be useful guides for decision making by providers in the field.

As noted in Exhibit 3-6, the challenge index would count the number of discrete challenges facing a family at intake, while the instability index would try to predict the degree of housing instability a family would experience a year after intake if kept in UC:

---

[25] Whisman, M. A., & McClelland, G. H. (2005). Designing, testing, and interpreting interactions and moderator effects in family research, *Journal of Family Psychology, 19*(1), 113.

[26] It may not be possible to form subgroups with statistically distinct impacts even if we find evidence from the continuous interactions that the indices do moderate impacts. This is because of the lower statistical power we have to detect distinct subgroup impacts compared to the power to detect significance of the continuous interaction terms. If we are unable to form subgroups with statistically distinct impacts when continuous interaction terms are significant, the guidance to practitioners will refer only to relative magnitude of the indices, rather than to specific cutpoints in the indices.

Exhibit 3-6: Moderators of Impact to be Examined in the Exploratory Analysis

- Number of economic and functional challenges facing a family at intake ("challenge index")
- Degree of housing instability anticipated a year after intake if given usual care ("instability index")

The challenge index will be constructed using data collected in the study's baseline survey. The instability index will be empirically derived from baseline predictors of subsequent instability in the UC group. We deliberately construct the challenge indicators and instability index from information that providers in the field would have readily available. This approach will assure that the indices will be simple, transparent, and feasible to use for service targeting. Providers in the field will then have the ability to calculate either challenge scores or instability indices for the families they are serving or considering serving and to use the scores and cutpoints in making service delivery decisions. Possible variables to be used in forming the challenge and instability indices are listed in Exhibit 3-7.

The analyses examining the moderating effects of these indices will produce estimates of how changes in the indices affect the size of impact in a given pairwise policy comparison, which is useful information both for service targeting and for policy simulations. The estimation model for each index will be:

$$(3) \qquad Y_i = \alpha^{q,r} + X_i\beta^{q,r} + S_i\gamma + T_i^{q,r}\delta^{q,r} + T_i^{q,r}S_i\chi^{q,r} + e_i^{q,r}$$

where outcome Y of family i is a function of a constant α, a vector of family characteristics X, the value of index S (the challenge or instability index), the assignment to group q or r, the interacted effect of group assignment and the index S, and a random error term e. In this model, the coefficient χ ("chi") is the amount that the impact of being assigned to group q rather than group r changes when index S changes by one unit. A statistically significant and positive (or negative) estimate of χ would imply that the impact of being assigned to group q rather than group r is larger (or smaller) for families facing more challenges at baseline or greater future housing instability.

If we find that the estimate of χ is statistically significant, we will use index S to form subgroups. The estimation model for determining whether effects differ in subgroups defined by an index will be:

$$(4) \qquad Y_i = \alpha^{q,r} + X_i\beta^{q,r} + S_i^{High}\gamma + T_i^{q,r}\delta^{q,r} + T_i^{q,r}S_i^{High}\lambda^{q,r} + e_i^{q,r}$$

where S_i^{High} is an indicator variable for having a high value of the index S (determined by some cutpoint), the coefficient λ ("lambda") is the amount that the impact of being assigned to group q rather than group r changes when subgroup membership changes from the *S Low* subgroup to the *S High* subgroup, and all other terms are defined as in equation (3). The test for statistical significance of λ is the test of whether impacts differ according to subgroup membership.[27]

[27] Equation (4) assumes that two subgroups will be formed using index S as their basis. If more than two subgroups are formed, equation (4) would be modified by adding an additional subgroup indicator term and an additional interaction term for each additional subgroup. While limiting the number of subgroups to two will provide the

The challenge index will be a count of the number of challenges that the family was facing at baseline and will be constructed using information collected at enrollment. The particular challenges to be counted will be based on the list of available baseline measures shown in Exhibit 3-7. We also intend to form an instability index by estimating a predictive model of the number of residential moves in the 13th to 18th month after random assignment (the interval over which the follow-up survey will measure housing status) for a random subset of the UC sample. Using a predictive model of residential moves rather than the actual number of moves over this period provides us the opportunity to learn how impacts differ for families predicted to be at different levels of housing instability absent special intervention by allowing *all* families—not just those actually getting UC—to be categorized in this way. We will estimate the predictive model using a random subset of families in the UC sample that are then excluded from the index-focused moderator analyses in order to not "overfit" the model. Fortunately, the UC randomization arm begins with the largest sample size, so little statistical precision will be lost to any policy comparison involving that arm (and none at all for the other comparisons) because of this exclusion.

Exhibit 3-7. Potential Challenge Indicators and Housing Instability Predictors to Be Used to Measure Impact Variation by Family Type

- Previous homelessness duration
- Family member with disability
- Last worked for pay more than "n" years ago
- From Housing Barriers battery
  - Not enough income to pay rent
  - Inability to pay security deposit or first/last month's rent
  - Lack of transportation to look for housing
  - Poor credit history
  - Not currently employed
  - No rent history or no local rent history
  - Past eviction
  - Past lease violations
  - Criminal record
  - Felony drug record
- Own medical condition
- Mental health/PTSD symptoms
- Own substance use

3.7 How Intervention Features Affect Impact Magnitudes

The last of the evaluation's research questions from Chapter 1 concerns the *origins* of effective homelessness interventions:

most statistical power to detect distinct impacts, we will examine the moderating relationship between the indices and the impacts before determining the most appropriate number of subgroups.

- What features of housing, services, and structure explain the effectiveness (or lack thereof) of various homelessness interventions?

The best evidence on this front will come from the main experiment, which will provide unbiased information on—for example—how the features that define PBTH compared to features that define CBRR influence the effectiveness of homelessness interventions in general. If the TOT impact estimate that compares these two "packages" of housing support, services, and structure to one another is significant in favor, say, of PBTH, we can conclude that if one is to undertake aggressive responses to homelessness for in-shelter families, the defining features of PBTH—a deep but temporary housing subsidy plus heavy support services—yield greater effectiveness than the defining features of CBRR—a temporary deep subsidy with light support services. Similarly, if a statistically significant TOT impact estimate emerges in the comparison of PBTH to SUB or from the comparison of CBRR to SUB we will know something more about configurations of features of housing, services, and structure that improve the effectiveness of homelessness interventions. All of this evidence will be produced with greater than usual reliability and rigor thanks to random assignment of families to the three intervention arms (plus the UC group), since random assignment ensures there are no factors besides differences in intervention features that could account for greater effectiveness—i.e., better outcomes for families—that differ from one intervention arm to another.

We will not be able to rely as directly on the randomized experiment to identify possible reasons for impact variations *within* a given intervention arm rather than between arms. Thus, if one wants to know whether the physical environment of a PBTH facility helps explain the effectiveness of that intervention, one needs to contrast families randomized into PBTH and served by providers that use apartments with families in that same arm served by providers who use dormitories. Since the providers involved may differ on other factors that influence program effectiveness, and the families served may have underlying differences, attempts to attribute within-arm impact differences to specific program features will suffer from selection bias. The experiment is still helpful, however, since it provides measures of the overall effectiveness of PBTH providers compared to UC in different sites that serve as the starting point for the attributional analysis. In other words, thanks to the experiment we will at least know that *we are trying to allocate the right (i.e., unbiased) impact measures at the site level* to their causal sources, including differences in intervention features within the PBTH model from site to site.

Getting from there to correct allocation is very difficult, as has been shown in attempt after attempt to "get inside the black box" of social experiments to determine what intervention features cause impacts to be larger for some participants than for others or for some sites than for others. The difficulties stem from three sources:

- *The number of potential causal factors to be considered* in interpreting variation in measured impacts. The influence of intervention features may easily be confounded with other family and community factors.
- *The small number of places where impacts can be calculated* in order to look at patterns of variation. We have only 12 study sites, and for some pairwise intervention comparisons (e.g., PBTH vs. CBRR), fewer than 12 sites.
- *The high variance of the available measures of impact at the different sites* due to small sample sizes by site for the different random assignment arms (around 50 on average). This will make it

difficult to *prove statistically* that any site-level factors are related to impacts, much less get the causal attribution right.

In light of these facts, we have serious doubts than any meaningful analysis can be conducted of how features of the housing, services, and structure of the individual interventions (e.g., CBRR) affect intervention effectiveness. It is our recommendation to HUD that such analyses not even be attempted, given the huge potential for false attribution when dozens of factors are at play in each of the 12 sites. These factors fall into four broad categories, each containing potentially many variables:

(i) Intervention features that we seek to isolate,

(ii) Background characteristics of participating families,

(iii) Local housing market and labor market conditions that will condition what families can accomplish and how much a given intervention can help, and

(iv) The existing social service environment, including housing assistance and other supports potentially available to the UC control group as well as to the other random assignment arms.

On one of these four fronts we do not have a problem: for exploring the influence of family characteristics on intervention effectiveness, the sample size of *families* in the study, not of sites, is what counts. This number is quite large, around 2,300. The other three categories of influences on intervention effectiveness pose severe problems, however. All of those categories—intervention features, local market conditions, and other services in the community—are defined only at the site level, with the exception of features of the PBTH intervention that may vary across providers within a site.[28] For this reason, our ability to control for local market conditions and service environments in order to isolate the causal role of intervention features in determining impact magnitudes will be extremely limited. In our view, any findings that emerged from this analysis would be more misinformation than meaningful policy guidance.

It is for this reason that we recommend HUD not undertake analysis of the last research question beyond what the cross-arm comparisons of the main study reveal—with experimental rigor—about how the features that define one intervention approach (e.g., SUB) compared to the features that define another intervention approach (e.g., PBTH) influence the effectiveness of homelessness interventions in general.

3.8 Analysis of Intervention Costs

The impact analysis will examine the causal impact of each of the three interventions and UC, relative to each other. However, estimates of causal impacts are only one input into decisions about homeless policy. The interventions studied are likely to vary widely in cost—and to each possibly exceed the cost of UC. Housing assistance provided in PBTH is temporary; subsidies in the SUB intervention will be indefinite but unlikely to last more than a few years, as most families with children do not remain in the voucher

[28] Several dozen PBTH providers participated in the evaluation in the 12 sites, while only one agency or at most two in each site provided CBRR and SUB. Even for PBTH, however, it is not possible to compute sub-site impact estimates by provider, since we do not know which families in the other random assignment arms to associate with the PBTH families served by a given provider. Allocation of families to providers was done after random assignment and thus did not happen for non-PBTH families.

program for long periods of time. Interventions without services will usually require fewer resources than those offering social services. Thus, if we were to conclude that there is little difference in impacts across a pair of interventions or between an intervention and UC, this might suggest choosing the lower-cost option. Conversely, if we were to find that an intervention had both bigger impacts and larger costs, then we would need to consider whether the bigger impacts justify the higher costs.

Cost-effectiveness analysis illuminates this question by computing the cost per unit of impact produced for the different interventions for some appropriate outcome indicator such as days in shelter over the 18-month follow-up period. However, in this case cost-effectiveness analysis and policy decision making will be complicated by the multiplicity of outcomes that our interventions hope to affect—and that the follow-up survey will measure—in the areas of homelessness and housing stability, self-sufficiency, adult well-being, child well-being, and family preservation. We will work with HUD to identify the most appropriate impact measures against which to calibrate costs besides housing stability (the central impact for the study) in a cost-per-unit-of-impact analysis. Indeed, it will be possible and may be maximally informative to calculate the cost per unit of impact for a variety of impact measures—each putting programs' costs into perspective within a particular area of the interventions' "return" to society.

Given differences in intervention intensity and cost, we might expect an intervention that is more intensive and expensive to produce a larger effect. For instance, we might expect that the PBTH intervention would produce better outcomes than the CBRR intervention, given the additional services and deeper subsidy provided by PBTH. We propose to pay special attention, therefore, to the upper confidence bound of each impact estimate when interpreting our results and relating them to costs. Otherwise, we have taken no account of the statistical uncertainty in the impact measures when relating them to costs. In the PBTH vs. CBRR example (assuming that outcomes under PBTH are in fact more favorable than outcomes under CBRR, though the example could be as easily reversed), the upper bound on the 90-percent confidence interval for the PBTH vs. CBRR impact estimate will tell us that outcomes under PBTH are unlikely to be better than outcomes under CBRR by an amount *greater than* this limit. To then relate its benefits in improving the particular outcome measure of interest, we can use the upper bound value to derive an estimate of the minimum cost per unit of impact that PBTH might entail compared to CBRR (since the maximum level of impact for a given cost produces the minimum cost per unit of impact). If policymakers decide that even this best-case scenario provides a return on investment from intervention expenditures insufficient to justify the more costly PBTH approach, there is little risk that this conclusion lacks a solid factual basis (a 10 percent risk).

3.9 Sample Sizes and Statistical Power

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 for some pairwise comparisons 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. 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 3-8 shows the MDEs by pairwise comparison for an impact analysis sample size of 1,729, assuming a 75 percent response rate for the follow-up survey in each random assignment arm (and hence for the full sample of 2,305 families). The MDEs presented are the minimum detectable differences in outcomes (in percentage points) between two randomly assigned groups with 80 percent power when we perform a two-sided[29] statistical test at a 10 percent level of significance, assuming a regression R^2 of 0.04[30] and no finite population correction.[31] The differences are shown for various average outcome levels for the second assignment group in each comparison.

The last column of the first row of Exhibit 3-8 shows that for a binary outcome with a mean of one half, 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 homelessness—from 50 percent to 41.7 percent, 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.

Our hypothesis is that the interventions to be tested in relation to the UC 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

29 While one-sided tests would decrease MDEs, 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.

30 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 MDEs.

31 Applying the finite population correction (FPC) would reduce the MDEs. 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.

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.[32] 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 did not (18 percent stable at five years) in the New York study. Thus, we conclude that an MDE of 8.0 to 10.2 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 UC group.

Exhibit 3-8: Minimum Detectable Effects for Prevalence Estimates by Pairwise Comparison

	Expected Number of Completed Follow-up Survey Interviews		MDE if Mean Outcome for the Second Assignment Group Is:		
Sample	First Assignment Group	Second Assignment Group	0.1 (or 0.9)	0.3 (or 0.7)	0.5
CBRR vs. UC	433	435	5.0 pp	7.6 pp	8.3 pp
SUB vs. UC	453	411	5.0 pp	7.6 pp	8.3 pp
PBTH vs. UC	272	257	6.4 pp	9.7 pp	10.6 pp
CBRR vs. SUB	290	329	5.9 pp	9.0 pp	9.8 pp
CBRR vs. PBTH	177	175	7.8 pp	11.9 pp	13.0 pp
SUB vs. PBTH	180	194	7.6 pp	11.5 pp	12.6 pp

Notes: (1) The MDEs 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 MDEs assume a 75 percent 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 3-8 labeled “MDE if Mean Control Group Outcome is: 0.3” likely the most relevant one.[33] Here, a smaller true impact can be detected with 80 percent assurance. The MDEs in Exhibit 3-8 are for analyses that are performed with the entire pooled sample. MDEs for split-sample analysis of discrete subgroups, were any conducted, would be larger than those shown here. As noted elsewhere, the study intends to explore how impacts differ by

32 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, two or three times—the observed 27 percentage point difference would mean an impact of 9–13 percentage points.

33 We note that Cowal et al. (2002), finds 44 percent of families with child separations over five years. We expect that child separations within the first 18 months will be somewhat lower.

family characteristics using the “challenge” and “instability” indices, the moderating roles of which can be examined without dividing the sample into smaller subsamples.[34]

Power Calculations for Earnings

Exhibit 3-9 shows the MDEs for earnings impacts by pairwise comparison. These MDEs 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 percent 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, many interventions focused directly on employment and training do not produce annual earnings impacts of this size. 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.

Exhibit 3-9: 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 MDEs 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 MDEs assume a 75 percent 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.

[34] “Challenge score” can be entered into the impact regression equation interacted with indicator variables for the different random assignment groups to see if the magnitude of effect from being assigned to a particular service package changes as the degree of family challenge rises, and the equation then estimated using all the data.

References

Allison, P. D. (2001). *Missing data*. Sage University Papers Series on Quantitative Applications in the Social Sciences, 07-136. Thousand Oaks, CA: Sage.

Angrist, J., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, *91,* 444-472.

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 U.S. Department of Housing and Urban Development.

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. 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.

Duffer, A. P., Lessler, J., Weeks, M., & Mosher, W. (1994). *Effects of incentive payments on response rates and field costs in a pretest of a national CAPI survey*. Research Triangle Institute, May.

Educational Testing Service. (1991). National adult literacy survey addendum to clearance package, Volume II: Analyses of the NALS field test (2-3).

Greene, W. H. (2003). *Econometric analysis: Fifth edition*. Upper Saddle River, NJ: Prentice-Hall, Inc.

Hendra, R., Ray, K., Vegeris, S., Hevenstone, D., & Hudson, M. (2011). *Employment, Retention and Advancement (ERA) demonstration delivery, take-up, and outcomes of an in-work training support for lone parents*. New York: MDRC.

Hendra, R., Riccio, J., Dorsett, R., Greenberg, D. H., Knight, G., Phillips, J., Robins, P., Vegeris, S., & Walter, J. (with Hill, A., Ray, K., & Smith, J.). (2011). *Breaking the low-pay, no-pay cycle: Final evidence from the UK Employment Retention and Advancement (ERA) Demonstration.* New York: MDRC.

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.

Little, R. J. A. (1986). Survey nonresponse adjustments for estimates of means. *International Statistical Review, 54*(2), 139-157.

Locke, G., Khadduri, J., & O'Hara, A. (2007). Housing models. In *Toward Ending Homelessness: The 2007 National Symposium on Homelessness Research*.

Maguire, S., Freely, J., Clymer, C., Conway, M., & Schwartz, D. (2010). Tuning in to local labor markets: Findings from the sectoral employment impact study. Philadelphia: Public/Private Ventures.

Michalopoulos, C., Wittenburg, D., Israel, D. A. R., Schore, J., Warren, A., Zutshi, A., Freedman, S., & Schwartz, L. (2011). *The accelerated benefits demonstration and evaluation project impacts on health and employment at twelve months*, Vol. 1. New York: MDRC.

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.

Orr, L. L. (1999). *Social experiments: Evaluating public programs with experimental methods*. Thousand Oaks, CA: Sage Publications.

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*. Prepared for the U.S. Department of Housing and Urban Development, Office of Policy Development and Research. Cambridge, MA: Abt Associates Inc. and National Bureau of Economic Research.

Puma, M., Bell, S., Cook, R., Heid, C., Shapiro, G., Broene, P., Jenkins, F., Fletcher, P., Quinn, L., Friedman, J., Ciarico, J., Rohacek, M., Adams, G., & Spier, E. (2010). *Head start impact study: Final report.* Washington, DC: U.S. Department of Health and Human Services.

Puma, M. J., Olsen, R. B., Bell, S. H., & Price, C. (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.

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.

Sanbonmatsu, L., Ludwig, J., Katz, L. F., Gennetian, L. A., Duncan, G. J., Kessler, R. C., Adam, E., McDade, T. W., & Lindau, S. T. (2011). *Moving to opportunity for fair housing demonstration program: Final impacts evaluation*. Prepared for the U.S. Department of Housing and Urban Development, Office of Policy Development and Research. Cambridge, MA: National Bureau of Economic Research.

Schochet, P. (2009). An approach for addressing the multiple testing problem in social policy impact evaluations. *Evaluation Review 33*(6), 539-567.

Scrivener, S., & Weiss, M. (with Teres, J.). (2009). *More guidance, better results? Three-year effects of an enhanced student services program at two community colleges.* New York: MDRC.

Shinn, M., Weitzman, B. C., Stojanovic, D., Knickman, J. R., Jimin,ez, 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.

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. (2008). *The third annual homeless assessment report to Congress.* Washington, DC.

U.S. Department of Housing and Urban Development. (2011). *The fifth annual homeless assessment report (AHAR) to Congress.* Washington, DC.

Whisman, M. A., & McClelland, G. H. (2005). Designing, testing, and interpreting interactions and moderator effects in family research. *Journal of Family Psychology, 19*(1), 111-120.

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.

Wood, R., McConnell, S., Moore, Q., Clarkwest, A., & Hsueh, J. (2010). *Strengthening unmarried parents' relationships: The early impacts of building strong families.* Princeton, NJ: Mathematica Policy Research.

U.S. Department of Housing and Urban Development
Office of Policy Development and Research
Washington, DC 20410-6000

U.S. DEPARTMENT OF HOUSING AND URBAN DEVELOPMENT

EQUAL HOUSING OPPORTUNITY

March 2013