

**Supporting Statement for Paperwork Reduction Act Submission,  
Capital Needs Assessments of Public Housing  
OMB Control Number 2528-New**

**B. COLLECTIONS OF INFORMATION EMPLOYING STATISTICAL METHODS**

**1. Describe (including a numerical estimate) the potential respondent universe and any sampling or other respondent selection method to be used. Data on the number of entities (e.g. establishments, State and local governmental units, households, or persons) in the universe covered by the collection and in the corresponding sample are to be provided in tabular form for the universe as a whole and for each of the strata in the proposed sample. Indicate expected response rates for the collection as a whole. If the collection had been conducted previously, provide the actual response rate achieved during the last collection.**

The information collected through this study will be used by the U.S. Department of Housing and Urban Development (HUD) to do the following:

1. Evaluate the approach to estimating capital needs and produce an estimate of capital needs for the total stock of public housing.
2. Develop and evaluate a selected predictive model using HUD data, primary survey data, and other data to estimate the capital needs of public housing.
3. Update the predictive model and evaluation based on the second survey iteration.

What follows presents a description of each information collection effort that the research team will perform: (1) sampling or respondent selection method, (2) the number of entities in the universe, (3) the sample planned to be achieved, and (4) expected response rates for each collection.

**Phase 1: Collect documentation and HUD administrative data, interview HUD staff and subject matter experts, and conduct literature review (not subject to the Paperwork Reduction Act (PRA)).**

As a first step, the research team will collect, review, and analyze documentation, HUD administrative data, and previously collected capital needs assessment (CNA) data available to the research team about CNAs. The research team will also interview HUD staff and talk to fewer than 10 subject matter experts and practitioners about best practices in CNAs. The data and documentation will be used to define capital needs, identify best practices in estimating capital needs, and evaluate the adequacy of HUD data to estimating capital needs.

**Phase 2: Identification of Respondents for First Survey of Public Housing Authorities (PHAs) (subject to PRA)**

The research team plans to administer a web-based survey to a stratified random sample of approximately 300 PHAs. There are currently 2,780 PHAs with active public housing programs. These PHAs include 6,523 housing projects with 940,330 units, or an average of 2.3 projects per PHA and 144 units per project (Table 1). The research team expects a response rate of 80 percent

(240 PHAs). The data from this survey will be used to describe how PHAs estimate their capital needs, evaluate their use of best practices, and identify perceived barriers to meeting their capital needs. The survey data will also be combined with data collected in Phase 1 to develop a model that will be used to estimate the capital needs of the total public housing stock.

**Phase 3: Identification of Respondents for Second Survey of PHAs (subject to PRA)**

The research team plans to administer the same web-based survey to a second stratified random sample of 500 PHAs. The second sample will follow the same stratification framework as the first. The second survey will ask many of the same questions as the first iteration of the survey. This effort will provide further validation and refinement of the CNA model developed and tested in Phase 2, a revised estimate of the capital needs of public housing, a revised assessment of PHAs’ use of best practices to estimate capital needs, and a revised identification of PHAs’ perceptions of barriers to meeting their capital needs. The research team expects a response rate of 80 percent (400 PHAs).

**Table 1. Universe of PHAs, Projects and Units, as of February 2022**

PHAs	Projects	Units
2,780	6,523	940,330

In addition to obtaining data from HUD and collecting primary data from PHAs, the research team also will analyze previously collected CNA data from a CNA provider that is a member of the research team. These data have been collected over the past 5 years, are not in HUD’s possession, are the property of the CNA provider, represent approximately 2,588 public housing and properties and 251,036 public housing units, and will be made available in Phase 1 data collection (Table 2).

**Table 2. Number of Public Housing Projects with Units, Previously Collected CNAs, 2016–2021**

Year	Number of Projects With CNAs	Approx. Number of Units
5 year	2,588	251,036

**2. Describe the procedures for the collection, including:**

- a. **The statistical methodology for stratification and sample selection;**
- b. **The estimation procedure;**
- c. **The degree of accuracy needed for the purpose described in the justification;**
- d. **Any unusual problems requiring specialized sampling procedures; and**
- e. **Any use of periodic (less frequent than annual) data collection cycles to reduce burden.**

- ***Statistical methodology for stratification and sample selection***

**Table 3. Summary of Methods for Stratification and Sample Selection**

Information Collection	Stratification and Sampling Methods
First Survey of PHAs	Sample approximately 300 PHAs. The research team will use stratification and random selection without replacement to select a

Information Collection	Stratification and Sampling Methods
	<p>sample of PHAs to survey about their estimates of capital needs, the practices they use to arrive at these estimates, and their perception of barriers to meeting their capital needs.</p> <p>Stratification variables will be identified in two stages. First, the research team will collect data on 13 variables such as property type, location, PHA size, and building age. Second, the research team will conduct a basic factor analysis on the stratification variables with the goal of reducing the variables to a smaller number of stratification factors.</p>
Second Survey of PHAs	<p>Sample a second, non-overlapping group of approximately 500 PHAs.</p> <p>The second iteration sample will allow revisions to the initial sampling approach based on actual data. The intention is to avoid overlap in PHAs between the two samples by eliminating the PHAs in the first sample from the population to be sampled for the second survey. The research team may alter the specific strata used in the first sample to stratify the second sample.</p>

- ***Estimation procedure***

Using data from the two separate surveys of PHAs, the research team will test the extent to which primary data provided by a representative sample of PHAs can be used in combination with HUD data and other secondary data available to estimate the capital needs of public housing. The two surveys will also gather information about the capital needs data that PHAs collect, ask PHAs to share those data with HUD, assess the use of best practices by PHAs for CNAs, and help identify barriers to meeting the capital needs of public housing.

### **First Survey of PHAs**

The main purpose of the first survey will be to develop and evaluate the performance of models using data collected from this survey and data collected in Phase 1 from HUD and other available sources. The research team will need to account for the differences between the survey data and the datasets collected in Phase 1. For example, the survey sample will more likely be representative of the population upon which it is based, whereas HUD and other data sources are less likely to be representative of all PHAs. In addition, the capital needs estimates obtained from the survey and secondary sources may be based on different approaches and/or methods of implementation. Finally, the research team will need to address nonresponse bias in the survey data, which could be a significant issue. To the extent possible, the research team will use data from Phase 1 to test the accuracy of out-of-sample predictions based on the survey-based model implementations.

### **Second Survey of PHAs**

The objective of this phase of data collection is to incorporate data from the second survey into the selected model specifications and assess the resulting improvements in predictive power. If any changes are made to the survey to address gaps/weaknesses in the first, the analyst will need to determine the implications for how the second sample data can be used.

- ***The degree of accuracy needed for the purpose described in the justification.***

### ***First Survey of PHAs***

The analyst will examine all variables and use descriptive statistics to identify outliers, data entry errors, and missing values; records with such anomalies will be flagged and addressed by removal, imputation, or other appropriate methods. Variables that need to be standardized or normalized will be constructed where needed.

Missing values in the survey data will require special attention. It is possible that some PHAs that respond to the survey may not be able to provide an estimate of their capital needs or choose not to do so. This could result in the following three categories of missing values:

- PHAs that do not respond to the survey.
- PHAs that respond but cannot provide estimates of their capital needs.
- PHAs that respond but choose not to provide estimates of their capital needs.

There are several ways to handle these nonresponses. First, the analyst could treat all three categories as a simple nonresponse and develop a single nonresponse adjustment factor for the sample weights. Another option would be to treat each category separately and develop three specific nonresponse adjustment factors (one for each category). For both approaches, the capital needs estimates for each stratum would be weighted by its corresponding sample weight and nonresponse adjustment factor to produce the desired population estimate for the strata. A potential issue with these approaches, however, is the likely possibility that many of the strata may not have enough data (or any at all) to yield credible estimates. In other words, the expectation is that there will be a lack of CNA data from many projects.

### ***Second Survey of PHAs***

The second iteration of the PHA survey will allow the analyst to test model(s) with new observations, comparing predicted-to-actual capital needs at a representative sample of projects. Results of these tests will be used to further refine the model(s). The finalized model(s) will be applied to the universe of eligible public housing projects to produce an estimate of total capital needs.

- ***Any unusual problems requiring specialized sampling procedures***

There are no unusual problems that require specialized sampling procedures.

- ***Use of periodic (less frequent than annual) data collection cycles to reduce burden.***

Each of the data proposed is a one-time occurrence; thus, a periodic cycle is unnecessary.

- 3. Describe the methods used to maximize response rates and to deal with non-response. The accuracy and reliability of the information collected must be shown to be adequate for the intended uses. For collections based on sampling, a special justification must be provided if they will not yield “reliable” data that can be generalized to the universe studied.**

### **First (Phase 2) Survey of PHAs**

The research team will recruit PHA and HUD staff members to pretest the data collection protocols and instruments to check the survey flow, clarity of questions, length, time to complete, and general data collection procedure. Based on the results of the pretest and in conjunction with HUD, the data collection protocols and survey instrument will be finalized. A pilot test or soft launch of the survey with a handful of PHAs will be conducted. Any key issues identified during the pilot test will be summarized and submitted to HUD to allow for revisions to the data collection protocol as necessary.

- Ensure that questions are clear and easy to understand and that the time needed to complete the survey is short. One method of doing so is to use close-ended questions as much as possible.
- To underscore the survey’s importance, requests to participate in the survey should come from a HUD official on HUD letterhead and a HUD email address.
- Contact public housing industry groups to convey the importance of this research to their members and encourage responses.
- Letter and email invitations and follow-up postcard and email reminders will be drafted in personalized form and addressed to the primary contact person. Call the primary contact to resolve substantially incomplete or problematic responses. If possible, HUD letterhead and logos will be included in these invitations and reminders.
- The invitation will clearly state why it will benefit the PHA to respond to the survey.
- To make it easy to respond, the online survey will be accessible via a variety of mediums (for example, tablets, smartphones, computers) and internet browsers.
- The online survey will clearly identify that it is sponsored by HUD with HUD’s logo.
- Suggest calling nonrespondents by telephone if the response rate is below 70 percent. Here, a two-step approach is recommended: The team will make a series of follow-up phone calls to PHAs that do not respond to the survey within 2 weeks of the original mail and email request. If the research team is not able to obtain the minimum number of acceptable responses for each group, the team will enlist the help of HUD staff to make one further follow-up call.

### **Second (Phase 3) Survey of PHAs**

- The second iteration of the survey will allow us to adjust our approach to fielding the survey as needed to increase response rates.

**Table 4. PHA Survey Fielding**

Proposed Survey	Universe	Sample	Target Completes	Response Rate (Minimum-Maximum)
First Survey (Phase 2) PHAs survey	2,780	300	240	80%
Second Survey (Phase 3) PHAs survey	2,480*	500	400	80%

\* The second survey will use a sample drawn from the population of PHAs with public housing projects excluding the 300 PHAs included in the sample for the first survey.

The research team will address nonresponse bias through comparison of known characteristics of respondents to nonrespondents and an analysis of the nonresponse sample at the end of fielding. Our nonresponse sample of 160, resulting in at least 640 completes, will identify statistically significant bias at a small effect size.

**4. Describe any tests of procedures or methods to be undertaken. Tests are encouraged as effective means to refine collections, but if ten or more test respondents are involved OMB must give prior approval.**

***Phase 2 Survey of PHAs and Phase 3 Survey of PHAs***

Multivariate regression analysis is the main technique the research team will employ to evaluate the use of primary and secondary data to develop national-level estimates of the capital needs of public housing. In this approach, the analyst will regress different measures of capital needs obtained from the survey on independent variables obtained from primary and secondary data sources, such as building age, number of units, and location.

If this possibility materializes, the research team proposes to take the following steps to address the significant lack of CNA data in the survey sample. The research team’s analyst will begin by implementing a post-stratification adjustment, which will entail (1) adding a new binary stratum that reflects the presence or absence of a CNA, and (2) using a raking algorithm to recalibrate the original weights. The analyst will then use propensity scores to match the survey observations without CNAs to observations in a larger separately constructed dataset of historical CNAs. Imputed values for the matched survey observations will then be drawn and inflated to the specified year of the analysis. In the final step, the analyst will extrapolate the sample capital needs estimates (including the imputed values) to the population using either the sample weights or a regression-based approach as described above.

**Table 5. Sample Stratification Variables**

Stratification Variable	PHA Coverage	Development/AMP Coverage
Information that HUD already possesses on CNAs and cost estimates.	Unknown	Unknown
Information that HUD needs to understand the full range of activities involved in a CNA.	Unknown	Unknown
CNAs and RAD conversions.	100%	100%
PHAs that have not performed a CNA or PNA.	7.5% to 45.0%	8.9% to 50.0%
PHAs that could not address all capital needs identified in a CNA or PNA.	Minimal	Minimal
Geography.	100%	100%
Building size.	100%	100%
Building age.	100%	100%
Market rates.	100%	100%
Urban–rural.	100%	100%
Single-family homes or duplexes.	100%	100%
Mixed-income projects with market-rate units.	100%	100%
HUD region.	100%	100%

Note: Building and development percentages are based on those with occupied low-rent units as of November 2021. Building age is based on the recorded date of first availability and may not reflect more recent rehabilitation or significant capital investment.

Given the potentially large number of regressors and the associated risk of an over-fitted model, the analyst will deploy a LASSO (Least Absolute Shrinkage and Selection Operator) regression as part of the analytic framework. This method utilizes a shrinkage algorithm to reduce a set of variables and improve both predictive accuracy and interpretability of the coefficients. As standard practice, several test statistics will also inform variable selection; these include a t-test or p-value for the statistical significance of the coefficients and the Akaike information criterion, which is a measure of model quality that balances model fit with simplicity. To ensure use of the best set of variables, the analyst will also conduct a principal component analysis.

Scatterplots between the dependent and independent variables, as well as analysis of residuals, will help determine the need for data transformations (for example, logging a variable) that will improve overall model fit. The analyst will also consider potential interactions between specific variables by adding interactive terms when specifying the model.

5. Provide the name and telephone number of individuals consulted on the statistical aspects of the design, and the name of the agency unit, contractor(s), grantee(s), or other person(s) who will actually collect and/or analyze the information for the agency.

**Table 6. Names, Affiliations, and Contact Information**

Role	Personnel	Phone Number/Email
<b>Econometrica, Inc. – Prime</b>		
Project Director	Dennis Stout	(301) 657-9883; <a href="mailto:DStout@EconometricaInc.com">DStout@EconometricaInc.com</a>
Survey Lead	Jennifer Stoloff	(301) 657-9883; <a href="mailto:JStoloff@EconometricaInc.com">JStoloff@EconometricaInc.com</a>
Data Analyst	Alex Thackeray	(301) 657-9883; <a href="mailto:AThackeray@EconometricaInc.com">AThackeray@EconometricaInc.com</a>
Data Analysis Lead	Chris Holleyman	(301) 657-9883; <a href="mailto:CHolleyman@EconometricaInc.com">CHolleyman@EconometricaInc.com</a>
Policy Options Lead	Margaret McGilvray	(301) 657-9883; <a href="mailto:MMcGilvray@EconometricaInc.com">MMcGilvray@EconometricaInc.com</a>

**Table 7. HUD Staff Who Advised on the Survey and Interview Instruments**

Name	Role	Phone Number/Email
Regina Gray	Co-COR	(202) 402-2876; <a href="mailto:Regina.C.Gray@hud.gov">Regina.C.Gray@hud.gov</a>
Mark Reardon	Co-COR	(202) 402-2133; <a href="mailto:Mark.A.Reardon@hud.gov">Mark.A.Reardon@hud.gov</a>