Supporting Statement for Paperwork Reduction Act Submissions
Validating Estimates of CPD Grantee Accrued Expenses
(OMB # 2506-New)

Supporting Statement

B. Collections of Information Employing Statistical Methods

1. Describe (including a numerical estimate) the potential respondent universe and any sampling or other respondent selection methods to be used. Data on the number of entities (e.g., establishments, State and local government 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, include the actual response rate achieved during the last collection.

    The proposed approach will be to sample information from 200 of the approximately 65,000 grant related disbursements issued for the most recently completed quarter (following approval) from HUD's HUDCAPS financial datamart. Based on prior research these are estimated to amount to approximately $2.75 billion. Negative disbursement records or records categorized as "payables" which were applicable to current or prior years will be excluded.

    The proposed approach is based on a September 30, 2014 OIG Audit Report, "Undocumented, Payable Liabilities in HUD's Financial Position." That report made use of a stratified sample by disbursement size, with multiple program-specific strata. The approach proposed here does not rely on within-program strata, but instead allocates sample to single program strata based on a desired level of precision at the program level. Within individual programs we propose to use simple random sampling.

    Upon selection of samples of disbursements from these individual programs, the grantees involved in the disbursements will be contacted via e-mail and telephone regarding the disbursements and asked to provide supporting financial information regarding the performance dates of activities for which the disbursements reflect repayment. Payments for activities performed in prior fiscal years will be identified as "accruals" and the share of current-period payments that are accruals will be calculated for each program by extrapolating from the sample estimate; a CPD-wide accruals percentage will be calculated using robust statistics taking the different programs' sampling probabilities into account.

	Lower Bound	Program Stratum Size	Sample	Sampling Weight	Assumed Population Proportion	Projected Sampling Error/Precision Level	FPC Used?
CDBG	All	45579	60	759.7	0.741176	+/- 9.3%	NO
HOME	All	6915	60	115.3	0.741176	+/- 9.3%	NO
COC	All	649	26	25.0	0.741176	+/- 14.13%	NO
HESG	All	6560	27	243.0	0.741176	+/- 13.87%	NO
HOPWA	All	6915	27	256.1	0.741176	+/- 13.87%	NO

The previous data collection, in 2014, resulted in a 100 percent response rate from grantees.

2. Describe the procedures for the collection of information including:

- *Statistical methodology for stratification and sample selection*

  Allocation of sample to programs is based on a "precision target" methodology, with different projected levels of precision (projected sampling error) by program. Projected sampling error is based on program stratum sizes and samples, a binomial distribution (a disbursement is an accrual or is not), an alpha level of 0.90 and a population proportion estimate of 0.74 (the latter two parameters following the prior OIG result). The CDBG and HOME programs receive allocations of sample resulting in a margin of error of approximately +/- 9%, with the remaining CPD programs given sample to target a level of precision under +/- 15%. The prior OIG report had an estimated level of precision of +/- 16.2%. None of the projected standard errors receive the finite population correction (FPC) as none of the program strata have a sampling fraction above 5%.

  Samples of individual program disbursements will be drawn using the sample() function in the R Statistical Environment (R Core Team, 2008).

- *Estimation procedure*

  The accrual rate for each program will be calculated by assigning an indicator (0-1) variable for accruals, which receives a value of "1" if a disbursement is identified as an accrual, and "0" if it is not; the point-estimate of the accrual rate at the program level is simply the average of the indicator variable for the program sample. The CPD-level accrual percentage will be calculated via a robust statistical method taking into account the program-stratum accrual rate and the stratum sampling weight (which varies by the program's disbursements' individual sampling probabilities). For the calculation of the overall CPD accrual percentage we propose to use the "survey" package for the R statistical computing environment, which permits the calculation of robust summary statistics from complex samples (Lumley 2004; 2014).

- *Degree of accuracy needed for the purpose described in the justification*

  The method described above should produce acceptable estimates of program accruals (+/- 9% to +/- 14%) with a 90% degree of confidence.

- *Unusual problems requiring specialized sampling procedures*

  None are anticipated.

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

  As there is no reason to believe that for well-established programs the accrual rate changes *apart from the first year of the program* (when on an annual period comparison basis there are no accruals, there being no prior periods) *and the conclusion of the final year of program funding* (after which *all* expenditures represent accruals), for most programs it will not be necessary to conduct this analysis annually. Instead, prior estimates can be used and the analysis repeated every 2-3 years.

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

   As part of the Office of Inspector General's (OIG) Financial Statement audit, they validated the Departments accrual data through a statistical sample which netted 100% response rate. CPD cannot use the data obtained through the OIG's review because of the Independence Principle. However, in the event that a grantee is unwilling or unable to provide supporting information to allow a determination of the program year in which an activity whose expenses are paid with a current disbursement occurred, another disbursement will be randomly selected (without replacement) from the same stratum. No "take all" or 100% sampled strata are proposed, so resampling without replacement will not present an issue.

4. Describe any tests of procedures or methods to be undertaken. Testing is encouraged as an effective means of refining collections of information to minimize burden and improve utility. Tests must be approved if they call for answers to identical questions from 10 or more respondents. A proposed test or set of test may be submitted for approval separately or in combination with the main collection of information.

   The information request to grantees consists of a communication asking them to provide all supporting financial materials (such as vendor invoices) for a specific, identified disbursement (or list of disbursements, if the grantee appears more than once in the sample). The information to be provided in each case will likely vary considerably from grantee to grantee based on their recordkeeping practices. Based on prior experience, it should not be necessary to conduct any cognitive or other testing on this request for information.

5. Provide the name and telephone number of individuals consulted on 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.

Statistical consultant and lead analyst:

Dr. Steven T. Sullivan
Senior Director and SME
Cloudburst Consulting Group, Inc.
8400 Corporate Drive, Suite 550
Landover, MD 20785
301-385-6693
steven.sullivan@cloudburstgroup.com

References

R Development Core Team, 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.
Lumley, T., 2014: Survey: Analysis of complex survey samples. R package version 3.30.
Lumley, T., 2004: Analysis of complex survey samples. Journal of Statistical Software 9(1): 1-19.