SUPPORTING STATEMENT PART B – COLLECTION OF INFORMATION EMPLOYING STATISTICAL METHODS

1. Respondent Universe and Sampling Methods

Below we describe the target population and sample frame, the sample design, the sample size, and the selection method. We also discuss postsurvey adjustments to the sample weights and variance estimation from survey data with a complex design.

a. Sample Design and Survey Implementation

Overview. The sample for the HHI is drawn in two stages: (1) agencies and (2) patients within agencies. The goal of the HHI sample design is to select a representative sample of 50 home health agencies that have patients who are potentially eligible for the demonstration in each of the three participating states (Colorado, Massachusetts, and Missouri), and to select a representative sample of such patients within each selected agency. Within each of the selected agencies, we plan to randomly select 5 patients who are potentially eligible for the demonstration, and for whom the agency will abstract the medical record information needed for this evaluation.

Eligibility. For sampling purposes, we have decided to include as "potentially eligible" those patients who have been receiving Medicare home health services from these agencies for two or more consecutive 60-day episodes, who use technical or human assistance to move, and who have three or more Activities of Daily Living (ADLs) for which they need assistance. An agency is considered to be eligible for this sample if it has five or more eligible patients.

Agency Sample. According to the most recent OASIS administrative data we have,[2] there are 60 eligible agencies in Colorado, 80 eligible agencies in Massachusetts, and 81 in Missouri. If these numbers are similar when looking at updated administrative data, we

will randomly select 50 in each of the three states. Within each state, we will randomly sort the eligible agencies by zip code, and select a sample of 50 in each, with probability proportional to the number of potentially eligible patients. We will use a sequential sampling technique based on a

[2] From January 2004 to June 2005.

procedure developed by Chromy.[3] If an agency is so large that its probability of selection is one or greater, we will select that agency with certainty and remove it from the random selection process. Based on the current data, 20 to 30 agencies per state will be selected with certainty.

Patient Sample. Within each of the selected agencies, we will obtain a list of potentially eligible patients, with information about their gender and age. We will sort each agency's list by gender, and then by age within gender, and select a Chromy sequential sample of 5 patients per agency with equal probability.

b. Weighting and Precision

After data collection, we will produce agency-level and patient-level analysis weights. At the agency level, the base weight will be the inverse of the probability of selection of the agency. The certainty selections will have an agency-level base weight equal to 1. If there is any ineligibility or nonresponse at the agency level, this will be accounted for in the weights. This weight should be used for agency-level analyses.

At the patient level, the weight will be the inverse of the patient's probability of

selection within agency. Should there be any missing medical record abstractions (for example, record not found), this will be accounted for in the weights as well. The cumulative patient-level weight, which should be used for patient-level analyses, will be the product of the agency and patient weights. For patients within non-certainty agencies, the cumulative weight should be approximately equal to 1.

[3]The Chromy procedure offers all the advantages of the systematic sampling approach but eliminates the risk of bias associated with that approach. It makes independent selections within each of the sampling intervals while controlling the selection opportunities for units crossing interval boundaries. Chromy, J.R. "Sequential Sample Selection Methods." Proceedings of the Survey Research Methods Section of the American Statistical Association, 1979, pp. 401-406.

c. Variance Estimation

At the agency level, the variance of estimates must account for the unequal weights across agencies, due to the probability-proportional-to-size selection methodology. Table C.1 shows the standard errors and 95 percent confidence intervals around estimates using all 50 agencies in a state, and around estimates from a subgroup of half these agencies. For example, if the outcome measure of interest is the proportion of agencies that report a certain characteristic, the in Massachusetts, if that proportion is .4, the standard error would be about .085, and the confidence interval would be .4 ± .170.

Table C.1. Precision of Agency-Level Estimates

Outcome measure – proportion of sample equal to:		.1	.2	.3	.4	.5
		.9	.8	.7	.6	.5
Colorado						
design effect[a]	1.089					
Sample size=50	std. error	.045	.060	.068	.073	.075
	conf. intvl.[b]	.090	.120	.137	.147	.150
Sample Size=25	std. Err.	.064	.085	.098	.104	.107

	conf. intvl.	.132	.176	.201	.215	.220
Massachusetts						
design effect	1.462					
Sample Size=50	std. error	.052	.069	.079	.085	.086
	conf. intvl.	.104	.139	.159	.170	.174
Sample Size=25	std. error	.074	.099	.113	.121	.123
	conf. intvl.	.153	.204	.233	.250	.255
Missouri						
design effect	1.233					
Sample Size=50	std. error	.048	.063	.073	.078	.079
	conf. intvl.	.096	.128	.146	.156	.159
Sample Size=25	std. error	.068	.091	.104	.111	.113
	conf. intvl.	.140	.187	.214	.229	.234

[a] Design effect due to unequal weighting.
[b] This number represents the half-width of a 95% confidence interval.

N.B. These figures are based on the currently available data. At the patient level, the variance of estimates must account for the unequal weights across patients, as well as the clustering effect of the multi-stage design (for patients selected within non-certainty agencies). Table C.2 shows the standard errors and 95 percent confidence intervals around estimates using all 250 patients in a state, and around estimates from a subgroup of half these patients. For an estimated proportion of about .2 made from all 250 patients selected in Colorado, the standard error would be about .034, and the confidence interval would be .2 ± .067.

Table C.2 Precision of Patient-Level Estimates

	proportion=	.1	.2	.3	.4	.5
	or	.9	.8	.7	.6	.5
Colorado						
Sample Size=250	std. error	.026	.034	.039	.042	.043
	conf. intvl.[a]	.050	.067	.077	.082	.084
Sample Size=125	std. err.	.036	.048	.055	.059	.060
	conf. intvl.	.071	.095	.109	.117	.119
Massachusetts						
Sample Size =250	std. error	.028	.037	.043	.046	.047
	conf. intvl.	.055	.074	.084	.090	.092
Sample Size=125	std. error	.040	.053	.061	.065	.066
	conf. intvl.	.079	.105	.120	.129	.131
Missouri						
Sample Size=250	std. error	.026	.035	.040	.043	.043
	conf. intvl.	.051	.068	.078	.084	.086
Sample Size=125	std. error	.037	.049	.056	.060	.061
	conf. intvl.	.073	.097	.111	.119	.121

[a] This number represents the half-width of a 95% confidence interval.

N.B. These figures are based on the currently available data.

When analyzing data resulting from a complex sample design, it is important to account for the design when calculating the variance of an estimate. Because of the unequal weighting and clustering, a specialized approach (such as Taylor Series or replication techniques) must be used to properly calculate the variances. These techniques are available in statistical packages such as SUDAAN and Stata.

2. Procedures for the Collection of Information

There will be three mailings to the home health agencies. The first will contain an advance letter that explains the purpose of the study and provides MPR's toll-free telephone number for agencies that have questions, a contact person, and information about the monetary compensation for participation. This letter will be mailed first class and sent one week before the first questionnaire mailing. A week later, a second mailing

will be sent, containing a cover letter, the questionnaire, and a return FedEx envelope. The cover letter will contain content similar to that of the advance letter and will address confidentiality concerns. We will send a second questionnaire mailing one month after the first questionnaire mailing. The cover letter will be modified to address issues of nonresponse and will include letters, if available, from the appropriate state and national home health associations. Both questionnaire mailings will be sent by priority mail. Agencies that return completed questionnaires will receive $50 in compensation for their time.

After questionnaires have been received and entered into the system, they will be quickly routed for data entry. We will develop procedures for identifying critical items and conducting follow-up calls to collect missing or inconsistent information. We will begin the telephone phase two weeks after the second questionnaire mailing. This phase will consist of two major activities:

(1) locating telephone numbers for HHAs, and (2) making telephone reminder calls to the agencies who have not return the mail instrument. We estimate that about 40 percent of the initial sample will be referred to the telephone phase for item follow-up and reminders. When we begin the telephone phase, we will update the telephone sample daily to eliminate, as much as possible, calls to respondents who have completed mail instruments.

3. Methods to Maximize Response Rates and Deal with Nonresponse

To assure the validity of the response to the survey, the survey seeks a response rate of 80 percent. To achieve the highest possible response, we will undertake the following

steps for the collection of information:

1. ***Pretesting.*** We pretested the instrument to assure that the language, questions, pathing, and format are readily comprehensible to the targeted population. Respondents are much less apt to refuse a question or, indeed, the entire questionnaire, when they can understand the task they are being asked to accomplish. We conducted this limited pretest of the entire questionnaire with 9 agencies resembling the targeted population.

1. 2. ***Support of home health associations.*** A key to achieving this response rate will be the support of state and national home health associations. We discussed this with the state associations and asked them to provide a letter of support. We have provided the home health associations an opportunity to comment on the survey instrument.
2. 3. ***Data collection materials.*** The focus of all respondent materials (advance and refusal conversion letters) will be to secure respondent cooperation through the clarity, simplicity, and thoroughness of the materials.
3. *4.* ***Data collection methods.*** Section C.2 described several data collection techniques that will minimize nonresponse: sending advance letters, offering an incentive, and calling agencies who have not returned completed questionnaires.

We will also assess and address any nonresponse in the postsurvey analysis phase. If the agencies which fail to respond to a survey would have provided systematically different answers from those who do respond, then the survey estimates obtained only from respondent data will be biased. We will calculate adjustments to the sampling weights to compensate for such bias.

Finally, survey data collected for this evaluation are subject to item nonresponse. Item nonresponse occurs when the beneficiary participates in the survey but is unable or unwilling to answer all the questions. Upon receipt of a completed survey, project staff will check the survey against a list of critical data elements. Any uncompleted items will trigger a call to the agency to obtain or clarify the missing/incomplete data. Remaining item non-response will be handled by including a missing category in the analysis.

4. Tests of Procedure or Methods to be Undertaken

To estimate completion time and uncover problems in questionnaire wording and logic, we conducted a pretest of the mail questionnaire with nine respondents. These respondents were representative of the full study sample. No significant problems were uncovered.

5. Persons Consulted on Statistical Aspects and Persons Collecting and/or Analyzing Data

Mathematica Policy Research, Inc. (MPR) is conducting this study, including collecting and analyzing the survey data, under contract to CMS (contract no: CMS 500-00-0033 (06)). The MPR person responsible for the statistical aspects of the sample design is Barbara Lepidus Carlson (609-275-2374). Karen CyBulski of MPR (609-936-2797) will direct the data collection effort. Valerie Cheh of MPR (609-275-2385) is the project director with overall responsibility for the project, and will lead the analysis. Ann Meadow of CMS (410-786-6602) is the technical Project Officer for the task order

.