Responses to 12-01-08 OMB Inquiries re: MCPSS (CMS-10097)

1.   Attachment 5 says that the stepwise regression was limited to section A and C because they were the most highly corrected with overall satisfaction. What were the correlations between the other sections and the overall satisfaction score?

>   We reviewed coefficients from a variety of regression models. Below are examples of the coefficients from the various survey sections (regressed on "overall satisfaction"):

	Range of coefficients across models (2008 data)	
	Low	**High**
Section A (Provider Inquiries):	0.53 to	**0.73**
Section B (Outreach & Education):	0.04 to	**0.08**
Section C (Claims Processing):	0.09 to	**0.38**
Section D (Appeals):	0.03 to	**0.07**
Section E (Provider Enrollment):	0.07 to	**0.10**
Section F (Medical Review):	0.07 to	**0.10**
Section G (Audit & Reimbursement):	0.04 to	**0.08**

2.   The AMA comment letter took issue with the way CMS depicts the level of satisfaction providers have with the contractors they deal with. When CMS makes statements like "In general, Medicare providers are highly satisfied with their contractors," what is this based on? Is this based on the results of this survey? The supporting statement currently implies that the results from this survey are not publicly reported but are, rather, used internally by CMS and contractors to improve their services. Please clarify.

>   The statement is based on the results of the survey. In 2008 the national average was 4.51 on a scale of 1 to 6, where 1 is not at all satisfied and 6 is completely satisfied. Scale points 2 through 5 are not labeled in the survey questionnaire. Both qualitative research done as part of the MCPSS contract and literature on use of such scales suggest that survey respondents do not perceive the intervals between the scale points as equal; respondents tend to use the upper part of the scale to make finer distinctions. Therefore, it is not appropriate to transform the scale directly to another metric, such as a 100-point scale with the traditional letter grade breaks.

>   MCPSS results have always been reported to the public. These reports include scores for each contractor as well as other relevant analysis. The public reports are available on
> https://www.mcpsstudy.org/srvy/rptsdocs.asp

These public reports provide high-level results from the survey. Results from more detailed analysis (including modeling of overall satisfaction) are provided in separate individual customized reports to CMS and the respective Contractors to improve the quality of services offered to providers.

OMB follow-up question: supporting statement part B says that reports provided to contractors provide information on cell sizes and standard errors for summary scores at all levels. Is there a reason why CMS has not provided this information for the reports released to the public? The report issued for 2008 (http://www.mcpsstudy.org/srvy/rptsdocs.asp) does not appear to include this information.

The target audience for the public report is health care providers, rather than researchers. We wanted to provide a brief summary with the information most important to them, the scores.

Also, the 2008 report states that a "complete survey" was defined as a survey where three core questions were answered. This ICR packages states that the three core questions is a change for the 2009 data collection and report. Please clarify. Was a revision request submitted to OMB prior to revising the definition of a "completed survey" or is this the first time approval has been sought?

The current OMB approval package is the first time the revised definition of a "complete survey" is submitted to OMB. The revised definition of "complete" is more stringent, than the previous OMB reviewed definition (three core questions vs. the previous two, only one of which was core). We performed regression analyses, to assess the implications of using core items, specific to contractor type, to define a completed survey and decided to incorporate the revised definition into the 2008 response rate year, publicly noting that we made a change in the definition of a "complete survey" in our analysis.

The 2008 report also states that the response rate was 70%. Did CMS conduct a non-response bias analysis? Does CMS provide this information to the public?

Yes, a nonresponse bias analysis was completed. It is included as an attachment at the end of this response.

This information was not made available to the public this year or in the past years. Again, our goal was to provide a brief, accessible summary, primarily for health care providers rather than researchers.

3.   If CMS is most interested in the "core questions," why not limit the survey questions to these items? Conversely, if all of the questions are

equally important for purposes of providing a composite score, why is a survey considered "complete" if only 3 core questions are responded to?

OMB has approved the following definition of a completed survey: Any one item in section C (claims processing) plus any other item in any other section of the survey. The result is that only two items were needed to consider the survey a complete. The new definition is now based on the actual data and seeks to focus only on those items that are the most important drivers of satisfaction.

Why only three items? In a series of regression analyses it was determined that three items explained the lion's share of variance and additional items had very little impact on the model. Example: For example, in the models for the Carriers, the first three variables entered into the model (A7, C1 and A5) explained 53.7% of the variance. Adding up to 7 other variables added only 2.9% to the total variance explained.

One reason that sections A (inquiries) and C (claims) are the biggest drivers of satisfaction is that these are business functions that all Medicare providers have experience with. Not all providers have an Appeal or Medical Review (for example) – so analytically these survey items are unlikely to be strongly correlated with overall satisfaction simply because not all providers ever use these business functions. However, while these functions are not highly correlated with overall reported satisfaction (at the aggregate level), these functions are still important and CMS must be able to monitor the Contractors and be able to provide them feedback for performance improvement.

4.   How does CMS handle item non-response?

Item nonresponse is ignored – data are not imputed.   Scores are based on aggregate data – the numerator of a section score is based on the aggregated responses to all items in the survey section, divided by the number of "valid" responses in that section. (Where valid is an answer on the 1-6 scale – responses of "don't know", "not applicable" and "refused" are excluded from both the numerator and denominator.)

5.   Please clarify what happens when a "core question" is skipped. If a core question is skipped, is the respondent merely reminded of the importance of the question? Is the whole survey thrown out as "incomplete"? Is the respondent not allowed to move onto the next question until the core question has been asked? Is there follow-up for respondents who leave the "core questions" blank?

Respondents are able to skip any and all questions in the web or paper surveys. Also, in the telephone survey a respondent may refuse to answer an item, or the respondent can indicate that they do not have sufficient information to answer the question (either they don't know or the item does not apply to them).

If we receive a web or paper survey where the respondent has skipped one or more "core question" we will have a telephone interviewer contact the respondent to ask the respondent any missing core items. If the telephone prompt is not successful, and core items remain missing, then the particular survey is considered incomplete - the survey is not counted in the numerator of the response rate and any data from the survey are excluded from the final analytic dataset.

OMB follow-up: This seems to run counter to what Attachment 5 says. Attachment 5 says "we do not recommend requiring non-missing data on all core items because it might jeopardize use of data from other items of the questionnaire. For example, the vast majority of those that are missing at least one core item did provide satisfaction data on at least 4 other items in sections A and C." Please clarify.

By "missing" we mean <u>entirely</u> missing – versus a response of "don't know" / "does not apply". So if the respondent indicates that he/she cannot answer because the item does not apply to him/her (they have not had any experience), then we will count this as "answered" (for purposes of defining a complete). Only if the core question is left entirely blank we call back to try to get a response. If the followup call is not successful then the survey is then considered incomplete (and the data is not included in the final analysis).

6.   Does prompting respondents to select a response, when they have previously indicated "don't know" or "refused to answer," inject bias into the responses? Does it make sense to limit the prompts to only those respondents who simply skipped the question altogether?

We only prompt answers that are entirely missing (that is, the respondent did not indicate that they did not know the answer, nor did they indicate that they refused to respond).

OMB follow-up: This seems to run counter to what Attachment 5 says. Attachment 5 says "If a respondent answers don't know, refuse or skips one of these items, there would be a follow-up screen that would re-ask the question. The question would emphasize the importance this item has for evaluating the contractor. No such follow-up would be made if the respondent reports the item is not applicable." Please clarify.

Attachment 5 is a "White Paper" with recommendations that were included with a prior submission. We decided not to implement this

particular recommendation, in part for the reason you suggest. The actual methodology that was used in 2008 is as follows:

-- If a respondent indicated they lack experience with the item to report a rating or otherwise indicated that they considered the item and did not provide a response, then we deemed that they have had an opportunity to report on the item and that further probing would be inappropriate .

-- If an item remained entirely blank, then we have an interviewer call the provider and attempt to complete the core items.

7.   What happens when respondents complete a paper version of the survey? How are they provided the same prompts as respondents who take the online survey or the interviewer-facilitated survey?

As mentioned in #5 above, any respondents submitting incomplete web or paper surveys receive a prompt from a telephone interviewer in an attempt to collect responses on all core survey items.

8.   B-17 on supporting statement A says that the OMB control number and expiration dates will not be printed on the forms. Please explain why the collection does not lend itself to the displaying of an expiration date.

The OMB control number is printed on the paper copy of the survey as well as the Web application. The expiration date has not printed, as the respondent may interpret that as the survey deadline.

OMB follow-up: OMB generally does not provide an exemption for displaying the expiration date for these types of reasons. Almost every survey administered by the federal government carries the expiration date, and this practice has not generated confusion about the survey deadline. What specific evidence does CMS have that the OMB expiration date would be confusing? To the extent that there is confusion, has CMS tried to include information (e.g. in the standard PRA blurb) that clarifies the difference between the survey deadline and the OMB expiration date?

We will add an OMB expiration date.

9.   On a number of question items, there appear to be secondary questions in boldface. (For example, item B11 starts with the main question and then has a secondary question, "What did 'communication with you' mean to you?"). What is the purpose of these questions and where/how are the respondents supposed to respond to them

The actual survey submitted to OMB do not have secondary questions. In a prior submission we included a copy of an example of a cognitive interview protocol - in this protocol we included examples of probes that might be asked of cognitive interview respondents.

Additional OMB Questions:

- How much missing data has CMS experienced previously?

  There are two main types of missing data in the MCPSS:

  Type 1: Missing due to an item and/or survey section not applying to a provider

  Type 2: Other missing (such as responses that are left entirely "blank" in the web survey)

  Certain survey sections (and items) have extremely high Type 1 missing -- an example is the Appeals section (this section would apply only to those providers who have appealed a claim denial).

  Generally the MCPSS has very low Type 2 missing data. Examples (from the 2008 data) are provided below:

  HOW SATISFIED ARE YOU WITH:

  A1 HOW QUICKLY YOU CAN REACH A REPRESENTATIVE TO MAKE A PROVIDER INQUIRY BY TELEPHONE
   Response 98.59
   Type 1 Missing 1.08
   Don't know 0.22
   Type 2 missing 0.09

  A2 RECEIVING THE CORRECT INFORMATION
   Response 98.98
   Type 1 Missing 0.66
   Don't know 0.2
   Type 2 missing 0.15

  C3 ACCURACY OF REMITTANCE ADVICES RECEIVED FROM YOUR CONTRACTOR
   Response 97.31
   Type 1 Missing 0.66
   Don't know 0.28
   Type 2 missing 1.75

- Does it expect missing data levels to remain at that level?

  We expect items missing to remain generally constant over time.

6

- Has CMS analyzed missing data to determine whether it is clustered by item or by respondent type?

  The Type 1 missing data are clustered by respondent type, only because certain activities may only apply to one set of respondents (e.g., section G -- Audit and Reimbursement -- only applies to the FI and RHHI providers ; or as noted above the appeals section would apply only to those providers who have appealed a claim denial).

  We have not seen evidence of clustering of the Type 2 missing data.

- Has CMS explored reasons for any such patterns?

  The patterns are as anticipated (because the nature of the services provided differs somewhat depending on the provider group / respondent).

<u>ATTACHMENT</u>

ANALYSIS OF POTENTIAL NONRESPONSE BIAS AND

WEIGHT ADJUSTMENTS TO REDUCE IT

ANALYSIS OF POTENTIAL NONRESPONSE BIAS AND

WEIGHT ADJUSTMENTS TO REDUCE IT

Nonresponse in surveys creates a potential for bias in the survey estimates. If response propensity is independent of substantive variables, e.g., satisfaction scores in this survey, then no bias would arise in the survey estimates. To reduce any potential bias, the sampling weights are usually adjusted for nonresponse and it is expected that the weighted estimates produced by these adjusted weights, will have much reduced bias, if any.

There are several methods to adjust the sampling weights for nonresponse. We used a response propensity method. The objective was, using the known characteristics of both respondents and nonrespondents, to identify subgroups of provider population within which the response propensity is independent of satisfaction. Thus, we attempted to form homogeneous subgroups with respect to response propensity using statistical modeling software[1]. After the subgroups were identified, the weights were adjusted for nonresponse within these subgroups.

We first provide a summary of the goals of the survey, its target population, and sample design. Then, we discuss the achieved response rates. In the final section, we describe the nonresponse adjustments applied to the sampling weights.

Summary of Sample Design and Data Collection Methods

The goal of the 2008 national implementation was to collect quantifiable data on provider satisfaction with the performance of Medicare Fee-for-Service (FFS) Contractors. The target population for the 2008 national implementation consisted of all Medicare providers served by ? different Medicare FFS Contractors. Some of the contractors provided more than one type of service. Thus, the total number of contractors by different contractor service types was 47. These Contractors were comprised of 21 Fiscal Intermediaries (FIs), 18 Carriers, four Regional Home Health Intermediaries (RHHIs) and four Durable Medical Equipment Contractors (DME-MACs).

The goal of sample design was to obtain valid and reliable estimates at contractor level and to conduct statistical tests for the differences in the mean satisfaction scores between the contractors. The targeted sample size was 400 completes for each contractor. The contractor sample sizes were allocated proportionately across the provider types. A small number of provider type strata and several jurisdiction and state strata within

---

[1]Göksel, H., Judkins, D.R., Mosher, W.D. (1992). Nonresponse adjustments for a telephone follow-up to a national in-person survey. *Journal of Official Statistics, Statistics Sweden,* 8(2).[?]

contractors had to be over-sampled to attain the target of a minimum number of 30 completes in each stratum.

The provider records were further stratified implicitly by sorting the records by additional provider characteristics. A sample of providers was drawn with equal probability and systematically within each state, jurisdiction, and provider type stratum within contractors.

Although Web was the primary mode of data collection, the 2008 national implementation was a multimode study. Initially, each sampled provider received a survey notification packet in the mail which provided information about the MCPSS and instructions on how to access and complete the online survey instrument. Providers also had the option to request a paper copy of the survey instrument any time during the study, and could mail or fax back their completed survey instruments. Westat followed up by telephone with providers who did not complete the Web survey or paper copy.

Regardless of the mode of data collection, all versions of the survey instrument contained the same questions, presented the questions in the same order. The survey instrument covered seven key areas of the interface between the providers and their Medicare FFS Contractors: provider inquiries, provider education and outreach, claims processing, appeals, provider enrollment, medical review, and provider audit and reimbursement. All the service areas were not relevant for all Medicare FFS Contractors. The survey instruments were hence designed to only inquire about the relevant services rendered by the Medicare FFS Contractor to their providers.

Response Rates

The 2008 national implementation achieved a final survey response rate of 69.6 percent. Table 1 shows the sample disposition and unweighted response rate and Table 2 shows the comparison of unweighted and weighted response rates by contractor types.

Table 1: 2008 National Implementation MCPSS Summary Sample Disposition

Total Sample	**35,886**
Completed Surveys	**20,251**
Partially Completed Surveys	**577**
Refusals	**1,359**
Ineligibles	**1,082**
Bundles	**5,163**
Other Non-Response	**4,535**
Unknown Eligibility	**2,919**
Response Rate	**69.6%**

Table 2: Unweighted and Weighted Response Rates by Contractor Type

	Response Rates	
Contractor Type	**Unweighted**	**Weighted**
FI	71.2	71.1
Carrier	65.6	66.1
RHHI	78.7	78.8
DME-MAC	77.6	76.6
Overall	**69.6**	**67.0**

The unweighted response rate was calculated using the following formula:

$$\frac{Completes}{Completes + Partial\ Completes + Refusal + Other\ Nonresponse + ((Unknown\ Eligibility) * Eligibility\ Rate)}$$

where Eligibility Rate was calculated as:

$$\frac{Completes + Partial\ Completes + Refusal + Other\ Nonresponse}{Completes + Partial\ Completes + Refusal + Other\ Nonresponse + Ineligible + Bundles}$$

The disposition categories listed in the above formulae were defined as follows:

- Completed surveys are cases where the respondent provided a survey response to at least one item in section C "Claims Processing" and at least one item in any other survey section.

- Partially completed surveys are cases where the respondent did not provide a survey response to <u>any</u> items in Section C "Claims Processing", but did provide a survey response to items in another or other sections.

- Refusals are cases where a respondent declined to participate in the 2008 national implementation study; thus the respondent was unwilling to provide any survey response to any survey items.

- Other Nonresponse cases are where we located correct contact information, but wasn't able to establish contact with the provider (e.g., ring no answers, answering machines, busy singles, etc.)

- Ineligibles are cases where:

  □ A respondent did not fit the eligibility criteria (e.g., has not had a Medicare claim in the past year); or
  □ A respondent is out of scope of the study (e.g., the facility has closed or its contract terminated).

- Bundles are cases where a respondent is affiliated with multiple facilities. If the respondent completes a single survey to represent multiple facilities, then all other facilities are linked to this completed survey;

- Unknown Eligibility are cases where:

  □ We had a telephone contact number, but was unable to communicate with the respondent due to language issues; or
  □ We didn't have any correct contact information (i.e., phone number nor mailing address information) available to use to contact the respondent. These cases are also known as nonlocatables.

The weighted response rate takes into account the effect of differential sampling rates. It also adjusts for multiple provider facilities that were associated with some of the satisfaction score reporting units in the survey.

Nonresponse Adjustments

The sampling weights were adjusted to reduce any potential bias caused by not obtaining a completed survey instrument from all the sample providers. To reduce this potential bias, a separate adjustment factor was computed in each nonresponse adjustment cell. A separate set of nonresponse adjustment cells were formed within each Contractor.

If response propensity is independent of satisfaction and other substantive variables within nonresponse adjustment cells, then nonresponse-adjusted weights yield unbiased estimates. There are several alternative methods of forming nonresponse adjustment cells to achieve this result. We used Chi-Square Automatic Interaction Detector (CHAID) software (SPSS, 1993[2]) to guide us in forming the cells. CHAID partitions data into homogenous subsets with respect to response propensity. To accomplish this, it first merges values of the individual predictors, which are statistically homogeneous with respect to the response propensity and maintains all other heterogeneous values. It then selects the most significant predictor (with the smallest p-value) as the best predictor of response propensity and thus forms the first branch in the decision tree. It continues applying the same process within the subgroups (nodes) defined by the "best" predictor chosen in the preceding step. This process continues until no significant predictor is found or a specified (about 20) minimum node size is reached. The procedure is stepwise and creates a hierarchical tree-like structure.

We developed two separate models (and thus a separate set of adjustment cells) to predict (1) propensity of determining eligibility among all sample cases, and (2) propensity of response among the eligible providers. The cases with undetermined eligibility included mostly nonlocatables. We believe that the provider characteristics influencing locatability and response after the provider is identified as eligible can be quite different.

All sample providers were classified into four major survey disposition categories based on the outcome of the survey. These categories were (1) respondent --completed instruments, (2) eligible nonrespondent (including refusals, other nonresponse, and partial completes), (3) ineligible (ineligibles and bundles), and (4) unknown eligibility (mostly nonlocatables).

---

[2] SPSS (1993), SPSS for Windows: CHAID, Release 6.0, User's Guide, Jay Magidson/SPSS Inc., 1993.

Variables employed in forming nonresponse adjustment cells had to be known for both responding and non-responding providers. The following variables were used as potential predictors in modeling response propensity with CHAID: Contractor, provider type stratum, Primary A/B MAC jurisdictions (for FIs and Carriers), Specialty MAC jurisdictions (for RHHIs and DME-MACs), State, Regional Office, claims size categories.

After creating two separate sets of adjustment cells, we carried out separate weight adjustments to compensate for those providers with unknown eligibility than for those nonresponding eligible providers. The weight adjustment factor for undetermined eligibility, within each adjustment cell, was computed as the ratio of the weighted (by the base weight) total number of sampled providers to the weighted number of providers, whose eligibility could be determined. This adjustment assumes that the rate of eligibility among the cases with unknown eligibility is the same as among the cases with known eligibility within each adjustment cell. The nonresponse adjustment factor was computed as the ratio of the weighted (after adjusting for undetermined eligibility) number of eligible (responding plus eligible nonresponding) providers to the responding providers within each nonresponse adjustment cell.

Although nonresponse adjustment can reduce bias, at the same time, it may increase the variance of estimates. Small adjustment cells and/or low response rates (or large nonresponse adjustment factors) may increase the variance and give rise to unstable estimates. In order to prevent an unduly increase in variance and thereby an adverse effect on the mean square error of the estimates, we attempted to limit the size of the smallest cell to a minimum and avoid large adjustment factors. Next, we discuss each weight adjustment in detail and present their formulae.

Adjusting the Weights for Cases with Unknown Eligibility

First, the weights were adjusted to compensate for cases with unknown eligibility. The adjustment factor for the adjustment class c (λ_c) was computed as:

$$\lambda_c = \frac{\sum_{i \in S_{1c}} W_{ci}^B + \sum_{i \in S_{2c}} W_{ci}^B + \sum_{i \in S_{3c}} W_{ci}^B + \sum_{i \in S_{4c}} W_{ci}^B}{\sum_{i \in S_{1c}} W_{ci}^B + \sum_{i \in S_{2c}} W_{ci}^B + \sum_{i \in S_{3c}} W_{ci}^B}$$

where,

S_{1c} is the set of responding cases (completes) in adjustment class c,

S_{2c} is the set of eligible nonresponding cases in adjustment class c,

S_{3c} is the set of ineligible cases in adjustment class c,

S_{4c} is the set of sampled cases with undetermined eligibility in adjustment class c,

W_{ci}^B is the base weight for provider record i in the adjustment class c.

Then, the weight adjusted for eligibility unknown cases for sampled record i in adjustment class c, (W_{ci}^U), was computed as:

$$W_{ci}^U = W_{ci}^B \times \lambda_c$$

Adjusting the Weights for Nonresponding Eligible Providers

After forming the nonresponse adjustment cells, the weights were adjusted to compensate for the eligible nonresponding providers. The nonresponse adjustment factor for cell α, δ_α was computed as:

$$\delta_\alpha = \frac{\sum_{i \in S_{1\alpha}} W_{\alpha i}^U + \sum_{i \in S_{2\alpha}} W_{\alpha i}^U}{\sum_{i \in S_{1\alpha}} W_{\alpha i}^U}$$

where,

$S_{1\alpha}$ is the set of responding providers (completes) in adjustment class α,

$S_{2\alpha}$ is the set of eligible nonresponding providers in adjustment class α.

$W_{\alpha i}^U$ is the weight adjusted for unknown eligibility cases for provider i in adjustment class α.

Then, we computed the final weight as the product of the weight that was adjusted for the unknown eligibility and the nonresponse adjustment factor. The final sample weight for provider i in nonresponse adjustment class α, $W_{\alpha i}^F$, was computed as follows:

$$W_{\alpha i}^{F} = W_{\alpha i}^{U} \times \delta_{\alpha}$$