OFFICE OF
THE INSPECTOR GENERAL

SOCIAL SECURITY ADMINISTRATION

PERFORMANCE MEASURE REVIEW: RELIABILITY OF THE DATA USED TO MEASURE PUBLIC KNOWLEDGE OF THE SOCIAL SECURITY ADMINISTRATION

February 2002 **A-02-01-11015**

EVALUATION REPORT

Mission

We improve SSA programs and operations and protect them against fraud, waste, and abuse by conducting independent and objective audits, evaluations, and investigations. We provide timely, useful, and reliable information and advice to Administration officials, the Congress, and the public.

Authority

The Inspector General Act created independent audit and investigative units, called the Office of Inspector General (OIG). The mission of the OIG, as spelled out in the Act, is to:

- Conduct and supervise independent and objective audits and investigations relating to agency programs and operations.
- Promote economy, effectiveness, and efficiency within the agency.
- Prevent and detect fraud, waste, and abuse in agency programs and operations.
- Review and make recommendations regarding existing and proposed legislation and regulations relating to agency programs and operations.
- Keep the agency head and the Congress fully and currently informed of problems in agency programs and operations.

To ensure objectivity, the IG Act empowers the IG with:

- Independence to determine what reviews to perform.
- Access to all information necessary for the reviews.
- Authority to publish findings and recommendations based on the reviews.

Vision

By conducting independent and objective audits, investigations, and evaluations, we are agents of positive change striving for continuous improvement in the Social Security Administration's programs, operations, and management and in our own office.

SOCIAL SECURITY

MEMORANDUM

Date: February 7, 2002 Refer To:

To: Jo Anne B. Barnhart
Commissioner

From: Inspector General

Subject: Performance Measure Review: Reliability of the Data Used to Measure Public Knowledge of the Social Security Administration (A-02-01-11015)

Following consultations with congressional committees, the Office of the Inspector General agreed to review the Social Security Administration's (SSA) performance indicators over a continuous 3-year cycle. We recently completed our first 3-year cycle. In conducting this work, we used the services of an outside contractor, PricewaterhouseCoopers (PwC), LLP, to assist us in our efforts.

For this report, we used PwC to conduct the review of one of the Agency's performance indicators related to the public's knowledge of SSA. The objective of the review was to assess the reliability of the data used to measure the level of public knowledge of SSA.

Please comment within 60 days from the date of this memorandum on corrective action taken or planned on each recommendation. If you wish to discuss the final report, please call me or have your staff contact Steven L. Schaeffer, Assistant Inspector General for Audit, at (410) 965-9700.

James G. Huse, Jr.

Attachment

PRICEWATERHOUSECOOPERS

Evaluation of Selected Performance Measures of the Social Security Administration:

Reliability of the Data Used to Measure Public Knowledge of SSA

Office of the Inspector General
Social Security Administration

INTRODUCTION

To evaluate the 11 performance indicators identified by the Social Security Administration (SSA) in its Fiscal Year (FY) 2001 Annual Performance Plan (APP), PricewaterhouseCoopers, LLP (PwC) was contracted to determine whether:

- Performance data used to calculate measured values for comparison to SSA's performance goals were complete, accurate, reasonable, consistent, and reliable;
- Performance data was properly calculated and reported;
- Key automated and manual internal controls related to systems and processes used to calculate and measure performance were adequate; and
- Performance measures were meaningful and in compliance with the Government Performance and Results Act of 1993 (GPRA).

This report is one of five separate stand-alone reports, corresponding to the following SSA process and performance measure (PM):

- Percent of public who are knowledgeable about Social Security programs (PM #11)
  FY 2000 Goal: 65 percent

This report reflects our understanding and evaluation of the process related to PM #11. To achieve its strategic goal "To strengthen public understanding of Social Security programs" SSA has developed several strategic objectives. One of these objectives is, "By 2005, nine out of ten Americans will be knowledgeable about the Social Security programs in five important areas:"

- Basic program facts;
- Financial value of programs to individuals;
- Economic and social impact of the programs;
- How the programs are financed today; and
- Financing issues.

One of the performance indicators cited in the plan is "Percent of public who are knowledgeable about Social Security programs." This indicator will be considered achieved if 65 percent of the public surveyed are knowledgeable about Social Security programs. SSA's FY 2001 APP contains one performance indicator developed to meet this objective as follows:

- *Percent of public who are knowledgeable about Social Security programs* - This indicator will be considered achieved if 65 percent of the public surveyed obtain a passing score on the survey

We performed our testing from September 21, 2000 through February 15, 2001. Our engagement was limited to testing at SSA's headquarters in Woodlawn, Maryland. The procedures that we performed were in accordance with the American Institute of Certified Public Accountants' Statement on Standards for Consulting Services, and are consistent with appropriate standards for performance audit engagements in *Government Auditing Standards* (Yellow Book, 1994 version). However, we were not engaged to and did not conduct an audit, the objective of which would be the expression of an opinion on the reliability or accuracy of the reported results of the performance measures evaluated. Accordingly, we do not express such an opinion. Had we performed additional audit procedures, other matters might have come to our attention that would have been reported to you.

BACKGROUND

This indicator has been created to measure the percent of the public who are knowledgeable about Social Security programs. The goal during FY 2000 is to have 65 percent of the public knowledgeable about SSA programs. SSA measures public understanding by conducting an annual survey. Below is an overview of the Public Understanding Measurement System (PUMS) survey.

Survey Objectives

PUMS is an annual survey conducted by SSA and its contractor, designed to determine what percent of the public is knowledgeable about SSA and the services it provides. The first year of the study established a baseline knowledge indicator that has been used in following years to track changes in the public's knowledge of SSA and its programs. Specifically, this study aims to measure:

- To what extent is the public's knowledge based on public education programs and products developed by SSA?
- What sources does the public use to obtain information on Social Security?
- How does the public prefer to receive information about Social Security?
- How easy and useful are the Social Security Statements?

Although not the data source for this Performance Measure, an additional study, Moving the Needle (MTN), was performed to further evaluate the effectiveness of various forms of public education and outreach efforts in raising public awareness and knowledge of Social Security. SSA anticipates using this information to design annual public education programs, which would target specific knowledge or performance gaps. Ultimately, it is expected that results from this survey will assist SSA in achieving the target goal that at least 90 percent of the public will be knowledgeable about SSA and its services by 2005.

Sample Design

The PUMS surveys have employed a stratified probability sampling design, where strata are defined based on SSA regions. Specifically, list-assisted Random Digit Dialing

(RDD) samples are selected to secure a minimum of 400 interviews in each of the 10 SSA regions.

The sampling design for the MTN involved a 2-stage process. In the first stage, 32 geographic areas were selected, reflecting population groups with the greatest potential for improvement with respect to knowledge of SSA services. These 32 areas were then paired based on demographic information to form 16 pairs of control-treatment groups. Upon consultation with SSA, eight of these pairs were retained, from which RDD samples of households were selected for participation in the survey. Prior to the survey administration, various methodologies were used to increase the knowledge of the public regarding the SSA services in each of the eight test areas.

Questionnaire Design

The PUMS questionnaire was developed with collaborative efforts from SSA, SSA's contractor, and other experts. The survey instrument consisted of questions dealing with the following three areas of inquiry:

- Public knowledge
- Sources of social security information
- Demographic information

In designing the questionnaire, SSA proposed a set of knowledge metrics that was tested in nine focus groups, which consisted of three age, income, and geographical groups. The Office of Communications (OComm) and Office of External Affairs conducted these focus groups in April of 1998. Consistent with Office of Management and Budget (OMB) restrictions, fewer than 10 respondents participated in each focus group. Following the focus groups, SSA provided all related materials to its contractor for their review. Senior researchers at the National Academy of Social Insurance also reviewed the measures.

Once essential knowledge indicators were identified, a scoring process was established to capture the trend in the public's knowledge. Specifically, it was decided to assign an equal measure of importance to each of the 23 awareness questions, with anyone scoring a 70 percent or higher identified as "knowledgeable." It should be noted that 1 of these 23 questions could secure 4 possible points, making the total possible points be 26. Ultimately, however, only 19 of the questions were used to measure knowledge, as 4 questions, including the 1 with a 4-point possible score were eliminated from the calculations. The following table, reproduced from the contractor's report[1], summarizes the evolution of the knowledge metric. Note that the contractor report did not explicitly define the question types shown below.

[1] The Gallup Organization, *PUMS Survey Technical Report,* March 11, 1999, page 11.

Table 1. Evolution of the Knowledge Metric for PUMS-I

Question Type	Initial Design, prior to Data Collection	Post Data Reliability Analysis	Further Revisions
Concept questions	22	21	17
Unaided awareness	4	4	0
Aided awareness	18	17	17
Specific factual questions	4	2	2
Total Points	**26**	**23**	**19**

For the most part, the questionnaires for PUMS-I and -II have remained unchanged, with the exception of a new section that was added to the PUMS-II survey. Starting with PUMS-III, however, a number of major changes were introduced for various policy and research issues. For example, 2 of the 19 questions contributing to the knowledge score were eliminated from the questionnaire, while another 3 questions were modified and demoted from the set of questions contributing to the knowledge metric.

The questionnaire used for the MTN survey is very similar to that used for PUMS-II. It contained all of the 19 questions related to the knowledge score, however, a number of other questions have been deleted. In addition, new questions inquiring about different public education and outreach programs by SSA were added to allow assessment of their effectiveness. The following table from the contractor's 1999 report provides a summary of the composition of these questionnaires.

Table 2. Composition of the Knowledge Indicator Questions For PUMS I, II, III, and MTN

Type of Question	Points Possible			
	PUMS-I	PUMS-II	MTN	PUMS-III
Concept questions	17	17	17	13
Unaided awareness	0	0	0	0
Aided awareness	17	17	17	13
Specific factual questions	2	2	2	1
Total Points	19	19	19	14
"Knowledgeable" Cut-off	13	13	13	10

Administration/Data Collection

The PUMS-I survey used a total of 19,283 telephone numbers to secure 4,009 completed interviews. While an 80 percent response rate was targeted, a response rate of only 33 percent was achieved for this survey. Upon conducting a nonresponse analysis, it was concluded that the resulting survey data were not subject to any nonresponse bias. No information has been provided regarding the number of telephone numbers used in each of the other studies, nor have we received any disposition reports that could be used to develop independent estimates of response rates for PUMS-II, III, and MTN.

Due to unavailability of technical reports for PUMS-II or PUMS-III, we cannot comment on whether changes have been introduced with respect to the administration of these

surveys. Starting in November 1999 and ending in January 2000, the MTN study was administered quarterly in 16 communities in Philadelphia, Atlanta, Chicago, and San Francisco. Each quarter 3,000 surveys were conducted, resulting in 12,000 completed interviews. All surveys were conducted in English and Spanish, with field periods as summarized in the following table.

Table 3. Field Periods for PUMS-I, II, III, and MTN Surveys

Data Collection	Start Date	End Date
PUMS-I	October 1998	November 1998
PUMS-II	November 1999	January 2000
MTN	November 1999	September 2000
PUMS-III	October 2000	January 2001

The following table provides a summary of the data collection activities, reflecting the extent of undisclosed information.

Table 4. Disposition Summary for PUMS-I, II, III, and MTN

Survey Characteristics	PUMS-I	PUMS-II	MTN	PUMS-III
	Stratified RDD	Stratified RDD	2-Stage RDD	Stratified RDD
Complete	4,009	4,000	12,000	4,000
Non-Target	2,832	Unavailable	Unavailable	Unavailable
Refusal	2,227	Unavailable	Unavailable	Unavailable
Disconnected	3,576	Unavailable	Unavailable	Unavailable
Total	19,283	Unavailable	Unavailable	Unavailable
Response Rate	33%	25.5%	Unavailable	Unavailable

Analysis and Report Generation

Prior to data analysis, survey data were weighted to project the findings to the population of interest. While improving the demographic representation of the resulting samples, the weighting process has made it possible to develop national estimates by compensating for the different regional sampling rates. As stated earlier, the PUMS-I knowledge metric began with 23 questions (26 possible knowledge points), however, it was reduced to only 19 questions with 19 possible points. The resulting survey data were analyzed using a 2-tiered approach: conceptual and factual knowledge.

Overall, the public performed well on the conceptual measure and poorly on the factual knowledge. Those respondents who could correctly answer 13 of the 19 aided questions were deemed knowledgeable about SSA. Accordingly, 55 percent of the public were estimated to be knowledgeable with the SSA services. Moreover, the following are some of the highlights of the analyses that were performed on the PUMS-I survey data.

- Education, age, income, and receiving a Personal Earnings and Benefits Estimate Statement were the four most important factors determining respondent's knowledge.
- There were no regional differences.

- Minorities, including women, indicated less knowledge of SSA programs compared to other groups.

Again, because there are no technical reports available for PUMS-II or PUMS-III, we cannot comment on whether the knowledge rating for these were calculated in a similar manner. Based on the PUMS-II survey, 57 percent of the public were estimated to be knowledgeable, which is the same rating obtained for the prior year via the PUMS-I survey. At the present time, the results from PUMS-III are not available.

Analogous to the PUMS surveys; the data from the MTN survey have been weighted to represent the demographic composition of the surveyed areas. Subsequently, quarterly estimates of knowledge ratings were calculated for each area and the Nation. According to the MTN survey results, it has been concluded that the public in the test areas has a significantly higher level of knowledge as compared to those in the control areas. This indicates that the additional efforts by SSA positively affect the public's knowledge. The following table summarizes the MTN estimates that have been obtained from the documents available for our evaluation.

Table 5. Knowledge Rating Based on the MTN Survey

Location Type	Percent Knowledgeable	
	Quarter I	Quarter IV
Treatment	58%	63%
Control	56%	56%

RESULTS OF EVALUATION

During the period of September 21, 2000 to February 15, 2001, we evaluated the current processes, systems and controls, which support the FY 2000 SSA performance measurement process. In addition, we determined the accuracy of the underlying performance measure data. Our evaluation of the information provided by SSA management and its contractor allowed us to determine that the preliminary reported FY 2000 results of the performance measure tested (shown below) was reasonably stated based on the methodology used by SSA.

Performance Measure	Reported Result
11. Percent of public who are knowledgeable about Social Security programs.	68 percent[2]

However, we did note the following four opportunities for improvement is SSA methodology:

1. Currently, there are no formal procedures in place to properly reflect the variance inflation due to weighting

2. Multicollinearity among the predictor variables may lead to incorrect results

[2] In its April 2001, *Performance Plan for Fiscal Year 2002,* SSA reported FY 2000 results of 75 percent for this PM.

3. The questionnaire design needs improvement

4. The survey results may be biased due to a significant rate of nonresponse

These items were noted as a result of our testing. We performed an evaluation of the survey methodology, including the sampling and questionnaire designs, data collection procedures, and data analysis and reporting. Because this performance measure is conducted each year, specific attention was given to each annual administration.

1. Currently, there are no formal procedures in place to properly reflect the variance inflation due to weighting.

Although producing the basic estimate of the proportion of the population aware of SSA activities is the primary goal, it is also important to understand how confident SSA can be in that estimate. Statisticians quantify this by measuring the sampling error in such estimates. Further, since weighting often increases sampling errors, use of standard variance calculation formulae with weighted data can result in misleading tests of significance.[3] That is, one might end up declaring significant improvements when the observed change might be attributable to sampling error. SSA stated that no special procedures have been used for variance estimation for the PUMS and MTN surveys. With weighted data, special procedures should be developed and implemented to properly reflect the variance inflation due to weighting.

In the case of complex sampling designs, such as the ones being used in the surveys of interest, research has shown that computed variances of survey estimates may under-represent the induced sampling errors. There are two general approaches for variance estimation for complex sampling designs involving weights. One is linearization, in which a nonlinear estimator (the type used in these two surveys) is approximated by a linear one, and then the variance of this linear proxy is estimated using standard variance estimation methods.[4] The second is replication, in which several estimates of the population parameters under the study are generated from different, yet comparable parts of the original sample. The variability of the resulting estimates is then used to estimate the variance of the parameters of interest.[5]

SSA should consider using one of these two variance estimation methods for future surveys.

2. Multicollinearity among the predictor variables may lead to incorrect results.

In order to identify factors that are highly relevant to increasing the knowledge of people about the services SSA provides, the contractor has used a statistical procedure called step-wise regression. This procedure, which is a special form of the ordinary regression analysis, uses the knowledge score as the dependent variable and a list of demographic

[3] Skinner, C., Holt, D. and Smith, T., Eds. (1989). Analysis of Complex Surveys. Wiley & Sons, NY.

[4] Wolter, K.M. (1985) Introduction to Variance Estimation. New York: John Wiley & Sons.

[5] Effron, B. (1982). The Jack-Knife, the Bootstrap, and Other Resampling Plans, CBMS Regional Conference Series in Applied Mathematics 38, Society for Industrial and Applied Mathematics, Philadelphia.

and other indicators as independent variables. This way, attempts are made to measure the relative importance of each independent variable in explaining the changes in the knowledge scores. Specifically, for each factor a measure is calculated that indicates how important (relevant) that factor is to the knowledge of individuals. However, it can be argued that the employed regression-based approach for this purpose might not be robust enough.

Most statistical procedures, such as regression, require that certain conditions be true for the procedure to perform effectively. All regression analyses involve two sets of variables: a left hand side variable, which is typically referred to as the dependent variable, and the right hand variables, which are typically referred to as independent variables. As the name implies, independent variables are supposed to be independent of each other for the regression model to produce reliable measures of importance for each of the independent variables. When this condition is not met, the results of a regression analysis can be questionable. This common anomaly, which results from the existing multicollinearity among the predictor variables, can lead to unstable results. That is, since the independent variables that have been used in this process are correlated (not independent), small changes can introduce significant fluctuations in the regression coefficients (i.e., the measure of importance for each factor).

There are well known methods to detect, assess and remedy multicollinearity, and we reference the book by Belsley, Kuh and Welsch[6] on this subject.

3. The questionnaire design needs improvement.

Based on our evaluation of the available questionnaires, we have identified a number of potential issues with the structure and wording of the questions, as follows:

- For the MTN questionnaire, the Question 19 series inquires about familiarity with three separate sources of receiving information about the SSA programs: news, public education, and campaigns. Throughout this series, however, inconsistent references are made to these three sources.
- The question series used within the PUMS and MTN questionnaire dealing with usefulness and impact of the SSA statement is based on a screening question (Q24) that simply asks respondents whether they have received the statement or not. That is, the existing sequence of questions does not determine if they have actually read the statement. It is preferable to establish prior reading of the statement, before inquiring about its usefulness.
- The answer categories used within the PUMS and MTN questionnaire to the demographic question series can be improved. For example, the employment status question does not provide an answer category for students and singles out maternity leave from all other types of leaves. This abbreviated set of employment categories can also confuse others, such as those disabled individuals who do not consider themselves unemployed. The marital status question, on the other hand, has a

[6] Belsley, D.A., Kuh, E., and Welsch, R. E., *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity,* John Wiley & Sons, 1980.

negative connotation to the order of its answer categories: divorced, separated, widowed, currently married, and single and never married.

Based on years of rigorous research, the United States Census Bureau has developed answer categories for the demographic type questions. A large number of survey research organizations use these categories when designing questionnaires. In addition to using a set of meticulously tested standards, use of the Census demographic categories enables researchers to use published population figures for weighting of survey data.

4. The survey results may be biased due to a significant rate of nonresponse

Initially, it was anticipated that the response rate to these surveys would be at least 80 percent. However, the secured response rates ranged between 25 and 34 percent. In light of such high rates of nonresponse, it is important to take remedial measures that would increase the response rate to these surveys. A response rate of 25 percent means that 75 percent of the targeted individuals have remained uncovered by the survey. If the group who decided to answer the survey differed, in some substantial way, from those who did not choose to respond, the results obtained could be far from the true proportion in the population as a whole. This uncertainty raises doubts about the credibility or usefulness of the findings of the survey. By weighting the data, SSA has attempted to remove some of the potential bias due to undercoverage; however, the employed methodology should be more outreaching. For example, in residential studies, typically the weighting procedure involves adjustment of the survey data along demographic and socioeconomic dimensions to make the respondent population match more closely these demographic or socioeconomic characteristics in the population as a whole.

According to the Paperwork Reduction Act of 1995 (PRA), Implementing Guide, Chapter VI Section E, samples that suffer from significant nonresponse cannot support valid statistical inferences.

The employed weighting process does not adjust the data with respect to any socioeconomic indicator. It is typical to use income or education as part of the weighting process.

CONCLUSIONS AND RECOMMENDATIONS

Our evaluation found that the reported FY 2000 results of the performance measure tested were reasonably stated. However, our evaluation noted various issues with the 2000 survey. We recommend that SSA take the following corrective actions:

1. SSA should obtain documentation to support the employed methodologies, survey administration protocols, and analysis made.

2. SSA should measure potential sampling errors in the estimates that reflect the employed sampling design, and incorporate the applied weights.

3. In order to establish an importance hierarchy among a set of factors, SSA should test the robustness of the regression methodology it is using to assure there is no multicollinearity. As mentioned above, this can be done by using the Belsley, Kuh and Welsch regression diagnostics. If multicollinearity is detected, we suggest alternative methods of determining the importance hierarchy such as factor analysis and principal components analysis, and we reference the book by Johnson and Wichern[7] on this subject.

4. SSA should consistently use the same set of information sources (news, public education, and campaigns) throughout the questionnaire.

5. SSA should establish prior reading of statements before inquiring about their usefulness.

6. SSA should use the employment categories that are used by the Bureau of Labor Statistics, instead of the list that is currently used with the PUMS III and MTN surveys.

7. SSA should adjust the data with respect to any socioeconomic indicator, and use income, or education as part of the weighting process.

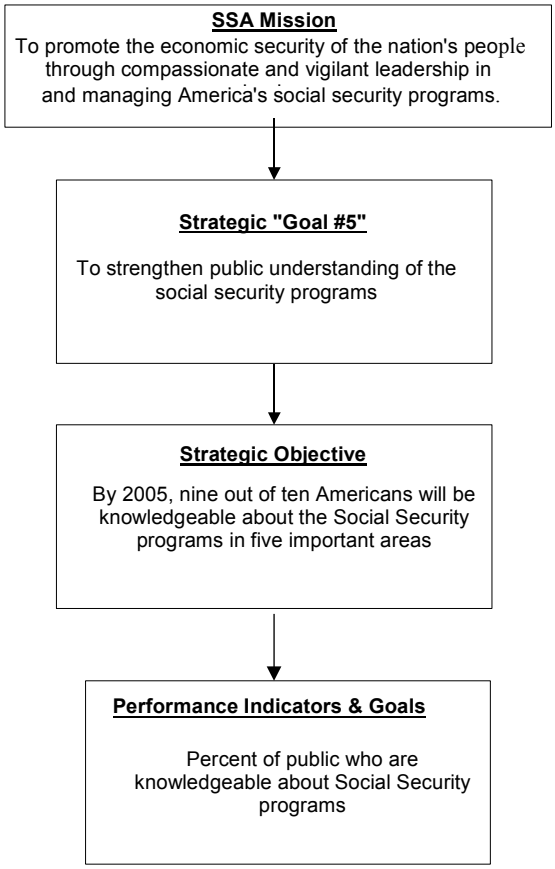
APPROPRIATENESS OF THE PERFORMANCE MEASURES

As part of this engagement, we evaluated the appropriateness of each of the performance measures with respect to GPRA compliance and SSA's APP. We determined whether the specific indicators and goals corresponded to the strategic goals identified in SSA's APP, determined whether each of these indicators accurately measure performance, and determined their compliance with GPRA requirements.

Performance Measure #11 aligns logically with the SSA Strategic Plan but still needs improvement.

The relationship between PM #11 and the applicable SSA Strategic Goal is depicted in the following figure:

[7] Johnson, R. and Wichern, D., *Applied Multivariate Statistical Analysis, Second Edition,* Prentice-Hall, 1988.



The SSA mission is supported by five strategic goals, including Goal 5, "To strengthen public understanding of Social Security programs." Goal 5, in turn, is supported by the single strategic objective, "By 2005, nine out of ten Americans will be knowledgeable about the Social Security programs in five important areas." PM #11 characterizes the public's level of knowledge about SSA programs. Assuming that the metric has strong performance measurement attributes, the diagram indicates that PM #11 logically aligns with SSA's strategic planning process.

Based on the taxonomy of performance measures included in Appendix F, PM #11 is a measure of accomplishment because it reports on a result (public awareness) achieved with SSA resources. It is further categorized as an outcome measure because it indicates the accomplishments or results (level of public awareness) that occur because of the services (public relations) provided. Furthermore, this measure of public awareness is similar to a measure of "public perceptions." As shown in Appendix F, measures of public perceptions are considered as outcome measures.

Within the framework of GPRA, Performance Measure #11 fits the intent of an outcome measure because it is "…a description of the intended result, effect, or consequence that will occur from carrying out a program or activity."[8] The intent of this performance measurement is to gauge public awareness (i.e., the effect) for the activity of providing

[8] OMB Circular A-11 *Preparation and Submission of Budget Estimates*, Section 200.2.

information to the public. A survey-based measurement of this type can be costly and takes time to implement. Nevertheless PM #11 is an appropriate and worthwhile GPRA performance indicator. It can be useful to both management and external stakeholders, as encouraged by OMB Circular A-11. However, there are a few inherent deficiencies in the current design of Performance Measure #11 that are worth noting:

Ideally, a performance metric should help the agency take action to affect the performance of the indicator being measured. In this case, the measurement system will not provide a clear indication of necessary action; this is because the analytical component of the public awareness survey is weak, since it utilizes simple statistical procedures. The performed analyses do not "read between the data lines." For the results to be more actionable, more advanced data analysis methodologies should be used to extract as much intelligence from the data as possible.

Performance Based Budgeting, the ultimate intent of GPRA, is an approach that relates budgets to outputs, and/or resources to performance. At a high level, this metric is well suited for performance based budgeting because stakeholders can evaluate the change in public awareness as a result of changing the total dollars spent on public relations. Where the current measurement system may fall short, however, is in indicating how dollars spent on specific types of educational programs or media impact public awareness. The MTN survey was intended to help clarify this. It is hoped that MTN can ultimately achieve this objective or that SSA can develop an alternative method for measuring the effectiveness of specific educational programs or media.

Recommendations

8) For the results to be more actionable, more advanced data analysis methodologies should be used to extract as much intelligence from the data as possible.

9) SSA should also work toward the successful implementation of the MTN survey or develop an alternative method for measuring the success of specific educational programs and/or media.

OTHER MATTERS

As part of this evaluation, we identified an issue that is peripheral to the engagement but, we believe, warrants SSA's attention. This point is discussed below.

1. Reporting FY 2001 Results (PUMS III).

While the idea of switching components of a knowledge indicator is not to be encouraged, it is understandable that, because of changes in policy, the definition of knowledge can change, requiring modifications to the questionnaire. However, it is notable that because of the introduced changes (e.g., use of 19 question in PUMS I & II instead of 14 questions in PUMS III) the knowledge scores have increased significantly, both for PUMS-I and II. These results are summarized in the following table.

Table 6. Changes in Knowledge Score Due to Changes in Measurement Method

Percent Knowledgeable	Based on 19 Questions	Based on 14 Questions
PUMS-I	55%	66%
PUMS-II	57%	68%
PUMS-III	Data not available	Data not available

Upon further evaluation, it appears that the above increase is partially due to elimination of questions that respondents have commonly scored low. The following table provides a summary of these questions.

Table 7. Summary of Questions Eliminated from the Knowledge Metric

Question		Reason for Deletion from Knowledge Score	Percent Correct (PUMS-II)
Q3c	Social Security pays for the food stamp program.	It does not help SSA or the public to know what services SSA does not provide	46%
Q5	What do you think is the youngest age someone can retire today, and start receiving FULL Social Security retirement benefits?	Confusion about what is considered "FULL" benefits	38%
Q7	Can a person retire early and still receive some Social Security retirement benefits?	Undisclosed	65%
Q14b	People on Social Security are living longer, so they cost the program more money.	Undisclosed	75%
Q14d	There is significant fraud and abuse by people who aren't entitled to benefits	Fraud and abuse are complex and hard to interpret	23%
MEAN			**49%**

Recommendation

10) If SSA plans to report the recalculated results for FY 1999 or FY 2000, it should ensure that the reported results include a description of the change in the knowledge calculation for these 2 years.

Agency Comments

SSA agreed with 9 of the 10 recommendations contained in this report. While agreeing with recommendation number eight, "…use more advanced data analysis methodologies to extract as much intelligence from the data as possible…." SSA stated that a lack of available resources has prevented it from completing more advanced analysis of the data collected in the PUMS survey. It noted that it is currently recruiting a staff person to help make such work possible in the future.

In disagreeing with recommendation number seven, “adjust the data with respect to any socioeconomic indicator, and use income or education as part of the weighting process,” SSA stated that it was satisfied with the procedures it has used throughout the PUMS survey. It believed that the changes suggested would not be an improvement and would make comparisons to previous years’ data difficult. The full text of the Agency’s comments is in Appendix C.

OIG Response

We appreciate the Agency’s comments to this report. The implementation of the recommendations will help to ensure for the efficient collection and use of the PUMS survey data.

We believe the data collected through PUMS would be more precise if SSA changed its current weighting methodology. There is a non-uniform response pattern across different demographic backgrounds in virtually all surveys. For instance, in household surveys, there are different response rates when comparing higher educated (more affluent) individuals with those at lower levels of education. The primary objective of weighting is to realign the composition of respondents so that they mimic that of the target population. Knowing that the socioeconomic composition of respondents (e.g., income or education) is almost always different from that of the target universe, it would benefit SSA to adjust (weight) the data along such indicators to reduce the skew that will otherwise bias the results. Key outcome measures of this survey are highly correlated with income and education. This further argues for adjusting the data with respect to these indicators, otherwise, the resulting data will be at the mercy of the mix of respondents they manage to contact. It is understandable that changing (improving) the weighting process will introduce some difficulties when it comes to comparing historical data. However, throughout the history of this survey significant changes have been introduced as deemed necessary; this change could be considered as yet another necessary adjustment.

APPENDICES

APPENDIX A – Scope and Methodology

APPENDIX B – Acronyms

APPENDIX C – Agency Comments

APPENDIX D – Performance Measure Summary Sheets

APPENDIX E – Performance Measure Process Maps

APPENDIX F – Performance Measure Taxonomy

SCOPE AND METHODOLOGY

The Social Security Administration (SSA) Office of the Inspector General (OIG) contracted PricewaterhouseCoopers to evaluate 11 SSA performance indicators identified in its Fiscal Year (FY) 2001 Annual Performance Plan (APP). We performed our testing from September 21, 2000 through February 15, 2001. Since FY 2001 performance results were not yet available as of the date of our evaluation, we performed tests of the performance data and related internal controls surrounding the maintenance and reporting of the results for FY 2000. Specifically, we performed the following:

1. Obtained an understanding of the Public Understanding Measurement System (PUMS) surveys.

2. Tested the reasonableness of the survey data.

3. Determined whether performance measures were meaningful and in compliance with the Government Performance and Results Act of 1993 (GPRA).

4. Identified findings relative to the above procedures and provided recommendations for improvement.

Our engagement was limited to testing at SSA's headquarters in Woodlawn, Maryland. The procedures that we performed were in accordance with the American Institute of Certified Public Accountants' Statement on Standards for Consulting Services, and are consistent with appropriate standards for performance audit engagements in *Government Auditing Standards* (Yellow Book, 1994 version). However, we were not engaged to and did not conduct an audit, the objective of which would be the expression of an opinion on the reliability or accuracy of the reported results of the performance measures evaluated. Accordingly, we do not express such an opinion. Had we performed additional audit procedures, other matters might have come to our attention that would have been reported to you.

1. Obtained an understanding of the PUMS surveys.

We obtained an understanding of the underlying process and procedures surrounding the implementation of the measure through interviews and meetings with the appropriate SSA and SSA's contractor personnel. Our evaluation of this performance measure involved a comprehensive evaluation of the survey methodology, including the sampling and questionnaire design, data collection procedures, and data analysis and reporting. Because this performance measure is conducted each year, specific attention has been given to each annual administration. In this process, we evaluated the following documents:

- **PUMS-I:**
  - National Report, 1998
  - Focus Group report, April 1998
  - Technical Report
  - Statement of Work
- **PUMS-II:**
  - Contract Requirements, May 1999
  - SSA's contractor Technical Proposal, June 1999
  - Survey Results
  - Summary of National Results
  - National Findings, April 2000
  - SSA Knowledge Indicators
- **Moving the Needle:**
  - Questionnaire
  - Regional Knowledge Tracking Survey: First Quarter (Nov 1999 – Sep 2000)
  - Knowledge Tracking Scorecard
- **PUMS-III:**
  - National Survey – Talking Points
  - Questionnaire
  - 14 Point Knowledge Point Discussion

2. Tested the reasonableness of the survey data.

To ensure the reasonableness of the number reported in the FY 2000 GPRA section of the SSA Annual Performance and Accountability Report, we evaluated the survey data for PUMS-I (FY 1999 survey) and PUMS-II (FY 2000 survey). Please note that the FY 2000 GPRA section of the SSA Annual Performance and Accountability Report only includes the results of the FY 1999 survey. Our evaluation included replicating the calculation of the knowledge score based on 26 and then 19 point scales. Once knowledge scores were calculated for each respondent, we calculated the overall percent of the population that are considered "knowledgeable."

As a result of this process, we were able to match those percents reported for PUMS-I as part the GPRA section, and for PUMS-II on SSA's internal reports. Moreover, survey weights were evaluated to ensure proper calculation of various weighting factors.

3. Determined whether performance measures were meaningful and in compliance with GPRA.

As part of this engagement, we evaluated the appropriateness of each of the performance measures with respect to GPRA compliance and SSA's APP. We determined whether the specific indicators and goals corresponded to the strategic goals identified in SSA's APP, determined whether each of these indicators accurately measure performance, and determined their compliance with GPRA requirements.

ACRONYMS

APP	Annual Performance Plan
FY	Fiscal Year
GPRA	Government Performance and Results Act
MTN	Moving the Needle
OMB	Office of Management and Budget
PM	Performance Measure
PUMS	Public Understanding Measurement System
PwC	PricewaterhouseCoopers LLP
RDD	Random Digit Dialog
SSA	Social Security Administration
SSI	Supplemental Security Income

AGENCY COMMENTS

SOCIAL SECURITY

MEMORANDUM

January 2, 2002 Refer To: S1J-3

To: James G. Huse, Jr.
Inspector General

Larry Dye
Chief of Staff

Subject: Office of the Inspector General (OIG) Draft Report, "Performance Measure Review: Reliability of the Data Used to Measure Public Knowledge of SSA" (A-02-01-11015)—INFORMATION

We appreciate OIG's efforts in conducting this review. Our comments on the report content and recommendations are attached.

Please let us know if we may be of further assistance. Staff questions may be referred to Dan Sweeney on extension 51957.

Attachment:
SSA Response

COMMENTS ON THE OFFICE OF THE INSPECTOR GENERAL (OIG) DRAFT REPORT, "PERFORMANCE MEASURE REVIEW: RELIABILITY OF THE DATA USED TO MEASURE PUBLIC KNOWLEDGE OF SSA" (A-02-01-11015)

Recommendation 1

Obtain documentation to support the employed methodologies, survey administration protocols, and analysis made.

Comment

We agree with this recommendation. Gallup has provided to us technical reports for all Public Understanding Measurement Surveys (PUMS) completed. Gallup will continue to provide technical reports for every survey they undertake.

Recommendation 2

Measure potential sampling errors in the estimates that reflect the employed sampling design, and incorporate the applied weights.

Comment

While we are confident that the statistical procedures used have fairly represented the United States population as a whole, we agree that the procedure could be improved. Specifically, as suggested by OIG, SSA and the contractor will in the future use software such as SUDAAN for all formal reporting.

Recommendation 3

Test the robustness of the regression methodology it is using to assure there is no multi-collinearity. If multi-collinearity is detected, consider alternative methods of determining the importance hierarchy such as factor analysis and principal components analysis.

Comment

Again, while we are confident that the methodology employed has been satisfactory, we agree with OIG's recommendation and will use the SUDAAN software to perform this function.

Recommendation 4

Consistently use the same set of information sources (news, public education, and campaigns) throughout the questionnaire.

Comment

As noted in the summary report of the audit findings, this recommendation largely concerns the questionnaire used in the Move the Needle (MTN) study, which did not figure in the computation of the national knowledge measure. While this was a "test" study, we agree that the questions could have been more valuable. As we do additional surveys that measure the outcome of specific public information campaigns, we will carefully consider the recommendation.

Recommendation 5

Establish prior reading of statements before inquiring about their usefulness.

Comment

We agree with this recommendation. In the fourth national PUMS survey, respondents who recalled receiving a *Statement* were asked the following question:

> Did you:
>
> 1) Glance at the statement
> 2) Read it carefully
> 3) Not look at it at all

Recommendation 6

Use the employment categories that are used by the Bureau of Labor Statistics (BLS) instead of the list that is currently used with the PUMS III and MTN surveys.

Comment

Most of the demographic categories used in the PUMS survey process (e.g., ethnicity and race) are the same as those used in the 1990 Census. However, there are some categories, such as the employment categories, that differ slightly. We are considering dropping the employment categories from the PUMS survey. However, if we decide to continue using them, we will consider using the BLS categories.

Recommendation 7

Adjust the data with respect to any socioeconomic indicator, and use income or education as part of the weighting process.

Comment

We do not agree with this recommendation, as we are satisfied with the procedures we have used throughout the PUMS surveys. Results of the PUMS survey have been adjusted to reflect age and race/ethnicity as well as probability of phone contact. We believe that the changes suggested would not be an improvement and would make comparison of past years' data difficult.

Recommendation 8

Use more advanced data analysis methodologies to extract as much intelligence from the data as possible.

Comment

While we agree with the intent of this recommendation, we must take into account resource implications. There is no doubt that we could get additional information from the data, but we have not had the resources to do so. Currently, our analysis centers around questions about who SSA's audience is and what they do and do not know. In addition, special analyses are done to help us understand what factors are associated with knowledgeable citizens and what types of citizens know which pieces of information. This information has historically been most pertinent to the regional offices. Additional analysis is completed on the few *Statement* questions that exist, although this is not the primary purpose of the PUMS survey.

The Office of Communications is in the process of recruiting a staff person to perform the kind of advanced data analysis that OIG suggests.

Recommendation 9

Work toward the successful implementation of the MTN survey or develop an alternative method for measuring the success of specific educational programs and/or media.

Comment

We agree. The MTN study has proven helpful and we are working with the contractor on a final report.

Recommendation 10

If reporting recalculated results for FY 1999 or FY 2000, ensure that the reported results include a description of the change in the knowledge calculation for those two years.

Comment

As we noted in the audit conference meeting last August, the changes made to the knowledge calculation were made solely as a result of our strategic planning process and the release in August 2000 of SSA's new strategic plan, "Mastering the Challenge." We have made this clear in our PUMS III informational materials.

PERFORMANCE MEASURE PROCESS MAP

PERFORMANCE MEASURE SUMMARY SHEET

Name of Measure	**Measure Type**	**Strategic Goal/Objective**
Percent of public who are knowledgeable about Social Security Programs.	Percentage	**Goal:** To strengthen public understanding of Social Security programs **Objective:** By 2005, nine out of ten Americans will be knowledgeable about the Social Security programs in five important areas: • Basic program facts • Financial value of programs to individuals • Economic and social impact of the programs • How the programs are financed today • Financing issues

Purpose	**Survey Frequency**
To assess the percent of the public who are knowledgeable about Social Security Programs, SSA will perform an annual Public Understanding Measurement System (PUMS) survey.	Annually

Target Goal	**How Computed**	**Data Source**
65%	Respondents who could correctly answer 13 of the 19 aided questions were deemed knowledgeable about SSA.	PUMS MTN

Designated Staff Members	**Division**
Rusty Toler Bernie Gonzales Lisa Jones	Office of Communication

Testing and Results

Our evaluation of this performance measure involved an evaluation of the survey methodology, including the sampling and questionnaire designs, data collection procedures, and data analysis and reporting. To obtain an understanding of PUMS, we evaluated the following documents:

- **PUMS-I:** National Report, 1998; Focus Group report, April 1998; Technical Report; and Statement of Work.
- **PUMS-II:** Contract Requirements, May 1999; Survey Results; Summary of National Results; National Findings, April 2000; and SSA Knowledge Indicators.
- **Moving the Needle:** Questionnaire, Regional Knowledge Tracking Survey: First Quarter (Nov 1999 – Sept 2000), and Knowledge Tracking Scorecard
- **PUMS-III:** National Survey – Talking Points, Questionnaire, and 14 Point Knowledge Point Discussion

To ensure the reasonableness of the number reported in the FY 2000 GPRA section of the accountability report, we performed the following:

- Replicated the calculation of knowledge score based on 26 and then 19 point scales.
- Once knowledge scores were calculated for each respondent, we calculated the overall percent of population that were considered "knowledgeable".
- Evaluated survey weights to ensure proper calculation of various weighting factors.

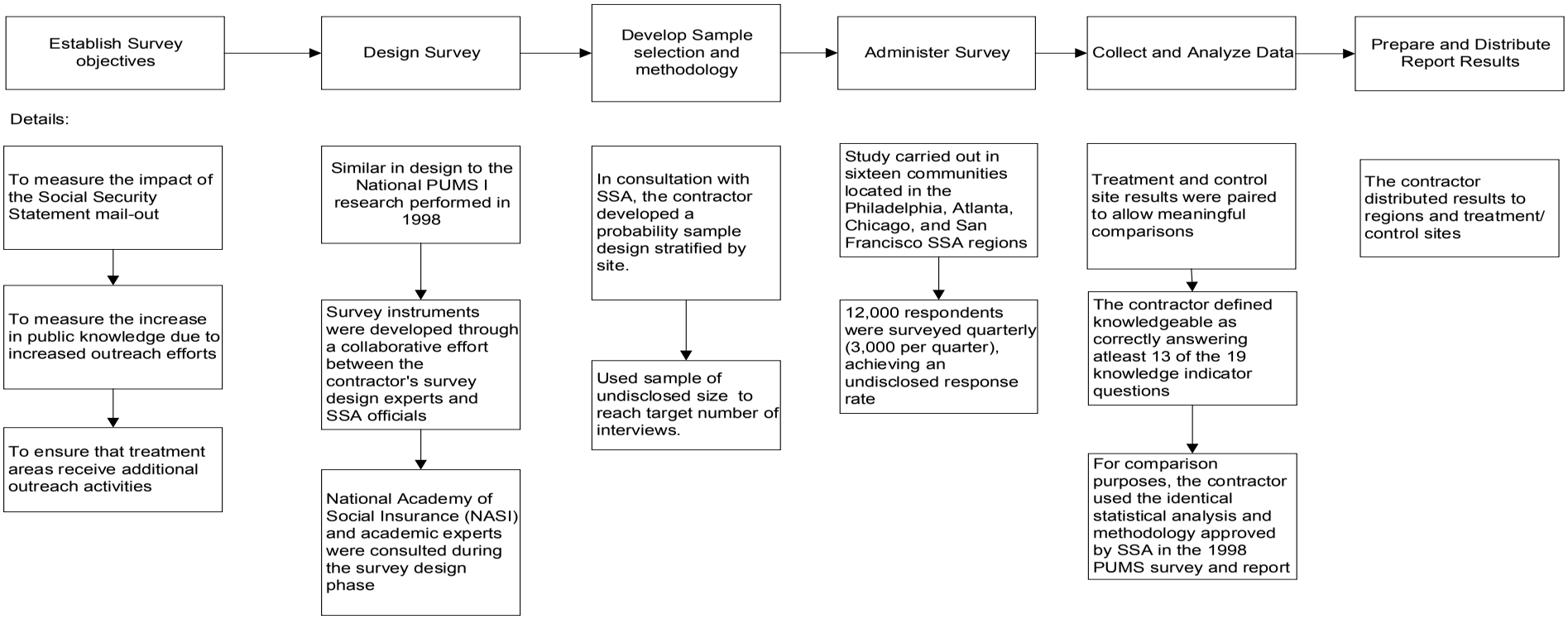
Refer to "Results of Evaluation" for a description of the findings.

Performance Measure # 11

Goal #5 is to strengthen public understanding of the Social Security Programs.

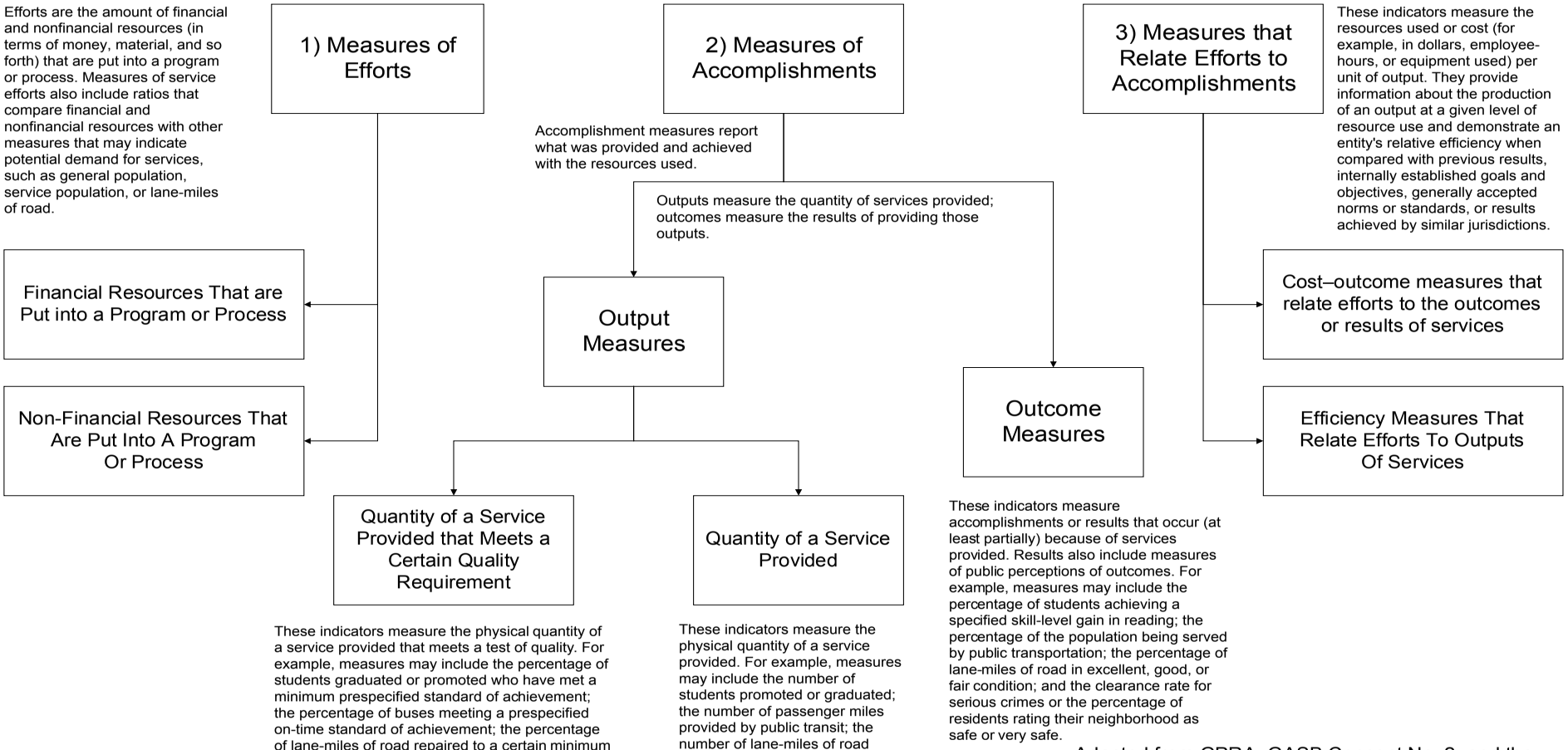
SSA developed the Public Understanding Measurement System (PUMS) to establish baseline data on the public's knowledge of Social Security programs, and to track changes in the public's level of knowledge through the year 2005. The PUMS initiative will provide data needed to design annual public education programs.

SSA hired a contractor to perform a one-year knowledge tracking study known as the "Move the Needle" study.



PERFORMANCE MEASURE TAXONOMY

Categories of Performance Measures:



Adapted from GPRA, GASB Concept No. 2, and the "Performance Measurement for Government" web site at Rutgers University www.rutgers.edu/Accounting/raw/seagov/pmg/index.html

DISTRIBUTION SCHEDULE

	No. of Copies
Commissioner of Social Security	1
Management Analysis and Audit Program Support Staff, OFAM	10
Inspector General	1
Assistant Inspector General for Investigations	1
Assistant Inspector General for Executive Operations	3
Assistant Inspector General for Audit	1
Deputy Assistant Inspector General for Audit	1
Director, Data Analysis and Technology Audit Division	1
Director, Financial Audit Division	1
Director, Western Audit Division	1
Director, Southern Audit Division	1
Director, Northern Audit Division	1
Director, General Management Audit Division	1
Issue Area Team Leaders	25
Income Maintenance Branch, Office of Management and Budget	1
Chairman, Committee on Ways and Means	1
Ranking Minority Member, Committee on Ways and Means	1
Chief of Staff, Committee on Ways and Means	1
Chairman, Subcommittee on Social Security	2
Ranking Minority Member, Subcommittee on Social Security	1
Majority Staff Director, Subcommittee on Social Security	2
Minority Staff Director, Subcommittee on Social Security	2
Chairman, Subcommittee on Human Resources	1
Ranking Minority Member, Subcommittee on Human Resources	1
Chairman, Committee on Budget, House of Representatives	1
Ranking Minority Member, Committee on Budget, House of Representatives	1
Chairman, Committee on Government Reform and Oversight	1
Ranking Minority Member, Committee on Government Reform and Oversight	1
Chairman, Committee on Governmental Affairs	1
Ranking Minority Member, Committee on Governmental Affairs	1

Chairman, Committee on Appropriations, House of Representatives	1
Ranking Minority Member, Committee on Appropriations, House of Representatives	1
Chairman, Subcommittee on Labor, Health and Human Services, Education and Related Agencies, Committee on Appropriations, House of Representatives	1
Ranking Minority Member, Subcommittee on Labor, Health and Human Services, Education and Related Agencies, Committee on Appropriations, House of Representatives	1
Chairman, Committee on Appropriations, U.S. Senate	1
Ranking Minority Member, Committee on Appropriations, U.S. Senate	1
Chairman, Subcommittee on Labor, Health and Human Services, Education and Related Agencies, Committee on Appropriations, U.S. Senate	1
Ranking Minority Member, Subcommittee on Labor, Health and Human Services, Education and Related Agencies, Committee on Appropriations, U.S. Senate	1
Chairman, Committee on Finance	1
Ranking Minority Member, Committee on Finance	1
Chairman, Subcommittee on Social Security and Family Policy	1
Ranking Minority Member, Subcommittee on Social Security and Family Policy	1
Chairman, Senate Special Committee on Aging	1
Ranking Minority Member, Senate Special Committee on Aging	1
Vice Chairman, Subcommittee on Government Management Information and Technology	1
President, National Council of Social Security Management Associations, Incorporated	1
Treasurer, National Council of Social Security Management Associations, Incorporated	1
Social Security Advisory Board	1
AFGE General Committee	9
President, Federal Managers Association	1
Regional Public Affairs Officer	1
Total	**97**

Overview of the Office of the Inspector General

Office of Audit

The Office of Audit (OA) conducts comprehensive financial and performance audits of the Social Security Administration's (SSA) programs and makes recommendations to ensure that program objectives are achieved effectively and efficiently. Financial audits, required by the Chief Financial Officers Act of 1990, assess whether SSA's financial statements fairly present the Agency's financial position, results of operations, and cash flow. Performance audits review the economy, efficiency, and effectiveness of SSA's programs. OA also conducts short-term management and program evaluations focused on issues of concern to SSA, Congress, and the general public. Evaluations often focus on identifying and recommending ways to prevent and minimize program fraud and inefficiency.

Office of Executive Operations

The Office of Executive Operations (OEO) supports the Office of the Inspector General (OIG) by providing information resource management; systems security; and the coordination of budget, procurement, telecommunications, facilities and equipment, and human resources. In addition, this office is the focal point for the OIG's strategic planning function and the development and implementation of performance measures required by the Government Performance and Results Act. OEO is also responsible for performing internal reviews to ensure that OIG offices nationwide hold themselves to the same rigorous standards that we expect from the Agency, as well as conducting employee investigations within OIG. Finally, OEO administers OIG's public affairs, media, and interagency activities and also communicates OIG's planned and current activities and their results to the Commissioner and Congress.

Office of Investigations

The Office of Investigations (OI) conducts and coordinates investigative activity related to fraud, waste, abuse, and mismanagement of SSA programs and operations. This includes wrongdoing by applicants, beneficiaries, contractors, physicians, interpreters, representative payees, third parties, and by SSA employees in the performance of their duties. OI also conducts joint investigations with other Federal, State, and local law enforcement agencies.

Counsel to the Inspector General

The Counsel to the Inspector General provides legal advice and counsel to the Inspector General on various matters, including: 1) statutes, regulations, legislation, and policy directives governing the administration of SSA's programs; 2) investigative procedures and techniques; and 3) legal implications and conclusions to be drawn from audit and investigative material produced by the OIG. The Counsel's office also administers the civil monetary penalty program.