

For Release
on Delivery
Expected at
9:30 a.m. EDT
Friday
May 18, 1990

SOCIAL SECURITY:
Service to the Public--Accuracy
of the 800 Phone Service

Statement of Joseph F. Delfico
Director, Income Security Issues
Human Resources Division

Before the
Special Committee on Aging
United States Senate



04852, 14/376

Mr. Chairman and Members of the Committee:

We are pleased to be here today to testify on the Social Security Administration's (SSA) methodology for evaluating the accuracy of information being provided to the public over its 800 phone system. My testimony will also address SSA efforts to stop withholding of Medicare catastrophic coverage premiums. I will first discuss SSA's 800 service.

BACKGROUND

Each year more than 60 million people call SSA's 800 number for a wide variety of reasons. While an individual can visit a social security office to obtain service, the 800 system is designed to be the public's initial point of contact for social security services. Most telephone inquiries, such as inquiries about social security office hours, can be taken care of immediately. Other matters, such as applying for benefits, are more complex and are referred to local offices to handle. Whatever the public's queries may be, accurate responses and good public service go hand-in-hand.

To ensure that the public receives good service, the accuracy of the information provided need to be measured and evaluated periodically. Though on the surface the measurement may appear straightforward, it is complex and can be done in a number of ways.

SSA has tried two ways to measure accuracy. When the 800 service started in 1988, SSA designed test questions that were typical for the 800 system with specific criteria to measure whether responses were accurate and complete. Posing as the public, SSA employees called the 800 number, asked the questions, and assessed the quality of the responses.

This approach made it relatively easy to assess the accuracy of the responses and reduced the subjectivity of the assessment. However, the approach had a number of drawbacks. There was some argument as to whether the questions were typical, and whether they were too hard or too easy. But most importantly, the results of the test procedure did not yield actual error rates for live calls, but instead only gave a reading of how well SSA did on certain types of predetermined questions. The results did little to illuminate how well SSA was serving the public in general.

Because of these drawbacks, SSA changed its approach and decided to monitor live calls nationwide beginning in October 1989. However, as we shall discuss, this approach also has its shortcomings.

Our testimony addresses SSA's current method for monitoring 800 service quality. We assessed how SSA designed and implemented

its study methodology and how it reports its results. We examined study instructions and other documents and discussed the study methodology and results with SSA officials. Also, as part of our review, we participated with SSA quality reviewers in monitoring live calls placed to SSA's Metro West facility in Baltimore. The monitoring took place during 8 days in February of this year. To assist us in this effort, we contracted with three former career SSA employees who had an average of 10 years experience at the supervisory level at SSA teleservice centers. Our contractors listened jointly with SSA to 188 calls involving 260 separate issues or questions.

Because of the small sample size, the results of our study cannot be projected nationally, regionally, or even to the Metro West complex for the days monitored. However, the sample did provide important insights into the implementation of SSA's methodology.

RESULTS OF OUR TESTING

One particular difficulty in monitoring live calls is that reviewers have only one opportunity to hear a call, interpret the facts as well as the context of the issues raised, and formulate an opinion of what constitutes a satisfactory response. While reviewers can take notes and use them to research manuals and program requirements at a later time, they can never replay

the conversation. Further complicating the difficulty of this task, it is common for calls to last 10 minutes or longer and involve more than one issue.

Given this inherent difficulty in monitoring calls, we wanted to assess the extent of variability in the judgments made by several reviewers listening to the same call. Therefore, we arranged to have two of the three GAO contractors listen concurrently with 1 SSA reviewer on all of the calls and have each of the GAO contractors and SSA reviewers record their judgments independently.

There was a high rate of agreement between the three reviewers. On 86 percent of the 260 issues, SSA and at least one GAO reviewer agreed on the accuracy of the response; and on 70 percent of the issues, the agreement was unanimous.

Next, we attempted to resolve differences among reviewers and assess the reliability of the initial decisions. To do this, we examined the reviewers' notes--which documented the basis for their decisions--and corroborated them with SSA operating instructions and manuals. Where GAO contractors disagreed with each other, we first arrived at a consensus among them. We then compared the GAO contractor position with SSA's position and discussed with SSA those calls on which we disagreed.

After the discussions with SSA, we reviewed all of the calls monitored to assure that the review criteria was consistently applied for the GAO contractor decisions. Further, we eliminated 28 of the 188 sample calls from our universe for two reasons. First, for 6 calls, SSA operating instructions were unclear as to how the teleservice representatives were to respond to the point; thus, we could not make a judgment on the call one way or another. Second, for 22 calls, the SSA and GAO reviewers did not agree on what the caller said. This left 160 calls to evaluate. For the rest of this analysis, we discuss the results on a call basis rather than on an issue basis because that is the way SSA compiles and reports its accuracy data. Computing error rates on an issue basis could result in a somewhat lower error rate.

The most serious error a teleservice representative can make is to provide an inaccurate response to a question that could adversely affect the caller's benefits. GAO and SSA agreed on 11 "payment/benefit" errors made during the test, but GAO found 3 additional errors. All 3 of these errors involve cases where the caller wanted to file a claim for benefits, but because SSA's computer systems were down, the caller was instructed to call back the 800 number or the local field office. SSA procedures require that in these circumstances, the telephone representatives should record the caller's intent to file for benefits, but the representative did not.

Perhaps of less importance, but still a measure of good public service, are workload/caller inconvenience errors. These are errors that cause SSA unwarranted workloads or inconvenience the public. GAO contractors found 55 calls with these errors, and SSA found 26, which is 29 less than GAO.

For about half of these 29 calls in which GAO and SSA differed, most of which involved requests for social security numbers, GAO and SSA basically agree that the call was not handled as well as it should have been, but differ as to whether the response was inaccurate or merely incomplete. We believe that SSA teleservice center guides and reviewer instructions could be clarified as to what constitutes an inaccurate or accurate but incomplete response.

Most of the other disputed judgments involved questions about obtaining Social Security cards. GAO contractors rated 12 of these calls inaccurate because the SSA employee failed to fully disclose the evidence needed to obtain a social security card. SSA headquarters officials argued that such disclosure (though required by their guide) was unnecessary because the social security card application, which is sent to the caller in almost all cases describes all the proofs needed.

The final category of responses we analyzed were those rated as accurate but incomplete. SSA rated 23 calls incomplete while we rated only 11 calls incomplete. The main reason for this difference is that we rated calls as inaccurate that SSA rated as incomplete. This was due to differences in interpretation of SSA instructions to its reviewers.

OBSERVATIONS ON SSA'S METHODOLOGY

In our opinion, SSA could take a number of steps to strengthen its methodology for measuring the accuracy and completeness of 800 service responses, and in our report to this Committee, we will focus on the following areas.

- . Clarifying the teleservice manual,
- . Integrating the accuracy checks with the quality assurance process,
- . Updating training to reflect trends in accuracy to specific responses, and
- . Recording conversations to make it easier to access accuracy.

DISCLOSURE OF DATA ON RESPONSE ACCURACY

Another aspect we would like to address is how SSA reports the results of its accuracy studies to the public and the Congress.

To date, SSA has only reported payment/benefit errors. It has not disclosed the extent to which callers are inconvenienced or receive incomplete information, both of which are public service issues. SSA officials told us that, in comparison with payment benefit errors, judgments on these types of responses are much more subjective and therefore they are not satisfied that such data are meaningful. While we agree that these judgments can be difficult, we believe that SSA should strive to refine and report on these matters in some fashion because they are important to measuring the quality of public service.

The other problem is how SSA computes the payment/benefit accuracy rate that it does report to the Congress. In testimony before the Congress earlier this year, and in its service quality report to the House and Senate Appropriations Committees dated March 1990, SSA said that its accuracy rate was 97 percent "regarding payment amounts or eligibility." The method used to compute the rate overstates accuracy, potentially to a significant extent.

We can illustrate this problem, using our test data (see attachment I). For the calls we monitored jointly with SSA, the agency found 11 payment/benefit errors. Using its current methodology, SSA would compute the payment/benefit error rate by dividing the 11 errors by 160, the total number of calls we

analyzed, regardless of whether these calls had anything to do with payments or benefits. This computation produces an error rate of 6.9 percent, or an accuracy rate of 93.1 percent for payment/benefit errors, and is not meaningful. We believe the rate would be meaningful if the calculation was made by dividing the errors by the universe of calls with a potential for having payment benefit errors. The error rate would then be 17.7 percent, with a corresponding accuracy rate of 82.3 percent.

ROLL BACK OF CATASTROPHIC COVERAGE
PREMIUMS COMPLETED ON SCHEDULE

I would now like to discuss another area you are interested in-- SSA's progress in stopping the withholding of Medicare catastrophic coverage premiums. The Medicare Catastrophic Coverage Act of 1988,¹ among other things, expanded the Medicare program to cover catastrophic medical expenses. Payments for this coverage were deducted from monthly Social Security payments beginning in January 1990. But almost immediately, the act came under fire from beneficiaries and others, and in December 1989, most provisions of the act were repealed effective January 1, 1990.²

¹P.L. 100-360, enacted on July 1, 1988.

²P.L. 101-234, enacted on December 13, 1989.

However, SSA was not able to stop withholding catastrophic coverage payments by the effective date without jeopardizing the timely processing of basic monthly Social Security benefits. Consequently, it continued to withhold premiums averaging \$5.30 from the monthly Social Security payments of 27 million beneficiaries through April 1990, until it reprogrammed its computers.

May's Social Security checks were the first not to have catastrophic coverage premiums withheld. Consequently, Treasury issued two refunds--one in February and another in April--returning the excess withholdings of about \$589 million.

SSA officials acknowledge that withholding premiums could have been stopped sooner if the agency's computer programs for catastrophic coverage premiums were better organized and easier to maintain. Given the limitations of SSA's computer systems, the agency's approach was the quickest way to stop the collection of the repealed catastrophic coverage premiums as well as return the money withheld. This unusual arrangement of collecting premiums while issuing refunds will cost the government about \$50 million.

SSA's difficulties in stopping the catastrophic coverage premiums are indicative of the long-standing computer systems problem the agency has, namely, the need to modernize its computer systems.

SSA has stated, and we have previously reported, that the complex structure of SSA's software programs makes them difficult to understand and maintain.³ These problems will only be corrected by its computer modernization effort, which will not be completed until the mid-1990s.

In our view, SSA should document the process that was followed to reverse withholding and keep current the computer programs that were developed. This could help save time and money in making future rate changes until SSA's effort to overhaul its existing computer system is completed.

- - - -

This concludes my prepared statement. I will be happy to answer any questions you and the committee members may have.

³Social Security Administration's Computer Modernization Effort May Not Achieve Planned Objectives (GAO/IMTEC-85-16, Sept. 30, 1985); Software Systems: SSA Encountering Significant Delays in Its Claims Modernization Project (GAO/IMTEC-87-8, Dec. 22, 1986); and Software Maintenance: SSA's Use of Its Software Maintenance Package (GAO/IMTEC-89-38, June 15, 1989).

COMPARISON OF SSA AND GAO TEST RESULTS
(Not Projectable)^a

	<u>SSA</u>	<u>GAO</u>
<u>Number of calls in sample^b</u>	160	160
<u>Payment/Benefit Errors</u>		
Number of calls with payment/benefit potential, as identified by SSA	62	62
Number of calls with payment/benefit errors	11	14
Percent of calls with payment/benefit errors	17.7	22.6
<u>Workload/Inconvenience Errors</u>		
Number of calls with workload/inconvenience potential	160	160
Number of calls with workload/inconvenience errors	26	55
Percent of calls with workload/inconvenience errors	16.3	34.4
<u>Percent of Call Universe With Errors</u> (Payment/Benefit and Workload/ Inconvenience)	23.1	43.1
<u>Number of Accurate But Incomplete Calls</u>	23	11

^aBecause of the small number of calls sampled, these results cannot be projected nationally, regionally, or even to the Metro West teleservice center for the period.

^bActual number of calls listened to was 188. Twenty-eight calls were eliminated because of lack of clarity and disagreement on what was said.