



February 2016

# SOCIAL SECURITY DISABILITY

## SSA Could Increase Savings by Refining Its Selection of Cases for Disability Review

Accessible Version

# GAO Highlights

Highlights of [GAO-16-250](#), a report to congressional requesters

## Why GAO Did This Study

To help ensure that only eligible individuals receive disability benefits, SSA conducts periodic CDRs to assess beneficiaries' medical condition. CDRs have historically saved the government money. However, in recent years, SSA has had difficulty conducting timely CDRs resulting in a backlog of over 900,000 CDRs in fiscal year 2014. With this backdrop, GAO was asked to study SSA's ability to conduct and manage timely, high-quality CDRs.

This report evaluates, among other things, how SSA selects which CDRs to conduct and the extent to which SSA reviews the quality of CDR decisions.

GAO analyzed CDR data for fiscal years 2003 through 2013 (the most recent year for which complete data were available); assessed SSA's models used to prioritize CDRs; reviewed relevant federal laws, regulations, and SSA documentation about CDR prioritization and accuracy review procedures; and interviewed SSA and state Disability Determination Services officials.

## What GAO Recommends

GAO recommends SSA, among other things, further consider cost savings as part of its prioritization of CDRs, analyze the root causes of CDRs with errors, and track date errors. SSA agreed with most of GAO's recommendations, but disagreed that there is a need to track date errors and to adjust its approach to sampling CDRs for quality review. GAO maintains actions are warranted and feasible as discussed in the report.

View [GAO-16-250](#). For more information, contact Daniel Bertoni at (202) 512-7215 or [bertonid@gao.gov](mailto:bertonid@gao.gov)

February 2016

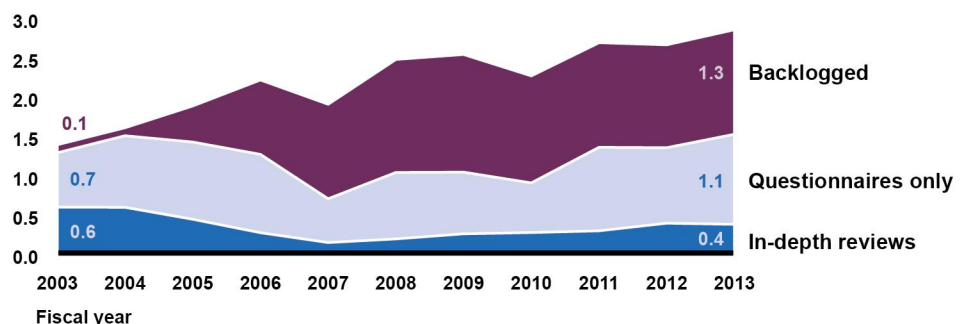
## SOCIAL SECURITY DISABILITY

### SSA Could Increase Savings by Refining Its Selection of Cases for Disability Review

## What GAO Found

The Social Security Administration (SSA) selects cases for continuing disability reviews (CDR) using several inputs, but it does not do so in a manner that maximizes potential savings. SSA first prioritizes CDRs required by law or agency policy such as those for children under 1 year old who are receiving benefits due in part to low birth weight. Then SSA uses statistical models to identify the remaining CDRs to be conducted each year. The models also determine which cases will receive an in-depth review of medical records by the Disability Determination Services—the state agencies that conduct CDRs—versus a lower-cost questionnaire sent directly to the beneficiary. As shown in the figure below, a growing number of cases have been set aside for future review (backlogged) over the last 10 years. Although SSA somewhat considers potential cost savings when selecting cases for in-depth reviews, its approach does not maximize potential savings for the government. For example, estimated average savings from conducting CDRs are higher for some groups of Disability Insurance (DI) beneficiaries than others, but SSA's selection process does not differentiate among these groups. As a result, it may be missing opportunities to efficiently and effectively use federal resources.

**CDRs Completed and Backlogged by SSA, Fiscal Years 2003-2013**  
Number of cases (in millions)



Source: GAO analysis of data from Social Security Administration (SSA) Continuing Disability Review (CDR) Annual Report and data provided by SSA's Office of Public Service and Operations Support. | GAO-16-250

SSA reviews a sample of CDRs for quality, but its analysis and reporting of errors is not comprehensive. Specifically, SSA randomly selects CDR decisions to check for a variety of potential errors. For example, SSA regularly monitors and reports on the frequency of errors that affect whether benefits are continued or ceased. However, contrary to federal internal control standards, SSA does not systematically analyze errors to detect and address root causes. Consequently, SSA lacks information that could help improve the quality of the reviews conducted by the Disability Determination Services. Further, in determining CDR accuracy rates, SSA does not count date errors, including incorrect cessation dates, which can affect disability benefit payments. As a result, decision makers do not have a complete picture of the CDR errors that affect disability payments.

---

# Contents

---

Letter	1	
	Background	2
	SSA Prioritizes CDRs Using Several Inputs but Does Not Fully Incorporate Potential Cost-Savings	6
	SSA Reviews a Sample of CDRs for Quality but Does Not Comprehensively Analyze and Report Errors	17
	SSA Reasonably Estimates CDR Cost Savings but Has Conducted Limited Sensitivity Analysis and Documentation	21
	Conclusions	23
	Recommendations for Executive Action	24
	Agency Comments and Our Evaluation	25
<hr/>		
Appendix I: Objectives, Scope, and Methodology		29
Appendix II: Comments from the Social Security Administration		34
Appendix III: GAO Contact and Staff Acknowledgments		39
	GAO Contact	39
	Staff Acknowledgments	39
<hr/>		
Appendix IV: Accessible Data	40	
	Agency Comment Letter	40
	Data Tables	46
<hr/>		
Figures		
	Figure 1: Continuing Disability Reviews Completed and Backlogged by SSA, Fiscal Years 2003-2013	5
	Figure 2: Social Security Administration's Process for Prioritizing Continuing Disability Reviews	8
	Figure 3: Percentage of Completed Full Medical Reviews by the Basis for their Selection, Fiscal Years 2003-2013	9
	Figure 4: Average Lifetime Savings per Full Medical Review by Beneficiary Group, Fiscal Years 2012-2013	12
	Figure 5: Average Lifetime Savings per Full Medical Review by Disability Insurance Beneficiary Groups, Fiscal Years 2012-2013	14

---

Figure 6: Simplified Hypothetical Scenarios Depicting Differences in Beneficiaries' Potential Lifetime Savings, Fiscal Year 2013	16
Figure 7: SSA's Annual Calculation of Cost Savings Rate from Continuing Disability Reviews	22

---

### Abbreviations

CDR	Continuing Disability Review
CMS	Centers for Medicare & Medicaid Services
DI	Disability Insurance
DICARS	Disability Case Adjudication and Review System
DDS	Disability Determination Services
SSA	Social Security Administration
SSI	Supplemental Security Income

---

---

This is a work of the U.S. government and is not subject to copyright protection in the United States. The published product may be reproduced and distributed in its entirety without further permission from GAO. However, because this work may contain copyrighted images or other material, permission from the copyright holder may be necessary if you wish to reproduce this material separately.



February 11, 2016

The Honorable Vern Buchanan  
Chairman  
Subcommittee on Human Resources  
Committee on Ways and Means  
House of Representatives

The Honorable Sam Johnson  
Chairman  
Subcommittee on Social Security  
Committee on Ways and Means  
House of Representatives

To help ensure that only eligible individuals receive disability benefits, the Social Security Administration (SSA) is generally required to conduct periodic continuing disability reviews (CDR) for certain individuals who are receiving benefits under one or both of its disability programs: Disability Insurance (DI) and Supplemental Security Income (SSI).<sup>1</sup> These reviews assess whether individuals remain eligible for benefits based on several criteria, including their current medical condition and ability to work. CDRs are an important check on the integrity of the disability programs and have consistently saved the government more money than they cost to conduct. However, in recent years, SSA has had difficulty completing timely reviews, and, as a result, amassed a backlog of over 900,000 CDRs by the end of fiscal year 2014.<sup>2</sup> In addition, the Disability Insurance Trust Fund has faced financial challenges in recent years.

With this backdrop, concerns have been raised about SSA's ability to conduct and manage timely, high-quality CDRs. You asked us to study this topic. This report examines (1) how SSA selects which CDRs to

---

<sup>1</sup> DI is an insurance program that provides monthly cash benefits to individuals who are unable to work because of severe long-term disability. SSI is a means-tested program that provides financial assistance to disabled, blind, or aged individuals.

<sup>2</sup> In 2012, we reported that hundreds of thousands of childhood CDRs were overdue by more than 3 years, including reviews for children expected to medically improve. GAO, *Supplemental Security Income: Better Management Oversight Needed for Children's Benefits*, [GAO-12-497](#) (Washington, D.C.: June 26, 2012).

---

conduct, (2) the extent to which SSA reviews the quality of CDR decisions, and (3) how SSA calculates cost savings from CDRs.

To examine SSA's process for selecting which cases receive a CDR, we reviewed legal requirements for conducting certain types of CDRs and analyzed data on the number and type of CDRs conducted for fiscal years 2003 through 2013 (the most recent year for which complete data were available). We also reviewed documentation about how SSA prioritizes which CDRs to conduct each year or backlog for future review. Specifically, we examined the statistical models SSA uses to help prioritize CDRs, and interviewed SSA officials about these practices. To evaluate the extent to which SSA reviews the quality of CDR decisions, we reviewed relevant federal laws, regulations, and program policies and analyzed SSA's monthly CDR accuracy data by state for June 2013 through April 2015. We also interviewed SSA officials and reviewed documentation about how SSA conducts CDR quality reviews and uses the results of these reviews to prevent errors. To evaluate SSA's approach to calculating cost savings from CDRs, we interviewed SSA actuaries and other officials about their process for estimating cost savings and reviewed their documentation and models. We assessed the reliability of the CDR case-processing and accuracy data used for our analyses by, among other things, reviewing related documentation and testing for obvious errors. We determined that the data were sufficiently reliable for assessing how SSA prioritizes CDRs and reviews their quality. We compared SSA's activities to federal internal control standards and generally accepted statistical and actuarial practices where relevant.

We conducted this performance audit from December 2014 to February 2016 in accordance with generally accepted government auditing standards. Those standards require that we plan and perform the audit to obtain sufficient, appropriate evidence to provide a reasonable basis for our findings and conclusions based on our audit objectives. We believe that the evidence obtained provides a reasonable basis for our findings and conclusions based on our audit objectives.

---

## Background

SSA administers two disability programs that provide monthly cash benefits to eligible individuals: (1) Disability Insurance (DI) for individuals (and their dependents) who have paid into the Disability Insurance Trust Fund and (2) Supplemental Security Income (SSI) for low-income individuals. To be eligible for DI or SSI benefits based on a disability, an individual must have a medically determinable physical or mental impairment that (1) prevents the individual from engaging in any

---

substantial gainful activity and (2) is expected to result in death or has lasted or is expected to last at least 1 year.

Federal law generally requires CDRs to be conducted at least once every 3 years for all DI beneficiaries whose disabilities are not considered permanent, and at intervals determined appropriate by SSA for those whose impairments are considered permanent.<sup>3</sup> For SSI, federal law generally requires SSA to (1) conduct CDRs for infants during their first year of life if they are receiving SSI benefits due in part to low birth weight, and at least once every 3 years for SSI children under age 18 if their impairments are considered likely to improve, and (2) review the cases of all SSI children beginning on their 18th birthday to determine whether they are eligible for disability benefits under adult disability criteria.<sup>4</sup> SSA may waive the requirement to conduct periodic legislatively-required CDRs on a state-by-state basis.<sup>5</sup> SSA may also conduct CDRs that are not required by law as it deems appropriate.

---

## CDR Process

SSA contracts with state Disability Determination Services (DDS) agencies to initially determine whether applicants are disabled and to conduct periodic CDRs to determine whether beneficiaries continue to be disabled. DDS examiners assess whether individuals are eligible for benefits based on several criteria, including their current medical condition and ability to work. At the time beneficiaries enter the DI or SSI programs or continue their benefits following a CDR, a DDS examiner determines beneficiaries' due dates for a subsequent CDR based on their potential for medical improvement. Beneficiaries classified as "medical improvement expected" are generally scheduled for a CDR within 6 to 18 months, beneficiaries classified as "medical improvement possible" are scheduled once every 3 years, and beneficiaries classified as "medical improvement not expected" are scheduled once every 5 to 7 years.

To cost-effectively manage its CDR workload, SSA conducts CDRs in different ways. In general, beneficiaries with a high likelihood of medical improvement are referred for a full medical review—an in-depth

---

<sup>3</sup> 42 U.S.C. § 421(i).

<sup>4</sup> 42 U.S.C. § 1382c(a)(3)(H).

<sup>5</sup> See 42 U.S.C. §§ 421(i)(2), 1383b(a).



---

assessment of a beneficiary's medical and vocational status. Beneficiaries with a low likelihood of medical improvement are, at least initially, sent a questionnaire known as a mailer. If beneficiaries respond to a mailer in certain ways, SSA may refer these individuals for a full medical review. In contrast to mailers, full medical reviews are more labor intensive and expensive.<sup>6</sup> Full medical reviews result in a decision to either cease or continue an individual's benefits. In fiscal year 2013, the cessation rate for CDRs involving full medical reviews was about 19 percent, whereas the cessation rate for all CDRs including mailers was about 5 percent.<sup>7</sup> Each year, SSA allocates a portion of its program integrity budget to CDRs, which affects the number of full medical reviews and mailers that the agency initiates during the year. When the number of cases due for a CDR exceeds SSA's capacity to conduct full medical reviews and mailers, the cases not initiated during the year are considered backlogged for future review.

The number of CDRs completed as full medical reviews, as mailers only, or backlogged varied during fiscal years 2003 through 2013 (see fig. 1). After the authority for special funding to process CDRs expired in fiscal year 2002,<sup>8</sup> backlogged CDRs increased from about 100,000 cases in fiscal year 2003 to more than 1 million in fiscal year 2007, reaching a peak of nearly 1.5 million in fiscal year 2009. At the same time, the number of full medical reviews fell from nearly 670,000 in fiscal year 2003 to less than 190,000 in fiscal year 2007 before rebounding to nearly 429,000 in fiscal year 2013.

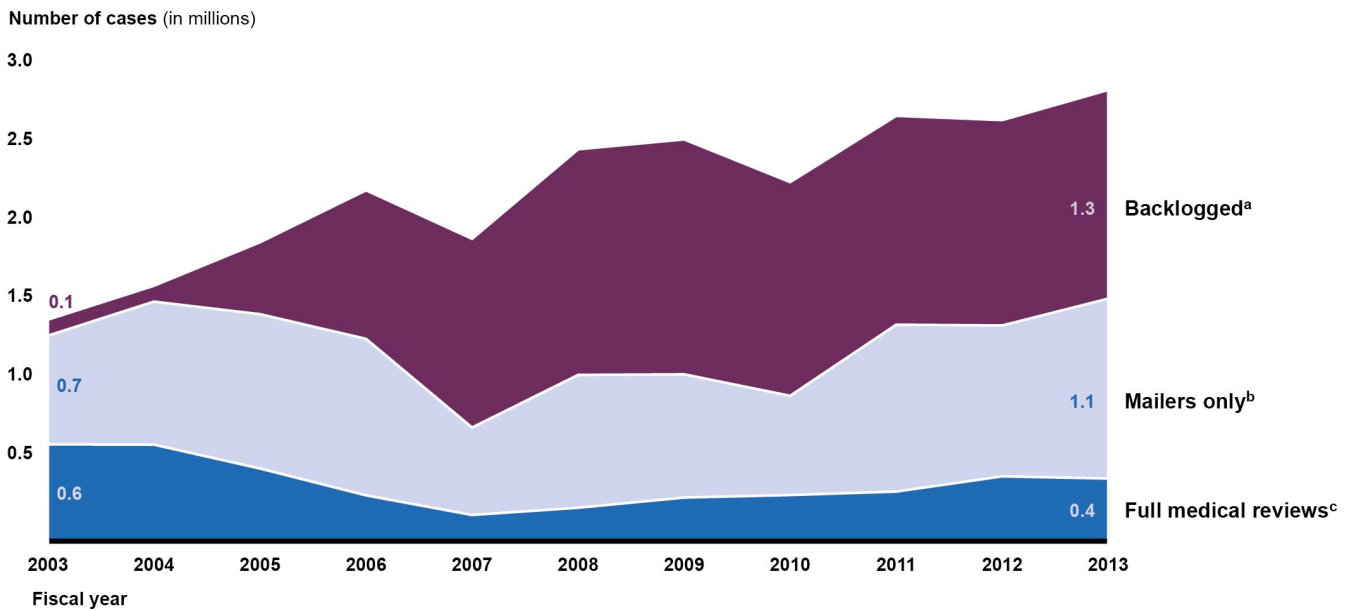
---

<sup>6</sup> According to SSA, in fiscal year 2013, the cost to process a mailer was \$13 and the cost to process a full medical review was \$1,031.

<sup>7</sup> These rates include SSA's forecast of appeals of CDR cessations and their results.

<sup>8</sup> The Contract with America Advancement Act of 1996 authorized about \$4.1 billion to be paid from the Old-Age and Survivors Insurance Trust Fund and the Disability Insurance Trust Fund to process CDRs in fiscal years 1996 through 2002. Pub. L. No. 104-121, § 103, 110 Stat. 847, 848. According to SSA, the agency used these funds to reduce the CDR backlog.

**Figure 1: Continuing Disability Reviews Completed and Backlogged by SSA, Fiscal Years 2003-2013**



Source: GAO analysis of data from Social Security Administration (SSA) Continuing Disability Review (CDR) Annual Report and data provided by SSA's Office of Public Service and Operations Support. | GAO-16-250

<sup>a</sup>When the number of cases due for a CDR exceeds SSA's capacity to conduct full medical reviews and mailers, the cases not initiated are considered backlogged for future review.

<sup>b</sup>Beneficiaries with a low likelihood of medical improvement are sent a questionnaire known as a mailer. If they respond to a mailer in certain ways, SSA may refer them for a full medical review and would not count them as "mailers only."

<sup>c</sup>Full medical reviews result in a decision to either cease or continue benefits.

SSA estimates the accuracy rate of CDRs and separately estimates the cost savings that result from CDRs. In addition to annually reporting the nationwide accuracy rate of all CDRs to the Congress, SSA internally tracks CDR accuracy rates by state and generates estimates for the accuracy of cessations and continuances, separately as well as combined.<sup>9</sup> For fiscal year 2013, SSA reported an accuracy rate of 97.2 percent

<sup>9</sup> SSA's accuracy estimates are derived from sample data and have sampling error associated with them. There are different ways to convey sampling error. A confidence interval is a range of values around the estimate that is likely to include the actual population value after repeated samples with a specified level of confidence, and it helps determine whether different estimates are significantly different from a statistical perspective. The margin of error is the maximum difference between the end point of a confidence interval and the estimate.

---

for CDR decisions.<sup>10</sup> For the same year, SSA reported an estimated ratio of federal program savings to costs for performing CDRs as \$15 to \$1. Savings from CDRs include federal benefits that would be paid to individuals were it not for a CDR that resulted in a cessation. Such benefits include those from Medicare and Medicaid because in certain situations individuals' eligibility for DI or SSI confers eligibility for these other programs.

---

## SSA Prioritizes CDRs Using Several Inputs but Does Not Fully Incorporate Potential Cost-Savings

### SSA Uses a Variety of Inputs, Including Legal Requirements and Statistical Models, to Prioritize CDRs

Because SSA does not complete all CDRs as scheduled due to competing priorities and existing resources, the agency must decide which cases will receive a full medical review. SSA uses a range of inputs to prioritize which CDRs to conduct, such as:

- **Statutory requirements:** Legal requirements to review SSI children beginning at age 18 to determine if they are eligible for benefits under adult disability criteria, and reviews of children up to 1 year old who are receiving SSI benefits due in part to low birth weight.<sup>11</sup>
- **SSA policies:** Rules established by SSA to guide prioritization. For example, SSA prioritizes cases with particular responses to its mailers and cases with a “medical improvement expected” designation that are coming up for review for the first time.

---

<sup>10</sup> SSA's most recent Annual Report on Continuing Disability Reviews is for fiscal year 2013. The estimated accuracy rate of 97.2 percent has a 95 percent confidence interval that ranges from 96.9 to 97.5 percent.

<sup>11</sup> SSA is also required to conduct periodic CDRs for certain other SSI children and DI cases due for review, but because SSA waives the requirement for many of these CDRs, when we refer to the “statutorily required” CDRs in this report, we exclude these types of cases.

- 
- **Statistical models:** A set of statistical models that score each case according to the likelihood of medical improvement, typically the sole criterion for ceasing benefits.<sup>12</sup>

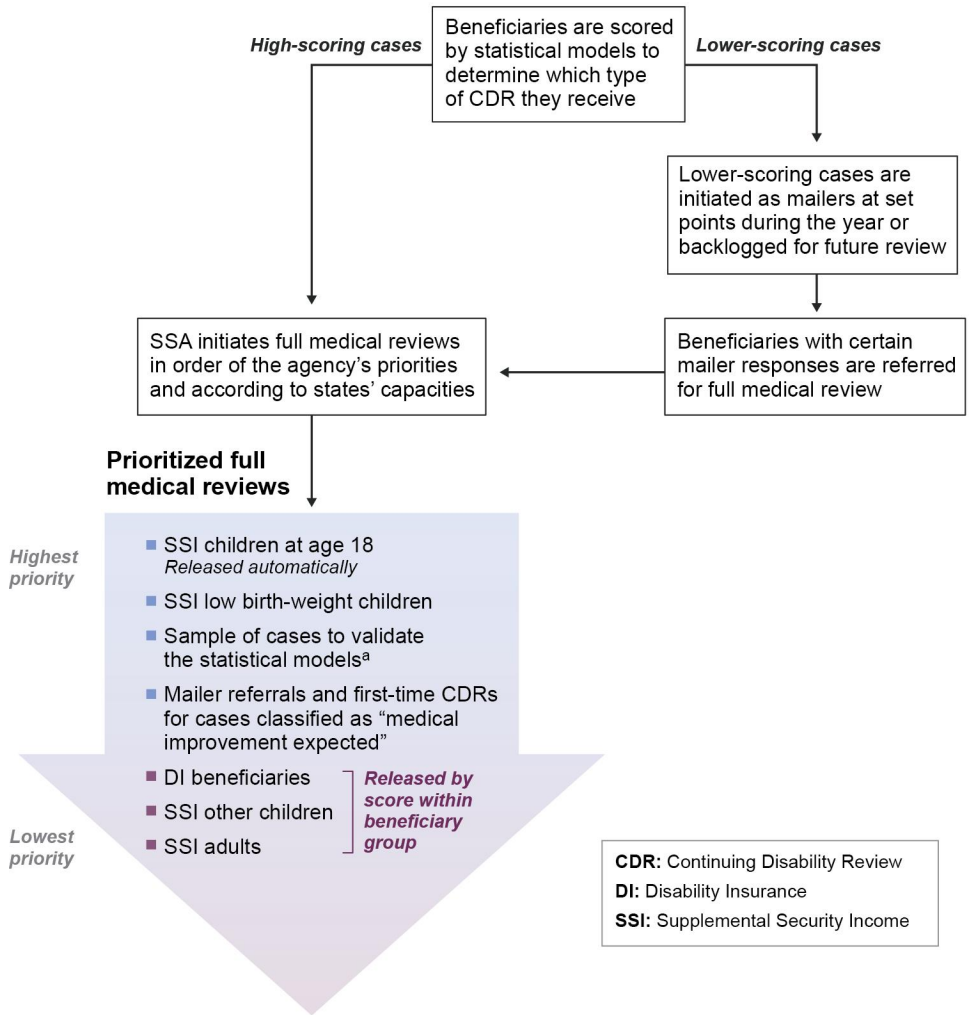
SSA's prioritization process determines which CDRs are initiated and in what form: full medical review or mailer. To begin, SSA initiates full medical reviews of cases that fall into two high-priority categories: first, statutory requirements, and then SSA policies. Once full medical reviews of all high-priority cases are initiated, SSA prioritizes the remaining cases by using its statistical models. Full medical reviews of cases with the highest scores (i.e., highest likelihood of medical improvement) are initiated as resources permit, first by beneficiary group (i.e., DI, SSI children, SSI adults) and then by the statistical scores of cases within the group (see fig. 2).<sup>13</sup> Cases with lower scores (i.e., lower likelihood of medical improvement) receive a mailer or are backlogged for future review.

---

<sup>12</sup> SSA uses five models, one for each of the following types of beneficiaries: (1) DI workers, who are disabled workers who paid into the Disability Insurance Trust Fund and meet certain eligibility criteria; (2) DI auxiliary beneficiaries, who are disabled adult children or surviving spouses of workers who paid into the Disability Insurance Trust Fund; (3) SSI adults, who are disabled adults and meet certain income requirements; (4) SSI children, who are disabled children from households that meet certain income requirements; and (5) concurrent beneficiaries, who receive both DI and SSI disability benefits.

<sup>13</sup> According to SSA officials, full medical reviews are generally conducted by DDSs in a beneficiary's state of residence. At the beginning of each fiscal year, SSA works with regional and state officials to forecast the number of reviews each state will conduct according to each state's resources and projected workloads. SSA sends cases to states on a monthly basis according to their capacity, first sending cases in high priority categories and then other cases according to beneficiary group and statistical score. DDSs initiate the review of cases from SSA in the order in which they are received.

**Figure 2: Social Security Administration’s Process for Prioritizing Continuing Disability Reviews**



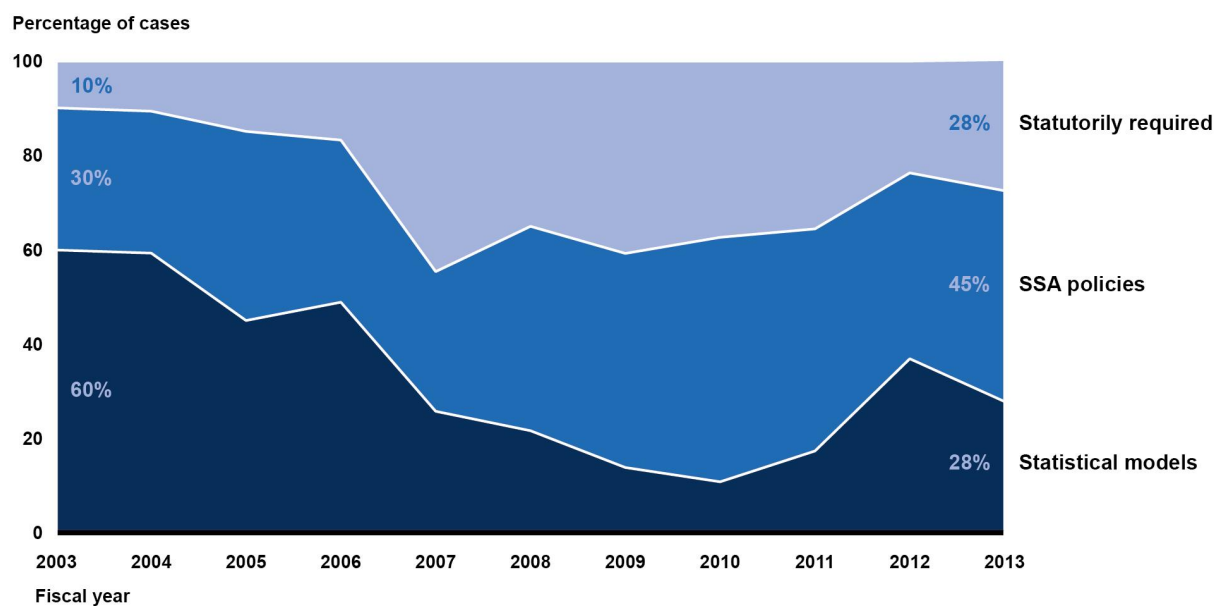
Source: GAO review of information from the Social Security Administration (SSA) Office of Public Service and Operations Support. | GAO-16-250

<sup>a</sup>SSA prioritizes a sample of about 60,000 cases annually to validate the statistical models.

The extent to which the statistical models have been used to select cases for full medical reviews has varied by year, but the models have been consistently used for determining who receives mailers. Specifically, we found that the extent to which SSA’s statistical models were used to select cases for full medical review was related to the combination of budget fluctuations for CDRs, SSA’s statutory requirements, and agency policies. We estimate that the statistical models were the basis for

selecting 11 to 60 percent of full medical reviews completed annually in fiscal years 2003 through 2013 (see fig. 3). In contrast, SSA has used the models consistently since 1993 to determine which cases should receive a mailer.

**Figure 3: Percentage of Completed Full Medical Reviews by the Basis for their Selection, Fiscal Years 2003-2013**



Source: GAO analysis of data from Social Security Administration (SSA) Continuing Disability Review Annual Reports and data from SSA's Office of Public Service and Operations Support. | GAO-16-250

Note: "Statutorily required" includes reviews of SSI children at age 18 and reviews of children up to 1 year old who are receiving SSI benefits due in part to low birth weight. "SSA policies" includes agency policies like ensuring that certain responses to mailers receive full medical reviews. SSA prioritizes a sample of about 60,000 cases annually to validate the statistical models; these cases are spread across all beneficiary groups except SSI reviews at age 18. To avoid potential double-counting, we excluded the cases used for validating the models from "SSA policies."

---

## SSA Tests the Accuracy of Its Models but Has Not Updated Them to Reflect Changing Conditions in Recent Years

Although SSA annually assesses its models' performance, the agency has not updated the models since 2007. The models' effectiveness depends on their ability to accurately predict beneficiaries' likelihood of medical improvement. To test the accuracy of the models, SSA conducts an annual validation process using a sample of completed cases to evaluate how well the models predicted medical improvement.<sup>14</sup> According to SSA officials, the model validation process has shown that the models' accuracy in predicting medical improvement has not degraded substantially in recent years.

In addition to model validation, in the past SSA has conducted periodic re-estimation of its statistical models to help ensure that they are up-to-date. In re-estimating its models, SSA updates the relationship between existing variables and medical improvement and tests whether new variables should be included. Re-estimation is particularly important when advances in medicine and assistive technology affect people's ability to work. In recent years, SSA has changed its classification of certain beneficiaries' impairments to reflect advances in medical knowledge. For example, in 2015, SSA revised its codes for cancer in light of new diagnoses and treatments. Because these codes can appear as variables in the statistical models, it is possible that the models are no longer accurately capturing the effect of cancer-related impairments on the likelihood of medical improvement. Although SSA officials believe that some advances would not markedly affect the accuracy of the models, the agency has not completed a re-estimation to confirm the effect of such changes. In addition, demographic changes in the underlying population of disability beneficiaries, which has grown substantially in recent years due in part to baby boomers reaching their disability-prone years, could also affect the accuracy of the models. The contractor that SSA hired to handle its last model re-estimation in 2007 provided SSA with a set of programs that would allow the agency to re-estimate the models in-house. Regular re-estimation and updating of predictive models is a best practice, and the contractor anticipated that SSA would do so at least every 3 years.

---

<sup>14</sup> To perform its annual validation, SSA generates a stratified random sample comprised of about 60,000 full medical reviews that it plans to initiate in the fiscal year. SSA compares predicted medical improvement with the actual improvement documented in completed cases.

---

Model accuracy leads to savings for SSA in two ways. First, model accuracy is important for identifying cases that are unlikely to result in medical improvement and can therefore be handled as mailers. According to SSA's contractor, the models' last re-estimation in 2007 increased the accuracy of the models while allowing SSA to process over 25,000 additional cases as mailers, potentially saving the agency over \$20 million by performing fewer full medical reviews. Second, greater model accuracy means the models are more likely to correctly assign high scores to cases most likely to demonstrate medical improvement, potentially leading to more medical cessations among beneficiaries who receive full medical reviews.

Although SSA acknowledged the importance of re-estimating its models again, it has yet to complete concrete actions toward doing so. In December 2015, SSA officials indicated that they were in the process of re-estimating the models, but the agency had not yet documented its efforts. In addition, according to SSA officials, the agency had not yet established plans to re-estimate the models on a regular basis. Without re-estimating its models on a regular basis, the agency risks losing the predictive accuracy of the models and could compromise its ability to use CDR resources efficiently.

---

### SSA Considers Cost Savings Information to a Limited Extent When Prioritizing CDRs

Although SSA considers cost savings when prioritizing CDR cases, it does not do so in a manner that will maximize potential savings. According to federal internal control standards, federal agencies should ensure effective stewardship of public resources.<sup>15</sup> The order in which SSA prioritizes beneficiary groups for CDRs generally aligns with the average savings per full medical review conducted for those groups in recent years.<sup>16</sup> For example, the two highest priority groups—statutorily required reviews for age-18 redeterminations and low birth-weight children—have the highest average savings in foregone disability benefits as a result of full medical reviews (see fig. 4).

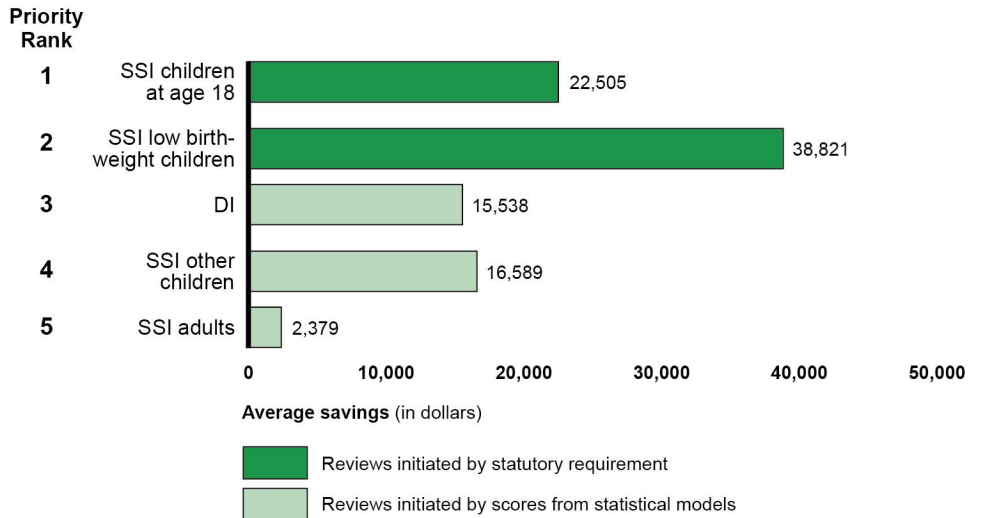
---

<sup>15</sup> GAO, *Standards for Internal Control in the Federal Government*, [GAO/AIMD-00.21.3.1](#) (Washington, D.C., November 1999).

<sup>16</sup> In comparing average cost-savings by beneficiary group, SSA considers varying savings from forgone benefits and does not consider costs because the agency's cost allocation system is not set up to determine the cost of administering full medical reviews for different groups. According to SSA officials, the cost of full medical reviews varies depending on the circumstances of individual cases, not by beneficiary group.



**Figure 4: Average Lifetime Savings per Full Medical Review by Beneficiary Group, Fiscal Years 2012-2013**



DI: Disability Insurance  
 SSI: Supplemental Security Income

Source: GAO analysis of data from the Social Security Administration (SSA) Office of the Chief Actuary. | GAO-16-250

Note: Average savings reflect the full medical reviews processed in fiscal years 2012 and 2013 and would be different if SSA had conducted a different mix of reviews in those years or if we had considered a different period of time for this analysis. We calculated average savings using weighted averages of the annual present value of the average lifetime savings per full medical review by beneficiary group.

However, the priority ranking of beneficiary groups is not exclusively reliant upon the average savings achieved from conducting full medical reviews, because the agency takes other factors into consideration. For example, the average savings per full medical review of children receiving SSI benefits on the basis of low birth weight is higher than that of SSI children at age 18. Although reviews of 18-year-olds are automatically initiated 2 months before the beneficiary’s 18th birthday, not all reviews of low birth-weight children are conducted as scheduled. Specifically, in fiscal years 2009 through 2014, approximately 3,900 to 21,700 low birth-weight reviews were backlogged annually. In addition, although the SSI Other Children group has a higher average savings in foregone disability benefits than DI beneficiaries, SSI Other Children are prioritized after DI beneficiaries. In fiscal year 2013, SSA conducted more than twice as many full medical reviews on DI beneficiaries as on SSI Other Children beneficiaries and backlogged tens of thousands more full medical reviews for SSI Other Children than for DI beneficiaries. DI cases have been given priority over SSI Other Children partly to protect the Disability

---

Insurance Trust Fund, which is the source of benefit payments to most DI recipients. However, recent action to address the solvency of the Disability Insurance Trust Fund somewhat mitigates this rationale.<sup>17</sup> If SSA had switched the number of full medical reviews conducted for these groups in 2013, it is possible that the agency would have generated over \$100 million more in savings.<sup>18</sup>

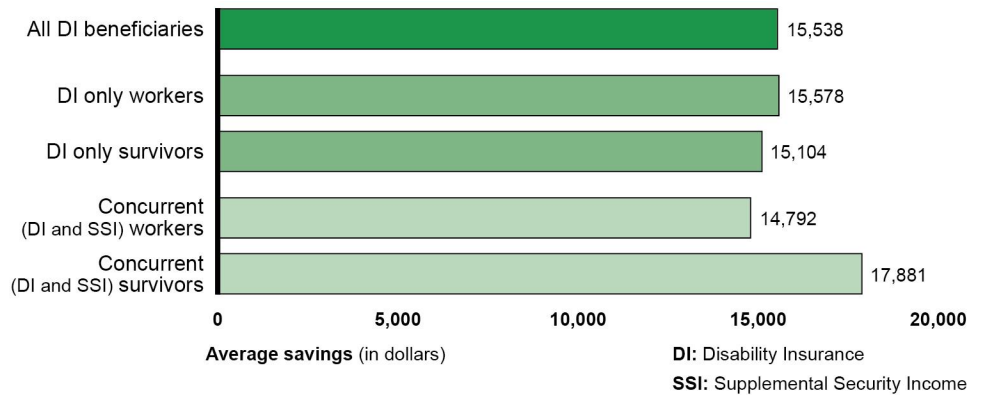
Furthermore, in focusing on beneficiary groups, SSA's prioritization process does not capture any differences among subgroups' average savings in foregone disability benefits as a result of full medical reviews. For example, the DI group can be split into four subgroups, and the average lifetime savings per full medical review among these subgroups differed by as much as approximately \$3,000 (or about 21 percent) in recent years (see fig. 5).

---

<sup>17</sup> According to the Congressional Budget Office, the Bipartisan Budget Act of 2015 is expected to delay the exhaustion of the Disability Insurance Trust Fund until fiscal year 2021. Previously, the fund was projected to be able to pay DI benefits in full on a timely basis until the fourth quarter of 2016. For more information about the solvency of the Disability Insurance Trust Fund, see GAO, *Social Security's Future: Answers to key questions*, [GAO-16-75SP](#) (Washington, D.C.: October 2015).

<sup>18</sup> In fiscal year 2013, SSA conducted about 149,000 full medical reviews of beneficiaries receiving DI only and about 53,000 full medical reviews of SSI Other Children. As a hypothetical, we estimated the savings had SSA conducted about 53,000 full medical reviews of beneficiaries receiving DI only and about 149,000 full medical reviews of SSI Other Children. For simplicity, this estimate assumes that the groups' average savings per full medical review did not change despite the different mix of reviews hypothetically conducted, which would have likely led to different average savings amounts for each group.

**Figure 5: Average Lifetime Savings per Full Medical Review by Disability Insurance Beneficiary Groups, Fiscal Years 2012-2013**



Source: GAO analysis of data from the Social Security Administration (SSA) Office of the Chief Actuary. | GAO-16-250

Note: Average savings reflect the full medical reviews processed in fiscal years 2012 and 2013 and would be different if SSA had conducted a different mix of reviews in those years or if we had considered a different period of time for this analysis. We calculated average savings using weighted averages of the annual present value of the average lifetime savings per full medical review by beneficiary group.

The aggregate mix of cases across different beneficiary groups reviewed during a fiscal year directly affects the agency’s total savings from conducting CDRs.<sup>19</sup> If SSA were to shift the mix of discretionary cases it reviews among subgroups within beneficiary groups while still taking likelihood of medical improvement into account, it could realize greater savings. For example, shifting the mix of DI cases reviewed to better align with historical average savings performance among different DI subgroups would likely increase SSA’s total savings. Furthermore, we reported in 2012 that certain subgroups of SSI children beneficiaries, such as those with speech and language disorders as well as other mental impairments, demonstrated higher rates of initial cessation (i.e., prior to the appeals process<sup>20</sup>) stemming from full medical review than other

<sup>19</sup> Within each beneficiary group or subgroup, there comes a point at which additional full medical reviews will be less likely to result in cessations (i.e., cases with medium or low scores from the statistical model, or cases at the lower end of high scores). The average savings from conducting full medical reviews for beneficiary groups will vary by year according to the number of cases in that group that ceased along with the value of foregone future disability benefits of ceased cases.

<sup>20</sup> Individuals who are dissatisfied with a determination have several opportunities for appeal within SSA, starting with a reconsideration; then a hearing before an SSA administrative law judge; and finally at the Appeals Council, which is SSA’s final administrative appeals level.

---

SSI children beneficiaries.<sup>21</sup> As a result, reviews of these subgroups are more likely to contribute to savings for the agency than other non-required SSI children reviews. Without considering average savings at the beneficiary subgroup level, SSA may not be maximizing the total savings it realizes from conducting full medical reviews.

In addition to differences in savings from shifts in the aggregate mix of cases receiving full medical reviews, savings can also differ when comparing individual cases. When an individual's benefits are ceased as the result of a CDR, the foregone benefits represent savings to the federal government. The amount of savings depends on various factors that affect how much SSA would have paid had the individual continued to receive disability benefits over time. These factors include the individual's age, life expectancy, and monthly benefit payment. For example, two individuals who are different ages but are otherwise similar (e.g., they live in the same state, have the same benefit amount, and have the same likelihood of medical improvement as determined by SSA's statistical models) would generate different expected savings from a CDR because the younger individual would likely receive benefits for a longer period of time. Similarly, two individuals who have different benefit amounts but are otherwise similar would generate different expected savings from a CDR because the individual with higher monthly benefits would likely receive greater total benefits over time. Prioritizing the CDR for the younger individual or the individual with a higher benefit level could result in greater savings for SSA. The simplified scenarios below illustrate this point; however, if SSA were to further incorporate such factors for individuals into its CDR prioritization process, a more complete set of inputs and assumptions would be needed (see fig. 6).

---

<sup>21</sup> In fiscal year 2011, the rate of initial cessation from full medical reviews was 38 percent for SSI children with speech and language delays, 39 percent for SSI children with personality disorders, and 32 percent for SSI children overall. GAO, *Supplemental Security Income: Better Oversight Needed for Children's Benefits*, [GAO-12-497](#). (Washington, D.C.: June 26, 2012).

**Figure 6: Simplified Hypothetical Scenarios Depicting Differences in Beneficiaries' Potential Lifetime Savings, Fiscal Year 2013**

When an individual's benefits are ceased through a Continuing Disability Review, the potential value of the foregone benefits depends on a variety of factors, including age and monthly benefit amount.

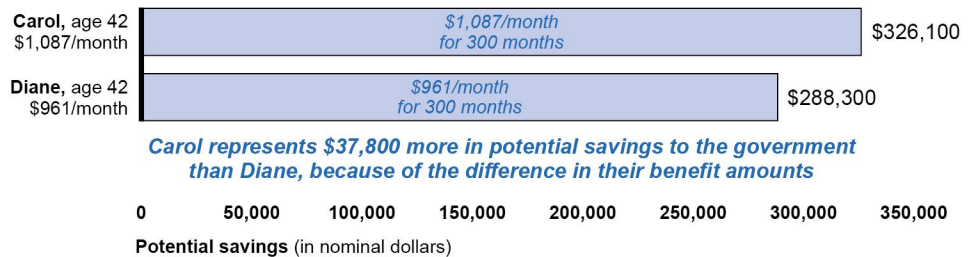
**Scenario 1**

Supplemental Security Income (SSI) beneficiaries with different ages and the same monthly benefits<sup>a</sup>



**Scenario 2**

Disability Insurance (DI) beneficiaries with the same age and different monthly benefits<sup>b</sup>



Source: GAO analysis of Social Security Administration's fiscal year 2013 statistical supplements on SSI and DI. | GAO-16-250

Note: To simplify these hypothetical scenarios, we made several assumptions: (1) all four beneficiaries have a high likelihood of medical improvement, and within each scenario, the likelihood of medical improvement is the same according to SSA's statistical models; (2) all beneficiaries have impairments that will not increase their likelihood of premature death (e.g., mental or musculoskeletal conditions); (3) all beneficiaries are expected to live beyond the age at which they would transition to retirement or aged benefits; and (4) all beneficiaries live in the same state. Our calculations include only federal disability benefits; they do not include any other programs (such as Medicare or Medicaid) for which beneficiaries might be eligible. Other factors, such as the possibility of returning to the disability rolls at a future date, were not taken into account in these calculations. The present value of future savings was not calculated in these examples for simplicity.

<sup>a</sup>The maximum monthly SSI disability benefit is set annually, and individuals' benefits are adjusted according to their income. At age 65, individuals become eligible for SSI due to their age regardless of a disability.

<sup>b</sup>The amount of monthly DI benefits is generally based on the workers' average income prior to disability, among other factors. For beneficiaries who were age 42 in fiscal year 2013, they would transition to the Social Security retirement program at their full retirement age of 67. SSA agreed that in this scenario, it could give priority to individuals with higher benefit amounts, but only in the event that both individuals have the same likelihood of medical improvement.

Despite the potentially substantial differences in savings among beneficiaries, SSA lacks a mechanism for factoring expected savings from benefit cessation into its CDR prioritization process on a case-specific basis. As a program integrity effort, CDRs are intended to assess the continued eligibility of beneficiaries to ensure that payments are made only to those individuals who should be receiving them, and SSA's statistical models use an appropriate proxy of eligibility—potential for

---

medical improvement—to prioritize cases for review. However, for beneficiaries with the same likelihood of medical improvement, SSA officials told us the agency does not further differentiate among individuals in the same beneficiary group on the basis of potential benefit savings. In SSA’s current prioritization process, the individuals depicted in the hypothetical scenarios in figure 6 would be equally likely to receive a full medical review because the agency does not consider the potential savings from individual cessations. As demonstrated in the analysis presented above, SSA could miss additional savings because it does not further consider beneficiaries’ potential savings when prioritizing cases for full medical review.

---

## SSA Reviews a Sample of CDRs for Quality but Does Not Comprehensively Analyze and Report Errors

To assess the ability of DDSs to correctly apply policy and fully document CDRs, SSA performs quality reviews of a sample of continuances and cessations.<sup>22</sup> The SSA quality reviewers who perform these reviews have guidance for checking specific elements of the decisions, and they are guided through a step-by-step computer program for conducting and documenting the reviews. SSA’s quality reviewers check CDRs for three types of errors: (1) decision errors, which include incorrect decisions or incomplete evidence to support a decision; (2) date errors, including incorrect benefit cessation dates;<sup>23</sup> and (3) administrative errors.<sup>24</sup> The reviewers return CDRs with decision errors to the DDS to perform additional work but generally correct those with date and administrative errors themselves.

SSA uses these quality reviews for multiple purposes. First, SSA estimates state, regional, and national CDR accuracy rates—the percentage of CDRs estimated to be accurate on the basis of a statistical sample. SSA also uses these accuracy rates to help monitor DDSs’ performance and shares this information with the DDSs. In addition, SSA

---

<sup>22</sup> Federal law requires SSA to review CDR continuances in the DI program to the extent necessary to ensure a high level of accuracy and provides SSA with the legal authority to review DI cessations. 42 U.S.C. § 421(c). SSA also has the authority to administer the SSI program in the same manner and subject to the same conditions set forth in the DI program. 42 U.S.C. § 1383b(a).

<sup>23</sup> Incorrect cessation dates are the most common type of date error on a CDR. According to SSA officials, other date errors, such as an incorrect onset date, are possible but rare.

<sup>24</sup> Examples of administrative errors include inputting an incorrect code for a medical impairment or adding a physician’s signature to an incorrect field.

---

uses the results of quality reviews to correct identified errors before the DDS decisions take effect.

Although SSA has reported high nationwide CDR accuracy rates in recent years,<sup>25</sup> we identified shortcomings in how SSA prevents errors, defines and reports accuracy, and samples CDRs for quality review:

- **Preventing errors:** Although SSA tracks the number and types of CDR decision errors and disseminates this information to state DDSs, it does not analyze the characteristics of CDR errors to help identify error trends associated with particular types of cases and address root causes. According to SSA officials, SSA probes CDR quality review data to uncover error trends by, for example, general groupings of impairments such as mental disorders. However, SSA does not analyze the data to uncover error trends for specific impairments, beneficiary types, or other characteristics. Federal internal control standards stipulate that management should assess the quality of performance over time and promptly resolve findings from audits and other reviews.<sup>26</sup> According to officials, SSA does not analyze the characteristics of CDRs with errors because CDR accuracy rates are generally high and resources are limited. In addition, officials stated that SSA does not have sufficient data to do statistical modeling for such analyses. However, it is possible to analyze the characteristics of CDRs with errors by comparing relevant percentages without modeling, using data from multiple years if necessary. According to SSA and DDS officials, certain types of cases may be more error-prone than others. For example, cases involving mental impairments are thought to be relatively error-prone because they can be more challenging to document. In addition, officials reported challenges in conducting CDRs of low birth-weight children receiving SSI benefits because of the lack of documentation of other impairments they may have. However, because SSA has not analyzed the incidence of inaccurate CDRs by impairment, beneficiary type, or other characteristics, it cannot efficiently identify common types of errors and their root causes to help the DDSs take steps to prevent them.

---

<sup>25</sup> From fiscal years 2010 through 2014, SSA reported a nationwide CDR accuracy rate of 97.2 percent to 97.8 percent.

<sup>26</sup> GAO, *Standards for Internal Control in the Federal Government*, [GAO/AIMD-00-21.3.1](#) (Washington, D.C.: November 1999).

- 
- **Defining and reporting accuracy:** In determining CDR accuracy rates, SSA does not include date errors, including incorrect cessation dates. As a result, decision makers do not have a complete picture of the CDR errors that affect disability payments. We have previously reported that to be useful, performance information must be complete, accurate, and valid, among other factors.<sup>27</sup> However, per SSA regulation, the agency does not consider date errors when calculating accuracy rates because date errors do not affect the decision to cease or continue benefits, according to officials.<sup>28, 29</sup> Nonetheless, such errors can affect the number of payments a beneficiary receives and thus SSA's costs. For example, cessation date errors in a CDR can result in some beneficiaries receiving payments for longer or shorter periods of time, and thus accruing overpayments or underpayments for the period in question.<sup>30</sup> Without including date errors in its reported accuracy rates, SSA does not provide its management and other decision makers and the public with complete information about errors that can affect disability payments. In addition, if SSA had counted date errors in CDR cessations, its accuracy rate for cessations in fiscal year 2014 would have fallen 1.6 percentage points from 95.5 percent to 93.9 percent. For some states, the effect of considering these errors is more pronounced. We examined SSA's fiscal year 2014 cessation accuracy rate estimates and found that for 13 states, the accuracy rates would have decreased by at least 2 percentage points had SSA counted date errors; and, in one state the accuracy rate for cessations would have fallen 7 percentage points, from 95.4 percent to 88.4 percent.

---

<sup>27</sup> GAO, *Managing For Results: GPRA Modernization Act Implementation Provides Important Opportunities to Address Government Challenges*, [GAO-11-617T](#) (Washington, D.C.: May 10, 2011).

<sup>28</sup> SSA regulations define accuracy in this context as the percentage of cases that do not have to be returned to state DDSs for further development or correction of decisions based on evidence in the files. See 20 C.F.R. §§ 404.1643, 416.1043. In explaining the accuracy standard, SSA stated that its primary purpose was to improve the initial claims process and ensure that only properly entitled claimants receive disability benefits and that its approach was to specify outputs (i.e., performance accuracy), rather than specifying all inputs that could go into the standard.

<sup>29</sup> SSA conducts stewardship reviews which examine the non-medical quality of various decisions related to benefit payments, including date designations. To do so, SSA reviews a sample of individuals receiving payments. In conducting and reporting on these reviews, however, SSA does not specifically focus on CDRs.

<sup>30</sup> SSA data do not allow us to quantify the total cost of these errors.



- 
- **Sampling CDRs for quality review:** SSA produces accuracy rate estimates by state DDS, but its sampling approach does not reliably and efficiently generate accuracy rate estimates for continuances and cessations separately in every state.<sup>31</sup> According to federal guidance for developing statistical estimates, agencies should develop a sampling plan that is reflective of the level of detail and precision needed of the key estimates.<sup>32</sup> CDR accuracy rates vary by state, and continuances are consistently more accurate than cessations. In fiscal year 2014, for example, the states' estimated CDR accuracy rates varied from 92.4 percent to 99.8 percent. In the same year, the estimated accuracy rate for continuances was 98.3 percent nationwide, whereas the equivalent for cessations was 95.5 percent. Moreover, the range of accuracy rates across states is much larger for cessations. In fiscal year 2014, state-level accuracy rates for cessations ranged from 78.3 to 100 percent, while the accuracy rates for continuances ranged from 92 to 100 percent. To monitor CDR accuracy, SSA randomly selects about 70 continuances and 70 cessations for quality review each quarter from each state.<sup>33</sup> Despite this sampling approach, SSA officials stated that their sampling design is not intended to produce precise estimates for continuations and cessations separately by state. However, precise accuracy rate estimates for continuations and cessations separately by state are needed to monitor DDS performance because of the difference in accuracy by decision type and because the state DDSs are managed separately.

In analyzing CDR workload and accuracy data, we found that SSA's sampling approach produced accuracy rate estimates with margins of error that were consistently wide in seven states and consistently narrow in six states for either one type of CDR decision or both. A wide margin of error occurs when there are not enough CDR decisions in the sample to produce a reliable estimate. In these instances, such as in Vermont and Wyoming, we found SSA could not produce an estimate with a margin of error of plus or minus 5

---

<sup>31</sup> SSA treats Washington, D.C. and Puerto Rico as states when monitoring CDR accuracy.

<sup>32</sup> Office of Management and Budget, *Standards and Guidelines for Statistical Surveys*, (Washington, D.C.: September 2006). An estimate's reliability depends on its precision. SSA considers an estimate to be reliable when it achieves a 95 percent confidence interval within plus or minus 5 percentage points.

<sup>33</sup> SSA generates state-level accuracy rates to monitor quality for a prescribed period that is usually 3 or 6 months.

---

percentage points using its current approach unless it sampled more CDR decisions. When SSA does not sample enough decisions and produces estimates with wide margins of error, decision makers may be relying on misleading information to assess CDR accuracy. Conversely, when SSA samples too many decisions and produces estimates with margins of error that are narrower than necessary to achieve reliable results, the agency may be wasting time and resources on such quality reviews.

---

## SSA Reasonably Estimates CDR Cost Savings but Has Conducted Limited Sensitivity Analysis and Documentation

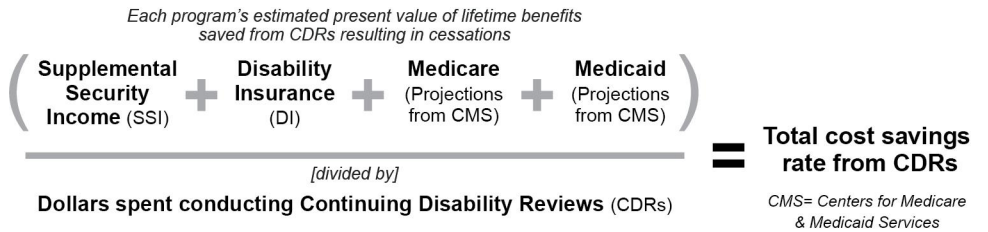
SSA's annual process for estimating the cost savings of CDRs—the estimated ratio of federal program savings to costs for performing CDRs—involves many steps. To calculate the federal program savings generated by CDRs in a particular year, SSA estimates the present value of expected future benefits over 40 years that are saved as a result of the reviews. In forecasting these savings, SSA considers benefits from programs administered by SSA (i.e., DI and SSI) as well as programs that are not administered by SSA (i.e., Medicare and Medicaid). To do so, SSA estimates the number of people whose benefits would be ceased by CDRs and considers the effect of appeals in determining these estimates.<sup>34</sup> SSA then estimates the savings associated with the cessations that are forecasted not to be overturned. It considers the age of individuals whose benefits would be ceased, and uses separate models to forecast savings from DI, SSI, Medicare, and Medicaid.<sup>35</sup> SSA also forecasts and accounts for the number and timing of beneficiaries who would stop receiving disability benefits regardless of a CDR. Similarly, it forecasts and excludes the number of former beneficiaries who will successfully reapply for benefits after, for example, a new disabling condition arises. To generate the overall CDR cost savings rate (i.e., the amount saved for every dollar invested in CDRs) for a particular year, SSA divides the present value of future benefit savings by SSA's actual cost of conducting CDRs during the relevant year (see fig. 7). To determine the cost of conducting CDRs, SSA considers its relevant expenses as well as those of the state DDSs.

---

<sup>34</sup> Specifically, to determine the number of expected cessations, SSA subtracts the number of initial cessation determinations that it forecasts will be reversed upon reconsideration or appeal.

<sup>35</sup> SSA relies on the Department of Health and Human Services' Centers for Medicare & Medicaid Services (CMS) to project savings for Medicare and Medicaid. CMS uses income distribution data from SSA to inform its savings projections.

**Figure 7: SSA’s Annual Calculation of Cost Savings Rate from Continuing Disability Reviews**



Source: GAO analysis of Social Security Administration (SSA) documents and interviews. | GAO-16-250

We determined that SSA’s methods and assumptions for estimating CDR cost savings were reasonable, but, in certain respects, inconsistent with guidance for conducting cost savings analysis of federal programs. Specifically, we identified two areas of weakness:

- Sensitivity analysis:** According to federal guidance for conducting cost savings analysis, “major assumptions should be varied and net present value and other outcomes recomputed to determine how sensitive outcomes are to changes in the assumptions.”<sup>36</sup> In reviewing SSA’s approach to estimating CDR cost savings for fiscal year 2012, we determined that SSA did not conduct sensitivity analysis of the overall cost savings rate. However, SSA separately performed some limited sensitivity analysis on savings from DI and SSI, which collectively represented about 82 percent of the savings that SSA forecasted for fiscal year 2012 CDRs.<sup>37</sup> SSA calculated the effect of using inputs from fiscal year 2011, such as the average benefit amount, on the savings estimates for fiscal year 2012. However, SSA did not vary its assumptions (e.g., from optimistic to pessimistic) to generate a range of estimated savings. In addition, SSA has not reported the effect of changing its assumptions about SSI and DI savings on the overall cost savings estimate. According to an SSA official, doing sensitivity analysis on the reported cost savings estimate would require additional coordination with CMS about Medicare and Medicaid. However, SSA

<sup>36</sup> Office of Management and Budget, Circular No. A-94 Revised, *Guidelines and Discount Rates for Benefit-Cost Analysis of Federal Programs*, October 29, 1992.

<sup>37</sup> Savings from Medicare and Medicaid represent the balance of SSA’s forecasted savings net of increased federal costs attributable to certain cessations under the Patient Protection and Affordable Care Act, according to SSA. The savings reflect present value as of the end of fiscal year 2012.

---

could conduct more complete sensitivity analysis by, for example, estimating a combined range of savings from DI and SSI without additional coordination. By not including a range of estimated savings for at least SSA's programs, decision makers lack data on the extent to which the estimates could vary under different assumptions.

- **Documentation:** According to federal guidance, models used in cost savings analysis should also be well documented and, where possible, available to facilitate independent review. SSA uses multiple complex models to estimate the cost savings of CDRs, but it has limited documentation about its methods, including data sources, assumptions, and limitations that factor into the estimates. For example, SSA has not documented how it estimates the number and timing of beneficiaries who stop receiving disability benefits because of a CDR, as well as the number of former beneficiaries who will successfully reapply for benefits. According to SSA officials, the agency has not yet documented the assumptions and procedures used to calculate CDR cost savings because of competing priorities and limited resources. Consequently, knowledge of SSA's models is limited to the few SSA actuaries who work with them, and this information is not readily available or transferrable to others, including external reviewers.

---

## Conclusions

In light of SSA's current backlog of CDRs and the long-term financial challenges of the Disability Insurance Trust Fund, conducting timely, high-quality, and cost-effective CDRs is particularly important. In an effort to use its resources efficiently, SSA applies several sound practices to help prioritize CDRs. However, without further integrating comparative cost savings information in its prioritization process, SSA is missing an opportunity not only to focus on CDRs that are likely to save the federal government the most money, but also to more efficiently use its resources for program integrity work. Maximizing cost savings is not the only goal of this work, but it is an important criterion to help SSA prioritize CDRs and ensure that beneficiaries are being more effectively selected for review.

Further, although SSA has an extensive process for reviewing the quality of CDR decisions and a high overall accuracy rate, until the agency systematically uses available data to identify error-prone cases and root causes, it will be hard-pressed to prevent similar errors from recurring. In addition, absent tracking all meaningful errors that it identifies, such as date errors, the agency and other stakeholders lack an accurate sense of the true error rate of CDRs. Similarly, SSA's current approach to sampling

---

state decisions means the agency may be relying on misleading performance information for making management decisions.

SSA has demonstrated that CDRs are cost-effective, and it applies sound methods and assumptions for estimating cost savings. However, because it does not vary the assumptions that it uses to estimate a range of potential returns on investment for CDRs, the Congress and other stakeholders do not have complete information on the precision of these estimates and the extent to which they could vary with changes in assumptions. Finally, SSA's limited documentation about its actuarial models leaves the agency vulnerable in the event of turnover of the few staff who use these models and challenges external reviewers' ability to understand and audit the integrity of its models.

---

## Recommendations for Executive Action

We recommend that the Acting Commissioner of Social Security:

1. Direct the Deputy Commissioner of Operations to further consider cost savings as part of its prioritization of full medical reviews. Such options could include considering the feasibility of prioritizing different types of beneficiaries on the basis of their estimated average savings and, as appropriate, integrating case-specific indicators of potential cost savings, such as beneficiary age and benefit amount, into its modeling or prioritization process.
2. Direct the Deputy Commissioner of Budget, Finance, Quality, and Management to complete a re-estimation of the statistical models that are used to prioritize CDRs and determine a plan for re-estimating these models on a regular basis to ensure that they reflect current conditions.
3. Direct the Deputy Commissioner of Budget, Finance, Quality, and Management to monitor the characteristics of CDR errors to identify potential root causes and report results to the Disability Determination Services. For example, SSA could analyze CDRs with and without errors to identify trends by impairment, beneficiary type, or other characteristics.
4. Direct the Deputy Commissioner of Budget, Finance, Quality, and Management to regularly track the number and rate of date errors, which can affect benefit payments (e.g., incorrect cessation dates), and consider including those errors in its reported CDR accuracy rates.

- 
5. Direct the Deputy Commissioner of Budget, Finance, Quality, and Management to adjust its approach to sampling CDRs to efficiently produce reliable accuracy rate estimates for continuances and cessations separately in each state.
  6. Direct the Chief Actuary to conduct sensitivity analyses on SSI and DI's contributions to CDR cost savings estimates and report the results reflecting a range of inputs (e.g., from optimistic to pessimistic).
  7. Direct the Chief Actuary to better document the methods including data sources, assumptions, and limitations that factor into its estimates of CDR cost savings.

---

## Agency Comments and Our Evaluation

We provided a draft of this report to SSA for review and comment, and its written comments are reproduced as appendix II in this report. SSA stated that it generally agreed with our recommendations, but noted that the level of program integrity funding it receives has affected the number of CDRs performed annually and, at times, the size of the CDR backlog. SSA also noted that our report implied that SSA is not focused on the CDRs that are most likely to save the government money and that the report did not fully convey the accuracy of the agency's statistical models and its treatment of CDR errors. We agree that SSA's CDR process is generally designed to use its program integrity resources efficiently and note in our report that SSA applies several sound practices to help prioritize CDRs including annually assessing its statistical models' performance. However, we maintain additional steps are warranted to ensure ongoing accuracy of the models and to maximize potential savings. We also note that the agency has a process in place to identify and evaluate errors, but maintain that additional steps could be taken to systematically analyze error trends and uncover root causes of errors. SSA agreed with four of our recommendations, partially agreed with one recommendation, and disagreed with two recommendations. The agency's specific concerns and our responses are described below:

- Regarding our recommendation to further consider cost savings as part of its prioritization of full medical reviews, SSA partially agreed. Although SSA agreed that it could look for ways to improve its return on conducting CDRs, it also stated that its statistical models and prioritization process already do much of what we recommend. For example, SSA stated that age is already a strong variable in its statistical models. However, these models predict medical improvement and are not designed to take expected cost savings into account. We continue to believe that to

---

maximize expected cost savings SSA could refine its prioritization process by factoring in actuarial considerations. For example, SSA could consider the effect of a beneficiary's age on expected costs savings, in addition to its existing statistical models that account for the effect of age on the likelihood of medical improvement.

- Regarding our recommendation to complete a re-estimation of the statistical models that are used to prioritize CDRs and determine a plan for re-estimating these models on a regular basis, SSA agreed and stated that it plans to complete its ongoing re-estimation and to document a process for determining when to re-estimate the models in the future.
- Regarding our recommendation to monitor the characteristics of CDRs with errors to identify root causes, SSA agreed and stated that it reports all errors to the relevant DDS for corrective action. SSA further stated that its identification of root causes is limited by the relatively few reviewed CDRs that have errors. However, in fiscal year 2014 as an example, SSA identified over 600 CDRs with errors. Although these CDRs make up a small percentage of the CDRs reviewed by SSA that year, the agency could analyze the characteristics of CDRs with errors by comparing relevant percentages without modeling. In addition, SSA could combine data from multiple years if it determined that considering more CDRs with errors would be helpful.
- Regarding our recommendation to track the number and rate of date errors and consider including them in its reported CDR accuracy rates, SSA disagreed and stated that, per SSA regulation, the agency does not consider date errors when calculating accuracy rates because date errors do not affect the decision to cease or continue benefits. SSA also stated its stewardship reviews examine the non-medical quality of benefit payment decisions. However, these reviews are not focused on CDRs, and SSA does not report results from them for CDRs specifically. SSA also explained that it does not track the number and rate of date errors because they are infrequent. However, SSA's regulations do not prevent the agency from tracking date errors, and until it does, SSA cannot definitively determine the frequency of these errors. In addition, we found that considering date errors substantially reduced some states' estimated CDR accuracy rates. Without tracking these errors, SSA cannot assess

---

their effect and consider whether including them in its reported CDR accuracy rates has merit.

- Regarding our recommendation to adjust its approach to sampling CDRs to efficiently produce reliable accuracy rate estimates for continuances and cessations separately in each state, SSA disagreed and stated that some states do not generate enough CDR decisions, particularly cessations, to generate statistically valid samples. However, for states with CDR samples that are consistently too small to produce reliable results, SSA could, for example, pool decisions from more months than it currently does to generate statistically valid samples by state. Conversely, for states with CDR samples that are consistently larger than necessary to efficiently achieve reliable results, SSA could, for example, reduce sample sizes. Because CDR accuracy rates vary by state and cessations are consistently less accurate than continuances, we maintain that SSA should adjust its approach to sampling CDRs.
- Regarding our recommendation to conduct sensitivity analyses on SSI and DI's contributions to CDR cost savings estimates and report the results reflecting a range of inputs, SSA agreed and stated that it will expand on its current sensitivity analyses as time and resources permit.
- Regarding our recommendation to better document the methods that factor into its estimates of CDR cost savings, SSA agreed and stated that it will improve and expand its existing documentation as time and resources permit.

SSA also provided technical comments, which we incorporated into the report as appropriate.

---

As agreed with your office, unless you publicly announce the contents of this report earlier, we plan no further distribution until 30 days from the report date. At that time, we will send copies to the appropriate congressional committees and the Acting Commissioner of Social Security. In addition, the report will be available at no charge on the GAO website at <http://www.gao.gov>.

If you or your staff have any questions about this report, please contact me at (202) 512-7215 or [bertonid@gao.gov](mailto:bertonid@gao.gov). Contact points for our



---

Offices of Congressional Relations and Public Affairs may be found on the last page of this report. GAO staff who made key contributions to this report are listed in appendix III.



Daniel Bertoni  
Director, Education, Workforce, and Income Security Issues

---

# Appendix I: Objectives, Scope, and Methodology

---

The objectives of this report were to examine (1) how the Social Security Administration (SSA) selects which Continuing Disability Reviews (CDR) to conduct, (2) the extent to which SSA reviews the quality of CDR decisions, and (3) how SSA calculates cost savings from CDRs.

---

## CDR Prioritization

To evaluate how SSA selects which CDRs to conduct, we reviewed relevant federal laws and interviewed SSA officials from the agency's offices of Public Service and Operations Support, Budget, and Quality Improvement. We also reviewed internal SSA documents on the agency's approach to prioritizing and processing CDRs, including its use of statistical models.

To evaluate the statistical models that SSA uses to help prioritize CDRs, we reviewed internal SSA documents about the statistical models, including lists of variables, tests of model fit, and detailed technical reports provided by the external contractor that last re-estimated the models. The technical reports provided by the contractor explained how each of the models was developed and tested, including the data sources and variables that were considered and used, how SSA impairment codes were aggregated into impairment groups, and how the functional form and interaction terms were identified. In 2007, the contractor compared the performance of models for adult beneficiaries against that of SSA's prior models—which had been estimated last in 2005—and only specifications and variables that improved model performance were retained. We evaluated the technical specifications and tests of model fit and predictive accuracy for the models for each of the beneficiary cohorts.

To estimate the proportion of full medical reviews completed because of their score from SSA's statistical models, we analyzed CDR annual report data from fiscal years 2003 through 2013. We assessed the reliability of these data by reviewing related documentation and interviewing knowledgeable agency officials, and we found these data sufficiently reliable for our purposes. To assist with the analysis, we obtained information from SSA's Office of Public Service and Operations Support on the number of SSA policy priority cases processed annually. From the total number of full medical reviews completed during a fiscal year, we subtracted completed full medical reviews that were prioritized because they were statutorily required (e.g., reviews of SSI children at age 18 and reviews of children under 1 year old who are receiving SSI benefits due in part to low birth weight) and because of SSA policy (e.g., mailers with certain responses and first-time reviews for beneficiaries in the "medical

---

improvement expected” diary category). To avoid potential double-counting, we did not count the sample of approximately 60,000 cases that SSA initiates annually to validate its statistical models among the policy priority cases because the sample consists of cases across all beneficiary groups, including statutorily required cases. As a result, our calculations may underestimate the number of required priority cases and overestimate the number of cases selected because of the statistical models.

To illustrate the impact of further incorporating cost savings into the prioritization process, we obtained and analyzed data from SSA on the average savings per full medical review by beneficiary group in each of fiscal years 2012 and 2013. Using weighted averages of these data, we calculated the average savings for fiscal years 2012 and 2013 separately and combined. We also developed two hypothetical scenarios that pair near-identical beneficiaries with different ages or monthly benefit payments to demonstrate the effect of considering cost savings on an individual basis. We used information from the fiscal year 2013 statistical supplements on the DI and SSI programs to develop reasonable ages and benefit levels for the hypothetical beneficiaries. We calculated the expected savings in foregone benefits after cessation for each beneficiary by multiplying the monthly benefit by the number of months until the beneficiaries would have aged out of the disability programs.<sup>1</sup>

---

## Quality Reviews of CDR Decisions

To understand the process that SSA uses to review the quality of CDR decisions, we reviewed relevant federal laws, regulations, policies, and procedures; interviewed SSA officials about these policies and procedures; and analyzed SSA’s CDR workload and decision accuracy data. This work included reviewing documentation of the Disability Case Adjudication and Review System (DICARS), the software program in which SSA completes quality reviews. We interviewed SSA officials about how the quality reviews are conducted and how the agency uses the results, and compared the agency’s policies and procedures to generally accepted statistical practices and federal internal control standards. We also interviewed state Disability Determination Services officials about

---

<sup>1</sup> For the DI beneficiaries, we assumed individuals aged 42 would reach full retirement age at 67 and would shift to Social Security retirement benefits at that point. For the SSI beneficiaries, we assumed they would shift to receiving SSI benefits as a result of age once they turn 65.

factors that challenge CDR quality. We assessed the reliability of SSA's CDR workload and decision accuracy data by performing data testing, reviewing related documentation, and interviewing agency officials, and we found the data to be sufficiently reliable for the purposes of this review.

To evaluate the extent to which SSA reviews the quality of CDR decisions, we analyzed SSA's CDR workload and decision accuracy data to determine whether its method for sampling CDRs and estimating CDR accuracy are consistent with generally accepted statistical practices and SSA's reporting goals. SSA reports accuracy rate estimates for each state every month using the most recent 3 or 6 months of quality review data.<sup>2</sup> Its goal is to produce estimates with 95 percent confidence intervals that are within plus or minus 5 percentage points of the estimate. We analyzed SSA's CDR workload and accuracy data, consistent with SSA's sampling and reporting methods, from June 2013 through April 2015. Specifically, we identified the number of continuance and cessation determinations in each state, the District of Columbia, and Puerto Rico; SSA's accuracy rate estimates for these determinations for each 6-month period;<sup>3</sup> and the 95 percent confidence interval margins of error for each estimate. To identify states that had estimates that consistently do not achieve the reporting goals, we compared the workload, accuracy rate estimates, and margins of errors to those specified in SSA's sample design and reporting goals. We calculated margins of error for estimates in which SSA did not provide them. We used a statistical formula that produces appropriate margins of error, including when standard formula do not apply, to determine and examine a margin of error for all estimates.<sup>4</sup> We chose this formula because in some cases CDR accuracy is so high or the sampling fraction is so large that the standard statistical formula used for these purposes would compute margins of error that are not appropriate, such as those resulting in confidence

---

<sup>2</sup> SSA generates state-level accuracy rates to monitor quality for a prescribed period that is usually 3 or 6 months. When states do not have enough decisions during a 3-month period to form reliable estimates, SSA may use 6 months of quality review data.

<sup>3</sup> An accuracy rate estimate for April 2015, for example, includes the quality reviews conducted during the 6 months from November 2014 through April 2015.

<sup>4</sup> We used the extreme proportions (or hypergeometric) formula, which is appropriate when normality does not apply and when the finite population correction factor, which is defined as the sample size divided by the population size, cannot be ignored. See William G. Cochran, *Sampling Techniques*, 3rd edition. (New York: John Wiley & Sons, 1977), 57.

---

intervals above 100 percent.<sup>5</sup> Since CDR accuracy cannot be greater than 100 percent, the standard formula is not appropriate. We also analyzed CDR workload data from fiscal years 2001 through 2014 to inform our evaluation of SSA's sampling method.

To determine the effect of date errors on accuracy rates, we analyzed data about CDR date errors and cessations. We considered data from fiscal years 2010 through 2014 to determine the frequency with which date errors occur. We calculated fiscal year 2014 cessation accuracy rate estimates for each state by combining the number of cessation decision errors and the number of date errors and dividing the total by the number of cessations SSA reviewed. SSA's date error data were not broken down by decision type (i.e., continuance or cessation), but we assigned these errors to cessations because of input from SSA and our corroborating analysis. According to SSA officials, the most common date error on a CDR is a cessation date error and other date errors, such as incorrectly inputting an onset date, can occur in a cessation or continuance but are rare. Our analysis corroborated this information. For example, in fiscal year 2014, of 127 date errors identified in CDRs nationally, 125 of them were cessation date errors. In addition, we calculated margins of error for each estimate to assess the statistical reliability of each estimate.<sup>6</sup> We used the statistical formula that produces appropriate margins of error, consistent with our approach to calculating margins of error in our analysis of SSA's sampling method.

---

## CDR Cost-Savings Estimate

To evaluate SSA's approach to calculating cost savings from CDRs, we compared SSA's estimation process to actuarial standards of practice and federal guidelines for benefit-cost analyses of federal programs. Specifically, we interviewed SSA actuaries about the models and methods they used to perform the cost-savings calculation for fiscal year

---

<sup>5</sup> The normality formula is a standard statistical formula used for calculating margins of error when the assumptions of normality apply.

<sup>6</sup> When the 95 percent confidence intervals of accuracy rate estimates do not overlap, accuracy rate estimates are considered statistically different at the 95 percent level of confidence. Using statistical significance to compare SSA and our cessation accuracy rates is difficult because the sampling method is not designed to detect differences of less than 10 percentage points (i.e., the designed margin of error is plus or minus 5 percentage points). Therefore, we did not assess the significance of the difference between SSA's cessation accuracy rate estimates by state and our respective estimates, which include date errors.

2012. We also reviewed portions of the programming code related to these models to corroborate the information from the actuaries. In addition, we examined the assumptions that SSA uses to calculate the present value of future benefits saved from ceasing a person's benefits as the result of a CDR by examining where and how SSA incorporates assumptions into its calculation process. Finally, we reviewed the fiscal year 2012 CDR cessation data and information SSA provided to the Centers for Medicare & Medicaid Services (CMS) that informed CMS's estimates of Medicare and Medicaid savings resulting from CDR cessations, but we did not review CMS's models.

We conducted this performance audit from December 2014 to February 2016 in accordance with generally accepted government auditing standards. Those standards require that we plan and perform the audit to obtain sufficient, appropriate evidence to provide a reasonable basis for our findings and conclusions based on our audit objectives. We believe that the evidence obtained provides a reasonable basis for our findings and conclusions based on our audit objectives.

# Appendix II: Comments from the Social Security Administration



**SOCIAL SECURITY**  
Office of the Commissioner

January 20, 2016

Mr. Daniel Bertoni  
Director, Education, Workforce,  
and Income Security Issues  
United States Government Accountability Office  
441 G Street, NW  
Washington, DC 20548

Dear Mr. Bertoni:

Thank you for the opportunity to review the draft report, "SOCIAL SECURITY DISABILITY: SSA Could Increase Savings by Refining Its Selection of Cases for Disability Review" (GAO-16-250). Please see our enclosed comments.

If you have any questions, please contact me at (410) 965-0520. Your staff may contact Gary S. Hatcher, Senior Advisor for Records Management and Audit Liaison Staff, at (410) 965-0680.

Sincerely,

A handwritten signature in blue ink, appearing to read "Frank Cristaudo".

Frank Cristaudo  
Executive Counselor to the Commissioner

Enclosure

SOCIAL SECURITY ADMINISTRATION BALTIMORE, MD 21235-0001

**COMMENTS ON THE GOVERNMENT ACCOUNTABILITY OFFICE (GAO) DRAFT REPORT, "SOCIAL SECURITY DISABILITY: SSA COULD INCREASE SAVINGS BY REFINING ITS SELECTION OF CASES FOR DISABILITY REVIEW" (GAO-16-250)**

**GENERAL COMMENTS**

While we generally agree with the recommendations, we do have some concerns with the context of the report as it relates to the discussion of our program integrity funding, our statistical predictive models, and our treatment of error cases. We outline our concerns in the sections below.

**Program Integrity Funding**

Our analysis shows, that for a long period, beginning in fiscal year (FY) 2006, the Continuing Disability Review (CDR) backlog began to grow due to limited funding, little hiring, and record increases in the initial disability claims workload, which competed for the same resources required for CDRs. Due to lack of resources, we were unable to complete many high priority CDRs, including some of the categories mentioned in this report. On the other hand, our analysis demonstrates with the seven-year plan to reduce the CDR backlog, that we are very capable of managing the CDR workload when we receive sustained, predictable program integrity funding levels.

In FY 2011, the Budget Control Act (BCA) authorized a sustained string of program integrity funding so we could rebuild our CDR capacity and work towards reducing our CDR backlog. However, Congress did not appropriate the full BCA level of funding until FY 2014. Although we made some progress in FY 2012 and FY 2013, the backlog lingered at 1.3 million. In FY 2014, we received sufficient program integrity funding to begin rebuilding CDR capacity. That year, we completed over 525,000 full medical reviews (including 89,000 other children), the largest number since FY 2005. In FY 2015, we completed 799,000 CDRs, including 224,000 other children CDRs, the most other children CDRs ever processed in one year. Of that 224,000, over 50,000 had a primary diagnosis code for speech and language delay, one of the cohorts this draft report mentions. We accomplished this using our current CDR prioritization process. By the end of FY 2015, we reduced the CDR backlog by almost 50 percent, and the other children backlog by almost 60 percent, relative to the backlog at the end of FY 2013.

**Statistical Predictive Models**

We are concerned with the characterization of the size and accuracy of our statistical model. The key purpose of our large program integrity sample is to provide ongoing measurement and validation of the accuracy, consistency, and reliability of our predictive model. The data shows that the performance of the CDR models did not indicate significant deterioration since our last re-estimation. The models independently predict a probability of medical improvement for each beneficiary, based on their individual case characteristics. The accuracy of the model should be independent of the number of beneficiaries in the program. Any implied conclusion that we are missing an opportunity to focus on CDRs that are likely to save the Federal government the most



money is inaccurate. Our medical CDR process is, in fact, designed to use scarce program integrity resources efficiently and effectively.

**CDR Errors**

GAO suggests that we analyze error trends by detailed characteristics on the CDR Quality Assurance (QA) sample. We have previously indicated to GAO, that the very small number of errors found each year does not provide sufficient data to conduct robust analysis, even if we were to pool a number of years together to analyze the CDR QA samples. With error rates of only 2-3 percent in the entire population, there are few, if any, “common” errors or types of cases that are “error prone” revealed in the QA sample. We do identify and analyze each and every error found through our CDR QA process and report all of these errors to the relevant Disability Determination Services (DDS) staff for corrective action. However, due to the very small number of errors found each year, our ability to systematically analyze CDR decisional accuracy errors is limited. All errors are returned to the State DDS for appropriate action or corrected by our quality reviewers. Our ongoing stewardship reviews focus on the accuracy of benefit amounts.

**RECOMMENDATION RESPONSES**

**Recommendation 1**

Further consider cost savings as part of its prioritization of full medical reviews. Such options could include prioritizing different types of beneficiaries on the bases of their estimated average savings, and integrating case-specific indicators of potential cost savings such as beneficiary age and benefit amount into its modeling or prioritization process.

**Response**

We agree we should constantly analyze our prioritization process and look for ways to improve it to get the best return.

Our predictive models and CDR prioritizing process already achieve much of what GAO recommends. Age is a very strong variable in our predictive models. For instance, in scenario 1 on page 16, GAO suggests that if all things are equal we should release the younger individual first due to the longer payback period for saved benefits. In fact, if everything about the two hypothetical individuals were truly the same except for age, our models would always score the younger person higher and the CDR for that person would always be released first.

On page 16, scenario 2, GAO shows two hypothetical beneficiaries who have the same likelihood of medical improvement, and recommends releasing the one with the higher benefit. In practice, it is highly unlikely that we would release one and not the other. We can certainly consider analyzing our data to see if it would be cost effective to use the benefit amount as a tiebreaker for CDRs scoring the same. However, we do not expect this would affect a large number of CDRs, and we do not recommend using benefit rates to prioritize CDRs with different scores.

Lastly, on page 14, GAO suggests prioritizing the DI beneficiary groups by average estimated savings. Concurrent survivors are the only group that could produce significantly higher projected savings. However, this group has low benefit amounts, and one of the reasons this group has higher savings projections is that it has slightly higher cessation rates than the other groups. Since we prioritize based on expected cessation rates, these cases are already among the first to be released from the Title II groups.

**Recommendation 2**

Complete a re-estimation of the statistical models that are used to prioritize CDRs, and determine a plan for re-estimating these models on a regular basis to ensure that they reflect current conditions.

**Response**

Although our large profile sample did not indicate any significant deterioration in the performance of the models, we agree that they should be periodically re-estimated. We started the re-estimation in FY 2015 and will provide a copy of our documentation when it is complete. We will also document our process for analyzing “when” to re-estimate the models.

**Recommendation 3**

Monitor the characteristics of CDR errors to identify potential root causes and report results to the Disability Determination Services. For example, SSA could analyze CDRs with and without errors to identify trends by impairment, beneficiary type, or other characteristics.

**Response**

We agree. We do identify and analyze each and every error found through our CDR quality assurance process and report all of these errors to the relevant DDS for corrective action. For identifying root causes, we are limited by the data available due to the low number of errors (2-3 percent) annually.

**Recommendation 4**

Regularly track the number and rate of date errors, which can affect benefit payments (e.g., incorrect cessation dates), and consider including those errors in its reported CDR accuracy rates.

**Response**

We disagree. Our CDR quality reviews examine the medical determination. Our Stewardship reviews examine the non-medical quality of benefit payment decisions. The Stewardship model provides non-medical payment accuracy performance measures for Title II and Title XVI payments. We model the accuracy rate data for CDRs after our initial QA data. The Code of Federal Regulations 404.1640 -404-1642 mandates the performance standards for the initial QA

accuracy rates. According to the regulations, the definition of performance accuracy includes the measurement of factors that have a potential for affecting a decision, as well as the correctness of the decision. We do not include onset and cessation date errors, because the medical decision in these cases is correct.

In addition, we do not track the number and rate of date errors because of their infrequency, especially at the State level. It is not cost beneficial to track and report date errors on a regular basis. As we previously shared with GAO, we can produce this data; however, we will not create and maintain a report that will likely be unused.

**Recommendation 5**

Adjust its approach to samplings CDRs to efficiently produce reliable accuracy rate estimates for continuances and cessations separately in each State.

**Response**

We disagree. States with smaller volumes of cases do not provide enough cases in the universe, especially cessations, for us to be able to select a statistically valid sample. We would have to conduct a 100 percent review of the cases, which does not constitute statistical sampling. A 100 percent review is a highly resource intensive undertaking, and would provide limited additional value to the agency's CDR QA accuracy reporting.

**Recommendation 6**

Conduct sensitivity analyses on SSI and DI's contributions to CDR cost savings estimates and report the results reflecting a range of inputs (e.g., from optimistic to pessimistic).

**Response**

We agree. As was shared with GAO during the audit, we already conduct some sensitivity analyses during the estimation process. However, in the future, we can try to expand on what is already done as time and resources permit.

**Recommendation 7**

Better document the methods including data sources, assumptions, and limitations that factor into its estimates of CDR cost savings.

**Response**

We agree. A certain amount of documentation of the estimation process already exists sufficient for us to complete this work. However, we do agree that the existing documentation could be improved and expanded, and we will try to do this going forward as time and resources permit.

---

# Appendix III: GAO Contact and Staff Acknowledgments

---

## GAO Contact

Daniel Bertoni, (202) 512-7215 or [bertonid@gao.gov](mailto:bertonid@gao.gov)

---

## Staff Acknowledgments

In addition to the contact named above, Erin Godtland (Assistant Director), Joel Green (Analyst-in-Charge), Susan Aschoff, James Bennett, Grace Cho, Alexander Galuten, Emei Li, Isabella Johnson, Rhiannon Patterson, Almeta Spencer, Daren Sweeney, Jeff Tessin, Kristen Timko, Frank Todisco, Sonya Vartivarian, and Shana Wallace made key contributions to this report.

---

# Appendix IV: Accessible Data

---

## Agency Comment Letter

---

### Text of Appendix II: Comments from the Social Security Administration

---

Page 1

SOCIAL SECURITY

Office of the Commissioner

January 20, 2016

Mr. Daniel Bertoni

Director, Education, Workforce, and Income Security Issues

United States Government Accountability Office

441 G Street, NW

Washington, DC 20548

Dear Mr. Bertoni:

Thank you for the opportunity to review the draft report, "SOCIAL SECURITY DISABILITY: SSA Could Increase Savings by Refining Its Selection of Cases for Disability Review" (GAO-16-250). Please see our enclosed comments.

If you have any questions, please contact me at (410) 965-0520. Your staff may contact Gary S. Hatcher, Senior Advisor for Records Management and Audit Liaison Staff, at (410) 965-0680.

Sincerely,

Frank Cristaudo

Executive Counselor to the Commissioner

---

Enclosure

---

Page 2

COMMENTS ON THE GOVERNMENT ACCOUNTABILITY OFFICE (GAO) DRAFT REPORT, "SOCIAL SECURITY DISABILITY: SSA COULD INCREASE SAVINGS BY REFINING ITS SELECTION OF CASES FOR DISABILITY REVIEW" (GAO-16-250)

#### GENERAL COMMENTS

While we generally agree with the recommendations, we do have some concerns with the context of the report as it relates to the discussion of our program integrity funding, our statistical predictive models, and our treatment of error cases. We outline our concerns in the sections below.

#### Program Integrity Funding

Our analysis shows, that for a long period, beginning in fiscal year (FY) 2006, the Continuing Disability Review (CDR) backlog began to grow due to limited funding, little hiring, and record increases in the initial disability claims workload, which competed for the same resources required for CDRs. Due to lack of resources, we were unable to complete many high priority CDRs, including some of the categories mentioned in this report. On the other hand, our analysis demonstrates with the seven-year plan to reduce the CDR backlog, that we are very capable of managing the CDR workload when we receive sustained, predictable program integrity funding levels.

In FY 2011, the Budget Control Act (BCA) authorized a sustained string of program integrity funding so we could rebuild our CDR capacity and work towards reducing our CDR backlog. However, Congress did not appropriate the full BCA level of funding until FY 2014. Although we made some progress in FY 2012 and FY 2013, the backlog lingered at 1.3 million. In FY 2014, we received sufficient program integrity funding to begin rebuilding CDR capacity. That year, we completed over 525,000 full medical reviews (including 89,000 other children), the largest number since FY 2005. In FY 2015, we completed 799,000 CDRs, including 224,000 other children CDRs, the most other children CDRs ever processed in one year. Of that 224,000, over 50,000 had a primary diagnosis code for speech and language delay, one of the cohorts this draft report mentions. We accomplished this using our current CDR prioritization process. By the end of FY 2015, we reduced the CDR backlog by almost 50 percent, and the other children backlog by almost 60 percent, relative to the backlog at the end of FY 2013.

---

### Statistical Predictive Models

We are concerned with the characterization of the size and accuracy of our statistical model. The key purpose of our large program integrity sample is to provide ongoing measurement and validation of the accuracy, consistency, and reliability of our predictive model. The data shows that the performance of the CDR models did not indicate significant deterioration since our last re-estimation. The models independently predict a probability of medical improvement for each beneficiary, based on their individual case characteristics. The accuracy of the model should be independent of the number of beneficiaries in the program. Any implied conclusion that we are missing an opportunity to focus on CDRs that are likely to save the Federal government the most

Page 3

money is inaccurate. Our medical CDR process is, in fact, designed to use scarce program integrity resources efficiently and effectively.

### CDR Errors

GAO suggests that we analyze error trends by detailed characteristics on the CDR Quality Assurance (QA) sample. We have previously indicated to GAO, that the very small number of errors found each year does not provide sufficient data to conduct robust analysis, even if we were to pool a number of years together to analyze the CDR QA samples. With error rates of only 2-3 percent in the entire population, there are few, if any, “common” errors or types of cases that are “error prone” revealed in the QA sample. We do identify and analyze each and every error found through our CDR QA process and report all of these errors to the relevant Disability Determination Services (DDS) staff for corrective action. However, due to the very small number of errors found each year, our ability to systematically analyze CDR decisional accuracy errors is limited. All errors are returned to the State DDS for appropriate action or corrected by our quality reviewers. Our ongoing stewardship reviews focus on the accuracy of benefit amounts.

### RECOMMENDATION RESPONSES

#### Recommendation 1

Further consider cost savings as part of its prioritization of full medical reviews. Such options could include prioritizing different types of beneficiaries on the bases of their estimated average savings, and integrating case-specific indicators of potential cost savings such as

---

beneficiary age and benefit amount into its modeling or prioritization process.

Response

We agree we should constantly analyze our prioritization process and look for ways to improve it to get the best return.

Our predictive models and CDR prioritizing process already achieve much of what GAO recommends. Age is a very strong variable in our predictive models. For instance, in scenario 1 on page 16, GAO suggests that if all things are equal we should release the younger individual first due to the longer payback period for saved benefits. In fact, if everything about the two hypothetical individuals were truly the same except for age, our models would always score the younger person higher and the CDR for that person would always be released first.

On page 16, scenario 2, GAO shows two hypothetical beneficiaries who have the same likelihood of medical improvement, and recommends releasing the one with the higher benefit. In practice, it is highly unlikely that we would release one and not the other. We can certainly consider analyzing our data to see if it would be cost effective to use the benefit amount as a tiebreaker for CDRs scoring the same. However, we do not expect this would affect a large number of CDRs, and we do not recommend using benefit rates to prioritize CDRs with different scores.

---

Page 4

Lastly, on page 14, GAO suggests prioritizing the DI beneficiary groups by average estimated savings. Concurrent survivors are the only group that could produce significantly higher projected savings. However, this group has low benefit amounts, and one of the reasons this group has higher savings projections is that it has slightly higher cessation rates than the other groups. Since we prioritize based on expected cessation rates, these cases are already among the first to be released from the Title II groups.

Recommendation 2

Complete a re-estimation of the statistical models that are used to prioritize CDRs, and determine a plan for re-estimating these models on a regular basis to ensure that they reflect current conditions.

Response



Although our large profile sample did not indicate any significant deterioration in the performance of the models, we agree that they should be periodically re-estimated. We started the re-estimation in FY 2015 and will provide a copy of our documentation when it is complete. We will also document our process for analyzing “when” to re-estimate the models.

#### Recommendation 3

Monitor the characteristics of CDR errors to identify potential root causes and report results to the Disability Determination Services. For example, SSA could analyze CDRs with and without errors to identify trends by impairment, beneficiary type, or other characteristics.

#### Response

We agree. We do identify and analyze each and every error found through our CDR quality assurance process and report all of these errors to the relevant DDS for corrective action. For identifying root causes, we are limited by the data available due to the low number of errors (2-3 percent) annually.

#### Recommendation 4

Regularly track the number and rate of date errors, which can affect benefit payments (e.g., incorrect cessation dates), and consider including those errors in its reported CDR accuracy rates.

#### Response

We disagree. Our CDR quality reviews examine the medical determination. Our Stewardship reviews examine the non-medical quality of benefit payment decisions. The Stewardship model provides non-medical payment accuracy performance measures for Title II and Title XVI payments. We model the accuracy rate data for CDRs after our initial QA data. The Code of Federal Regulations 404.1640 -404-1642 mandates the performance standards for the initial QA

accuracy rates. According to the regulations, the definition of performance accuracy includes the measurement of factors that have a potential for affecting a decision, as well as the correctness of the decision. We do not include onset and cessation date errors, because the medical decision in these cases is correct.

In addition, we do not track the number and rate of date errors because of their infrequency, especially at the State level. It is not cost beneficial to track and report date errors on a regular basis. As we previously shared with GAO, we can produce this data; however, we will not create and maintain a report that will likely be unused.

#### Recommendation 5

Adjust its approach to samplings CDRs to efficiently produce reliable accuracy rate estimates for continuances and cessations separately in each State.

#### Response

We disagree. States with smaller volumes of cases do not provide enough cases in the universe, especially cessations, for us to be able to select a statistically valid sample. We would have to conduct a 100 percent review of the cases, which does not constitute statistical sampling. A 100 percent review is a highly resource intensive undertaking, and would provide limited additional value to the agency's CDR QA accuracy reporting.

#### Recommendation 6

Conduct sensitivity analyses on SSI and DI's contributions to CDR cost savings estimates and report the results reflecting a range of inputs (e.g., from optimistic to pessimistic).

#### Response

We agree. As was shared with GAO during the audit, we already conduct some sensitivity analyses during the estimation process. However, in the future, we can try to expand on what is already done as time and resources permit.

#### Recommendation 7

Better document the methods including data sources, assumptions, and limitations that factor into its estimates of CDR cost savings.

#### Response

We agree. A certain amount of documentation of the estimation process already exists sufficient for us to complete this work. However, we do agree that the existing documentation could be improved and expanded, and we will try to do this going forward as time and resources permit.

## Data Tables

**Data Table for Highlights Figure: CDRs Completed and Backlogged by SSA, Fiscal Years 2003-2013**

Number of cases (in millions)

Fiscal year	In-depth reviews	Questionnaires only	Backlogged
2003	0.617	0.692	0.102
2004	0.615	0.911	0.101
2005	0.463	0.983	0.458
2006	0.292	0.997	0.946
2007	0.167	0.557	1.202
2008	0.213	0.846	1.437
2009	0.278	0.785	1.496
2010	0.294	0.632	1.361
2011	0.316	1.063	1.33
2012	0.413	0.961	1.308
2013	0.399	1.146	1.329

Source: GAO analysis of data from Social Security Administration (SSA) Continuing Disability Review (CDR) Annual Report and data provided by SSA's Office of Public Service and Operations Support. | GAO-16-250

**Data Table for Figure 1: Continuing Disability Reviews Completed and Backlogged by SSA, Fiscal Years 2003-2013**

Number of cases (in millions)

Fiscal year	Full medical reviews <sup>c</sup>	Mailers only <sup>b</sup>	Backlogged <sup>a</sup>
2003	0.617	0.692	0.102
2004	0.615	0.911	0.101
2005	0.463	0.983	0.458
2006	0.292	0.997	0.946
2007	0.167	0.557	1.202
2008	0.213	0.846	1.437
2009	0.278	0.785	1.496
2010	0.294	0.632	1.361
2011	0.316	1.063	1.33
2012	0.413	0.961	1.308

Fiscal year	Full medical reviews <sup>c</sup>	Mailers only <sup>b</sup>	Backlogged <sup>a</sup>
2013	0.399	1.146	1.329

Source: GAO analysis of data from Social Security Administration (SSA) Continuing Disability Review (CDR) Annual Report and data provided by SSA's Office of Public Service and Operations Support. | GAO-16-250

**Accessible Text for Figure 2: Social Security Administration's Process for Prioritizing Continuing Disability Reviews**

1. Beneficiaries are scored by statistical models to determine which type of Continuing Disability Review (CDR) they receive
2. Lower-scoring cases are initiated as mailers at set points during the year or backlogged for future review
  - a. Beneficiaries with certain mailer responses are referred for full medical review
3. For high-scoring cases, SSA initiates full medical reviews in order of the agency's priorities and according to states' capacities
4. Full medical reviews, prioritized from highest to lowest
  - a. Supplemental Security Income (SSI) children at age 18 are released automatically
  - b. SSI low birth-weight children
  - c. Sample of cases to validate the statistical models<sup>a</sup>
  - d. Mailer referrals and first-time CDRs for cases classified as "medical improvement expected"
  - e. Disability Insurance (DI) beneficiaries [Released by score within beneficiary group]
  - f. SSI other children [Released by score within beneficiary group]
  - g. SSI adults [Released by score within beneficiary group]

Source: GAO review of information from the Social Security Administration (SSA) Office of Public Service and Operations Support. | GAO-16-250

**Data Table for Figure 3: Percentage of Completed Full Medical Reviews by the Basis for their Selection, Fiscal Years 2003-2013**

Percentage of cases

Fiscal year	Statistical models	SSA policies	Statutorily required
2003	59.94	30.14	9.92
2004	59.28	30.1	10.62

Fiscal year	Statistical models	SSA policies	Statutorily required
2005	44.99	40.1	14.91
2006	48.87	34.38	16.76
2007	25.74	29.62	44.64
2008	21.62	43.34	35.04
2009	13.82	45.4	40.78
2010	10.79	51.84	37.38
2011	17.32	47.1	35.58
2012	36.88	39.41	23.71
2013	27.8	44.7	27.8

**Data Table for Figure 4: Average Lifetime Savings per Full Medical Review by Beneficiary Group, Fiscal Years 2012-2013**

Category	Priority Rank	Reviews initiated by	Average savings (in dollars)
Supplemental Security Income (SSI) children at age 18	1	Statutory requirement	22,505
SSI low birth-weight children	2	Statutory requirement	38,821
Disability Insurance	3	Scores from statistical models	15,538
SSI other children	4	Scores from statistical models	16,589
SSI adults	5	Scores from statistical models	2,379

Source: GAO analysis of data from the Social Security Administration (SSA) Office of the Chief Actuary. | GAO-16-250

**Data Table for Figure 5: Average Lifetime Savings per Full Medical Review by Disability Insurance Beneficiary Groups, Fiscal Years 2012-2013**

Category	Average savings (in dollars)
All Disability Insurance (DI) beneficiaries	15,538
DI only workers	15,578
DI only survivors	15,104
Concurrent (DI and SSI) workers	14,792
Concurrent (DI and SSI) survivors	17,881

Source: GAO analysis of data from the Social Security Administration (SSA) Office of the Chief Actuary. | GAO-16-250

**Accessible Text for Figure 6: Simplified Hypothetical Scenarios Depicting Differences in Beneficiaries' Potential Lifetime Savings, Fiscal Year 2013**

When an individual's benefits are ceased through a Continuing Disability Review, the potential value of the foregone benefits depends on a variety of factors, including age and monthly benefit amount.

Scenario 1: Supplemental Security Income (SSI) beneficiaries with different ages and the same monthly benefits<sup>a</sup>

Individual	Age	Monthly benefit	Duration	Potential savings (in nominal dollars)
"Alan"	51	\$550	168 months	\$92,400
"Bob"	43	\$550	264 months	\$145,200

Bob represents \$52,800 more in potential savings to the government than Alan, because of their different ages

Scenario 2: Disability Insurance (DI) beneficiaries with the same age and different monthly benefits<sup>b</sup>

Individual	Age	Monthly benefit	Duration	Potential savings (in nominal dollars)
"Carol"	42	\$1,087	300 months	\$326,100
"Diane"	42	\$961	300 months	\$288,300

Carol represents \$37,800 more in potential savings to the government than Diane, because of the difference in their benefit amounts

Source: GAO analysis of Social Security Administration's fiscal year 2013 statistical supplements on SSI and DI. | GAO-16-250

**Accessible Text for Figure 7: SSA's Annual Calculation of Cost Savings Rate from Continuing Disability Reviews**

Total cost savings rate from CDRs=

Supplemental Security Income (SSI) +

Disability Insurance (DI) +

Medicare (Projections from Centers for Medicare & Medicaid Services) +

Medicaid (Projections from Centers for Medicare & Medicaid Services)

[divided by]

---

**Dollars spent conducting Continuing Disability Reviews (CDRs)**

Source: GAO analysis of Social Security Administration (SSA) documents and interviews. | GAO-16-250

---

---

## GAO's Mission

The Government Accountability Office, the audit, evaluation, and investigative arm of Congress, exists to support Congress in meeting its constitutional responsibilities and to help improve the performance and accountability of the federal government for the American people. GAO examines the use of public funds; evaluates federal programs and policies; and provides analyses, recommendations, and other assistance to help Congress make informed oversight, policy, and funding decisions. GAO's commitment to good government is reflected in its core values of accountability, integrity, and reliability.

---

## Obtaining Copies of GAO Reports and Testimony

The fastest and easiest way to obtain copies of GAO documents at no cost is through GAO's website (<http://www.gao.gov>). Each weekday afternoon, GAO posts on its website newly released reports, testimony, and correspondence. To have GAO e-mail you a list of newly posted products, go to <http://www.gao.gov> and select "E-mail Updates."

---

## Order by Phone

The price of each GAO publication reflects GAO's actual cost of production and distribution and depends on the number of pages in the publication and whether the publication is printed in color or black and white. Pricing and ordering information is posted on GAO's website, <http://www.gao.gov/ordering.htm>.

Place orders by calling (202) 512-6000, toll free (866) 801-7077, or TDD (202) 512-2537.

Orders may be paid for using American Express, Discover Card, MasterCard, Visa, check, or money order. Call for additional information.

---

## Connect with GAO

Connect with GAO on [Facebook](#), [Flickr](#), [Twitter](#), and [YouTube](#).  
Subscribe to our [RSS Feeds](#) or [E-mail Updates](#).  
Listen to our [Podcasts](#) and read [The Watchblog](#).  
Visit GAO on the web at [www.gao.gov](http://www.gao.gov).

---

## To Report Fraud, Waste, and Abuse in Federal Programs

Contact:

Website: <http://www.gao.gov/fraudnet/fraudnet.htm>

E-mail: [fraudnet@gao.gov](mailto:fraudnet@gao.gov)

Automated answering system: (800) 424-5454 or (202) 512-7470

---

## Congressional Relations

Katherine Siggerud, Managing Director, [siggerudk@gao.gov](mailto:siggerudk@gao.gov), (202) 512-4400, U.S. Government Accountability Office, 441 G Street NW, Room 7125, Washington, DC 20548

---

## Public Affairs

Chuck Young, Managing Director, [youngc1@gao.gov](mailto:youngc1@gao.gov), (202) 512-4800  
U.S. Government Accountability Office, 441 G Street NW, Room 7149  
Washington, DC 20548