



Report to the Chairman, Committee on
Veterans' Affairs, House of
Representatives

November 2014

VETERANS' DISABILITY BENEFITS

Improvements Could Further Enhance Quality Assurance Efforts

GAO Highlights

Highlights of [GAO-15-50](#), a report to the Chairman, Committee on Veterans' Affairs, House of Representatives

Why GAO Did This Study

With a backlog of disability compensation claims, VBA faces difficulties in improving the accuracy and consistency of the claim decisions made by staff in its 57 regional offices. To help achieve its goal of 98 percent accuracy by fiscal year 2015, VBA recently implemented a new way of measuring accuracy and changed several quality assurance activities to assess the accuracy and consistency of decisions and to provide feedback and training to claims processors. GAO was asked to examine VBA's quality assurance activities.

This report evaluates (1) the extent to which VBA effectively measures and reports the accuracy of its disability compensation claim decisions and (2) whether VBA's other quality assurance activities are coordinated and effective. GAO analyzed VBA claims and STAR accuracy data from fiscal year 2013 (the most recent fiscal year for which complete data are available); reviewed relevant federal laws, VBA guidance, and other documents relevant to quality assurance activities; and interviewed VBA staff from headquarters and four VBA regional offices (selected to achieve variety in geography, workload, and accuracy rates), as well as veteran service organization officials.

What GAO Recommends

GAO is making eight recommendations to VA to improve its measurement and reporting of accuracy, review the multiple sources of policy guidance available to claims processors, enhance local data systems, and evaluate the effectiveness of quality assurance activities. VA concurred with all of GAO's recommendations.

View [GAO-15-50](#). For more information, contact Daniel Bertoni at (202) 512-7215 or bertonid@gao.gov.

November 2014

VETERANS' DISABILITY BENEFITS

Improvements Could Further Enhance Quality Assurance Efforts

What GAO Found

The Veterans Benefits Administration (VBA)—within the Department of Veterans Affairs—measures and reports the accuracy of its disability compensation claim decisions in two ways: (1) by claim and (2) by disabling condition, though its approach has limitations. When calculating accuracy rates for either measure through its Systematic Technical Accuracy Review (STAR), VBA does not always follow generally accepted statistical practices, resulting in imprecise performance information. For example, VBA does not adjust its accuracy estimates to reflect that it samples the same number of claims for review from each regional office—despite their varying workloads—and thus produces imprecise estimates of national and regional accuracy. Further, VBA reviews about 39 percent (over 5,000) more claims nationwide than is necessary to achieve its desired precision in reported accuracy rates, thereby diverting limited resources from other important quality assurance activities, such as targeted reviews of error-prone cases. In addition to issues with its statistical practices, VBA's process for selecting claims for STAR review creates an underrepresentation of claims that are moved between regional offices, which may inflate accuracy estimates because these claims have had historically lower accuracy rates. Finally, VBA has not clearly explained in public reports the differences in how its two accuracy measures are calculated or their associated limitations, as suggested by best practices for federal performance reporting.

VBA has taken steps to enhance and coordinate its other quality assurance activities, but GAO found shortcomings in how VBA is implementing and evaluating these activities. To improve local accuracy, VBA created regional office quality review teams (QRTs) with staff dedicated primarily to performing local accuracy reviews. QRTs assess individual claims processor performance and conduct special reviews to forestall certain types of errors. In addition, VBA began using questionnaires for assessing decision-making consistency, which are more efficient to administer than VBA's prior approach to conducting consistency studies. VBA also coordinates quality assurance efforts by disseminating national accuracy and consistency results, trends, and related guidance to regional offices for use in training claims processors. Further, VBA uses STAR results to inform other quality assurance activities, such as focusing certain QRT reviews on commonly made errors. However, GAO identified implementation shortcomings that may detract from the effectiveness of VBA's quality assurance activities. For example, contrary to accepted practices for ensuring the clarity and validity of questionnaires, VBA did not pre-test its consistency questionnaires to ensure the clarity of questions or validity of the expected results, although VBA officials indicated that they plan to do so for future questionnaires. In contrast with federal internal control standards that call for capturing and distributing information in a form that allows people to efficiently perform their duties, staff in the four regional offices that we visited had trouble finding the guidance they needed to do their work, which could affect the accuracy as well as the speed with which staff decide claims. Federal standards also call for knowing the value of efforts such as quality assurance activities and monitoring their performance over time; however, VBA has not evaluated the effect of its special QRT reviews or certain consistency studies on improving targeted accuracy rates, and lacks clear plans to do so.

Contents

Letter		1
	Background	3
	VBA's Approach to Measuring and Reporting Accuracy of Claim Decisions Has Limitations	6
	VBA Has Enhanced and Coordinated Its Quality Assurance Activities, Though Gaps in Implementation May Limit Their Effectiveness	15
	Conclusions	23
	Recommendations for Executive Action	24
	Agency Comments and Our Evaluation	25
Appendix I	Objectives, Scope and Methodology	27
Appendix II	Statistical Sampling Methodology	31
Appendix III	Comments from the Department of Veterans Affairs	35
Appendix IV	GAO Contact and Staff Acknowledgments	38
Tables		
	Table 1: Regional Offices Selection Criteria	30
	Table 2: STAR Monthly Sample Results for VBA's Boston Regional Office, Fiscal Year 2013	34
Figures		
	Figure 1: Effect of Weighting on Regional Office Claim-Based Accuracy Rankings, Fiscal Year 2013	7
	Figure 2: Ranking of Weighted Estimates of Claim-Based Accuracy Rates with 95 Percent Confidence Intervals, by Regional Office, Fiscal Year 2013	9
	Figure 3: Claim-Based and Issue-Based Accuracy Rates with 95 Percent Confidence Intervals by Number of Issues Claimed, Fiscal Year 2013	13

Abbreviations

IRR	inter-rater reliability
OIG	Office of Inspector General
QRT	quality review team
RVSR	Rating Veterans Service Representative
STAR	Systematic Technical Accuracy Review
VA	Department of Veterans Affairs
VBA	Veterans Benefits Administration
VSR	Veterans Service Representative

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.



November 19, 2014

The Honorable Jeff Miller
Chairman
Committee on Veterans' Affairs
House of Representatives

Dear Mr. Chairman:

The Department of Veterans Affairs' (VA) disability compensation program provides cash benefits to veterans for disabling conditions incurred or aggravated while in military service. In fiscal year 2013, VA paid \$53.6 billion in disability compensation to 3.6 million veterans. Within VA, the Veterans Benefits Administration (VBA), which is charged with processing disability compensation claims, faces a backlog of claims, due in part to the recent wars in Iraq and Afghanistan and the increasing number of servicemembers leaving the military. At the same time, VBA set a goal of achieving 98 percent accuracy in fiscal year 2015 for compensation claim decisions, which are made by staff in 57 VBA regional offices. Accurate claim decisions can help ensure that VBA is paying disability benefits only to those entitled to such benefits, in the correct amounts. Meanwhile, consistent decisions help ensure that veterans' claims receive comparable treatment, regardless of which VBA adjudicator or regional office processes the claim.

Questions have been raised about recent changes in the calculation of VBA's national accuracy rate, which is based on its national Systematic Technical Accuracy Review (STAR), and whether such changes reflect reliable measures of accuracy and VBA's commitment to serving veterans. GAO and VA's Office of Inspector General (OIG) have also previously reported on shortcomings in VBA's quality assurance activities.¹ This report examines (1) the extent to which VBA effectively measures and reports the accuracy of compensation claim decision-

¹ See, for example, *VA Disability Claims Processing: Preliminary Observations on Accuracy Rates and Quality Assurance Activities*, [GAO-14-731T](#) (Washington, D.C.: July 14, 2014), and VA Office of Inspector General, *Audit of Veterans Benefits Administration Compensation Rating Accuracy and Consistency Reviews*, (Washington D.C.: March 12, 2009).

making, and (2) whether VBA's other quality assurance activities are coordinated and effective.

To determine the extent to which VBA effectively measures the accuracy of compensation claim decisions, we reviewed STAR guidance, reports, and data and interviewed cognizant staff. We assessed VBA's sampling methodology and analyzed STAR and other VBA data on claims processed and reviewed from October 2012 through September 2013. We focused on the STAR process for reviewing disability compensation claims that were evaluated by VBA.² We did not review quality assurance efforts involving pension claims or appealed cases.³ To assess how VBA reports accuracy, we reviewed relevant VBA performance reports and compared VBA practices with legal requirements for agency performance reporting and related GAO work.⁴ To determine whether VBA's quality assurance activities are coordinated and effective, we reviewed VBA quality assurance policies, reports and guidance to identify key quality assurance activities, and then examined each activity's function and process by reviewing relevant guidance and policy documents and interviewing central office officials. We also interviewed VBA officials from four regional offices to gain their perspectives on how quality assurance activities are implemented at the regional office level, as well as how information is shared among quality assurance activities.⁵ We compared VBA's quality assurance activities against its internal guidance and

² Most of these are initial or reopened claims for benefits. A veteran may reopen a claim, for example, for increased benefits based on a new service-connected disability, or a worsening existing disability. We did not review quality assurance efforts for other types of disability compensation actions, referred to by VBA as authorizations. These include, for example, changes to benefit payments for additional dependents.

³ We did not include pension claims because VBA is reviewing its approach to the accuracy assessment of pension claims, which represent a small proportion of VBA's disability benefits workload. As of August 23, 2014, VBA had an inventory of about 10,000 pending pension claims among a total inventory of approximately 546,000 claims awaiting a rating.

⁴ See 31 U.S.C. § 1116 for legal requirements.

⁵ We visited the Newark, New Jersey and Oakland, California VBA regional offices and conducted telephone interviews with Nashville, Tennessee and Waco, Texas staff. We selected these offices to achieve variety in each of the following criteria: (1) number of claims processed annually; (2) geography (at least one regional office in each of VBA's four geographic divisions); (3) claims-based accuracy rates; and (4) issue-based accuracy rates. For each location, we interviewed managers, quality assurance staff, and veteran service organization representatives.

standards for internal control in the federal government.⁶ We also reviewed VBA's methods for designing and implementing its consistency studies against generally accepted practices in survey and questionnaire development. We assessed the reliability of VBA data used for all our analyses and determined that they were sufficiently reliable for the purposes of providing information on trends in claims decisions. For additional details on our objectives, scope, and methodology, see appendix I.

We conducted this performance audit from September 2013 to November 2014 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

VA pays monthly disability compensation to veterans with service-connected disabilities (i.e., injuries or diseases incurred or aggravated while on active military duty) according to the severity of the disability.⁷ VBA staff in 57 regional offices process disability compensation claims.⁸ These claims processors include Veterans Service Representatives (VSR) who gather evidence needed to determine entitlement, and Rating Veterans Service Representatives (RVSR) who decide entitlement and the rating percentage. Veterans may claim more than one medical condition, and a rating percentage is assigned for each claimed medical

⁶ GAO, *Standards for Internal Control in the Federal Government*, [GAO/AIMD-00-21.3.1](#) (Washington, D.C.: November 1999).

⁷ 38 U.S.C. § 1101 et seq. VA's ratings are awarded in 10-percent increments, from 0 to 100 percent. Generally, VA does not pay disability compensation for disabilities rated at 0 percent. As of December 2013, basic monthly payments ranged from \$130.94 for a veteran with 10-percent disability rating and no dependents, and to \$3,134 for a veteran with a 100-percent disability rating, a spouse and one child.

⁸ For quality assurance purposes, VBA counts one of its sub-offices as a separate regional office, in addition to its 56 regional offices. Thus, for reporting purposes, we refer to 57 offices.

condition, as well as for the claim overall.⁹ In fiscal year 2013, VBA decided more than 1 million compensation claims.

Since fiscal year 1999, VBA has used STAR to measure the decisional accuracy of disability compensation claims. Through the STAR process, VBA reviews a stratified random sample of completed claims, and certified reviewers use a checklist to assess specific aspects of each claim.¹⁰ Specifically, for each of the 57 regional offices, completed claims are randomly sampled each month and the data are used to produce estimates of the accuracy of all completed claims. VA reports national estimates of accuracy from its STAR reviews to Congress and the public through its annual performance and accountability report and annual budget submission. VBA also produces regional office accuracy estimates, which it uses to manage the program. Regional office and national accuracy rates are reported in a publicly available performance database, the Aspire dashboard.¹¹

Prior to October 2012, VBA's estimates of accuracy were claim-based; that is, claims free of errors that affect veterans' benefits were considered accurate and, conversely, claims with one or more errors that affect benefits were considered inaccurate.¹² Beginning in October 2012, VBA also began using STAR data to produce issue-based estimates of accuracy that measure the accuracy of decisions on the individual medical conditions within each claim. For example, a veteran could submit one claim seeking disability compensation for five disabling medical conditions. If VBA made an incorrect decision on one of those conditions, the claim would be counted as 80 percent accurate under the

⁹ In this report, we use the terms "medical condition" and "medical issue" interchangeably.

¹⁰ The STAR review has two major components. The benefit entitlement review assesses whether the correct steps were followed in addressing all issues in the claim, collecting appropriate evidence, and whether the resulting decision was correct, including effective dates and payment rates. Accuracy performance measures are calculated based on the results of the benefit entitlement review. The STAR review also assesses whether claims processors appropriately documented the decision and notified claimants.

¹¹ The Aspire dashboard is an online report of VBA's performance by program. Data are updated monthly and available by regional office and nationally. See http://www.benefits.va.gov/REPORTS/Aspire_dashboard.asp.

¹² VBA does not count errors that do not affect veterans' benefits, but it notes them during STAR reviews and works to correct them. Examples of such errors include missing signatures and lacking decision notification.

new issue-based measure. By comparison, under the existing claim-based measure, the claim would be counted as 0 percent accurate unless the error did not affect benefits when considered in the context of the whole claim. In March 2014, VBA reported a national estimate of issue-based accuracy in its fiscal year 2015 annual budget submission and plans to update this estimate in VA's next performance and accountability report. VBA also produces issue-based estimates by regional office, and reports them in the Aspire dashboard. For fiscal year 2013, the regional office claim-based accuracy rates ranged from an estimated 78.4 to 96.8 percent, and the issue-based accuracy rates ranged from an estimated 87.0 to 98.7 percent.

Beyond STAR, VBA has programs for conducting regional office quality reviews and for measuring the consistency of decisions. In March 2012, VBA established quality review teams (QRT) with one at each regional office. A QRT conducts individual quality reviews of claims processors' work for performance assessment purposes. The QRT also conducts in-process reviews before claims are finalized to help prevent inaccurate decisions by identifying specific types of common errors. Such reviews also serve as learning experiences for staff members. Since fiscal year 2008, VBA has also conducted studies to assess the consistency of disability claims decisions across regional offices. Initially, this initiative used inter-rater reliability (IRR) studies to assess the extent to which a cross-section of claims processors from all regional offices agree on an eligibility determination when reviewing the entire body of evidence from the same claim. In 2013, VBA revised its approach and began using questionnaires as its primary means for assessing consistency. A questionnaire includes a brief scenario on a specific medical condition for which claims processors must correctly answer several multiple-choice questions.

VBA's Approach to Measuring and Reporting Accuracy of Claim Decisions Has Limitations

VBA Does Not Follow Accepted Statistical Practices and Thus Generates Imprecise Accuracy Data

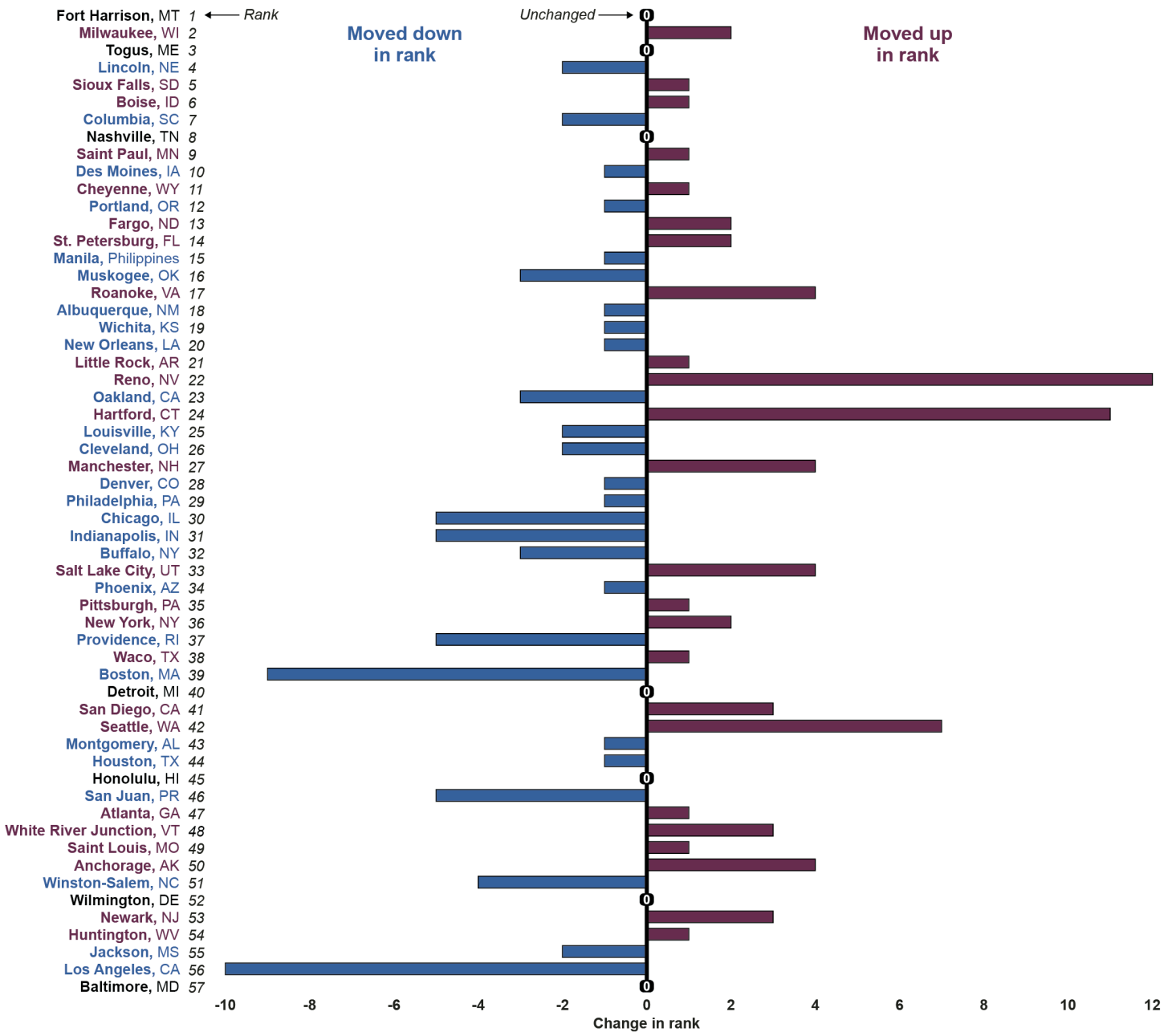
When calculating accuracy rates, VBA does not always follow generally accepted statistical practices. For example, VBA does not weight the results of its STAR reviews to reflect its approach to selecting claims by regional office, which can affect the accuracy of estimates.¹³ According to our analysis of VBA data, weighting would have resulted in a small change to VBA's nationwide claim-based accuracy rate for fiscal year 2013—from 89.5 to 89.1 percent.¹⁴ At the regional level, 29 of the 57 offices would have experienced a somewhat greater increase or decrease in their accuracy rates.¹⁵ Without taking weighting into consideration, regional office accuracy performance may be misleading and VBA management may focus corrective action or positive recognition on the wrong offices. For example, by taking weighting into account for the 57 regional offices in fiscal year 2013, the Reno regional office would have improved in relative accuracy by 12 places (from 34th to 22nd place), whereas the Los Angeles office would have declined in relative accuracy by 10 places (from 46th to 56th place) (see fig. 1).

¹³ VBA samples about the same number of claims from each regional office regardless of the offices' varying sizes. Thus, smaller regional offices are disproportionately represented, and the set of all claims reviewed nationally does not comprise a simple random sample of all claims processed by regional offices. Weighting adjusts for this fact and yields more correct estimates.

¹⁴ The estimated accuracy rate of 89.5 percent has a 95 percent confidence interval that ranges from 89 to 90 percent. The estimated accuracy rate of 89.1 has a 95 percent confidence interval that ranges from 88.4 to 89.8 percent.

¹⁵ In comparing the weighted accuracy estimates that we computed to the unweighted estimates that VBA reported for regional offices in fiscal year 2013, we found that weighting would increase the accuracy rate more than .4 percent for 17 offices and decrease the accuracy rate more than .4 percent for 12 offices. Weighting would increase the accuracy estimates for regional offices by as much as 2.1 percent and decrease the estimates by as much as 3.6 percent.

Figure 1: Effect of Weighting on Regional Office Claim-Based Accuracy Rankings, Fiscal Year 2013



Source: GAO analysis of Systematic Technical Accuracy Review (STAR) data of the Veterans Benefits Administration. | GAO-15-50

VBA also does not calculate the confidence intervals associated with the accuracy estimates that it generates, which prevents a complete understanding of trends over time and comparisons among offices.¹⁶ Accuracy estimates for different regional offices, or for the same office over time, are considered statistically different from each other when their confidence intervals do not overlap. As such, meaningful comparisons could be made on the basis of our analysis between, for example, Fort Harrison's estimated claim-based accuracy rate (ranked #1) and New York's estimated claim-based accuracy rate (ranked #36) because their confidence intervals did not overlap in fiscal year 2013 (see fig. 2). Conversely, comparisons between Fort Harrison's and Milwaukee's or Pittsburgh's estimated claim-based accuracy rates (ranked #2 and #35 respectively)—which had overlapping confidence intervals in fiscal year 2013—require a statistical test to determine if their differences are statistically meaningful.¹⁷ In effect, the claim-based accuracy rate of Fort Harrison and those of the regional offices with the next 34 highest reported accuracy rates may not be meaningfully different despite being ranked 1 through 35 of 57. Similarly, according to agency officials, VBA also does not calculate the confidence intervals associated with its newer issue-based accuracy estimates, which prevents meaningful comparisons between those estimates as well. Because VBA produces issue-based estimates using the same sample drawn to produce claim-based estimates, it would have to take extra steps to calculate the associated confidence intervals.¹⁸ As with the claim-based accuracy estimates, not computing the confidence intervals associated with issue-based

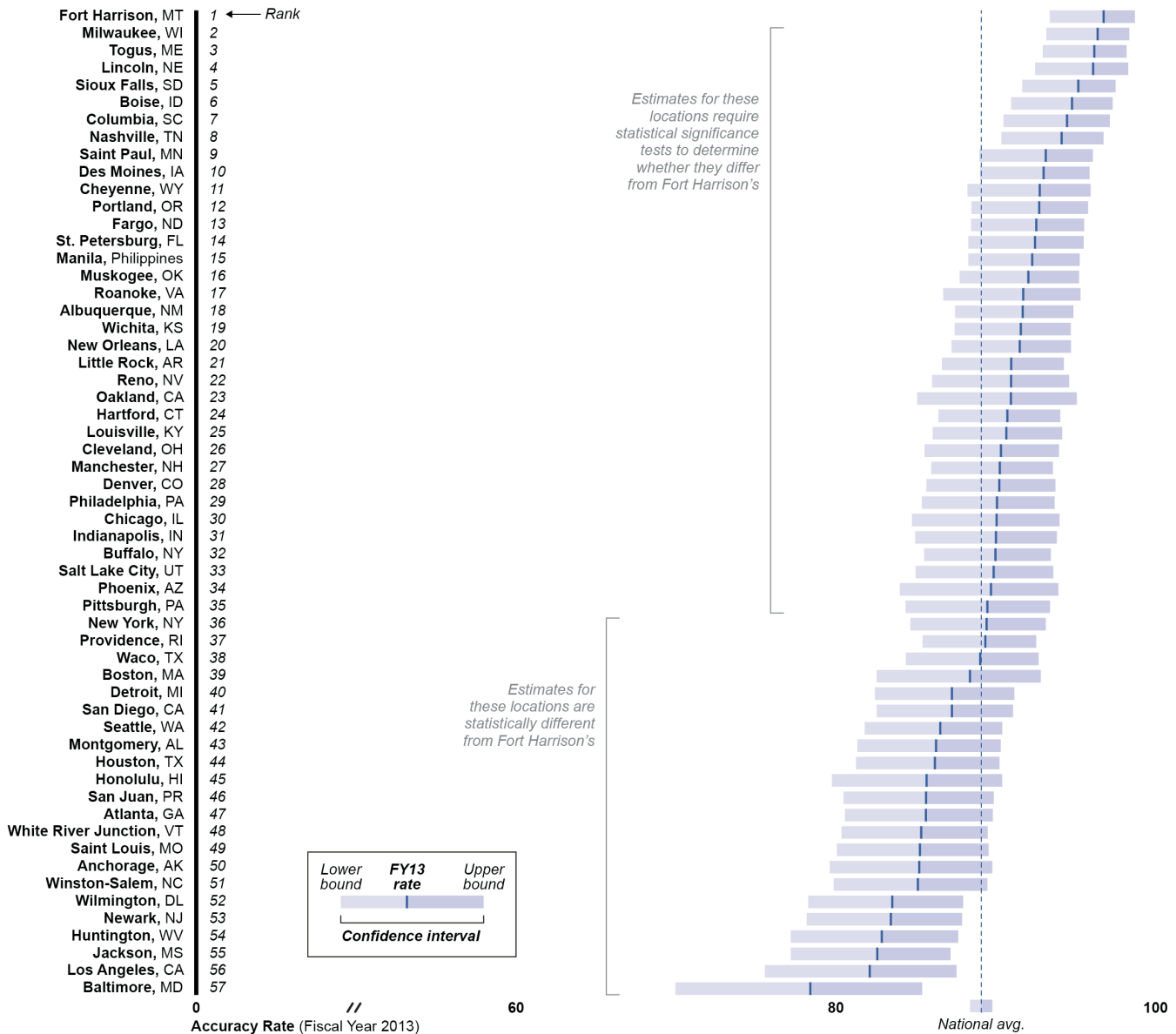
¹⁶ STAR accuracy estimates are derived from sample data and have sampling error associated with them. The confidence interval is a range of values around the estimate which is likely to include the actual population value, and helps determine whether different estimates are significantly different from a statistical perspective. The margin of error is the maximum of the difference between the lower bound of the confidence interval and the estimate, and the difference between the upper bound of the confidence interval and the estimate.

¹⁷ The required statistical test is called a t-test, which is a statistical hypothesis test that can be used to determine if two estimates are statistically different from each other. It is calculated by dividing the difference of the two estimates by the standard error of the difference.

¹⁸ Specifically, without changing its sampling approach for issue-based accuracy reviews, VBA would need to use a statistical technique called ratio estimation because the current sampling approach is based on claims, and not issues. For more information, see appendix II.

estimates limits VBA's ability to monitor its regional offices' relative performance and its overall performance over time.

Figure 2: Ranking of Weighted Estimates of Claim-Based Accuracy Rates with 95 Percent Confidence Intervals, by Regional Office, Fiscal Year 2013



Source: GAO analysis of Systematic Technical Accuracy Review (STAR) data of the Veterans Benefits Administration. | GAO-15-50

Note: STAR accuracy estimates are derived from sample data and have sampling error associated with them. The confidence interval is a range of values around the estimate, which is likely to include the actual population value.

VBA's approach to measuring accuracy is also inefficient because it reviews more claims than are statistically required to estimate accuracy. VBA randomly selects about 21 claims per month from each of its regional offices for STAR review, regardless of the offices' varying workloads and historical accuracy rates. According to VBA, this uniform approach allows the agency to achieve a desired level of precision of its accuracy estimates for each regional office.¹⁹ However, accepted statistical practices would allow for fewer cases to be reviewed at regional offices where the number of claims processed has been relatively small or accuracy has been high. According to our analysis of fiscal year 2013 regional office workload and accuracy results, VBA could reduce the overall number of claims it reviews annually by about 39 percent (over 5,000 claims) and still achieve its desired precision for its regional office accuracy estimates.²⁰ More efficient sampling could allow VBA to select fewer cases for review and free up limited resources for other important quality assurance activities, such as additional targeted accuracy reviews on specific types of error-prone or complex claims. Specifically, reviewing about 5,000 fewer claims could free up about 1,000 staff days because, according to VBA officials, STAR staff review at least 5 claims per day.

Calculating weighted estimates and confidence intervals, and adjusting sampling according to shifting workloads and accuracy rates, requires use of statistical methodology. According to VBA officials we interviewed, although STAR management used a statistician to help develop the way in which they measure accuracy, it currently does not use a statistician to, for example, weight STAR results and calculate confidence intervals for accuracy estimates. Further, VBA officials said they did not consult a statistician when developing the new issue-based accuracy measure, but rather relied on the same sampling methodology and approach for

¹⁹ VBA arrived at its sample size—246 rating claims per regional office per year—based on an assumed accuracy of rate of 80 percent for each regional office, and a desired precision that reflects sampling error of plus or minus 5 percentage points at the 95 percent level of confidence in accuracy estimates for each regional office. For more information, see appendix II.

²⁰ The precise reduction in total sample size from the current level is dependent on regional office workload and accuracy performance in the baseline year or years used for the calculations. We determined an overall reduction of over 5,000 claims or about 39 percent in the required sample size for STAR was possible using fiscal year 2013 regional office workload and accuracy data as the baseline for our calculations. Only for one regional office did we find that VBA would need to increase the number of claims currently reviewed to achieve its desired level of sample precision.

estimating accuracy as for the claim-based measure. We have previously reported that to be useful, performance information must meet users' needs for completeness, accuracy, consistency, and validity, among other factors.²¹ In response to our draft July 2014 testimony based on preliminary work, VBA officials stated they are exploring alternatives to their current methodology for estimating accuracy.²²

Beyond not following generally accepted statistical practices, VBA's STAR review systematically excludes certain claims, which may inflate accuracy rate estimates. Specifically, according to VBA officials, when a claim moves from one regional office to another, because a veteran has moved or workloads are redistributed, the database VBA uses to select claims for STAR review does not always reflect the office responsible for making the final determination for the claim.²³ As a result, STAR staff often select for review, then subsequently de-select, claims that have changed regional office jurisdiction.²⁴ Of the 14,286 rating claims randomly selected initially by VBA for review in fiscal year 2013, about 10 percent were de-selected because of a change in jurisdiction and replaced with other randomly selected claims. Those de-selected claims are not eligible for STAR review for the regional office that was ultimately responsible for the claim, thereby causing an underrepresentation of these claims in the STAR sample. Such underrepresentation may inflate VBA's reported accuracy rate because redistributed claims have historically had lower accuracy rates than non-redistributed claims.²⁵ In responding to our draft report, VBA indicated it is revising its procedures

²¹ GAO, *Managing For Results: GPRA Modernization Act Implementation Provides Important Opportunities to Address Government Challenges*, [GAO-11-617T](#) (Washington, D.C.: May 10, 2011).

²² [GAO-14-731T](#).

²³ VBA refers to redistributing workloads from backlogged regional offices to other locations as "brokering." In fiscal year 2013, VBA brokered about 10 percent of rating claims between regional offices.

²⁴ When cases are deselected from a regional office's sample, STAR staff commensurately increase the number of claims to be selected for review for that office in the following month according to VBA officials.

²⁵ According to VBA, in fiscal year 2012 redistributed rating claims had an average accuracy rate of 82.6 percent whereas non-redistributed rating claims had an average accuracy rate of 86.5 percent. Fiscal year 2012 was the last full year that redistributed claims were decided by separate processing centers, and that STAR staff reviewed redistributed claims separately from non-redistributed claims.

to ensure that claims selected for STAR review are included in the accuracy rate of the responsible regional office regardless of whether a change of jurisdiction occurred.

VBA Does Not Report Key Information to Help Users Understand Accuracy Metrics

Federal agencies should report clear performance information to the Congress and the public to ensure that the information is useful for decision making. In prior work, we identified clarity as a key attribute to a successful performance measure, meaning that the measure is clearly stated and its associated methodology is identified.²⁶ Measures that lack clarity may confuse or mislead users, and not provide a good picture of how well the agency is performing. We have also reported on best practices in implementing related federal performance reporting requirements, such as those in the GPRA Modernization Act of 2010.²⁷ Specifically, agencies must disclose information about the accuracy and validity of their performance information in their performance plans, including the sources for their data and actions to address any limitations.²⁸

VBA's accuracy reporting lacks methodological details that would help users understand the distinction between its two accuracy measures and their associated limitations. While VBA's new issue-based measure provides some additional perspective on the quality of claim decisions to date, VBA has not fully explained in its public reports how the issue-based and claim-based measures differ. For example, the issue-based measure tends to be higher than the claim-based measure because the former allows for claims to be considered partially correct, whereas the claim-based measure does not. According to VBA officials, the issue-based estimate provides a better measure of quality because veterans' claims have increasingly included multiple medical issues.²⁹ Our analysis of STAR data confirms that as the number of issues per claim increases,

²⁶ GAO, *Tax Administration: IRS Needs to Further Refine Its Tax Filing Season Performance Measures*, [GAO-03-143](#) (Washington, D.C.: Nov. 22, 2002).

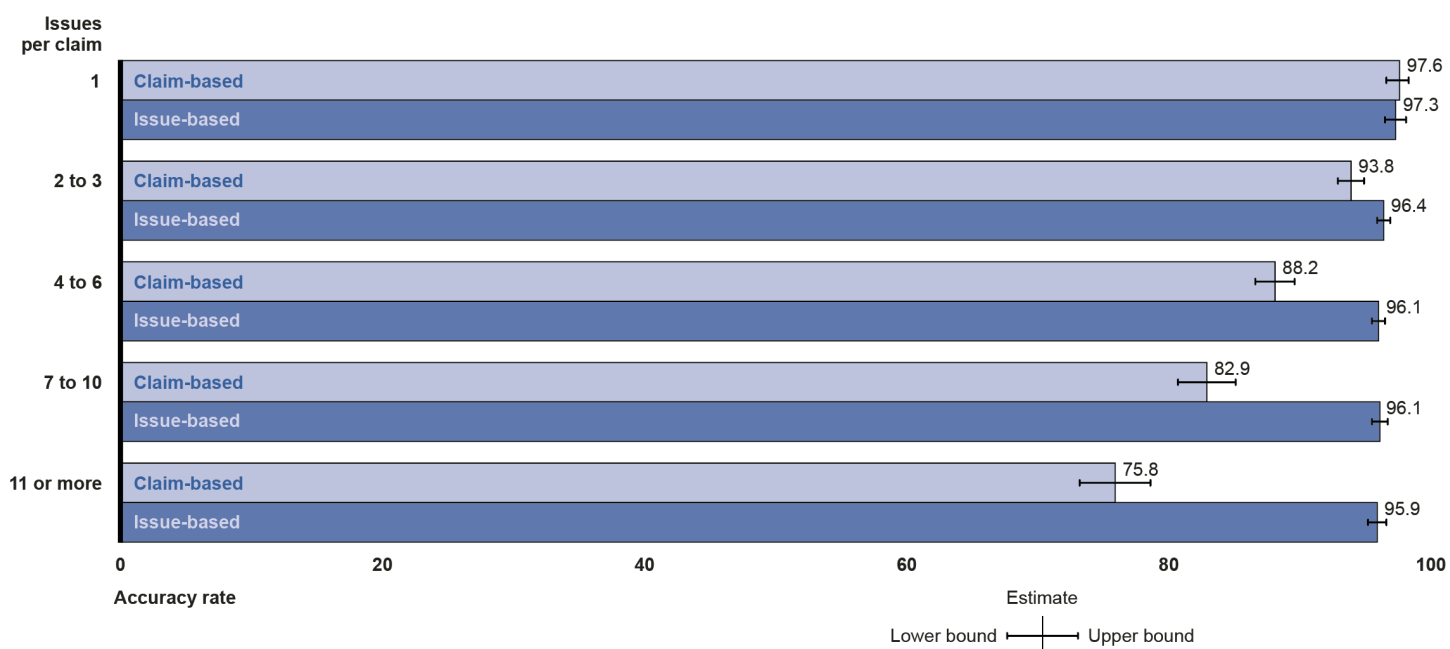
²⁷ GAO, *Managing For Results: GPRA Modernization Act Implementation Provides Important Opportunities to Address Government Challenges*, [GAO-11-617T](#) (Washington, D.C.: May 10, 2011).

²⁸ 31 U.S.C. § 1116.

²⁹ According to VBA, the average number of issues per rating claim increased from 2.8 in fiscal year 2009 to 4.9 in fiscal year 2013.

the chance of at least one issue being decided incorrectly within a single claim increases because there are more opportunities for error (see fig. 3). However, VA did not report in its fiscal year 2015 budget request how these measures are calculated and why the issue-based measure might be higher than the claim-based measure. VA has also not reported these distinctions in its Aspire dashboard.³⁰

Figure 3: Claim-Based and Issue-Based Accuracy Rates with 95 Percent Confidence Intervals by Number of Issues Claimed, Fiscal Year 2013



Source: GAO analysis of Systematic Technical Accuracy Review (STAR) data of the Veterans Benefits Administration. | GAO-15-50

VBA also counts claims processing errors differently under its claim-based measure than it does under its issue-based measure but does not report these distinctions, which raises questions about the transparency and consistency of VBA’s accuracy measures. For both measures, VBA differentiates between benefit entitlement errors that may financially affect the veteran and other errors, such as documentation and administrative

³⁰ VA’s most recent performance and accountability report does not contain issue-based accuracy data, but VA plans to include issue-based data in its next performance and accountability report.

errors that do not financially affect the veteran. For claim-based accuracy, VBA counts errors that financially affect the veteran now, but does not count errors that may financially affect the veteran in the future, although it works to correct both types of errors. For example, if one of several claimed medical conditions was rated incorrectly (e.g., 10 percent instead of 20 percent), but this error did not immediately affect the overall rating of the claim, VBA would not consider the claim in error because it did not affect the benefits that the veteran would receive.³¹ For the issue-based accuracy measure, however, VBA would count this as an error even if the error did not immediately affect the veteran's benefits. Unlike claim-based accuracy, issue-based accuracy may also include errors that would never affect future payments. For example, an incorrect effective date that is within the same month as the correct effective date does not affect benefits, but is counted as an error in VBA's issue-based accuracy measure. Conversely, according to VBA officials, this is not counted as an error in its claim-based measure. According to our analysis of STAR data, up to 6.9 percent of reviewed claims in fiscal year 2013 had these types of errors (i.e., benefit entitlement errors that do not immediately and may never affect benefits), and if they were all counted as errors, VBA's unweighted claim-based accuracy rate would have decreased by about 2 percent.³²

Further, VA has not explained in public reports that its accuracy measures are estimates that have distinct confidence intervals and limitations. Users should be aware of these confidence intervals to make meaningful comparisons, for example, between the two measures or over time for the same measure. In terms of each accuracy measure's limitations, the claim-based measure does not provide a sense of the proportion of issues that the agency decides correctly because the measure counts an entire claim as incorrect if any error is found. On the other hand, the issue-based measure does not provide a sense of the proportion of claims that the agency decides with no errors.

³¹ Such an error could affect the veteran's benefits if it is not corrected, the veteran were to claim new or worsened conditions in the future, and the subsequent re-calculation of the overall rating was affected by the error, according to agency officials.

³² The estimate of 6.9 percent has a 95 percent confidence interval that ranges from 6.5 to 7.3 percent. The estimated accuracy rate decrease is 1.7 percent and has a 95 percent confidence interval that ranges from 1.5 to 1.9 percent.

VBA Has Enhanced and Coordinated Its Quality Assurance Activities, Though Gaps in Implementation May Limit Their Effectiveness

VBA Has Taken Steps to Enhance and Coordinate Key Quality Assurance Activities

In addition to its STAR reviews, VBA's quality assurance framework includes other complementary activities, which have been enhanced to help meet its goal of 98 percent accuracy in fiscal year 2015. Specifically, VBA (1) established quality review teams (QRT) in March 2012 in regional offices as a means of strengthening its focus on quality where claims are processed, and (2) enhanced efforts to assess the consistency of decisions.

Although regional offices were previously responsible for assessing individual performance, QRTs represent a departure from the past because QRT personnel are dedicated primarily to performing these and other local quality reviews.³³ In addition, VBA requires QRT staff to pass a skills certification test annually—similar to VBA requirements for STAR staff and in contrast to requirements for claims processors who must pass a test every 2 years. In July 2013, VBA issued national guidance to ensure consistent QRT roles and practices across regional offices. For example, it included guidance on selecting individual quality review claim samples and conducting additional reviews for claims processors who do not meet their accuracy goals.³⁴ In addition to conducting individual

³³ Prior to QRTs, accuracy of individual claims processors was assessed against targets for each employee. However, these reviews were generally performed by the claims processors' supervisors. VBA expected that having QRT members perform the individual reviews would allow supervisors to focus more on performance management.

³⁴ QRT reviewers review an average of 5 randomly-selected claims per claims processing staff member per month. For claims processing staff members found in need of accuracy improvement, 10 reviews per claims processing staff member per month may be performed.

quality reviews, QRT personnel are charged with conducting in-process reviews of claims that are not yet finalized, looking for specific types of common errors. Quality reviewers are also responsible for providing feedback to claims processors on the results of their quality reviews, typically as reviews are completed, including formal feedback from the results of individual quality reviews and more informal feedback from the results of in-process reviews. In addition, at the four offices we contacted, quality reviewers are available to answer questions and provide guidance to claims processors as needed.

VBA's efforts to assess consistency of claims decisions have also expanded in recent years. Up until 2013, VBA largely relied on inter-rater reliability (IRR) studies to assess consistency, which to date have been time consuming and resource intensive. Claims processors typically required about 4 hours to review an entire claim. The process was administered by proctors in the regional offices and the results were hand-graded by national VBA staff. Given the resources involved, IRR studies have been typically limited to 300-500 (about 25-30 percent) claims processors, randomly selected from the regional offices. In 2009, VBA expanded its consistency program to include questionnaires, which it now relies on more heavily to assess consistency. The more streamlined consistency questionnaires require less staff time to complete because, in addition to a brief scenario on a specific condition, participants have 10 or fewer multiple-choice questions to answer. The questionnaires are administered electronically through the VA Talent Management System, removing the need to proctor or hand-grade the tests, which has allowed VBA to significantly increase employee participation. A recent consistency questionnaire was taken by about 3,000 claims processing employees—representing all employees responsible for rating claims. Further, VBA now administers these studies more frequently, from about 3 to 24 per year. According to VBA officials, they plan to further expand the use of consistency studies from two questionnaires per month to six to eight per month, pending approval of additional quality assurance staff.³⁵

VBA also has taken steps to coordinate its quality assurance efforts in several ways, such as systematically disseminating information on national accuracy and consistency results and trends to regional office

³⁵ VBA would like to increase the number of consistency questionnaires to target additional claims processing positions such as the Claims Assistant position, according to officials.

management and QRTs, which in turn share this information with claims processing staff. With respect to STAR, in addition to receiving monthly updates on overall accuracy performance, regional offices receive quarterly reports with analyses of accuracy performance including information by error type. QRT reviewers also participate in monthly conference calls with STAR staff during which they discuss error trend information. While claims processing staff learn about errors they made on claims directly from STAR, managers or QRT members at each of the regional offices we contacted noted that they also share STAR trend data with claims processors during periodic training focused on STAR error trends. With respect to consistency studies, regional offices receive national results; regional office-specific results; and, since February 2014, individual staff results. Officials at each of the four regional offices we visited told us QRT staff share the results of consistency studies with staff and inform claims processors of the correct answers to the questions.

Coordination also occurs when QRT personnel disseminate guidance and support regional office training based on error trends identified through STAR and other quality assurance activities. Two of the four offices we contacted cited instances where they have used consistency study results for training purposes. At one office, the results from a consistency study were used to provide training on when to request an exam for certain conditions, such as tinnitus. In general, at each of the four offices, officials told us that QRT reviewers conduct, or work with regional office training coordinators to conduct, periodic training forums for claims processors. Regional offices we contacted also supplement training with other communications informed by quality review results. For example, QRTs at three of the four regional offices we contacted produce periodic newsletters for regional office claims processors, which include guidance based on errors found in all types of reviews. Specifically, at one office, a newsletter was used to disseminate guidance on ensuring that a rating decision addresses all issues in a claim. The need for this guidance was identified on the basis of STAR and local quality review results.

Lastly, VBA coordinates its quality assurance activities by using STAR results to guide other quality assurance efforts. According to VBA officials, the agency has used STAR data to identify error trends associated with specific medical issues, which in turn were used to target efforts to assess consistency of decision-making related to those issues. Recent examples are (1) the August 2013 IRR study, which examined rating percentages and effective dates assigned for diabetes mellitus (including peripheral neuropathy); and (2) a February 2014 study on obtaining correct disability evaluations on certain musculoskeletal and

respiratory conditions. In addition, according to VBA, the focus of in-process reviews performed by QRTs has been guided by STAR error trend data. VBA established in-process reviews in March 2012 to help the QRTs identify and prevent claim development errors related to medical examinations and opinions, which it described as the most common error type. More recently, VBA has added two more common error types—incorrect rating percentages and incorrect effective benefit dates—to its in-process review efforts. VBA officials stated that they may add other common error types based on future STAR error analyses.

Some Gaps in Implementation Persist and the Effectiveness of Quality Assurance Activities Is Unclear

While QRTs reflect VBA's increased focus on quality, during our site visits we identified shortcomings in QRT practices and implementation that could reduce their effectiveness. Specifically, we identified the following shortcomings: (1) the exclusion of claims processed during overtime to assess individual performance; (2) the inability to correct errors identified before a claim is finalized in certain situations; and (3) a lack of pre-testing of consistency questionnaires.

Regarding the first shortcoming, we learned that three of the four offices we contacted had agreements with their local unions that prevented QRT personnel from reviewing claims processed during overtime to assess individual performance.³⁶ As a result, those regional offices were limited in their ability to address issues with the quality of work performed during overtime. Centrally, VBA officials did not know which or how many regional offices excluded claims processed during overtime, or the extent to which excluding cases worked during overtime occurred nationally. According to VBA data, claims processed on overtime represented about 10 percent of rating-related claims completed nationally in fiscal year 2013. After we reported this finding,³⁷ VBA issued guidance in August 2014 to regional offices stipulating inclusion of claims processed on overtime, and that the regional offices work with their local unions to rescind any agreements that exclude such claims from review.

Second, officials at four regional offices we contacted told us that they face a challenge in conducting individual quality and in-process reviews

³⁶ To help reduce its claims backlog, VBA has required claims processors to work 20 hours per month of mandatory overtime during portions of fiscal years 2013 and 2014.

³⁷ [GAO-14-731T](#).

as expected³⁸ because VBA's Veterans Benefits Management System lacks the capability to briefly pause the process and prevent claims from being completed while a review is still underway.³⁹ VBA officials acknowledged that this was a problem for regional offices in completing reviews, based on anecdotal information from regional offices, but did not have information on the extent to which this occurred. VBA officials noted that reviews could be performed after a claim is completed; however, if an error is found, the regional office might need to rework the claim and provide the veteran with a revised decision. The officials also noted that VBA is working toward modifying its Veterans Benefits Management System to address this issue, but is at the initial planning stage of gathering requirements and could not provide a time frame for completion.

Thirdly, although VBA has developed a more streamlined approach to measuring consistency, VBA officials told us that consistency questionnaires were developed and implemented without any pre-testing, which would have helped the agency determine whether the test questions were appropriate for field staff and were accurately measuring consistency. Pre-testing is a generally accepted practice in sound questionnaire development for examining the clarity of questions or the validity of the questionnaire results. In the course of our review, VBA quality assurance officials noted that they plan to begin pre-testing consistency questionnaires as a part of a new development process. Specifically, after each questionnaire has been developed, two to three quality assurance staff who have claims processing experience, but were not involved in the questionnaire's development, would be targeted to pre-test it. Quality assurance staff responsible for the consistency studies would then adjust the questionnaire if necessary before it is administered widely. While initially slated to occur in July 2014, VBA quality assurance staff now anticipate pre-testing to begin in September 2014.

Beyond these implementation shortcomings, staff in each of the four offices we contacted said that several key supports were not sufficiently updated to help quality review staff and claims processors do their jobs

³⁸ A regional office is expected to perform in-process reviews equivalent to 10 percent of their expected claims decisions per month, according to VBA guidance.

³⁹ VBA's Veterans Benefits Management System is intended to help streamline the claims process by allowing for paperless claims processing, including electronic claims files.

efficiently and effectively. Staff at these offices consistently described persistent problems with central guidance, training, and data systems.

- **Guidance:** Federal internal control standards highlight the need for pertinent information being captured and distributed in a form that allows people to perform their duties efficiently.⁴⁰ However, regional office quality review staff said they face challenges locating the most current guidance among all of the information they are provided. Managers or staff at each of the regional offices we contacted said that VBA's policy manuals are outdated. As a result, staff must search numerous sources of guidance to locate current policy, which is time-consuming and difficult. This, in turn, could affect the accuracy with which they decide claims. One office established a spreadsheet to consolidate guidance because the sources were not readily available to claims processors. VBA officials acknowledged that there are several ways it provides guidance to regional offices. In addition to the existence of relevant regulations and VBA's policy and procedures manual, VBA provides guidance to claims processors through policy and procedures letters, monthly quality calls and notes from these calls, various bulletins, and training letters and other materials maintained on VBA's intranet site. While agreeing that having multiple sources of guidance could be confusing to staff, VBA officials noted they face challenges in updating the policy manual and other available guidance materials to ensure that they are as current as possible. After we reported on this issue,⁴¹ VBA officials noted that they are considering streamlining the types of guidance provided. They also plan to develop a system of consolidated links to guidance documents by alphabetized topic to help claims processors access the information more efficiently; however, VBA officials acknowledge that developing a single repository will be a challenging project and have not yet dedicated adequate resources for this effort.
- **Training:** Staff in the offices we contacted also said that in some cases national training has not been updated to reflect the most current guidance, which in turn makes it difficult to provide claims processors with the information they need to avoid future errors. For example, staff from one regional office noted that training modules on

⁴⁰ GAO, *Standards for Internal Control in the Federal Government*, [GAO/AIMD-00-21.3.1](#) (Washington, D.C.: November 1999).

⁴¹ [GAO-14-731T](#).

an error-prone issue—Individual Unemployability and related effective dates of benefits—had not been updated to reflect all new guidance, the sources of which included conference calls, guidance letters, and frequently asked questions compiled by VBA’s central office.⁴² Further, officials at regional offices we contacted expressed concern that VBA limits their flexibility to update out-of-date course materials. In response to these concerns, VBA training officials explained that that they are continually updating national training to reflect new guidance, but how long it takes is a function of the extent of the policy change. These officials noted that updating the Individual Unemployability training was particularly delayed because of numerous, unanticipated changes in policy and related guidance that resulted in their setting aside previously updated course materials and starting over. VBA training officials also explained that while VBA does not allow changes to the contents of courses in its catalog, regional offices can propose courses for the catalog, based on their needs identified through quality reviews.

- **Data systems:** Regional office quality review staff also told us that they are required to log errors into three systems or databases that do not “speak to one another” and two lack the capability to fully track errors trends, thereby limiting their ability to take corrective actions. At the regional office level, quality assurance information is entered into three different databases or systems.⁴³ Staff at each of the four offices we contacted said that the Automated Standardized Performance Elements Nationwide system used for tracking individual accuracy for performance management purposes lacks functionality to create reports on error trends by claimed medical issue or reasons for specific types of errors. As a result, three offices maintain separate spreadsheets to identify error trends related to individual accuracy. Regional office staff also noted that one of the two systems used to track in-process reviews does not help track error trends, for example, by employee, resulting in two offices maintaining additional spreadsheets to track this information. At the national level, VBA

⁴² Individual Unemployability is a part of VA’s disability compensation program that allows VA to pay benefits at the 100 percent level to certain veterans whose service-connected disabilities prevent them from maintaining substantial gainful employment.

⁴³ Specifically, individual performance reviews are entered into the Automated Standardized Performance Elements Nationwide system, whereas in-process reviews are entered into either a WebLogon database or a SharePoint database depending on the type of error being reviewed.

central office has made some improvements in reporting and now has the ability to analyze regional office information on errors by medical issue. According to VBA officials, they share this information with regional office managers and quality staff during training calls. VBA officials stated that a planned replacement for its Automated Standardized Performance Elements Nationwide system would have addressed reporting limitations at the local level, but was halted. As of September 2014, VBA did not have a timeframe for restarting the process for acquiring a new system.

Finally, VBA's efforts to evaluate the effectiveness of its quality assurance activities have been limited. Specifically, VBA officials told us that although they have not seen an increase in the national accuracy rate in fiscal year 2014, the number of errors related to claim development has declined, demonstrating the success of QRT reviews and training in targeting these errors.⁴⁴ Also, VBA identified 13 regional offices whose issue-based accuracy rates improved between the first and third quarters of fiscal year 2014, attributing these improvements to actions taken by quality assurance staff in fiscal year 2014.⁴⁵ However, it was not clear from the documentation VBA provided whether and how it monitored the effectiveness of these actions for all regional offices. With respect to consistency studies, VBA also has not evaluated—and lacks plans to evaluate—the efficacy of using consistency questionnaires relative to the more resource-intensive IRR studies. According to a VBA official, the consistency questionnaires have helped identify regional offices and individuals in need of further training on the basis of the percentage of incorrect answers, as well as the need for national training. However, officials could not provide data or evaluations indicating that consistency questionnaires have improved accuracy rates in the areas studied.

VBA officials noted that they are considering a new data system that would combine all local and national quality assurance data—including STAR, in-process reviews, and individual quality reviews—and allow for

⁴⁴ The STAR review assesses whether adequate evidence was developed to support the rating decision. Possible development errors include failure to obtain sufficient medical records, including a medical examination or opinion.

⁴⁵ VBA cited several actions including fielding in-process reviews for additional error types, performing consistency studies to identify claims processors needing training, holding quality calls with regional office staff, and releasing clarifying guidance to the regional offices.

more robust analyses of root causes of errors. Specifically, they expect the system will show relationships across the results of various quality assurance reviews to determine employee competence with various aspects of claims processing. According to VBA officials, this system would also enable them to more easily evaluate the effectiveness of specific quality assurance efforts. Evaluation can help to determine the “value added” of the expenditure of federal resources or to learn how to improve performance—or both. It can also play a key role in strategic planning and in program management, informing both program design and execution.⁴⁶ Continuous monitoring also helps to ensure that progress is sustained over time.⁴⁷ However, VBA officials indicated that this proposal is still in the conceptual phase and requires final approval for funding and resources.

Conclusions

VBA’s dual approach for measuring accuracy is designed to provide additional information to better target quality improvement efforts, but its methods and practices lack rigor and transparency, thereby undermining the usefulness and credibility of its measures. By not leveraging a statistician or otherwise following statistical practices in developing accuracy estimates, VBA is producing and relying on inaccurate estimates to make important internal management decisions. Similarly, by using a one-size sampling methodology, VBA is unnecessarily expending limited resources that could be used elsewhere. The systematic exclusion of redistributed claims and those moved between offices further calls into question the rigor of its accuracy estimates. Lastly, VBA’s reporting of its two accuracy metrics lacks sufficient transparency to help members of Congress and other stakeholders fully understand the differences and limitations of each, and thus may undermine their trust in VBA’s reported performance.

VBA has enhanced and coordinated other aspects of its quality assurance framework, but shortcomings in implementation and evaluation detract from their overall effectiveness. For example, although VBA is disseminating the results of national STAR reviews and consistency studies, and local QRTs are using those results to focus related training or guidance to claims processing staff, until centralized guidance is

⁴⁶ GAO, *Designing Evaluations: 2012 Revision*, [GAO-12-208G](#) (Washington, D.C.: January 2012).

⁴⁷ [GAO/AIMD-00-21.3.1](#).

consolidated and streamlined, staff lack ready access to information that will help them prevent errors. Moreover, absent adequate system capabilities to support local quality reviews, QRTs are unable to stop incorrect decisions from being finalized, and may not be aware of error trends that could be mitigated through training or other corrective action. Finally, although some of its quality assurance activities are relatively new, VBA lacks specific plans to evaluate their effectiveness and may miss opportunities to further improve or target these activities to more error-prone areas. In general, unless VBA takes steps to improve the rigor of all its quality assurance methods and practices, VBA may find progress toward achieving its goal of 98 percent accuracy in fiscal year 2015 illusive—especially in the face of challenging workloads, limited resources, and expectations of timely claim decisions.

Recommendations for Executive Action

To help improve the quality of VBA's disability compensation claim decisions, we recommend that the Secretary of Veterans Affairs direct the Under Secretary for Benefits to:

- Leverage appropriate expertise to help VBA do each of the following:
 - weight its accuracy estimates to reflect the sample design for reviewed claims;
 - determine and report the confidence intervals associated with its reported accuracy estimates; and
 - re-examine its approach to calculating the regional office sample size for STAR.
- Take steps to ensure that redistributed claims and those moved between regional offices are not underrepresented in the STAR sample.
- Increase transparency in explaining how the claim-based and issue-based accuracy rates are calculated as well as their key limitations when publicly reporting these metrics.
- Review the multiple sources of policy guidance VBA provides to determine ways to consolidate them or otherwise improve their availability and accessibility for use by staff in regional offices.

-
- Take steps to ensure that any future upgrades to local data systems allow QRTs to pause the claims process when errors are detected and enable QRTs to better track error trends.
 - Take additional steps to evaluate the effectiveness of quality assurance activities to identify opportunities to improve or better target these activities.

Agency Comments and Our Evaluation

We provided a draft of this report to VA for review and comment, and its written comments are reproduced as appendix III in this report. VA generally agreed with our conclusions and concurred with all of our recommendations. The agency outlined how it plans to address our recommendations as follows:

- Regarding our recommendations to leverage appropriate expertise to improve its measurement and reporting of accuracy, VA stated that a VBA statistician has begun developing a revised sampling methodology that takes into consideration output and claims processing accuracy at each regional office to determine sample sizes. VBA also plans to appropriately weight accuracy estimates and calculate the margins of error based on the revised sampling methodology. VBA intends to report results based on this new methodology beginning in March 2015.
- Regarding our recommendation to take steps to ensure that redistributed claims and those moved between regional offices are not underrepresented in the STAR sample, VA stated that VBA's revised sampling methodology will be based on the office completing the claim, and that no claims will be excluded from samples due to changes in jurisdiction. VBA intends to implement this revised sampling methodology by the end of March 2015.
- Regarding our recommendation to increase transparency in explaining how the claim-based and issue-based accuracy rates are calculated, VA stated that VBA will describe its sampling, assessment criteria, calculation, and reporting methodologies for claim and issue-level accuracy as part of future performance documents and public reports. VBA anticipates implementing this recommendation by the end of March 2015.
- Regarding our recommendation to review the multiple sources of policy guidance VBA provides to regional office staff, VA stated that in September 2014, VBA began improving the availability and

accessibility of policy guidance, as well as consolidating references to this guidance. VBA anticipates completing this project by the end of April 2015.

- Regarding our recommendation to take steps to ensure that any future upgrades to local data systems allow QRTs to pause the claims process when errors are detected and enable QRTs to better track error trends, VA stated that VBA is designing a new database that will incorporate all types of quality reviews (i.e., regional office reviews, STAR, and consistency studies) and provide VBA with more data analysis capabilities. Although VA did not outline specific steps VBA plans to take to upgrade local data systems so that QRTs may pause the claims process, VBA plans to implement this recommendation by the end of June 2015.
- Regarding our recommendation to take additional steps to evaluate the effectiveness of quality assurance activities to identify opportunities to improve or better target these activities, VA stated that VBA's new database will enable VBA to do so by the end of June 2015.

VA also provided technical comments, which we incorporated as appropriate.

We are sending copies of this report to the appropriate congressional committees and the Secretary of Veterans Affairs. In addition, the report is 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. 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 IV.

Sincerely yours,



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

Appendix I: Objectives, Scope and Methodology

The objectives of this report were to examine (1) the extent to which the Veterans Benefits Administration (VBA) effectively measures and reports the accuracy of compensation claim decision-making, and (2) whether VBA's other quality assurance activities are coordinated and effective.

Review of Systematic Technical Accuracy Review (STAR)

To assess VBA's measurement and reporting of the accuracy of compensation claim decision-making, we focused on the STAR process for reviewing disability compensation claims that VBA identifies as rating-related—that is, requiring a decision on the claimant's eligibility for benefits and the monthly benefit amount. We did not review quality assurance over disability compensation claims that did not involve a rating, including adjustments for additional dependents. We also did not review quality assurance efforts involving appealed cases, aspects of which fall under the Board of Veterans' Appeals. Finally, we did not review pension claims, which represent a small portion of VBA's disability benefits workload, because VBA is reviewing its approach to the accuracy assessment of pension claims.¹

To determine the extent to which STAR appropriately reflects the accuracy of claims, we reviewed VBA policy manuals, the STAR checklist, and other tools used in VBA's STAR review. We interviewed VBA and Office of Inspector General (OIG) officials to learn whether there are claim types that are omitted from STAR review and, if so, the reasons for these omissions. To determine how errors are identified and counted under STAR, we examined the ways in which the checklist and other STAR procedures are used to quantify errors. We visited VBA's office in Nashville, Tennessee, where the STAR reviews are conducted to observe the review process and program methodology in action. We reviewed checklists used to assess accuracy of claims and identified information VBA uses on the basis of these checklists to calculate accuracy rates.

To assess the extent to which VBA uses generally accepted statistical practices to generate accuracy rates, we analyzed VBA data on claims processed and reviewed from October 2012 through September 2013. In analyzing STAR data, we calculated the weighted claim-based annual accuracy rate for each regional office and nationwide. We then calculated the 95 percent confidence intervals associated with these estimated

¹ As of August 25, 2014, VBA had an inventory of about 10,000 pending pension claims among a total inventory of approximately 545,000 claims awaiting a rating.

accuracy rates. We applied a statistical sample size formula suitable for use in a stratified random sample and analyzed the differences this approach produced compared to VBA's sample size estimation methodology for regional offices.

We assessed the reliability of VBA's STAR data by performing electronic data testing, reviewing related documentation, and interviewing knowledgeable agency officials. We also assessed the reliability of VBA's claim processing data by interviewing knowledgeable agency officials about the data. To electronically assess the reliability of the STAR data, we tested for duplicate benefit records, tested the claim disposition date field to ensure we only analyzed STAR claims from fiscal year 2013, checked the benefit claim end product code to ensure we only included benefit claims with end product codes eligible for inclusion in the STAR accuracy sample, checked for missing data in key analysis variables, and examined the range of values in key variables to check for outliers. We determined that the data were sufficiently reliable for our purposes.

To assess how VBA reports accuracy, we identified and reviewed relevant VBA performance reports, such as VA's Performance and Accountability Report and Aspire dashboard data. We also interviewed VBA officials about the rationale for creating the issue-based accuracy measure, and the agency's plans for reporting its performance on accuracy and consistency. We compared VBA practices with legal requirements for agency performance reporting such as the GPRA Modernization Act of 2010 and related GAO work (e.g., GAO, *Managing For Results: GPRA Modernization Act Implementation Provides Important Opportunities to Address Government Challenges*, [GAO-11-617T](#), Washington, D.C.: May 10, 2011).

Coordination and Effectiveness of Quality Assurance Activities

To determine whether VBA's quality assurance activities are coordinated and effective, we reviewed VBA quality assurance policies, reports, and guidance to identify key quality assurance activities. Based on this review, we focused on quality review teams (QRT), which are located in each regional office and responsible for local quality assurance, as well as on VBA's consistency program that is administered by VBA's centralized quality assurance staff. We then examined each activity's function and process by reviewing relevant guidance and policy documents and interviewing central office officials. Specifically:

- We reviewed VBA policy and procedure documents for quality review teams (QRT) to learn the purposes of, and the information generated by, these efforts. In addition, we interviewed VBA central office and

regional office officials to gather their perspectives on any redundancy or gaps between quality assurance efforts. We compared the functions of and information yielded by quality assurance components with the framework laid out in VBA's Quality Assurance Program Plan, as well as standards for internal control in the federal government (see GAO, *Standards for Internal Control in The Federal Government*, [GAO/AIMD-00-21.3.1](#), Washington, D.C. : November 1999). In addition, we interviewed VBA regional office officials to learn about processes QRTs follow and how these procedures may vary across regional offices. We also reviewed and compared VBA criteria for QRT staff, STAR reviewer, and claims processor certification.

- We reviewed documents and interviewed VBA officials to learn more about the recent changes to the agency's approach to assessing consistency. More specifically, we explored the rationale for the change from using inter-rater reliability (IRR) studies to using consistency questionnaires. We assessed the development and implementation of the recent consistency questionnaires by, for example, examining VBA's consideration of pre-testing the instruments using generally accepted survey procedures, and how pre-testing may affect the resulting measures of consistency. Finally, to further determine how consistency questionnaires are complementary with other quality assurance efforts, we reviewed VBA's process for determining topics for consistency questionnaires. Specifically, we asked about the methods used to select and prioritize topics, including the extent to which officials use findings from QRTs and STAR.

To further determine what and how information is shared among quality assurance components and how this coordination helps to identify problem areas, we interviewed VBA regional office officials to gather their perspectives on how information is shared from STAR, QRT, consistency studies, and regional office compliance visits and how that information-sharing could be improved. We interviewed officials at the regional level to gain their perspectives on coordination and effectiveness of all of VBA's quality assurance activities. At each office, we spoke with service center managers and quality assurance staff, as well as representatives of local veteran service organizations. The regional offices were selected to reflect a range of characteristics related to: (1) geography (at least one regional office in each of VA's four areas), (2) number of claims processed annually, (3) claim-based accuracy rates, and (4) issue-based accuracy rates. We did not identify specific quality assurance pilots or initiatives being tested in regional offices. We selected 4 of VBA's 57

regional offices for review. We visited the Oakland and Newark regional offices and conducted telephone interviews with Nashville and Waco regional office staff. Table 1 provides information about the regional offices we selected to visit.

Table 1: Regional Offices Selection Criteria

Office	VBA Area	Compensation caseload	STAR rating accuracy (claims)	STAR rating accuracy (issues)
Oakland, CA	Western	21st	20th	34th
Newark, NJ	Eastern	41st	54th	55th
Nashville, TN	Southern	6th	7th	5th
Waco, TX	Central	2nd	38th	36th

Source: GAO analysis of VA data.

Appendix II: Statistical Sampling Methodology

This appendix provides additional technical details on ratio estimation for producing issue-based accuracy rates, as well as the audit work we did to re-estimate the regional office Systematic Technical Accuracy Review (STAR) sample sizes using a formula for stratified random probability samples.

Ratio Estimation

Because STAR is designed to sample claims and produce an estimate of the claim-based accuracy rate and because the number of medical issues per claim varies, ratio estimation should be used to develop issue-based accuracy rates. Furthermore, during their review of sampled claims, STAR reviewers may find that one or more inferred issues were missed or, conversely, that the review process included one or more issues inappropriately. Thus, the STAR sample of claims must be used to estimate both the total number of issues as well as the number of issues that were processed correctly. With respect to STAR, ratio estimation takes the form shown below.

$$\hat{R} = \frac{\hat{t}_a}{\hat{t}_m} = \frac{\sum_{i=1}^{57} \sum_{j=1}^{12} \sum_{k=1}^{n_{i,j}} W_{i,j} a_{i,j,k}}{\sum_{i=1}^{57} \sum_{j=1}^{12} \sum_{k=1}^{n_{i,j}} W_{i,j} m_{i,j,k}}$$

In the formula, the subscript i represents the regional office, the subscript j represents the month of the fiscal year, $n_{i,j}$ represents the monthly sample size for regional office i in month j , $W_{i,j}$ represents the stratum sampling weight for regional office i in month j , $a_{i,j,k}$ represents the number of issues adjudicated correctly on claim k in month j and regional office i and $m_{i,j,k}$ represents the total number of issues on claim k in month j and regional office i .

The ability to calculate a ratio estimate and its associated confidence interval are available in most statistical software applications.

Sample Size Re-Estimation

Each month the Veterans Benefits Administration (VBA) selects a random sample of benefit claims within each VA regional office to review under the STAR program. The measure of interest is the estimated percent of claims that were processed correctly by VBA regional office staff. The sample size formula used by VBA to derive the number of claims to select in each VBA regional office is shown below.

$$n = \left(\frac{Z^2}{E^2} \right) \times (PQ)$$

In the formula, Z = the quantile from the Normal distribution for the desired level of confidence. The desired margin of sampling error is denoted by E . The assumed percent of accuracy in the population is denoted by P , and Q is defined as $Q = (1 - P)$. For their calculations, VBA uses the following values:

$$Z = 1.96 \quad E = 0.05 \quad P = 0.8 \quad \text{and} \quad Q = 0.20.$$

When these values are plugged into the equation, $n = 246$. This is VBA's target annual sample size for each VA regional office. With 57 regional offices, this translates into 14,022 claims selected nationally per fiscal year in the STAR sample. On a monthly basis, when divided by 12, $246/12 = 20.5$ which rounds up to 21. Thus, VBA's monthly sample size for each regional office is 21 claims. By definition, the sample frame for each month is the set of veteran benefit claims completed by the regional office within the previous month.

The standard statistical formula for the sample size calculation with a stratified random sample is shown below. We applied this formula to determine an annual total sample size for a regional office in the coming fiscal year using observed monthly accuracy rates and monthly number of claims completed from the previous fiscal year.

$$n_0 = \left(\frac{Z^2}{E^2} \right) \sum_{j=1}^{12} W_j P_j Q_j$$

In turn, this initial sample size is adjusted with the finite population correction factor. The formula for the adjusted sample size is shown below.

$$n = \frac{n_0}{1 + \frac{n_0}{N}}$$

In these formulas, the terms Z and E are defined as before in the sample size formula currently used by VBA. The term P_j is the observed historical monthly accuracy rate for the regional office in the prior fiscal year. The term Q_j is defined as $1 - P_j$. The term W_j represents the monthly fraction of the annual total number of claims processed by the regional office in the prior fiscal year. The term N is the total number of claims processed by the regional office in the prior fiscal year. Because STAR is intended for monitoring benefit claim processing, we re-set the value of P_j to a value of 0.90 for any month where $P_j = 1$ in order to ensure a minimum monthly sample allocation. In order to demonstrate how this formula works in practice, data for the Boston regional office are shown in table 2 as an example.

Table 2: STAR Monthly Sample Results for VBA’s Boston Regional Office, Fiscal Year 2013

Month	Claims Processed	Stratum	Monthly Accuracy Rate
October	581	1	0.905
November	453	2	0.941
December	537	3	0.905
January	671	4	1.000
February	562	5	0.950
March	971	6	0.905
April	930	7	0.850
May	1,207	8	0.842
June	1,408	9	0.900
July	1,668	10	0.955
August	2,023	11	0.818
September	1,575	12	0.818
Total	12,586		0.899

Source: GAO analysis of Systematic Technical Accuracy Review (STAR) data of the Veterans Benefits Administration (VBA).

Here are the calculations for the Boston regional office using the data in table 1.

$$n_0 = \left(\frac{1.96^2}{0.05^2} \right) \times \left(\left(\frac{581}{12586} \right) (0.905 * 0.095) + \left(\frac{453}{12586} \right) (0.941 * 0.057) + \dots + \left(\frac{1575}{12586} \right) (0.818 * 0.182) \right)$$

Calculation gives $n_0 = 159.40$ and $n=158$ after applying the correction for sampling from a finite population.

The effect of applying a stratified random sample formula, which uses historical observed monthly accuracy rates from the prior fiscal year as well as accounting for the population size from the prior fiscal year, is a reduction in the needed annual STAR sample size from 246 to 158 claims for the Boston regional office.

Appendix III: Comments from the Department of Veterans Affairs



DEPARTMENT OF VETERANS AFFAIRS
WASHINGTON DC 20420

October 28, 2014

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

Dear Mr. Bertoni:

The Department of Veterans Affairs (VA) has reviewed the Government Accountability Office's (GAO) draft report, "**VETERANS' DISABILITY BENEFITS: Improvements Could Further Enhance Quality Assurance Efforts**" (GAO-15-50). VA generally agrees with GAO's conclusions and concurs with GAO's recommendations to the Department.

The enclosure specifically addresses GAO's recommendations and provides an action plan for each, and provides technical comments to the draft report. VA appreciates the opportunity to comment on your draft report.

Sincerely,

A handwritten signature in black ink, appearing to read "Jose D. Riojas".

Jose D. Riojas
Chief of Staff

Enclosure

Enclosure

Department of Veterans Affairs (VA) Response to
Government Accountability Office (GAO) Draft Report
***“VETERANS’ DISABILITY BENEFITS: Improvements Could Further Enhance
Quality Assurance Efforts”***
(GAO-15-50)

GAO Recommendation: To help improve the quality of VBA’s disability compensation claim decisions, GAO recommends that the Secretary of Veterans Affairs direct the Under Secretary for Benefits to:

Recommendation 1: Leverage appropriate expertise to help VBA do each of the following:

- weight its accuracy estimates to reflect the sample design for reviewed claims;
- determine and report the confidence intervals associated with its reported accuracy estimates; and
- re-examine its approach to calculating the regional office sample size for STAR.

VA Comment: Concur. A statistician from the Veterans Benefits Administration (VBA) Office of Performance Analysis and Integrity (PA&I) is working to revise the sampling and reporting of VBA’s compensation and pension claims processing accuracy program. In August 2014, VBA initiated a review of the accuracy sampling and reporting process. As a result, VBA is developing a sample methodology to consider output and claims processing accuracy at each station to determine sample sizes. Additionally, all work completed will be sampled according to the regional office that completed the claim, eliminating cases that have historically been excluded from quality review due to brokering. VBA will calculate margins of error and appropriately weigh accuracy estimates based on the revised sampling methodology.

In addition, VBA has completed a thorough review of the current Systematic Technical Accuracy Review (STAR) process and will make necessary programming changes and test the new sampling methodology in December 2014. The new quality samples will be created beginning in January 2015 for cases completed in December 2014. Reporting for these cases will begin in March 2015 and will include the confidence intervals for each regional office. Target Completion Date: March 31, 2015.

Recommendation 2: Take steps to ensure that redistributed claims and those moved between regional offices are not underrepresented in the STAR sample.

VA Comment: Concur. VBA’s revised sampling methodology will be based on the office completing the claim. No claims will be excluded from the samples due to changes in jurisdiction. VBA will capture both the office of original jurisdiction as well as the office completing the claim to ensure that this work, known as “brokered work,” is not underrepresented and meets the same high-quality expectations of our compensation and pension programs. Target Completion Date: March 31, 2015.

Enclosure

Department of Veterans Affairs (VA) Response to
Government Accountability Office (GAO) Draft Report
***“VETERANS’ DISABILITY BENEFITS: Improvements Could Further Enhance
Quality Assurance Efforts”***
(GAO-15-50)

Recommendation 3: Increase transparency in explaining how the claim-based and issue-based accuracy rates are calculated as well as their key limitations when publicly reporting these metrics.

VA Comment: Concur. VBA will create an abstract describing its sampling, assessment criteria, accuracy calculation, and reporting methodologies for claim and issue-level accuracy. This abstract will accompany future performance documents and public reports to explain key differences between the claim-based and issue-based accuracy rates. Target Completion Date: March 31, 2015.

Recommendation 4: Review the multiple sources of policy guidance VBA provides to determine ways to consolidate them or otherwise improve their availability and accessibility for use by the field.

VA Comment: Concur. In September 2014, VBA began the process of improving the availability and accessibility of policy guidance, as well as consolidating all policy guidance references. VBA anticipates completing the project by April 2015. Target Completion Date: April 30, 2015.

Recommendation 5: Take steps to ensure that any future upgrades to local data systems include improvements to allow for the pausing of processing claim decisions identified by QRTs as incorrect, and to enable QRTs to better track error trends.

VA Comment: Concur. VBA is currently designing a new database that will incorporate all types of quality reviews, to include local regional office reviews, STAR, and consistency studies, and capture data at various stages of the claims process. The database will provide VBA with increased data analysis capabilities for accuracy review and improved tracking of error trends. Target Completion Date: June 30, 2015.

Recommendation 6: Take additional steps to evaluate the effectiveness of quality assurance activities to identify opportunities to improve or better target these activities.

VA Comment: Concur. As stated in the response to Recommendation 5, VBA is currently designing a new database that will incorporate all types of quality reviews and capture data at various stages of the claims process. The database will provide VBA with increased data analysis capabilities, to include improved tracking of error trend analysis. This will allow VBA to evaluate the effectiveness of the quality assurance activities and identify opportunities for improvement. Target Completion Date: June 30, 2015.

Appendix IV: GAO Contact and Staff Acknowledgments

GAO Contact

Daniel Bertoni, bertonid@gao.gov, or (202) 512-7215

Staff Acknowledgments

In addition to the contact named above, Michele Grgich (Assistant Director), Dana Hopings (Analyst-In-Charge), Carl Barden, James Bennett, David Chrisinger, Alexander Galuten, Joel Green, Avani Locke, Vernetta Shaw, Almeta Spencer, Walter Vance, and Greg Whitney made key contributions to this report.

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](#). Visit GAO on the web at 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

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

Congressional Relations

Katherine Siggerud, Managing Director, 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, (202) 512-4800 U.S. Government Accountability Office, 441 G Street NW, Room 7149 Washington, DC 20548

