**Benefits Claims Decision Support System (BCDSS)**

**00031AB Comparative Performance Analysis of Ear Models**

**February 2017**



**Version 1.0**

Contents

[1 Executive Summary 1](#_Toc474412597)

[2 Overview 1](#_Toc474412598)

[2.1 Introduction 1](#_Toc474412599)

[2.2 Supporting Claims Data 2](#_Toc474412600)

[2.3 BCDSS Model Objectives and Basis for Ear Model Disaggregation 2](#_Toc474412601)

[3 Comparative Performance Analysis 4](#_Toc474412602)

[3.1 Overview and Summary Findings 4](#_Toc474412603)

[3.2 Performance Comparison between BCDSS Ear Model (1.1) to GHL Model (1.1) 5](#_Toc474412604)

[3.3 Observations 6](#_Toc474412605)

[3.4 Impact of Rating Calculators on Model Accuracy 7](#_Toc474412606)

[3.5 Recommendations Based on Comparative Analysis 9](#_Toc474412607)

[3.6 General Hearing Loss Accuracy and Throughput Deficiencies 9](#_Toc474412608)

[3.6.1 Inaccuracy 9](#_Toc474412609)

[3.6.2 Throughput 10](#_Toc474412610)

[3.7 Tinnitus Inaccuracy 11](#_Toc474412611)

# Executive Summary

The Benefits Claims Decision Support System (BCDSS) team was tasked by the then acting Under-Secretary for Benefits, Tom Murphy, to decompose the aggregate statistical model developed to statistically adjudicate supplemental claims for increased compensation for Ear disabilities. The previous aggregate approach produced a Combined Disability Determination (CDD) for a suite of ear-related disabilities, as opposed to a single disability determination for a specific ear-related disability. The BCDSS team’s analysis of model performance established that the very high percentage of Tinnitus Claims for Increase either alone or in combination with General Hearing Loss (GHL) Claims skewed the results of the Ear Model. This resulted from the fact that Tinnitus original claims are rated overwhelmingly at 10 percent and almost never granted increases in subsequent claims whereas the GHL outcomes are more randomly distributed. The team proceeded to decompose the aggregate Ear Model into separate models for GHL and Tinnitus.

A comparative analysis of the models that resulted from the disaggregation of the Ear Model indicates that:

* The Tinnitus and GHL model increased accuracy and throughput across a common set of claims (in the case of Tinnitus, increasing throughput by over 50 percent).
* The diagnosis-code specific models were applied to 291,885 claims and to produce 289,010 calculated decisions – a 32 percent increase in the volume of calculated decisions.
* Performance comparisons are complicated by the fact that CDDs produced by the Ear Model contained rating decisions for diagnosis codes other than Tinnitus or General Hearing Loss in at least 36,319 cases, accurately predicting CDDs that could NOT be related directly to the Tinnitus or GHL decisions.

This document describes the performance delivered by the GHL and Tinnitus models in greater detail – individually and collectively, and compares the performance to that delivered by the Ear Model. The document also provides observations and recommendations on a path forward for future models.

# Overview

## Introduction

This ***Comparative Performance Analysis of Ear Models*** provides an analysis comparing the performance of diagnostic code-specific models (GHL Model No. 1.1 and Tinnitus Model No. 1.2) with the performance delivered by the original BCDSS Ear Model No. 1.0.

The GHL and Tinnitus models represent two models within a portfolio of predictive models for BCDSS. BCDSS is a proof-of-concept project jointly sponsored by the Department of Veterans Affairs (VA) Center for Innovation (VACI) and the Veterans Benefits Administration (VBA) to demonstrate the feasibility of using automation and predictive models to calculate the disability rating for certain types of claims for disability compensation benefits. The BCDSS project includes both the development of candidate models for rating claims and a software platform which provides the functionality and environment for the models to run against Veteran and Claims data in order to produce a model recommended rating.

## Supporting Claims Data

Data used to support model development, testing, and comparative analysis were derived from the same dataset used by MITRE Corporation[[1]](#footnote-1) and the BCDSS team to develop the earlier ear- and knee-related models. This data encompassed claims received by VBA between 2005 through 2014.[[2]](#footnote-2) The distribution of claims that were deemed “eligible” for use in modeling are provided in Table 1.

Table : Distribution of Claims Deemed Eligible for Modeling



Claims are deemed eligible based on whether the claim is a supplemental claim (End Product 02X), the claimant is making a relevant contention (defined as Contention Classification Codes 2200, 3140, 3150, 4700, 5710, or 6850), and VBA previously promulgated a service connected prior decision for the specific diagnosis code (Codes 6100 for GHL, or 6260 for Tinnitus).

## BCDSS Model Objectives and Basis for Ear Model Disaggregation

The objective for all BCDSS models is to accurately determine the rating for the maximum number of specified disability conditions[[3]](#footnote-3) (e.g., hearing loss) contained within supplemental claims for increased disability compensation. There are two distinct metrics for this objective:

* **Accuracy** refers to the model calculating the same rating as that determined by the Rating Veteran Service Representative (RVSR) through VBA’s traditional claims adjudication process.
* **Through-put** refers to the supplemental claims containing at least one specified condition for which the rating can be calculated by the model (stated as a percentage).

The analysis completed by MITRE[[4]](#footnote-4) demonstrated that accuracy rates of greater than 90 percent were possible for aggregate ear-related conditions in just over 50 percent of eligible supplemental claims. Initial BCDSS hearing model (BCDSS Model No. 1.0) development efforts for aggregate ear-related conditions achieved accuracy rates of 94.38 percent for 60.81 percent of eligible supplemental claims. Per the Sponsor’s direction, the initial Hearing Loss Model No. 1.0 was developed using the hearing-related codes referenced in footnote 2.

A performance analysis of the initial BCDSS Ear Model 1.0 results[[5]](#footnote-5) (Figure 1) revealed three key findings:

Figure : Distribution of Modeling Results by Diagnostic Code

1. 93 percent of all claims eligible to be rated using BCDSS Hearing Model No. 1.0 were rated using the General Hearing Loss (6100) or Tinnitus (6260) diagnosis codes, i.e., 23 percent of the claims contained hearing loss code ONLY, 14 percent contained Tinnitus code ONLY, and 57 percent contained claims for both GHL and Tinnitus codes.
2. Tinnitus, once substantiated and rated, must receive a rating of 10 percent. As a result, the rating for supplemental claims where the claimant has previously received a 10 percent rating for Tinnitus will remain unchanged at 10 percent.
3. Because of the dominance and predictability of Tinnitus in the population of hearing eligible supplemental requests for increased claims, VACI requested that the team proceed to decompose the model into its constituent diagnostic codes, then redevelop and re-evaluate the resulting models to determine if separate models provide improved results for accuracy and throughput. The specific objective for the GHL Model is to demonstrate that ***a condition-specific GHL*** ***Model can yield an accuracy rate equal to or greater than 95 percent across more than 60 percent of eligible claims.***

# Comparative Performance Analysis

## Overview and Summary Findings

VACI requested that the BCDSS Team evaluate the performance that might be achieved from separate General Hearing Loss and Tinnitus models. The original BCDSS Ear Model was capable of determining the Combined Disability Determination (CDD) for combined disabilities codes in 60.81 percent of eligible supplemental claims with an accuracy rate of 94.38 percent. A comparative performance analysis of separate, diagnosis code-specific models uses these metrics as benchmarks against which to assess relative performance.

The combined number of eligible claims for the tested dataset (for both Tinnitus and GHL) was [**291,885**](#Table1). This data set included **147,640** claims that contained both diagnostic codes, **48,140** claims containing GHL only diagnosis codes, and **96,105** claims containing only Tinnitus diagnosis codes. The GHL and Tinnitus models collectively addressed a total of **213,636** claims (or **73.19** percent) at a combined accuracy rate of **94.24** percent. A detailed explanation for why the remaining 78,249 were unaddressed is provided in [Section 3.6.2](#_Throughput).

As noted above, 147,640 claims contained both Tinnitus and GHL rating decisions. Because the same claim contained multiple decisions, the models individually produced separate decisions for Tinnitus and GHL. The total number of Tinnitus and GHL decisions associated with the 291,885 claims was **367,259**. The two models accurately predicted **275,578** of these rating decisions; an accuracy rate of **95.7** percent and a throughput of **78.69** percent. The specific distribution and performance level by Claim Category and Decision Type, is presented in Table 2.

Table : Diagnosis Codes within Claims



## Performance Comparison between BCDSS Ear Model (1.1) to GHL Model (1.1)

The original BCDSS Ear Model was developed using claims from the same dataset as those used to develop subsequent BCDSS models. However, the dataset used to develop the GHL and Tinnitus models was composed of all available claims, as opposed to the subset originally used to develop the Ear Model (the factors that determined claim eligibility were identified during the process of formulating the Ear Model and consequently some eligible claims were not included in the data set used to the model).

The BCDSS Ear Model produced Combined Disability Determinations (CDDs), a calculation that encompassed one or more rating decisions for the following diagnostic codes: 6100, 6200, 6201, 6202, 6204, 6205, 6207, 6209, 6210, 6211, and 6260. The GHL and Tinnitus models, by definition, calculated single the ratings for diagnostic codes 6100 (GHL) and 6260 (Tinnitus). As a result, each model calculated the same volume of results as the Ear Model (or, when combined, almost twice the number of results) at comparable or better accuracy and throughput rates.

Moreover, in at least 36,319 cases, eligible claims contained at least one decision for an Ear related diagnosis that was not General Hearing Loss or Tinnitus. As a result, comparing the results from the two efforts produces a range of permutations. For example, four distinct permutations for accurate results are possible, including:

* Cases where both the Ear Model and GHL Model accurately calculated a change in the CDD rating for the claim decision.
* Cases where the Ear Model accurately calculated a change in CDD and where GHL accurately calculated that the rating was confirmed and continued (in such cases, one or more ratings for other Ear related diagnosis codes changed while the rating for GHL may have remained unchanged).
* Cases where the Ear Model accurately calculated that the CDD remained unchanged while the GHL model accurately determined that change in the rating for GHL was warranted (in such cases, the rating for another diagnosis code changed and offset the change in the rating for GHL).
* Cases where the Ear and GHL models accurately determined that both the CDD for the ear, and individual rating for GHL, should be confirmed and continued.

The same permutations exist for inaccurate claim results, as well as results for unaddressed claims. These 12 possible permutations that exist for each diagnostic code incorporated within an aggregated model highlight the complications that arise from models that are designed to output a single CDD for a collection of diagnostic codes. A complete comparison between the GHL Model and the Ear Model is presented in Table 3.

Table : GHL Model and Ear Model Comparison



## Observations

Observations relating to the comparison of performance between the two models include:

1. Although the accuracy of the Ear Model across the common set of claims 98.41 percent, throughput is only 46.28 percent (below the 50 percent threshold). Comparable performance metrics for the GHL Model are 90.23 percent accuracy with a throughput of 59.71 percent.
2. The Ear Model accurately identified the CDD for 11,534 claims where a rating increase was promulgated by the RVSR. In all of these cases, the GHL Model accurately identified that the rating decision for GHL was Confirmed and Continued (C&C). The Ear Model also accurately identified the CDD for another 9,538 claims where the rating for GHL was C&C that the GHL model did not address. This permutation accounts for the differences in performance between the two models in accurately predicting the rating when a change is warranted, and the complexity in comparing the performance of two models where one produces an aggregate rating.
3. The examples cited above can be explained by the inclusion of other ear related service-connected rating decisions that were warranted a rating increase, and consequently increased the CDD calculated by the Ear Model (or perhaps multiple decisions, some warranting an increase, and others warranting a decrease, but ultimately resulting in a change in the CDD from the previous decision profile).
4. This permutation is magnified in comparisons between the Ear Model and the Tinnitus Model. In this case (encompassing a claim set common to both of 127,778 claims), the Tinnitus Model yields an accuracy rate of 99.87 percent across 100 percent of claims, compared to the Ear Model which produced an accuracy rate of 98.69 percent but only addressed 45.26 percent of the claims. This data is presented in Table 4.

Table : Ear Model and Tinnitus Model Comparison



## Impact of Rating Calculators on Model Accuracy

VACI and VBA requested that the BCDSS Team evaluate the potential impact of the introduction of automated “calculators” on the consistency of ratings. These calculators are now used widely and are an integral part of the Veterans Benefits Management System (VBMS). They deliver proposed ratings by diagnostic code based on evidence-supported input provided by RVSRs and subsequently support calculation of the claimant’s overall disability rating. The calculators were integrated into the rating determination process as a means to improve the quality and consistency of rating determinations. The Office of Compensation Services noted that rating calculators were made mandatory on November 1, 2010 for GHL and more broadly in July 2011 with the advent of the evaluation builder.

Given the nature of Tinnitus claims, the focus of the analysis was on GHL. Data compiled for GHL decisions promulgated between 2006 and 2014 identified a marked increase in claims for an increase in disability compensation (from 7,438 in 2006 to 56,641 in 2011). However, the percentage of those C&C remained relatively constant at approximately 71 percent. 2006 and 2013 were exceptions, marked by a low of 68.64 percent and a high of 73.81 percent, respectively. This data is presented in Figure 2.



Figure : Rating Decision Type for General Hearing Loss Decisions

BCDSS model performance statistics closely parallel these results, consistently identifying claim decisions that are C&C at rates in excess of 99 percent. Consequently, the implementation of calculators does not appear to have made an appreciable contribution to an increase in the consistency of RVSR decisions where the decision was C&C. However, the BCDSS Team noted a significant increase in the accuracy of the model for claims where an increase was granted for decisions that occurred after 2010. This is illustrated in Figure 3 below. The model inherently reflects the consistency of RVSR decisions, and the BCDSS Team notes that the most commonly repeated and consistently accurate fact patterns (Pattern IDs within the model) for claims where an increase was granted all occur beginning in 2010. ***This suggests that the calculators did have a significant impact on decision consistency.***



Figure : Model Accuracy for Changes in Rating in General Hearing Loss Decisions

## Recommendations Based on Comparative Analysis

The BCDSS Team recommends that condition-specific models (GHL, Tinnitus, etc.) be incorporated into the BCDSS system for Ear Conditions—either in parallel with the existing Ear Model or in lieu of the original Ear Model. This recommendation is based on:

1. **Improved Accuracy and Through-put:**  Diagnostic code specific models deliver a 13 percent increase in through-put at comparable accuracy levels.
2. **Improved Clarity and Traceability of the Modeling Result:** Qualitative benefits are also accrued from diagnosis code-specific models. BCDSS users (whether RVSRs, Quality Assurance personnel, or modeling agents) are able to evaluate the specific historical background of the issue within the claim, the nature and circumstances of relating to both the specific Veteran contention and the rating determination, and considerations that may have led to a difference between the rating returned by the model and that assigned by the RVSR. Models that generate a CDD for a group of issues do not allow for such an evaluation as there is no mechanism to evaluate the modeling result (unless, of course, only one diagnosis code formed the basis of the CDD).

## General Hearing Loss Accuracy and Throughput Deficiencies

### Inaccuracy

The General Hearing Loss model inaccurately identified 11,636 claims (6 percent of the total eligible claim volume, and 10 percent of the volume addressed by the model). 11,619 (or 99.8 percent) of these claims were claims where a change in rating was determined by the RVSR to be warranted. Notable characteristics relating to these claims include:

* The claimant in 10,847 (93 percent) of these cases was older than 60 at the time the claim was filed.
* The claimant had received only one previous decision related to GHL in 9,743 (or 84 percent) of these claims, and the average age of the decision (the period of time that had elapsed since the previous decision was instituted) was 3.49 years.
* The claimant, in 86 percent of these claims, had either received only one other ear-related decision for Tinnitus (6,818 cases) or no other ear-related decisions (2, 873 cases). In fact, the claimant also had not received a decision for any of the other most frequently co-occurring conditions.
* Over half (6,196) were for a 10 percent increase from an original 0 percent rating. Most of the remaining changes resulted in either a 20 or 30 percent rating

The data suggests that the absence of a general disability benefit filing history (and in particular, a hearing-related filing history), despite being retirement age, contributed significantly to the inaccuracies.

### Throughput



Figure : Distribution of Unaddressed Claims

The model was unable to assign a rating in 78,249 eligible claims within the target data set. These claims were generally equally divided between those that were C&C and those that were found to merit a change in rating (see Figure 4). Notable characteristics relating to these claims include:

* The claimant was age 70 or older in 73,160 (or 93 percent) of these claims.
* The claimant had only received ear-related decisions for GHL and Tinnitus in 38,228 (49 percent) of these claims or only for GHL in 21,321 (27 percent) of these claims.
* The average age of the prior GHL decision was 3.83 years.
* The claimant had not received a prior decision in any of the most frequently co-occurring condition at the time of filing in 66,992 (86 percent) of these claims.

As was the case for those claims where the model returned an inaccurate rating, the data suggests that a viable, repeatable pattern could not be obtained because of the absence of a general disability benefit filing history (and in particular, a hearing-related filing history).

## Tinnitus Inaccuracy

Only 883 of 243,745 eligible claims containing a Tinnitus decision were rated as something other than 10 percent for that diagnosis; virtually all of these were rated 0. As indicated in the Tinnitus model manual (BCDSS Model No. 1.0), “[t]he Office of Compensation Service suggested that the 0 percent ratings may well have been assigned in error or resulted from a 0 percent evaluation, i.e., the exam revealed that the claimant did not suffer from Tinnitus.”

1. See MITRE Study completed under Contract No. VA118A13J0421/VBA OSP COMPENSATION SERVICES CLIN 0005 IFCAP 101-J47030 and related “Statistical Adjudication Engineering Notebooks.” [↑](#footnote-ref-1)
2. Provided by VBA’s Office of Performance Analysis & Integrity (PA&I). Table names are ah4929\_rating\_decision, ah4929\_rating\_corp\_claim, and ah4929\_person. [↑](#footnote-ref-2)
3. Conditions coded as Diagnosis 6100 (Hearing Loss), 6200 (Otitus Media (Chronic Ear Infection)), 6201 (Otitis Media), 6202 (Otosclerosis), 6204 (Labyrinthitis), 6205 (Meniere's Syndrome), 6207 (Loss or partial loss of Ear), 6209 (Benign growth of Ear), 6210 (Hearing Loss), 6211 (Perforated ear drum), and 6260 (Tinnitus). [↑](#footnote-ref-3)
4. See footnote 1. [↑](#footnote-ref-4)
5. Categorization is based on a prior service-connected decision for the diagnostic code and not on a decision for the diagnosis code within the subject claim. [↑](#footnote-ref-5)