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

**December 2016**



**Version 1.0**

Contents

[1 Overview 1](#_Toc470010387)

[1.1 Introduction 1](#_Toc470010388)

[1.2 Tinnitus Model Objectives 1](#_Toc470010389)

[2 Tinnitus Model Results 3](#_Toc470010390)

[2.1 Tinnitus Model Performance Analysis 3](#_Toc470010391)

[3 BCDS Modeling - Business Process Integration 4](#_Toc470010392)

[Appendix A: Technical Specifications 6](#_Toc470010393)

[Appendix B: BCDSS Model Architecture 8](#_Toc470010394)

[B.1 BCDSS Pattern-Matching Architecture 11](#_Toc470010395)

[Appendix C: Summary of MITRE Corporation’s Methodology 14](#_Toc470010396)

# Overview

## Introduction

This Tinnitus Model Manual describes the methodology, specifications, performance parameters, and statistical limitations of the Benefits Claims Decision Support System (BCDSS) Tinnitus Model (BCDSS Model No. 004). The Tinnitus Model is one of a portfolio of predictive models supporting 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.

## Tinnitus Model Objectives

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

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

An analysis by MITRE Corporation 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 1.0 ) development efforts for aggregate ear related conditions achieved accuracy rates of 94.38 percent for 60.81 percent of eligible supplemental claims.

**Footnote for Ear-related conditions:** 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).

A performance analysis of the BCDSS 1.0 Ear Model results (Figure 1) revealed two key findings:

Figure : Distribution of Modeling Results by Diagnostic Code

1. 93 percent of all claims eligible to be rated using BCDSS Hearing Model 1.0 were rated using the General Hearing Loss (6100) or Tinnitus (6260) diagnosis codes that is 23 percent of the claims contained hearing loss code ONLY, 14 percent contained Tinnitus code ONLY, and 57 percent contained claims for both General Hearing Loss and Tinnitus codes.
2. Recurrent Tinnitus, once substantiated and demonstrated to be service connected, must receive a rating of 10 percent (See Section 2). As a result, the rating for supplemental claims where the claimant has previously received a 10% rating for Tinnitus will remain unchanged at 10%
3. Because of the dominance and predictability of Tinnitus in the population of hearing eligible supplemental request for increase claims,, the BCDSS business team proposed 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 Tinnitus model is to demonstrate that ***a condition-specific Tinnitus model can yield an accuracy rate greater than 95 percent across ALL eligible claims (a through-put of 100 percent).***

# Tinnitus Model Results

The Tinnitus model 2.2 is designed to first establish whether a claim is [eligible](#_Eligibility) [[1]](#footnote-1)to be rated using the model, and then to establish that a prior rating for Tinnitus has been granted. Upon confirmation, the model simply assigns the issue within the claim a rating of 10 percent consistent with the rating schedule, which specifies:

“6260 Tinnitus - Recurrent 10 Percent

* Note (1): A separate evaluation for tinnitus may be combined with an evaluation under diagnostic codes 6100, 6200, 6204, or other diagnostic code, except when tinnitus supports an evaluation under one of those diagnostic codes.
* Note (2): Assign only a single evaluation for recurrent tinnitus, whether the sound is perceived in one ear, both ears, or in the head.
* Note (3): Do not evaluate objective tinnitus (in which the sound is audible to other people and has a definable cause that may or may not be pathologic) under this diagnostic code, but evaluate it as part of any underlying condition causing it.”

Similarly, the VBA rating calculator generates the following in its award explanation:

We have assigned a 10 percent evaluation based on recurrent tinnitus. A single evaluation for recurrent tinnitus is assigned, whether the sound is perceived in one ear, both ears, or in the head. This is the highest schedular evaluation allowed under the law for tinnitus, recurrent.

## Tinnitus Model Performance Analysis

The team tested the Tinnitus model to verify that the business logic, when applied to eligible claims, performed within established specifications. Model performance analysis identified **243,751** eligible claims containing a prior rating for tinnitus between 2008 and 2014. Application of the model had an accuracy rate of **99.64 percent** for **100 percent** of claim throughput.

The 883 claims containing Tinnitus diagnosis that differed from the specified business rule (i.e., the assigned rating was NOT 10 percent) broke down as follows:

* 880 were assigned a rating of 0 percent.
* Two were assigned a rating of 20 percent.
* One was assigned a rating of 100 percent.

The 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.

# BCDS Modeling - Business Process Integration

VBA’s current claims adjudication process requires claimants to file a supplemental claim for increased disability compensation when they believe they are eligible based on a change in condition, information, or circumstances. Similar to the original claim for benefits, the claimant must present medical evidence sufficient to substantiate the condition, if they believe the condition has worsened and to what degree. After the veteran submits the necessary evidence, a decision is made and promulgated after review and authorization by appropriately credentialed VBA staff.

Statistical claims adjudication uses historical data captured in the filing patterns and substantiating diagnostic information of millions of claimants to establish a statistical basis for predicting the rating for a claim containing similar factual attributes in the future Using Statistical Adjudication where model performance is acceptable could allow VBA to calculate the rating of a claim without requesting additional substantiating medical evidence. However, as to supplemental claims for increase, service connection and an initial fully substantiated rating must first be established by the claimant. As a result, BCDSS models currently address only supplemental claims (EP-02X) where the claimant has received at least one prior service-connected rating for the modeled condition.

The Office of Compensation Service originally envisioned using the BCDSS calculated rating to provide a claimant the opportunity to accept his/her statistically adjudicated rating (where a change might be granted) or electing to have his/her claim adjudicated through the traditional process.

Analysis of the claims data by the BCDSS team revealed that only 40 percent of hearing related supplemental claims for increased compensation are granted. Of these, 27 percent are for general hearing loss and the balance are the vast majority of the balance are for hearing conditions other than Tinnitus. 243,123 of 251,567 (or 96.6 percent) evaluated supplemental Tinnitus claims were continued at the current level of disability compensation (the increase was denied).

The question arises as to whether VBA will be able to use statistical adjudication to identify and rate the vast majority of hearing related supplemental requests for increase claims which are confirmed and continued: that is essentially denied from the Veteran’s viewpoint. However, although rating these claims using statistical adjudication would require statutory and regulatory changes, the potential benefits are enormous. Veterans would see almost instantaneous feedback on their claim and accelerated adjudication where a positive decision is rendered. The VA would also realize efficiencies through reduced workload. In order to be viable in the VBA environment, more analysis and modeling needs to conducted to develop models which can address claims which contain claims for multiple distinct disability conditions Hence, VACI and VBA have identified four potential uses for BCDSS in its current form:

1. Delivering an automated rating to the RVSR (to inform the final claim rating)–In this case, the rating would be calculated immediately following establishment and passed to the RVSR following claim development via the BCDSS user-interface.
2. Delivering an automated rating to the Decision Review Officer (DRO) or Senior Veteran Service Representative (VSR) during claim authorization (to assist in the review and verification of the manually calculated rating)–In this case, the rating would be calculated immediately following establishment and passed to the Authorizer following award generation by the responsible VSR.
3. Delivering large sets of BCDSS statistically adjudicated claim decisions to VBA quality assurance staff as means to evaluate the accuracy of claim adjudication decisions, or trends in accuracy across multiple Regional Offices.
4. Providing BCDSS access to RVSRs to support training.

# Appendix A: Technical Specifications

BCDSS Model requirements stipulate that the three core components be defined for each BCDSS model: 1) the Decision Determination Matrix (DDM); 2) the Target Claim Feature-Set Data and Computational Specifications; and 3) the Claimant Feature-Set Data and Computational Specifications. The Tinnitus Model varies from this requirement because of its simplicity. Target Claim and Claimant feature and computational specifications are still required, but a single binary business rule is substituted from the DDM. Each of these components are summarized below.

# Tinnitus Model Core Business Rule (Substitution for DDM)

Tinnitus, once diagnosed and medically substantiated, is rated at 10%. As a result, the core business rule applied to all Tinnitus (Code 6260) issues in claims contending Tinnitus can be stated as:

If ELIGIBILITY is NOT NULL then RATING=10

# Eligibility

Claim eligibility is established by verifying that the claim is a re-opened end-product (end\_product\_type\_cd=‘02%’) and that the claimant is contending a relevant condition. The contention classification codes that determine relevance are provided in table 1 below.

**Table 1: Contention Codes**

| **Feature** | **Definition** |
| --- | --- |
| 2200 | Ear Condition |
| 3140 | hearing loss |
| 3150 | hearing loss, sensorineural |
| 4700 | Non-specific Ear Condition |
| 5710 | Ringing in Ears |
| 6850 | Tinnitus |

SQL Statement:

**Select** ptcpnt\_vet\_id, bnft\_claim\_id, end\_product\_type\_cd, date\_of\_claim, cntntn\_clsfn\_id, prfil\_dt

**From** XXXX\_rating\_corp\_claim

**where** end\_product\_type\_cd is like ‘02%’ and

(cntntn\_clsfn\_id=’2200’ or cntntn\_clsfn\_id=’4700’ or cntntn\_clsfn\_id=’3140’ or cntntn\_clsfn\_id=’5700’ or cntntn\_clsfn\_id=’6860’)

# Claimant Data Computations

Only one claimant data based feature is necessary to calculate Tinnitus ratings and that is to establish that a service-connected (or 1151 Granted) prior rating was awarded for Tinnitus (Diagnostic Code 6260).

SQL Statement

**Select** ptcpnt\_vet\_id, prfil\_dt, prcnt\_nbr, max(begin\_dt)

**From** XXXX\_rating\_decision

**Where** diagnosis\_code=‘6260’ and

(narrative\_codesheet\_ind=‘N’ or narrative\_codesheet\_ind=’C’) and (dsblty\_decn\_type\_cd='SVCCONNCTED' or dsblty-decn-type\_cd='1151GRANTED') and

prcnt\_nbr is not null

Then Rating=10

# Appendix B: BCDSS Model Architecture

The BCDSS team evaluated MITRE’s analytical methodology for statistically adjudicating claims for scalability and the ease with which the classifiers could be automated. The BCDSS team’s evaluation determined that a simpler approach using a pattern-matching architecture could streamline rating calculation and prediction, while providing enhanced scalability. This approach retains the ability to apply Ordinal Logistic Regression (OLR) or other advanced statistical algorithms should such analytical capabilities become necessary to satisfy performance parameters. The BCDSS team based this hypothesis on the following:

* A cursory analysis of supplemental claims for increased compensation containing at least one ear-related contention indicated that most claimants file an average of less than two such claims. The team concluded event-specific analysis might yield just as accurate a result given the limited number of relevant events contained within most claimants’ adjudication history.
* Establishing an architecture that can use and modify a claimant’s adjudicatory history, as well as training data to derive a single Combined Disability Determination (CDD), seemed unnecessarily cumbersome. It might also present significant performance challenges when there are a high numbers of users and transaction requests. The final design concept needs to be repeatable, efficient, and highly scalable.
* MITRE applied machine-learning algorithms to establish predictive adjudicatory patterns from historical training and test data and then applied the pattern against target claims to calculate the likely CDD. Current computer processing speeds can execute pattern-matching routines far faster and with greater efficiency than machine-learning algorithms, even with inventories of more than 100,000 patterns.
* A basic modeling architecture that relies on matching a fixed set of claim attributes for a specified claim that includes the features contained within an inventory of patterns is extremely scalable. It can also exclude invalid patterns. This would be far more challenging to accomplish with statistical adjudication models.

Based on this, the team developed and tested a more streamlined and scalable design concept. This design concept assumes that the attributes related to specific conditions and the decisions of VA personnel who adjudicate claims for those conditions form repeated and predictable patterns. These patterns are captured within the values of a predefined suite of predictive characteristics or features. A sample pattern inventory is shown in Figure 2:



Figure : Sample Pattern Inventory

The BCDSS team tested its hypothesis with the same data that MITRE used to conduct its analysis and the same feature set:

1. The team extracted claims that included one or more ear-related contentions from the initial set of 1.2 million claims.
2. The data were grouped based on unique combinations of values for the set of features.
3. The patterns were numbered and aligned with the resulting CDD for the ear (the calculated current CDDs were rounded to the nearest 10 percent or quantified based on MITRE’s analysis).
4. The team executed an automated routine to match the same set of claims with the pattern inventory to determine whether the logic was self-repeating (i.e., the claims used to identify the patterns could again be matched to provide an accurate result).
5. The initial test indicated that fewer than 65,000 patterns were required to determine the CDD at accuracy rates above 90 percent. Based on these initial results, the team adopted the pattern-matching architecture for preliminary development and testing purposes.

The BCDSS team used a four phase technical approach (Figure 2 below) consistent with previous models.



Figure 2: BCDS Model Development Approach

* **During Phase 1,** the team extracted and segmented data to ensure sufficient claimants with relevant claimed contentions and adjudication decisions are included to form distinct development and test data sets. The team then synthesized the data within the development data set to form a high-resolution feature set (i.e., values were calculated to the highest level of precision in the underlying data: in this case, time periods were calculated in days where defined by starting and end dates).
* **During Phase 2,** the team aggregated the development data to form unique, numbered patterns, and aligned the patterns with the corresponding CDDs. Sensitivity and optimization analyses were conducted on each feature to identify the relevant impact of the feature on accuracy, and the extent to which the feature could be aggregated to increase the rate at which a pattern is re-used across test data without compromising the model’s predictive accuracy. For example, storing the period between the claim date and last CDD in years rather than days.
* **During Phase 3,** the team optimized the model to maximize “rate of use” within the context of target through-put and accuracy performance parameters. The confidence level of pattern sets are a function of repetition (or “rate of use”) relative to the size of the sample (number of claims eligible for modeling).
* **During Phase 4,** the team tested the model against an entirely separate set of verification data to verify that the performance results initially obtained during optimization were replicated, and confidence levels were sustained. The government was also provided the model to allow its analysts to conduct independent analyses of the models.

## B.1 BCDSS Pattern-Matching Architecture

The BCDSS pattern-matching modeling architecture relies on replicated fact patterns for similar conditions and claim adjudication histories, and the consistent promulgation of ratings for such fact patterns by VA employees to reliably predict ratings (without subjecting the claim to additional substantiation and human evaluation). A relatively small number of claim attributes and established facts contained in the claimant’s filing history comprise these patterns. The rating of a specific condition within the target claim is obtained by establishing the fact pattern of the claim using a predefined set of attributes, and matching it to a specific historical pattern and the rating that most often results. This BCDSS pattern-matching architecture contains three major components:

1. A set of claim-specific features and the associated specifications for calculating the values for these features. These features relate to the claim that will be subject to predictive modeling—the target claim.
2. A set of claimant specific features and the associated specifications for calculating the values for these features. These features relate to attributes of the claimant’s filing history, including decisions contained in prior claims for similar conditions, the period of elapsed time between the target claim and prior decisions, the place of the target claim in the sequence of relevant claims in the claimant’s adjudication history, etc.
3. The Decision Determination Matrix (DDM) or set of predictive patterns and related CDDs for the modeled condition. The columns of the DDM constitute the various features as well as the CDD and their performance attributes. The rows are the unique combinations of values that constitute the predictive patterns.

The basic logic described above is depicted in Figure 3.



Figure 3: Pattern-Matching Logic

The BCDSS provides an automated environment in which authorized users select one or more target claims to adjudicate, one or more models to apply, and one or more output formats preferred by the user. The system then executes automated routines in accordance with the user selections to derive the results. Three inter-related modeling engines support these functions as depicted in Figure 4.



Figure 4: BCDS System Workflow and Modeling Architecture

Upon user selection, the BCDSS retrieves appropriate model(s) from the model repository. The ingest engine verifies that the claim satisfies the eligibility requirement and executes the required calculations against the target claim and related claimant data to create a composite feature set. The modeling engine then conducts the pattern matching and either assigns the associated CDD (where a match is found), or identifies that the CDD could not be established. Finally, the output engine integrates descriptive information about the claim (and the CDD, where appropriate) and formats the data in accordance with the user’s selection.

Appendix C: Summary of MITRE Corporation’s Methodology

The methodology used by MITRE to complete its analyses is documented in a series of engineering notebooks (completed under Contract No. VA118A13J0421/VBA OSP COMPENSATION SERVICES CLIN 0005 IFCAP 101-J47030). MITRE’s methodology was as follows:

1. VA supplied MITRE (and later, the BCDSS team) with a set of claims and veteran data for the purpose of the analysis. These VBA claims and rating decision data were first conditioned, which involved aligning the claims data with the corresponding decision data to produce a chronological claimant adjudication history and feature vector. This feature vector constitutes a recurring sequence of claim attributes and the trajectory of related decisions that result in common outcomes.
2. Machine-learning algorithms or “classifiers” (e.g., Random Forest, Logistic Regression, Ordinal Logistic Regression (OLR), and Auto-encoder) quantified the relative importance of a suite of common claim attributes (or features) for determining the CDD for subject conditions within supplemental claims, across samples of the above described claimant adjudication histories. Those features with greatest predictive value were subjected to testing and predictive modeling.
3. The classifiers were also applied to evaluate their relative performance in predicting the CDD of randomly selected (eligible) claims. The above-mentioned dataset, once conditioned, was segmented into a Training Set and a Test Set. Training data supplied the analytical basis for determining correlations between features to calculate the predicted CDD. Separate sample claims were then used to test the predictive capabilities.
4. Results were measured based on throughput (i.e., the ratio of the number of claims processed by the classifier to the total number of claims) and accuracy (i.e., the percentage of predicted CDDs that are equal to the CDD assigned by the responsible Rating Veterans Service Representative (RVSR).

MITRE’s work established the relative importance of specific features and identified methods that can quantify their importance. These analyses also provide performance evaluations of various classifiers MITRE applied to statistically adjudicate claims that match specific fact patterns. These patterns include the features identified as having significant predictive value, as well as prior claimant adjudication decisions, time periods between the subject claim and prior decisions, the diagnostic information used to describe those decisions, attributes of the claimant (age at the time of filing, etc.), and data within the subject claim (e.g., the contention classification codes used to describe the Veteran’s new claimed disability).

MITRE categorized its set of features as either numeric or text. Numeric values were calculated where necessary (e.g., time periods, counts, etc.), rounded to the nearest decile (e.g., 35 would be stored as 40), and otherwise stored as true/false values (e.g., whether a specific diagnostic code or contention code was used). The same true/false logic was applied to the calculation of text values (i.e., the presence of specified text strings in specified fields within the data).

1. Eligibility is determined by:

   * Establishment as an End Product (EP-02X)
   * The presence of contention codes 2200, 3140, 3150, 4700, 5710, or 6850
   * Confirmation of a prior rating. The feature set, in this case, is limited to the prior rating for service connected issue identified for diagnostic code 6260.

   [↑](#footnote-ref-1)