**Benefits Claims Decision Support System (BCDSS) General Hearing Loss Model**

Contents

[Section 0: Overview 2](#_Toc468094180)

[Introduction 2](#_Toc468094181)

[Purpose 2](#_Toc468094182)

[Section 1: Background 3](#_Toc468094183)

[Statistical Adjudication of Disability Compensation Claims 3](#_Toc468094184)

[Observations and Limitations of Using Statistical Models to Adjudicate Claims for Disability Compensation Benefits 3](#_Toc468094185)

[Summary of MITRE Corporation’s Methodology 5](#_Toc468094186)

[Section 2: Conceptual Design and Model Development Approach 6](#_Toc468094187)

[BCDS Model - Conceptual Design 6](#_Toc468094188)

[Section 3: Findings from the BCDS Ear Model Performance Testing & Analysis 10](#_Toc468094189)

[Section 4: Technical Approach 11](#_Toc468094190)

[BCDS Pattern-Matching Architecture 13](#_Toc468094191)

[Section 4: Data Requirements 16](#_Toc468094192)

[Data Source and Structure Overview 16](#_Toc468094193)

[BCDS General Hearing Loss Model Data Segmentation 16](#_Toc468094194)

[Section 5: Model Development and Optimization 17](#_Toc468094195)

[General Hearing Loss Model Performance Targets 17](#_Toc468094196)

[Hypothesis-driven Model Formulation Process 17](#_Toc468094198)

[Model Optimization Process for Co-Occurring Conditions 19](#_Toc468094199)

[Testing and Performance Analysis of Co-Occurring Condition Models 19](#_Toc468094200)

[Expanded General Hearing Loss Model Development, Optimization and Testing 20](#_Toc468094201)

# Section 0: Overview

## Introduction

The Benefits Claims Decision Support System (BCDSS) project will demonstrate the feasibility of using automation and predictive models to calculate the disability rating for claims for general hearing loss contained within Veteran claims for disability compensation benefits. Such capabilities can help the Department of Veterans Affairs (VA) improve the quality and consistency of its current claims decisions and potentially deliver the services based on the outcome of those decisions to Veterans much faster than today. Both outcomes, if achieved, will significantly improve the Veteran Experience and efficiency of VA’s disability benefits claims process.

## Purpose

The BCDSS General Hearing Loss Model determines the rating for hearing loss contained within supplemental claims for increased disability compensation benefits (claims). This General Hearing Loss Model Manual describes the methodology, specifications, performance parameters, and statistical limitations of the General Hearing Loss Model (BCDSS Model No. 0003). The manual also provides:

* An overview of the background and origins of the BCDSS project (Section 1) and the prior analytical work conducted by MITRE Corporation that gave rise to the project
* A general description of the conceptual design and model development approach adopted by the BCDSS project team, and its relation to MITRE’s previous work (Section 2);
* A description of the BCDSS Team’s initial findings relating to its Ear Model, and the analytical basis for deconstructing and segmenting this model into two distinct predictive models: General Hearing Loss and Tinnitus (Section 3).
* A description of the data requirements, structure, and segmentation used to develop the General Hearing Loss Model as well as a listing of the claim attributes selected for used to develop the model (Section 4)
* A description of the General Hearing Loss Model performance parameters, and the development and optimization process used to satisfy those parameters (Section 5)
* A description of the General Hearing Loss Model testing and evaluation process, and the statistical results of these processes (Section 6).

Sections 1 and 2 apply generally to all BCDS models, while Sections 3 through 6 are specific to the General Hearing Loss Model.

**NOTE:**Sections 0 through 2 are descriptive of the methodology and consequently are the same in each Model Manual.

# Section 1: Background

## Statistical Adjudication of Disability Compensation Claims

Between 2013 and 2014, the Veterans Benefits Administration’s (VBA) Office of Compensation Service asked the MITRE Corporation to determine the feasibility of using predictive models to adjudicate disability compensation claims with little or no human intervention. This research project demonstrated that Combined Disability Determination (CDD) for hearing- and knee-related disability conditions can be calculated using machine learning algorithms and other advanced statistical analysis methods. MITRE called this capability “statistical adjudication.” VBA, in collaboration with VA Center for Innovation (VACI), seeks to demonstrate the feasibility of using an automated information technology system to apply similar methods at scales and performance parameters similar to those experienced during normal VBA claims processing operations.

## Observations and Limitations of Using Statistical Models to Adjudicate Claims for Disability Compensation Benefits

Statistical analysis and modeling techniques require a substantial statistical basis or foundation for identifying valid meaningful correlations between variables and establishing predictive patterns to determine the calculated outcomes. There are two broad claims groups:

* Original—the first claim filed by a claimant for disability benefits
* Supplemental—any subsequent claim filed by the claimant seeking to adjust his/her benefit or disability status.

The claimant’s service history and medical information to substantiate disability decisions for the subject disability condition form this statistical foundation. For this reason, only supplemental claims where prior information exists (and not original claims) are eligible for statistical adjudication methods at this time. Related limitations include the following:

* The subject claim must contain at least one fully promulgated prior disability determination for the specific condition being statistically adjudicated. In many cases, original claims do not include the condition that is the subject of the supplemental claim (e.g., hearing loss may not have been claimed by the Veteran in the original claim, but it forms the basis of the supplemental claim for an increase in the Veteran’s overall disability rating).
* By definition, the claim and prior claimant data must align with general conditions and data for claimants filing similar claims. Unique conditions may be eligible, but are un-addressable (as distinguished from inaccurately adjudicated). For example, a claim may resemble a known predictive pattern but the model will return an inaccurate disability determination, while the facts relating to an un-addressable claim inadequately resemble those of claims supplying the statistical basis of the model.

The potential operational efficiencies offered to VBA through the automation of supplemental claims adjudication are enormous, notwithstanding these limitations. Supplemental claims compose 61 percent of all claims requiring a “rating” (the substantiation and evaluation of medical evidence to make an award determination). The claims are similarly labor and time intensive, often requiring additional medical examinations, records collection, and evaluations.

## 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,” a process that involves 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) were used to quantify 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 (the “ratio of the number of claims processed by the classifier to the total number of claims”) and accuracy (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 be used to 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).

# Section 2: Conceptual Design and Model Development Approach

## BCDSS Model—Conceptual Design

The BCDSS team evaluated MITRE’s analytical methodology for statistically adjudicating claims for scalability and the ease with which the classifiers—as described in the Engineering Notebooks—could be automated. The BCDSS team’s evaluation suggested that a simpler approach using a pattern-matching architecture could streamline rating calculation and prediction, and provide enhanced scalability. This approach retains the ability to apply OLR or other classifiers 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 increase containing at least one ear-related contention indicated that most claimants file, on average, 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 claimant’s adjudication history.
* Establishing an architecture that would ingest and condition at runtime a claimant’s adjudicatory history, as well as sufficient conditioned training data to derive a single CDD, seemed unnecessarily cumbersome. It might also present significant performance challenges when servicing high numbers of users and transaction requests. The final design concept would need to be repeatable, efficient, and highly scalable.
* The analytical methodology used by MITRE relied on applying machine-learning algorithms to establish predictive adjudicatory patterns from historical “training” data and then applying the pattern against target (or “test”) claims to calculate the likely CDD. The number of such patterns required to accurately capture a significant fraction of claims for specific conditions is limited. Current computer processing speeds can execute pattern-matching routines far faster and with greater efficiency than more sophisticated 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 (or target claim) with the features contained within an inventory of patterns is extremely scalable in an operational setting, and can be maintained over time to EXCLUDE invalid patterns. This would be far more challenging to accomplish with statistical adjudication models.

Accordingly, the team developed and tested a more streamlined and scalable design concept. This design concept assumes that claimant filing 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”).

For example, a 60 year old Veteran files for increased disability compensation based on the perceived worsening of his hearing impairment. His original claim, filed five years previously, was granted at 10 percent based on a medically substantiated diagnosis of tinnitus. A relevant feature set can be distilled from this information composed of claimant age, prior ear CDD, age of prior ear CDD, and prior diagnosis code. The resulting pattern for this specific Veteran is illustrated in the table below.

Table : Sample Pattern

|  |  |  |  |
| --- | --- | --- | --- |
| **Claimant Age (Years)** | **Prior Ear CDD (%)** | **Age of Prior Ear CDD (Years)** | **Diagnosis Code** |
| ***60*** | ***10*** | ***5 years*** | ***Tinnitus*** |

An example set of patterns are provided in Figure 1 below.



Figure 1: Sample Pattern Inventory

MITRE categorized its set of features as being either numeric or text. In this case, numeric values were calculated where necessary (e.g., time periods, counts, etc.), rounded to the nearest year (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).

The BCDSS team tested its hypothesis with the same data that MITRE used to conduct its analysis and the same feature set. Data from 1.2 million claims were first extracted from the set where the claimant had filed at least one claim with an ear-related contention. The data were next aggregated to define the underlying predictive patterns: unique combinations of values for the set of features. The patterns were numbered and aligned with the resulting CDD for the Ear (the calculated current CDDs were rounded to the nearest 10% or quantized consistent with MITRE’s analysis). Finally, 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). 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.

# Section 3: Findings from the BCDS Ear Model Performance Testing & Analysis

The BCDSS team developed the Ear Model (001), which provided a combined disability determination (or CDD) for ear related conditions (defined as diagnosis codes: 6100, 6200, 6201, 6202, 6204, 6205, 6207, 6209, 6210, 6211, and 6260). The model addressed 94.38 percent of eligible claims, with an accuracy rate of 60.81 percent.



Figure 2: Distribution of Modeling Results by Diagnostic Code

A performance analysis of the Ear Model results (Figure X above) revealed two key findings:

1. 93 percent of all eligible claims were diagnosed using the General Hearing Loss (6100) or Tinnitus (6260) diagnosis codes: 23 percent for hearing loss ONLY, 14 percent for Tinnitus ONLY, and 57 percent for both General Hearing Loss and Tinnitus.
2. Tinnitus, once substantiated and rated, must receive a rating of 10 percent. As a result, supplemental claims where the claimant has previously received a rating will almost always be rated at 10 percent.

Based on these findings, VACI requested that the Ear Model be “disaggregated” into its constituent components, redeveloped, and evaluated to determine if separate models provide improved results for accuracy and throughput.

# Section 4: Technical Approach

The BCDSS team used a four phase technical approach (Figure 3 below) consistent with previous models.



Figure 3: BCDSS Model Development Approach

During Phase 1, the team will extract and segment data to ensure sufficient claimants with relevant claimed contentions and adjudication decisions are included to form distinct development and test data sets. The team will next synthesize the data within the development data set to form a high-resolution feature set (i.e., calculate values to the highest level of precision offered from the underlying data: in this case, calculate time periods in days where defined by starting and end dates).

During Phase 2, the team aggregates the development data to form unique, numbered patterns, and aligns the patterns with the corresponding CDDs. Sensitivity and optimization analyses will be conducted on each feature to identify the relevant impact of the feature on accuracy and the extent to which the feature can be aggregated to increase the rate at which a pattern is re-usable 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 will optimize the model to maximize “rate of use” within the context of target throughput 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 will test the model against an entirely separate set of verification data to confirm the initial performance results during optimization as well as maintaining sustained confidence levels. The government will be provided the model so its analysts can conduct independent analyses of the models.

## BCDS Pattern-Matching Architecture

The BCDSS pattern-matching modeling architecture relies on replicated fact patterns for similar conditions and claim adjudicatory 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 BCDSS team obtains the rating of a specific condition within the target claim 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 is composed of 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, i.e., 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 4.



Figure 4: 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 based on the user’s selections to derive the results. Three interrelated modeling engines support these functions as depicted in Figure 5.



Figure 5: BCDSS System Workflow and Modeling Architecture

Upon user selection, the BCDSS retrieves the 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 associate 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.

# Section 4: Data Requirements

## Data Source and Structure Overview

As stated previously, the data used to produce the BCDSS General Hearing Loss Model (003) was the same data provided to MITRE to support its analysis. This database was provided by VBA’s Office of Performance Analysis and Integrity (PA&I) on December 12, 2016. The data was composed of over 41 million claims processing transactions under various end-products in eight tables. Of these, three contain the bulk of relevant data: Rating\_Corp\_Claim containing claim specific information, Rating\_Decision containing claimant decision data, and Rating\_Decision\_Spec\_Issue containing rating decision data.

## BCDS General Hearing Loss Model Data Segmentation

The traditional method for developing predictive models is to segment data that will form the basis of the analysis by defining one subset for model development and reserving a second subset for testing and model analysis. **213,342** eligible claims were segmented and used to develop and optimize the initial General Hearing Loss model. The BCDSS team segmented these claims based on the presence of at least one hearing loss-related contention (Contention Classification Codes 3140, 3150, 2200 or 4700) while at least one eligible ear-related issue must be service-connected and a rating promulgated within a previously adjudicated claim for benefits.

# Section 5: Model Development and Optimization

## General Hearing Loss Model Performance Targets

## The original BCDSS Ear Model demonstrated that the Combined Disability Determination (CDD) for claims for ear-related conditions has an accuracy rate of **94.38 percent**, and can do so for relevant issues contained within **60.81 percent** of eligible supplemental claims. This performance parameter, stated as 61% throughput at 94% accuracy, is applicable to the BCDSS General Hearing Loss Model.

## Hypothesis-driven Model Formulation Process

The team first worked with VBA subject matter experts to define analytical hypotheses that may provide an empirical basis for predicting a rating for general hearing loss based on prior evidence contained in VBA corporate claims data. The two hypotheses deemed most likely to provide predictive indicators that would deliver satisfactory modeling performance were:

1. The co-occurrence of medical conditions in the claimant’s filing history that might indicate a co-relatable “trajectory” in the claimant’s general hearing loss condition. For example, analyses might reveal that both general hearing loss and Traumatic Brain Injury (TBI) increase (or not) in parallel over time. For example, should the increase occur at multiple times in the claimant’s history at measurable intervals and increases, sufficient statistical evidence might form accurate predictive patterns.
2. The occurrence of text strings within a prior diagnostic narrative that might, in conjunction with previously used predictive features, provide sufficient evidence to form reliable predictive patterns.

The BCDSS team first established what conditions co-occurred most frequently with general hearing loss. The 10 most frequent co-occurring conditions are presented below.

The team next compiled three predictive features for eligible general hearing loss claims: the rating assigned for the condition in the most recent prior claim for benefits, the period of time that elapsed since that rating was determined, and whether and to what extent the rating changed over time. The same values were calculated for General Hearing Loss, as well as features from the earlier Ear Model that were found to have the greatest predictive value. The complete set of initial General Hearing Loss Features is:

Claimant Age

9411 Prior Rating (PTSD)

Prior 6100 Rating

Age of Prior 9411 Rating

Age of Prior 6100 Rating

Change in Prior 9411 Rating

Contention Code 2200

7913 Prior Rating (Diabetes mellitus)

Contention Code 3140

Age of Prior 7913 Rating

Contention Code 3150

Change in Prior 7913 Rating

Contention Code 4700

5237 Prior Rating (Lumbosacral or Cervical Strain)

6260 Prior Rating (Tinnitus)

Age of Prior 5237 Rating

Age of Prior 6260 Rating

Change in Prior 5237 Rating

Change in Prior 6260 Rating



Data were then aggregated to identify all unique combinations of the defined features (or patterns) for eligible claims within the development dataset. For the Initial General Hearing Loss build, the number of unique patterns was slightly more than 95,000.

The team next assigned unique identifiers to each pattern and aligned the patterns with unique instances of resulting condition-specific Ratings. In some cases, the same pattern may result in a different rating. In these cases, the number of occurrences for each different rating is calculated. The most common rating for the specific pattern is then assigned for the model. Once each pattern was assigned a unique rating (and the pattern-set or model is complete), the team applied the model to the original data to calculate initial throughput, accuracy, and re-use rates.

## Model Optimization Process for Co-Occurring Conditions

The team next optimized the model to determine the set of features that yields the highest throughput, accuracy, and pattern re-use rates. The more often a fact pattern is applicable (its rate of use), the higher its predictive value. Adding and subtracting features from the overall feature set increases pattern re-use rates. Similarly, aggregating individual features increases the applicability of the results patterns to targeted claims (e.g., converting days to years rounded to whole integers allows the same pattern to be applied to more claims). Each different configuration forms a unique model scenario. The team next conducted sensitivity analysis to quantify:

* The contribution of each individual feature to the overall accuracy and throughput of the model; and
* The impact of feature aggregation on accuracy and throughput.

The team narrowed the number of viable model configuration scenarios to a defined set of preliminary models based those that yield the highest rates of use within throughput and accuracy thresholds. This more limited set of preliminary models formed the basis for testing and performance analysis.

**NOTE:** Early in the optimization process, the BCDSS Team determined that the rating for general hearing loss changed very infrequently (in only 27.9 percent or 54,675 of the claims within the development data) from what the rating was previously. This suggested that an exceptions-based model might prove more accurate and reliable (i.e., identifying the patterns that determine the direction and extent of change, when warranted, as opposed to identifying patterns that accurately predict all ratings, regardless of whether a change is warranted).

## Testing and Performance Analysis of Co-Occurring Condition Models

The team next test the preliminary ear models against the segmented test data. The models varied by feature-set (e.g., inclusion of all or some of the contention text or diagnostic text features) and pattern-specific accuracy levels (90%, 80%, etc.). Multiple preliminary models were tested to determine which model best replicated its results, and to quantify pattern rate of use and confidence levels. Finally, the models were updated and re-run to determine the performance levels of the model following refinement. The final results are presented in Table 2.

Table : General Hearing Loss Testing and Optimization Results



Testing revealed that patterns composed of co-occurring features calculated the rating for general hearing loss at a maximum accuracy rate of 78 percent, marginally better than merely assigning the same rating to all general hearing loss diagnostic codes when a prior rating had been assigned (as stated earlier, roughly 72 percent). Based on these results, the BCDSS team determined that additional features beyond co-occurring conditions would need to be added to the General Hearing Loss model.

## Expanded General Hearing Loss Model Development, Optimization and Testing