Department of Veterans Affairs (VA)   
  
Benefits Claims Decision Support System (BCDSS)   
  
Ear Model



**July, 2016  
Version 1.0**

Revision History

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Date | Version | Description | Author | Reviewer(s) |
| 07/31/2016 | 1.0 | Included the Revision History section | Dave |  |
|  |  |  |  |  |
|  |  |  |  |  |

Contents

[Section 0: Overview 2](#_Toc449638391)

[Introduction 2](#_Toc449638392)

[Purpose 2](#_Toc449638393)

[Section 1: Background 3](#_Toc449638394)

[Statistical Adjudication of Disability Compensation Claims 3](#_Toc449638395)

[Observations and Limitations of Using Statistical Models to Adjudicate Claims for Disability Compensation Benefits 3](#_Toc449638396)

[Summary of MITRE Corporation’s Methodology 5](#_Toc449638397)

[Section 2: Conceptual Design and Model Development Approach 6](#_Toc449638398)

[BCDS Model - Conceptual Design 6](#_Toc449638399)

[Technical Approach 10](#_Toc449638400)

[BCDS Patter-Matching Architecture 11](#_Toc449638401)

[Section 3: Data Requirements 15](#_Toc449638402)

[Data Source and Structure Overview 15](#_Toc449638403)

[BCDS Ear Model Data Segmentation 15](#_Toc449638404)

[Section 4: Model Development and Optimization 16](#_Toc449638405)

[Knee Model Performance Targets 16](#_Toc449638406)

[Model Development Process 16](#_Toc449638408)

[Model Optimization Process 17](#_Toc449638409)

[Testing and Performance Analysis 18](#_Toc449638410)

[Quantitative Characterization of Addressed Claims 19](#_Toc449638411)

[Description and Quantitative Overview of Unaddressed Claims 22](#_Toc449638412)

[Section 5: Model Verification Testing 22](#_Toc449638413)

[Verification Assessment 23](#_Toc449638414)

[Appendix 25](#_Toc449638415)

[Acronyms 25](#_Toc449638416)

[Glossary 25](#_Toc449638417)

# Section 0: Overview

## Introduction

The Benefits Claims Decision Support (BCDS) project will demonstrate the feasibility of using automation and predictive models to calculate the combined disability determination (CDD) for specific conditions contained within Veteran claims for disability compensation benefits. Such capabilities can help VA improve the quality and consistency of its current claims decisions, and potentially deliver 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 BCDS Ear Model determines the CDD for ear conditions contained within supplemental claims for increased disability compensation benefits (claims). This Ear Model Manual describes the methodology, specifications, performance parameters, and statistical limitations of the Ear Model (BCDS Model No. 0001). The manual also provides:

* An overview of the background and origins of the BCDS project (Section 1) and the prior analytical work conducted by Mitre Corporation (MITRE) that gave rise to the project:
* A general description of the conceptual design and model development approach adopted by the BCDS project team, and its relation to MITRE’s previous work (Section 2);
* A description of the data requirements, structure, and segmentation used to develop the Ear Model, as well as a listing of the claim attributes selected for used to develop the model (Section 3);
* A description of the Ear Model performance parameters, and the development and optimization process used to satisfy those parameters (Section 4);
* A description Ear Model testing and evaluation process, and the statistical results of these processes (Section 5).

Sections 1 and 2 apply generally to all BCDS models, while Section 3 through 5 are specific to the Knee Model.

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

# Section 1: Background

## Statistical Adjudication of Disability Compensation Claims

Between 2013 and 2014, the Veterans Benefits Administration’s (VBA) Office of Compensation Service initiated a research project with the MITRE Corporation to determine the feasibility of using predictive models to adjudicate disability compensation claims with little or no human intervention. Conducted in a laboratory setting, this research project demonstrated that CDDs 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 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 substantial statistical basis or foundation for identifying valid meaningful correlations between variables and establishing predictive patterns to determine the calculated outcomes. Claims are broadly grouped into to one of two types: original – the first claim filed by a claimant for disability benefits; and 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 his/her original claim, but 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). In the one case, 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 is summarized below.

1. VA supplied MITRE (and later, the BCDS 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 next applied to quantify the relative importance of a suite of common claim attributes (or features) in 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 then 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 data set, 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 through-put (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 RVSR).

MITRE’s work established the relative importance of specific features, and identifies 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 are composed of the features identified as having significant predictive value – including 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 (his/her age at the time of filing, etc.), and finally, 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

## BCDS Model - Conceptual Design

The BCDS 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. MITRE identified 32 features (of an initial set of 249) found to be of greatest predictive value. The 32 features are presented in Table 1 below.

Table : 32 Feature Set Defined by MITRE Corp. for Ear Classification

|  |  |  |
| --- | --- | --- |
| **Feature** |  | **Feature** |
| Diagnostic Code 6100 | Days since last change in CDD |
| Prior CDD for Ear Issues | Days between the claim date and the most recent ear-related CDD begin date |
| Age of Claim | “Hearing” in Diagnosis Text |
| Date of Birth | “Tinnitus” in Diagnosis Text |
| Number of Ear Contentions | Diagnostic Code 6204 |
| Number of Contentions | Date the claim was filed |
| “Left” in Diagnosis Text | Contention Code 6850 |
| “Bilateral” in Contention Text | The presence of an ear diagnosis code in the narrative text. |
| Prior CDD of All Issues | Diagnostic Code 6260 |
| “Left” in Contention Text | “Loss” in Diagnosis Text |
| “Hearing” in Contention Text | “Tinnitus” in Contention Text |
| “Loss” in Contention Text | Diagnostic Code 6200 |
| “Right” in Contention Text | “6100” in Diagnosis Text |
| Contention Classification Code 3140 | Number of Rating Profiles |
| Begin Date for Most Recent Ear CDD | Rating Profile Date |
| “Right” in Diagnosis Text | Regional Office responsible for adjudicating claim |

The BCDS team’s evaluation suggested that a simpler approach that used a pattern-matching architecture might be adopted to streamline CDD calculation and provide enhanced scalability. This approach uses MITRE’s identified features as the underlying basis for calculating CDDs. The approach also retains the ability to apply OLR or other classifiers should such analytical capabilities become necessary to satisfy performance parameters. This hypothesis was based 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 2 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 run-time, a claimant’s adjudicatory history, as well as sufficient conditioned training data, all to derive a single CDD, seemed unnecessarily cumbersome and might 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 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 patterns that were determined to be invalid – something that 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 is premised on the fact 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 5 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 32 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 BCDS team tested its hypothesis by using the same data used by MITRE 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 (note – the calculated current CDD’s 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 from which the patterns were identified could be again 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%. Based on these initial results, the team adopted the pattern-matching architecture for preliminary development and testing purposes.

## Technical Approach

The BCDS team adopted a four phase technical approach to fully develop BCDS models consistent with the pattern matching architecture. This approach is illustrated in Figure 2 below.



Figure 2: BCDS Model Development Approach

During Phase 1, the team extracts and segments data to ensure sufficient claimants with relevant claimed contentions and adjudication decisions are included to form distinct development and test data sets. The team next synthesizes the data within the development data set to form a high-resolution feature set (i.e., values are calculated to the highest level of precision offered from the underlying data: in this case, time periods were calculated 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 are next 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-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 optimizes 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 tests the model against an entirely separate set of verification data to verify that the performance results initially obtained during optimization are replicated, and confidence levels are sustained. The government is also provided the model to afford its analysts an opportunity conduct independent analyses of the models.

## BCDS Pattern-Matching Architecture

The BCDS pattern-matching modeling architecture relies on replicated fact patterns for similar conditions and claim adjudicatory histories, and the consistent promulgation of CDDs for such fact patterns by VA employees to reliably predict CDDs (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 CDD of specified conditions 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 CDD that most often results.

This BCDS 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 – 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 claimants 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 BCDS system provides an automated environment in which authorized users select one or more target claims that are to be adjudicated, one or more models that are to be applied, 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 BCDS 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 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 users selection.

# Section 3: Data Requirements

## Data Source and Structure Overview

As stated previously, the data used to produce the BCDS Ear Model (002) was the same data provided to MITRE to support its analysis. This data base was provided by VBA’s Office Performance Analysis and Integrity (PA&I) on December 12, 2016. The data was composed of over 41 million claim 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 Ear Model Data Segmentation

The traditional method for developing predictive models is to segment data that will form the basis of the analysis; defining one subset for model development and reserving a second subset for testing and model analysis. 3,800,000 claims and associated data were extracted from the total data set. This data constituted the claims where the claimant’s adjudication history contained at least one claim containing an ear related contention (see the above stated criteria). Of these data, **2,742,885 claims** were segmented for model development and optimization, while the balance of the claims **(1,057,716)** were set aside for testing and refinement. Of the 2.7M development claims, **258,324** were found to be eligible for modeling based on the defined eligibility criteria. To be “Eligible” for BCDS auto-adjudication of ear conditions, claims must be coded as End Product (EP) 02X (where X may be 0 through 9), contain at least one Ear related contention (Contention Classification Codes 2200, 2210, 2220, 3140, 3150, 4130, 4210, 4700, 4920, 5000, 5010, 5710, or 6850), and at least one eligible ear related issue must have been found to be service connected and a rating promulgated within a previously adjudicated claim for benefits. This claim segmentation for both development and test claims is illustrated below in Figure 5.



Figure 5: Segmentation of Claims where claimant had filed at least one ear related claim

# Section 4: Model Development and Optimization

## Ear Model Performance Targets

## MITRE established that predictive models can replicate the CDD for ear related conditions determined by authorized VA employees **90 percent** of the time, and can do so for issues contained within **50 percent** of eligible claims. This performance parameter, stated as 50% throughput at 90% accuracy, is applicable to the BCDS Ear Model.

## Model Development Process

The team first calculated the feature-specific values for all claims within the defined set of development data. Resulting condition-specific CDDs are also calculated at this time. Data are then aggregated to identify all unique combinations of the defined features (or patterns) for eligible claims within the development data set. For the Initial Ear Model build, the number of unique patterns was slightly more than 78,144.

The team next assigns unique identifiers to each pattern and aligns the patterns with unique instances of resulting condition-specific CDDs. In some cases, the same pattern may result in different CDDs. In these cases, the number of occurrences for each different CDD is calculated. The most common CDD for the specific pattern is then assigned the CDD for the model. When each pattern is assigned a unique CDD (and the pattern-set or model is complete), the team applies the model to the data from which it was originally derived to calculate initial through-put, accuracy, and re-use rates.

## Model Optimization Process

The team next optimizes the model to determine the set of features that yields the highest throughput, accuracy, and pattern re-use rates. The more a fact pattern is found to be 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 conducts 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 narrows 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 (in the case of the ear, 50% and 90% respectively). This more limited set of preliminary models form the basis for testing and performance analysis.

## Testing and Performance Analysis

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 3.

Table : Ear Model Testing and Optimization Results



As indicated in Figure 5, accurately rated claims are attributable to a small volume of high-frequency pattern sets (41%).



Figure 5: Relationship between Throughput, Accuracy, and Pattern-count of the Ear Model

## Quantitative Characterization of Addressed Claims

Historically, **40** percent of eligible supplemental claims result in a change in the CDD for the ear. As indicated in Figure 6 below, **43.8** percent of these claims are accurately calculated by the model. Claimant data were insufficient to reliably calculate the CDD for those claims that were unaddressed. These claims were primarily characterized as being assigned the general hearing loss diagnostic code (6100), which alone or in combination with Tinnitus comprise over 94 percent of eligible claims.

The team also noted that Tinnitus, when diagnosed, received a CDD over 95% of the time. The impact of these findings on accuracy is illustrated in Figure 6.



Figure 6: Composition and accuracy of CDDs assigned to eligible Supplemental Claims.

The sequential number of the claim in the claimant’s filing history– a function of the age of the Ear CDD within the claimant’s adjudicatory history – is second only to the prior Ear CDD itself in accurately calculating the Ear CDD for the target claim. As indicated in Figure 7 below, the accuracy declines with each supplemental claim.



Figure 7: Accuracy Relative to Sequential Claim Order

Although the model’s accuracy remained relatively consistently above 90 percent, the percent of unaddressed claims declines over time from a high in 2014 of 49 percent to a low of 29% in 2005. See Figure 8 below.



Figure 8: Claims processed results

## Description and Quantitative Overview of Unaddressed Claims

The team subjected **289,092** eligible ear-related supplemental claims to testing and evaluation as part of model both model development and testing process (258,324 during development, and 30,768 during optimization and testing). **62,986** eligible claims were “unaddressed” by the model. These unaddressed claims are largely focused within claims for general hearing loss (alone or in combination with Tinnitus, as indicated earlier).

Based on this analysis, VACI and VBA are exploring options for developing a model for this diagnosis that uses a separate set of, as yet undefined, features.

# Section 5: Model Verification Testing

The final step in the development process is to apply the model to an entirely new set of claims to verify that performance levels can be replicated within acceptable tolerances, and where variations exist, explain such variance to the satisfaction of the government.

The team extracted a data set containing **562,927** claims for claimants who have filed at least one claim with an ear related issue from a new database provided by the government to verify the model. Computations to derive claim and claimant features were then run against the data and matched to the Ear Model DDM.

A specific data trigger was found by Compensation Services subject matter experts to have exaggerated the number of claims where the CDD was “severed” (and consequently reduced from some value to NULL in the data). Based on feedback from Compliance Services subject matter experts, the data trigger was eliminated, the model was recompiled using the original 3.8 million claim set, and finally subjected to verification testing. The final results are provided below.

Table : Verification Testing Results



## Verification Assessment

The Verification Test results indicated less than a 2% drop in accuracy to 94.38% and 15% drop in throughput. Significantly, high-frequency patterns (those used more than 5 times, with an average of 24.89 matches per pattern) deliver 48% throughput over the validation data set. Low frequency patterns account for the remaining 13% of the claims rated by the model.

# Appendix

## Acronyms

Benefits Claims Decision Support (BCDS)

Combined Disability Determination (CDD)

## Glossary