ReedGroup®

MDGuidelines® | The measure of health

DURATION BENCHMARKS METHODOLOGY

Updated 2021

ReedGroup MDGuidelines

Contents

INTRODUCTION

A significant body of scientific evidence has shown that return-to-work and/or activity are associated with health benefits to the patient (Brassil, 2013). Understanding the expected path and time frame for returning to normal lifestyle may help providers set recovery goals for returning patients to normal living. MDGuidelines provides users with different ways to compare a population's return-to-activity performance against benchmarks that are based on large data sets and physician guidance. In this document, we will present the methodology behind the MDGuidelines Physiological and Population Duration Benchmarks.

PHYSIOLOGICAL BENCHMARK

The **MDGuidelines Physiological Benchmark** provides recommended disability durations that represent the physiological healing time for uncomplicated cases (herein called "physiological durations"). Developed by the MDGuidelines Medical Advisory Board, the physiological durations are based on clinical expertise and informed by real world claims. These physiological durations do not represent the absolute minimum or maximum lengths of disability at which an individual must or should return to work. Rather, they represent important points in time at which, if recovery has not occurred, additional evaluation (and possible intervention) should take place.

PHYSIOLOGICAL DURATIONS DEVELOPMENT

MDGuidelines employs a two-step process in the development of the physiological duration tables. Using real-world case data and previously released physiological duration tables, the senior staff create statistical profiles that are reviewed and revised by a medical advisory board who apply their experience and research as a corrective, when necessary, to the statistical profiles. The evidence of the population data coupled with the consensus of expert medical practitioners provides an evidence-based, iterative process to create the physiological duration tables. This Modified Delphi approach combines the depth of MDGuidelines proprietary data with the breadth of expert medical judgment.

The first phase of the Modified Delphi approach involves a panel who flags and "corrects" durations that are skewed by factors such as selection bias. These "corrected" durations are subjected to the second phase of independent scrutiny. This scrutiny includes two levels of bias protection. First, a panel of experts must deliberate on the proposed ("corrected") durations—drawing solely upon their clinical experience and without recourse to the reference data. Thus, this group of experts does not merely replicate the steps established in the first phase. Instead, they approach the durations from another angle, with the result that any lingering discrepancies highlight the need further investigation. The second protection against bias occurs because this panel of experts operate independently of each other's input, insulating them from premature consensus.

The third phase requires a consolidation of professional opinions. The scrutinized and clinically modified durations are weighed against each other and against the reference data. This entire cycle is repeated when necessary. In this respect, duration guidelines follow the principles of evidence-based medicine: they result from clinical judgment and experience informed by statistical data, and they provide a baseline that is both humane and rigorous.

HOW TO INTERPRET PHYSIOLOGICAL DURATIONS

The physiological duration tables provide approximate return-to-activity timelines for injured or ill employees so that they can obtain the greatest health and productivity, according to physiological healing times. The physiological duration tables assume a) uncomplicated cases; and b) return to full duty.

Table 1. Example physiological duration table

JOB CLASS	MINIMUM	OPTIMUM	MAXIMUM
Sedentary	1	7	42
Light	3	14	42
Medium	14	21	56
Heavy	21	32	84
Very Heavy	28	48	91

The MDGuidelines Physiological Benchmark provides minimum, optimum, and maximum recovery time by job classifications. While "return to full duty" is assumed in the physiological duration tables for consistency, in many cases the injured individual may return to activity in a restricted capacity. When activity is restricted (i.e., return to work accommodations), the exertion level of the new job description should be followed in the physiological duration tables. For example, an employee may go out with a medium exertion level but be brought back to a sedentary desk position. Using Table 1 optimum durations as a further illustration, an employee with a medium job class could be brought back to work at seven days for a sedentary job, moved to a light duty job at 14 days, and finally brought to full duty at 21 days. Therefore, the physiological duration tables provide milestones for helping employees progress towards full duty.

Physiological duration tables are most useful when envisioned as a continuum in the case management process. These values do not represent the absolute minimum or maximum lengths of disability at which an individual must or should return to work. Rather, they represent important points in time at which, if full recovery has not occurred, additional evaluation should take place.

Users may find that some MDGuidelines physiological duration tables contain the term “indefinite”. This indication implies the potential for an indefinite disability. In these cases, it is possible that a return to work may not be compatible at the same activity level.

In many physiological duration tables, five job classifications are displayed. These job classifications are based on the amount of physical effort required to perform the work. The classifications correspond to the Strength Factor classifications described in the United States Department of Labor’s *Dictionary of Occupational Titles*. The Department of Labor job classifications focus on physical effort only. This may not be relevant to the duration of some disabilities as many factors go into the length of disability.

POPULATION BENCHMARK

The **MDGuidelines Population Benchmark** provides summary statistics on disability durations drawn from the MDGuidelines Population Database of real-world disability records. These statistics include the frequency of conditions, their average lengths (herein called “population durations”), and the probability of return to full duty. The statistics underlying the Population Benchmark represent the actual observed experience of individuals across the spectrum of physical conditions, in a variety of industries, and with varying levels of case management. The Population Benchmark also reflects various psycho-social factors (e.g., individual’s motivation and benefit structure) that may affect return to normal activity.

DATA SOURCES

The MDGuidelines Population Database includes more than seven million disability leave records for over 18,000 unique conditions (Figure 1) with information on length of time from date of absence to return to full duty, sex, age, job class (level of job exertion), and coexisting conditions. These records were provided to MDGuidelines from employers, insurers, healthcare providers, and government agencies. Both short-term disability and workers’ compensation records were used for the Population Benchmark. The Population Database records are primarily from the United States (95%) across all 50 states (Figure 2).

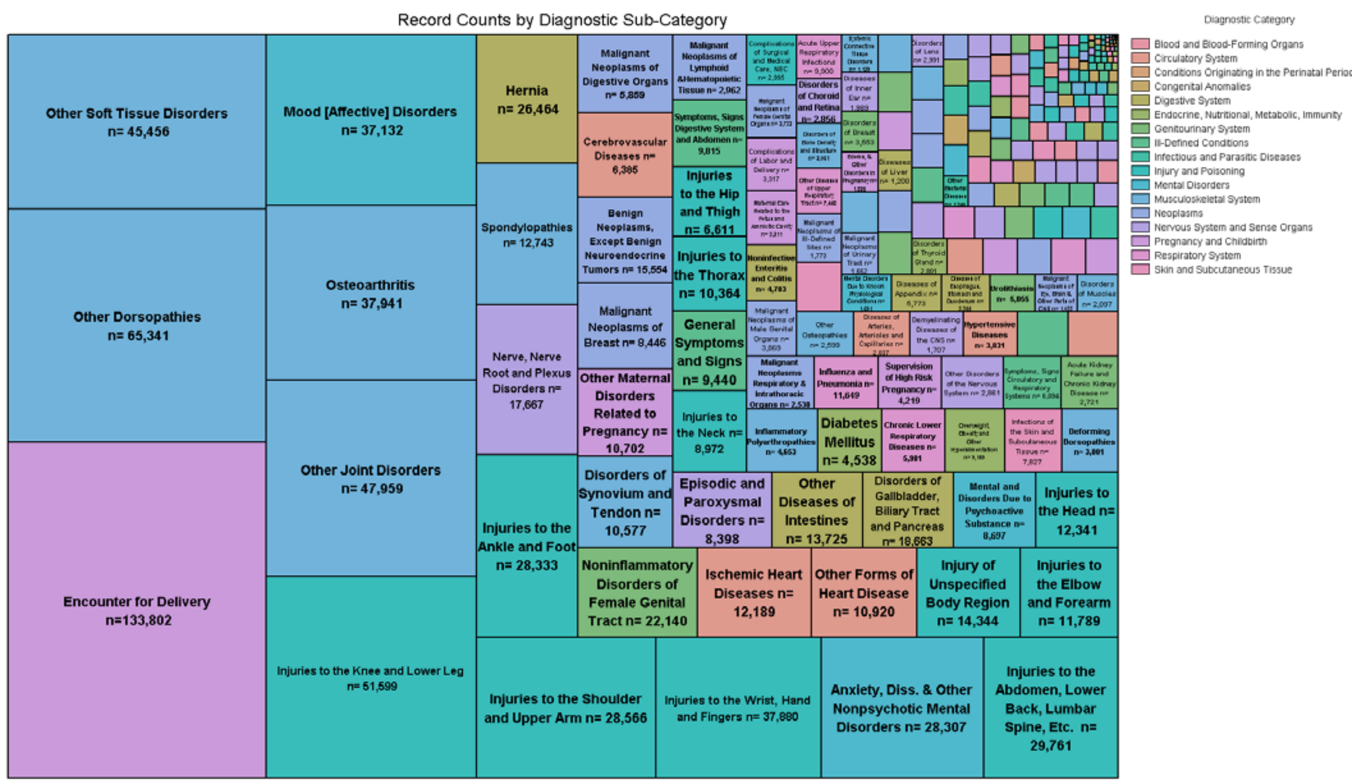

Figure 1. Breakdown of diagnostic sub-categories in Population Database. Colors indicate diagnostic category.

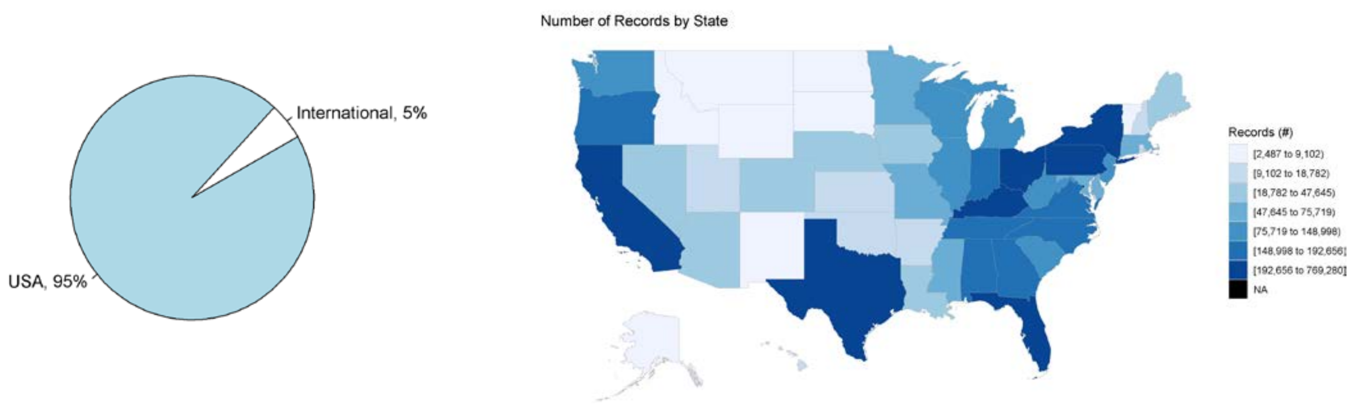

Figure 2. Geographic distribution of records in the MDGuidelines Population Database

DATA QUALITY ASSURANCE

Extensive data cleaning and validation is performed prior to using the data for any analysis. The following steps are part of the data quality assurance protocol:

1. Medical Code Validation Checks:
    a. Medical code is a valid, billable medical code
    b. Medical code corresponds to record's sex (e.g., remove records with obstetric diagnoses and male sex)
    c. Medical code corresponds to record's age (e.g., remove records with pediatric diagnoses)
2. Date Validation:
    a. A first absence date and a follow-up date. The follow-up date is typically the return to full duty date, but could also be the last date the record was tracked or the date the individual transferred from STD to LTD.
    b. Follow-up date is not before first absence date
    c. Follow-up date is not after receipt of data (e.g., return to work dates cannot be in the future)
3. Claim Demographics Validation:
    a. Perform previously mentioned medical code validation checks on record comorbidities
    b. Standardize variables across all data sets (i.e., all females mapped to "F" in sex column)

POPULATION BENCHMARK STATISTICS

Diagnoses recorded with ICD-9-CM codes and ICD-10-CM codes were both used in the Population Benchmark statistics. We mapped each claim with a diagnosis recorded using ICD-9-CM to all the applicable ICD-10-CMs using the Centers for Medicare and Medicaid general equivalency mapping (GEM) tables. If the GEM tables included a "choice list" variable, we only mapped an ICD-9-CM to ICD-10-CM if the choice list was either "0"or "1."

The specificity of ICD-10-CMs were meant for medical coding, not developing duration statistics. Therefore, we removed unwanted specificity using the following methods:

1. Collapse the “7th character” codes.
    a. These include codes that specify initial encounters, subsequent encounters, and sequela visits. Although there are many 7th character codes depending on the type of injury. For example, for fractures there’s a “subsequent encounter for fracture with routine healing” and “subsequent encounter for fracture with delayed healing”
    b. Example of collapse: S11.011 (Laceration without foreign body of larynx) contains S11.011A (initial encounter), S11.011D (subsequent encounter), and S11.011S (sequela).
2. Collapse laterality codes.
    a. Some codes specify whether the right or left extremity had the diagnosis. And some are marked “unspecified” or “bilateral.” Without knowing the claimant’s own laterality, these codes are less informative for disability duration.
    b. Example of collapse: G56.0 (Carpal tunnel syndrome) contains G54.00 (unspecified upper limb), G54.01 (right upper limb), G54.02 (left upper limb), and G54.03 (bilateral upper limbs)

Typical to return-to-work (RTW) data, the MDGuidelines Population Database contains records for individuals that do not have a date specifying when the individual returned to full duty, but do have a follow-up date after their first absence date noting they were still on disability. There may be multiple reasons for this including that the individual never returned to full duty because they transferred to LTD, dropped out of the workforce, or died. A missing full duty date may also be because of incomplete data. However, since we have partial information of the time an individual was absent from work up until a certain point in time (called “right-censored” date in statistical terms), we must use this partial information and account for those individuals where we do not have a full duty date. If we do not account for those without a full duty date, we would bias the data towards only the most straightforward cases, those that left on a disability and returned to full duty.

To create more accurate and complete Population Benchmark statistics, we included information from all available records. In instances where we do not know what happened at the end of a case (did not return to full duty, died) we used a statistical method to utilize that information without giving it the same weight as a complete record. This statistical method, called a Kaplan-Meier estimation of the survival curve, was applied to STD and WC cases together to calculate the following duration statistics:

Table 2. Example Population Benchmark statistics table

Medical Code	Case Frequency	Mean	Percentile					% Records Returning to Full Duty
			5th	25th	Median	75th	95th	
G56.01	High	62	13	32	56	112	602	87%

The definition of each statistic:

Case Frequency = a field that describes the number of records by condition in the Population Database.

Mean = the geometric mean of disability durations for the condition in the Population Database

5th %ile = the 5th percentile disability durations for the condition in the Population Database. For example, if there are 100 records for a medical code and the 5th percentile was 13 days, five out of 100 records would have a disability duration of 13 days or less.

25th %ile = the 25th percentile of disability durations for the condition in the Population Database. For example, if there are 100 records for a medical code and the 25th percentile was 32 days, 25 out of 100 records would have a disability duration of 32 days or less.

Median = the median or 50th percentile of disability durations for the condition in the Population Database. For example, if there are 100 records for a medical code and the median was 56 days, 50 out of 100 records would have a disability duration of 56 days or less.

75th %ile = the 75th percentile disability durations for the condition in the Population Database. For example, if there are 100 records for a medical code and the 75th percentile was 112 days, 75 out of 100 records would have a disability duration of 112 days or less.

95th %ile = the 95th percentile disability durations for the condition in the Population Database. For example, if there are 100 records for a medical code and the 95th percentile was 602 days, 95 out of 100 records would have a disability duration of 602 days or less.

% of Records Returning to Full Duty = The percentage of records that returned to full duty within the follow-up time (transferred to LTD from STD, dropped out of work force, etc.).

Note: If a population statistic indicates “Indefinite”, then the records at that percentile and above never returned to full activity and we cannot give a definitive duration. For example, if the 75th percentile indicates “Indefinite” then at least 25% of the records in the database for that condition did not return to full duty.

CASE-LEVEL COMPARISON

We developed predictive models using STD and WC records in the MDGuidelines Population Database to facilitate case-level comparison. We excluded external causes and visit codes as primary diagnoses in the models. Analysis was restricted to records from individuals 16 years and older. Finally, we removed all conditions where the disability duration was greater than two years. For disability durations with a zero duration, we randomly gave a duration between zero and one day (required when using a logarithmic model). After all exclusion criteria, the model used approximately seven million records.

POTENTIAL PREDICTORS

The following variables were tested for their ability to predict disability duration:

1. Age in years (16 to 99 inclusive)
2. Sex (binomial, 0 = male, 1 = female)
3. Job class as defined by U.S. Department of Labor's Dictionary of Occupational Titles. The job classes include "Sedentary", "Light", "Medium", "Heavy", and "Very Heavy" work (ordinal variables).
4. Program type (binomial, 0 = STD, 1 = WC)
5. Coexisting conditions. Coexisting conditions that fit within comorbidity groupings as defined by Quan et al. (2005) were grouped (binomial, 0/1) and the individual ICD-10-CM codes within the groupings were removed. Additional coexisting conditions that did not fit within the comorbidity groupings were used individually as binomial variables within the model. A co-morbidity was only considered if there were at least ten records for that condition or the comorbidity grouping.

Variables missing data in more than 25% of the records per model were removed as potential predictors. Missing data for predictors (<25% missing) was imputed using the observed variable distribution.

STATISTICAL METHODS

To create predictive models, we used survival models to account for the right-censored records (individuals that do not return to full duty) in the data. We leveraged information across sub-classes of ICD-10-CM codes by building a model for each sub-class with specific conditions represented by indicator variables (binomial- yes/no). For example, ICD codes related to venous embolism and thrombosis (ICD-10-CM codes starting with I82) were analyzed in a single model. As an illustration, say the disability records contain 75 records with Budd-Chiari syndrome (ICD-10-CM = I82.210) and 25 records of thrombophlebitis migrans (ICD-10-CM = I82.411), all 100 records would be used in the survival model with two indicator variables (binomial, 0/1) indicating whether the individual had ICD-10-CM = I82.210 or I82.411. Further, if the number of records in a particular sub-class were less than 40, we combined all the conditions within a diagnostic subcategory to build the model, also only using if more than 40 records. The advantage of grouping similar conditions together is that we have more statistical power to detect associations between demographic predictors (e.g., age, sex) and RTW durations. Individual indicator variables for the specific ICD-10-CM were also only included if at least 20 records were present. Finally, we did not build a predictive model if the return to work probability was less than 60%.

We used the least absolute shrinkage and selection operator method (Lasso) method with a Cox-Proportional Hazard kernel to determine the predictors of the prognostic model (Tibshirani, 1997). Using 10-fold cross-validation, the Lasso method penalizes the negative log of the partial likelihood across a range of values for a regularization parameter (lambda). The final model and selected predictors were chosen using the largest value of lambda such to minimize the error. This procedure was implemented using the *cv.glmnet* function from the glmnet package (Friedman, Hastie, & Tibshirani, 2010; Simon, Friedman, Hastie, & Tibshirani, 2011) using R version 3.3.1 (R Core Team, 2016). Figure 3 illustrates cross-validation and how Lasso picks significant predictors.

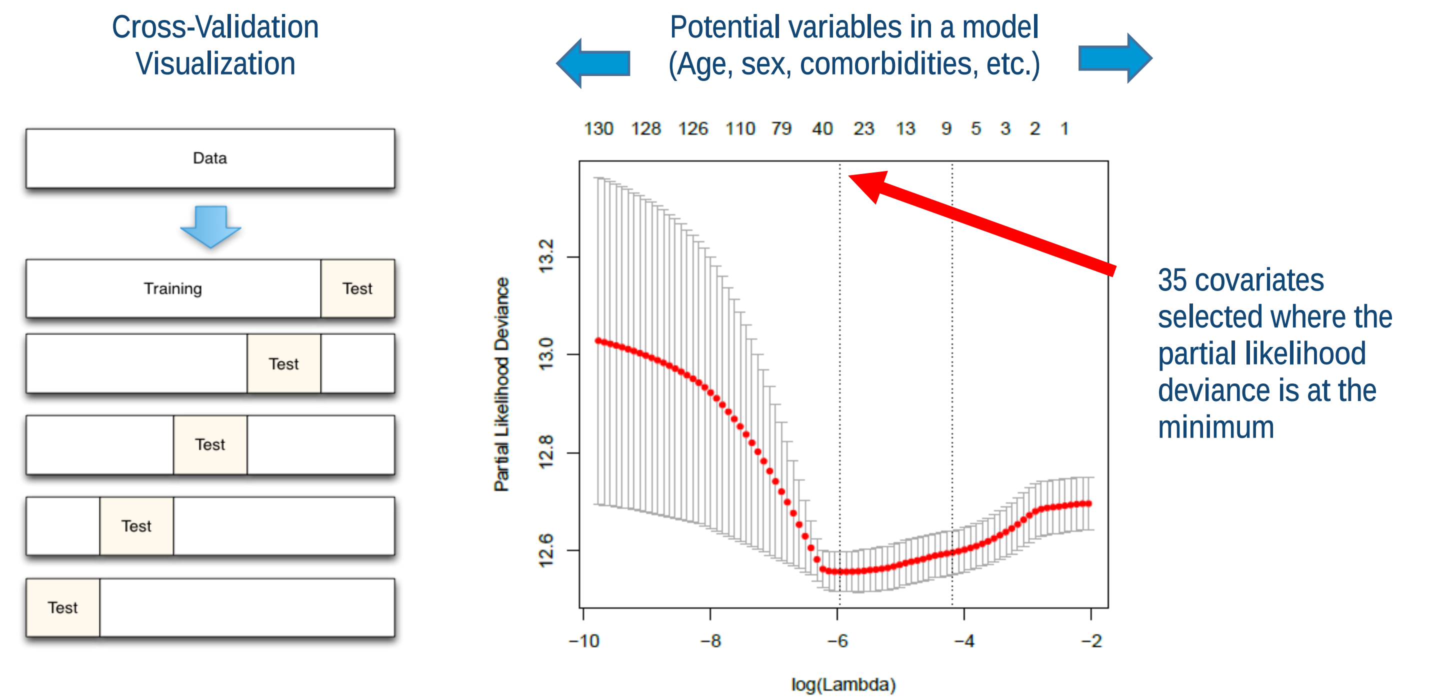

Figure 3. Example of Lasso cross-validation and variable selection. The left figure illustrates how the data set is repeatedly split into a training and test, where a model is built in the "training" set and the performance is checked in the "test" set. The right figure illustrates that cross-validation is applied across different combinations of variables and the final model is selected as the combination of variables that produces the minimum prediction error.

The population durations generally followed a log-normal distribution more closely than a gamma or exponential distribution; therefore, we input the significant predictors from the lasso procedure into a log-normal parametric survival model to predict case durations. To further optimize the models, we performed a backward stepwise regression procedure removing variables with a p-value > 0.2. Finally, if a comorbidity grouping or individual coexisting condition reduced the total predicted case duration (protective effect), we removed that condition assuming that coexisting conditions should not theoretically improve prognosis.

REFERENCES

Brassil, E. B. (2013). AMA Guides[TM] to the Evaluation of Work Ability and Return to Work (2nd ed.), edited by James B. Talmage, J. Mark Melhorn, and Mark H. Hyman. *Medical Reference Services Quarterly*, *32*(4), 476–478. http://doi.org/10.1080/02763869.2013.837746

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, *33*(1), 1–22. Retrieved from http://www.jstatsoft.org/v33/i01/

Quan, H., Sundararajan, V., Halfon, P., & Fong, A. (2005). Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 Administrative Data, *43*(11). Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/16224307

R Core Team. (2016). R: A Language and Environment for Statistical Computing. Vienna, Austria. Retrieved from https://www.r-project.org/

Simon, N., Friedman, J., Hastie, T., & Tibshirani, R. (2011). Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent. *Journal of Statistical Software*, *39*(5), 1–13. Retrieved from http://www.jstatsoft.org/v39/i05/

Tibshirani, R. (1997). The Lasso Method for Variable Selection in the Cox Model, *16*(March 1995), 385–395. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/9044528