Dec 18, 2020

Statistical Analysis Plan for Development of Injury Prediction Models for the MP3 Return to Duty Study

Dan Rhon[1,2], Tina Greenlee[1], Jeff Sorensen[3], Garrett Bullock[4,5]

[1]Department of Rehab Medicine, Brooke Army Medical Center, San Antonio, TX;
[2]Department of Rehab Medicine, Uniformed Services University of Health Sciences, Bethesda, MD; [3]NUBLU LLC, Salt Lake City, Utah;
[4]Centre for Sport, Exercise and Osteoarthritis Research Versus Arthritis, University of Oxford, UK;
[5]Nuffield Department of Orthopaedics, Rheumatology, and Musculoskeletal Sciences, University of Oxford, Oxford, UK

 dx.doi.org/10.17504/protocols.io.bihfkb3n

Dan Rhon

THIS PROTOCOL ACCOMPANIES THE FOLLOWING PUBLICATION

Developing predictive models for return to work using the Military Power, Performance and Prevention (MP3) musculoskeletal injury risk algorithm: a study protocol for an injury risk assessment programme. Inj Prev . 2018 Feb;24(1):81-88. doi: 10.1136/injuryprev-2016-042234.

DOI

dx.doi.org/10.17504/protocols.io.bihfkb3n

PROTOCOL CITATION

Dan Rhon, Tina Greenlee, Jeff Sorensen, Garrett Bullock 2020. Statistical Analysis Plan for Development of Injury Prediction Models for the MP3 Return to Duty Study. **protocols.io** https://dx.doi.org/10.17504/protocols.io.bihfkb3n

MANUSCRIPT CITATION please remember to cite the following publication along with this protocol

Developing predictive models for return to work using the Military Power, Performance and Prevention (MP3) musculoskeletal injury risk algorithm: a study protocol for an injury risk assessment programme. Inj Prev . 2018 Feb;24(1):81-88. doi: 10.1136/injuryprev-2016-042234.

CREATED

Jul 12, 2020

LAST MODIFIED

Dec 18, 2020

PROTOCOL INTEGER ID

39175

DISCLAIMER:

The view(s) expressed herein are those of the author(s) and do not reflect the official policy or position of Brooke Army Medical Center, the U.S. Army Medical Department, the U.S. Army Office of the Surgeon General, the Department of the Air Force, the Department of the Army, the Department of Defense or the U.S. Government.

1 ADMINISTRATIVE INFORMATION

The SAP below is written following the Guidelines for Content of Statistical Analysis Plans in Clinical Trials (adapted for the purposes of this cohort study) JAMA Reference: http://dx.doi.org/10.1001/jama.2017.18556

1.1 Title and Trial Registration

Predictive Models for Spine and Lower Extremity Injury After Discharge from Physical Rehabilitation: Statistical Analysis Plan

Clinicaltrials.gov #NCT02776930
Protocol Publication PMID: 27884941

Daniel Rhon (DR)
Tina Greenlee (TG)
Garrett Bullock (GB)
Jeff Sorensen (JS)

1.2 SAP Version

2.0

1.3 Protocol Version

1.0

1.4 SAP Revisions

A decision was made to adjust the analytic approach to align with best practices and advances in prediction modeling techniques. One key difference being the approach to identify potential predictors based on a univariate analysis approach, which has potentially many limitations that would introduce bias. At the time the original SAP was determined (2013), this was thought the best approach by the research team, and was also the approach used in previous studies by this team at the time. These revisions were made before any assessment or analysis of the data were completed (before the data scientists even received the final dataset for analysis), and this protocol was written and published prior to the start of any data analysis steps.

SAP in the original protocol initially approved by the IRB:
"Subjects will be dichotomized as injured or non-injured based on the injury surveillance data. Key demographic, physical performance (FMS, YBT, SFMA, Hop Test, MSRT, & Shuttle Run), and self-report measures will be examined for group differences. Independent t- tests will be used for continuous variables and chi-square tests for categorical variables. Variables with a significance level of $P \leq .10$ will be retained as potential predictors. This significance level was selected to increase the likelihood that no potential predictor variables would be overlooked. For continuous variables with a significant univariate relationship, sensitivity and specificity values will be calculated for all possible cutoff points and then plotted as a receiver operating characteristic (ROC) curve. The point on the curve nearest the upper left-hand corner will represent the value with the best diagnostic accuracy, and this point will be selected as the cutoff defining a positive test. Sensitivity, specificity, odds ratios, and positive likelihood ratios (LRs) will be calculated for potential predictor variables. Potential predictor variables will be entered into a backward stepwise logistic regression model to determine the most accurate set of variables predictive of musculoskeletal injury status. A significance level of .10 will be required for removal from the equation to minimize the likelihood of excluding potentially helpful variables. Variables retained in the regression model will comprise the computerized algorithm for predicting those Service Members that are likely to experience a musculoskeletal injury. The Hosmer-Lemeshow summary goodness-of-fit statistic will be used to assess the fit of the model to the data and test the hypothesis that the model fits the data.

We will further analyze the data to determine whether weighting individual predictors according to the relative size of the beta coefficients increases the prognostic accuracy of the model. Weights will be calculated by taking the beta coefficient for each variable in the final model and dividing it by the lowest beta coefficient and then rounding to the nearest integer. After the weight is formulated, an ROC curve will be used to identify the cutoff value that represented the best diagnostic accuracy. Sensitivity, specificity, and positive LRs, as well as corresponding 95% confidence intervals, will be calculated for the cutoff point that maximized the diagnostic utility of the weighting system.

Risk stratification (low, moderate, or high) will be based on LRs associated with the clinical prediction rule for injury outlined above. A positive LR > 10 will place the individual as high risk, a LR between 2 and 10 would place the individual as moderate risk. Those with a positive LR less than 2 will be listed as low risk."

The new statistical analysis plan is outlined below and was fully developed and published here before starting any of the data analysis steps.

1.5 Roles and Responsibilities

DR: Primary Investigator
TG: Data cleaning and data validation
GB and JS: Data scientists - responsible for finalizing SAP, agreement on best practices and statistical methods for primary and secondary aims and outcomes.

Introduction

2 INTRODUCTION

2.1 Background and Rationale

History of prior injury is one of the primary risk factors for musculoskeletal injury in Army service members, also commonly referred to as tactical athletes. Lower extremity and back injuries have stood out for their majority contributions to disability in the military. Deficits in movement quality following injury have been posited to mechanistically contribute to re-injury. The ability to identify which military tactical athletes are at higher risk to sustain a future injury is of high priority to military leaders and policymakers. Therefore, we chose to examine a group of service members who had sustained lower extremity or thoracolumbar injuries and had been cleared to return to full duty following those injuries. For the one year following their return to duty, we tracked new and recurrent injuries and related time off from work. Demographic, psychosocial, health, and physical movement measures captured at the time of discharge from medical care, at the beginning of that year, would then be analyzed to examine which combinations of variables would be able to predict future injury. This information can aid providers in care decisions when returning military tactical athletes to full duty following an injury.

This document outlines the statistical analysis plan for the study. It was fully developed and published here in its final form before any analytical work was begun on the models.

2.2 Objectives

The primary objective is to identify criteria that can help predict the likelihood of future injury or re-injury and fit those variables into a robust prediction model that can aid clinicians and policy-makers in assessing the risk of future musculoskeletal injury in service members.

Study Methods

3 STUDY METHODS

3.1 Trial Design:

N/A - This is not a clinical trial

Citation: Dan Rhon, Tina Greenlee, Jeff Sorensen, Garrett Bullock (12/18/2020). Statistical Analysis Plan for Development of Injury Prediction Models for the MP3 Return to Duty Study. https://dx.doi.org/10.17504/protocols.io.bihfkb3n

This is a cohort study designed to develop models to predict risk of musculoskeletal injury.

3.2 Randomization:

Not Applicable

3.3 Sample Size:

The a priori plan was to enroll 480 Service Members. With an expected injury rate of 20-50% and a study completion rate of 75%, we planned for 180 subjects (360 total) that successfully completed the study for each of the 2 target conditions (lower extremity and lumbar/thoracic spine injuries) at one year. This would allow for enough subjects to be enrolled to account for any problems with baseline data collection and over/under-sampling based on the unit-based data collection procedures used in this study. Based on recruitment from previous work, we anticipated a 75%-85% recruitment rate, and a 75% one-year completion rate (75% in Prevention of Low Back Pain in the Military [POLM - PMID: 23608562] trial and 85% in the initial MP3 study [PMID: 32134698]). Estimates from a healthy population estimate that approximately 20% of male and 40-50% of females Service Members will sustain at least one time-loss musculoskeletal injury over a one-year period. Since prior history of injury is one of the strongest predictors for re-injury, and all of our subjects will already have this variable present, we moved our conservative male injury rate from 20% to 30%, and expected to have successful completion of 135 subjects per specific cohort. We further estimated based on these numbers, that approximately 40 Service Members per cohort would sustain a time-loss injury, thus providing the potential for a robust 5-10 variable regression model per cohort. At the same time, in order to test our hypothesis based on sex, we planned to keep enrollment open to female subjects until reaching a minimum target of 100.

3.4 Framework:

Development of a multivariable prediction model

3.5 Statistical Interim Analysis and Stopping Guidance:

N/A - No interim analysis conducted and no guidance was provided for stopping as this was not an interventional study

3.6 Timing of Final Analysis

All outcomes will be analyzed collectively

3.7 Timing of Outcome Assessments

The outcome could have occurred at any time point during the 1-year surveillance period. Injury occurrence, type of injury, and the number of work-loss days were extracted from relevant electronic databases at the end of the entire study period (batch extraction), by utilizing the entire 1-year period of surveillance as the date/time parameters for data extraction. This occurred > 90 days after the last individual completed their 1-year surveillance period.

Statistical Principles

4 STATISTICAL PRINCIPLES

4.1 Confidence Intervals and P values:

Prediction models will be evaluated through calibration and discrimination. However, specific predictor significance will also be evaluated. Predictor 95% confidence intervals and P Values will be reported for statistical models to aid in interpretation. Overal prediction models will be reported with intercept, coefficients, and standard errors for future reproducibility. There are not multiple primary analyses; thus, adjustment for multiplicity is not needed; however, multiple sensitivity analyses will be performed

Citation: Dan Rhon, Tina Greenlee, Jeff Sorensen, Garrett Bullock (12/18/2020). Statistical Analysis Plan for Development of Injury Prediction Models for the MP3 Return to Duty Study. https://dx.doi.org/10.17504/protocols.io.bihfkb3n

to investigate the robustness of the results.

4.2 Adherence and Protocol Deviations

This was an observational study, not an intervention trial and therefore adherence to delivery of an intervention is not applicable.

4.3 Analysis Populations

The plan is to analyze all participants enrolled who met the criteria and where complete data was available (per missing data criteria established below).

Trial Population

5 TRIAL POPULATION

5.1 Screening Data:

Participants were recruited from a pool of patients that had just been cleared to return to full military duty without any restrictions. At the time of release to return to full duty, they were approached for participation in the study and screened for eligibility. The final report will provide details on the number of total subjects screened in reference to the total number enrolled, as well as the most common reasons for exclusion during screening.

5.2 Eligibility

Inclusion criteria

1. US active duty military service member.
2. Age 18–45 years (or emancipated minor).
3. A lower extremity or lumbar/thoracic spine injury was the primary reason the patient was seeking care.
4. Patient deemed ready to return to work without any limitations by their medical provider.

Exclusion criteria

1. Service members who plan on leaving the military in the following 12 months after enrolment in the study (separation or retirement from the military, or medical board), which will be the full period of injury surveillance.
2. A concomitant injury for which the patient is already seeking or planning to seek medical care.
3. Any type of restricted or modified work program due to a musculoskeletal injury; must be returning to work without any limitations.
4. Service members pending a Medical Evaluation Board.
5. Trauma or polytrauma that results in amputation of any limbs or appendages.
6. Injuries from high-velocity incidents, such as motor vehicle injury, etc.
7. Pregnancy, or recently pregnant within the last 6 months— subjects who become pregnant during the course of the study will be withdrawn based on the different injury risk factors that may be associated with musculoskeletal injury during pregnancy.

5.3 Recruitment

CONSORT flow chart of participants from baseline through completion of study will be present in the final report.

5.4 Withdrawal/Follow-up

The level of consent withdrawal will be tabulated (classified as “consent to continue follow-up and data collection” “consent to continue data collection only”, “complete – no further follow-up or data collection”).

Citation: Dan Rhon, Tina Greenlee, Jeff Sorensen, Garrett Bullock (12/18/2020). Statistical Analysis Plan for Development of Injury Prediction Models for the MP3 Return to Duty Study. https://dx.doi.org/10.17504/protocols.io.bihfkb3n

We will report rates and reason for withdrawals, as well as lost-to-follow-up rates and account for these individuals accordingly in the analysis (remove from final analysis). We will also remove participants found to not be eligible after prior healthcare utilization is pulled (were found to not be fully cleared for duty, and therefore not eligible for inclusion).

The timing of withdrawal will also be reported for participants that withdrew.

5.5 Baseline Patient Characteristics

Participants will be described with respect to age, sex, BMI, ethnicity, marital status, military rank, years spent in the military, smoking habits, activity level, and injury category. We will present overall characteristics, as well as between main groups (e.g. thoracolumbar spine injuries and lower extremity injuries).

Analysis

6 ANALYSIS

Participant statistics will be described using mean (standard deviation) for continuous normally distributed variables, median (interquartile range) for non-normally distributed continuous variables, and frequencies and percentages for categorical variables. Injury incidence will be calculated per 1000 athlete exposures.

Prior to model development, continuous variables will be assessed for non-linearity in relation to injury. Non-linearity will be assessed through restricted cubic splines at three, four, and five knots. Non-linearity criterion will be based on Akaike's Information Criteria (AIC), residuals, visual inspection, and biological plausibility.

6.1 Outcome Definitions:

Co-Primary Outcomes:
1) Sustained a time-loss injury during the 1-year period of surveillance - YES or NO (dichotomous variable)
2) Limited duty (work) days due to MSK injury during the 1-year period of surveillance (continuous variable)

A time-loss injury is defined as an injury resulting in at least 1 day of restricted military duty
The 1-year period of surveillance ran from the day of enrollment through the following 364 days (total of 365)

Secondary Outcomes:
- Time to first injury following return to full duty (continuous variable)
- MSK-related medical costs during the 1-year period of surveillance (continuous variable)
- Injury location - upper extremity, lower extremity or spine (categorical variable)

6.2 Analysis Methods:

Primary Model Development

Citation: Dan Rhon, Tina Greenlee, Jeff Sorensen, Garrett Bullock (12/18/2020). Statistical Analysis Plan for Development of Injury Prediction Models for the MP3 Return to Duty Study. https://dx.doi.org/10.17504/protocols.io.bihfkb3n

The criteria for transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) were followed for all model development. A prediction model using logistic regression will be performed, with the occurrence of a time-loss injury as the binary outcome. Following primary model development, an elastic net analysis will be performed. Elastic net is a penalized method that assesses multicollinearity and shrinks the coefficients (down to zero) to reduce the risk of overfitting. Further, elastic net allows for predictor selection and incorporates these results into the overall model, and has been found to have improved results compared to more traditional statistical methods.To find the optimal alpha and lambda shrinkage parameters, tenfold cross-validation with ten iterations per fold will be performed, with the greatest Kappa used to determine the best tuning parameters. Models will then be internally validated to decrease optimism through bootstrapping with 2000 iterations or 10 fold cross-validation with 10 iterations per fold. Sensitivity analyses will include: logistic regression for participants without pain, logistic regression for spine and lower extremity cohorts separately, logistic regressions for injury severity, and survival analysis in relation to time to injury.

Model Performance

Model performance will be investigated by assessing Nagelkerke R-squared, calibration, and discrimination. Calibration measures agreement between predicted risks from the model and what was observed, while discrimination evaluates how well the model differentiates between those with and without the outcome. Calibration will be quantified by the calibration slope (with 95% confidence intervals) and graphically by plotting the predicted risk against the observed outcome using a loess smoother. Discrimination will be evaluated by the Area Under Receiver Operating Characteristic Curve (AUC). An AUC of 0.5 equates to correctly predicting 50% of outcomes, which is no better than random guessing. An AUC of 1.0 demonstrates perfect (100%) outcome prediction.

6.3 Statistical Methods – adjustment for covariates

All predictors will be considered for creating the most precise and accurate prediction models. Predictors will be chosen through clinical reasoning, review of the literature, and collinearity assessment.

6.4 Statistical Methods – subgroup analyses

As this is not a clinical trial, the standard approach to subgroup analysis will not apply here. In other sections, we have discussed our intent to account for differences in the models based on sex and injury type (both original injury type and future injury type).

6.5 Missing Data:

All data will be investigated for missingness prior to analyses. Missing data will be categorized as missing completely at random, missing at random, or missing not at random. If missing data is below 5%, a complete case analysis will be performed. If missing data is above 5%, multiple imputation will be performed. Imputation iterations will be at a minimum of 20 iterations, and will proportionally increase with the percentage of greatest missing data (e.g. 30% missing data will entail 30 imputations). Missing outcomes will be marked prior to imputation, with outcomes imputed. However, imputed outcomes will be excluded from analyses. It should be noted if data is categorized as missing not at random, multiple imputation will not be performed on these data, and will be potentially excluded from analyses.

6.6 Additional Analyses:

Citation: Dan Rhon, Tina Greenlee, Jeff Sorensen, Garrett Bullock (12/18/2020). Statistical Analysis Plan for Development of Injury Prediction Models for the MP3 Return to Duty Study. https://dx.doi.org/10.17504/protocols.io.bihfkb3n

Machine Learning
Following primary model development, internal validation, and assessing model performance, four machine learning models (Random Forest, Support Vector Machine Regression, Gradient Boosting Machine, and Artificial Neural Networks) will be developed to predict injury. All machine learning models will incorporate all the same predictors used to develop the logistic regression model. An iterative matrix process will be performed to assess optimal tuning hyperparameters for all machine learning models, with root mean square error utilized to determine optimal tuning. All models will then be internally validated to reduce overfitting with 10 fold cross-validation. Machine learning models will also be used to determine the importance and weight of predictors to help inform clinical practice and decision making.

Decision Curves
Decision curve analyses will be performed to determine whether incorporating the prediction model into clinical practice improves military outcomes. The net benefit is the fraction of true positives gained by making decisions based on predictions over a range of plausible risk thresholds. Within this study, the threshold probability will be defined as the population risk of injury within military personnel. As injury risk can vary between baseball players, the net benefit was calculated through a range of predicted risks. The net benefit of the prediction model will then be compared through decision curve analyses to assume all are at high risk ('treat all') and assume all are at low risk ('treat none'). Further, the prediction model's net benefit will be compared to current military evidence-based practice guidelines.

6.7 **Harms:**

N/A - this was an observational cohort; not a randomized clinical trial

6.8 **Statistical Software:**

The computer software programs R and Python will be used for all analyses.

6.9 **References**

1. Balabin RM, Lomakina EI. Support vector machine regression (SVR/LS-SVM)—an alternative to neural networks (ANN) for analytical chemistry? Comparison of nonlinear methods on near infrared (NIR) spectroscopy data. *Analyst.* 2011;136(8):1703-1712. PMID: 21350755
2. Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. *Statistics in medicine.* 2016;35(2):214-226. PMID: 26553135
3. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Bmj.* 2015;350:g7594. PMID: 25569120
4. Durrleman S, Simon R. Flexible regression models with cubic splines. *Statistics in medicine.* 1989;8(5):551-561. PMID: 2657958
5. Friedman JH. Greedy function approximation: a gradient boosting machine. *Annals of statistics.* 2001:1189-1232.
6. Hassoun MH. *Fundamentals of artificial neural networks*: MIT press; 1995.
7. Jain AK, Mao J, Mohiuddin KM. Artificial neural networks: A tutorial. *Computer.* 1996;29(3):31-44.
8. Liaw A, Wiener M. Classification and regression by randomForest. *R news.* 2002;2(3):18-22.
9. Philp F, Al-shallawi A, Kyriacou T, Blana D, Pandyan A. Improving predictor selection for injury modelling methods in male footballers. *BMJ open sport & exercise medicine.* 2020;6(1). PMID: 32095267
10. Ridgeway G. Generalized Boosted Models: A guide to the gbm package. *Update.* 2007;1(1):2007.
11. Royston P, Sauerbrei W. *Multivariable model-building: a pragmatic approach to regression anaylsis based on fractional polynomials for modelling continuous variables.* Vol 777: John Wiley & Sons; 2008.
12. Ruder S. An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747.* 2016.
13. Vapnik V. *The nature of statistical learning theory*: Springer science & business media; 2013.
14. Vickers AJ, Elkin EB, Steyerberg E. Net reclassification improvement and decision theory. *Statistics*

in medicine 2009;28(3):525-26. PMID: 17907248
15. Vickers AJ. Decision analysis for the evaluation of diagnostic tests, prediction models, and molecular markers. *The American Statistician* 2008;62(4):314-20. PMID: 19132141
16. Zou H, Hastie T. Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology).* 2005;67(2):301-320.
17. Fulton J, Wright K, Kelly M, Zebrosky B, Zanis M, Drvol C, Butler R. Injury risk is altered by previous injury: a systematic review of the literature and presentation of causative neuromuscular factors. Int J Sports Phys Ther. 2014 Oct;9(5):583-95. PMID: 25328821; PMCID: PMC4196323.
18. Jones, Bruce H., Veronique D. Hauschild, and Michelle Canham-Chervak. 2018. “Musculoskeletal Training Injury Prevention in the U.S. Army: Evolution of the Science and the Public Health Approach.” *Journal of Science and Medicine in Sport / Sports Medicine Australia* 21 (11): 1139–46. PMID: 29602720
19. Molloy, Joseph M., Timothy L. Pendergrass, Ian E. Lee, Michelle C. Chervak, Keith G. Hauret, and Daniel I. Rhon. 2020a. “Musculoskeletal Injuries and United States Army Readiness Part I: Overview of Injuries and Their Strategic Impact.” *Military Medicine*, March. https://doi.org/10.1093/milmed/usaa027. PMID: 32175566
20. Molloy, Joseph M., Timothy L. Pendergrass, Ian E. Lee, Keith G. Hauret, Michelle C. Chervak, and Daniel I. Rhon. 2020. “Musculoskeletal Injuries and United States Army Readiness. Part II: Management Challenges and Risk Mitigation Initiatives.” *Military Medicine*, February. https://doi.org/10.1093/milmed/usaa028. PMID: 32107561
21. Protocol Publication (Open Access): Rhon, Daniel I., Deydre S. Teyhen, Scott W. Shaffer, Stephen L. Goffar, Kyle Kiesel, and Phil P. Plisky. 2016. “Developing Predictive Models for Return to Work Using the Military Power, Performance and Prevention (MP3) Musculoskeletal Injury Risk Algorithm: A Study Protocol for an Injury Risk Assessment Programme.” *Injury Prevention: Journal of the International Society for Child and Adolescent Injury Prevention*, November. https://doi.org/10.1136/injuryprev-2016-042234. PMID: 27884941
22. Teyhen, Deydre S., Scott W. Shaffer, Robert J. Butler, Stephen L. Goffar, Kyle B. Kiesel, Daniel I. Rhon, Jared N. Williamson, and Phillip J. Plisky. 2015. “What Risk Factors Are Associated With Musculoskeletal Injury in US Army Rangers? A Prospective Prognostic Study.” *Clinical Orthopaedics and Related Research* 473 (9): 2948–58. PMID: 26013150
23. Teyhen, Deydre S., Scott W. Shaffer, Stephen L. Goffar, Kyle Kiesel, Robert J. Butler, Daniel I. Rhon, and Phillip J. Plisky. 2020. “Identification of Risk Factors Prospectively Associated With Musculoskeletal Injury in a Warrior Athlete Population.” *Sports Health*, March, 1941738120902991. PMID: 32134698

Citation: Dan Rhon, Tina Greenlee, Jeff Sorensen, Garrett Bullock (12/18/2020). Statistical Analysis Plan for Development of Injury Prediction Models for the MP3 Return to Duty Study. https://dx.doi.org/10.17504/protocols.io.bihfkb3n