

Evaluating Machine Learning Techniques for Early PTSD prognosis after Trauma using synthetic data.

Author:

Thomas J Wise

Student Number:

6664202

Programme:

Methodology and Statistics for the Behavioural, Biomedical and Social Sciences

Supervisors:

Dr. Mirjam van Zuiden

Prof. dr. Rens van de Schoot

Proposed Journal of Publication:

Multivariate Behavioural Research (MBR): Taylor and Francis

Ethics Approval

Status Accepted: #20-0039; Date: 09/10/2020

Status Accepted: #20-0140; Date: 27/10/2020

Introduction

Post-Traumatic Stress Disorder (PTSD) develops from exposure to traumatic events (accidents, mass violence, war or natural disasters), and is characterized through the re-experiencing of the event, increased physiological arousal, event associated stimuli avoidance, negative alterations in mood and cognition, culminating in a significant impairment to daily life (Association & others, 2013; Augsburger & Galatzer-Levy, 2020). Symptom progression of PTSD can be categorized into four distinct trajectories; resilient, recovered, chronic and delayed (Galatzer-Levy, Huang, & Bonanno, 2018; Schoot et al., 2018). Given the significant impairment resulting from PTSD, it is paramount that clinicians are able to deliver effective interventions. One proposed method to improve this, is through the application of trajectory specific interventions. In order to deliver this however, a validated prognostic screening instrument is required, which can appropriately classify a patient into one of these trajectories.

The proposed application of such a prognostic screening instrument is widespread. Especially when considering that the global lifetime exposure to any traumatic event is 70.4% (Benjet et al., 2016), and the lifetime prevalence of PTSD being 11% for women and 5.5% for men (Kessler, Sonnega, Bromet, Hughes, & Nelson, 1995), there is a widespread need for such an instrument. In particular, this project is directly aimed towards the trajectory prediction of those who have recently been exposed to a traumatic event, through its use within emergency departments. However, to develop this prognostic screening instrument, first the identification and development of an appropriate statistical model is required.

To develop this model, Machine learning (ML) techniques, have the greatest potential, due to their capacity to identify complex underlying patterns within data. In particular this has been demonstrated through prognostic research in physical (Kononenko, 2001; Kourou, Exarchos, Exarchos, Karamouzis, & Fotiadis, 2015; Rajkomar, Dean, & Kohane, 2019, p. @cruz2006applications) and psychiatric (Passos et al., 2016; Webb et al., 2020) medicine, with previous research also displaying its efficacy within PTSD prognosis research (Galatzer-Levy, Karstoft, Statnikov, & Shalev, 2014; Ramos-Lima, Waikamp, Antonelli-Salgado, Passos, & Freitas, 2020). Although this multidisciplinary support exists; current research into PTSD prognosis is limited, particularly in technique comparison across different sample sizes. For PTSD this is extremely important, with ML models being typically developed using large data sets, which are classically unattainable within PTSD research. To date, the largest proposed data set, is not set to exceed 5,000 participants (McLean et al., 2020). Even when considering international collaboration efforts, the noted constitutional and sociocultural factors involved in the development of PTSD (Yehuda, 2004) places limitations on the sample sizes which can be collected and analysed effectively.

Therefore, to overcome the challenges presented regarding the limited available sample size, alongside the additional statistical issues through the application of multiple comparative statistical tests upon one data set, the use of simulated, synthetic data is proposed. The use of simulated data has been demonstrated as appropriate across the medical research field, for the usage of developing and testing prognostic and statistical models, during times when the real data cannot solely be used (Burton, Altman, Royston, & Holder, 2006, pp. @ambler2002simplifying, @morris2019using). Although understanding that the use of simulated data is required in understanding which ML technique is most appropriate, the way in which data is simulated is also of importance. Since, if the simulated data is not able to sufficiently mimic the behaviour of the original data set, then the models developed will have limited generalisability, beyond that of the data simulated, limiting the conclusions which can be drawn. Therefore, it is important to first explore which simulation method is most appropriate, between what is classified as “conventional wisdom” by Skron dal (2000), and the proposed counter through meta-models, both with and without accounting for missing data through multiple imputation. Through comparing these three simulation methods, further support can be provided for the recommendations made regarding the most appropriate ML technique, for the development of this prognostic screening instrument.

Within this intermediary report, further illumination will be provided into the planned methodological and analytical plans which can be summarized through figure 1. In particular, each section of the figure will be expanded, with key discussion points outlined.

INSERT FLOW DIAGRAM 1

Methodology

Original Data

The data, which will provide both the foundation for the planned simulations and the data which the recommended ML technique will be applied to; is the TraumaTips data set (Mouthaan et al., 2014). Collected between September 2005 to March 2009, this data set contains a total of 852 patients each who experienced a traumatic injury. Each participant in the study was requested to complete a series of PTSD self-report instruments, referred to as Index Tests (SPAN (Meltzer-Brody, Churchill, & Davidson, 1999; Zlotnick, Davidson, Shea, & Pearlstein, 1996), IES-R (Marmer & Weiss, 1997), TSQ (Foa, Riggs, Dancu, & Rothbaum, 1993)) alongside a semi-structured interview of the CAPS, based upon DSM-IV criteria (Blake et al., 1995). These measures were then taken at 5 time points (23 days (mean), 1 month, 3 months, 6 months and 12 months), after admission into the study. This data set was chosen given its reflection of the wider population of PTSD patients. . . .

For the purpose of this study, since an understanding of PTSD symptoms progression is required, only those participants who participated at two or more time points were included within the study, reducing the total population to 522 participants. Additionally to this, other data cleaning methods were applied, these can be divided into the reclassification of variables, removal of *Non-stable questionnaires* (those which were collected with hindsight (self-report but with retrospect)) and time based-grouping.

Reclassification of Variables One of the first methods applied was the classification of string variables into categorical groups, allowing a more effective be utilization in the simulation methods and to identify their role in PTSD trajectory. In particular, reclassification was applied to:

- Sport participation before their traumatic event. Reclassified using (???) classification of sports according to their static and dynamic components.
- Employment or Profession at time of traumatic event. Reclassified using Centraal Bureau voor de Statistiek (CBS) employment classification (???)
- Psychotropic medication usage at time of traumatic event, or historically. Classification of medication in line with their general pharmaceutical classification.
- Personal or family history of psychiatric conditions. Reclassified using major DSM-V (Association & others, 2013) themes.
- Relation of family members with noted psychiatric conditions. Reclassified under the broad categories of first, second or third degree relation
- Others

Removal of Non-Stable Questionnaires Removal of those variables/questionnaires which were determined with hindsight (thus invalidating them) EXPAND UPON THIS

Time Based-Grouping Finally participants were regrouped according to specific time windows. This decision was based upon the idea that, if it is believed symptom progression has a temporal basis, then it is important to group together those who were recorded within specific temporal periods. For this study, as the major and desired temporal periods were, 23 days, 1 month, 3 months, 6 months and 12 months. It is proposed that the following time period (T) groupings are formed: 0-30 days (T1, <1 month), 31-90 days (T2, 1-3 months), 91-180 days (T3, 3-6 months), 181-365 (T4, 6-12months) and beyond 365 days (T5, 12 < months). The use of these groups will illustrate where PTSD symptom progression is influenced by temporal factors.

Data Simulation

Through three different data simulation methods, it is possible to determine which is most appropriate for this research, in addition to which has the greatest potential for wider application in this clinical research field. As previously mentioned, the first method will demonstrate the *conventional wisdom* (Skrondal, 2000), wherein variables are considered as independent units, and thus simulated in isolation. This will be completed through firstly determining which probability distribution best describes the variables, **through comparing key indicators such as skewness and kurtosis, EXPLAIN IN MORE DETAIL**. Before simulating it using the specific parameters from the original data provided. **Provide an example??**.

By contrast, the second and third model will be based upon meta-models (Skrondal, 2000), which assumes observations are the product of functions between variables. An example being that a participants IES-R Score, could be predicted by a participants age, gender, traumatic incident type and severity of accident. These interconnected functions will then be able to simulate the data constructing a more appropriate simulation of an individual observation.

In addition to the use of a meta-model within the third model, multiple imputation will be applied to the cleaned data. Since a large amount of data will still be missing, therefore to understand whether more complete data will better mimic the original, multiple imputation through the R package MICE (Buuren & Groothuis-Oudshoorn, 2010) will be applied.

Overall, not only will this highlight which simulation method is appropriate within this clinical field, but it will also further replicate the previous work in the field of meta-modeling and data simulation, applying it to this novel data set.

Comparative Testing: Data Simulation Methods

To conclude which simulation method is most appropriate for simulating the original TraumaTips data set, a series of comparative tests will be completed. And used to determine which data set both mimics and behaves similarity during statistical testing as the original. It is important to note that these tests will only be used as a proof of principle, validating the conclusions made and therefore will regard only the general outcome, with no optimization or interpretations made as a result, only comparisons of similarity will be made between the original and simulated data sets.

When considering the influence of bias, each data set will be simulated 100 times, to reduce the potential influence. Additionally, the simulated data sets will not only match the size of the original data set, but also be generated with a much smaller (50 observations) and much larger (10,000 observations) sample than the original. With the intention, that this range of tested samples will further support that the applied simulation methods are sufficiently scalable, and that they are correctly deployed when simulating both large and small data sets.

The planned comparative tests can be divided into two types: descriptive statistical tests and applied statistical tests. The first, will allow the comparison between the overall distribution of variables, whereas the applied tests will confirm the similarities in behaviour of the tests when ML techniques are applied.

Descriptive Statistical Tests To examine whether the simulated data sets are similar in their distributions and results, descriptive statistical tests will be applied to a sample of the variables within the data sets. This will entail, take a subset of 10% of the variables from the original data set and comparing their core descriptive statistics (mean, sd, skewness. . . etc) of their simulated counterparts. [Maybe consider more??]

Applied Statistical Tests To examine how data behaves when statistical tests are applied to them, this will be done through actively applying a ML technique (such as a decision tree, regression or SVM) to the original data, before applying the same model, using matching parameters to the simulated data sets in turn. From this, the core measures of performance (sensitivity, specificity and AUC) will be compared.

Those simulated data sets most similar to the original, will have performance measure scores which are similar (when averaged across the multiple iterations and sample sizes) to the original.

Since this method, may introduce bias from the ML technique applied, as one technique may be better fitted between the models, multiple different techniques will be tested. For this research, at least two techniques will be applied, a decision tree and a SVM, these will be applied as decision trees present a simple approach to classification, whereas there is a large amount of literature supporting the use of SVMs in PTSD prognosis.

From these applied tests, it would be expected to output a test like that displayed in figure 2.

INSERT MOCK UP CHART HERE

Reduction of Dimensionality

As the aim of the wider project, is to create a screening instrument for those entering psychiatric care shortly after a traumatic event, it is important to reduce the dimensions of the selected simulated data. Through this reduction, only those variables which will be available to the clinical professionals will be considered as variables of interest, for developing the ML model. At this time, it is proposed that these variables of interest will be those easily self-reported without clinical interventions, such as the previously highlighted index tests (IES-R, SPAN, TSQ), core demographic information (age, sex, incident type). Alongside these variables of interest, it will also be important to retain the trajectory classification, alongside other outcome variables which will be used in the development, training and testing of the ML model.

Although it could be proposed that this reduction of dimensionality, after data simulation will be computationally intensive (as more data will be required to be simulated). It should be noted that though ensuring these planned removed variables are included in the simulation process, will help to provide a more comprehensive simulation, in particular for those developed using a meta-model.

Application of Machine Learning Techniques

To make sufficient recommendations regarding which ML technique should be used on the original TraumaTips data set, before a screening instrument is developed. A variety of different techniques which have been demonstrated as successfully across PTSD prognosis research, as well as in both psychiatric and wider physical medical research. The examined techniques can be divided into classification or regression techniques, based upon their outcome measure. The planned techniques, their R packages as well as additional notes can be summarized in Table 1.

TABLE 1

In practical terms, these techniques will be examined both in their non-optimized and optimized forms, in order to illustrate the methodological process behind how the results were gathered. This is particularly important to ensure that the process and conclusions drawn can be understood by those looking to expand upon this research, and develop the screening instrument hereafter. Furthermore, similar to the procedure outlined for comparing simulated data sets, these ML techniques will be applied across a variety of sample sizes (between 50-10,000) in order to examine how key performance indicators (detailed in section: Performance Metrics) are impacted by the sample size. Additionally, this process will be repeated up to 1,000 prior to formal publication, with at least 100 iterations being completed before publication of the final report. The use of such a large amount of iterations is important to reduce random effects and therefore the influence of bias on the prediction model.

GENERAL TABLE CONTENTS

Proposed Techniques (general)

- SVM: e1071
- Regression: functions: glm(), lm() (Additional functions should be considered)

- K-nearest Neighbour (KNN): class
- Decision Trees:
 - Decision Trees (Recursive Partitioning): tree
 - Random Forests: randomForest / ranger
 - XGBoosted Trees: xgboost
- Neural Networks: neuralnet ??
- Naive Bayes (//) -

Additional Notes:

Ensure to illustrate my understanding that each method is likely to generate a different “type” of outcome.
 - For example, SVM = Classification (SVM) or Regression (SVR), Decision Trees/Random Forests, can be either/or

Outcome Results

1. Detail/Review the proposed outcome results.
2. Highlight individually the adv/disadv of each
 - 2a. Class only, + Typically used in the field - Limited on cat only (no wiggle room)
 - 2b. Class from Probability + includes wiggle room, useful for clinical practice - must consider specific thresholds, for example 33% vs 32% is this significantly different
 - 2c. Probability only + Wiggle room, advanced analysis, - Difficult clinical use, may be challenging to determine outcome results
 - 2d. Final Distribution of variables, + good for seeing overall trends, - may not reflect true nature of the individual(s)
- 3.

(Evenly split distribution - may not matter individually - adaptive/mal)

Performance Metrics

Outline what metrics will be used, and why they will be used

Sensitivity / Specificity / AUC - All used frequently within Psychiatric/PTSD ML Research Detail how these will be interpreted / what results or outcomes we will be specifically looking for.

Discuss thresholds of interest - “we hope to see... below/above specific threshold”.

Results

Discussion

References

- Ambler, G., Brady, A. R., & Royston, P. (2002). Simplifying a prognostic model: A simulation study based on clinical data. *Statistics in Medicine*, 21(24), 3803–3822.
- Association, A. P., & others. (2013). *Diagnostic and statistical manual of mental disorders (dsm-5)*. American Psychiatric Pub.
- Augsburger, M., & Galatzer-Levy, I. R. (2020). Utilization of machine learning to test the impact of cognitive processing and emotion recognition on the development of ptsd following trauma exposure. *BMC Psychiatry*, 20(1), 1–11.
- Benjet, C., Bromet, E., Karam, E., Kessler, R., McLaughlin, K., Ruscio, A., . . . others. (2016). The epidemiology of traumatic event exposure worldwide: Results from the world mental health survey consortium. *Psychological Medicine*, 46(2), 327–343.
- Blake, D. D., Weathers, F. W., Nagy, L. M., Kaloupek, D. G., Gusman, F. D., Charney, D. S., & Keane, T. M. (1995). The development of a clinician-administered ptsd scale. *Journal of Traumatic Stress*, 8(1), 75–90.
- Burton, A., Altman, D. G., Royston, P., & Holder, R. L. (2006). The design of simulation studies in medical statistics. *Statistics in Medicine*, 25(24), 4279–4292.
- Buuren, S. van, & Groothuis-Oudshoorn, K. (2010). Mice: Multivariate imputation by chained equations in r. *Journal of Statistical Software*, 1–68.
- Cruz, J. A., & Wishart, D. S. (2006). Applications of machine learning in cancer prediction and prognosis. *Cancer Informatics*, 2, 117693510600200030.
- Foa, E. B., Riggs, D. S., Dancu, C. V., & Rothbaum, B. O. (1993). Reliability and validity of a brief instrument for assessing post-traumatic stress disorder. *Journal of Traumatic Stress*, 6(4), 459–473.
- Galatzer-Levy, I. R., Huang, S. H., & Bonanno, G. A. (2018). Trajectories of resilience and dysfunction following potential trauma: A review and statistical evaluation. *Clinical Psychology Review*, 63, 41–55.
- Galatzer-Levy, I. R., Karstoft, K.-I., Statnikov, A., & Shalev, A. Y. (2014). Quantitative forecasting of ptsd from early trauma responses: A machine learning application. *Journal of Psychiatric Research*, 59, 68–76.
- Kessler, R. C., Sonnega, A., Bromet, E., Hughes, M., & Nelson, C. B. (1995). Posttraumatic stress disorder in the national comorbidity survey. *Archives of General Psychiatry*, 52(12), 1048–1060.
- Kononenko, I. (2001). Machine learning for medical diagnosis: History, state of the art and perspective. *Artificial Intelligence in Medicine*, 23(1), 89–109.
- Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. *Computational and Structural Biotechnology Journal*, 13, 8–17.
- Marmer, C., & Weiss, D. (1997). The impact of event scale–revised. *Assessing Psychological Trauma and PTSD*. Guilford, New York, 399–411.
- McLean, S. A., Ressler, K., Koenen, K. C., Neylan, T., Germine, L., Jovanovic, T., . . . others. (2020). The aurora study: A longitudinal, multimodal library of brain biology and function after traumatic stress exposure. *Molecular Psychiatry*, 25(2), 283–296.
- Meltzer-Brody, S., Churchill, E., & Davidson, J. R. (1999). Derivation of the span, a brief diagnostic screening test for post-traumatic stress disorder. *Psychiatry Research*, 88(1), 63–70.
- Morris, T. P., White, I. R., & Crowther, M. J. (2019). Using simulation studies to evaluate statistical methods. *Statistics in Medicine*, 38(11), 2074–2102.
- Mouthaan, J., Sijbrandij, M., Luitse, J. S., Goslings, J. C., Gersons, B. P., & Olf, M. (2014). The role of acute cortisol and dheas in predicting acute and chronic ptsd symptoms. *Psychoneuroendocrinology*, 45, 179–186.

- Passos, I. C., Mwangi, B., Cao, B., Hamilton, J. E., Wu, M.-J., Zhang, X. Y., ... others. (2016). Identifying a clinical signature of suicidality among patients with mood disorders: A pilot study using a machine learning approach. *Journal of Affective Disorders*, 193, 109–116.
- Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. *New England Journal of Medicine*, 380(14), 1347–1358.
- Ramos-Lima, L. F., Waikamp, V., Antonelli-Salgado, T., Passos, I. C., & Freitas, L. H. M. (2020). The use of machine learning techniques in trauma-related disorders: A systematic review. *Journal of Psychiatric Research*, 121, 159–172.
- Schoot, R. van de, Sijbrandij, M., Depaoli, S., Winter, S. D., Olf, M., & Van Loey, N. E. (2018). Bayesian ptsd-trajectory analysis with informed priors based on a systematic literature search and expert elicitation. *Multivariate Behavioral Research*, 53(2), 267–291.
- Skrondal, A. (2000). Design and analysis of monte carlo experiments: Attacking the conventional wisdom. *Multivariate Behavioral Research*, 35(2), 137–167.
- Webb, C. A., Cohen, Z. D., Beard, C., Forgeard, M., Peckham, A. D., & Björgvinsson, T. (2020). Personalized prognostic prediction of treatment outcome for depressed patients in a naturalistic psychiatric hospital setting: A comparison of machine learning approaches. *Journal of Consulting and Clinical Psychology*, 88(1), 25.
- Yehuda, R. (2004). Risk and resilience in posttraumatic stress disorder. *The Journal of Clinical Psychiatry*.
- Zlotnick, C., Davidson, J., Shea, M. T., & Pearlstein, T. (1996). Validation of the davidson trauma scale in a sample of survivors of childhood sexual abuse. *Journal of Nervous and Mental Disease*.