Journal of Psychiatric Research 96 (2018) 15–22

Contents lists available at ScienceDirect

Journal of Psychiatric Research

journal homepage: www.elsevier.com/locate/psychires

Predicting posttraumatic stress disorder following a natural disaster

Anthony J. Rosellini [a], [b], *, Francisca Dussaillant [c], José R. Zubizarreta [a], [d], Ronald C. Kessler [a], Sherri Rose [a]

[a] Department of Health Care Policy, Harvard Medical School, Boston, MA, USA
[b] Center for Anxiety and Related Disorders, Boston University, Boston, MA, USA
[c] School of Government, University of Desarrollo, Santiago, Chile
[d] Division of Decision, Risk and Operations, and Department of Statistics, Columbia University, New York, NY, USA

ARTICLE INFO

Article history:
Received 1 February 2017
Received in revised form
23 July 2017
Accepted 7 September 2017

Keywords:
Posttraumatic stress disorder
Machine learning
Natural disaster
Earthquake
Risk score

ABSTRACT

Earthquakes are a common and deadly natural disaster, with roughly one-quarter of survivors subsequently developing posttraumatic stress disorder (PTSD). Despite progress identifying risk factors, limited research has examined how to combine variables into an optimized post-earthquake PTSD prediction tool that could be used to triage survivors to mental health services. The current study developed a post-earthquake PTSD risk score using machine learning methods designed to optimize prediction. The data were from a two-wave survey of Chileans exposed to the 8.8 magnitude earthquake that occurred in February 2010. Respondents (n = 23,907) were interviewed roughly three months prior to and again three months after the earthquake. Probable post-earthquake PTSD was assessed using the Davidson Trauma Scale. We applied super learning, an ensembling machine learning method, to develop the PTSD risk score from 67 risk factors that could be assessed within one week of earthquake occurrence. The super learner algorithm had better cross-validated performance than the 39 individual algorithms from which it was developed, including conventional logistic regression. The super learner also had a better area under the receiver operating characteristic curve (0.79) than existing post-disaster PTSD risk tools. Individuals in the top 5%, 10%, and 20% of the predicted risk distribution accounted for 17.5%, 32.2%, and 51.4% of all probable cases of PTSD, respectively. In addition to developing a risk score that could be implemented in the near future, these results more broadly support the utility of super learning to develop optimized prediction functions for mental health outcomes.

Earthquakes are a common and deadly natural disaster that can result in both ground shaking and tsunami waves. According to the Center for Research on the Epidemiology of Disasters (2016), earthquakes affected roughly one hundred million people and resulted in over 700,000 deaths worldwide between 2000 and 2015. Although earthquake exposure has been associated with several adverse psychosocial consequences (e.g., depression; suicidality), posttraumatic stress disorder (PTSD) is typically found to be the most prevalent negative mental health outcome (North, 2014). A recent meta-analysis of 46 studies of earthquake survivors found an overall post-earthquake PTSD incidence of 23.7% (Dai et al., 2016).

Although a large literature has identified risk factors associated with post-earthquake (including post-tsunami) PTSD (Cairo et al., 2010; Chen et al., 2014; Cheng et al., 2014; Dell'Osso et al., 2013; Kun et al., 2009, 2013; Lai et al., 2004; Priebe et al., 2009; Rosendal et al., 2014; Sattler et al., 2014; Tural et al., 2004; van Griensven et al., 2006; Wang et al., 2009, 2011; Wen et al., 2012; Zhang et al., 2011), few studies have examined how to combine risk factor information into a risk score that can be used to predict who is most likely to develop post-earthquake PTSD. The development and use of clinical tools to identify individuals at high risk of PTSD is consistent with the American Red Cross PsySTART program (Schreiber et al., 2014). In PsySTART, aid workers meet with survivors in the immediate aftermath of a natural disaster to complete a risk factor checklist. Decisions about triaging survivors to mental health interventions are determined based on the total number of 13 risk factors that are present (American Red Cross, 2012). Given that regression coefficients vary widely across risk

---

* Corresponding author. Center for Anxiety and Related Disorders, Department of Psychological and Brain Sciences, Boston University, 648 Beacon Street, 6th Floor, Boston, MA 02215, USA.

E-mail address: ajrosellini@gmail.com (A.J. Rosellini).

https://doi.org/10.1016/j.jpsychires.2017.09.010

factors for post-earthquake PTSD (Cheng et al., 2014; Tural et al., 2004; Zhang et al., 2011), assuming all predictors equally contribute to PTSD risk may not result in optimal prediction. The one existing study to develop a regression-based risk score for post-earthquake PTSD used rounded main terms coefficients (no interactions) from a logistic regression of 11 risk factors (Liu et al., 2012). However, it is unclear if maximum prediction accuracy was achieved given evidence of interactions among predictors of post-earthquake PTSD (Dell'Osso et al., 2013; Fan et al., 2011).

Another way to develop a post-earthquake PTSD risk score would be through machine learning methods designed to optimize prediction. Machine learning has been used to develop risk scores for PTSD onset related to other types of traumatic events (Galatzer-Levy et al., 2014, 2017; Karstoft et al., 2015a; Karstoft et al., 2015b). A number of popular machine learning algorithms are described in Table 1. There are several reasons why machine learning algorithms might outperform standard parametric method. In comparison to conventional main terms regression examining the direct effect of predictors on an outcome, for example, there are machine learning algorithms available that automatize identification of interactions and non-linearities (e.g., multivariate adaptive regression splines, Friedman, 1991; random forests, Breiman, 2001; Bayesian trees, Chipman et al., 2010). In addition, whereas a conventional regression based on highly correlated independent variables (e.g., injury and amputation, Liu, et al., 2012) might have good prediction accuracy in the sample which it was developed but perform poorly in independent samples (model overfit), machine learning methods can be employed to reduce the likelihood of overestimating prediction performance. For example, penalized regression algorithms (i.e., *regularization*) prevent overfit by shrinking coefficients among collinear variables (Friedman et al., 2010).

Ensembling methods refer to a type of machine learning in which multiple algorithms are consolidated into a single algorithm with improved prediction performance. Super learning (van der Laan et al., 2007; van der Laan and Rose, 2011) is an ensembling method that is particularly well suited to develop risk scores (see Rose, 2013 for a mortality risk score example) because of its flexibility in generating a consolidated algorithm from a number of different prediction approaches. In other words, a super learner algorithm is able to simultaneously (i) capture the relationships of predictors with an outcome (e.g., using conventional regression), even if predictors are highly correlated (e.g., using penalized regression), and (ii) detect interactions and nonlinear associations (e.g., using decision-tree or spline algorithms).

Super learning has been used in one study to develop a risk score for PTSD related to any type of traumatic event (Kessler et al., 2014). In this study, the super learner algorithm outperformed a select number of individual algorithms (including logistic regression) in predicting PTSD based on several hundred risk factors. However, that study was limited as the data were cross-sectional, relied on retrospective reports of PTSD symptoms and risk factors, and a small proportion of the sample had disaster-related PTSD. Accordingly, the goal of the current study was twofold: (i) to demonstrate how machine learning methods can be used to develop a more accurate post-earthquake PTSD risk score than conventional regression methods, and (ii) to develop a preliminary model-based risk score for post-earthquake PTSD that could be expanded or adapted in future epidemiological disaster research.

1. Materials and methods

1.1. Sample

The data came from a two-wave household survey of individuals living in Chile at the time of the 8.8 magnitude earthquake occurring on February 27, 2010. The survey was conducted by Chile's Ministry of Planning and Cooperation (Division Observatorio Social, 2010) using fully-structured face-to-face interviews. The pre-earthquake survey was a biennial nationally representative socioeconomic-health survey conducted between November and December 2009. In order to understand the public health impact of the earthquake, including the tsunami that occurred in some coastal areas, a subsample of baseline respondents were re-interviewed between May and June 2010. Of the 27,000 households asked to participate in the post-earthquake survey, 22,456 agreed (response rate = 83.2%). Additional details of the survey design are available elsewhere (Division Observatorio Social and AGCI Ministerio de Relaciones Exteriores, 2015). The sample used here consisted of all 23,907 adults who participated in both surveys and completed the post-earthquake PTSD assessment.

1.2. Outcome measure

Probable DSM-IV PTSD was assessed in the post-earthquake survey using the Davidson Trauma Scale (DTS; Davidson et al., 1997). The DTS was administered in Spanish to assess past-week PTSD symptoms specifically in relation to the earthquake and tsunami. The DTS assesses the frequency and severity of all 17 DSM-IV PTSD symptoms using a 0–4 scale (total score 0–136). The reliability and factor structure of the DTS has been supported among Chileans exposed to the 2010 earthquake (Leiva-Bianchi and Araneda, 2013). Comparison of the English DTS to independent structured interview-based PTSD diagnosis suggests that a DTS total score ≥40 indicates a probable PTSD diagnosis (Davidson et al., 1997). PTSD research conducted in Spanish speaking countries suggests that the ≥40 cut score demonstrates good concurrent validity with a PTSD diagnosis using the Spanish Clinician Administered PTSD Scale ($\kappa = 0.78$; Coronas et al., 2008). Several studies of Spanish-speaking samples have applied this cut score (e.g., Leiva-Bianchi and Araneda, 2013; Ruiz-Parraga and Lopez-Martinez, 2014). In the current sample, 9.9% of respondents lived in areas where the earthquake could not be felt (or no tsunami waves). We identified cases of probable PTSD based on (i) living in an area affected by the disaster (conservative confirmation of PTSD Criterion A1), and (ii) having a DTS total score ≥40 (13.3% had probable PTSD, $n = 3182$).

1.3. Independent variables

We reviewed two areas of literature to identify risk factors for PTSD among adults: (i) studies of risk factors specifically relevant to post-earthquake PTSD (cited in the introduction), and (ii) systematic reviews of risk factors for PTSD related to *any* type of natural disaster (Goldmann and Galea, 2014; Norris et al., 2002) and *any* type of traumatic event (Ozer et al., 2003; Sayed et al., 2015). Consistent with the literature, we organized risk factors into pre-earthquake factors (present *before* the trauma), peri-earthquake factors (objective and subjective experiences-severity immediately surrounding the trauma), and post-earthquake factors (present *after* the trauma).

We identified all survey questions that could be used to operationalize risk factors identified in the literature review. As the purpose of the surveys was not to study the complete range of all previously identified PTSD risk factors, especially peri-earthquake factors, we supplemented the survey data with other publically available data about the severity-impact of the earthquake (described below). The goal of the analysis was to optimize prediction of probable PTSD, not to test a conceptual model. We consequently operationalized as many risk factors as possible, regardless of if they had been directly (e.g., sex, age, property damage, job loss, Kun et al., 2013; Tural et al., 2004; Wang et al., 2011; Wen et al., 2012; Zhang et al., 2011) or indirectly associated with PTSD (e.g., individuals living in areas with stronger ground shaking more likely to experience fear or injury, which is associated with increased PTSD risk, Tural et al., 2004; Wang et al., 2009; Zhang et al., 2011). In total, 67 variables were created (see Supplemental Table 1 for complete descriptions-coding).

Pre-earthquake. Pre-earthquake variables included (i) dichotomous indicators of sex, marital status, and living situation, (ii) several dichotomous and categorical employment and education-specific variables (e.g., employment status; years of school), (iii) categorical variables representing the condition of the walls and roof of the home (rated by the interviewer), and (iv) health variables, including a categorical overall health rating and several dichotomous variables representing 30-day health and long-term health problems.

Peri-earthquake. Earthquake severity was operationalized using peak ground acceleration (PGA) data obtained from the United States Geological Survey (USGS). PGA is an *objective* measurement of ground shaking intensity in a given geographic area. A continuous PGA value was assigned to each respondent, based on the municipality[1] where they lived, using an inverse distance weighted average of the three closest USGS PGA grid estimates (range = 0.0–32.0, Wald et al., 1999; see Zubizarreta et al., 2013 for additional details). We also used USGS data on the date, location (municipality), and intensity of aftershocks. The USGS assesses aftershocks using the “did you feel it” method (Atkinson and Wald, 2007; United States Geological Survey, 2016a), a *subjective* rating of shaking intensity coded according to the Modified Mercalli Scale (range = 2.0 to 7.9) (United States Geological Survey, 2016b). We used aftershock data to create categorical variables representing the frequency and intensity of aftershocks in the week following the earthquake.

Tsunami severity was operationalized using data on wave height and horizontal wave inundation stored in the National Geophysical Data Center's Global Historical Tsunami Database (National Geophysical Data Center, 2016). Continuous variables were defined from these data to represent maximum wave height (range = 0.0–29.0 m) and deepest horizontal inundation (range = 0.0–1032.3 m) in each of the 65 coastal municipalities in Chile. We also defined death rate due to the earthquake in the municipality where the respondent was living. This was a continuous variable (range = 0.00–0.36/10,000 residents) coded using data on number of disaster-related deaths obtained from the Statistics Unit at the Chilean Forensic Services Department (López and Insunza, 2013) and municipality population at the time of the earthquake (Instituto Nacional de Estadísticas, 2015).

In addition, select questions from the post-earthquake survey were used to define additional indicators of disaster severity. Specifically, questions rated by the interviewer were used to define categorical variables reflecting the post-disaster condition of the home; extent of damage to the home, and estimated cost of damage to the home.

Post-earthquake. Given the goal of developing a PTSD risk score that could (eventually) be used in the immediate aftermath of an earthquake (consistent with PsySTART), we restricted the post-earthquake predictors to those that would be feasible to assess within roughly one week of earthquake occurrence. A limited number of questions from the post-earthquake survey were used to create three dichotomous variables representing whether the respondent was permanently displaced, had an educational disruption, or had a job disruption.

1.4. Analysis method

All analyses were conducted in R (R Core Team, 2015). Super learning (van der Laan et al., 2007; van der Laan and Rose, 2011), implemented using the SuperLearner package (Polley et al., 2016), was used to generate the prediction function for probable post-disaster PTSD. The goal of super learning is to develop a consolidated algorithm with optimized mean squared error (MSE). Briefly, super learning is implemented by identifying a user-specified “library” of algorithms, implementing each algorithm in the library using 10-fold cross-validation (CV) to generate predicted values, regressing the outcome onto the CV *predicted values* from each algorithm in the library (to determine the best weighted combination of the individual algorithms), and fitting each algorithm in the library on the full dataset and combining these fits with the weights generated in the prior step (i.e., to estimate the super learner predicted values; see Supplemental Table 2 for additional details).

Algorithms. The library included 39 algorithms in total; 13 algorithms each implemented using three different sets of predictors. Based on recent recommendations (LeDell et al., 2016; Rose, 2013), and because it was unknown which algorithm would result in optimized prediction, we selected several different types: logistic regression (Hosmer et al., 2013), five penalized regressions (for an overview of penalized regression methods see Friedman et al., 2010), two spline regressions (adaptive splines, see Friedman, 1991; adaptive polynomial splines, see Stone et al., 1997); two decision tree methods (random forests, see Breiman, 2001; Bayesian additive regression trees, see Chipman et al., 2010), and three support vector machines (linear, polynomial, and radial kernals, see Steinwart and Christmmann, 2008). Detailed descriptions of the algorithms and their R packages can be found in Table 1 (see also Supplemental Table 3).

The 13 algorithms were implemented using the full set of independent variables as well as two smaller variable sets because (i) we wanted to determine how crucial it would be to assess all 67 independent variables, and (ii) it was possible that use of fewer independent variables might improve performance of some algorithms (e.g., if there was collinearity among the full set of variables). We consequently applied super learning screening methods (built into the R package) to define two restricted predictor sets. First, consistent with how Liu et al. (2012) selected variables for their logistic regression risk score, we used *screen.ttest* to identify the 10 independent variables with the strongest bivariate associations with probable PTSD based a rank-order of univariate *t*-test p-values (“*t*-test set”). Second, we used *screen.glmnet* to identify the unknown number of non-collinear (i.e., non-redundant) independent variables that would *not* have their coefficients shrunk to zero in a lasso regression (“lasso set”).

Evaluating algorithm performance. To evaluate and compare the performance of the super learner to the 39 individual algorithms, we used the CV predicted values from each algorithm to conduct receiver operating characteristic curve analysis in 1000 bootstrap replicates to estimate area under the curve (AUC) and its 95% confidence interval (CI) (Robin et al., 2011). AUC is a measure of overall prediction accuracy, with values 0.50–0.70 generally interpreted as indicating poor prediction; 0.70–0.79 as acceptable; 0.80–0.89 as excellent; and >0.90 as outstanding (Hosmer et al., 2013). We calculated AUCs to four decimals to increase precision of comparisons (AUCs reported in text should be multiplied by 10^{-2}). To test if the AUC achieved by the super learner was

---

[1] Chile's territorial organization consists of municipalities, provinces and regions. Municipalities are the smallest territorial units (346 total municipalities; 203 which were affected by the earthquake); provinces group several municipalities (from 2 to 32); and regions group several provinces (from 2 to 8). Chile had 15 regions at the time of the earthquake.

Table 1
Brief descriptions of common machine learning algorithms.

Algorithm	R package	Description
Conventional Logistic regression	*stats*	• Traditional parametric logistic regression • Prone to overfit if independent variables are highly collinear • Optimal functional form of independent variables unknown (e.g., linear versus non-linear)
Regularization (Friedman et al., 2010) Ridge Elastic net Lasso	*glmnet*	• Penalized regression reduces overfit due to collinear independent variables • Ridge regression shrinks coefficients for collinear independent variables *toward* zero, but does not fully-eliminate any independent variable • Elastic net regression allows various penalties where coefficients for collinear independent variables are shrunk *toward* zero (but *not* eliminating contributions to the predicted probability) and/or *to* zero (eliminating their contributions to the predicted probability) • Mixing parameter penalty (alpha) is set somewhere between 0.01 and 0.99. Three elastic net algorithms were examined here (mixing parameter penalty set to 0.25, 0.50, and 0.75) • Lasso regression shrinks coefficients for collinear covariates to zero, eliminating their contributions to the predicted probability
Spline Adaptive splines (Friedman, 1991) Adaptive polynomial splines (Stone et al., 1997)	 *earth* *polspline*	• Adaptive spline regression flexibly captures interactions and linear and non-linear associations • Linear segments (splines) of varying slopes are connected and smoothed to create piece-wise curves (basis functions) • Final fit is built using a stepwise procedure that selects the optimal combination of basis functions • Earth and polymars are generally similar, but differ in the order which basis functions (e.g., linear versus nonlinear) are added to build the final model
Decision tree Random forest (Breiman, 2001) Bayesian additive regression trees (Chipman et al., 2010)	 *randomFroest* *BayesTree*	• Decision tree methods capture interactions and non-linear associations • Independent variables are partitioned (based on values), stacked to build decision trees, and ensembled into an aggregate "forest" • Random forests builds numerous trees in bootstrapped samples and generates an aggregate tree by averaging across trees (reducing overfit) • Bayesian trees are based on an underlying probability model (priors) for the structure and likelihood for data in terminal nodes; the aggregate tree is generated by averaging across tree posteriors (reducing overfit)
Support vector machines (Steinwart &Christmmann, 2008) Linear kernel Polynomial kernel Radial kernel	*e1401*	• Support vector machines treats each independent variable as a dimensions in high dimensional space and attempts to identify the best hyperplane to separate the sample into classes (e.g., cases and non-cases) • Goal is to find the hyperplane with the maximum margin between the two closest points in space • Captures linear associations, but alternate kernels can be used to capture nonlinearities (polynomial and radial basis kernels were used here)

statistically superior to the individual algorithms, the Hanley and McNeil method (Hanley and McNeil, 1983) was implemented in 1000 bootstrap replicates (Robin et al., 2011).

AUC is a measure of overall prediction accuracy; it does not provide information about prediction accuracy specifically among individuals who might be identified as high-risk and offered a preventive intervention. If a PTSD risk score were used after a natural disaster, a cut-point would need to be applied to identify exactly who should receive preventive intervention. However, the ideal cut-point(s) for identifying high-risk individuals would vary depending on post-disaster resources and the cost-effectiveness of available preventive interventions. Accordingly, risk scores are often used to create tiers of risk (e.g., Kim et al., 2015; Yeoh et al., 2011). As optimal cut-points were unknown, we evaluated the operating characteristics for a variety of different risk tiers. We created 20 risk tiers by rank ordering the predicted values from the best performing algorithm (hypothesized to be the super learner) and dividing respondents into 20 equal groups. We evaluated algorithm performance by calculating operating characteristics within each risk tier. We prioritized sensitivity (the proportion of all true positive cases of probable PTSD among individuals classified as high-risk) and positive predictive value (PPV; among individuals classified as high-risk, the proportion who ultimately developed probable PTSD) over specificity and negative predictive value because the consequence of not offering prevention to a true positive case of PTSD is greater than the consequence of unnecessarily offering prevention to a false positive. Following this approach allowed us to compare our results to Bromet et al. (2017), who created a preliminary logistic regression risk score for PTSD related to *any* type of disaster and also evaluated sensitivity and PPV across 20 risk tiers.

2. Results

2.1. Super learner versus logistic regression

The distributions of all 67 independent variables along with their bivariate associations (odds ratios) with probable PTSD are presented in Supplemental Tables 4-6. Nearly 80% of the 67 variables had significant associations with probable PTSD. The super learner algorithm had a lower MSE (9.94×10^{-1}) than conventional logistic regression using the *t*-test set (10 best variables, $\text{MSE} = 10.31 \times 10^{-2}$), lasso set (49 non-collinear variables, $\text{MSE} = 10.25 \times 10^{-2}$), and all 67 independent variables ($\text{MSE} = 10.25 \times 10^{-2}$) (Table 2). The super learner also achieved significantly better AUC (79.04×10^{-2}, 95% CI = 78.27–79.81) than the three specifications of conventional logistic regression (*t*-test set $\text{AUC} = 76.51 \times 10^{-2}$, 95% CI = 75.70–77.33, lasso set $\text{AUC} = 77.24 \times 10^{-2}$, 95% CI = 76.45–78.03, full set $\text{AUC} = 77.22 \times 10^{-2}$, 95% CI = 76.43–78.02) according to the bootstrap method of comparing paired AUCs ($Ds = 8.9–11.5$, $ps < 0.001$).

2.2. Super learner versus the other algorithm

The super learner algorithm also had a lower MSE than the other 36 individual algorithms. Bayesian additive regression trees (BART) was the best individual algorithm, achieving an MSE of 10.07×10^{-2} and AUC of 78.36×10^{-2} (95% CIs = 77.59–79.13) using either the full set or lasso set of predictors. However, the super learner had significantly better AUC than all 36 non-logistic algorithms ($Ds = 4.7–41.2$; $ps < 0.001$). Nonzero super learner weights were estimated for only nine (23%) of the individual algorithms.

Table 2
Overall performance of the cross-validated individual algorithms and the cross-validated super learner.

	MSE x 10^{-1}	AUC x 10^{-2}	(95% CI)
Super learner	9.94	79.04	(78.27–79.81)

Algorithm	Full set (67 variables)				Lasso set (51 variables)				T-test set (10 variables)			
	MSE x 10^{-2}	AUC x 10^{-2}	(95% CI)	SL Wgt	MSE x 10^{-2}	AUC x 10^{-2}	(95% CI)	SL Wgt	MSE x 10^{-2}	AUC x 10^{-2}	(95% CI)	SL Wgt
Logistic	10.25	77.22*	(76.43–78.02)	–	10.25	77.24*	(76.45–78.03)	–	10.31	76.51*	(75.70–77.33)	–
Ridge	10.24	77.39*	(76.60–78.18)	–	10.24	77.35*	(76.56–78.14)	–	10.31	76.60*	(75.79–77.41)	–
Elastic net (MPP = 0.25)	10.23	77.35*	(76.55–78.14)	–	10.24	77.28*	(76.49–78.08)	–	10.30	76.53*	(75.72–77.34)	–
Elastic net (MPP = 0.50)	10.22	77.33*	(76.54–78.12)	–	10.25	77.27*	(76.48–78.07)	–	10.30	76.52*	(75.71–77.34)	–
Elastic net (MPP = 0.75)	10.24	77.30*	(76.51–78.10)	–	10.25	77.27*	(76.48–78.07)	–	10.30	76.52*	(75.71–77.33)	–
Lasso	10.24	77.29*	(76.50–78.09)	–	10.25	77.27*	(76.48–78.06)	–	10.30	76.52*	(75.70–77.33)	–
Adaptive splines	10.17	77.67*	(76.87–78.46)	–	10.16	77.72*	(76.93–78.52)	0.06	10.22	77.10*	(76.29–77.91)	–
Adaptive polynomial splines	10.56	73.64*	(72.79–74.49)	–	10.45	74.96*	(74.12–75.81)	0.02	10.35	76.19*	(75.38–77.01)	–
Random forests	10.15	77.66*	(76.86–78.47)	0.28	10.29	77.03*	(76.21–77.85)	0.05	11.17	72.30*	(70.84–73.76)	0.15
Bayesian additive trees	10.07	78.36*	(77.59–79.13)	0.23	10.07	78.36*	(77.59–79.13)	0.03	10.09	77.90*	(77.11–78.68)	0.15
SVM (Linear)	11.57	52.69*	(51.60–53.79)	–	11.53	51.86*	(50.78–52.95)	–	11.43	58.59*	(57.53–59.65)	–
SVM (Polynomial)	11.22	64.42*	(63.30–65.54)	–	11.28	63.20*	(62.07–64.34)	–	11.54	50.21*	(48.68–50.90)	–
SVM (Radial)	10.91	68.60*	(67.59–69.62)	0.03	10.97	67.57*	(66.51–68.62)	–	11.50	56.27*	(55.18–57.36)	–

*Significantly different AUC (worse performance) compared to the super learner ($p < 0.001$) using the Hanley and McNeil (1983) method with 1000 bootstrap replicates.
Abbreviations: CV, 10-fold cross validation; MSE, mean squared error; AUC, area under the receiving operating characteristic curve; CI, confidence interval; MPP, mixing penalty parameter (alpha); SVM, support vector machines; SL Wgt, super learner weight; – indicates that the super learner weight was zero.

2.3. *Tiers of risk*

We then examined the operating characteristics across the 20 tiers of super learner predicted risk. Respondents in the top 5%, 10%, and 20% of the predicted risk distribution respectively accounted for 17.5%, 32.2%, and 51.4% of all cases of probable PTSD (Fig. 1). PPV in these risk tiers was 46.6%, 42.8%, and 34.2%, respectively (Fig. 2). Specificity and negative predictive value in the top 5%, 10%, and 20% of risk were 96.9/88.4%, 93.4/90.0%, and 84.8/91.2%, respectively.

3. Discussion

This is the first study to develop an optimized risk score for post-disaster PTSD using machine learning. As it was unclear a priori what type of algorithm would result in the most accurate prediction, we applied the ensembling method known as super learning to develop the risk score from many different types of algorithms. The super learner algorithm had significantly better AUC than all 39 of the individual algorithms from which it was developed. The relative improvement of the super learner compared to the individual algorithms varied widely. Based on AUC, super learning was 35–52% better than linear support vector machines (the worst performing algorithm), 1% better than BART (the best performing algorithm), and 2–3% better than logistic regression.

A super learner algorithm should not perform worse than the best performing individual algorithm from which it is developed. Thus, although one may question the relative improvement in AUC compared to a (simpler) single algorithm (e.g., logistic regression), super learning should be considered whenever the goal is maximizing prediction accuracy. Further, for debilitating conditions like PTSD, improvements in prediction accuracy of *any* magnitude may be important. For example, whereas 557 true-positive cases of probable PTSD were in the top 5% of risk based on the super learner, 487–502 true-positive cases were in the top 5% of risk based on logistic regression - logistic regression would miss over 50 cases. It is also possible for super learning to have even larger improvements over individual algorithms than observed here (Bergquist et al., in press; van der Laan and Rose, 2011).

Performance of our super learner can be compared to two other post-disaster PTSD risk scores developed using logistic regression. Liu and colleague's (2012) 11-variable post-earthquake PTSD risk score was developed in a training subsample (70% of the total sample) and achieved an AUC of 0.77 when applied to the remaining validation subsample (sensitivity and PPV were not reported because no high-risk cut-point was operationalized). This AUC is lower than our super learner and identical to the AUC achieved in our CV logistic regression algorithms. These AUC values reflect the upper-end of *fair* prediction. Bromet and colleague's (2017) risk score for PTSD related to *any* type of disaster, which was also formed using 10-fold CV, achieved a much lower AUC (0.63). Although comparisons should be made cautiously given their small sample and use of retrospective reports of disasters and PTSD risk factors, Bromet et al.'s risk score also had a lower PPV (20.4%) but *higher* sensitivity (44.5%) among individuals in the top 5% risk tier.

Although additional research and development is needed prior to implementation, the current study provides a foundation for future research that should aim to expand and adapt super learning risks scores in other disaster settings. With sufficient external validation in samples of earthquake survivors from different countries/cultures, it could be possible to develop an app-based interface that would allow disaster aid workers to use a super learner risk score to make recommendations for preventive interventions. Whereas a multi-week cognitive-behavioral prevention program might be offered to individuals in the top one or two tiers of risk, a brief one-session psychoeducation-based program might be offered to individuals in middle risk tiers. Prior to large scale implementation of a risk tool, it would be necessary to determine disaster-specific cost and resource effective cut-point(s) for triaging survivors to different mental health services.

Three limitations must be acknowledged. First, the outcome was *probable DSM-IV* PTSD, assessed using a self-report questionnaire. Although research suggests the Spanish translation of the DTS can accurately approximate a *DSM-IV* PTSD diagnosis, algorithm performance may have been different had the outcome been a *DSM-5* PTSD diagnosis assessed using a structured clinical interview (e.g., Clinician-Administered PTSD Scale, Weathers et al., 2013). Second, the survey was conducted over five years ago, in a sample of individuals from a single country, and was not designed to assess the full range of risk factors for post-disaster PTSD. The use of cross-validation does not preclude external validation and expansion in other samples of earthquake survivors as performance may be

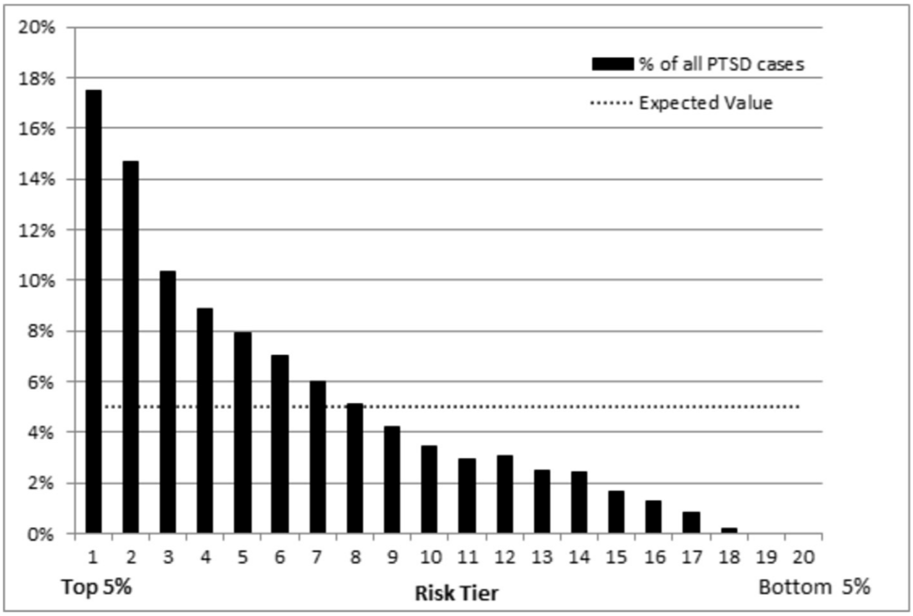

Fig. 1. Proportion of probable PTSD cases within ventiles of predicted risk (sensitivity) based on the super learner.
Ventiles of predicted risk were created by rank ordering the super learner predicted probabilities and creating 20 equally sized groups. The bars show the observed proportions of all PTSD cases in each ventile of risk.

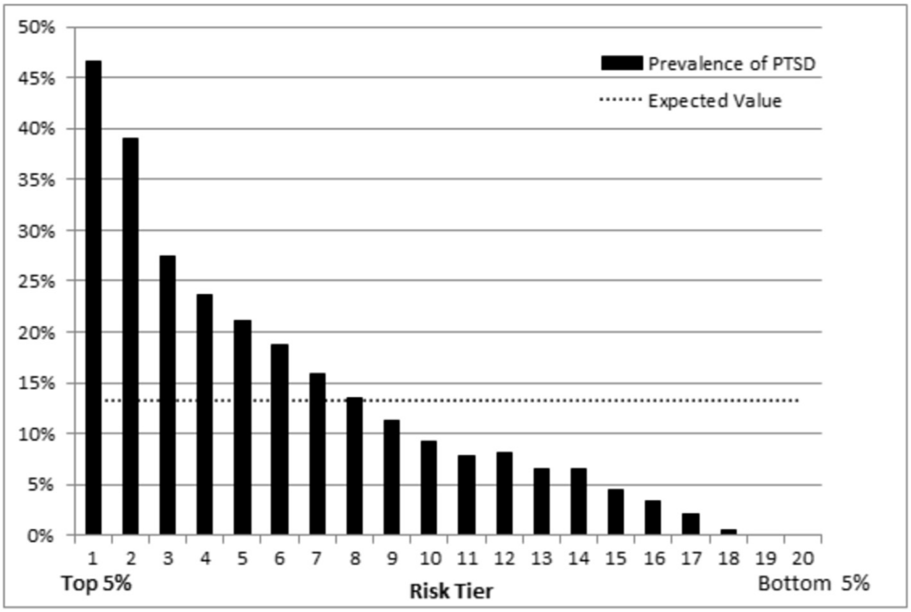

Fig. 2. Prevalence of probable PTSD within ventiles of predicted risk (positive predictive value) based on the super learner.
Ventiles of predicted risk were created by rank ordering the super learner predicted probabilities and creating 20 equally sized groups. The bars show the prevalence of PTSD cases in each ventile of risk.

improved with data on other peri- and post-earthquake factors (e.g., significant others suffering injuries, Kun et al., 2013; Tural et al., 2004; Wang et al., 2009; Wen et al., 2012). Performance also may be improved with inclusion of biological risk factors (e.g., neuroendocrine response, Galatzer-Levy et al., 2017), which could ultimately guide future pharmacological preventive interventions. Third, there are limitations associated with the use of super learning – it takes increased computational time compared to using only individual algorithms, though this limitation will be less problematic as technology advances. Super learner performance is nonetheless contingent on the performance of the individual algorithms included in the library.

Within the context of these limitations, we demonstrated that super learning can be used to create a risk score that predicts PTSD more accurately than conventional regression methods. Expansion and further validation of the current super learning algorithm for post-earthquake PTSD, and developing super learner algorithms for other types of disasters (including man-made disasters), could be of great value to mental health professionals attempting to identify individuals in need of PTSD preventive interventions. This is especially true given the growing PTSD prevention literature (Amos et al., 2014; Forneris et al., 2013; Kliem and Kroger, 2013).

Conflict of interest

Dr. Kessler has been a consultant for Hoffman-La Roche, Inc. and Johnson & Johnson Wellness and Prevention. Dr. Kessler has served on advisory boards for Mensante Corporation, Johnson & Johnson Services Inc. Lake Nona Life Project, and U.S. Preventive Medicine. Dr. Kessler is a co-owner of DataStat, Inc. The other authors declare nothing to disclose.

Acknowledgement

The authors also thank the Ministry of Planning (currently the Social Development Ministry – Ministerio de Desarrollo Social de Chile), intellectual proprietor of the surveys, for having permitted access to the data.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.jpsychires.2017.09.010.

References

American Red Cross, 2012. Disaster Mental Health Handbook: Disaster Services. http://www.cdms.uci.edu/PDF/Disaster-Mental-Health-Handbook-Oct-2012.pdf.

Amos, T., Stein, D.J., Ipser, J.C., 2014. Pharmacological interventions for preventing post-traumatic stress disorder (PTSD). Cochrane Database Syst. Rev. 7.

Atkinson, G.M., Wald, D.J., 2007. "Did you feel it?" intensity data: a surprisingly good measure of earthquake ground motion. Seismol. Res. Lett. 78, 362–368.

Bergquist, S., Brooks, G., Keating, N., Landrum, M.B., Rose, S., (in press). Classifying lung cancer severity with ensemble machine learning in health care claims data. Proceedings of Machine Learning Research.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

Bromet, E.J., Atwoli, L., Kawakami, N., Navarro-Mateu, F., Piotrowski, P., King, A.J., Aguilar-Gaxiola, S., Alonso, J., Bunting, B., Demyttenaere, K., Florescu, S.E., De Girolamo, G., Gluzman, S., Haro, J.M., De Jonge, P., Karam, E.G., Lee, S., Kovess-Masfety, V., Medina-Mora, M.E., Mneimneh, Z., Pennell, B.E., Posada-Villa, J., Salmeron, D., Takeshima, T., Kessler, R.C., 2017. Post-traumatic stress disorder assoicated with natural and human-made disasters in the World Mental Health Surveys. Psychol. Med. 47 (2), 227–241.

Cairo, J.B., Dutta, S., Nawaz, H., Hashmi, S., Kasl, S., Bellido, E., 2010. The prevalence of posttraumatic stress disorder among adult earthquake survivors in Peru. Disaster Med. Public Health Prep. 4 (1), 39–46.

Center for Research on the Epidemiology of Disaster (CRED), 2016. Disaster Profiles. http://www.emdat.be/disaster_profiles/index.html. (Accessed 27 July 2016).

Chen, H., Chen, Y., Au, M., Feng, L., Chen, Q., Guo, H., Li, Y., Yang, X., 2014. The presence of post-traumatic stress disorder symptoms in earthquake survivors one month after a mudslide in southwest China. Nurs. Health Sci. 16, 39–45.

Cheng, Y., Wang, F., Wen, J., Shi, Y., 2014. Risk factors of post-traumatic stress disorder (PTSD) after Wenchuan earthquake: a case control study. PLoS One 9 (5), e96644.

Chipman, H.A., George, E.I., McCulloch, R.E., 2010. BART: Bayesian additive regression trees. Ann. Appl. Statistics 4 (1), 266–298.

Coronas, R., García-Parés, G., Viladrich, C., Santos, J.M., Menchón, J.M., 2008. Clinical and sociodemographic variables associated with the onset of posttraumatic stress disorder in road traffic accidents. Depress Anxiety 25, E16–E23.

Dai, W., Chen, L., Lai, Z., Li, Y., Wang, J., Liu, A., 2016. The incidence of post-traumatic stress disorder among survivors after earthquakes:a systematic review and meta-analysis. BMC Psychiatry 16, 188.

Davidson, J.R., Book, S.W., Colket, J.T., Tupler, L.A., Roth, S., David, D., Hertzberg, M., Mellman, T., Beckham, J.C., Smith, R.D., Davison, R.M., Katz, R., Feldman, M.E., 1997. Assessment of a new self-rating scale for post-traumatic stress disorder. Psychol. Med. 27 (1), 153–160.

Dell'Osso, L., Carmassi, C., Massimetti, G., Stratta, P., Riccardi, I., Capanna, C., Akiskal, K.K., Akiskal, H.S., Rossi, A., 2013. Age, gender and epicenter proximity effects on post-traumatic stress symptoms in L'Aquila 2009 earthquake survivors. J. Affect Disord. 146 (2), 174–180.

Division Observatorio Social, 2010. Encuesta Post Terremoto. Ministerio de Desarrollo Social, Gobierno de Chile.

Division Observatorio Social, AGCI Ministerio de Relaciones Exteriores, 2015. Consulta Interactiva de Datos: Manuales. http://www.redatam.org/redchl/mds/encpt/. (Accessed 1 October 2015).

Fan, F., Zhang, Y., Yang, Y., Mo, L., Liu, X., 2011. Symptoms of posttraumatic stress disorder, depression, and anxiety among adolescents following the 2008 Wenchuan earthquake in China. J. Trauma Stress 24 (1), 44–53.

Forneris, C.A., Gartlehner, G., Brownley, K.A., Gaynes, B.N., Sonis, J., Coker-Schwimmer, E., Jonas, D.E., Greenblatt, A., Wilkins, T.M., Woodell, C.L., Lohr, K.N., 2013. Interventions to prevent post-traumatic stress disorder: a systematic review. Am. J. Prev. Med. 44 (6), 635–650.

Friedman, J.H., 1991. Multivariate adaptive regression splines. Ann. Statistics 19 (1), 1–141.

Friedman, J., Hastie, T., Tibshirani, R., 2010. Regularization paths for generalized linear models via coordinate descent. J. Stat. Softw. 33 (1), 1–22.

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. J. Psychiatr. Res. 59, 68–76.

Galatzer-Levy, I.R., Ma, S., Statnikov, A., Yehuda, R., Shalev, A.Y., 2017. Utilization of machine learning for prediction of post-traumatic stress: a re-examination of cortisol in the prediction and pathways to non-remitting PTSD. Transl. Psychiatry 7 (3), e0.

Goldmann, E., Galea, S., 2014. Mental health consequences of disasters. Annu. Rev. Public Health 35, 169–183.

Hanley, J.A., McNeil, B.J., 1983. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology 148 (3), 839–843.

Hosmer, D.W., Lemeshow, S., Sturdivant, R.X., 2013. Applied Logistic Regression, third ed. Wiley, Hoboken, New Jersey.

Instituto Nacional de Estadísticas, 2015. Demográficas y Vitales, Productos Estadísticos. Comunas: Actualización de Población 2002-2012 y Proyecciones 2012-2020. http://www.ine.cl/canales/chile_estadistico/familias/demograficas_vitales.php. (Accessed 1 October 2015).

Karstoft, K.I., Galatzer-Levy, I.R., Statnikov, A., Li, Z., Shalev, A.Y., members of Jerusalem Trauma, O., Prevention Study, g, 2015a. Bridging a translational gap: using machine learning to improve the prediction of PTSD. BMC Psychiatry 15, 30.

Karstoft, K.I., Statnikov, A., Andersen, S.B., Madsen, T., Galatzer-Levy, I.R., 2015b. Early identification of posttraumatic stress following military deployment: application of machine learning methods to a prospective study of Danish soldiers. J. Affect Disord. 184, 170–175.

Kessler, R.C., Rose, S., Koenen, K.C., Karam, E.G., Stang, P.E., Stein, D.J., Heeringa, S.G., Hill, E.D., Liberzon, I., McLaughlin, K.A., McLean, S.A., Pennell, B.E., Petukhova, M., Rosellini, A.J., Ruscio, A.M., Shahly, V., Shalev, A.Y., Silove, D., Zaslavsky, A.M., Angermeyer, M.C., Bromet, E.J., de Almeida, J.M., de Girolamo, G., de Jonge, P., Demyttenaere, K., Florescu, S.E., Gureje, O., Haro, J.M., Hinkov, H., Kawakami, N., Kovess-Masfety, V., Lee, S., Medina-Mora, M.E., Murphy, S.D., Navarro-Mateu, F., Piazza, M., Posada-Villa, J., Scott, K., Torres, Y., Carmen Viana, M., 2014. How well can post-traumatic stress disorder be predicted from pre-trauma risk factors? An exploratory study in the WHO World Mental Health Surveys. World Psychiatry Official J. World Psychiatric Assoc. 13 (3), 265–274.

Kim, D.H., Cha, J.M., Shin, H.P., Joo, K.R., Lee, J.I., Park, D.I., 2015. Development and validation of a risk stratification-based screening model for predicting colorectal advanced neoplasia in Korea. J. Clin. Gastroenterol. 49 (1), 41–49.

Kliem, S., Kroger, C., 2013. Prevention of chronic PTSD with early cognitive behavioral therapy. A meta-analysis using mixed-effects modeling. Behav. Res. Ther. 51 (11), 753–761.

Kun, P., Han, S., Chen, X., Yao, L., 2009. Prevalence and risk factors for posttraumatic stress disorder: a cross-sectional study among survivors of the Wenchuan 2008 earthquake in China. Depress Anxiety 26 (12), 1134–1140.

Kun, P., Tong, X., Liu, Y., Pei, X., Luo, H., 2013. What are the determinants of post-traumatic stress disorder: age, gender, ethnicity or other? Evidence from 2008 Wenchuan earthquake. Public Health 127 (7), 644–652.

Lai, T.J., Chang, C.M., Connor, K.M., Lee, L.C., Davidson, J.R., 2004. Full and partial PTSD among earthquake survivors in rural Taiwan. J. Psychiatr. Res. 38 (3), 313–322.

LeDell, E., van der Laan, M.J., Peterson, M., 2016. AUC-maximizing ensembles through metalearning. Int. J. Biostat. 12 (1), 203–218.

Leiva-Bianchi, M.C., Araneda, A.C., 2013. Validation of the Davidson Trauma Scale in its original and a new shorter version in people exposed to the F-27 earthquake in Chile. Eur. J. Psychotraumatol 4.

Liu, X., Ma, X., Hu, X., Qiu, C., Wang, Y., Wang, Q., Zhang, W., Zhang, J., Li, T., 2012. A risk score for predicting post-traumatic stress disorder in adults in a Chinese earthquake area. J. Int. Med. Res. 40 (6), 2191–2198.

López, E.N., Insunza, J.V., 2013. El terremoto/tsunami en Chile. Una mirada a las estadísticas médico legales. Investig. Forense 2, 113–129.

National Geophysical Data Center (NGDC), & World Data Service (WDS), 2016. Global Historical Tsunami Database: National Geophysical Data Center, NOAA. https://www.ngdc.noaa.gov/hazard/tsu_db.shtml. (Accessed 1 March 2016).

Norris, F.H., Friedman, M.J., Watson, P.J., Byrne, C.M., Diaz, E., Kaniasty, K., 2002. 60,000 disaster victims speak: Part I. An empirical review of the empirical literature, 1981-2001. Psychiatry 65 (3), 207–239.

North, C.S., 2014. Curr. Psychiatry Rep. 16, 481.

Ozer, E.J., Best, S.R., Lipsey, T.L., Weiss, D.S., 2003. Predictors of posttraumatic stress disorder and symptoms in adults: a meta-analysis. Psychol. Bull. 129 (1), 52–73.

Polley, E.C., LeDell, E., van der Laan, M.J., 2016. SuperLearner: Super Learner Prediction (Version 2.0–19) [R Package].

Priebe, S., Grappasonni, I., Mari, M., Dewey, M., Petrelli, F., Costa, A., 2009. Post-traumatic stress disorder six months after an earthquake: findings from a community sample in a rural region in Italy. Soc. Psychiatry Psychiatr. Epidemiol. 44 (5), 393–397.

R Core Team, 2015. R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.C., Muller, M., 2011. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinforma. 12, 1–8.

Rose, S., 2013. Mortality risk score prediction in an elderly population using machine learning. Am. J. Epidemiol. 177 (5), 443–452.

Rosendal, S., Mortensen, E.L., Andersen, H.S., Heir, T., 2014. Use of health care services before and after a natural disaster among survivors with and without PTSD. Psychiatr. Serv. 65 (1), 91–97.

Ruiz-Parraga, G.T., Lopez-Martinez, A.E., 2014. The contribution of posttraumatic stress symptoms to chronic pain adjustment. Health Psychol. 33 (9), 958–967.

Sattler, D.N., Assanangkornchai, S., Moller, A.M., Kesavatana-Dohrs, W., Graham, J.M., 2014. Indian Ocean tsunami: relationships among posttraumatic stress, posttraumatic growth, resource loss, and coping at 3 and 15 months. J. Trauma Dissociation 15 (2), 219–239.

Sayed, S., Iacoviello, B.M., Charney, D.S., 2015. Risk factors for the development of psychopathology following trauma. Curr. Psychiatry Rep. 17 (70), 1–7.

Schreiber, M.D., Yin, R., Omaish, M., Broderick, J.E., 2014. Snapshot from superstorm sandy: american red cross mental health risk surveillance in lower New York state. Ann. Emerg. Med. 64 (1), 59–65.

Steinwart, I., Christmmann, A., 2008. Support Vector Machines. Springer, New York.

Stone, C.J., Hansen, M., Kooperberg, C., Truong, Y.K., 1997. Polynomial splines and their tensor products in extended linear modeling (with discussion). Ann. Statistics 25, 1371–1470.

Tural, U., Coskun, B., Onder, E., Corapcioglu, A., Yildiz, M., Kesepara, C., Karakaya, I., Aydin, M., Erol, A., Torun, F., Aybar, G., 2004. Psychological consequences of the 1999 earthquake in Turkey. J. Trauma Stress 17 (6), 451–459.

United States Geological Survey (USGS), USGS Earthquake Hazards Program, 2016a. Earthquake Hazards Program. http://earthquake.usgs.gov/earthquakes/search/. (Accessed 20 April 2016).

United States Geological Survey (USGS), USGS Earthquake Hazards Program, 2016b. ShakeMap Scientific Background. http://earthquake.usgs.gov/earthquakes/shakemap/background.php#accmaps. (Accessed 20 April 2016).

van der Laan, M.J., Rose, S., 2011. Targeted Learning: Causal Inference for Observational and Experimental Data. Springer, New York.

van der Laan, M.J., Polley, E.C., Hubbard, A.E., 2007. Super learner. Stat. Appl. Genet. Mol. Biol. 6. Article25.

van Griensven, F., Chakkraband, M.L., Thienkrua, W., Pengjuntr, W., Lopes Cardozo, B., Tantipiwatanaskul, P., Mock, P.A., Ekassawin, S., Varangrat, A., Gotway, C., Sabin, M., Tappero, J.W., Thailand Post-Tsunami Mental Health Study, G, 2006. Mental health problems among adults in tsunami-affected areas in southern Thailand. Jama 296 (5), 537–548.

Wald, D.J., Quitoriano, V., Heaton, T.H., Kanamori, H., 1999. Relationships between peak ground acceleration, peak ground velocity, and modified mercalli Intensity in California. Earthq. Spectra 15 (3), 557–564.

Wang, L., Zhang, Y., Wang, W., Shi, Z., Shen, J., Li, M., Xin, Y., 2009. Symptoms of posttraumatic stress disorder among adult survivors three months after the Sichuan earthquake in China. J. Trauma Stress 22 (5), 444–450.

Wang, B., Ni, C., Chen, J., Liu, X., Wang, A., Shao, Z., Xiao, D., Cheng, H., Jiang, J., Yan, Y., 2011. Posttraumatic stress disorder 1 month after 2008 earthquake in China: wenchuan earthquake survey. Psychiatry Res. 187 (3), 392–396.

Weathers, F.W., Blake, D.D., Schnurr, P.P., Kaloupek, D.G., Marx, B.P., Keane, T.M., 2013. The Clinician-administered PTSD Scale for DSM-5 (CAPS-5). Interview available from the National Center for PTSD at www.ptsd.va.gov.

Wen, J., Shi, Y., Li, Y., Yuan, P., Wang, F., 2012. Quality of life, physical diseases, and psychological impairment among survivors 3 years after Wenchuan Earthquake: a population based survey. PLoS One 7 (8), e43081.

Yeoh, K.G., Ho, K.Y., Chiu, H.M., Zhu, F., Ching, J.Y., Wu, D.C., Matsuda, T., Byeon, J.S., Lee, S.K., Goh, K.L., Sollano, J., Rerknimitr, R., Leong, R., Tsoi, K., Lin, J.T., Sung, J.J., Asia-Pacific Working Group on Colorectal, C, 2011. The Asia-Pacific Colorectal Screening score: a validated tool that stratifies risk for colorectal advanced neoplasia in asymptomatic Asian subjects. Gut 60 (9), 1236–1241.

Zhang, Z., Shi, Z., Wang, L., Liu, M., 2011. One year later: mental health problems among survivors in hard-hit areas of the Wenchuan earthquake. Public Health 125 (5), 293–300.

Zubizarreta, J.R., Cerda, M., Rosenbaum, P.R., 2013. Effect of the 2010 Chilean earthquake on posttraumatic stress: reducing sensitivity to unmeasured bias through study design. Epidemiology 24 (1), 79–87.