HEDG

Health, Econometrics and Data Group

The University of York

WP 20/20

Economic Consequences of Road Traffic Injuries. Application of the Super Learner algorithm

Ieva Sriubaite; Anthony Harris;

Andrew M. Jones and Belinda Gabbe

November 2020

http://www.york.ac.uk/economics/postgrad/herc/hedg/wps/

Economic Consequences of Road Traffic Injuries. Application of the Super Learner algorithm*

Ieva Sriubaite† Anthony Harris‡ Andrew M. Jones§
Belinda Gabbe¶

Nov 2020

Abstract

We perform a prediction analysis using methods of supervised machine learning on a set of outcomes that measure economic consequences of road traffic injuries. We employ several parametric and non-parametric algorithms including regularised regressions, decision trees and random forests to model statistically challenging empirical distributions and identify the key risk groups. In addition to a traditional outcome of interest – health care costs – we predict net monetary benefits from treatment, and productivity losses measured by the probability to return to work after the injury. Using the predictions of each selected algorithm we construct an ensemble machine learning algorithm - the Super Learner algorithm. Our findings demonstrate that the Super Learner is effective and performs best in predicting all outcomes. Further analysis of predictions by different groups of patients play an important role in the understanding of key risk factors for higher costs and poorer outcomes and offers a deeper understanding of risk in the health care sector.

Keywords: Prediction and classification, super learner, machine learning, health-care costs, patient outcomes, road traffic injuries.

JEL Classifications: I11, I19, C14, C38, C53.

---

*We thank to participants of the Virtual Essen Health Economics Seminar Series for valuable comments. Financial support from the DAAD/Go8 funding by the "Bundesministerium für Bildung und Forschung (BMBF)", German Federal Ministry of Education and Research is gratefully acknowledged. The Victorian State Trauma Registry is a State Government of Victorian Department of Health and Human Services and Transport Accident Commission funded initiative.

†Corresponding Author: Centre for Health Economics, Monash Business School, Monash University Level 5, Building H, Caulfield Campus, 900 Dandenong Road, Caulfield East VIC 3145, Australia. E-mail: ieva.sriubaite@monash.edu.

‡Centre for Health Economics, Monash Business School, Monash University, Australia

§Department of Economics and Related Studies, University of York, UK and Centre for Health Economics, Monash Business School, Monash University, Australia

¶School of Public Health and Preventive Medicine, Monash University, Australia. Belinda Gabbe was supported by an Australian Research Council Future Fellowship (FT170100048)

1 Introduction

Budget pressures on governments generate considerable incentives for healthcare payers to develop statistical models that accurately predict future healthcare expenditures. Such models are extensively applied in both public and private insurance payment systems to reduce inefficiencies associated with adverse selection and to design effective provider reimbursement systems. For instance, under prospective payments providers have a particular interest in these models as inaccurate predictions of potential care costs puts them at risk of underpayment. Risk adjustment methodology is designed to remedy these disadvantages by aligning the prospective or capitated payment made to a provider with an individual's expected healthcare costs and thus reduce incentives to under treat or reject high cost individuals (Kapur *et al.*, 2000; Cucciare and o'Donohue, 2006). Yet the theoretical design does not often reflect in practice.

Historically, risk adjustment has relied almost exclusively on patient risk factors such as age and gender to predict individuals' healthcare costs for the following year (Ellis *et al.*, 2018). More recently, case mix funding models for hospitals incorporated diagnostic and treatment-related complexities such as the patient's classification to diagnosis-related groups, inlier/outlier models identifying the length of stay (Duckett, 1998; Pirson *et al.*, 2006; Schreyögg *et al.*, 2006) and the prediction of any hospital acquired complications to reflect any extra costs of a hospital admission (IHPA, 2019). A major limitation of this approach is its failure to account for potential moral hazard associated with the cost-quality trade-offs available to some providers (Ellis and McGuire, 1996; Jürges and Köberlein, 2015). When clinical risk is low, the provider is incentivized to adjust their provided quality of care in order to reduce costs and maintain a revenue surplus in low return cases. Thus, to be efficient the payment system requires not only accurately predict potential treatment costs, but also inform public funders about patient clinical outcomes. In this paper we extend the work of Rose (2016) and consider additional patient outcomes in the risk adjustment framework.

Commonly applied models in the risk adjustment literature are standard linear regressions that include a small set of patient demographic characteristics and clinical diagnoses to adjust for potential risk factors (Iezzoni, 2012). These are largely outdated given the significant improvements in statistical modelling. The statistical profile of such outcomes are often characterized by non-normal distributions; data are often strongly skewed and exhibit a particularly long tail representing patients at the highest risk for high costs and poor outcomes (Jones, 2011). Thus, modelling these outcomes comes with a number of statistical challenges. The literature on cost modeling has evolved from applications of various parametric approaches (see, e.g., Duan, 1983; Jones, 2000; Manning *et al.*, 2005;

Dixon *et al.*, 2011; Curtis *et al.*, 2014; Jones *et al.*, 2014), but traditional methods often suffer from problems caused by the presence of significant correlations between the selected covariates (James *et al.*, 2013) and could potentially provide misleading information about the key risk-factors. In addition, the performance of such models is often better when data is transformed using a deterministic function (e.g. logarithmic), that adds a level of complexity to the interpretation of results. To tackle these challenges researchers have explored a variety of semi-parametric approaches (see, e.g., Deb and Burgess, 2003; Gilleskie and Mroz, 2004; Manning *et al.*, 2005; Mullahy, 2009; McDonald *et al.*, 2013; Jones, 2011; Jones *et al.*, 2014, 2015), though the performance of such models is mixed (Jones and Lomas, 2016).

As medical information has become more and more detailed, the focus has shifted towards a variety of new statistical approaches utilised in the field of "big data". Data science methods such as supervised Machine Learning (ML) offer functional flexibility and the ability to fit difficult data patterns without imposing prior assumptions. Several data science techniques, such as testing out-of-sample performance, have already been adopted in the modelling literature (Bertsimas *et al.*, 2008; Jones *et al.*, 2014; Jones and Lomas, 2016). The main advantages of ML methods are the ability to uncover complex structures not known/specified in advance (Mullainathan and Spiess, 2017) and to account for potential multicollinearity when controlling for a large set of covariates (James *et al.*, 2013). Thus, they enable fitting very flexible functional forms without overfitting the data and can perform better at out-of-sample predictions than standard regression analysis (Chu and Zhang, 2003).

A small but growing literature has used such methods in research related to risk adjustment and prediction of health care resource use, service utilisation and various clinical outcomes (e.g., Bertsimas *et al.*, 2008; Lahiri, 2014; Arandjelović, 2015; Einav *et al.*, 2016; Rose, 2016; Pyrkov *et al.*, 2018; Burnham *et al.*, 2018; Kan *et al.*, 2019). In this paper we contribute to this literature by utilising such methods to investigate the main risk factors for high costs and poor outcomes of road traffic injuries. Around 1.35 million people died worldwide in 2016 due to a road traffic related injury, which is currently the eighth leading cause of death and the leading cause of death among children and adults younger than 30 years old (WHO, 2018). Alongside high fatality rates, road traffic injuries leave around 50 million people worldwide with non-fatal injuries that are likely to become lifelong disabilities with a heavy economic burden to victims and their families due to costly treatment and rehabilitation care (Chen *et al.*, 2019).

In Australia costs of road trauma are estimated to have been approximately A$22.2 billion in 2015, equivalent to 1.3 per cent of GDP. Although fatality rates have been decreasing over time, the number of hospitalised injuries is rising. The loss of life, health and well-being account for the largest item of total economic costs of road injuries (42%) while the specialised care and rehabilitation for persons who were disabled during injury is the third largest item with an additional 10 per cent of total costs (AAA, 2017). In an environment of rapidly increasing health care expenditure it is particularly important for researchers and policy makers to understand the main risk factors for high costs and poor outcomes. This will ensure cost-effective delivery of services by weighing the costs of treatment and long-term care against the gains in health and well-being.

This paper employs a rich patient-level dataset, the Victorian State Trauma Registry (VSTR) and complements it with the comprehensive insurance claims data provided by the Transport Accident Commission. In Victoria, the Transport Accident Commission (TAC) is a government-owned organisation that offers compulsory third party insurance for those who were injured in transport collisions. With a goal to promote road traffic safety and improve the state's trauma system, the insurer collects funds via a TAC charge, a component of the vehicle registration yearly fee. The insurance scheme offers ongoing financial support for medical treatment of incident-related injuries as well as the loss of income during the recovery time. Due to medical heterogeneity of each trauma case and complex and expensive long-term recovery pathways it is particularly important to find efficient ways to distribute these resources. Hence, to stay financially sustainable organisations such as the TAC have considerable interest in understanding the main factors and predictors of health care expenditures and other long-term outcomes.

To contribute to the growing literature on risk-adjustment in health care markets, we employ a Machine Learning (ML) based algorithm – the Super Learner – to predict the economic consequences of injury. The Super Learner algorithm is based on multiple parametric and non-parametric algorithms and selects an optimal weighted combination of them to find the best predictive model. Proposed by van der Laan *et al.* (2007) the algorithm has demonstrated significant potential in research related to health care that predicts various clinical patient outcomes (Kessler *et al.*, 2014; Pirracchio *et al.*, 2015). Evidence using the Super Learner in the economic context is provided in recent research by Rose (2016) and Rose *et al.* (2017) that focus on plan payment risk adjustment and the prediction of the unprofitability of health insurance enrollees.

In this paper we contribute to this emerging evidence of the application of the Super Learner in the economic context and consider a wider scope of policy relevant outcomes for risk-adjustment. In addition to a traditional measure of healthcare spending, we consider other non-medical spending associated with injury such as adaptation to physical disability and equipment costs that are covered by the TAC insurance. To shed light on the cost-effectiveness of such treatment we adopt the concept proposed by Stinnett and Mullahy (1998) and estimate the net benefit based on the traditional measure of the

Quality of Life Years (QALYs) in the cost-effectiveness literature. Net benefits from the treatment and TAC insurance coverage provide policy relevant information about the societal value of health expressed in the monetary value of the QALY. Lastly, as in most cases traffic on roads involves the working age population, return to work is a key indicator of successful rehabilitation following a road trauma. For this reason, we consider victims' probability to return to work after the injury.[1] This paper contributes to the current risk adjustment literature in predicting health and social care related costs. First, in addition to a measure of resource use, we consider other societal values of health care and employment that inform health economists and policy makers about the quality of care. Second, we adapt advanced statistical methods to improve the predictive power of traditional regression-based approaches and employ an ensemble ML algorithm, the Super Learner, to construct the best predictive model.

This paper is organised as follows. The next section provides a summary of the Victorian State Trauma System as well as the government-owned insurance scheme (TAC) in the state of Victoria. Section 3 describes the data, while Section 4 outlines the methodology used to estimate the economic consequences of the injury and introduces the prediction methods in detail. Section 5 reports prediction results and evaluates them based on a number of evaluation criteria, Section 6 discusses potential prediction errors and Section 7 concludes.

2 Victorian State Trauma System and TAC insurance

Victoria is the second most populous state in Australia with a population of approximately 6.5 million.[2] The state operates a regionalized trauma system to ensure that injured patients receive the best possible medical treatment and specialized hospital care. The system is categorized into three levels of care: three major trauma services [MTS] (two adult and one paediatric), that provide definite care to most of the state's trauma patients either through primary triage or secondary transfer; metropolitan trauma services [MeTS] and regional trauma services are the second level of care that also provide immediate care when MTS cannot be reached in time; and Metropolitan Primary Care Services offers the third level of care (DHHS, Feb 2014). Around 80% of trauma cases receive care at a designated trauma centre for major trauma services and nearly 90% of

[1]To date the most common statistical approaches to predict the return to work have been logistic regression and Cox proportional hazard models (see, e.g., Ip *et al.*, 1995; Nielsen *et al.*, 2010; Kong *et al.*, 2011; Van Patten *et al.*, 2016).

[2]as of September 2018, The Australian Bureau of Statistics, 3101.0 - Australian Demographic Statistics Catalogue.

road injuries are treated at a MTS (VSTR, 2014). The trauma system is monitored using a population-based trauma registry that collects data about all major trauma patients irrespective of the admitting hospital. The registry is used for research purposes in order to continuously enhance the quality of trauma management and improve patient outcomes after an injury (see, e.g. DHHS, Feb 2014; DHHS, Jul 2014; VSTR, 2014; Gabbe *et al.* (2012, 2014); Beck *et al.* (2016)).

Multiple sources of funding exist in Victoria to finance the treatment of trauma. While Australia's publicly funded universal health care insurance scheme (Medicare) provides health care coverage for all Australian citizens and permanent residents, nearly all care for road injuries is funded by a publicly-owned organisation, the Transport Accident Commission (TAC). The TAC operates on a "no-fault" basis and provides financial and rehabilitation support for Victorians who were injured in transport incidents. The compensation covers out-of-pocket medical and non-medical costs[3] and life-back-on-track expenses including income assistance, rehabilitation, return to work programs, travel and funeral costs as well as costs for specialised equipment such as wheelchairs and modified vehicles to support patients who acquired disability due to injury (TAC, 2018). The TAC is funded through annual vehicle registration payments and works closely with the Roads Corporation of Victoria (VicRoads) to improve the safety on Victoria's roads. In addition to providing financial assistance to injured individuals TAC invests significant funds to install various safety measures in high-risk incident locations and provides public education to encourage safe driving in the community.

3 Data

The empirical analysis uses tdata from the VSTR that includes information on all major trauma patients in Victoria.[4] It provides a wide range of patient characteristics such as age, gender, socio-economic status as well as comprehensive medical and non-medical information about patient injuries. The registry records various characteristics at the scene of the incident such as cause, place, and mechanism of the injury, that allow us to control for potentially important differences between injuries. Detailed clinical information recorded at the time of admission as well as during a hospital stay provides specific information about the patient's treatment and recovery. After discharge from the hospital

[3]For example, non-medical costs could be related to travelling to medical appointments, accommodation, any legal or administrative costs associated with the claims reimbursement and disability-related health support.

[4]Using the ICD-10-AM information, major trauma is defined if any of the following criteria are met (i) Death (at scene of injury or in-hospital) related to injury; (ii) an Injury Severity Score >12; (iii) admission to an intensive care unit (ICU) for > 24h and requiring mechanical ventilation for at least part of their ICU stay; and (iv) urgent surgery is performed.

each patient is followed up by a telephone interview at 6, 12 and 24 months after injury. An interviewer collects detailed information about patients' recovery, level of physical function, health-related quality of life and return to work.[5] To identify post-discharge deaths the registry is linked with the state's deaths register.

Our sample of interest is restricted to patients that experienced a major trauma as a result of a road traffic crash in Victoria during 2009–2017. We identify patients aged above 15 years who were funded by the TAC and link it with the insurance claims data. The TAC insurance database includes information on around 8 million instances of claims paid to patients who were injured in a road incident. The study uses 10 years of data and treatment-related information as well as paid benefits for losses of earnings due to the injury. Descriptive statistics for the sample are presented in Table A.1 in the Appendix.

4 Methods

4.1 Outcomes

Direct costs of the injury

Using detailed information from the TAC insurance claims data, for each patient i we compute *total direct costs of injury (D_i)* attributed to the initial treatment at the hospital as well as all subsequent costs a patient had within 24 months after discharge from the hospital. Here D_i is a function of both medical, M_i, (ambulance care, inpatient care hospital stay, outpatient care, rehabilitation and prescription drugs) and non-medical, N_i, (travel, accommodation, legal, administrative and disability related health support) expenses, but is not subject to any impairment annuity or loss of earnings.[6]

$$D_i = M_i + N_i \tag{1}$$

Net benefits of treatment

Due to a traumatic event patients with severe injuries often experience a lengthy and complicated recovery process with life-long consequences (McCullough *et al.*, 2014). To better understand benefits associated with the TAC insurance coverage, we follow the framework developed by Stinnett and Mullahy (1998). We calculate the utility gained

[5]The response rate of the follow up study at 6 and 12 months was around 70% for full data collection; 10% partial data collection and around 5% reported death before the study; at 24 months the respective figures were 65%, 20% and 7% respectively.

[6]We exclude any expenses related to impairment annuity, loss of earnings or death benefits paid to a spouse or other family member to avoid double counting with health-related quality of life used to calculate net benefits of treatment below.

from ongoing medical and non-medical support after injury provided by TAC as the difference between the quality-adjusted-life-years (QALYs) with treatment and the expected QALYs without treatment. For each patient i the*Net Health Benefit* (NHB) is defined as the following:

$$NHB_i = QALY_i^{TAC} - QALY_i^{NON} \tag{2}$$

here NHB is expressed in units of 'benefit' gained from treatment such as QALYs. $QALY^{TAC}$ is the benefit gained from the treatment provided by the TAC insurance, whereas $QALY^{NON}$ is the outcome without treatment for the life-threatening injury, which likely would be fated.

To measure QALYs associated with the treatment received, $QALY^{TAC}$, we rely on information about Australian life expectancy by age and gender provided by the Australian Institute of Health and Welfare[7] and the 3-level EuroQol five dimensions questionnaire (EQ-5D-3L),[8] that was included in the follow up survey conducted 2 years after the injury. EQ-5D-3L instrument comprises five dimensions based on patient's mobility, self-care, usual daily activities, experience of pain or discomfort and anxiety or depression. Each dimension is ranked in order of increasing severity according to: *no problems*, *some problems* or *extreme problems*. Using this information we estimate each patient's health utility scores using all five dimensions of EQ-5D, representing patient's health state two years after the injury occurred. A number of studies have concluded that an individual who had a traumatic injury is likely to suffer from long-term consequences from physical and psychologic impairment, acquired disability or Post-traumatic Stress Disorder. The highest risk for long-term effects was found in patients who were hospitalised or required critical medical care. (Holbrook *et al.*, 2001a,b, 2005; Sluys *et al.*, 2005; Holtslag *et al.*, 2007). Following this evidence, we estimate lifelong NHB for patients who encountered a life-threatening injury and, due to the severity and extensiveness of their injury, were admitted to the Intensive Care Unit (ICU). Using the information recorded during the in-hospital stay, we select a sample of patients who spent at least 1 day in the ICU and were mechanically ventilated. If patients did not receive this medical intervention, the counterfactual state would likely be death. Thus it is reasonable to assume that their $QALY^{NON}$ would be equal to zero and all units of 'benefit', $QALY_i^{TAC}$, would be gained from the initial medical intervention and ongoing financial support provided by TAC.

[7]The life expectancy tables can be found here: https://www.aihw.gov.au/reports/life-expectancy-death/deaths-in-australia/contents/age-at-death [Accessed 14.04.2020]. The data were collected from the AIHW 17 Jul 2019 report. We count the expected years of life subject to patient's age and gender. As the patient's age was recorded at the time of injury, we adjust it by adding two years in line with the time when the follow up study was conducted.

[8]For a detailed documentation about EQ-5D-3L questionnaire see https://euroqol.org/eq-5d-instruments/; and more information on use of EQ-5D-3L in injury setting see Derrett *et al.* (2009).

Considering that patients have positive time preference we follow the recommendation by the National Institute for Health and Care Excellence and discount NHB to current values at a rate of 3.5 % per year (Whitehead and Ali, 2010).

To express NHB in monetary terms we compute a *Net Monetary Benefit* (hereafter: NMB_i) including the direct costs. For each QALY gained from treatment, Huang *et al.* (2018) propose that an individual is willing to pay approximately A$67,000[9] for a sustained health improvement. NMB_i is then defined by

$$NMB_i = NHB_i * WTP^{QALY} - D_i \tag{3}$$

with patient's Willingness to Pay for each QALY, WTP^{QALY}, gained from treatment.

Return to work

We exclude any expenses related to loss of earned income from previous measures on the grounds that its inclusion alongside a utility measure of physical and mental well-being, that already includes aspects of work related utility, would lead to some double counting. However, health-related quality of life may not capture all of the utility related to paid work, thus we consider the return to work (RTW) as another relevant outcome to complement the risk adjustment model. Individuals who are unable to return to work after an injury experience greater physical difficulties and poorer mental health (Hoffman *et al.*, 2007; Iles *et al.*, 2008), thus understanding the main barriers resuming employment could potentially mitigate patients' economic losses in the long run. As the most prevalent group among road-traffic injuries are of working age who have many years to participate in the labour market, RTW is a crucial indicator of potential consequences of road injuries.

We model a binary response outcome - patient's likelihood to return to work within one-year of the injury. This information was collected in the questionnaire in the follow up study subject to the condition that patient worked (for income) before the injury. The working population in this study accounts for around 70% of all road-traffic injuries.

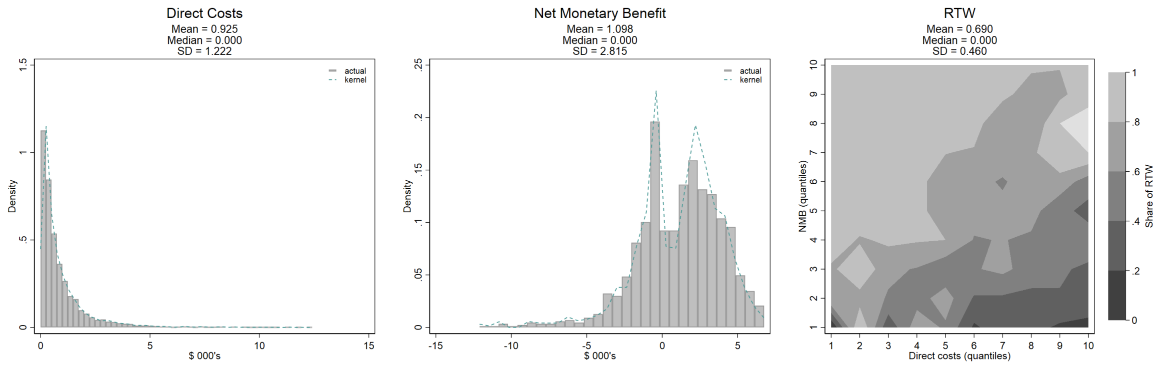
Observed Outcomes

Figure 1 presents all outcomes considered in this study. *Direct Costs* features a non-normal and positively skewed distribution and a particularly heavy tail. This is typical of health care spending where the vast majority of patients exhibit low costs and a few have

[9]The selected willingness to pay for one QALY is a high value in comparison with other international estimates. However, in this paper we are primarily interested in the relative contribution of risk factors, thus any change in the constant monetary value for QALY does not have an impact on the interpretation of the estimation results.

extremely high costs more than 10 times higher than the population average. The middle panel illustrates the distribution of *Net Monetary Benefit*. Recall, that this outcome is estimated for a sub-sample of patients who were admitted to the ICU.[10] Similar to *Direct Costs*, this distribution is non-normal. With a significant proportion of patients having a positive net benefit from treatment, the outcome is skewed left and has a lower kurtosis than the *Direct Costs*. Patients in the left tail either died soon after the discharge or reported very poor outcomes are classified by the EQ5D as "worse than death". The distribution has several modes that adds an additional complexity to the modelling.

Figure 1.
The Presentation of Prediction Outcomes



Note.— Figure presents the empirical distribution of *Direct costs* of injury on the left panel, the distribution of *Net Benefit* in the middle panel (both expressed in 100 thousands AU$) and the interaction of the former and the latter with *Return to Work* on the right panel. *Return to Work* is averaged over five quantiles.

The right panel of Figure 1 illustrates the variation in the binary outcome, *RTW*, with respect to the *Direct Costs* (presented on the x *axis*) and *Net Monetary Benefit* (presented on the y *axis*) using a sub-sample of patients who worked prior to the injury. Patients who were less likely to return to work within one-year of injury are presented by the darker shaded area while those who were more likely to return to work by the brighter areas. Figure 1 shows that in most cases patients who return to work have higher net benefits of treatment irrespective of their treatment costs suggesting that gains in QALYs outweigh higher costs. These patients presumably recover well or have sufficient support to return to work and other social activities. Patients with a more significant and long-lasting disability have the lowest probability of returning to work with low utility and high costs. However, a number of patients are situated away from this pattern and while they return to work, they still have comparatively low utility and high costs of initial treatment or ongoing supports. Predicting the pattern for these groups is the challenge

[10]This sub-sample has slight differences in estimated *Direct costs*. While the distributional properties and extreme values of treatment costs are statistically similar, the costs of care for these patients were on average higher due to an expensive treatment at the ICU.

in this paper.

4.2 Prediction methods

To perform the predictions we utilise an ensemble machine learning (ML) framework – the Super Learner algorithm. The Super Learner utilises various selected algorithms and builds a prediction function as a weighted combination of them. This makes the Super Learner a very versatile algorithm that often outperforms any single (ML) algorithm (Laan and Rose, 2011) and, similar to other ML algorithms, has the ability to model difficult data patterns with a better out-of-sample performance (Chu and Zhang, 2003; James *et al.*, 2013; Mullainathan and Spiess, 2017). The algorithm was first suggested by van der Laan *et al.* (2007) and has been used in previous research to predict various clinical patient outcomes such as post-traumatic stress disorder (Kessler *et al.*, 2014) and mortality in intensive care units (Pirracchio *et al.*, 2015). In the context of risk adjustment in health insurance markets, Rose (2016) has used the algorithm to predict health care expenditures and the construction of plan payment risk adjustment model in the U.S. market and Rose *et al.* (2017) focused on the prediction of the unprofitability of health insurance enrollees.

To allow for flexibility of the prediction function we consider both parametric as well as non-parametric methods. We set up the super learner based on the following menu of six prediction algorithms. First, we employ a regression model[11] with the set of all covariates shown in Table A.1. Due to the large number of intercorrelated covariates, least squares estimators might suffer from high variance and over prediction. We supplement our menu with several other algorithms based on regularisation methods. There are two types of regularization methods: *L1-regularization* augments the OLS loss function with a tuning parameter for all non-zero coefficients that penalizes the sum of coefficients' absolute values, whereas *L2-regularization* introduces a penalty for the sum of squared coefficients. While a high L2-penalty shrinks covariates towards zero, a very high L1-penalty sets them to be zero and in this way drops the covariate from the best fitting model (Tikhonov and Arsenin, 1977; Tibshirani, 1996; Zou and Hastie, 2005; Tikhonov *et al.*, 2013). We add two regularisation algorithms: lasso which is a penalized regression with a tuning parameter λ chosen via an internal 10-fold cross validation; and an elastic net regularized regression method with α and λ values selected via an internal 10-fold cross validation. In addition, we consider unpenalized lasso regression, that is a linear regression with a sub-set of covariates that are selected in the first step using L1-regularization.

[11] We employ linear regression models when modelling the continuous outcomes: *Direct Costs* and the *Net Monetary Benefit*; and logistic regression models when modelling the binary response outcome - *RTW*.

To better map non-linear relationships of the predicted outcomes we supplement our menu with several non-parametric methods. We first set up a tree-like model, a common machine-learning data structuring approach, that can be visualised as a flow-chart process. A Decision Tree splits the data into a set of subsamples defined by a classification rule and represented by a tree branch. Each branch could either lead to another sub-tree or have a leaf/terminal node with an assigned decision label. By applying this data splitting method each observation can be split using tree branches. To boost the accuracy and stability of the tree, the tree is pruned by setting constraints on the model parameters (Breiman *et al.*, 1984; Biggs *et al.*, 1991; Mola, 1998; Breiman, 2001; Scornet *et al.*, 2014; Maimon and Lior, 2014). We set a constraint of at least 50 observations in the terminal node and estimate a decision tree model. Additionally we supplement the analysis by introducing a Random Forest, an ensemble learning method that is based on sampling multiple decision trees (Maimon and Lior, 2014). Similarly we constrain each tree to have at least 50 observations in the terminal node and grow 500 random trees to estimate bootstrapped standard errors.

Employing this diverse set of algorithms, we follow the strategy outlined in Rose *et al.* (2017) and specify the super learner algorithm as follows:

$$\Psi(P_0) \;=\; \alpha_1\hat{\psi}_{reg} + \alpha_2\hat{\psi}_{L1reg} + \alpha_3\hat{\psi}_{lasso} + \alpha_4\hat{\psi}_{enet} + \alpha_5\hat{\psi}_{tree} + \alpha_6\hat{\psi}_{forest} + \epsilon \quad (4)$$

and estimate it using least squares method. We select a comprehensive collection of predictors to find the best performing prediction function. We choose predictors based on risk factors associated with the severity of injury and the cost of routine treatments that lead to poor health and labour market outcomes. The full set of covariates before regularization contains various patient demographic characteristics (age, gender, residential region, SES quintiles); clinical treatment-related characteristics (Injury Severity Score [ISS], Glasgow Comma Scale [GCS], number of days in ICU, number of ventilated hours, number of commorbidities as well as the comorbidity index); injury-specific controls (injury group, mechanism, activity, cause and place); health related behavioural covariates (if alcohol/drug/substance use, if any mental issues, if mood or neurotic disorders); admitted hospital, year and month fixed effects and a large set of binary main diagnosis variables.

4.3 Performance evaluation

Metrics

We employ several statistical metrics to evaluate the performance of each algorithm including the Super Learner. When modelling the continuous outcomes D_i and B_i we estimate the coefficient of determination, R^2, that evaluates the proportion of the variance explained by the selected set of covariates, and the mean squared error (MSE) measuring the prediction error (Wooldridge, 2020). These metrics are defined as follows:

$$MSE(\psi^j) = \frac{1}{J}\sum_j (y_i - \hat{\psi}_i^j)^2 \tag{5}$$

$$R^2(\psi^j) = \frac{\sum_{i,j}(y_i - \hat{\psi}_i^j)^2}{\sum_i (y_i - \bar{y})^2} \tag{6}$$

for each outcome y of patient i predicted by algorithm j. These metrics are evaluated based on the cross validation re-sampling procedure outlined below.

The prediction of a qualitative response, such as a binary outcome RTW, is known as a classification exercise in the ML literature. To evaluate the performance of the classification we set up a confusion matrix that represents a number of true positives (TP), a number of false positive (FP), a number of true negatives (TN) and a number of false negatives (FN) as outlined in the matrix below:

Table 1.
Confusion matrix

		Observed	
		Positive(1)	*Negative*(0)
Classified	*Positive*(1)	TP	FP
	Negative(0)	FN	TN

Based on the confusion matrix we estimate two accuracy measures for a binary classifier: the *Sensitivity* - a true positive rate that is a rate of correctly classified positive outcomes and the *Specificity* - a true negative rate or a rate of correctly classified negative outcomes defined as the following

$$Sensitivity = \frac{TP}{TP + FN} \tag{7}$$

$$Specificity = \frac{TN}{TN + FP} \tag{8}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

We plot a *receiver-operating-characteristics* (ROC) curve that evaluates the performance of the classifier summarized over various classifying thresholds. Due to the unbalanced structure of our selected binary response outcome we chose a set of classifying thresholds that fluctuates around the mean value of the outcome and calculate the accuracy measures using each of these thresholds. The ROC curve plots *Sensitivity* on the vertical axis against $(1 - Specifity)$ on the horizontal axis and represents the overall performance of the prediction by the area under the ROC curve. The larger the size of the area, the better the prediction performance. Using this, we classify the final prediction using a threshold with the largest area under the ROC curve. Additionally, we consider the *Accuracy*, that describes the prediction accuracy in percentages (James *et al.*, 2013; Maimon and Lior, 2014).

Cross-Validation

We perform two types of cross-validation. First, we perform an external cross-validation by randomly dividing the sample into two parts: *a training sample* that is used to fit each of the algorithms and *a validation sample* used to predict outcomes, validate the predictions and to estimate the Super Learner algorithm. This re-sampling procedure is based on 60:40 % split. Second, we implement 10-fold internal cross-validation when fitting regularized regressions and non-parametric models. We partition the training sample (selected via the external cross-validation) into inner training and validation sets and repeat the process 10 times (folds), with each of the randomly selected validation sample used only once to evaluate the prediction. The results from all folds are then averaged (van der Laan and Duboit, 2003; James *et al.*, 2013).

All results in this paper are presented for the validation sample selected via external cross-validation.

5 Prediction results

In this section we discuss prediction results using the set of algorithms outlined in Section 4. We first employ the training sample via 10-fold cross-validation to fit each single algorithm and then obtain predictions using a leave-out sample to validate the prediction performance. Results for both continuous outcomes are shown in Figure 2 and Figure 3. The upper panel illustrates the distribution of predicted values, whereas the bottom panel reports statistical measures for goodness of fit to evaluate the performance of each algorithm.

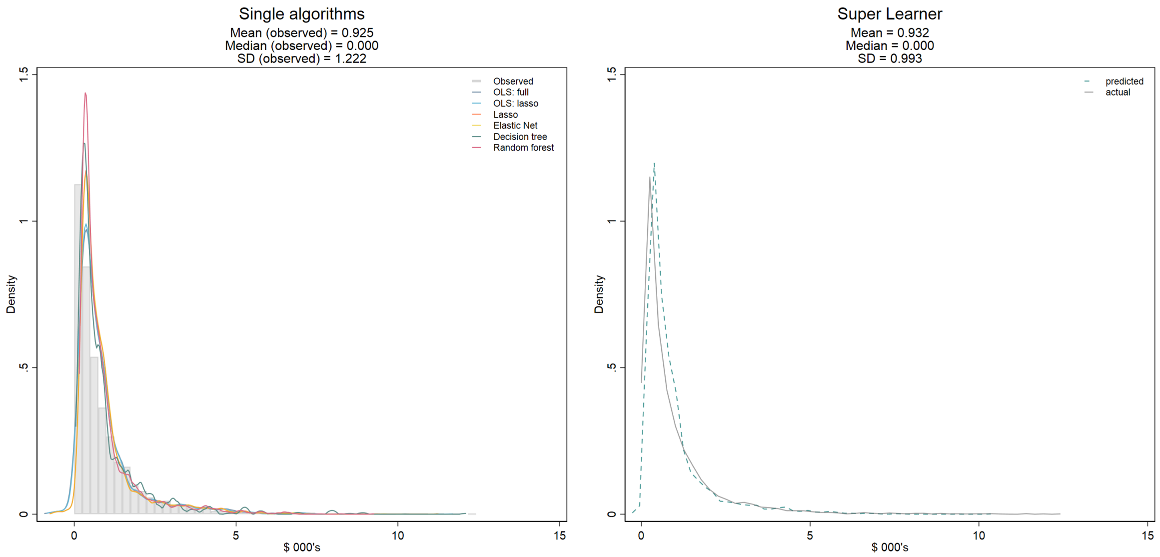
As shown in the upper left panel of Figure 2, all single algorithms have captured the positive skewness of the data, but several of them (particularly the OLS regression with a full set of covariates) (mis)predict negative values for patients with very low treatment costs. Only non-parametric models such as the Regression Tree and the Random Forest perform better in this particular feature by predicting only positive values.[12] However, only the Regression tree performs well when describing the long tail of the distribution that represents patients with very high costs. In the case of the Random Forest, a poor prediction of high costs comes at the expense of higher predicted levels of low costs patients visualized by a spike at low values. The bottom panel of Figure 2 reports statistical metrics as defined in (5) and (6) to evaluate the overall performance. Based on these measures, all single algorithms, except for the Regression Tree, perform similarly with an R^2 value equal to 0.64 and a MSE of 0.52. The Regression tree, while performing better in predicting the tail, failed to accurately predict low costs patients leading to a significantly lower R^2 value of 0.54 and a higher MSE of 0.68. The Elastic Net is the best performing single algorithm with R^2 value equal to 0.65 and a MSE of 0.52. The Super learner algorithm has remarkably outperformed all single algorithms considered in this paper with predictions shown in the upper right panel of the Figure 2 with R^2 value equal to 0.79 and MSE equal to 0.49.

A more perceptible difference in the performance of single algorithms is noted in Figure 3 that presents the prediction results for *Net Monetary Benefit*. Here parametric models perform better in quantifying the left tail of the distribution. However, particularly the Lasso regression and the Elastic Net predicts significantly higher levels of patients with low positive net benefits ranging from 0 to around \$4000. Non-parametric models are better of predicting the bimodal shape of the distribution. Based on the goodness of fit measures reported in the bottom panel, the Random Forest is the best performing single algorithm with a R^2 value equal to 0.45 and a MSE of 4.39. A similar performance is noted for the Lasso Regression (R^2: 0.44 and MSE: 4.46) and the Elastic Net (R^2: 0.44 and MSE: 4.45). Similarly as in the case of *Direct Costs* the Super Learner had best overall performance in predicting the distribution of the *Net Monetary Benefit* with the highest R^2 equal to 0.54 and the lowest MSE value equal to 4.18. For more detailed evaluation of each algorithm, refer to the additional evidence provided in the Appendix.[13]

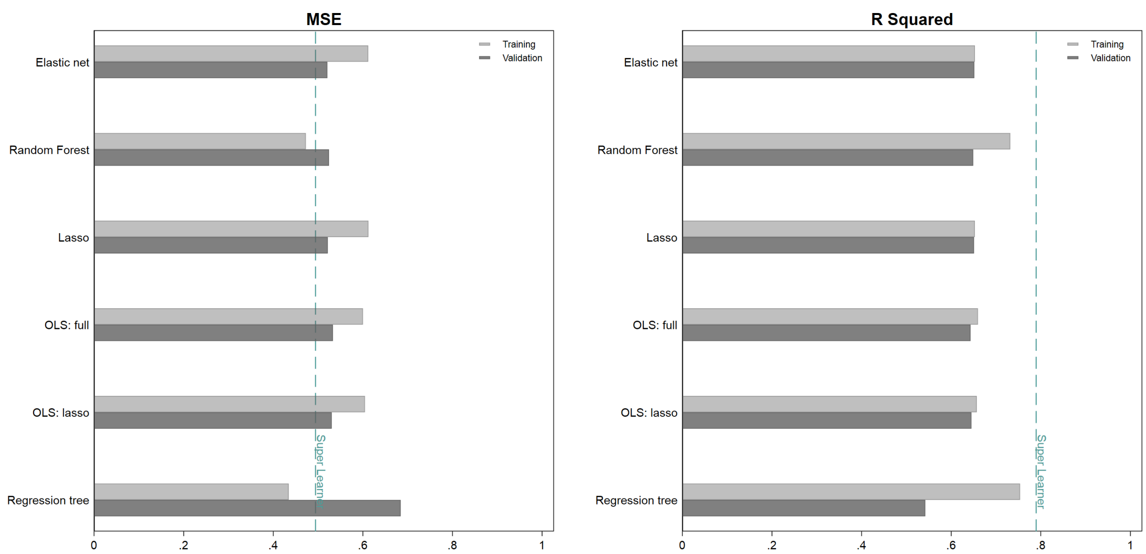
[12]Descriptive statistics of all predictions can be found in Table A.2 in the Appendix.

[13]Additional graphical evidence for the performance of each single algorithm as well as the Super Learner is presented in the Section 7. We show empirical distributions of predicted values in Figure A.1 and Figure A.2; quantile-quantile scatter plots that plots the quantiles of predicted values against the quantiles of the observed values in Figure A.4 and Figure A.5 and prediction errors for each observation in the scatter plots presented in Figure A.6 and Figure A.7.

Figure 2.
Outcome: *Direct Costs*



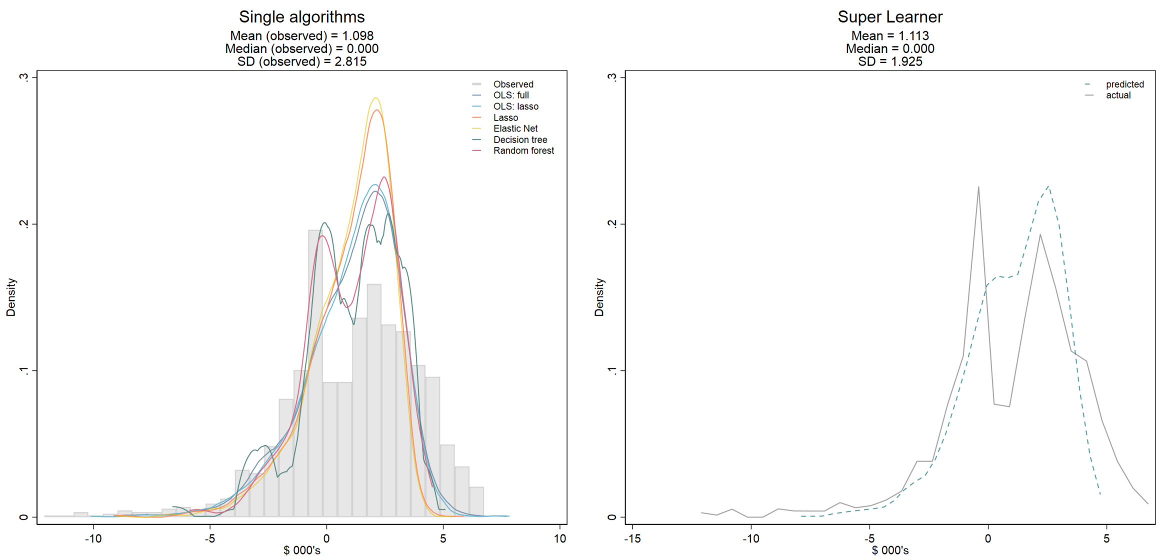
The distribution of observed outcomes and predictions



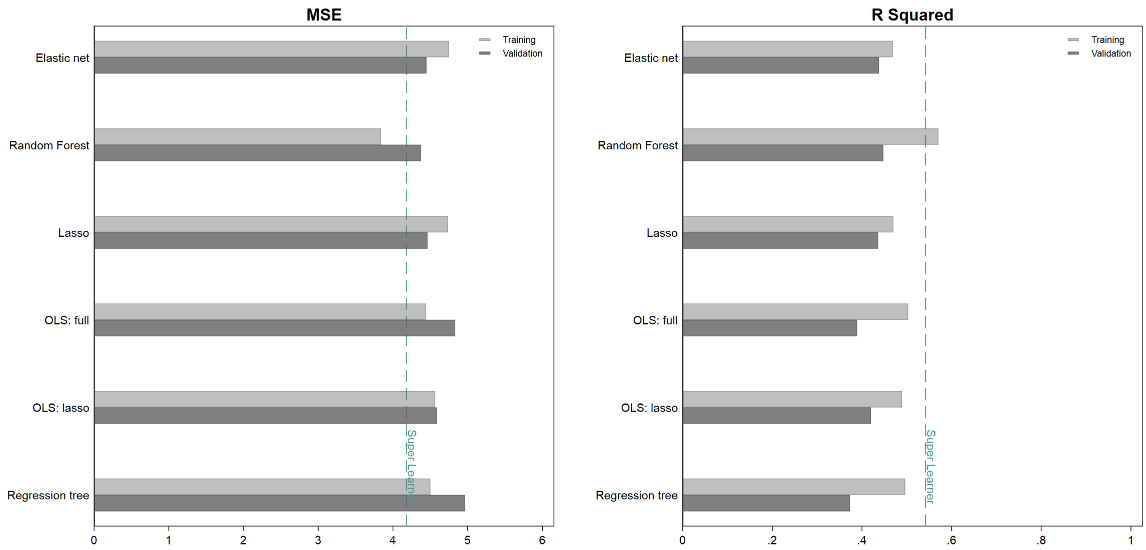
The Goodness of fit: Mean Squared Error and R^2

Note.— Figure presents the prediction results for the outcome *Direct Costs*: the empirical distributions (restricted to the validation sample only) of predicted values in the upper panel and statistical measures on Goodness of fit in the bottom panel. Both MSE and R^2 statistical measures to evaluate the Super Learner algorithm are estimated using a reduced form model and the validation sample.

FIGURE 3.
Outcome: *Net Benefit*



The distribution of observed outcomes and predictions



The Goodness of fit: Mean Squared Error and R^2

NOTE.— FIGURE PRESENTS THE PREDICTION RESULTS FOR THE OUTCOME *NET MONETARY BENEFIT*: THE EMPIRICAL DISTRIBUTIONS (RESTRICTED TO THE VALIDATION SAMPLE ONLY) OF PREDICTED VALUES IN THE UPPER PANEL AND STATISTICAL MEASURES ON GOODNESS OF FIT IN THE BOTTOM PANEL. BOTH MSE AND R^2 STATISTICAL MEASURES TO EVALUATE THE SUPER LEARNER ALGORITHM ARE ESTIMATED USING A REDUCED FORM MODEL AND THE VALIDATION SAMPLE. ANALYSIS SAMPLE IS RESTRICTED TO PATIENTS ADMITTED TO THE ICU.

Table 2 presents estimated contributions of each single algorithm as specified in (4). From the left to the right each column reports results from a full (indicated by *I*) and a reduced (indicated by *II*) form of the Super Learner for each outcome discussed in Section 4.1. The reduced form model is an equivalent to a model specified in (4) with a preceding step of selecting a subset of potential predictors using the L1-regularization method. Recall, that the specification of the Lasso regression is conceptually similar to a linear regression, but, in contrast to the traditional least-squares estimation, it augments the loss function and introduces a penalty for model parameters. Moreover, the Elastic Net is a related technique that has a flexibility to generate close to zero coefficients along with a variable selection when zero-valued coefficients are eliminated from the model. Thus, in practice, these models often make statistically similar predictions when selected tuning parameters for optimum performance cause one algorithm to resemble the other. Similarly as depicted in Figure 2 and Figure 3, we observe a strong collinear relationship[14] between these algorithms in Table 2 and, as a result, we perform a reduced form of the Super Learner with selected predictors via L1-regularization to avoid over-fitting the data.

Table 2.
Estimation of the Super Learner

	(1) Direct costs	(2) Direct costs	(3) NMB	(4) NMB	(5) RTW	(6) RTW
	I	II	I	II	I	II
Lasso	-1.095*		-0.752		27.12***	
	(-2.31)		(-1.64)		(6.80)	
Elastic net	1.024*		0.872	0.182	-48.21***	
	(2.08)		(1.80)	(0.88)	(-11.61)	
OLS/Logit: full	0.205	0.214	-0.213		2.086*	
	(1.50)	(1.57)	(-1.59)		(1.97)	
OLS/Logit: lasso	0.331	0.267	0.548*	0.278	8.219***	
	(1.76)	(1.92)	(2.51)	(1.75)	(5.89)	
Decision tree	0.0665**	0.0736**	0.0465	0.0583	-0.491	
	(2.70)	(3.02)	(0.68)	(0.86)	(1.75)	
Random forest	0.475***	0.449***	0.5568***	0.542***	11.88***	1.464***
	(10.44)	(11.36)	(5.94)	(5.72)	(19.55)	(28.01)
Observations	4650	4650	1379	1379	2948	2948
R^2	0.79	0.79	0.54	0.54		
MSE	0.494	0.494	4.170	4.179		
Accuracy					0.778	0.759

Note.— The predictions of *Direct Costs* and *Net Monetary Benefit* are based on a linear specification of lasso, Elastic Net and OLS regressions, while the predictions of *RTW* are estimated using a logistic regression specification. All estimations performed on validation sample. Columns *I* report the full specification, while Columns *II* report estimation results from reduced form models with a preceding step of the L-1 regularization to select the predictors. Models with *RTW* outcome restricted to a subsample of patient who worked prior to the injury and models with *Net Monetary Benefit* are restricted to patients who were admitted to the ICU during their hospital stay. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

[14] Further evidence on the collinearities between the predicted values are shown in the correlation matrixes presented in Figure A.9, Figure A.10, Figure A.11 and in the descriptive statistics of predictions reported in Table A.2 in the Appendix.

With respect to the outcome *Direct Costs* reported in Columns (1) and (2), we note that the reduced form model does not significantly alter the estimated contributions and does not affect the overall prediction performance with a steady R^2 value equal to 0.79 and a MSE of 0.49. The contribution of the Random Forest algorithm is the highest in magnitude with an estimate of 0.45, following by the OLS specification with lasso selected covariates and the OLS specification with a full set of covariates, respectively. The lowest, albeit statistically significant, contribution is estimated for the Regression Tree. Next, Columns (3) and (4) show estimation results on the outcome *Net Monetary Benefit* using a subsample of patients admitted to the ICU units. Unlike in the case of *Direct Costs*, in addition to the Lasso regression the prior L1-regularization suggests to eliminate the OLS full specification. It noticeably reduces the contribution of the Elastic Net and the OLS specification with lasso selected covariates, but only slightly alters the contribution of the Random forest, that is also the highest in magnitude and the strongest in terms of the statistical significance.

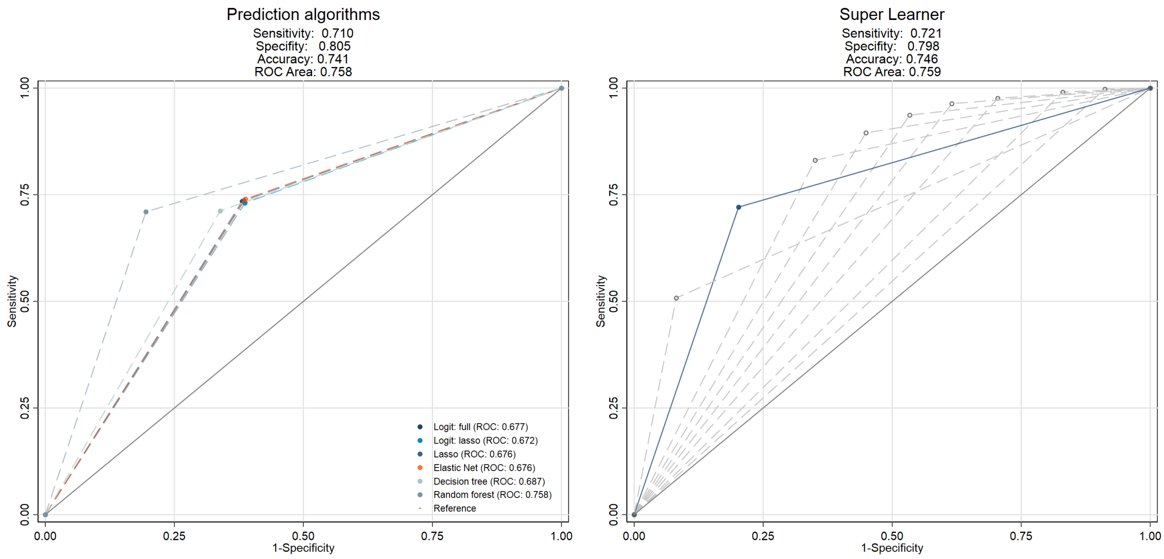
Lastly, Columns (5) and (6) in the Table 2 outline the Super Learner specification for classifying the binary response outcome *RTW*. Similarly, as in the case of previously discussed outcomes, the Super Learner is likely affected by high collinearities between single algorithms that is reflected by reversed signs of algorithms' contributions. It is expected, that due to high positive correlation, as in the case of the parametric models such as the logistic regression, Lasso and the Elastic Net, one algorithm withdraws the contribution from the other. Thus, using similar techniques, we perform L1-regularization and re-estimate the the Super Learner logistic regression in the reduced form. L1-regularization here suggests that selecting the Random Forest with a contribution of 1.464 leads to the best classifying model. This result goes in line with the existing literature on the performance of the Random Forest in the classification that has been demonstrated to have improved prediction accuracy in comparison to other supervised learning methods (Breiman, 2001; Svetnik *et al.*, 2003; Kuhn and Johnson, 2016).

Additional evidence on the classification performance by each single algorithm as well as the Super Learner is presented in Figure 4. Recall, that the ROC curve plots the *Sensitivity* on the y-axis against (*1-Specificity*) on the x-axis and the main measure of interest is the calculated area under the ROC curve. The larger the area, the better overall classification performance. The left panel of Figure 4 reports these measures for each single algorithm with a selected classification rule that leads to the largest area under the ROC curve, accordingly.[15] These results, again, demonstrate that the Random Forest algorithm outperforms other single algorithms and provide further support to the L1-

[15]To find the best fit for each single algorithm we perform a classification with a number of different thresholds fluctuating around the mean of the outcome. For more detailed evidence on the performance using various selected classification rules refer to Figure A.8 in the Appendix.

regularization performed in the reduced form specification. Only a slight improvement in the classification performance is noted by the Super Learner specification. In comparison to the Random Forest the estimated area under the ROC curve is nearly the same as in the case of the Super Learner, however the Accuracy reported in the panel header shows that the Super Learner has a slightly better prediction performance with an accuracy rate of approximately 75%. For more detailed information on the empirical distribution of predicted probabilities, refer to Figure A.3 in the Appendix.

FIGURE 4.
The ROC curves



NOTE.— FIGURE PRESENTS THE GOODNESS OF FIT MEASURES FOR THE PREDICTION OF *RTW*. IN THE LEFT PANEL EACH LINE INDICATES THE SPECIFICITY/SENSITIVITY MEASURES FOR EACH SINGLE ALGORITHM USING A CLASSIFICATION RULE THAT LEADS TO THE LARGEST AREA UNDER THE ROC CURVE. THE PANEL HEADER REPORTS THE STATISTICAL METRICS IN DETAIL FOR THE BEST PERFORMING SINGLE ALGORITHM. THE RIGHT PANEL PRESENTS THE PREDICTION RESULTS OF THE SUPER LEARNER USING SEVERAL SELECTED THRESHOLDS THAT FLUCTUATE AROUND THE MEAN VALUE OF THE OUTCOME. THE BLUE LINE INDICATES THE METRICS USING THE CLASSIFICATION RULE WITH THE BEST PERFORMANCE AND THE PANEL HEADER REPORTS ITS STATISTICAL EVALUATION METRICS. THE RESULTS ARE SHOWN ON THE VALIDATION SAMPLE AND ARE FURTHER RESTRICTED TO A SUB-SAMPLE OF PATIENTS WHO WORKED PRIOR TO THE INJURY

6 Risk adjustment errors by different groups

We next study the extent of prediction errors from the Super Learner in risk adjustment. Using the reduced form specification presented in Table 2 in Section 5, we analyse the differences between the observed and predicted outcomes for different group of patients. This provides a better understanding who are at risk for high costs and worse outcomes and may form a group of policy interest. Figure 5 and Figure 6 provide graphical illustration for outcomes *Direct Costs* and *Net Monetary Benefit*, respectively.

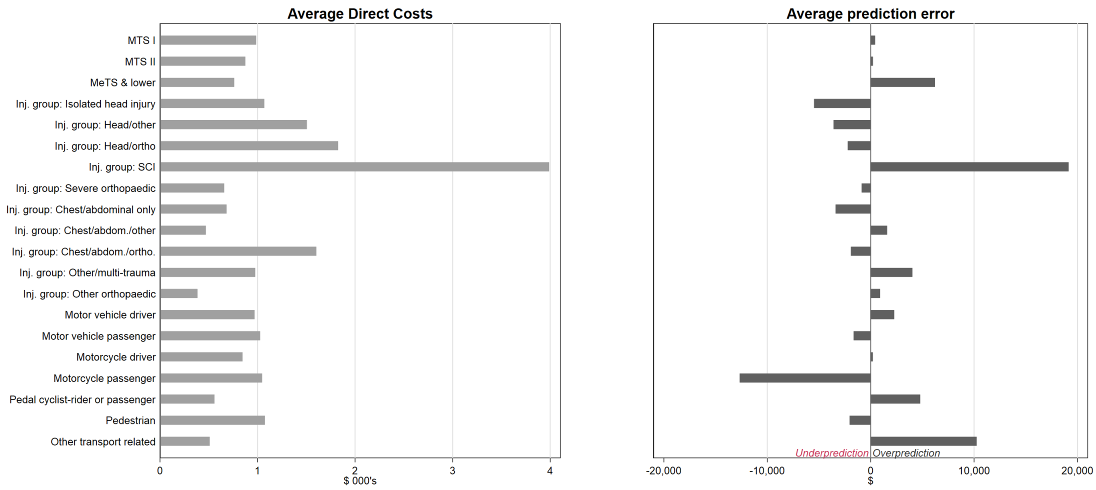
The upper panel of Figure 5 reports average injury costs by selected injury & treatment-related characteristics and their corresponding average prediction error. Treatment costs are on average higher at the Major Trauma Services, that is expected as this type of hospitals offers the highest level of trauma care in Victoria and, in most cases, treat very severely injured patients. A vast majority of patients (around 90%) are treated in a MTS with a considerable variation of type of injuries as well as patients characteristics providing enough evidence to make accurate predictions. The Super Learner demonstrates a high performance in predicting treatment costs for MTS, but exhibit a positive prediction error for patients treated in hospitals offering lower levels of trauma care, often located in more regional and rural areas. Regional variations in patients' clinical and socio-demographic characteristics as well as the quality of care provided are likely determinants of such differences in prediction errors. A similar pattern we observe when looking at the type of residential location and the socioeconomic status shown in the bottom panel of Figure 5. While costs of care are on average lower for patients living in lower socioeconomic status households and in metropolitan areas, the prediction error is much the same as for patients residing in more remote areas and does not exceed an overprediction of AU$ 2,000. More significant errors we note for patients living in higher socioeconomic status households and for those who were injured in Victoria but permanently residing outside the state and likely have more unobserved characteristics in the registry.

One of the most complex groups of patients to predict costs of care is patients who experience spinal cord injury, the most expensive road traffic injury. The variation in treatment costs for these patients is substantial, with some patients having very high costs and others – significantly lower. This is partially driven by lower chances of survival during hospital treatment as well as after discharge that causes total treatment costs to be lower than the algorithm predicts. In addition, age is also a significant factor as older individuals often have lower costs because of their lower chances of successful recovery that often result in assisted living without long-lasting and expensive rehabilitation services. For similar reasons the prediction error is high in absolute terms for patients with an isolated head injury. However, in this case, the prediction error is negative. With additional and more detailed clinical information about the extent and severity of the injury these errors could be addressed in the risk adjustment. This result signals the importance of discussed characteristics in the prediction of treatment costs, in particular for the youngest and the oldest groups of patients with the most severe injuries such as the spinal cord injury and those who are treated and reside in more rural and remote areas.

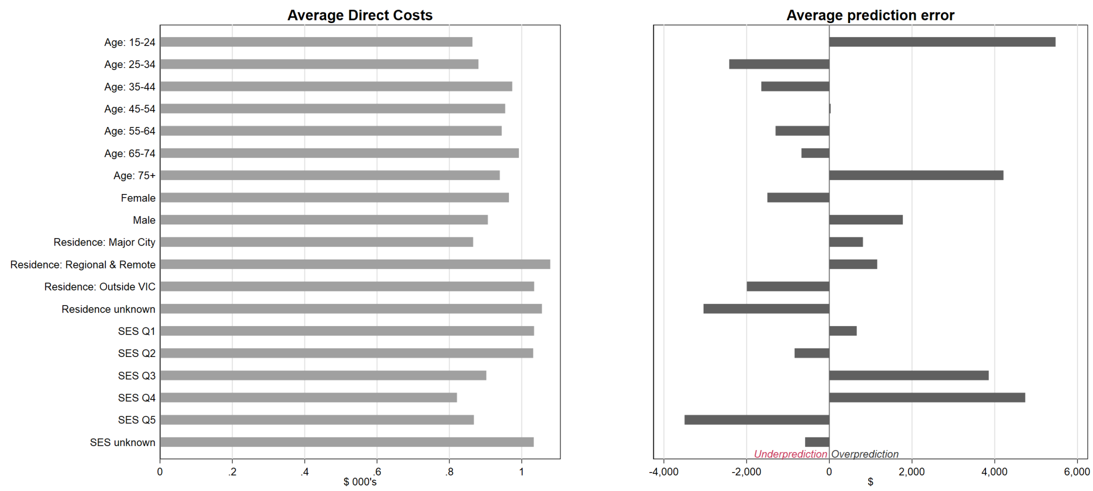
A comparison of averages in *Net Monetary Benefit* and it's respective prediction errors are presented in Figure 6 and tells a similar story; patients with spinal cord injuries have the highest benefits from treatment, but is one of the most complex groups of patients to obtain accurate predictions of. In addition, residents living in regional and remote areas as well as outside Victoria have higher prediction error. Unlike in the case of *Direct Costs*, the error is mostly positive, meaning that the algorithm predicts greater benefits from treatment than they are observed. It is an important result for policy makers targeting various groups of patients such as residents of regional and remote areas. These patients are more cost-burdened and, in addition, have lower observed benefits from treatment than residents living in metropolitan areas. The benefits from treatment are often overpredicted, thus it is important to consider this group of patients when applying recoupment adjustment to hospital payments. Moreover, it is worth noting, that the patient group with the lowest and negative benefits from treatment are patients with chest and abdominal injuries, suggesting the urgency to target this group for potential improvements in their recovery process. These results, again, illustrate a significant variation of costs and benefits among different group of patients defined by their type of injury and residential location. However, considering that only a moderate proportion (54%) of variance in the *Net Monetary Benefit* is explained using the Super Learner, we acknowledge the need for further observable characteristics to improve the accuracy of the prediction.

Similar to the error analysis of the continuous outcomes, Table 3 reports details about the classification errors made by the Super Learner when classifying *RTW*. Based on the classification error rate, the algorithm performs better in classifying a negative outcome. From a total of 944 patients who did not return to work, the algorithm correctly classified 753, implying a misclassification of every 5th patient. From a total of 2004 patients who returned to work, the algorithm suggested 1445 positive outcomes with an approximate error rate of 28%. The Super Learner appears to be more sensitive to the prediction of positive outcome that is more common for these patients. Two out of three patients successfully return to work after the injury, but the prediction algorithm suggests a slightly lower success rate. Among the group of misclassified patients approximately two-thirds reside in metropolitan areas and are the youngest group of patients aged 15-24 years (results not reported in the table). More commonly, the misclassification of a negative outcome is made for patients with severe and moderate orthopaedic injuries, while a positive outcome is more incorrectly specified for patients with head, chest and abdominal injuries. One possible reason for this result could be that patients with orthopaedic injuries to treat, though are not among the most costly groups of patients, require longer rehabilitation care. On the other hand, patients with head, chest and abdominal injuries are one of the most expensive injuries, suggesting a higher severity of a clinical case and poorer outcomes than otherwise expected. Interestingly, we do not observe any differences in misclassification across socioeconomic statuses.

Figure 5.
Prediction errors in *Direct Costs*



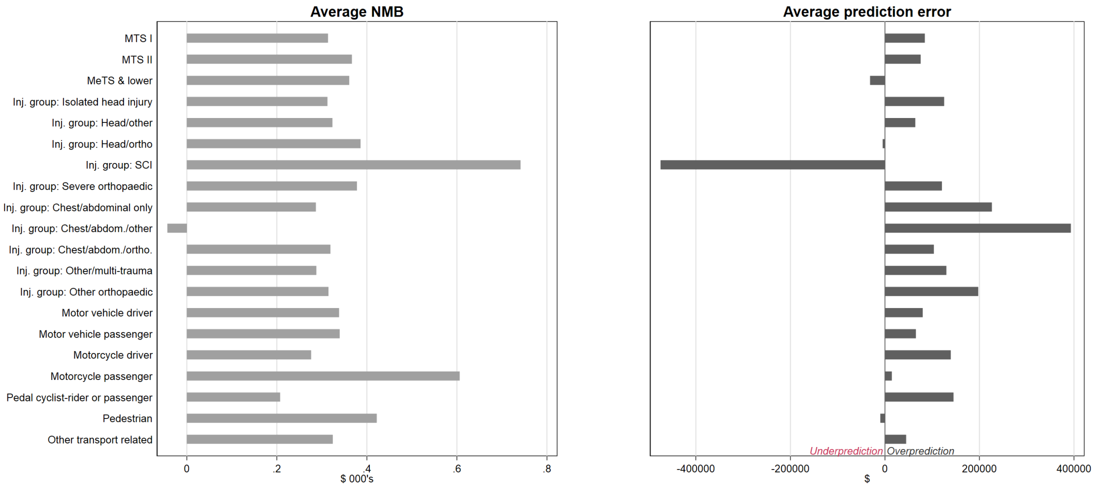
by injury & treatment-related characteristics



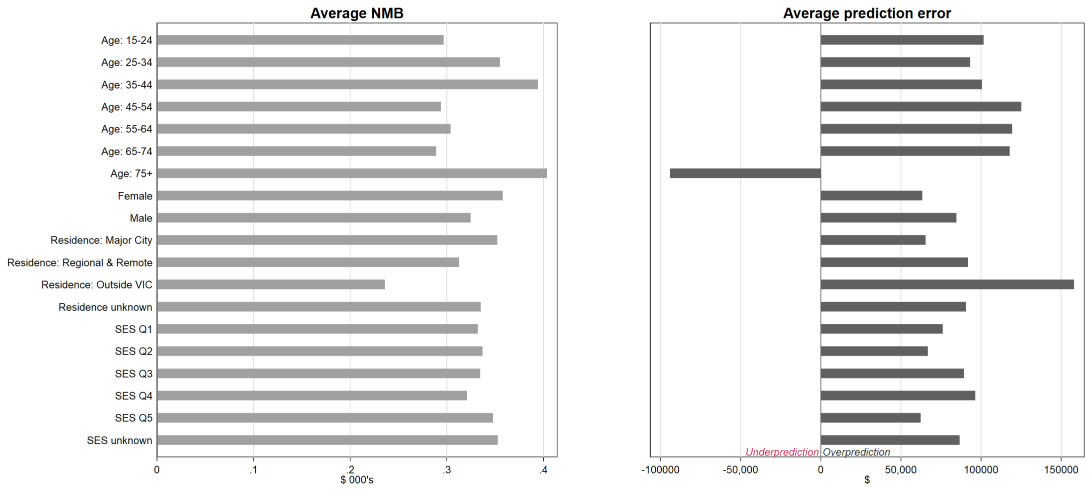
by patient characteristics

Note.— Figure presents average prediction errors by selected groups of patients that are made by the Super Learner algorithm in the prediction of *Direct Costs*. Here MTS refers to the Major Trauma Services - the highest level of trauma care in Victoria. Patients considered in this analysis were admitted to two different MTS, that were de-identified using indicator I and II. MeTS refers to Metropolitan Trauma services; lower levels of care include regional trauma services and rural healthcare services. SES refers to the Social Economic Status as defined buy the Index of Relative Socio-Economic Advantage and Disadvantage. All prediction errors are reported on the validation sample.

Figure 6.
Prediction errors in *Net Monetary Benefit*



by injury & treatment-related characteristics



by patient characteristics

Note.— Figure presents average prediction errors by selected groups of patients that are made by the Super Learner algorithm in the prediction of *Net Monetary Benefit*. Here MTS refers to the Major Trauma Services - the highest level of trauma care in Victoria. Patients considered in this analysis were admitted to two different MTS, that were de-identified using indicator I and II. MeTS refers to Metropolitan Trauma services; lower levels of care include regional trauma services and rural healthcare services. SES refers to the Social Economic Status as defined buy the Index of Relative Socio-Economic Advantage and Disadvantage. All prediction errors are reported on the validation sample. Analysis sample is restricted to patients admitted to the ICU.

Table 3.
Prediction errors in *RTW*

	(1) Observed: 0	(2) Observed: 1	(3) Total
Classified: 0	753	559	1312
Classified: 1	191	1445	1636
Total	944	2004	2948
Error rate	**0.20**	**0.28**	**0.25**

Note.— Table presents the classification errors made by the Super Learner algorithm. Error rate denotes a share of misclassified outcomes. Estimation performed on validation sample and restricted to a subsample of patient who worked prior to the injury.

7 Conclusion

In this paper we employ supervised machine learning algorithms to construct a powerful risk adjustment model for injury related resource use. In addition to traditional health care costs we consider risk adjustment to other policy relevant outcomes that are particularly important for the understanding of the societal value of health and work in recovery from injury. We employ a comprehensive patient-level dataset of Victorian State Trauma Registry that incorporates all major trauma patients in Victoria. We link this dataset to detailed insurance claims records provided by the Transport Accident Commission and compute health care costs for each patient who suffered a major trauma in a road-traffic related injury. To better inform policy makers about the potential patient outcomes after treatment and account for potential moral hazard, we estimate the net monetary benefit gained from treatment and support funded by TAC insurance. This measure relies on the concept of Quality Adjusted Life Years used in the cost-effectiveness literature and offers a tool for interpretation of the societal value of health (Stinnett and Mullahy, 1998). Lastly, to inform policy makers about potential losses in labour market we consider patient's probability to return to work after suffering from a major trauma. To predict the latter outcomes we utilise both parametric and non-parametric algorithms to construct an ensemble machine learning framework - the Super Learner - and predict the economic consequences of road traffic injury: *Direct Costs*, *Net Monetary Benefit* and *Return to Work*.

Our findings demonstrate that the Super Learner is effective and performs well in predicting all outcomes considered in this paper. In addition to high overall performance in predicting outcomes for patients with a mild and a moderate severity of an injury, it performs well in describing patients with uncommon characteristics and is able to classify patients with the highest health care costs and the lowest net benefits gained from treatment. The algorithm only slightly outperforms the Random Forest prediction of a binary response outcome that is often referred to as the best performing classification algorithm in the machine learning literature. We extend our prediction analysis by examining in detail the Super Learner's performance by different groups of patients. This analysis reveals further information about sensitive groups and has a strong relevance for future budget planning and reimbursement for health care providers. Injury groups such as a spinal cord injury and chest and abdominal injuries are one of the most complex groups to get an accurate prediction of potential future costs indicating a need of particularly detailed information about the treatment of these patients to ensure an adequate remuneration. Average cost and net benefits from treatment vary widely across injury types and patient characteristics but in a way that is largely predictable. The algorithms used here predict over half of the variation in cost and net benefits suggesting that adjustment to capitation or prospective payments are feasible. Thus, our results suggest that payments for health care providers should take into account these risk factors when adjusting the reimbursement system. The Super Learner is a powerful tool in predicting economic consequences of the road traffic injury including the health care spending, giving a strong advice to be considered in health policies for the near future.

References

Arandjelović, O. (2015). Prediction of health outcomes using big (health) data. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 2543–2546.

Beck, B., Bray, J., Cameron, P., Cooper, D. and Gabbe, B. (2016). Trends in severe traumatic brain injury in Victoria, 2006–2014. *The Medical Journal of Australia*, **204**, e1–e6.

Bertsimas, D., Bjarnadóttir, M., Kane, M., Kryder, J., Pandey, R., Vempala, S. and Wang, G. (2008). Algorithmic prediction of health care costs and discovery of medical knowledge. *Operations Research*, **56** (6), 1382–1392.

Biggs, D., Ville, B. D. and Suen, E. (1991). A method of choosing multiway partitions for classification and decision trees. *Journal of Applied Statistics*, **18** (1), 49–62.

Breiman, L. (2001). Random forests. *Machine Learning*, **45**, 5–32.

—, Friedman, J., Stone, C. and Olshen, R. (1984). *Classification and Regression Trees*. The Wadsworth and Brooks/Cole Statistics: Probability Series, Taylor & Francis.

Burnham, J. P., Lu, C., Yaeger, L. H., Bailey, T. C. and Kollef, M. H. (2018). Using wearable technology to predict health outcomes: a literature review. *Journal of the American Medical Informatics Association*, **25** (9), 1221–1227.

Chen, S., Kuhn, M., Prettner, K. and Bloom, D. E. (2019). The global macroeconomic burden of road injuries: estimates and projections for 166 countries. *The Lancet*, **3** (9), E390–E398.

Chu, C.-W. and Zhang, G. P. (2003). A comparative study of linear and nonlinear models for aggregate retail sales forecasting. *International Journal of Production Economics*, **86** (3), 217–231.

Cucciare, M. and o'Donohue, W. (2006). Predicting Future Healthcare Costs: How Well Does Risk-Adjustment Work? *Journal of Health Organization and Management*, **20**, 150–162.

Curtis, K., Lam, M., Mitchell, R., Black, D., Taylor, C., Dickson, C., Jan, S., Palmer, C. S., Langcake, M. and Myburgh, J. (2014). Acute costs and predictors of higher treatment costs of trauma in New South Wales, Australia. *Injury*, **45** (1), 279–284.

Deb, P. and Burgess, J. (2003). A quasi-experimental comparison of econometric models for health care expenditures. *Economics Working Paper Archive at Hunter College*, **212**.

Department of Health and Human Services (Feb 2014). *Trauma towards 2014 - Review and future directions of the Victorian State Trauma System*. Retrieved from:

https://www2.health.vic.gov.au/about/publications/policiesandguidelines/Trauma-towards-2014---Review-and-future-directions-of-the-Victorian-State-Trauma-System# [accessed 14.04.2020], Victoria State Government.

Department of Health and Human Services (Jul 2014). *Victorian State Trauma System and Registry Report 2014-15*. Retrieved from: https://www2.health.vic.gov.au/about/publications/annualreports/victorian-state-trauma-registry-summary-report-2014-15 [accessed 14.04.2020], Victoria State Government.

Derrett, S., Black, J. and Herbison, G. (2009). Outcome after injury – a systematic literature search of studies using the EQ-5D. *Journal of Trauma*, **67** (4), 883–890.

Dixon, J., Smith, P., Gravelle, H., Martin, S., Bardsley, M., Rice, N., Georghiou, T., Dusheiko, M., Billings, J., Lorenzo, M. D. and Sanderson, C. (2011). A person based formula for allocating commissioning funds to general practices in england: development of a statistical model. *BMJ*, **343**.

Duan, N. (1983). Smearing estimate: a nonparametric retransformation method. *Journal of the American Statistical Association*, **78**, 605–610.

Duckett, S. J. (1998). Casemix funding for acute hospital inpatient services in Australia. *Medical Journal of Australia*, **169** (S1), S17–S21.

Economic Connections (2017 September). *Cost of Road Trauma in Australia*. Retrieved from: https://www.aaa.asn.au/knowledge-centre/page/7/?category&tag=reports [accessed 14.04.2020], Australian Automobile Association.

Einav, L., Finkelstein, A., Kluender, R. and Schrimpf, P. (2016). Beyond Statistics: The Economic Content of Risk Scores. *American Economic Journal: Applied Economics*, **8** (2), 195–224.

Ellis, R., Martins, B. and Rose, S. (2018). *Risk Adjustment for Health Plan Payment*, pp. 55–104.

Ellis, R. P. and McGuire, T. G. (1996). Hospital response to prospective payment: Moral hazard, selection, and practice-style effects. *Journal of Health Economics*, **15** (3), 257–277.

Gabbe, B., Lyons, R., Fitzgerald, M., Judson, R., Richardson, J. and Cameron, P. (2014). Reduced population burden of road transport–related major trauma after introduction of an inclusive trauma system. *Annals of surgery*, **261**.

—, Simpson, P., Sutherland, A., Wolfe, R., Fitzgerald, M., Judson, R. and Cameron, P. (2012). Improved functional outcomes for major trauma patients in a regionalized, inclusive trauma system. *Annals of Surgery*, **255**, 1009–1015.

Gilleskie, D. and Mroz, T. (2004). A flexible approach for estimating the effects of covariates on health expenditures. *Journal of Health Economics*, **23**, 391–418.

Hoffman, B., Papas, R., Chatkoff, D. and Kerns, R. (2007). Meta-analysis of psychological interventions for chronic low back pain. *Health psychology : official journal of the Division of Health Psychology, American Psychological Association*, **26** (1), 1–9.

Holbrook, T., Hoyt, D. and Anderson, J. (2001a). The impact of major in-hospital complications on functional outcome and quality of life after trauma. *The Journal of trauma*, **50**, 91–95.

—, —, Coimbra, R., Potenza, B., Sise, M. and Anderson, J. (2005). Long-term posttraumatic stress disorder persists after major trauma in adolescents: New data on risk factors and functional outcome. *The Journal of trauma*, **58**, 764–769.

—, —, Stein, M. and Sieber, W. (2001b). Perceived threat to life predicts posttraumatic stress disorder after major trauma: Risk factors and functional outcome. *The Journal of trauma*, **51**, 287–292.

Holtslag, H., Beeck, E., Lindeman, E. and Leenen, L. (2007). Determinants of long-term functional consequences after major trauma. *The Journal of trauma*, **62**.

Huang, L., Frijters, P., Dalziel, K. and Clarke, P. (2018). Life satisfaction, qalys, and the monetary value of health. *Social Science & Medicine*, **211**, 131–136.

Iezzoni, I. (2012). *Risk Adjustment for Measuring Healthcare Outcomes*. IL: Health Administration Press, 4th edn.

IHPA (2019). *Pricing and funding for safety and quality. Risk adjustment model for Hospital Acquired Complications*. Retrieved from: https://www.ihpa.gov.au/sites/default/files/publications/pricing_and_funding_for_safety_and_quality_-_risk_adjustment_model_for_hospital_acquired_complications_2018-19.pdf [accessed 02.11.2020].

Iles, R., Davidson, M. and Taylor, N. (2008). Psychosocial predictors of failure to return to work in non-chronic non-specific low back pain: A systematic review. *Occupational and Environmental Medicine*, **65**, 507–517.

Ip, R. Y., Dornan, J. and Schentag, C. (1995). Traumatic brain injury: factors predicting return to work or school. *Brain Injury*, **9** (5), 517–532.

James, G., Witten, D., Hastie, T. and Tibshirani, R. (2013). *An Introduction to Statistical Learning: with Applications in R*. Springer.

Jones, A. M. (2000). Health econometrics. In A. J. Culyer and J. P. Newhouse (eds.), *Handbook of Health Econometrics*, vol. 1, Elsevier: Amsterdam, pp. 265–344.

— (2011). Models for health care. In M. P. Clements and D. F. Hendry (eds.), *Oxford Handbook of Economic Forecasting*, Oxford: Oxford University Press, pp. 625–654.

— and Lomas, J. (2016). A quasi-monte-carlo comparison of parametric and semiparametric regression methods for heavy-tailed and non-normal data: an application to healthcare costs. *Journal of the Royal Statistical Society*, **179** (Part 4), 951–974.

—, — and Rice, N. (2014). Applying beta-type size distributions to healthcare cost regressions. *Journal of Applied Econometrics*, **29**, 649–670.

—, — and — (2015). Healthcare cost regressions: going beyond the mean to estimate the full distribution. *Health Economics*, **24**, 1192–1212.

Jürges, H. and Köberlein, J. (2015). What explains DRG upcoding in neonatology? The roles of financial incentives and infant health. *Journal of Health Economics*, **43**, 13–26.

Kan, H. J., Kharrazi, H., Chang, H.-Y., Bodycombe, D., Lemke, K. and Weiner, J. (2019). Exploring the use of machine learning for risk adjustment: A comparison of standard and penalized linear regression models in predicting health care costs in older adults. *PLOS ONE*, **14**, e0213258.

Kapur, K., Young, A. and Murata, D. (2000). Risk adjustment for high utilizers of public mental health care. *The Journal of Mental Health Policy and Economics*, **3**, 129–137.

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. C., 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. and 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*, **13** (3), 265–274.

Kong, W., Tang, D., Xiaoyuan, L., Yu, I., Liang, Y. and He, Y. (2011). Prediction of return to work outcomes under an injured worker case management program. *Journal of Occupational Rehabilitation*, **22**, 230–240.

Kuhn, M. and Johnson, K. (2016). *Applied Predictive Modeling*. 5th edition, Springer Science+Business Media New York 2013.

Laan, M. and Rose, S. (2011). *Targeted Learning: Causal Inference for Observational and Experimental Data*. Berlin, Heidelberg, New York: Springer.

Lahiri, B. (2014). *Predicting Healthcare Expenditure Increase for an Individual from Medicare Data*. Retrieved from: https://www.academia.edu/7836580/Predicting_Healthcare_Expenditure_Increase_for_an_Individual_from_Medicare_Data [accessed 21.04.2020].

Maimon, O. and Lior, R. (2014). *Data Mining With Decision Trees: Theory And Applications (2nd Edition)*. Series In Machine Perception And Artificial Intelligence, World Scientific Publishing Co. Pte. Ltd: Singapore.

Manning, W. G., Basu, A. and Mullahy, J. (2005). Generalized modeling approaches to risk adjustment of skewed outcomes data. *Journal of Health Economics*, **24** (3), 465–488.

McCullough, A. L., Haycock, J. C., P., F. D. and Moran, C. G. (2014). Early management of the severely injured major trauma patient. *British Journal of Anaesthesia*, **113** (2), 234–241.

McDonald, J. B., Sorensen, J. and Turley, P. A. (2013). Skewness and kurtosis properties of income distribution models. *Review of Income and Wealth*, **59** (2), 360–374.

Mola, F. (1998). Classification and regression trees software and new developments. In A. Rizzi, M. Vichi and H.-H. Bock (eds.), *Advances in Data Science and Classification*, Springer Berlin Heidelberg, pp. 311–318.

Mullahy, J. (2009). Econometric modeling of health care costs and expenditures: A survey of analytical issues and related policy considerations. *Medical Care*, **47** (7), S104–S108.

Mullainathan, S. and Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, **31**, 87–106.

Nielsen, M. B., Madsen, I. E. H., Bültmann, U., Christensen, U., Diderichsen, F. and Rugulies, R. (2010). Predictors of return to work in employees sick-listed with mental health problems: Findings from a longitudinal study. *European Journal of Public Health*, **21**, 806–811.

Pirracchio, R., Petersen, M. L., Carone, M., Rigon, M. R., Chevret, S. and van der Laan, M. J. (2015). Mortality prediction in intensive care units with the Super ICU Learner Algorithm (SICULA): a population-based study. *The Lancet. Respiratory Medicine*, **3** (1), 42–52.

Pirson, M., Martins, D., Jackson, T., Dramaix, M. and Leclercq, P. (2006). Prospective casemix-based funding, analysis and financial impact of cost outliers in all-patient refined diagnosis related groups in three Belgian general hospitals. *The European Journal of Health Economics*, **7**, 55–65.

Pyrkov, T. V., Slipensky, K., Barg, M., Kondrashin, A., Zhurov, B., Zenin, A., Pyatnitskiy, M., Menshikov, L., Markov, S. and Fedichev, P. O. (2018). Extracting biological age from biomedical data via deep learning: too much of a good thing? *Scientific Reports*, **8** (5210).

Rose, S. (2016). A machine learning framework for plan payment risk adjustment. *Health Services Research*, **51**, 2358–2374.

—, Bergquist, S. L. and Layton, T. J. (2017). Computational health economics for identification of unprofitable health care enrollees. *Biostatistics*, **18** (4), 682–694.

Schreyögg, J., Stargardt, T., Tiemann, O. and Busse, R. (2006). Methods to determine reimbursement rates for diagnosis related groups (DRG): A comparison of nine European countries. *Health Care Management Science*, **9**, 215–223.

Scornet, E., Biau, G. and Vert, J.-P. (2014). Consistency of random forests. *The Annals of Statistics*, **43**.

Sluys, K., Häggmark, T. and Iselius, L. (2005). Outcome and quality of life 5 years after major trauma. *The Journal of Trauma*, **59**, 223–232.

Stinnett, A. A. and Mullahy, J. (1998). Net health benefits: A new framework for the analysis of uncertainty in cost-effectiveness analaysis. *Medical Decision Making*, **18** (2), S68–S80.

Svetnik, V., Liaw, A., Tong, C., Culberson, J. C., Sheridan, R. P. and Feuston, B. P. (2003). Random forest: a classification and regression tool for compound classification and QSAR modeling. *Journal of Chemical Information and Computer Sciences*, **43** (6), 1947–1958.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, **58** (1), 267–288.

Tikhonov, A., Goncharsky, A., Stepanov, V. and Yagola, A. (2013). *Numerical Methods for the Solution of Ill-Posed Problems*. Mathematics and Its Applications, Springer Netherlands.

Tikhonov, A. N. and Arsenin, V. Y. (1977). *Solutions of Ill-Posed Problems*. Winston & Sons, Washington.

Transport Accident Commision (2018/19). *TAC Annual Report*. Retrieved from: https://www.tac.vic.gov.au/about-the-tac/media-and-events/news-and-events/2019/tac-annual-report-tabled-in-parliament [accessed 14.04.2020], Victoria State Government.

van der Laan, M. J. and Duboit, S. (2003). Unified cross-validation methodology for selection among estimators and a general cross-validated adaptive epsilon-net estimator: finite sample oracle inequalities and examples. *Technical Report 130: Division of Biostatistics, University of California, Berkeley*.

—, Polley, E. C. and Hubbard, A. E. (2007). Super learner. *Statistical applications in genetics and molecular biology*, **6**.

Van Patten, R., , Z. C., Merz, Mulhauser, K. and Fucetola, R. (2016). Multivariable prediction of return to work at 6-month follow-up in patients with mild to moderate acute stroke. *Archives of Physical Medicine and Rehabilitation*, **97** (12), 2061–2067.

Victorian State Trauma Registry (Apr 2014). *Victorian State Trauma Registry Special Focus Report. Review of the Case Review Group Indicators – Addendum to Report*. Retrieved from: https://trauma.reach.vic.gov.au/sites/default/files/VSTS%20guideline%20Ver%202.0.pdf [accessed 14.04.2020], Victoria State Government.

Whitehead, S. J. and Ali, S. (2010). Health outcomes in economic evaluation: the QALY and utilities. *British Medical Bulletin*, **96** (1), 5–21.

WHO (2018). *Global Status Report on Road Safety 2018*. Retrieved from: https://www.who.int/publications-detail/global-status-report-on-road-safety-2018 [accessed 14.04.2020], World Health Organisation.

Wooldridge, J. M. (2020). *Introductory Econometrics: A Modern Approach*. 7th edition, Boston: Cengage.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B*, **67** (5), 768–768.

Appendix: Additional tables and figures

TABLE A.1.
Descriptive statistics of prediction covariates

	Mean	Sd	Min	Max
— Sample restriction variables —				
Training sample	0.60	0.49	0	1
If admitted to the ICU	0.30	0.46	0	1
If worked prior to the incident	0.67	0.47	0	1
— Outcomes —				
Direct costs of injury, $ 000's	0.95	1.29	0	17
Net Monetary Benefit, $ 000's	2.50	2.57	-17.45	6.75
Return to Work within 1 year	0.68	0.47	0	1
— for the computation of outcomes —				
Net Health Benefit	0.25	0.32	-0.22	1.43
Quality-Adjusted Life years	24.10	17.99	0	67
Quality-Adjusted Life years (discounted)	5.15	2.92	0	10
— Patient demographics —				
If male	0.68	0.47	0	1
Age: 15-24	0.21	0.41	0	1
Age: 25-34	0.18	0.38	0	1
Age: 35-44	0.16	0.36	0	1
Age: 45-54	0.14	0.35	0	1
Age: 55-64	0.11	0.32	0	1
Age: 65-74	0.09	0.28	0	1
Age: 75+	0.11	0.31	0	1
Education: Tertiary	0.46	0.50	0	1
Education: Secondary	0.42	0.49	0	1
Education: Primary	0.03	0.16	0	1
Education: Other	0.01	0.10	0	1
Education: Unknown	0.09	0.28	0	1
Marital Status: Single - Never married	0.15	0.35	0	1
Marital Status: Currently married	0.17	0.37	0	1
Marital Status: Separated	0.02	0.12	0	1
Marital Status: Divorced	0.02	0.15	0	1
Marital Status: Widowed	0.03	0.16	0	1
Marital Status: Living with partner (defacto relationship)	0.06	0.23	0	1
Marital Status: Partnered but not living together	0.03	0.18	0	1
Marital Status: Other	0.00	0.01	0	1
Marital Status: Unknown	0.53	0.50	0	1
Type of residence: Major City	0.71	0.45	0	1
Type of residence: Regional & Remote	0.24	0.43	0	1
Type of residence: Outside VIC	0.03	0.16	0	1
Type of residence: Unknown	0.02	0.14	0	1
Region: Barwon South West	0.09	0.28	0	1
Region: Gippsland	0.06	0.23	0	1
Region: Grampians	0.04	0.19	0	1
Region: Hume	0.06	0.24	0	1
Region: Loddon Mallee	0.04	0.20	0	1
Region: Eastern Metro	0.14	0.34	0	1
Region: Northern Metro	0.16	0.37	0	1
Region: Southern Metro	0.23	0.42	0	1
Region: Western Metro	0.14	0.35	0	1

Continued on next page

Table A.1 – *Continued from previous page*

	Mean	Sd	Min	Max
Region: Overseas	0.00	0.04	0	1
Region: Unknown in Victoria	0.01	0.09	0	1
Region: Unknown outside Victoria	0.00	0.01	0	1
Region: New South Wales	0.01	0.12	0	1
Region: Queensland	0.00	0.06	0	1
Region: South Australia	0.00	0.06	0	1
Region: Western Australia	0.00	0.04	0	1
Region: Tasmania	0.00	0.03	0	1
Region: Northern Territory	0.00	0.03	0	1
Region: Australian Capital Territory	0.00	0.03	0	1
SES: Q1	0.18	0.38	0	1
SES: Q2	0.17	0.37	0	1
SES: Q3	0.19	0.39	0	1
SES: Q4	0.20	0.40	0	1
SES: Q5	0.25	0.43	0	1
SES: Unknown	0.02	0.14	0	1
— *Clinical treatment-related* —				
If patient died in hospital	0.04	0.19	0	1
ISS < 12	0.28	0.45	0	1
ISS > 12	0.53	0.50	0	1
ISS unknown	0.19	0.39	0	1
CCI $= 0$	0.71	0.45	0	1
CCI $= 1$	0.22	0.41	0	1
CCI > 1	0.07	0.26	0	1
Days in ICU	2.22	5.48	0	140
Hours ventilated	29.56	99.43	0	3089
— *Injury-related characteristics* —				
MTS I	0.52	0.50	0	1
MTS II	0.41	0.50	0	1
MeTS & lower	0.07	0.26	0	1
Unintentional	0.97	0.16	0	1
Intentional-self harm	0.01	0.11	0	1
Assault/Maltreatment	0.00	0.06	0	1
Intent cannot be determined	0.00	0.07	0	1
Intentional-other	0.01	0.08	0	1
AIS: Upper extremity	0.75	0.94	0	4
AIS: Lower extremity	1.16	1.32	0	5
AIS: Thorax	0.47	0.50	0	1
AIS: Head	0.96	1.49	0	6
AIS: Spine	0.86	1.18	0	6
AIS: Face	0.39	0.72	0	4
AIS: Abdominal pelvis	0.50	1.08	0	6
AIS: Neck	0.07	0.42	0	5
AIS: External burns	0.10	0.32	0	5
Inj. group: Isolated head injury	0.02	0.14	0	1
Inj. group: Head/other	0.03	0.17	0	1
Inj. group: Head/ortho	0.12	0.33	0	1
Inj. group: SCI	0.01	0.11	0	1
Inj. group: Severe orthopaedic injuries	0.24	0.43	0	1
Inj. group: Chest/abdominal injuries only	0.01	0.11	0	1
Inj. group: Chest/abdo/other	0.00	0.05	0	1
Inj. group: Chest/abdo/ortho	0.05	0.21	0	1
Inj. group: Other/multi-trauma	0.27	0.45	0	1
Inj. group: Other orthopaedic injuries	0.24	0.43	0	1

Continued on next page

Table A.1 – *Continued from previous page*

	Mean	Sd	Min	Max
Cause: Motor vehicle driver	0.36	0.48	0	1
Cause: Motor vehicle passenger	0.14	0.34	0	1
Cause: Motorcycle driver	0.26	0.44	0	1
Cause: Motorcycle passenger	0.01	0.09	0	1
Cause: Pedal cyclist-rider or passenger	0.08	0.26	0	1
Cause: Pedestrian	0.15	0.36	0	1
Cause: Other transport related circumstance	0.01	0.10	0	1
Place: Home	0.01	0.10	0	1
Place: Residential Institution	0.00	0.02	0	1
Place: School, public admin area	0.00	0.03	0	1
Place: Medical hospital	0.00	0.05	0	1
Place: Athletics and sports area	0.01	0.07	0	1
Place: Road, street, or highway	0.91	0.29	0	1
Place: Trade or service area	0.01	0.11	0	1
Place: Industrial or constructional area	0.00	0.02	0	1
Place: Farm	0.01	0.09	0	1
Place: Place for recreation	0.01	0.08	0	1
Place: Other specified place	0.03	0.18	0	1
Place: Place unknown	0.01	0.10	0	1
Activity: Sports	0.02	0.15	0	1
Activity: Leisure	0.06	0.23	0	1
Activity: Working for Income	0.01	0.10	0	1
Activity: Education	0.00	0.01	0	1
Activity: Other Work	0.00	0.07	0	1
Activity: Being Nursed	0.00	0.02	0	1
Activity: Vital activity	0.01	0.07	0	1
Activity: Other activity	0.61	0.49	0	1
Activity: Activity unknown	0.29	0.46	0	1
— *Health-related behaviour* —				
Alcohol condition	0.06	0.23	0	1
Drug conditions	0.02	0.15	0	1
Substance use condition	0.08	0.26	0	1
Any Mental condition	0.10	0.29	0	1
Mood disorders	0.01	0.11	0	1
Neurotic disorder conditions	0.01	0.11	0	1
— *Insurance coverage-related characteristics* —				
TAC indicator for catastrophic injury	0.05	0.22	0	1
TAC division: Independence	0.08	0.28	0	1
TAC division: Rapid Recovery	0.87	0.34	0	1
TAC division: Supported Recovery	0.05	0.21	0	1
TAC division: Unknown/Other	0.00	0.05	0	1
Vehicle premium risk zone: high	0.39	0.49	0	1
Vehicle premium risk zone: medium	0.17	0.38	0	1
Vehicle premium risk zone: low	0.21	0.40	0	1
Vehicle premium risk zone: unknown	0.24	0.43	0	1
TAC premium insurance class: Passenger vehicle	0.47	0.50	0	1
TAC premium insurance class: Goods vehicle	0.08	0.27	0	1
TAC premium insurance class: Motorcycles	0.18	0.38	0	1
TAC premium insurance class: Other	0.03	0.17	0	1
TAC premium insurance class: Unknown	0.24	0.43	0	1
— *Time* —				
Year 2009	0.10	0.30	0	1
Year 2010	0.10	0.30	0	1
Year 2011	0.12	0.32	0	1

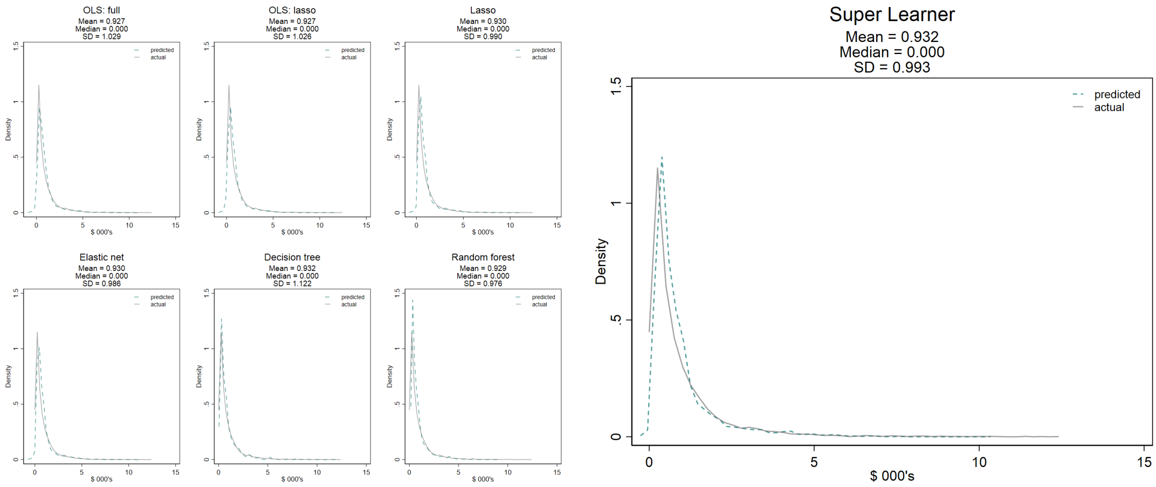
Continued on next page

Table A.1 – *Continued from previous page*

	Mean	Sd	Min	Max
Year 2012	0.11	0.31	0	1
Year 2013	0.11	0.32	0	1
Year 2014	0.11	0.31	0	1
Year 2015	0.12	0.32	0	1
Year 2016	0.12	0.33	0	1
Year 2017	0.11	0.31	0	1
Observations	11625			

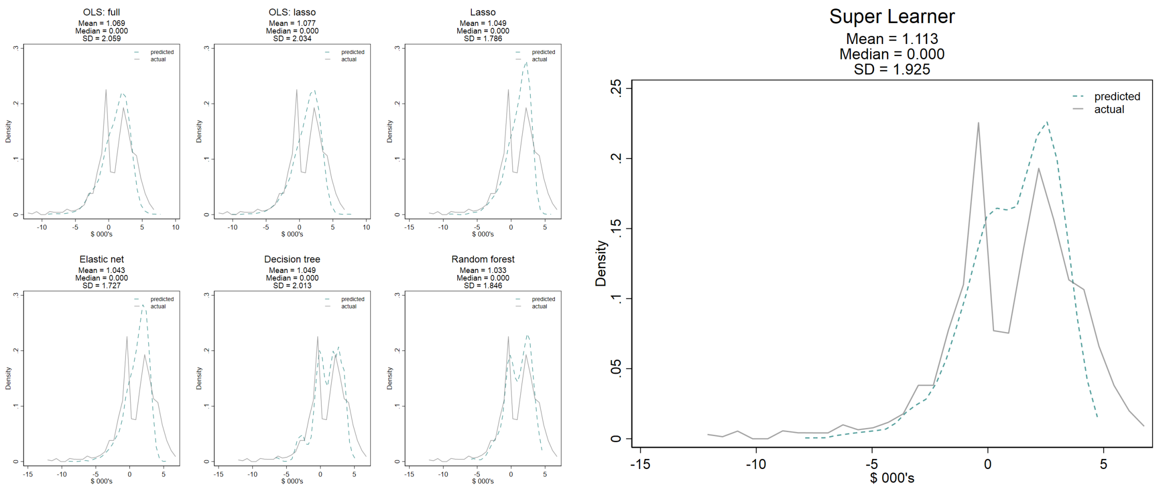
Note.— Table presents descriptive statistics of variables used to restrict the sample, generate the prediction outcomes and covariates included in the prediction models. In most cases all models include a set of dummy indicators divided into groups: patient demographics, clinical treatment-related, injury-related characteristics, health-related behaviour and insurance coverage-related characteristics. Here SES is the Social Economic Status as defined by the Index of Relative Socio-economic Advantage and Disadvantage; MTS refers to Major Trauma Services, MeTS - Metropolitan Trauma Services, lower levels of care include Regional Trauma Services and Rural Healthcare Services; ISS refers to the Injury Severity Score; CCI - Charlson Comorbidity Index and AIS - the Abbreviated Injury Scale. In addition, in all prediction models we an extensive set of dummy indicators for the main diagnosis. This is not reported in the table.

Figure A.1.
The Distribution of Predicted values: Outcome *Direct Costs*



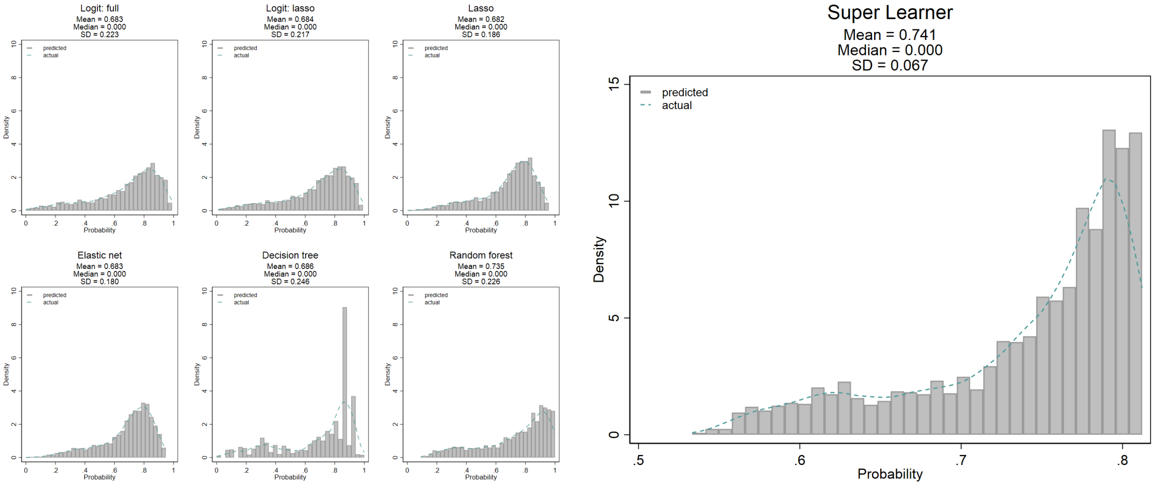
Note.— Figure presents the Kernel Density estimation of predicted values of the *Direct costs* within two years of the injury by each single algorithm and the Super Learner. Grey solid line refers to the observed values of the outcome, while the dashed green line to the predicted values. The subtitles report the main statistical measures of the predictions. Statistics shown on the validation sample.

Figure A.2.
The Distribution of Predicted values: Outcome *Net Monetary Benefit*



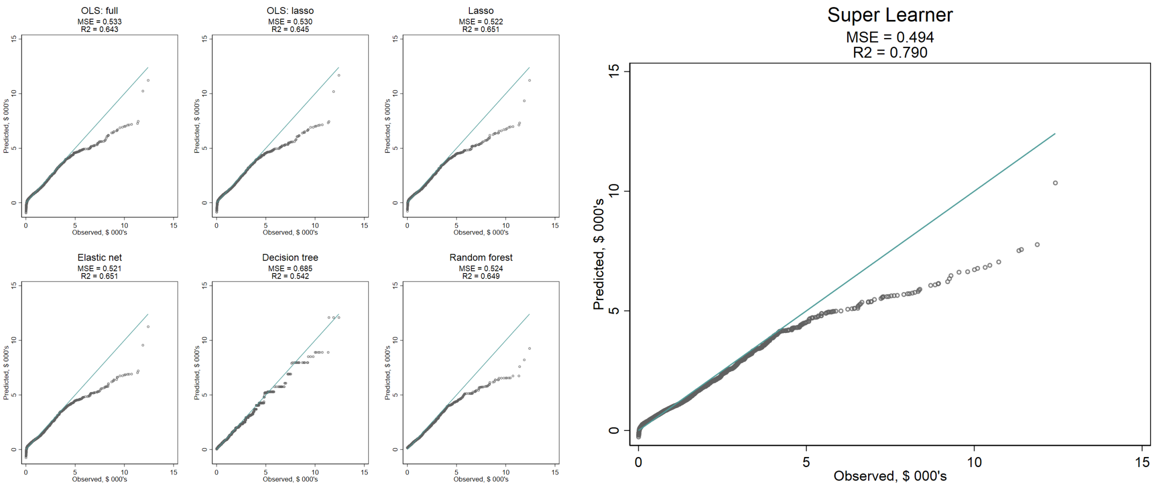
Note.— Figure presents the Kernel Density estimation of predicted values of the *Net Monetary Benefit* following 2 years after the injury by each single algorithm and the Super Learner. Grey solid line refers to the observed values of the outcome, while the dashed green line to the predicted values. The subtitles report the main statistical measures of the predictions. Statistics shown on the validation sample and are further restricted to a sub-sample of patients who were admitted to the ICU.

Figure A.3.
The Distribution of Predicted values: Outcome *RTW*



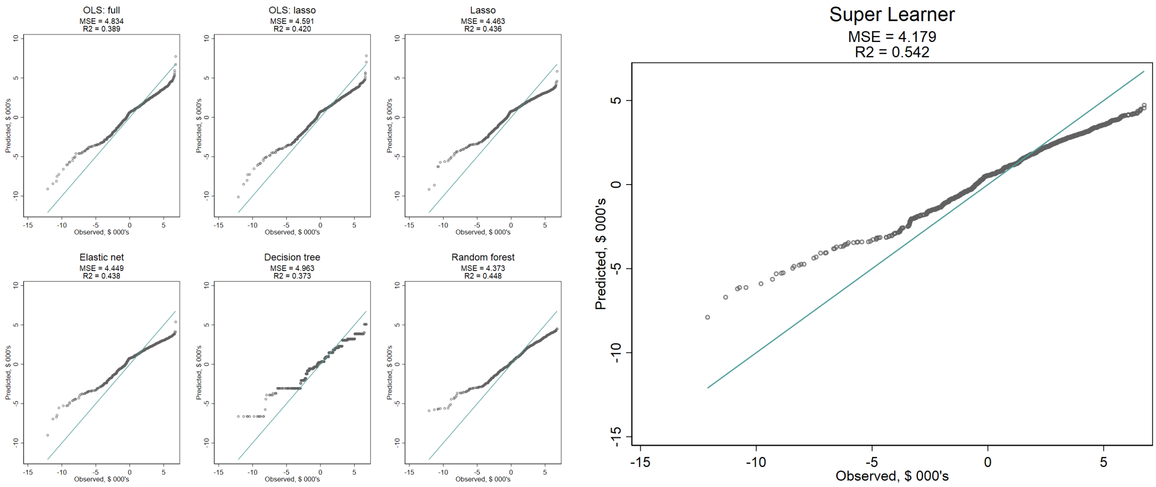
NOTE.— FIGURE PRESENTS EMPIRICAL DISTRIBUTIONS OF PREDICTED VALUES OF THE *RTW* BY EACH SINGLE ALGORITHM. THE SUBTITLES REPORT THE MAIN STATISTICAL MEASURES OF THE PREDICTIONS ON THE VALIDATION SAMPLE. STATISTICS SHOWN ON THE VALIDATION SAMPLE AND ARE FURTHER RESTRICTED TO A SUB-SAMPLE OF PATIENTS WORKED PRIOR TO THE INJURY.

Figure A.4.
The Quantile-Quantile plot: Outcome *Direct Costs*



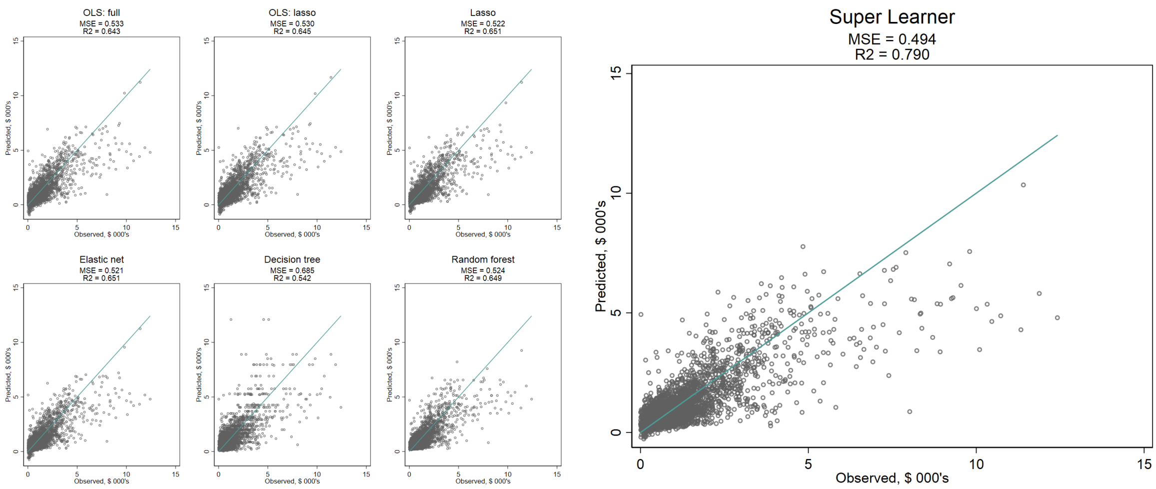
NOTE.— FIGURE PRESENTS THE QUANTILE - QUANTILE PLOT THAT PLOTS QUANTILES OF THE OBSERVED VALUES AGAINST THE PREDICTED VALUES FOR THE OUTCOME *DIRECT COSTS*. THE SUBTITLE REPORTS THE GOODNESS OF FIT STATISTICAL MEASURES OF THE PREDICTIONS ON THE VALIDATION SAMPLE. QUANTILES ESTIMATED ON THE VALIDATION SAMPLE.

Figure A.5.
The Quantile-Quantile plot: Outcome *Net Monetary Benefit*



Note.— Figure presents the Quantile - Quantile plot that plots quantiles of the observed values against the predicted values for the outcome *Net Monetary Benefit*. The subtitle reports the Goodness of Fit statistical measures of the predictions on the validation sample. Quantiles estimated on the validation sample and further restricted to a sub-sample of patients who were admitted to the ICU.

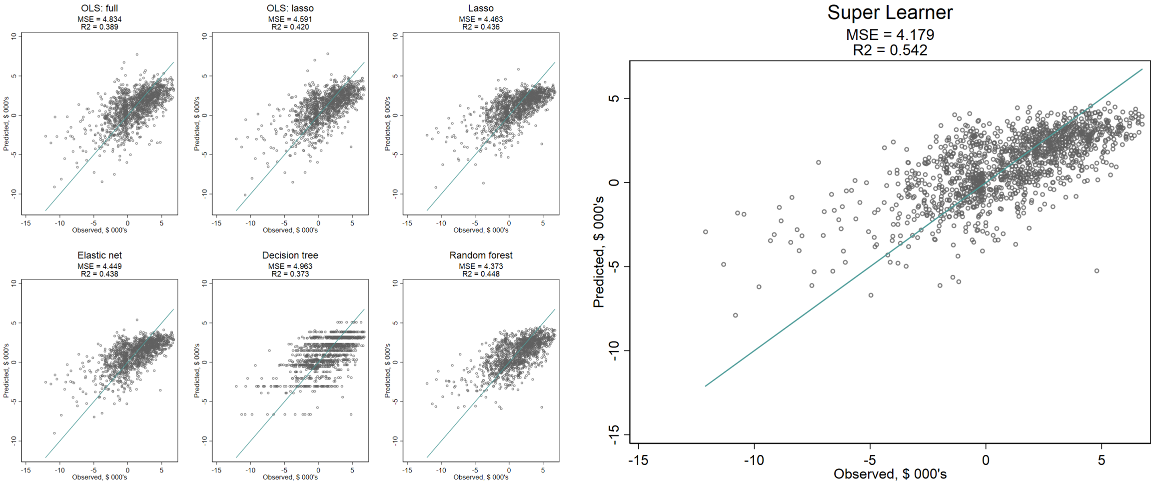
Figure A.6.
The Prediction Error for each observation: Outcome *Direct costs*



Note.— Figure presents the prediction error for each observation of the outcome *Direct Costs*. The reference line denotes a perfect prediction. The observations shown on the validation sample.

Figure A.7.

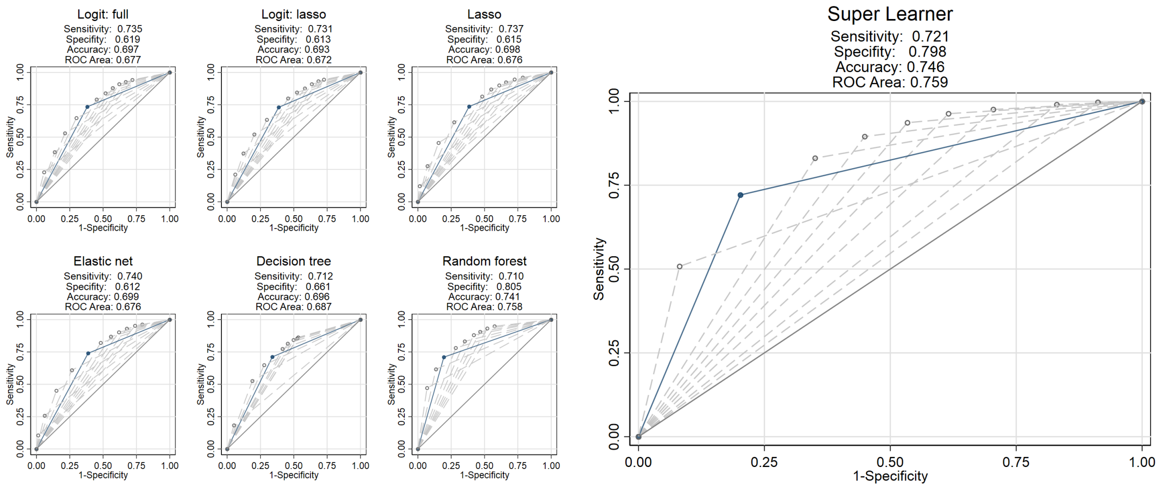
The Prediction Error for each observation: Outcome *Net Monetary Benefit*



Note.— Figure presents the prediction error for each observation of the outcome *Net Benefit*. The reference line denotes a perfect prediction. The observations shown on the validation sample and further restricted to a sub-sample of patients who were admitted to the ICU.

Figure A.8.

The Distribution of Predicted values: Outcome *RTW*



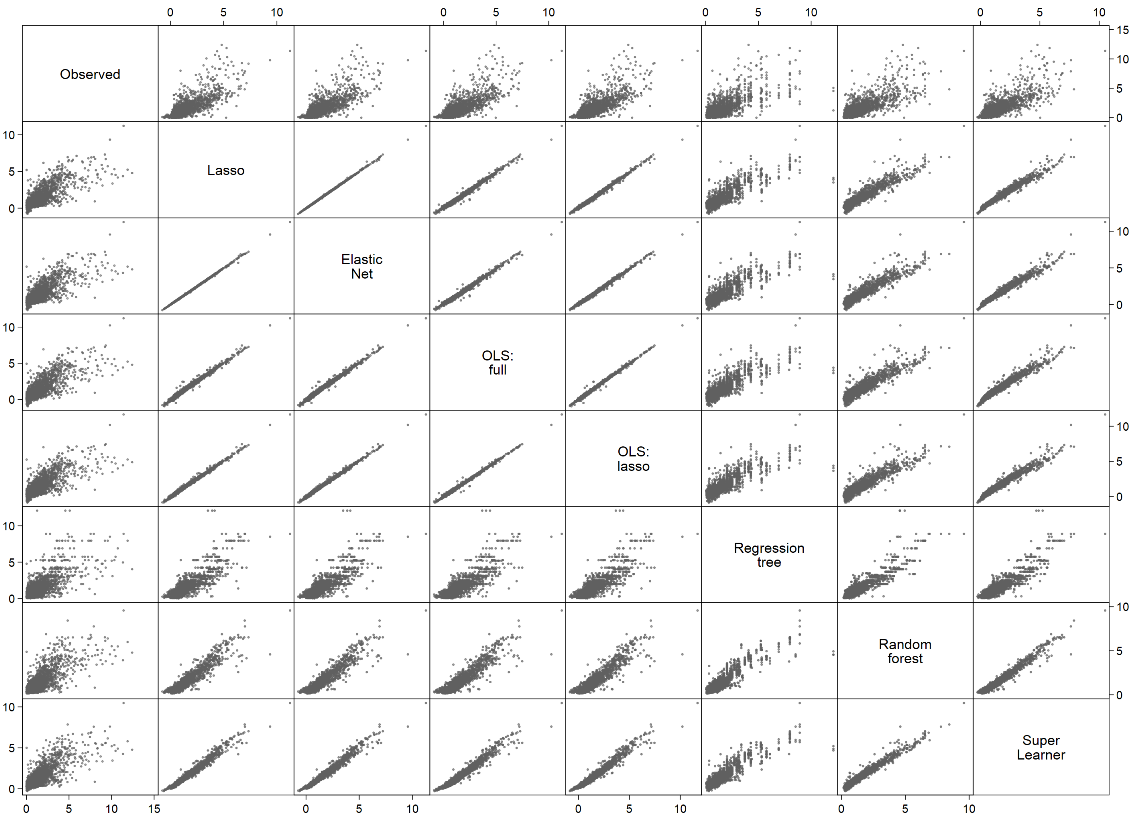
Note.— Figures present the Goodness of Fit measures for the prediction of *RTW* by each single algorithm (on the left panel) and the Super Learner (on the right panel). In each graph each line represents the Specificity/Sensitivity statistical measures for different selected classification thresholds that fluctuate around the mean of the outcome. The blue lines indicate the best prediction performed by each algorithm according to the estimated area under the ROC curve and the subtitle reports its statistical metrics in detail. The results are shown on the validation sample and are further restricted to a sub-sample of patients who worked prior to the injury.

Table A.2.
Descriptive statistics of predictions

	mean	sd	min	median	max
—Direct costs—					
Observed	0.92	1.22	0.00	0.51	12.41
Lasso	0.93	0.99	-0.76	0.62	11.23
Elastic Net	0.93	0.99	-0.75	0.62	11.26
OLS: full	0.93	1.03	-0.93	0.62	11.23
OLS: lasso	0.93	1.03	-0.85	0.62	11.69
Regression tree	0.93	1.12	0.05	0.57	12.11
Random forest	0.93	0.98	0.16	0.59	9.27
Super Learner	0.93	0.99	-0.27	0.60	10.35
Observations	4650				
—Net Monetary Benefit—					
Observed	1.10	2.81	-12.11	1.39	6.75
Lasso	1.05	1.79	-9.14	1.45	5.87
Elastic Net	1.04	1.73	-9.00	1.41	5.40
OLS: full	1.07	2.06	-9.09	1.38	7.74
OLS: lasso	1.08	2.03	-10.11	1.43	7.84
Regression tree	1.05	2.01	-6.61	1.51	5.10
Random forest	1.03	1.85	-5.89	1.26	4.53
Super Learner	1.11	1.93	-7.89	1.42	4.73
Observations	1379				
—RTW—					
Observed	0.69	0.46	0.00	1.00	1.00
Lasso	0.68	0.19	0.02	0.73	0.97
Elastic Net	0.68	0.18	0.02	0.73	0.97
Logit: full	0.69	0.22	0.00	0.75	0.99
Logit:: lasso	0.69	0.22	0.01	0.75	0.99
Regression tree	0.69	0.24	0.00	0.80	1.00
Random forest	0.74	0.23	0.10	0.82	1.00
Super Learner	0.74	0.07	0.54	0.77	0.81
Observations	2948				

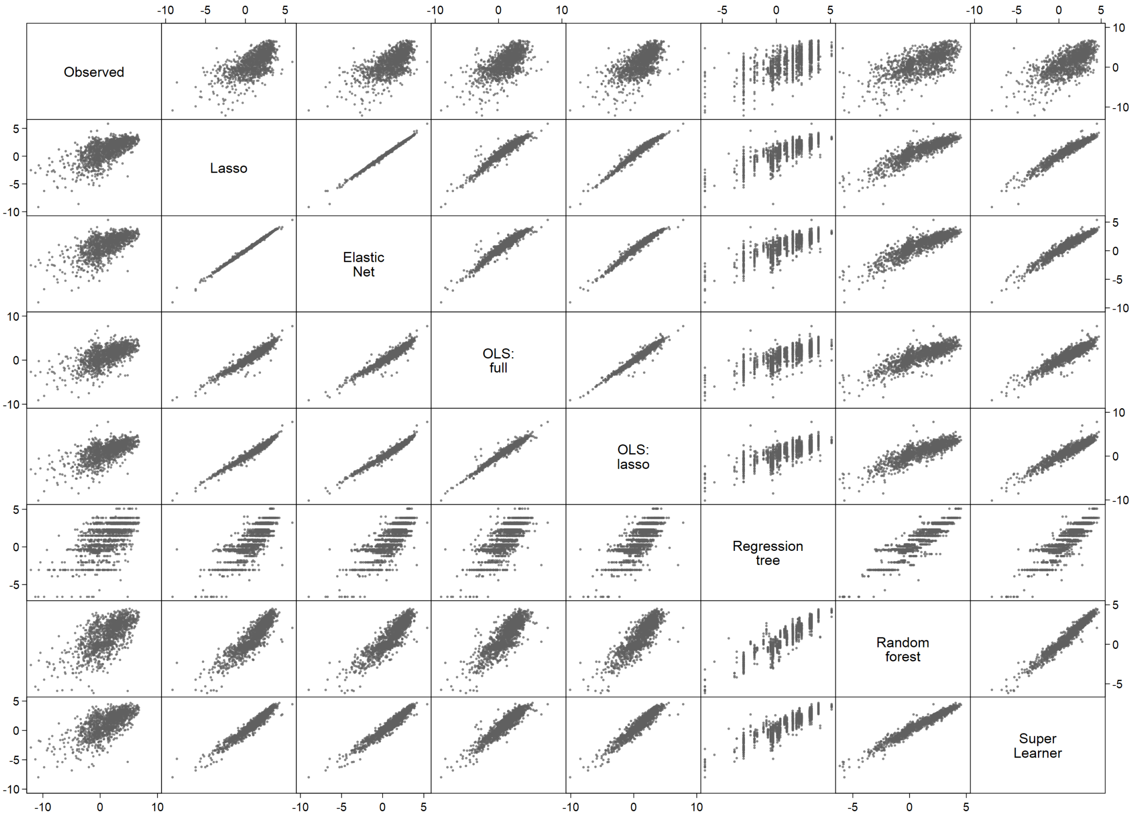
Note.— Table presents the descriptive statistics of predictions made using single algorithms and the Super Learner. The predictions of *Direct Costs* and *Net Monetary Benefit* are based on a linear specification of lasso, elastic net and OLS regressions, while the predictions of *RTW* are estimated using a logistic regression specification. All estimations performed on validation sample. Models with *RTW* outcome restricted to a subsample of patient who worked prior to the injury and models with *Net Monetary Benefit* are restricted to patients who were admitted to the ICU during their hospital stay.

Figure A.9.
Correlation Matrix: Outcome *Direct costs*



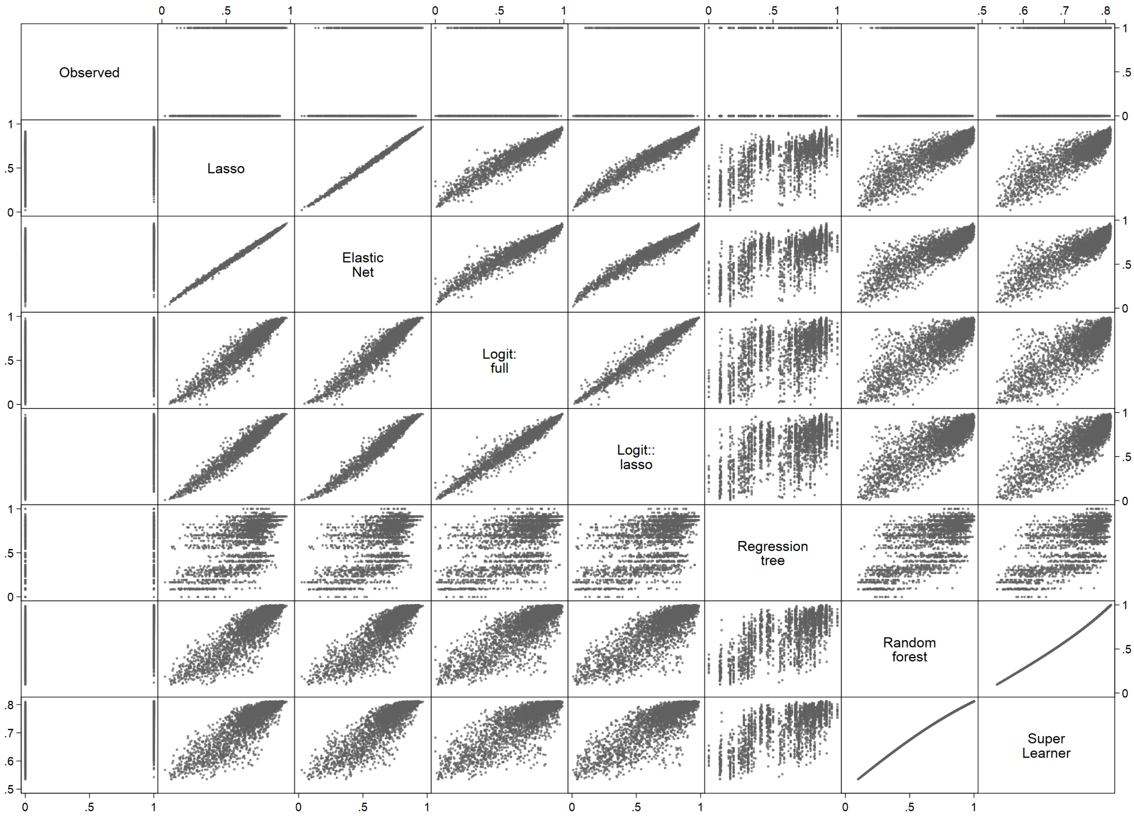
Note.— Figure presents the correlation matrix of predictions made by each single algorithm in the prediction of the outcome *Direct costs*. The results are shown on the validation sample.

FIGURE A.10.
Correlation Matrix: Outcome *Net Monetary Benefit*



NOTE.— FIGURE PRESENTS THE CORRELATION MATRIX OF PREDICTIONS MADE BY EACH SINGLE ALGORITHM IN THE PREDICTION OF THE OUTCOME *NET MONETARY BENEFIT*. SAMPLE IS RESTRICTED TO PATIENTS WHO WERE ADMITTED TO THE ICU DURING THEIR HOSPITAL STAY AND THE RESULTS ARE SHOWN ON THE VALIDATION SAMPLE.

Figure A.11.
Correlation Matrix: Outcome *RTW*



Note.— Figure presents the correlation matrix of predictions made by each single algorithm in the prediction of the outcome *Net Monetary Benefit*. Sample is restricted to patients who worked prior to the injury and the results are shown on the validation sample.