

---

# WHY INTERPRETABLE CAUSAL INFERENCE IS IMPORTANT FOR HIGH-STAKES DECISION MAKING FOR CRITICALLY ILL PATIENTS AND HOW TO DO IT

---

Harsh Parikh<sup>\*1</sup>, Kentaro Hoffman <sup>\*3</sup>, Haoqi Sun <sup>\*2</sup>, Wendong Ge<sup>2</sup>, Jin Jing<sup>2</sup>, Rajesh Amerineni<sup>2</sup>, Lin Liu<sup>4</sup>, Jimeng Sun<sup>2</sup>, Sahar Zafar<sup>2</sup>, Aaron Struck<sup>5,6</sup>, Alexander Volfovsky<sup>† 1</sup>, Cynthia Rudin <sup>† 1</sup>, and M. Brandon Westover <sup>†2</sup>

<sup>1</sup>Duke University

<sup>2</sup>Massachusetts General Hospital

<sup>3</sup>University of North Carolina at Chapel Hill

<sup>4</sup>Harvard University

<sup>5</sup>University of Wisconsin-Madison

<sup>6</sup>William S Middleton Veterans Hospital

## ABSTRACT

Many fundamental problems affecting the care of critically ill patients lead to similar analytical challenges: physicians cannot easily estimate the effects of at-risk medical conditions or treatments (which is problematic for treatment decisions) because the causal effects of medical conditions and drugs are entangled. They also cannot easily perform studies: there are not enough critically ill patients for high-dimensional observational causal inference analysis, and randomized controlled trials often cannot ethically be conducted. However, mechanistic knowledge is available, including how drugs are absorbed into the body, and the combination of this knowledge with the limited data could potentially suffice – if we knew how to combine them. In this work, we present a framework for interpretable estimation of causal effects for critically ill patients under exactly these complex conditions: interactions between drugs and observations over time, patient data sets that are not large, and mechanistic knowledge that can substitute for lack of data. Our framework incorporates pharmacokinetics and pharmacodynamics with interpretable matching methods to adjust for confounders such as patients’ drug response, medical history, and demographic variables. We apply this framework to an extremely important problem affecting critically ill patients, namely the effect of seizures and other potentially harmful electrical events in the brain (called epileptiform activity – EA) on outcomes. EA is a key indicator of whether the patient will suffer long term severe

---

<sup>\*</sup>Co-first authors

<sup>†</sup>Co-senior authors

neurological disability or death. Given the high-stakes involved, and the high noise in the data, interpretability is critical for troubleshooting such complex problems. Interpretability of our matched groups allowed neurologists to perform chart review to verify the quality of our causal analysis. For instance, our work indicates that a patient who experiences a high level of seizure-like activity (75% high EA burden), and is untreated for a six-hour window, has, on average, a 16.7% increased chance of adverse outcomes such as severe brain damage, lifetime disability, or death. We were also able to show that patients with mild-but-long-lasting EA (average EA burden  $\geq 50\%$ ) have their risk of an adverse outcome increased by 11.2%. This information is essential to any neurologist who treats critically-ill patients.

## 1 Introduction

Caring for critically ill patients is extremely challenging: the decisions are high stakes, there are difficult causal questions (will this patient respond to available drugs?), and decisions about drug dosage are entangled with observations that physicians are making about the patient over time.

Experiments (clinical trials) on such patients are difficult, observational datasets are noisy and small, and there may be potential important variables, such as drug absorption rates, and the severity of the patient’s condition, which typically are not recorded in a database. Ignoring these variables can lead to biased estimates of treatment effects, a naïve statistical analysis is doomed to fail, and the use of black box models in either analysis or decision making could easily lead to erroneous conclusions and cause harm. Ideally, we need an *interpretability-centered* framework for these types of high-stakes causal analyses: a physician should be able to verify the quality of every single step in the analysis, from how a current patient compares to past patients (case-based reasoning), how drug absorption and response is modeled, and an understanding of the relative importance of variables.

This paper introduces a general framework that can help estimate heterogeneous causal effects from high-dimensional patient data with complex time-series interactions, low signal-to-noise ratio and where treatments are not randomly assigned. Each step of the framework is designed to be interpretable. Importantly, we leverage established interpretable pharmacokinetic-pharmacodynamic (PK/PD) models to describe personalized clinical-decision-physiological-response interactions, allowing us to identify individuals who might react similarly to treatments. We learn a flexible distance metric on the space of covariates to perform matching for estimating the medium- and long-term causal effects of both the clinical decisions and physiological responses; the matched group we construct for each patient can be validated, or possibly, criticized. In the context of medical data, this validation can be performed via a chart review that provides a qualitative assessment of the matches in terms of information that was not directly used in the matching procedure.

Using this framework, we perform the first causal analysis of a common form of potentially harmful electrical activity in the brain known as “epileptiform activity” (EA, also called “ictal-interictal-injury continuum activity”, see [Hirsch et al., 2021](#)). EA is common to critically ill patients suffering from brain injury ([Lucke-Wold et al., 2015](#)), cancer ([Lee et al., 2013](#)), organ-failure ([Boggs, 2002](#)), and infections such as Covid-19 ([Nikbakht et al., 2020](#); [Lin et al., 2021](#)), affecting more than half of patients who undergo electroencephalography ([Gaspard et al., 2013](#)). Prolonged EA is associated with increased in-hospital mortality, and survivors often suffer from a functional and cognitive disability

(Ganesan and Hahn, 2019; Rossetti et al., 2019; Kim et al., 2018). While there is a growing body of literature indicating that EA is *associated* with poor outcomes (Oddo et al., 2009), there is still a debate as to whether (a) EA is part of a causal pathway that worsens a patient's outcomes and thus requires aggressive treatment, *or* (b) the worsened outcomes are due to mechanisms other than EA such as the side-effects of medications or the inciting medical illness, with EA occurring as an epiphenomenon. (Chong and Hirsch, 2005; Rubinos et al., 2018; Osman et al., 2018; Johnson and Kaplan, 2017; Tao et al., 2020; Cormier et al., 2017).

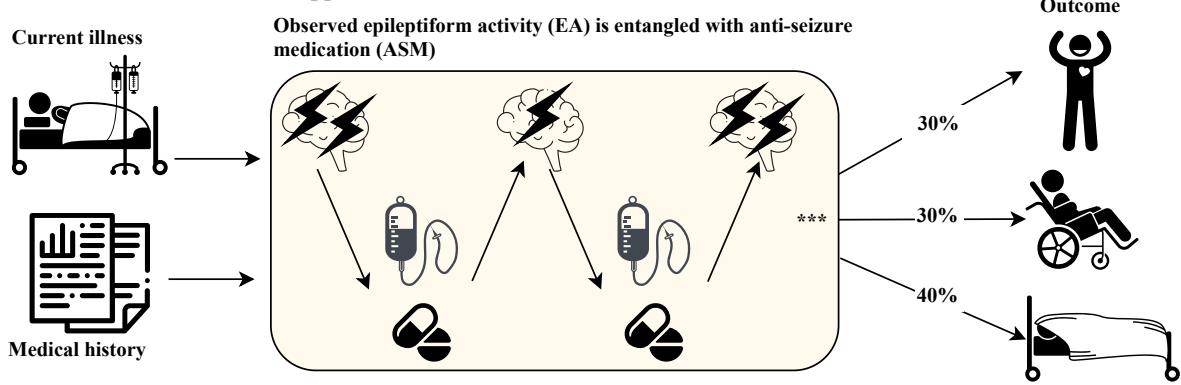
However, the study of EA suffers from a variety of limitations. First, a hypothetical clinical trial studying EA would need to randomly induce EA in patients while limiting their drug treatments, which is neither plausible nor ethical. Second, as it requires a physician to order an EEG and trained technologist to monitor the device, sample sizes for EA datasets tend to be no larger than a few hundred to a thousand and of limited time windows. Worst of all, what complicates the study of EA is its complex interactions with anti-seizure medications (ASM). Medical caregivers administer ASMs based on patients' EA, and in turn, EA is affected by ASMs. Therefore, this creates an entanglement (see Figure 1) between the EA (treatment) and ASMs (confounder), potentially obscuring the true causal effect of EAs.

The study of EA has been a case where scientists have been using *predictive* models to answer a *causal* question despite strong confounding factors. Researchers have used regression models to adjust for the patient's medical history and demographic factors (Payne et al., 2014; De Marchis et al., 2016; Zafar et al., 2018; Muhlhofer et al., 2019), and have interpreted the resulting regression coefficient for EA as the causal effect of EA on a patient's outcome. While this approach is appealing for its simplicity and is widely used, *it is not appropriate to interpret regression coefficients as causal in the presence of strong confounding interactions*. Using conventional prognostic modeling approaches can put one at the risk of misinterpreting the association between high levels of ASM, EA, and poor outcomes as causal even if no causal link exists.

Our framework is different than other approaches in that it aims to tightly match patients on *all known relevant confounding factors* such as medical history and diagnosis, pharmacological characteristics and demographics. We adjust for important pharmacokinetic/pharmacodynamic (PKPD) parameters to better characterize individualized responses to anti-seizure medications; this mechanistic information helps compensate for our not-large sample size and limited EEG observation time window. The interpretability of our framework also gives important medical insights into the EA process which are easier for practicing clinicians to incorporate.

We can thus finally provide the first high quality causal analysis of EA. We find that higher EA burden indeed leads to worse neurologic outcomes (Figure 3), in a way that depends on the intensity (max burden over 6 hours) and duration (average burden over 24 hours) of EA. Specifically, those with a max burden between 0.75 to 1 are on average 13.4% more likely to be discharged with a poor outcome (as defined by modified Rankin Scale of Burn, 1992) than those with max burden less than 0.25. Additionally, we find that patients with central nervous system infection or toxic metabolic encephalopathy are affected by EA more than the average level in this cohort. Importantly, the validity of the estimate is supported by a detailed clinical chart review of the matched groups, which could only be accomplished because of the interpretability of our framework.

### Observational data: What Happened



### Counterfactual: What would happen if the patient experienced different level of EA

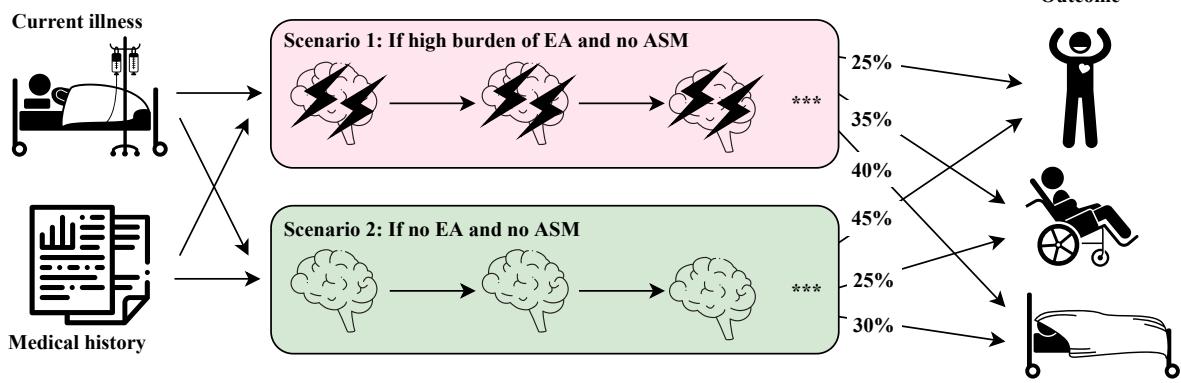


Figure 1: *Upper:* Illustration showing that observed epileptiform activity (EA) and treatment decision form a feedback loop, that is also influenced by current illness and medical history (left). The entire time-series of EA and ASM influence patient outcomes. Possible outcomes include return to normal health, disability, or death at the time of hospital discharge (right). *Lower:* Our goal is to estimate the effect of EA on patient outcomes. The effect is obtained by comparing the patient outcome across counterfactual scenarios. Scenario 1 is where every patient in this cohort had certain (high) level (or burden) of EA but no ASM is given; Scenario 2 is where every patient had no EA and also no ASM could be given. (Note that the probabilities given here are illustrative, and not taken from data.)

## 2 Framework

The general framework is shown in Figure 2. The first step of our framework is the identification of physiological phenomena that might affect long-term health outcomes. Importantly, these phenomena are frequently not recorded directly, and instead relevant patterns must be extracted from raw waveforms. Examples include monitoring blood pressure and serial blood cultures in patients with sepsis; heart rhythms, blood pressure, oxygen levels, and serial blood electrolyte levels in patients with life threatening arrhythmias like atrial fibrillation or atrial flutter; urine output, body weight, and blood electrolytes in patients with acute kidney failure; intracranial pressure and brain tissue oxygen levels in patients with severe traumatic brain injury; or, as in the example that we analyze in this paper, detecting EA from EEG signals. *Our framework focuses on estimating the long-term effects of these patterns.* However, the raw waveform data rarely exists in settings without clinical interventions: we must control for the effects of interventions, for example, in the medical scenarios mentioned above: effects of medications to increase blood pressure and antibiotics given in sepsis; medications to abort arrhythmias and raise blood pressure in patients with atrial fibrillation/flutter; electrolyte infusions, diuretic drugs, and hemodialysis given to patients with acute kidney failure; high concentration saline or

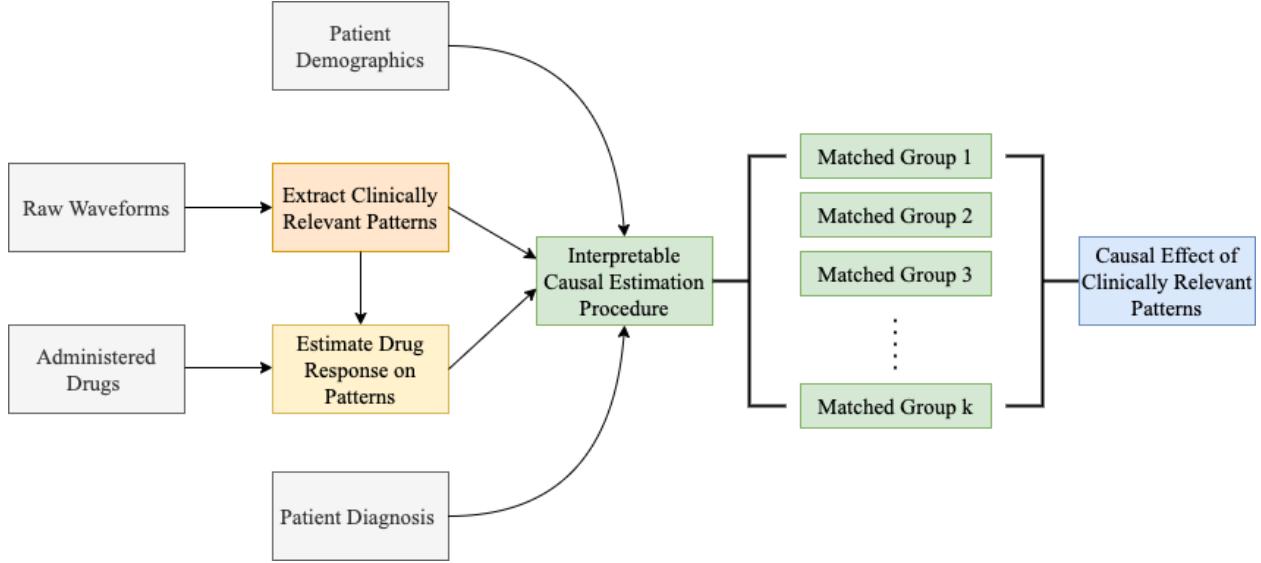


Figure 2: Flowchart demonstrating the working of our framework for interpretable inference of causal effects

surgical treatments given to reduce intracranial pressure in patients with brain trauma; or the amount of antiseizure medication given how well it was absorbed in patients treated for EA.

As the goal is to identify the long-term effects of observed patterns, a patient who was never treated is not comparable to a patient who was. Thus, we combine the patient’s demographic variables (e.g., age, weight) and patient characteristics within a pharmacodynamic/pharmacokinetic model to estimate drug response parameters for each patient. The patient data, including drug response parameters, are all used for high-quality matching; each patient is matched *almost exactly* to past patients with similar characteristics, medical history, and estimated drug response parameters. Almost-exact matching (Parikh et al., 2020) matches patients directly on potential confounders (not, for instance, on proxies such as the propensity score). The matched groups permit case-based reasoning, and allow us to estimate the effects of both seizure-like activity and drugs meant to reduce seizure-like activity on patient outcomes. In addition to these matched groups being almost-exact, domain experts can perform chart review for each patient’s matched groups to evaluate their quality. As these charts contain not only the quantitative factors used for matching but also qualitative information such as doctors’ notes, they allows for a holistic assessment that might lead to unobserved confounding.

### 3 The Causal Study of EA

As discussed, EA affects more than half of critically ill patients on EEG (Gaspard et al., 2013), and understanding its effects can help prevent severe brain damage. In what follows, we outline our approach to EA analysis following the framework discussed above.

**Patient Cohort** Our study is a retrospective cross-sectional analysis of patients admitted to the Massachusetts General Hospital (MGH) between September 2011 and February 2017. Institutional review boards at MGH, Duke University, and University of North Carolina at Chapel Hill approved the retrospective analysis without requiring written informed consent. Inclusion criteria included (1) admission to the hospital, (2) monitoring with continuous electroencephalography (EEG) for more than 2 hours, and (3) availability of drug administration data from the hospital’s

electronic records. Patients who had poor quality of EEG signal for more than 30% of the total recording length or those missing discharge outcome were excluded from the study. For patients with multiple visits to the hospital, we only analyzed their first visit. A flowchart of the full patient selection procedure can be seen in Figure 8. The final cohort contained 995 critically ill patients.

For each patient, we collected a variety of variables about their medical history including demographics (gender, marital status, and age), clinical factors (substance abuse, history of seizures or epilepsy, chronic kidney disease, etc.), and what disease(s) they were diagnosed with (cancer, subarachnoid hemorrhage , or central nervous system infection). As this information concerns factors that are fixed before admission to the hospital for treatment, these are referred to as the *pre-admission variables*. A full list of these pre-admission variables can be found in Table 2.

**Outcomes of Interest** Once a patient has stabilized (or passed away), they are discharged from the hospital. The level of disability at discharge is quantified on a 0 to 6 ordinal scale called the Modified Rankin Scale (mRS). In the literature the post-discharge outcome is frequently binarized into those with ( $mRS \geq 4$ ) and without ( $mRS \leq 3$ ) serious disabilities (Zafar et al., 2018). Our work also uses this binarized Modified Rankin Scale as the outcome of interest, with  $Y$  equal to 1 representing a patient discharged with serious disabilities and 0 representing a patient without serious disabilities.

**Complex Time Series Interactions: Drug treatments and EA** After treatment is started, patients are kept under close observation including frequent visits by physicians and nurses, and continuous brain monitoring using electroencephalography (EEG). Based on these observations, physicians update a patient’s treatment by adjusting the types and doses of anti-seizure medications (ASMs). This observation-treatment cycle results in: (1) a univariate time series of the average proportion of time the  $i$ -th patient experienced EA in the past  $\omega$  hours( $\{Z_{i,t}^\omega\}_{t=1}^T$ ) based on an EEG sampling rate of 2 seconds and (2) a 6-dimensional vector time-series ( $\{W_{i,t}\}_{t=1}^T$ ) representing the dose of 6 most commonly used ASMs (Lacosamide, Levetiracetam, Midazolam, Phenobarbital, Propofol, and Valproate) received by  $i$ -th patient at time-step  $0 \leq t \leq T$ . We use  $\omega = 6$  hours as it is a reasonable amount of time to observe the effects of the ASMs on EA and for physicians to adjust a patient’s ASM regimen (Garoud et al., 2006). Details on how EA signals were identified in the EEG recordings can be found in Appendix B.

**Clinically Relevant Summaries of EA Burden Over Time.** We summarize the EA time series  $\{Z_{i,t}^6\}_{t=1}^T$  in two clinically relevant ways, which we refer to as an *EA burden*:

1. *Mean EA burden* ( $E_{i,\text{mean}}$ ) measures the average proportion of time a patient experiences EA in the first 24 hour recording period.
2. *Max EA burden* ( $E_{i,\text{max}}$ ) measures the 6 hour sliding window with the highest proportion of EA within the first 24 hour recording period.

The former measures the prevalence of EA while the second summary provides insights into the most intense periods of EA over a short period of time. By quantifying EA burden in these two different ways, we seek to separately understand the potential harm caused both by brief periods of intense EA and by prolonged periods of less intense EA burden.

**Estimands of Interest.** We would like to estimate the degree to which untreated epileptiform activity (of different intensities) can cause worse neurological outcomes. The potential outcomes of interest are a function of the full time series of EA burden and drug exposures  $Y(\{E_{i,t}, W_{i,t} : t = 1, \dots, T\})$  and we make the simplifying assumption that they vary only according to the clinically relevant summaries of EA burden and whether drugs are present or absent. That is, we say that  $Y(\{E_{1,t}, W_{1,t} : t = 1, \dots, T\}) = Y(\{E_{2,t}, W_{2,t} : t = 1, \dots, T\})$  if  $E_{\max}(\{E_{1,t} : t = 1, \dots, T\}) = E_{\max}(\{E_{2,t} : t = 1, \dots, T\})$  and  $\bar{W}_1 = \bar{W}_2$ , where  $\bar{W}_i = \mathbf{1}[(\sum_t \sum_j W_{i,t,j}) > 0]$ . Thus,  $\bar{W}_i = 1$  if any drugs are administered otherwise  $\bar{W}_i = 0$ . Our estimand of interest is the probability a patient is discharged with severe disability if the patient has EA burden (either  $E_{\max}$  or  $E_{\text{mean}}$ ) equal to  $e$  and was not treated with ASMs. This can be represented as:

$$Pr[Y_i(E_{i,\max} \in e, \bar{W}_i = 0) = 1] \text{ and } Pr[Y_i(E_{i,\text{mean}} \in e, \bar{W}_i = 0) = 1]. \quad (1)$$

Here,  $Y_i$  is the binarized post-discharge outcome,  $e$  is the binned EA burden with  $e \in \{\text{mild, moderate, severe, very severe}\}$  and  $W_{i,t} = 0 \forall t$  indicates that no ASMs were ever administered. We are interested in estimating the potential outcome when there are no administered ASMs because this allows us to disentangle the effects of EA on outcome from the effect of drugs. For interpretability, we bin EA burden ( $e$ ) into 4 levels – mild (0% to 25%), moderate (25% to 50%), severe (50% to 75%), very severe (75% to 100%) – see Table 4 in the Appendix for the number of patients in each category. The choice of cutoffs was influenced by animal models which showed that an EA burden of 50% serves as an important indicator of when EA begins to damage the brain (Trinka et al., 2015). A sensitivity analysis to these choices is provided in Appendix C.

**The variables we Control for: Pre-admission Covariates and Drug-response Covariates.** In the ASM observation-treatment procedure, we observed two large sources of potential confounding. First, those with different diagnoses and patient characteristics may receive more or less ASM treatment from physicians, potentially confounding the estimated harm caused by EA with the harm due to diagnosis or patient characteristics. To address this, a collection of 70 pre-admission covariates that could potentially influence ASM treatment were selected by a group of practicing neurologists and were controlled for via the matching algorithm, Matching After Learning To Stretch (MALTS).

A second source of potential confounding comes from a patient’s drug response. Due to differing past medical history, current medical conditions, age, and other factors, some patients respond well to some ASMs while other patients respond less. This in turn, can directly affect the amount and number of ASMs that a patient receives and their final outcome. To account for this, we modeled each patient’s response to ASM drugs via a one-compartment Pharmacokinetic/Pharmacodynamic (PK/PD) model, and controlled for each patient’s drug responsiveness parameters using MALTS.

## 4 Result: EA Burden have a Direct Causal Effect on Survival

With the EA data summarized as above, our framework can now provide the first causal analysis of the effect of seizure-like activity on the possibility of severe brain damage.

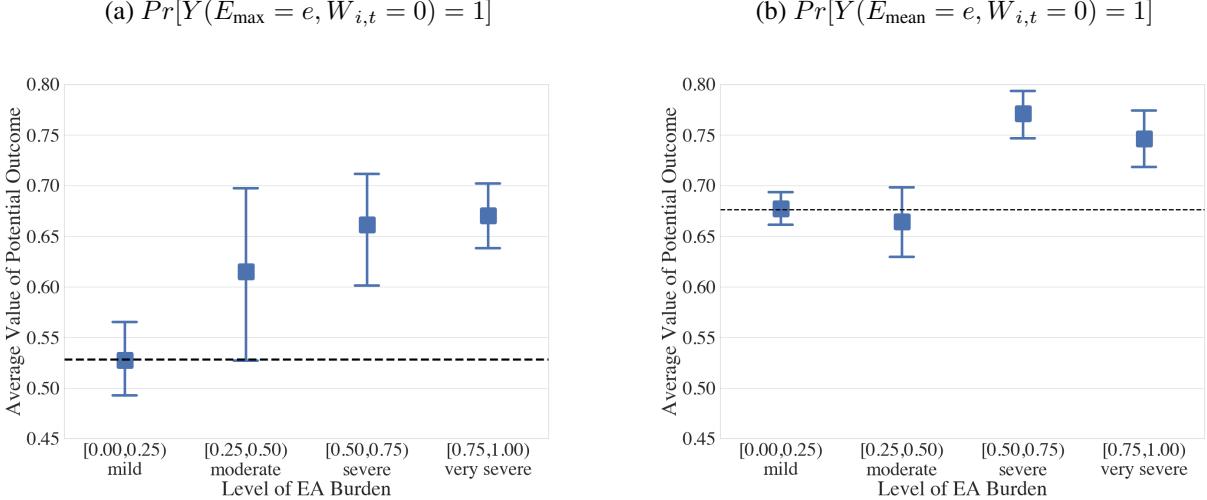


Figure 3: The probability of a poor outcome mRS for either Mild, Moderate, Severe, or Very Severe EA burden. EA Burden is quantified as (Left): Maximum EA in a 6-hour moving average window; (Right): Mean EA in a 6-hour moving average window. In both scenarios, an increase in EA burden leads to a worse outcome for the patient. Outcome worsens monotonically for  $E_{\max}$ , whereas for  $E_{\text{mean}}$ , there is a jump at approximately 0.5. In both plots the horizontal line represents the baseline median average potential outcome for mild case. Note that these baselines need not be equal due to the marginalization over  $\bar{W}_{i,t}$ .

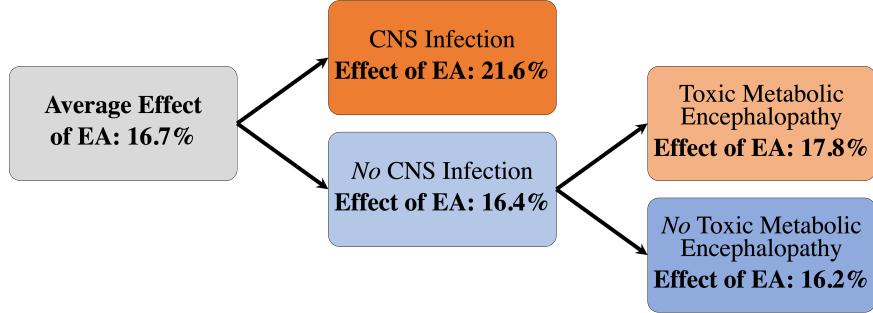
**Average Effect of Max EA Burden on Patient Outcomes.** Figure 3(a) illustrates our first main result: those with higher levels of  $E_{\max}$  are at higher risk of poor neurologic outcomes. Moreover, the risk of a poor outcome increases monotonically as the EA burden increases, culminating in *an average increase of 16.7% in probability of a poor outcome when a patient's untreated EA burden increases from mild (0 to 0.25) to very severe (0.75 to 1)*.

**Average Effect of Mean EA Burden on Patient Outcomes.** Figure 3(b) shows our other main result: those with higher levels of  $E_{\text{mean}}$  are also at higher risk of being discharged with poor outcomes. However unlike  $E_{\max}$ , the risk caused by increasing  $E_{\text{mean}}$  spikes up when a patient goes above even a moderate EA burden, [0.25, 0.50]. Our results indicate that *severe and very severe prolonged EA burden (over 24 hours) increase the risk of worse outcome by 11.2%* as compared to mild or moderate prolonged EA burden.

**Heterogeneity in Effects for Max EA Burden.** While increases in EA burden tend to lead to worse outcomes overall, we found also that there is significant heterogeneity in the size of the effect due to each patient's pre-admission covariates. We can quantify the relative change in outcome from a very severe max EA as the ratio of expected outcomes for those with high EA burden over the expected outcome of those with low EA burden minus one:

$$\text{Average Effect of EA} = \frac{\Pr[Y(E_{\max} \geq 0.75, \bar{W} = 0) = 1] - \Pr[Y(E_{\max} < 0.25, \bar{W} = 0) = 1]}{\Pr[Y(E_{\max} < 0.25, \bar{W} = 0) = 1]}. \quad (2)$$

Thus if the average effect of EA is zero, a very severe max EA is no worse than a mild max EA while an average effect of EA of one would represent a 100% increase in the probability of a bad outcome. Based on this relative effect, we observe that *those with central nervous system infections or with toxic metabolic encephalopathy are at higher risk of a worse outcome if they had a large increase in  $E_{\max}$  burden*. We conjecture that this may be the result of a central nervous system infection and EA leading to a higher inflammatory response in the patient, potentially leading to or



(a) Recursive partitioning of covariate space and respective relative effects of very severe EA

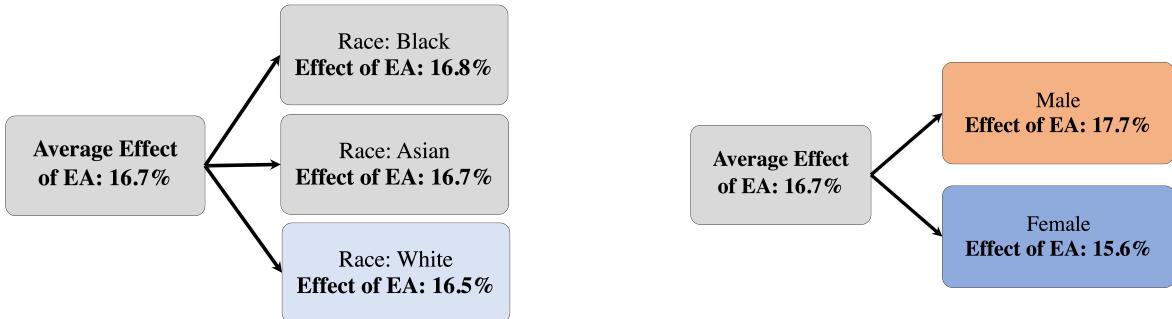


Figure 4: Heterogeneity in the average effect of EA, stratified by: (a) Recursive partitioning on the entire covariate space using Gini splitting to find the most important splits; (b) Partitions the space according patients’ race. The remaining race classes (other, undisclosed, and missing) are rare representing 0.5%, 5%, and 8.4% of the total population. (c) Partitions the space across to patients’ gender. Orange coloring in the boxes implies that the subgroup experiences a larger average estimated causal effect of EA on neurologic outcomes than the cohort mean, and blue implies a smaller causal effect. Subgroups in orange fare worse as a result of a higher EA burden.

exacerbating damage to the brain. Figure 4(a) uses a decision tree to break down the population into subpopulations with differing conditional average treatment effects.

We further examined race and gender as possible effect modifiers of EA burden. Figure 4(b) shows that race does not seem modify the risk from increases in  $E_{max}$ . By contrast, sex does appear to modify the risk: male patients appear to be more susceptible to very severe  $E_{max}$  worsening the chances of recovery compared to female patients (see Figure 4(c)).

## 5 Methodology

Let us describe how we applied the framework from Section 2 to analyze the EA data to obtain the results in Section 4. We divide the estimation pipeline into three stages (Figure 5):

1. In the *first stage* (Section B), we calculate  $E_{max}$  and  $E_{mean}$ . To do this, we need to first identify segments of the EEG signal containing seizure-like EA behavior. Doing this using human annotators would be extremely time consuming, so we use a convolutional neural network (CNN) trained on human annotators’ classifications of

10 second windows into non-EA and EA in a semi-supervised fashion (Ge et al., 2021b; Zafar et al., 2021; Jing et al., 2016). We use the predictions to compute EA time series ( $Z_t^\omega$ ). As described in Section 4,  $E_{\max}$  and  $E_{\text{mean}}$  are computed directly from  $Z_t^\omega$ . Details are in the appendix in Section B.

2. In the *second stage* (Section 5.1), we fit a personalized pharmacokinetic/pharmacodynamic (PK/PD) model to each patient’s response to ASM (Hill, 1909).
3. In the *third stage* (Section 5.2) we combine the pre-admission covariates, such as baseline demographic data and data related to the nature and severity of the present illness, and the PK/PD parameters estimated in the second stage, to adjust for potential confounding and to estimate the potential outcomes of interest. We learn a distance metric to create high-quality matched groups using an *interpretable and accurate* matching method, Matching After Learning to Stretch for EA effect estimation (MALTS, Parikh et al., 2020).

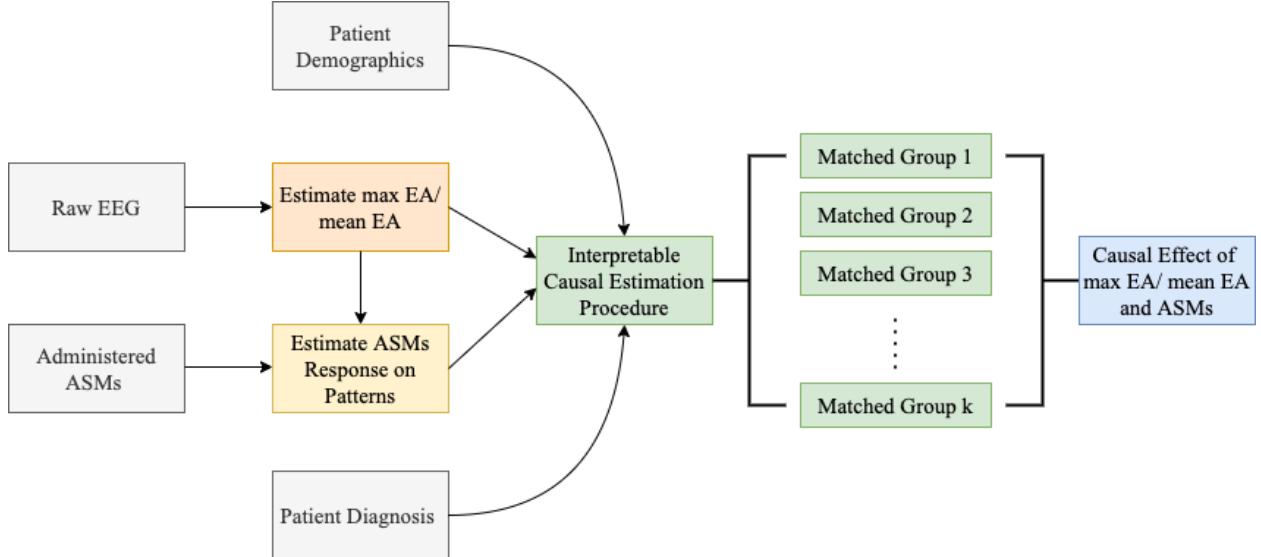


Figure 5: The overall analysis framework, consisting of three parts (indicated by different colors): EA burden computation, individual PK/PD modeling, and MALTS matching and effect estimation.

## 5.1 Mechanistic Pharmacological Model

As noted in section 3, doctors dynamically modify the type and dosage of ASM using the current EA observation, previous treatment, and patient’s responsiveness to these treatments. This cyclical relationship potentially confounds the relationship between EA and a patient’s final outcome. The heterogeneity in a patient’s responsiveness to ASMs can be due to a variety of factors such as past medical history, current medical conditions, age, etc. However, the infrequency of some rare medical conditions makes it difficult to learn a nonparametric model of drug response that incorporates all relevant medical factors. To account for this, we leveraged the domain knowledge from pharmacology and use a one-compartment Pharmacokinetic/Pharmacodynamic (PK/PD) mechanistic model to estimate drug response as a function of ASM dose. The parameters of the PK/PD model can be interpreted as high-dimensional propensity scores that summarize a patient’s responsiveness to a drug regime, such that any two patients with similar PK/PD parameters will exhibit similar responses under identical drug regimes.. To account for the effect of past medical history and cur-

rent medical conditions on drug responsiveness, these factors and the parameters from the PK/PD model are controlled for via a matching procedure as described in Section 5.2.

We use a single-compartment PK model to estimate the bloodstream concentration  $D_{i,t,j}$  of ASM  $j$  in patient  $i$  at time  $t$  (drug PK), and Hill's PD model (Hill, 1909) to estimate a short-term response to drugs:

$$\frac{dD_{i,t,j}}{dt} = -\frac{1}{\kappa_j} D_{i,t,j} + W_{i,t,j}, \quad (3)$$

$$Z_{i,t} = 1 - \sum_j \frac{D_{i,t,j}^{N_{i,j}}}{D_{i,t,j}^{N_{i,j}} + ED_{50,i,j}^{N_{i,j}}}. \quad (4)$$

Here  $\kappa_j$  is the average half-life of the drug (see Appendix 3 for half-lives),  $W_{i,t,j}$  is the body weight-normalized drug administration rate in units of mg/kg/h,  $N_{i,j}$  represents how responsive the patient is to drug  $j$ , and  $ED_{50,i,j}$  is the dosage required to reduce the patient's EA burden by 50%. Since  $N_{i,j}$  (the Hill coefficient) is constrained to be non-negative, a positive correlation between drug concentration and EA burden results in an  $N_{i,j}$  value of 0. The PD parameters were fit using *scipy*'s nonlinear least squares function. The estimated PD parameters reflect wide heterogeneity across patients as well, and indicate clearly which patients responded well to ASMs (shown in Figure 6).

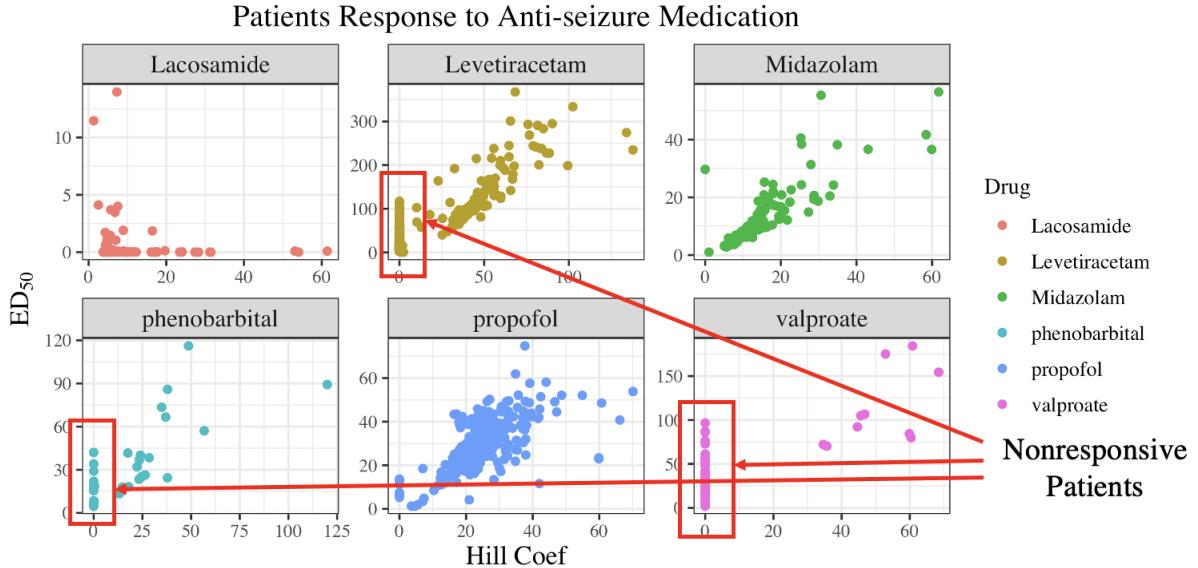


Figure 6: Hill coefficient vs.  $ED_{50}$  for the six drugs. Each point is a patient. The non-responsive patients with Hill coefficient of zero are highlighted.

## 5.2 Interpretable-and-Accurate Causal Inference

In this section, we discuss the causal inference method used to estimate the potential outcomes. Given the stakes involved and the high level of noise in the data, we chose an interpretable-and-accurate causal inference method, MALTS, to estimate cause-effect relationships. MALTS is an *honest* matching method that learns a distance metric using a subset of data as training set. Further, the learned metric is used to produce high-quality matched groups on the rest of the units (also called as estimation set). These matched groups are used to estimate heterogeneous causal

effects with high accuracy. Previous work on MALTS shows that it performs on-par with contemporary black-box causal machine learning methods while also ensuring interpretability (Parikh et al., 2020, 2019).

The conventional objective function of MALTS, described in Parikh et al. (2020), was designed to estimate the contrast of potential outcomes under binary “treatment.” In this paper, we adapt it to estimate conditional average potential outcomes for n-ary “treatment.” For our problem there are  $4 \times 2$  “treatment” arms – four levels of EA burden crossed by whether or not drugs were administered. We construct the matched group  $G_i$  for each patient  $i$  by matching on  $X_i = [\{C_{i,j}\}_j, \{N_{i,j}\}_j, \{ED_{50,i,j}\}_j]$  - the vector of pre-admission covariates and PD parameters. We estimate  $Pr[Y(e, \delta) = 1 | X = X_i]$  by averaging the observed outcomes for units in the matched group  $G_i$  with  $E_{\max}$  equals  $e$  and  $\bar{W}$  equals  $\delta$ . We use an analogous estimator for  $E_{mean}$ .

MALTS’ estimates of the conditional average potential outcome are *interpretable* because it is computed with the units in the matched groups. These matched groups can be investigated by looking at the raw data to examine their cohesiveness. One might immediately see anything that may need troubleshooting, and easily determine how to troubleshoot it. For instance, if the matched group does not look cohesive, the learned distance metric might need troubleshooting. Or, processing of the EEG signal might need troubleshooting if the max EA burden values do not appear to be correct. Or, the PK/PD parameters might need troubleshooting if patients who appear to be reacting to drugs quickly are matched with others whose drug absorption rates appear to be slower, when at the same time, the PK/PD parameters appear similar. We will demonstrate this with a matched group analysis in the next section.

## 6 Interpretable Matched Group Analysis

In this section, we provide an assessment of the quality of the matched groups. These types of analyses determine trust of the causal conclusions.

### 6.1 Stretch Coefficients Give Insight into the Matching Process

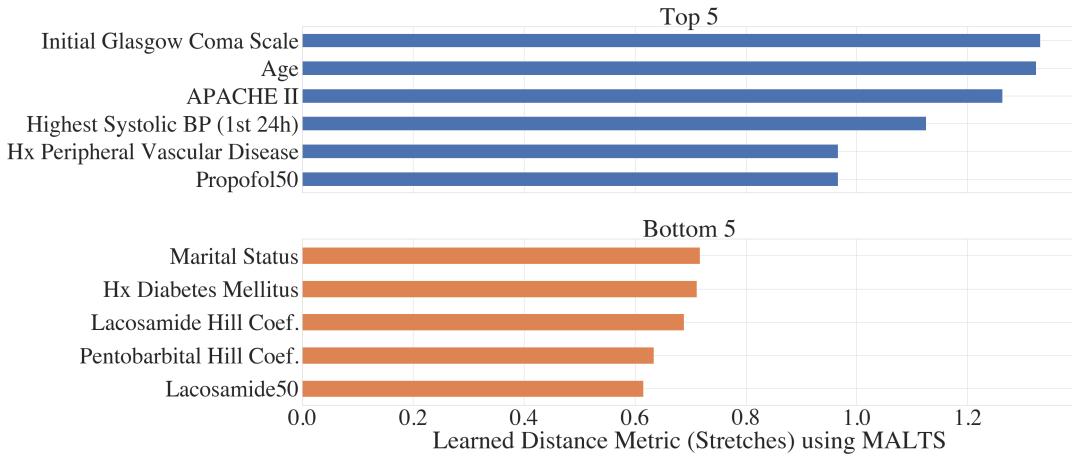


Figure 7: The top and bottom 5 variables, based on the average stretching weights in MALTS, when we are studying the effect of the maximum EA burden  $E_{\max}$ . BP = blood pressure; Coef = coefficient.; Lacosamide50 = concentration of Lacosamide that reduces EA burden by 50%; Propofol50 = concentration of propofol that reduces EA burden by 50%; Hx = History.

Through visualizing the stretch coefficients, one can gain insight into the relative importance of variables in the MALTS matching procedure. For max EA burden, one can see in Figure 7 that two medical scoring systems were both heavily weighted, with the iGCS Score being the most important variable and APACHE II score being the third most important variable. These two scoring systems capture a patient's level of consciousness and severity of illness. When considering that age and systolic blood pressure were the second and fourth most important variables, this shows that our matched groups essentially must consist of individuals that agree on these medical scores and biomarkers representing overall health and current level of neurologic impairment. In Figure 7, one can also see that the three least important variables to match on are Hill coefficients and  $ED_{50}$  parameters from one of the anti-seizure medications. This stands in contrast with the  $ED_{50}$  parameter for Propofol, which was one of the top five *most* important variables. This presents an interesting result: perhaps Propofol, a potent intravenous anesthetic drug used to treat seizures, including information about how it is prescribed, may be much more important in understanding the effect of seizure burden than its fellow anti-seizure medications, most of which are less potent.

## 6.2 Matched Groups are Validated by Neurologist's Chart Review

To ensure the validity of our causal conclusions, it is crucial that the matching process does not overlook major unobserved confounding factors. Inspired by similar approaches in the social sciences (Hasegawa et al., 2019), one can check for unobserved confounders by having a domain expert perform a post-facto analysis of the matched groups. If a domain expert who has access to all of a patient's medical information finds the patients in each matched group to be qualitatively comparable, this gives us confidence that we are controlling for all relevant sources of confounding.

This approach to considering unobserved confounding is well suited for medical data. In addition to factors that are easy to quantify, such as APACHE II scores, it is common for a patient to have a large volume of qualitative information along with quantitative data in the form of doctor's notes and documentation. As doctors' chart reviews are not restricted to quantitative information, this ensures that we are checking for qualitative and quantitative sources of unobserved confounding.

For our matched groups analysis, three practicing neurologists, Chart Reviewers 1, 2, and 3 (CR 1-3), were sent 3 randomly chosen matched groups and asked to perform a manual chart review of the selected patients. Based on these charts, the neurologists were asked to independently make a qualitative analysis of the patients within the matched groups and to report their outcome prognosis (chance of a poor neurologic outcome) and likelihood of experiencing a high EA burden.

From the results of the post-hoc analysis, as in Table 1, the three neurologists found no problematic sources of confounding, therefore validating our causal effect estimate. Moreover, from the reviewer's qualitative analyses, we can observe which factors each matched group was matched tightly on. For example, group three is tightly matched with all patients having similar APACHE II scores and all but one having relatively good prognoses. This contrasts with group one, where patients are tightly matched on acute neurological injuries at the cost of a tighter match on APACHE II score. Viewing what is tightly matched in each group provides a holistic evaluation of which factors have been properly controlled for, such as age, and which factors are either unimportant or lack the sample size to tightly match upon, such as many of the less common diseases.

## 7 Discussion

We presented four main points in this work: (1) First, we developed a novel framework that combines mechanistic modeling with a distance metric learning-based matching method to adjust for complex time-series confounders. (2) Second, we have provided, for the first time (to the best of our knowledge), an estimate of the causal effect of epileptiform activity (EA) on post-discharge outcomes in patients with critical illness. We find that higher EA burden indeed leads to worse neurologic outcomes (Figure 3), in a way that depends on the intensity (max burden over 6 hours) and duration (average burden over 24 hours) of EA. (3) Third, our results provide insights into individualized potential outcomes. For example, we show that patients with central nervous system infection or toxic metabolic encephalopathy are affected by EA to a higher extent compared to the average level in this cohort. (4) Finally, we leveraged the interpretability of our approach to validate our matched patients via chart review with the help of three neurologists. The general consensus in the chart review found that the matches were of high quality, matching together patients with similar prognoses.

**Clinical Implications.** Our findings have two primary implications for treatment of EA: (1) First, treatment should be based on EA duration as well as intensity. We find that intense periods of EA burden (max EA), even if relatively brief (6 hours) lead to worse outcomes. By contrast, sustained periods of EA (mean EA burden) show a binary relationship with outcome: EA < 50% has minimal effect, but EA  $\geq$  50% causes worse outcome. This suggests that interventions should put higher priority on patients with mean EA burden higher than 50%, while treatment intensity should be low and conservative when EA intensity is low. (2) Second, treatment policies should be based on admission profile, because the potential for EA to cause harm depends on age, past medical history, reason for admission, and other characteristics. For example, as our results suggest, patients with central nervous system infection or toxic metabolic encephalopathy should be monitored more closely with more robust treatment. By contrast, current treatment protocols used in hospitals tend to be generic, recommending that treatment be tailored based on the intensity or duration of EA (e.g., more aggressive treatment for status epilepticus), but providing little guidance on how to take other patient characteristics into account. As a result, treatment approaches vary widely between doctors. This suggests an opportunity to improve outcomes by personalizing treatment approaches.

**Results in context.** Our work builds on prior results demonstrating associations between EA, treatments, and neurologic outcomes. [Oddo et al. \(2009\)](#) studied a cohort of 201 ICU patients where 60% had sepsis as a primary admission diagnosis. They found that EA (seizures and periodic discharges) were associated with worse outcomes, after performing a regression adjustment for age, coma, circulatory shock, acute renal failure, and acute hepatic failure. However, these authors did not adjust for treatment with ASM, including phenytoin (given to 67% of patients), levetiracetam (62% of patients), lorazepam (57% of patients), and four other drugs. [Tabaeizadeh et al. \(2020\)](#) found that the maximum daily burden of EA/seizures, together with their discharge frequency, are associated with higher risk of poor outcome (mRS at hospital discharge 4–6) in 143 patients with acute ischemic stroke. However, they did not control for ASMs which were given to 83% of patients. Lack of adjusting for drug use is also found in the pediatric literature on EA ([Ganesan and Hahn, 2019](#)). Not adjusting for treatment is problematic because a growing number of studies suggest that aggressive treatment with ASM may be harmful. One example is the use of therapeutic coma for status epilepticus, where anesthetics such as pentobarbital or propofol are used to temporarily place the brain into a state of profoundly suppressed activity to stop EA while giving treatments of the underlying cause of EA to take effect. Recent

evidence shows that use of therapeutic coma is associated with worse outcomes, including a recent retrospective study of 467 patients with incident status epilepticus of [Marchi et al. \(2015\)](#) which found that therapeutic coma was associated with poorer outcome, higher prevalence of infection, and longer hospital stay ([Lin et al., 2017](#); [Rossetti et al., 2005](#)). However, because more aggressive treatment is reserved for more severely ill patients, these studies have also come under criticism for failing to adequately adjust for the type and severity of medical illness, and for the burden of epileptiform activity. Adequately adjusting for these factors has been challenging before now because of the complex interactions and feedback loops involved. However, without adjusting for these factors, it remains unclear whether the association between EA and poor outcomes is due to over-treatment, the underlying illness, or the direct effects of EA. Without an answer to this question, it has remained unclear whether current treatment approaches are helping or hurting patients.

We addressed this gap by introducing an analytic approach that is able to simultaneously account for the entwined and time-varying effects drug and EA burden, and their interactions with patient characteristics. One key component of our approach is adjusting for patients' pharmacodynamic (PD) parameters to account for heterogeneity among patients. Critically ill patients can be different in many ways including measured and unmeasured variables. PK/PD parameters provide a way to quantify the dynamics of the propensity of experiencing EA. The PK/PD parameters are important to take into account since they create spurious correlations impacting both the propensity of having high EA burden and the clinical outcome. By accounting for PK/PD parameters, we were able to adjust for exposure to anti-seizure drugs, such as phenytoin and pentobarbital, where the medications themselves may worsen outcomes. Because prior studies did not disentangle the potential harmful effects of EA and seizures from anti-seizure drugs, the field remained worried but uncertain. Another key innovation is our application of an advanced methodology designed specifically for causal inference using observational data. In the prior studies cited above, multivariate regression was used to adjust for potential confounders. The nature of observational data and multivariate regression (model misspecification) have made it impossible to establish a causal link between seizures and other EA vs. clinical outcome. The matching approach in MALTS, being a causal inference method, achieves both the flexibility of being free of model misspecification (non-parametric) and the interpretability of the learned weights, therefore creating less biased estimates of the causal effects. With this new approach, we are able to provide, for the first time, credible estimates of how much harm EA causes and in which types of patients. Moreover, MALTS comes with the additional advantage that one can easily perform post-hoc analyses of the matched groups, ensuring that the causal claims are accounting for potential unobserved confounders that an expert may be able to identify.

Our approach has several limitations that could be improved in future work. When evaluating the EA burden, it would be worthwhile in future work to consider the subtype of EA (GPD/LPD/LRDA), discharge frequency for periodic discharge patterns, the morphological features (such as seizure with/without triphasic waves), and the spatial extent of EAs. We currently do not have high quality human labels at the necessary resolution to pursue these tasks. On the other hand, the automatic EA annotator, based on a trained convolutional neural network, although not perfect, achieves similar inter-rater reliability as that of experts for the six normal/EA/seizure patterns ([Ge et al., 2021a](#)). To reduce this uncertainty in this study, we grouped these EA patterns into binary EA (seizure/GPD/LPD/LRDA) vs. non-EA categories (GRDA/normal/artifact). The definition of EA burden is also relatively coarse compared to those defined by [Ganesan and Hahn \(2019\)](#). The PK/PD model can be further improved by including more mechanistic or physiological detail, such as a context sensitive half-life for propofol ([Hughes et al., 1992](#)).

In summary, our results present a data-driven statistical causal inference approach to quantify the harm of EA in ICU. We not only confirm that EA burden (adjusted for ASM) are indeed harmful and worsen patients' neurologic outcomes, but careful analysis illustrates that there exist important subgroups of patients that are more affected by EA. Based on this, a future direction is to learn an interpretable optimal treatment policy for EA burden to improve patient outcomes.

Table 1: Three randomly chosen matched groups. We include doctors' prognosis and their qualitative estimate of risk of high EA burden. The notes column presents neurologists' remarks during chart review, based on medical notes not used for matching. The last column contains neurologists' notes on quality of matched groups: the third group is the tightest, followed by the second and the first. Crucially, this ordering matches an automatic measure of tightness produced by MALTS.

| Age                 | Gender | APACHE-II | Doctor's Poor Outcome Prognosis                     | Doctor's Estimate of High EA                       | Patient Summary   | Matched Group Analysis  |
|---------------------|--------|-----------|---|--|---|---|
| (a) Matched Group 1 |        |           |   |  |   |   |
| 57                  | Male   | 15        | CR1: CR2: 40% - 60%<br>CR3: 20% - 40%               | CR1, CR2, CR3: <b>40% - 60%</b>                    | History of coronary artery disease and tongue cancer. Admitted due to cardiac arrest. On arrival to the hospital, he is comatose.   | All three reviewers noted that the patients were similar in age with high risk of EA due to acute neurological injuries (ruptured brain aneurysm, brain tumor). However, as some patients following a cardiac arrest tend to carry a worse prognosis, while patients with refractory epilepsy have a very good chance of recovering, the range of APACHE-II scores (6-15) is broad, this group is not tightly matched.  |
| 54                  | Male   | 7         | CR1: CR3: 40% - 60%<br>CR2: 60% - 80%               | CR1: 40% - 60%<br>CR2: 60% - 80%<br>CR3: 20% - 40% | History of rectal cancer. Admitted due to a ruptured brain aneurysm. On arrival to the hospital, he is mildly confused.   |   |
| 57                  | Female | 4         | CR1: 80% - 100%<br>CR2: 60% - 80%<br>CR3: 40% - 60% | CR1, CR2: 60% - 80%<br>CR3: 40% - 60%              | With a brain tumor, admitted due to a seizure. Brain tumor has grown larger and is causing swelling in the brain  |   |
| 59                  | Male   | 6         | CR1, CR2, CR3: <b>80% - 100%</b>                    | CR1: 60% - 80%<br>CR2, CR3: 80% - 100%             | With epilepsy, admitted due to generalized convulsive seizures multiple times per day. Between seizures, he is cognitively normal.  |   |
| (b) Matched Group 2 |        |           |   |  |   |   |
| 62                  | Female | 3         | CR1, CR2: 80% - 100%<br>CR3: 60% - 80%              | CR1, CR3: 60% - 80%<br>CR2: 80% - 100%             | History of a benign brain tumor (meningioma), complicated by epilepsy, treated with anti-seizure medication. Admitted due to recurrent generalized convulsive seizures.   | Four of the patients are similar in age. The other patient is much younger, but her history of severe chronic illness makes her comparable to the other patients. All patients have relatively high risk for seizures / EA (risk factors: brain tumor, brain blood vessel malformation, brain hemorrhage, brain tumor, and epilepsy). Based on data available at hospital admission, the three patients with history of epilepsy or relatively static neurological injury (treated AVM with minor hemorrhage, remote meningioma, refractory epilepsy) have good short term prognosis compared to those with large intraparenchymal tumor with cerebral edema and midline shift. |
| 28                  | Female | 3         | CR1, CR3: 60% - 80%<br>CR2: 80% - 100%              | CR1: 0% - 20%<br>CR2: 80% - 100%<br>CR3: 60% - 80% | With multiple previous episodes of severe pneumonia and a large brain blood vessel malformation (arteriovenous malformation) that has caused focal seizures. Admitted due to pain and bleeding in the right eye.      |   |
| 54                  | Male   | 7         | CR1, CR3: 40% - 60%<br>CR2: 60% - 80%               | CR1: 40% - 60%<br>CR2: 60% - 80%<br>CR3: 20% - 40% | With rectal cancer in the past, admitted to the hospital because of a ruptured brain aneurysm. On arrival to the hospital he is mildly confused.  |   |
| 64                  | Male   | 6         | CR1: 60% - 80%<br>CR2: 80% - 100%<br>CR3: 40% - 60% | CR1, CR3: 60% - 80%<br>CR2: 80% - 100%             | History of seizures, admitted due to generalized convulsive seizure. At admission, there is a large frontal brain tumor.  |   |
| 59                  | Male   | 6         | CR1, CR2, CR3: <b>80% - 100%</b>                    | CR1: 60% - 80%<br>CR2, CR3: 80% - 100%             | With epilepsy, admitted due to generalized convulsive seizures multiple times per day. Between seizures, he is cognitively normal.  |   |
| (c) Matched Group 3 |        |           |   |  |   |   |
| 48                  | Male   | 4         | CR1, CR3: 60% - 80%<br>CR2: 80% - 100%              | CR1, CR3: 40% - 60%<br>CR2: 80% - 100%             | Post headaches led to discovery of a brain tumor, admitted to have the brain tumor surgically removed.  | The age range of these patients is relatively broad, however they have similar APACHE II scores, and similar levels of consciousness on arrival to the hospital. All of the patients in the group have a similar risk for seizures based on their baseline information: tumors or bleeding in the brain or (suspected) seizures. Based on data available at the time of admission, all of the patients have at least a moderately high chance ( $\geq 60\%$ ) of a good neurological outcome.   |
| 61                  | Female | 5         | CR1, CR2: 80% - 100%<br>CR3: 40% - 60%              | CR1, CR2: 80% - 100%<br>CR3: 60% - 80%             | Epilepsy since childhood. Free of seizures for several years. Admitted since she began having generalized convulsive seizures again.  |   |
| 49                  | Male   | 4         | CR1, CR3: 60% - 80%<br>CR2: 80% - 100%              | CR1, CR2: 80% - 100%<br>CR3: 40% - 60%             | Unexplained adult-onset blindness, neuropathy, vertigo, problematic coordination, and migraines. Admitted due to suspected seizures; where he suddenly develops tunnel vision and becomes unresponsive for 5 minutes. |   |
| 55                  | Male   | 4         | CR1: 60% - 80%<br>CR2: 80% - 100%<br>CR3: 40% - 60% | CR1, CR3: 60% - 80%<br>CR2: 80% - 100%             | Past hepatitis C and a benign neck tumor, admitted due to balance problems and a fall. Brain imaging shows bleeding around the brain.   |   |
| 55                  | Male   | 4         | CR1: 60% - 80%<br>CR2, CR3: 80% - 100%              | CR1, CR3: 40% - 60%<br>CR2: 80% - 100%             | Past hypertension, diabetes, recent back surgery, and a fever. Admitted due to unable to talk. No evidence of stroke, suspect he has had a seizure  |   |
| 42                  | Male   | 3         | CR1, CR2: 60% - 80%<br>CR3: 40% - 60%               | CR1, CR2, CR3: <b>60% - 80%</b>                    | History of lung cancer spread to brain, causing occasional seizures. Admitted due to difficulty to wake up after nap. Several brain tumors are bleeding.  |   |
| 61                  | Male   | 3         | CR1, CR2: 60% - 80%<br>CR3: 40% - 60%               | CR1, CR2, CR3: <b>60% - 80%</b>                    | History of lung cancer spread to liver. Admitted due to confusion. A new mass in brain. Suspect that cancer has spread to brain and may cause seizure.  |   |

## Acknowledgment

MBW was supported by NIH R01NS107291, R01NS102190, R01NS102574, RF1AG064312, R01AG062989; Department of Defense through a subcontract from Moberg ICU Solutions, Inc; and the Glenn Foundation for Medical Research and American Federation for Aging Research (Breakthroughs in Gerontology Grant); American Academy of Sleep Medicine (AASM Foundation Strategic Research Award); Football Players Health Study (FPHS) at Harvard University. MBW is the co-founder of Beacon Biosignals. AV, CR and HP were supported by NIH R01EB025021 and NSF FAIN 2147061. AV was supported by NSF DMS 2046880. KH is partially supported by NSF-DMS-1929298.

## References

- Benbadis, S. R. (2006). Introduction to sleep electroencephalography. *Sleep: A Comprehensive Handbook*, pages 989–1024.
- Blackwell, M. (2014). A selection bias approach to sensitivity analysis for causal effects. *Political Analysis*, 22(2):169–182.
- Boggs, J. (2002). Seizures and organ failure. In *Seizures*, pages 71–83. Springer.
- Burn, J. (1992). Reliability of the modified rankin scale. *Stroke*, 23(3):438–438.
- Chong, D. J. and Hirsch, L. J. (2005). Which EEG patterns warrant treatment in the critically ill? reviewing the evidence for treatment of periodic epileptiform discharges and related patterns. *Journal of Clinical Neurophysiology*, 22(2):79–91.
- Cormier, J., Maciel, C. B., and Gilmore, E. J. (2017). Ictal-interictal continuum: when to worry about the continuous electroencephalography pattern. In *Seminars in Respiratory and Critical Care Medicine*, volume 38, pages 793–806. Thieme Medical Publishers.
- De Marchis, G. M., Pugin, D., Meyers, E., Velasquez, A., Suwatcharangkoon, S., Park, S., Falo, M. C., Agarwal, S., Mayer, S., Schmidt, J. M., et al. (2016). Seizure burden in subarachnoid hemorrhage associated with functional and cognitive outcome. *Neurology*, 86(3):253–260.
- Ganesan, S. L. and Hahn, C. D. (2019). Electrographic seizure burden and outcomes following pediatric status epilepticus. *Epilepsy & Behavior*, 101:106409.
- Garoud, F., Lequeux, P., Bejjani, G., and Barvais, L. (2006). The influence of the dose on the time to peak effect of propofol. *European Journal of Anaesthesiology*.
- Gaspard, N., Manganas, L., Rampal, N., Petroff, O. A., and Hirsch, L. J. (2013). Similarity of lateralized rhythmic delta activity to periodic lateralized epileptiform discharges in critically ill patients. *JAMA Neurol*, 70(10):1288–1295.
- Ge, W., Jing, J., An, S., Herlopian, A., Ng, M., Struck, A. F., Appavu, B., Johnson, E. L., Osman, G., Haider, H. A., et al. (2021a). Deep active learning for interictal ictal injury continuum EEG patterns. *Journal of Neuroscience Methods*, 351:108966.
- Ge, W., Jing, J., An, S., Herlopian, A., Ng, M., Struck, A. F., Appavu, B., Johnson, E. L., Osman, G., Haider, H. A., Karakis, I., Kim, J. A., Halford, J. J., Dhakar, M. B., Sarkis, R. A., Swisher, C. B., Schmitt, S., Lee, J. W.,

- Tabaeizadeh, M., Rodriguez, A., Gaspard, N., Gilmore, E., Herman, S. T., Kaplan, P. W., Pathmanathan, J., Hong, S., Rosenthal, E. S., Zafar, S., Sun, J., and Brandon Westover, M. (2021b). Deep active learning for interictal ictal injury continuum EEG patterns. *Journal of Neuroscience Methods*, 351:108966.
- Hasegawa, R. B., Webster, D. W., and Small, D. S. (2019). Evaluating Missouri's handgun purchaser law: A bracketing method for addressing concerns about history interacting with group. *Epidemiology*, 30(3):371–379.
- Hill, A. V. (1909). The mode of action of nicotine and curari, determined by the form of the contraction curve and the method of temperature coefficients. *The Journal of Physiology*, 39(5):361–373.
- Hirsch, L. J., Fong, M. W., Leitinger, M., LaRoche, S. M., Beniczky, S., Abend, N. S., Lee, J. W., Wusthoff, C. J., Hahn, C. D., Westover, M. B., et al. (2021). American clinical neurophysiology society's standardized critical care eeg terminology: 2021 version. *Journal of Clinical Neurophysiology*, 38(1):1–29.
- Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). Densely connected convolutional networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2261–2269.
- Hughes, M. A., Glass, P. S., and Jacobs, J. R. (1992). Context-sensitive half-time in multicompartment: pharmacokinetic models for intravenous anesthetic drugs. *The Journal of the American Society of Anesthesiologists*, 76(3):334–341.
- Jing, J., Dauwels, J., Rakthanmanon, T., Keogh, E., Cash, S., and Westover, M. (2016). Rapid annotation of interictal epileptiform discharges via template matching under dynamic time warping. *Journal of Neuroscience Methods*, 274:179–190.
- Johnson, E. L. and Kaplan, P. W. (2017). Population of the ictal-interictal zone: the significance of periodic and rhythmic activity. *Clinical Neurophysiology Practice*, 2:107–118.
- Kim, J. A., Boyle, E. J., Wu, A. C., Cole, A. J., Staley, K. J., Zafar, S., Cash, S. S., and Westover, M. B. (2018). Epileptiform activity in traumatic brain injury predicts post-traumatic epilepsy. *Annals of Neurology*, 83(4):858–862.
- Lee, M. H., Kong, D.-S., Seol, H. J., Nam, D.-H., and Lee, J.-I. (2013). Risk of seizure and its clinical implication in the patients with cerebral metastasis from lung cancer. *Acta neurochirurgica*, 155(10):1833–1837.
- Lin, J. J., Chou, C. C., Lan, S. Y., Hsiao, H. J., Wang, Y., Chan, O. W., Hsia, S. H., Wang, H. S., Lin, K. L., Group, C. S., et al. (2017). Therapeutic burst-suppression coma in pediatric febrile refractory status epilepticus. *Brain and Development*, 39(8):693–702.
- Lin, L., Al-Faraj, A., Ayub, N., Bravo, P., Das, S., Ferlini, L., Karakis, I., Lee, J. W., Mukerji, S. S., Newey, C. R., Pathmanathan, J., Abdennadher, M., Casassa, C., Gaspard, N., Goldenholz, D. M., Gilmore, E. J., Jing, J., Kim, J. A., Kimchi, E. Y., Ladha, H. S., Tobochnik, S., Zafar, S., Hirsch, L. J., Westover, M. B., and Shafi, M. M. (2021). Electroencephalographic abnormalities are common in covid-19 and are associated with outcomes. *Annals of Neurology*, 89(5):872–883.
- Lucke-Wold, B. P., Nguyen, L., Turner, R. C., Logsdon, A. F., Chen, Y.-W., Smith, K. E., Huber, J. D., Matsumoto, R., Rosen, C. L., Tucker, E. S., et al. (2015). Traumatic brain injury and epilepsy: underlying mechanisms leading to seizure. *Seizure*, 33:13–23.

- Marchi, N. A., Novy, J., Faouzi, M., Stähli, C., Burnand, B., and Rossetti, A. O. (2015). Status epilepticus: impact of therapeutic coma on outcome. *Critical Care Medicine*, 43(5):1003–1009.
- Muhlhofer, W. G., Layfield, S., Lowenstein, D., Lin, C. P., Johnson, R. D., Saini, S., and Szaflarski, J. P. (2019). Duration of therapeutic coma and outcome of refractory status epilepticus. *Epilepsia*, 60(5):921–934.
- Nikbakht, F., Mohammadkhani Zadeh, A., and Mohammadi, E. (2020). How does the covid-19 cause seizure and epilepsy in patients? the potential mechanisms. *Multiple sclerosis and related disorders*, page 102535.
- Oddo, M., Carrera, E., Claassen, J., Mayer, S. A., and Hirsch, L. J. (2009). Continuous electroencephalography in the medical intensive care unit. *Critical Care Medicine*, 37(6):2051–2056.
- Osman, G. M., Araújo, D. F., and Maciel, C. B. (2018). Ictal interictal continuum patterns. *Current Treatment Options in Neurology*, 20(5):1–20.
- Parikh, H., Rudin, C., and Volfovsky, A. (2019). An application of matching after learning to stretch (MALTS) to the ACIC 2018 causal inference challenge data. *Observational Studies*, 5:118–130.
- Parikh, H., Rudin, C., and Volfovsky, A. (2020). MALTS: Matching after learning to stretch. [arXiv 1811.07415](https://arxiv.org/abs/1811.07415).
- Payne, E. T., Zhao, X. Y., Frndova, H., McBain, K., Sharma, R., Hutchison, J. S., and Hahn, C. D. (2014). Seizure burden is independently associated with short term outcome in critically ill children. *Brain*, 137(5):1429–1438.
- Rossetti, A. O., Hirsch, L. J., and Drislane, F. W. (2019). Nonconvulsive seizures and nonconvulsive status epilepticus in the neuro icu should or should not be treated aggressively: A debate. *Clinical Neurophysiology Practice*, 4:170–177.
- Rossetti, A. O., Logroscino, G., and Bromfield, E. B. (2005). Refractory status epilepticus: effect of treatment aggressiveness on prognosis. *Archives of Neurology*, 62(11):1698–1702.
- Rubinos, C., Reynolds, A. S., and Claassen, J. (2018). The ictal–interictal continuum: to treat or not to treat (and how)? *Neurocritical Care*, 29(1):3–8.
- Sun, H., Jia, J., Goparaju, B., Huang, G.-B., Sourina, O., Bianchi, M. T., and Westover, M. B. (2017). Large-scale automated sleep staging. *Sleep*, 40(10).
- Tabaeizadeh, M., Nour, H. A., Shoukat, M., Sun, H., Jin, J., Javed, F., Kassa, S., Edhi, M., Bordbar, E., Gallagher, J., et al. (2020). Burden of epileptiform activity predicts discharge neurologic outcomes in severe acute ischemic stroke. *Neurocritical Care*, 32(3):697–706.
- Tao, J. X., Qin, X., and Wang, Q. (2020). Ictal-interictal continuum: a review of recent advancements. *Acta Epileptologica*, 2(1):1–10.
- Trinka, E., Cock, H., Hesdorffer, D., Rossetti, A. O., Scheffer, I. E., Shinnar, S., Shorvon, S., and Lowenstein, D. H. (2015). A definition and classification of status epilepticus—Report of the ILAE Task Force on Classification of Status Epilepticus. *Epilepsia*, 56(10):1515–1523.
- Zafar, S. F., Postma, E. N., Biswal, S., Boyle, E. J., Bechek, S., O'Connor, K., Shenoy, A., Kim, J., Shafi, M. S., Patel, A. B., et al. (2018). Effect of epileptiform abnormality burden on neurologic outcome and antiepileptic drug management after subarachnoid hemorrhage. *Clinical Neurophysiology*, 129(11):2219–2227.

Zafar, S. F., Rosenthal, E. S., Jing, J., Ge, W., Tabaeizadeh, M., Aboul Nour, H., Shoukat, M., Sun, H., Javed, F., Kassa, S., et al. (2021). Automated annotation of epileptiform burden and its association with outcomes. *Annals of neurology*, 90(2):300–311.

## Appendix A Data

Table 2: Covariates (C) being matched

| Variable  | Value           |
|---|-----------------|
| Age, year, median (IQR)                                 | 61 (48 – 73)    |
| Male gender, n (%)                                      | 475 (47.7%)     |
| Race  |                 |
| Asian, n (%)  | 33 (3.3%)       |
| Black / African American, n (%)                         | 72 (7.2%)       |
| White / Caucasian, n (%)                                | 751 (75.5%)     |
| Other, n (%)  | 50 (5.0%)       |
| Unavailable / Declined, n (%)                           | 84 (8.4%)       |
| Married, n (%)  | 500 (50.3%)     |
| Premorbid mRS before admission, median (IQR)            | 0 (0 – 3)       |
| APACHE II in first 24h, median (IQR)                    | 19 (11 – 25)    |
| Initial GCS, median (IQR)                               | 11 (6 – 15)     |
| Initial GCS is with intubation, n (%)                   | 415 (41.7%)     |
| Worst GCS in first 24h, median (IQR)                    | 8 (3 – 14)      |
| Worst GCS in first 24h is with intubation, n (%)        | 511 (51.4%)     |
| Admitted due to surgery, n (%)                          | 168 (16.9%)     |
| Cardiac arrest at admission, n (%)                      | 79 (7.9%)       |
| Seizure at presentation, n (%)                          | 228 (22.9%)     |
| Acute SDH at admission, n (%)                           | 146 (14.7%)     |
| Take anti-epileptic drugs outside hospital, n (%)       | 123 (12.4%)     |
| Highest heart rate in first 24h, /min, median (IQR)     | 92 (80 – 107)   |
| Lowest heart rate in first 24h, /min, median (IQR)      | 71 (60 – 84)    |
| Highest systolic BP in first 24h, mmHg, median (IQR)    | 153 (136 – 176) |
| Lowest systolic BP in first 24h, mmHg, median (IQR)     | 116 (100 – 134) |
| Highest diastolic BP in first 24h, mmHg, median (IQR)   | 84 (72 – 95)    |
| Lowest diastolic BP in first 24h, mmHg, median (IQR)    | 61 (54 – 72)    |
| Mechanical ventilation on the first day of EEG, n (%)   | 572 (57.5%)     |
| Systolic BP on the first day of EEG, mmHg, median (IQR) | 148 (130 – 170) |
| GCS on the first day of EEG, median (IQR)               | 8 (5 – 13)      |
| History   |                 |

|  |             |
|--|-------------|
| Stroke, n (%)  | 192 (19.3%) |
| Hypertension, n (%)                                      | 525 (52.8%) |
| Seizure or epilepsy, n (%)                               | 182 (18.3%) |
| Brain surgery, n (%)                                     | 109 (11.0%) |
| Chronic kidney disorder, n (%)                           | 112 (11.3%) |
| Coronary artery disease and myocardial infarction, n (%) | 160 (16.1%) |
| Congestive heart failure, n (%)                          | 90 (9.0%)   |
| Diabetes mellitus, n (%)                                 | 201 (20.2%) |
| Hypersensitivity lung disease, n (%)                     | 296 (29.7%) |
| Peptic ulcer disease, n (%)                              | 50 (5.0%)   |
| Liver failure, n (%)                                     | 46 (4.6%)   |
| Smoking, n (%)   | 461 (46.3%) |
| Alcohol abuse, n (%)                                     | 231 (23.2%) |
| Substance abuse, n (%)                                   | 119 (12.0%) |
| Cancer (except central nervous system), n (%)            | 180 (18.1%) |
| Central nervous system cancer, n (%)                     | 85 (8.5%)   |
| Peripheral vascular disease, n (%)                       | 41 (4.1%)   |
| Dementia, n (%)  | 45 (4.5%)   |
| Chronic obstructive pulmonary disease or asthma, n (%)   | 139 (14.0%) |
| Leukemia or lymphoma, n (%)                              | 22 (2.2%)   |
| AIDS, n (%)  | 12 (1.2%)   |
| Connective tissue disease, n (%)                         | 47 (4.7%)   |
| Primary diagnosis  |             |
| Septic shock, n (%)                                      | 131 (13.2%) |
| Ischemic stroke, n (%)                                   | 85 (8.5%)   |
| Hemorrhagic stroke, n (%)                                | 163 (16.4%) |
| Subarachnoid hemorrhage (SAH), n (%)                     | 188 (18.9%) |
| Subdural hematoma (SDH), n (%)                           | 94 (9.4%)   |
| SDH or other traumatic brain injury including SAH, n (%) | 52 (5.2%)   |
| Traumatic brain injury including SAH, n (%)              | 21 (2.1%)   |
| Seizure/status epilepticus, n (%)                        | 258 (25.9%) |
| Brain tumor, n (%)                                       | 113 (11.4%) |
| CNS infection, n (%)                                     | 64 (6.4%)   |
| Ischemic encephalopathy or Anoxic brain injury, n (%)    | 72 (7.2%)   |
| Toxic metabolic encephalopathy, n (%)                    | 104 (10.5%) |
| Primary psychiatric disorder, n (%)                      | 35 (3.5%)   |
| Structural-degenerative diseases, n (%)                  | 35 (3.5%)   |
| Spell, n (%)   | 5 (0.5%)    |

|                                       |             |
|---------------------------------------|-------------|
| Respiratory disorders, n (%)          | 304 (30.6%) |
| Cardiovascular disorders, n (%)       | 153 (15.4%) |
| Kidney failure, n (%)                 | 65 (6.5%)   |
| Liver disorder, n (%)                 | 30 (3.0%)   |
| Gastrointestinal disorder, n (%)      | 18 (1.8%)   |
| Genitourinary disorder, n (%)         | 34 (3.4%)   |
| Endocrine emergency, n (%)            | 28 (2.8%)   |
| Non-head trauma, n (%)                | 13 (1.3%)   |
| Malignancy, n (%)                     | 65 (6.5%)   |
| Primary hematological disorder, n (%) | 24 (2.4%)   |

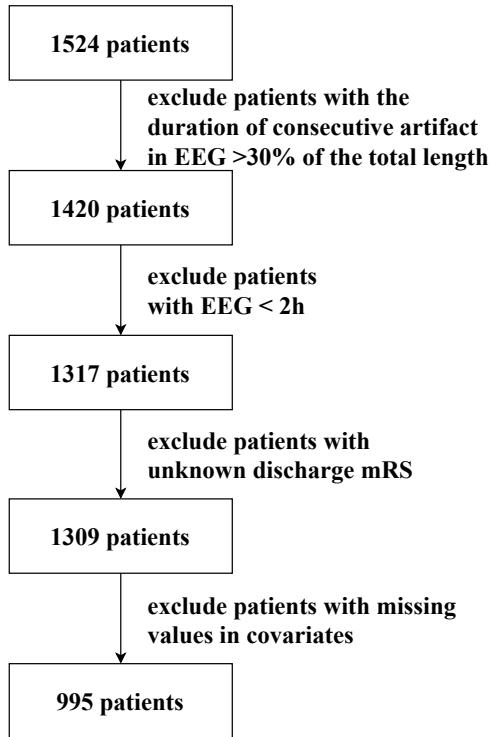


Figure 8: Data flowchart showing the preprocessing of patients

### A.1 Anti-Seizure Medications

Six drugs were studied: propofol, midazolam, levetiracetam, lacosamide, phenobarbital, and valproate. Propofol and midazolam are sedative antiepileptic drugs (SAEDs) which are given as continuous infusion, while the others are non-sedative antiepileptic drugs (NSAEDs) which are given as bolus. Only the period when there is EEG recording is used. The dose is normalized by body weight (kg). We use the half-lives from the literature (see Table 3) for calculating the drug concentrations  $D_{i,t,j}$  in the blood using the PK model.

Table 3: Half life for the anti-seizure medications used in the PD modeling.

| Drug          | Half Life  |
|---------------|------------|
| Propofol      | 20 minutes |
| Midazolam     | 2.5 hours  |
| Levetiracetam | 8 hours    |
| Lacosamide    | 11 hours   |
| Phenobarbital | 79 hours   |
| Valproate     | 16 hours   |

## A.2 Binning of EA Burden

For statistical efficiency and interpretability, we bin the EA burden ( $e$ ) into 4 levels – mild, moderate, severe, very severe – see Table 4.

Table 4: Binning of EA burden into 4 levels

| EA Burden                                 | Mild      | Moderate    | Severe      | Very Severe |
|---|-----------|-------------|-------------|-------------|
| $E_{\max}$ or $E_{\text{mean}}$           | 0 to 0.25 | 0.25 to 0.5 | 0.5 to 0.75 | 0.75 to 1   |
| Number of patients with $E_{\max}$        | 272       | 130         | 107         | 451         |
| Number of patients with $E_{\text{mean}}$ | 661       | 134         | 88          | 77          |

## A.3 Summary of Notation

Table 5: Primary table of notations.

| Symbol              | Description   |
|---------------------|---|
| $C_i$               | Vector pre-admission covariates such as age, vital signs, and medical history |
| $W_{i,t}$           | Sequence of ASMs administered during their stay in the hospital               |
| $E_{i,\max}$        | Worst 6 hour epoch of EA burden within a 24 hour period                       |
| $E_{i,\text{mean}}$ | Average amount of time a patient experiences EA in a 24 hour period           |
| $Y_i$               | Binarized post-discharge outcome (0 if mRS $\leq 3$ and 1 if mRS $> 3$ )      |
| $Y_i(e, w)$         | Potential outcome if EA burden is $e$ and total ASMs administered is $w$      |

## Appendix B Extracting EA Patterns from EEG

**Expert Labeling of EEG Signals.** The EEG signals of 1309 patients at Massachusetts General Hospital who met the inclusion criteria (described in Section 3) were recorded from September 2011 to February 2017. Of these, 82 randomly selected patients had their EEG signals re-referenced into 18 channels via a standard double banana bipolar montage (Benbadis, 2006) to create a time-frequency representation of a patient’s neurological state. These time-frequency representations were then segmented by domain experts using the labeling assistance tool *NeuroBrowser* (Jing et al., 2016) to identify occurrences of EA patterns. These 82 patients served as the training set for a semi-supervised procedure to create an neural network to automatically identify EA patterns.

**Neural Network Based Labeling of EEG Signals.** For the cEEG signal labeling procedure, the time-frequency representation was split into 10-second sliding windows with an 8-second overlap. These windows were then converted into an 8-bit color image and used as inputs to the recursive convolutional neural network DenseNet (Huang et al.,

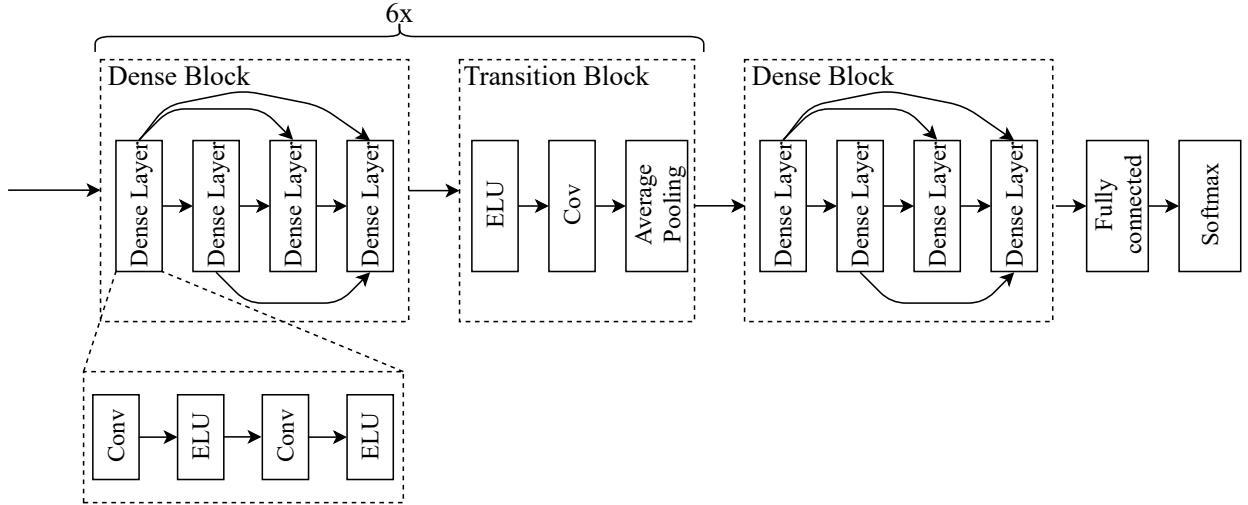


Figure 9: Structure of the DenseNet for automatic EA labeling.

2017); a Hidden Markov model was added to smooth the outputs (Ge et al., 2021a). By treating this as an image classification problem, this closely mimics the procedure performed by the domain experts using *NeuroBrowser*. DenseNet classified each 10-second window as either normal brain activity or one of 4 types of common EA patterns: (1) generalized periodic discharges (GPD), (2) lateralized periodic discharges (LPD), (3) lateralized rhythmic delta activity (LRDA) and (4) Seizure (Sz), as defined by the American Clinical Neurophysiology Society (Hirsch et al., 2021). The trained automatic EA annotator demonstrated accuracy for Seizure at 39% (human inter-rater agreement 42%), GPD at 62% (62%), LPD at 53% (58%), LRDA at 38% (38%), GRDA at 61% (40%), and normal brain-activity/artifact at 69% (75%), therefore, closely matching human performance up to the level of uncertainty one would get from interrater reliability studies.

**Operationalizing DenseNet.** We used DenseNet with 7 blocks (Figure 9). Each block included 4 dense layers. Each dense layer is comprised of 2 convolutional layers and 2 exponential linear unit (ELU) activations. In between each dense block was a transition block consisting of an ELU activation, a convolutional layer, and an average pooling layer. There were 6 transition blocks in total. The last two layers of DenseNet were a fully connected layer followed by a softmax layer. The loss function includes Kullback-Leibler divergence inversely weighted by the class ratio to account for imbalance among the EA classes. After fitting, it was observed that DenseNet’s classifications were much more volatile than the original data, with predictions abruptly changing from normal brain activity to EA patterns. This highlighted a limitation of traditional EEG classification from images, as the images were fed independently with no context about neighboring images beyond the 10-second window given. To correct for this volatility, the results of DenseNet were smoothed using a Hidden Markov Model. To smooth to a similar degree as the human labeled data, the probabilities of the transition matrix were fit on the 82 human-labeled patients. These probabilities were then used as the hidden state to smooth the output from DenseNet. We made the HMM first order due to precedent of first order HMMs providing good smoothing for other EEG problems (Sun et al., 2017).

The results of the automatic EA annotator resulted in accuracy for Seizure at 39% (human inter-rater agreement 42%), GPD at 62% (62%), LPD at 53% (58%), LRDA at 38% (38%), GRDA at 61% (40%), and others/artifact at 69%

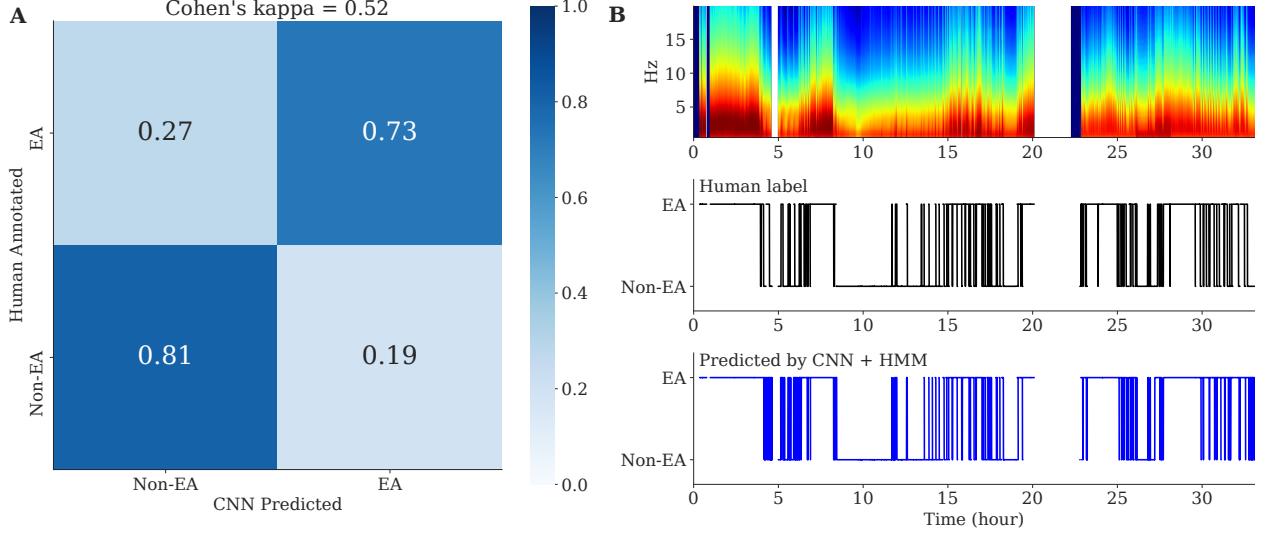


Figure 10: (A) Confusion matrix for the CNN prediction vs. human annotation, where each row represents the fraction of 2-second segments classified into EA (seizure/GPD/LPD/LRDA) or Non-EA (GRDA/other/artifact). The overall Cohen’s kappa is 0.52. (B) The top panel shows the spectrogram of the EEG signal of one example subject; the middle panel shows EA patterns annotated by a human expert for every 2 second interval. The bottom panel shows the EA pattern annotated by the CNN followed by HMM smoothing.

(75%). Therefore matching human performance. We further combined the classification into binary classes, EA (seizure/GPD/LPD/LRDA) vs. non-EA (GRDA/other/artifact) (Figure 10) to reduce the chance of error since these patterns are intrinsically on a continuous spectrum.

### Appendix C Sensitivity to the definition of EA burden

Throughout the analysis, the summaries of EA burden,  $E_{\max}$  and  $E_{\text{mean}}$  are quantized into four equally sized groups. This is done in accordance with clinician recommendations. In this section we evaluate the sensitivity of our analysis to these decisions. Specifically, we consider  $E_{\max} \in \{[0, \rho_1), [\rho_1, 0.5), [0.5, \rho_2), [\rho_2, 1.0]\}$  where the analysis in the paper specifies  $\rho_1 = 0.25$  and  $\rho_2 = 0.75$ . The interpretation of these parameters is as follows: the *mild* EA burden category allows for no more than  $100 \times \rho_1$  percent of a six hour window to be spent with EA and the *very severe* EA burden category allows for no less than  $100 \times \rho_2$  percent of a six hour window to be spent with EA. By varying these parameters we redefine which individuals are considered mild versus very severe EA during the analysis.

From sensitivity analysis to definition of EA burden, we observe following (see Figure 11):

- The potential outcome under mild EA burden ( $\mathbb{E}[Y([0.0, \rho_1), 0)]$ ) is mildly sensitive to changes in  $\rho_2$  which is expected. Further, we observe that the gradient of the same with respect to  $\rho_1$  is relatively flat, and  $\mathbb{E}[Y([0.0, \rho_1), 0)]$  is bounded between 0.525 and 0.6 for  $\rho_1 \in [0.1, 0.4]$ .
- Analogously the potential outcome under mild EA burden ( $\mathbb{E}[Y([\rho_2, 1.0], 0)]$ ) is mildly sensitive to changes in  $\rho_1$  and its gradient with respect to  $\rho_2$  is relatively flat, and  $\mathbb{E}[Y([0.0, \rho_1), 0)]$  is bounded between 0.645 and 0.705 for  $\rho_1 \in [0.6, 0.9]$ .
- The point estimates of  $\mathbb{E}[Y([0.0, \rho_1), 0)]$  are always strictly less than the point estimates of  $\mathbb{E}[Y([\rho_2], 1.0)]$

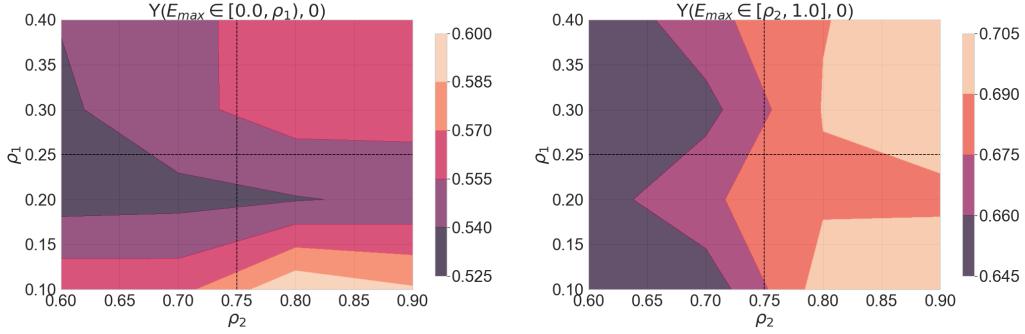


Figure 11: Sensitivity to quantization of EA burden into four levels.  $\rho_1$  is the boundary between mild and moderate EA burden and  $\rho_2$  is the boundary between severe and very severe EA burden. The contour plot shows estimated average potential outcomes –  $Y([0, \rho_1], 0)$  and  $Y([\rho_2, 1], 0)$  – for a range of  $\rho_1$  and  $\rho_2$ . We find that the gradient of contours is more or less flat and the estimates do not change by a large amount as the sensitivity parameters change.

## Appendix D Missingness Pattern

To check for possible selection bias, we compared the discharge mRS in patients with different missing conditions in Figure 12 where some of them were excluded in this cohort. We used the Mann-Whitney U test (nonparametric t-test) to compare the medians, since mRS does not follow a normal distribution.

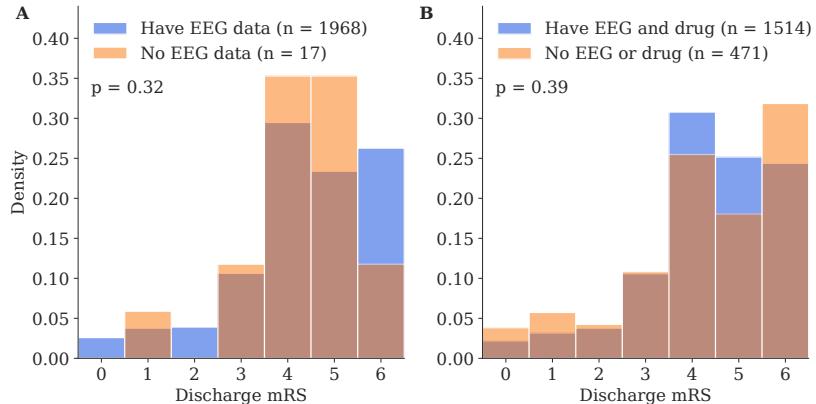


Figure 12: (A) The histogram of patients' discharge mRS (possible values are 0,1,2,3,4,5,6). The two subsets that are compared are patients who have EEG data ( $n = 1968$ ) vs. patients who do not have EEG data ( $n = 17$ ). To make the subsets comparison, the y-axis shows the density instead of the count. The p-value is from the Mann-Whitney U test of the two subsets. (B) Similar to A, but for patients who have EEG and drug data ( $n = 1514$ ) vs. patients who do not have EEG or drug data ( $n = 471$ ).

The results show that the medians of discharge mRS in patients with EEG, versus that in patients without EEG, are not significantly different; similarly, the medians in patients with both EEG and drug data, versus that in patients without EEG or drug data, are not significantly different neither. Therefore the missingness pattern can be considered as not influencing our results, hence the selection bias is negligible.

## Appendix E Robustness to causal assumptions

In providing our estimate of average potential outcome, our causal approach makes several important assumptions including: 1) pre-admission covariates and PD parameters are both potential sources of confounding and thus need to be controlled for 2) the post-discharge outcome,  $Y$ , is directly affected by **both** the level of EA burden,  $E_{max}/E_{mean}$  and the presence of Anti-seizure medications  $\bar{W}$ . In this section, we demonstrate how the estimation of potential outcome can vary with these assumptions.

### E.1 Assumption 1): The need to control for pre-admission covariates and PD parameters

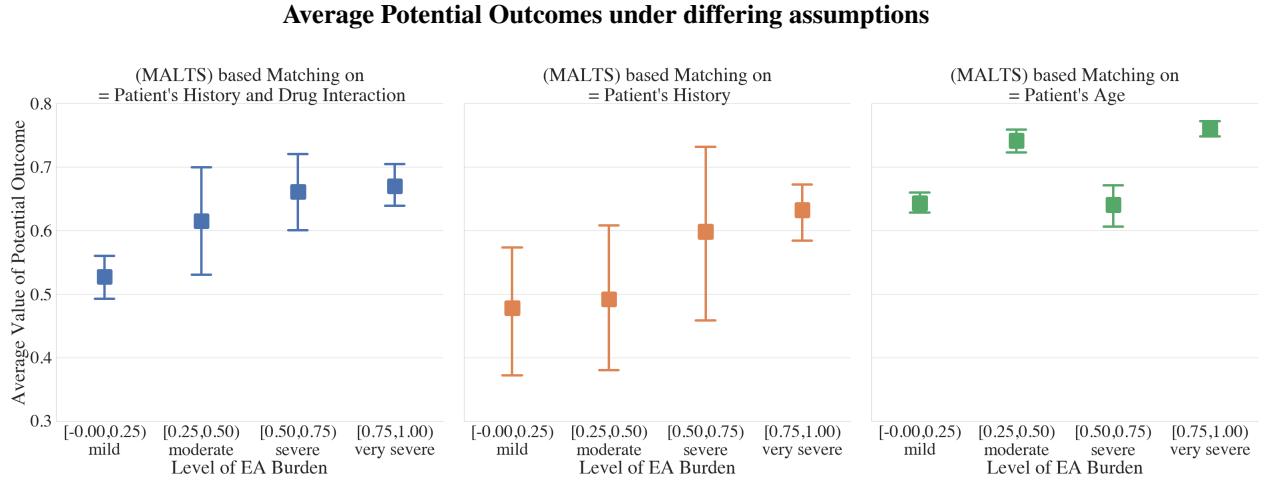


Figure 13: Estimated average potential outcome for different levels of  $E_{max}$  by matching on (left) all pre-admission covariates and PD parameters, (middle) all pre-admission covariates, and (right) only age of the patients.

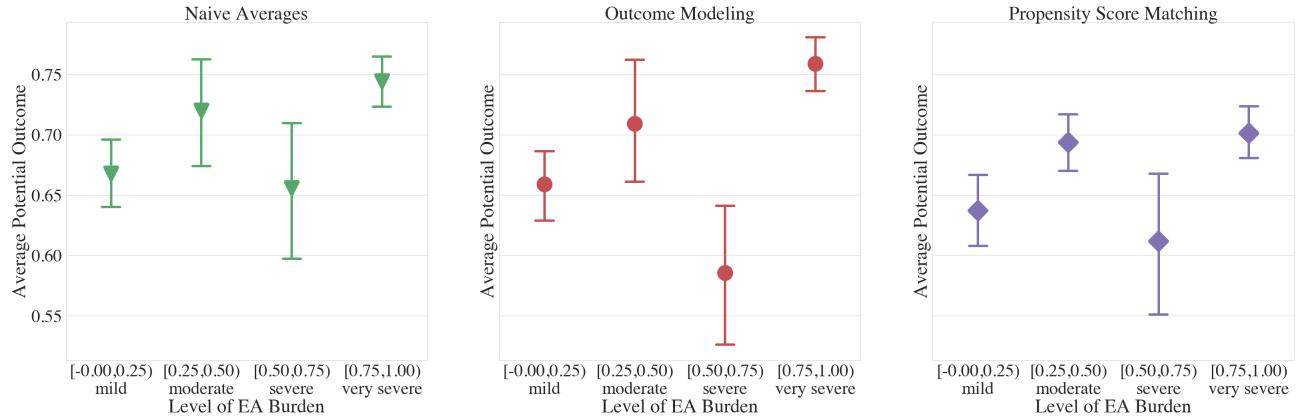


Figure 14: Estimated average potential outcomes computed using (left) Naive Average approach, (middle) Outcome modeling approach, and (right) Propensity Score matching.

Previously, it was posited that pre-admission covariates such as age and diagnosis and PD parameters could be large sources of confounding in the estimation of average potential outcomes. In this section, we investigate this assumption by having MALTS create matched groups based on fewer and fewer factors and comparing the resulting average potential outcomes.

The left side of Figure 13 shows the estimated average potential outcome when MALTS controls for only one, albeit important, variable, age. The results do not show a monotonic relationship between EA burden and average potential outcome. When matching on all pre-admission covariates but no PD parameters using MALTS, while the monotonic relationship between EA burden and average potential outcome is now clear, the uncertainty in the estimates and the shape of the trend differs. In particular, without adding in the information from the ASM's PK/PD models, one tends to underestimate the probability that a patient would leave the hospital impaired or dead.

## **E.2 Assumption 2): Post-discharge outcome are a function of the level of EA burden and the presence of Anti-seizure medications**

In this section, we compare our method, which posits that both the level of EA burden and the presence of Anti-seizure medications are the only two causal factors with a “Naive Average” approach which posits that EA burden is the **only** causal factor, an “Outcome Modeling” approach that treats all of the factors in our study as having a direct causal effect on the outcome, and a Propensity score approach, which performs a causal estimation, albeit under differing assumptions.

In the Naive Average approach, at each EA burden,  $\frac{1}{3}$  of the data is left out and the probability of leaving the hospital impaired is computed on the remaining  $\frac{2}{3}$  of the data. This procedure is repeated 15 times and the mean and standard deviation of the replicates are reported as the left-most figure in Figure 14. The choice of 15 and  $\frac{2}{3}$  was done to match as closely as possible the 15 replicates and 2:1 training to testing ratio that was used by MALTS.

In the outcome modeling approach, which takes up the middle of Figure 14, we perform a logistic regression where we regress the post-discharge outcome against EA burden, the presence of anti-seizure medications, and all of the factors that MALTS matched such as pre-admission covariates and PD parameters. Note that this approach makes the assumption that there are no interactions between the regressors, which goes contrary to our understanding of the treatment procedure, as factors such as age and diagnosis have a known interaction with a patient's response to anti-seizure medications. Like the naive averages approach, we perform 15 replicates of a logistic regression with the same 2:1 train/test used in the Naive Averages approach and MALTS approach.

On the right of Figure 14, we have the average potential outcome computed with a common approach to causal estimation, propensity score matching. Unlike MALTS which matches together patients directly on their covariates, propensity score matching is based on matching together patients based on their probability of being within the treatment or control arm. This makes the stronger assumption that the probability of being within the treatment or control arm can be modeled parametrically, in this case as using a logistic regression.

The results of these three approaches all yield similar results, showing an approximately sinusoidal relationship between EA burden and average potential outcome. This differs from the original MALTS result in the top left of Figure 13 which shows a clear monotonic relationship between EA burden and average potential outcome. As MALTS is the only method that takes a causal approach without making the strong parametric assumptions in propensity score matching, this seems to hint that perhaps the lack of control for confounding variables has been throwing off the regression based approaches to analyzing the damage caused by EA burdens.

## Appendix F Sensitivity Analysis for Unobserved Confounding

In this section, we study how sensitive our inferences are to unobserved confounding. In particular, we study the sensitivity to an unobserved confounder that correlates patients' post-discharge outcome with  $E_{max}$ . We would like to see if the presence of an unobserved confounder we failed to control for could have biased our inferences. We can encode the effect of an unobserved confounder using a selection bias function  $q(e)$  with sensitivity parameter  $\psi$ . This approach is similar to the one proposed in [Blackwell \(2014\)](#). We parameterize  $q(\cdot)$  as a logarithmic function of  $e$ .

$$\begin{aligned} q(e) &= \mathbf{E}[Y_i((e, 0)) | E_{max,i} = e, \bar{W}_i = 0] - \mathbf{E}[Y_i((e, 0)) | E_{max,i} \neq e, \bar{W}_i = 0] \\ &= \psi \ln(e + 1) \end{aligned}$$

When  $\psi$  is positive (negative), this indicates that patients with observed *bad* (*good*) outcomes also have high observed EA burden. This parametric form also assumes that a patient with low  $E_{max}$  is affected less by an unobserved confounder  $U$  compared to a unit with higher  $E_{max}$  with the marginal increase tapering off as the  $E_{max}$  increases. This is congruent with the neurologist's intuition that a perfectly healthy individuals with normal brain activity will be affected less by an unobserved confounder  $U$ .

To perform the sensitivity analysis, we apply the following debiasing to the observed outcome and re-estimate the average potential outcomes:

$$Y_i^{debiased} = Y_i - q(E_{max,i})(1 - P(E_{max,i}|X = X_i)).$$

If the unobserved confounding does not large impact on the estimation of average potential outcome, then the estimated potential outcome under very severe EA burden ([0.75,1.0]) will be more than average potential outcome under mild EA burden ([0.0,0.25]).

Our sensitivity analysis found that point estimate of potential outcome under very severe EA burden is always worse than the potential outcomes under mild EA burden for a range of sensitivity parameter  $\psi$  between [-1,1]. We further find that our inference is statistically significant for a wide range of  $\psi$ :  $-1.0 \leq \psi \leq 0.50$ . The sensitivity highlights that the conclusions from our study and analysis are insensitivity to high levels of unobserved confounding.

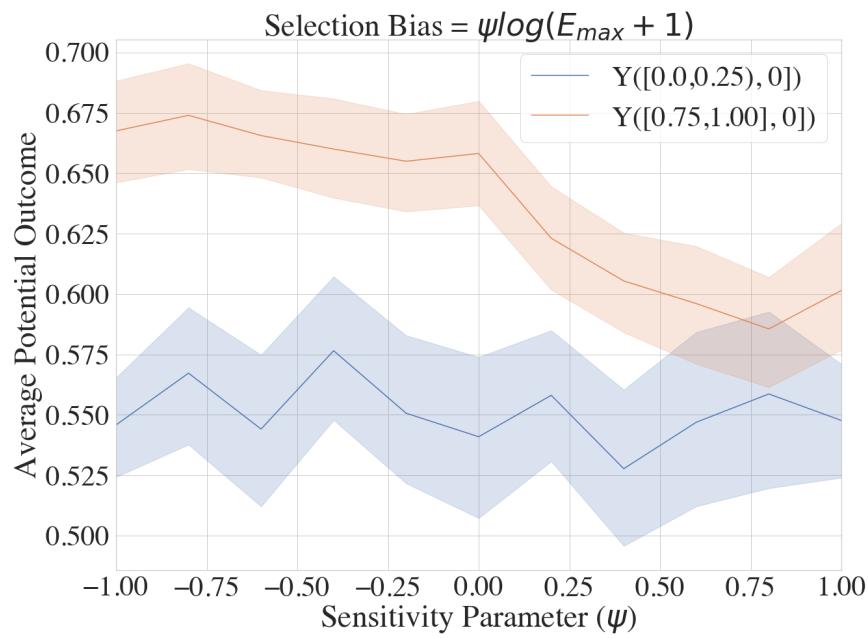


Figure 15: Sensitivity to unobserved confounding The results show that even at very high levels of selection bias, the effect of EA burden is not lost, indicating a degree of robustness in our results.