

QUANTIFYING THE HARM CAUSED BY SEIZURE-LIKE BRAIN ACTIVITY IN CRITICALLY ILL PATIENTS, USING INTERPRETABLE CAUSAL INFERENCE COMBINING MATCHING AND MECHANISTIC MODELS

Harsh Parikh*, Kentaro Hoffman*, Haoqi Sun*, Wendong Ge, Jin Jing, Rajesh Amerineni, Lin Liu, Jimeng Sun, Sahar Zafar, Alexander Volfovsky**, Cynthia Rudin**, M. Brandon Westover**

* Co-first authors, ** Co-senior authors.

Abstract

Seizures and seizure-like brain activity (“Epileptiform activity”, EA) are common in critically ill patients. Multiple studies have shown an association between high burden of EA and worsened clinical outcomes. However, without better control of factors such as anti-seizure medication administration and individual pharmacodynamic characteristics, it is currently unknown if EA causes adverse outcomes or is merely a side effect of critical illness. This problem is one of a class of complex observational causal inference problems common to many areas of healthcare, where patient treatment changes over time in response to newly generated data. We propose a framework for handling these types of complex causal inference problems, which uses a combination of individualized pharmacokinetics and pharmacodynamics parameters, and leverages interpretable matching methods to control for confounders including demographic variables, medical history, and pharmacodynamics parameters. Interpretability is extremely useful for troubleshooting complex problems, and our full process and all of its parameters are interpretable. We use our framework to show that, in a single-center retrospective study of 995 critically ill

patients, an untreated six-hour window with EA burden reaching more than 75% increases the probability of a poor neurological outcome (post-discharge Mmodified Rankin Score ≥ 4) by 16.7% on average (normalized to baseline), while mild but lasting EA with an average burden $\geq 50\%$ increases the risk of a poor outcome by 11.2%. Augmented by its interpretability, the estimated effect on outcome is validated by neurologists. This work provides causal inference-based and medically validated evidence that neurological outcomes for severely ill patients could be improved by reducing EA burden.

1 Introduction

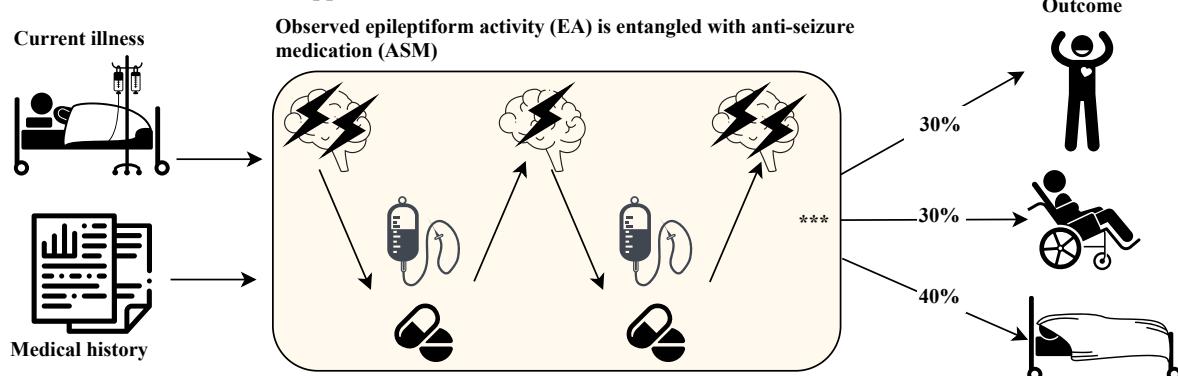
A large number of critically ill patients suffering from stroke, organ failure, cardiac arrest, brain trauma, or other serious conditions experience seizure-like brain activity called “epileptiform activity” (EA, also called ictal-interictal continuum activity, see [Hirsch et al., 2021](#)). There is a growing body of literature indicating that EA is *associated* with poor outcomes ([Oddo et al., 2009](#)). For example, prolonged EA is associated with increased in-hospital mortality, and survivors often suffer from a functional and cognitive disability ([Lalgudi Ganesan S, 2019; Rossetti et al., 2019; JA et al., 2018](#)). However, there is still a debate as to whether (a) EA is part of a causal pathway that worsens a patient’s outcomes that requires immediate treatment, *or* (b) the worsened outcomes is due to mechanisms other than EA such as the medications, and some EA can be tolerated ([Chong DJ, 2005; Rubinos et al., 2018; Osman et al., 2018; Johnson and Kaplan, 2017; Tao et al., 2020; Cormier et al., 2017](#)).

The question of whether EA leads to worse outcomes is an example of inferring causality from time-series with exposure-confounder feedback using observational data. In critical ill patients, the patient’s EA can be tracked as a time-series with continuous electroencephalogram (EEG) monitoring, which is involved in a feedback: anti-seizure medications (ASM) is given based on patient’s EA, and in turn EA is affected by ASM. To answer if EA has a causal effect on outcome, one would need to randomly “assign” different levels of EA while keeping the ASM constant, and follow up the patient’s outcome. Such randomized clinical trial would be neither possible or ethical, which limits our ability to infer the causal effects.

Existing approaches use prognostic modeling to determine if EA is part of a causal pathway affecting

patient's outcome. Researchers use regression models to control for the patient's medical history and demographic factors (Payne et al., 2014; De Marchis et al., 2016; Zafar, 2018; Muhlhofer et al., 2019). The resulting regression coefficient for EA is often interpreted 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 sufficient to control for the entanglement between EA and ASM due to their feedback (see Figure 1). Using conventional prognostic modeling approaches can put one at the risk of misinterpreting the association between *high levels of ASM* and *bad outcomes* to be causal, as well as bias due to model misspecification.

Observational data: What Happened



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

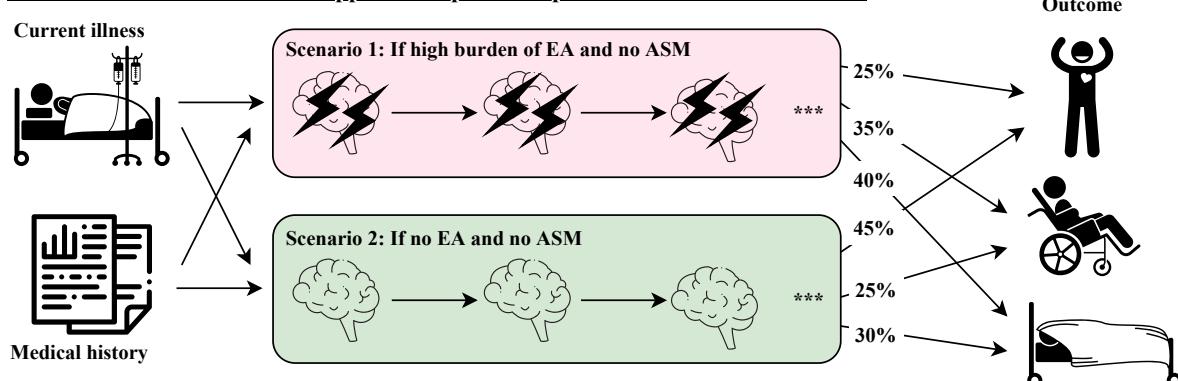


Figure 1: *Upper*: Illustration showing that observed epileptiform activity (EA) and treatment decision form a feedback loop, that is also influenced by current illness and medical history (left). The entire time-series of EA and ASM influence patient outcomes. Possible outcomes include return to normal 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.

In the presence of strong confounding factors in an observational study, it is common to estimate causal

effects using matching methods (Rosenbaum and Rubin, 1983; Imbens and Rubin, 2015; Stuart, 2010; Diamond and Sekhon, 2013; Wang et al., 2021). Matching procedures pair *similar* patients across treatment arms to compute causal effect(s). Here, we extend the matching approach to incorporate time-series data with exposure-confounder feedback using an interpretable method, Matching After Learning to Stretch (MALTS) (Parikh et al., 2020). We also control for important mechanistic pharmacokinetic/pharmacodynamic (PKPD) parameters to better characterize individualized response to medications. We aim at achieving a less biased estimation of the causal effect of EA on outcome, harvesting the power of interpretable causal inference and mechanistic modeling. Importantly, the validity of the estimate is supported by detailed clinical chart review of the matched groups.

2 Background and Problem Definition

Patient Characteristics 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 additional 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 EEG signal quality 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 C).

For each patient, we collected information from a variety of sources including demographic (gender, marital status, and age), clinical factors (substance abuse, history of seizures or epilepsy, or chronic kidney disease), 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 they are admitted to the hospital for treatment, these are referred to as the *pre-admission variables*. A full list of pre-admission variables can

be found in Table 2.

Outcomes of interest Once a patient has stabilized (or has passed away), they are officially 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, 2018). Our work also uses this dichotomized Modified Rankin Scale as our outcome of interest, with Y equal to 1 representing a patient discharged with serious disabilities and 0 representing a patient without serious disabilities.

Time series data: drug treatments and EA over time. After initial 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 EA experienced by the i -th patient ($\{Z_{i,t}^\omega\}_{t=1}^T$) with a ω hours sampling frequency and (2) a 6-dimensional vector ($\{W_{i,t}\}_{t=1}^T$) representing when the i -th patient received one of the 6 most common ASMs, Lacosamide, Levetiracetam, Midazolam, Phenobarbital, Propofol, and Valproate. We use $\omega = 6$ hours as it is a reasonable amount of time for a patient to be visited and to observe the effects of the ASMs (Garoud1 et al., 2006).

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 hours recording period.

The former measures the prevalence of EAs 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 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 timeseries of EA burden and drug exposures $Y(\{E_t, W_t : t = 1, \dots, T\})$ 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 $\bar{W}_i = 1_{\sum_t W_{i,t} > 0}$. Our estimand of interest is the probability a patient is discharged with severe disabilities if the patient has EA burden (either E_{\max} or E_{mean}) equal to e and was not treated with ASMs. This can be represented as:

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

Here, Y 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 numbers of patients in each category. The choice of cutoffs were influenced by animal models which showed that an EA burden of 50% serves as an important indicator of damaging epileptiform activity (Trinka et al., 2015). A sensitivity analysis to these choices is provided in Appendix C.

The variables we control for: preadmission 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 the physi-

cian, potentially confounding the estimated harm caused by EA with the harm due to diagnosis or patient characteristics. To address this, a collection of 70 preadmission 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 are much less responsive. This in turn, can directly affect the amount and number of ASMs that a patient receives as well as 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. Details of the model can be found in Section 4.1.

3 Main Results: Adverse Effects of EA Burden

Average Effect of Max EA Burden. Figure 2(a) illustrates one of our main results, that those with higher levels of E_{max} are at higher risk of being discharged with poor 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 (0 to 0.25), mild, to (0.75 to 1), very severe.

Average Effect of Mean EA Burden. Figure 2(b) shows our other main result, that those with higher levels of E_{mean} are also at higher risk of being discharged with poor outcomes. However unlike E_{max} , the risk caused by increasing E_{mean} spikes up when a patient goes above 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.

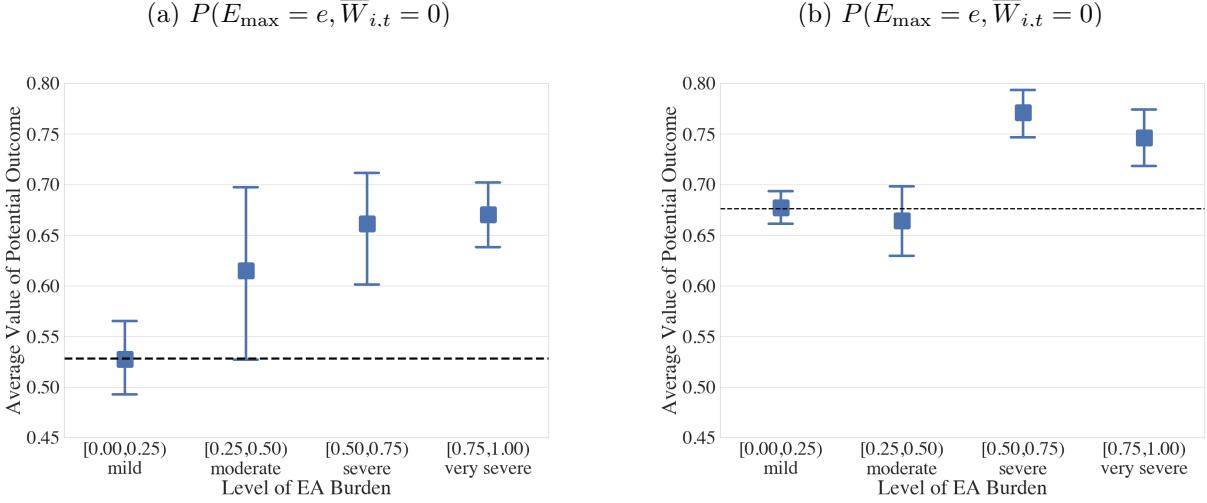


Figure 2: The probability of a bad outcome mRS for either Mild, Moderate, Severe, or Very Severe EA burden. EA Burden is quantified as (Left): Maximum EA in a 6-hour moving average window; (Right): Mean EA in a 6-hour moving average window. In both scenarios, an increase in IIC burden leads to a worse outcome for the patient. Outcome worsens linearly for E_{\max} , whereas for E_{mean} , there is a jump at approximately 0.5.

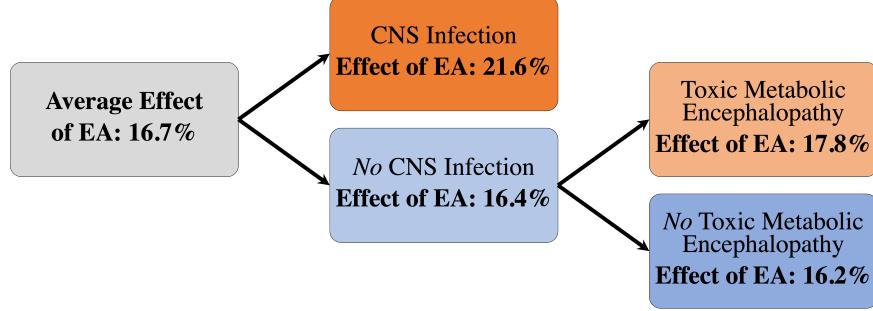
Heterogeneous Effects of Max EA Burden. While increases in EA burden tend to lead to worse outcomes, there is significant heterogeneity in the size of the effect due to each patient’s pre-admission covariates. Let us quantify the average effect of the max EA burden τ as follows. τ is the ratio of expected outcomes for those with high EA burden over the expected outcome of those with low EA burden minus one:

$$\tau = \frac{\mathbb{E}[Y(E_{\max} \geq 0.75, \bar{W} = 0)]}{\mathbb{E}[Y(E_{\max} < 0.25, \bar{W} = 0)]} - 1. \quad (2)$$

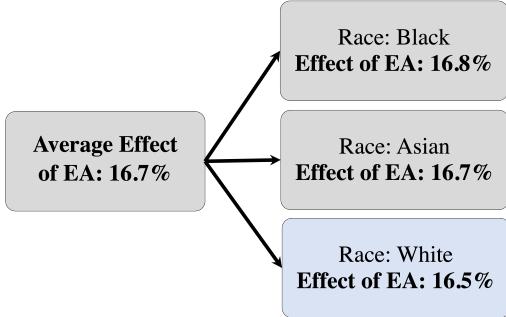
Figure 3(a) shows the results of recursively partitioning the population level τ based on a variety of important neurological factors. Here, we partition on medical history and diagnosis using a decision tree. We can see that those with an infection in the central nervous system and those with toxic metabolic encephalopathy are at higher risk of a worse outcome if they had a large increase in E_{\max} burden. This is thought to be the result of a CNS infection and EAs leading to a higher inflammatory response in the patient, potentially leading to permanent damage.

We also examined race and gender as possible effect modifiers of EA burden. Figure 3(b) shows that race

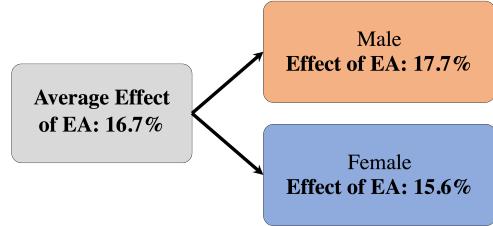
does not seem modify the risk from increases in E_{max} . On the contrary, we observe that male patients are more susceptible to very severe E_{max} worsening the chances of recovery compared to female patients (see Figure 3(c)).



(a) Recursive partitioning of covariate space and respective estimates of conditional average effects of EA



(b) Effect of EA on patients from different races



(c) Effect of EA on male and female patients

Figure 3: Estimated conditional average effects of EA burden in sub-group created by partition of covariate space based on neurological diagnosis at the time of admission. (a) Recursively partitions the space using Gini splitting to find the most important splits; (b) Partitions the space into patients according their 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 into male and female patients. Orange fill in the boxes implies that the subgroup has higher average estimated causal effect from EA than the cohort mean and blue implies lower. Subgroups in orange are likely fare worse as a result of a higher EA burden.

4 Framework

In this section, we discuss the methodology used to estimate the average potential outcomes from Section 3.

We divide the estimation pipeline into three stages (Figure 4):

1. In the *first stage* (Section B), we calculate E_{max} and E_{mean} . To do this, we need to classify EEG

signal into seizure-like EA behavior, and non-seizure like behavior. Doing this using human annotators would be extremely expensive, so we use a convolution neural network (CNN) trained on the human annotators' classifications of each 10 second window into non-EA brain behavior and EA in a semi-supervised fashion. We use the predictions to compute EA time series (Z_t^ω). As described in Section 3, E_{\max} and E_{mean} are computed directly from Z_t^ω . Details are in the appendix in Section B.

2. In the *second stage* (Section 4.1), we fit a personalized pharmacokinetic/pharmacodynamic (PK/PD) model to each patient's response to ASM (Hill, 1909).
3. Finally, in the *third stage* (Section 4.2), we combine the pre-admission covariates, such as baseline demographic data and the nature and severity of the present illness, as well as the PK/PD parameters estimated in the second stage, to control for potential confounding and 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 (MALTS, Parikh et al., 2020).

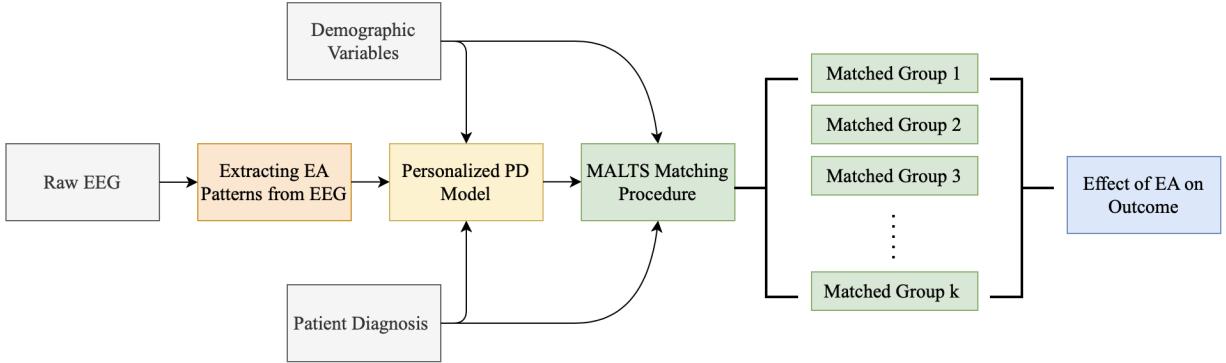


Figure 4: The overall analysis framework, consisting of four parts indicated by different colors: EA burden computation, individual PK/PD modeling, MALTS matching, and effect estimation.

4.1 Mechanistic Pharmacological Model

As noted in section 2, as ASM treatments come from an observation-treatment cycle which in turn allows doctors to make modifications to the type and dosage of ASMs being administered, this can potentially

confound the relationship between ASMs and a patient's final outcome. The responsiveness of a patient to ASMs can be due to a variety of factors such as past medical history, current medical conditions, age. However, the infrequency of some rare medical conditions makes it difficult to learn a satisfactory model of drug response that incorporates all of the relevant medical factors. To account for this, we leveraged knowledge from the mechanistic modeling community to create a one-compartment Pharmacokinetic/Pharmacodynamic (PK/PD) model modeling 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. To account for the effect of past medical history and current medical conditions on drug responsiveness, these factors and the parameters from the PK/PD model are controlled via a matching procedure in as described in Section 4.2.

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

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

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

Here κ_j is the average half-life of the drug (see Appendix 3 for half-lives), $W_{i,j,t}$ 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 as a clear indication of which patients responded well to ASMs (shown in Figure 5).

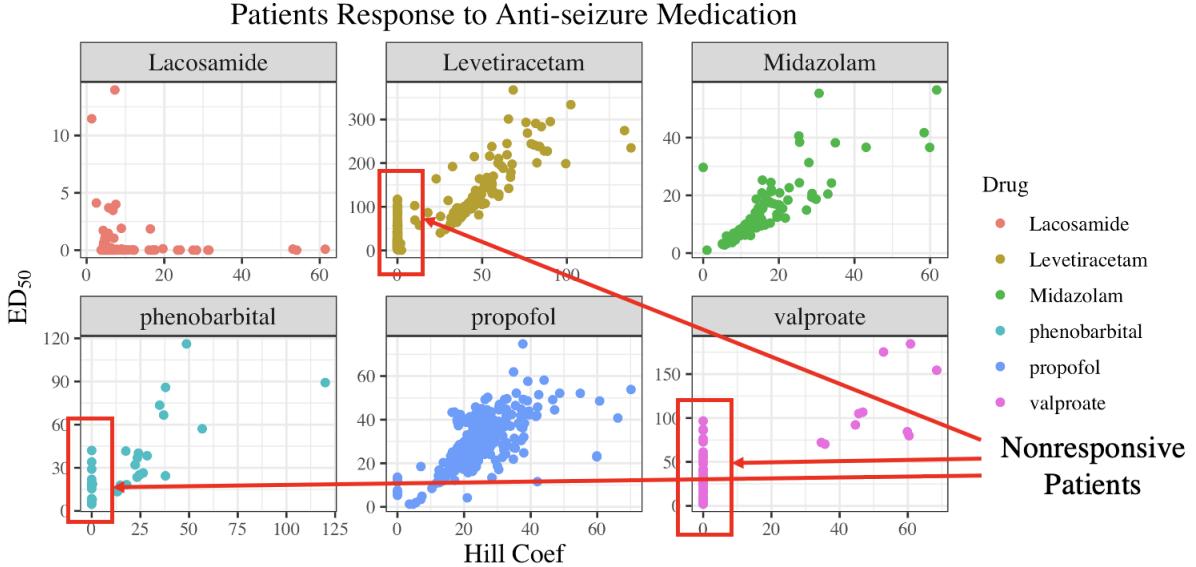


Figure 5: Hill coefficient vs. ED_{50} for the six drugs. Each point is a patient. The non-responsive patients with Hill coefficient of zero are highlighted.

4.2 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 the 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 of interest, there are 4×2 “treatment” arms – four levels of EA burden based 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 the conditional average potential outcomes $\mathbb{E}[Y(e, \delta)|X = X_i]$ by averaging the observed outcomes for units in the matched group G_i with E_{\max} equals e and \bar{W} equals δ . We use an analogous estimator for E_{mean} .

MALTS' estimates of 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 its 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.

5 Matched Group Analysis

5.1 Stretch Coefficients Give Insight into the Matching Process

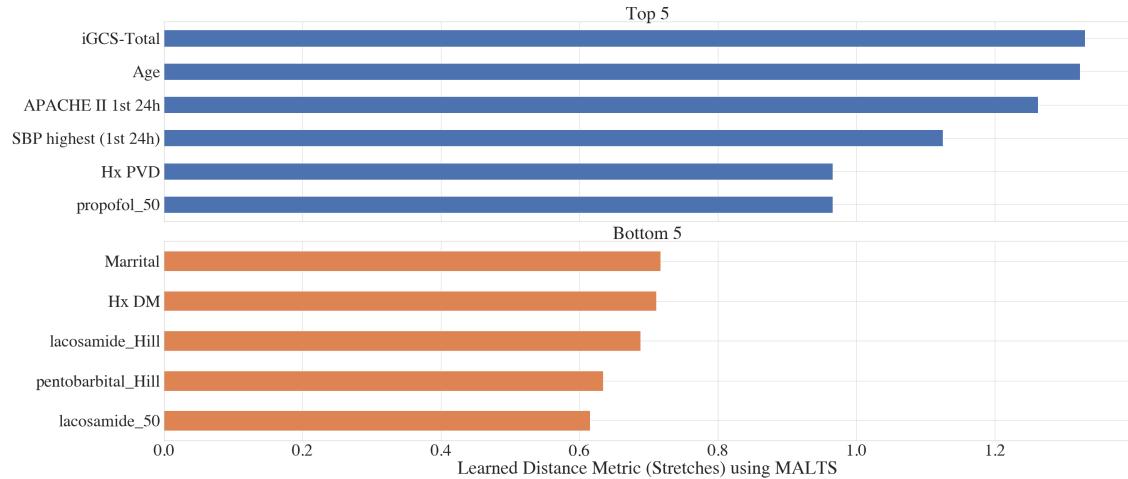


Figure 6: The top and bottom 5 variables, based on the average stretching weights in MALTS, when we are studying the effect of the maximum EA burden E_{\max} .

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 6 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 the level of consciousness and the risk of death for a current patient. 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 consciousness. In Figure 6, 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 may be much more important in understanding the effect of seizure burden than its fellow anti-seizure medications.

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 second and third groups are tight while the first is not as tight.

Age	Gender	APACHE-II	Doctor's Prognosis (chance of bad outcome)	Doctor's Assessment of High EA burden Risk	Notes	Matched Group Analysis
(a) Matched Group 1						
57	Male	15	40% - 60%	40% - 60%	With past medical history of coronary artery disease and tongue cancer, admitted to the hospital because of a cardiac arrest. On arrival to the hospital, he is comatose.	These patients are similar in age. Two of them were previously treated for cancer, and one has active cancer (a brain tumor). Only one of the patients has coma. All members of this matched group are at elevated risk for EA due to acute neurological injuries (anoxic brain injury, ruptured brain aneurysm, brain tumor) or in the patient because of epilepsy. Because coma following cardiac arrest generally carries a worse prognosis than the others' conditions, and because the range of APACHE-II scores (6-15) is broad, this group is not tightly matched.
54	Male	7	20% - 40%	40% - 60%	With past medical history of rectal cancer, admitted to the hospital because of a ruptured brain aneurysm. On arrival to the hospital he is mildly confused.	
57	Female	4	20% - 40%	80% - 100%	With a brain tumor, admitted to the hospital because of a seizure. Brain imaging done on admission to the hospital shows that the brain tumor has grown larger and is causing swelling in the brain tissue surrounding the tumor. On arrival to the hospital she is moderately sleepy and confused.	
59	Male	6	0% - 20%	80% - 100%	With epilepsy, admitted because he began having generalized convulsive ("grand mal") seizures multiple times per day. In between seizures, he is cognitively normal.	
(b) Matched Group 2						
62	Female	3	0% - 20%	80% - 100%	With past medical history of a benign brain tumor (meningioma), complicated by epilepsy, treated with anti-seizure medication. Admitted to the hospital because of recurrent generalized convulsive ("grand mal") 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 patients have similar risk of a poor neurologic outcome.
28	Female	3	0% - 20%	60% - 80%	With multiple previous episodes of severe pneumonia and a large brain blood vessel malformation (arteriovenous malformation) that has caused focal seizures in the past. Admitted to the hospital with pain and bleeding inside the right eye.	
54	Male	7	20% - 40%	40% - 60%	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	0% - 20%	60% - 80%	With past medical history of seizures, admitted to the hospital because of a generalized convulsive seizure. Brain imaging done at the time of hospital admission shows a large frontal brain tumor.	
59	Male	6	0% - 20%	80% - 100%	With epilepsy, admitted because he began having generalized convulsive ("grand mal") seizures multiple times per day. In between seizures, he is cognitively normal.	
(c) Matched Group 3						
48	Male	4	0% - 20%	60% - 80%	1. Presented with headaches 2. RT lesion suspicious for brain tumor	
61	Female	5	0% - 20%	80% - 100%	1. Refractory epilepsy 2. Presented with seizures	
49	Male	4	0% - 20%	60% - 80%	1. Diagnosed with onset blindness 2. Has vertigo and cerebellar ataxia 3. Presented with spells of behavioral arrests	
55	Male	4	0% - 20%	60% - 80%	1. Has hepatitis and right jugular foramen epidermoid 2. Presented with intoxication, disequilibrium and right subdural hematoma	
55	Male	4	20% - 40%	60% - 80%	1. Has hypertension and diabetes; had a recent back surgery 2. Presented with acute aphasia and fever; imaging negative for stroke	
42	Male	3	20% - 40%	60% - 80%	1. Lung cancer metastasis to brain and epilepsy 2. Presented with somnolence and multiple small intracerebral brain hemorrhage	
61	Male	3	20% - 40%	60% - 80%	1. Lung cancer metastasis to liver 2. Presented with confusion and new brain mass	

5.2 Matched Groups are Validated by Neurologist’s Chart Review

One way to check that no important information has been lost in the matching process is to have a domain expert perform a post-facto analysis of the MALTS matched groups. If a domain expert who has access to all the information does not find any reason why all the patients within a matched group should not have a similarly estimated prognosis, then the matching procedure passes this test for quality. Thus, for this experiment, a practicing neurologist (MBW) was randomly given several matched groups and asked to perform a manual chart review of the selected matched patients. Based on these charts, the neurologist made a qualitative expert judgment of whether the patients within a matched group seem like they should broadly have a similar prognosis. The results indicate that the matched group creation process did not miss any obvious important information that a neurologist thinks is crucial for prognostics.

In the table above, we present chart review results for 3 randomly selected match groups. In all three groups, it was deemed that patients seemed roughly appropriate to be matched together and merited a similar prognosis based on information available within the first day of hospital admission. In particular, with a few exceptions, we can see that the patients within the 3 matched groups are tightly matched on Age and APACHE II score, which are the 2nd and 3rd most important variables according to MALTS’s stretch matrix. In particular, we can see that while 3/4 of matched group 1 consists of cancer patients, only 1/5 of matched group 2 consists of cancer patients. This division gives us confidence in the tightness of our matched groups and the quality of our estimates.

6 Discussion

We presented four main points in this work. First, we assembled a large set of continuous EEG recordings and clinical variables, including detailed histories of how these patients were treated with anti-seizure medications. Because doctors currently do not follow a standardized treatment approach for EA burden or seizure-like patterns, these data provide a ‘natural experiment’ in which similar patients receive different treatments, providing an opportunity to analyze the specific effects of EA on neurologic outcome. Sec-

ond, we developed a new analytic approach that combines mechanistic pharmacokinetic/pharmacodynamic modeling with a new matching procedure to disentangle the effects of illness, drug interventions, and the burden of epileptiform activity (EA) on neurologic outcomes. Third, we have provided, for the first time, an estimate of the causal effect of epileptiform activity on neurologic outcomes in patients with critical illness. We find that higher EA burden indeed leads to worse neurologic outcome outcomes at hospital discharge (Figure 2), in a way that depends on the intensity and duration of EA. Finally, our results provide insights into individualized potential outcomes. For example, we show that patients with malignancy or respiratory infections are affected by EA to a higher extent compared to the average level in this cohort (although this cohort already has high risk of poor outcome).

Our findings have two main implications for treatment of EA. First, treatment should be based on EA duration and 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 step-like relationship with outcome: EA < 50% has minimal effect, but EA > 50% causes worse outcome. This suggests that interventions should put higher priority on patients with EA burden higher than 50%; while treatment intensity should be low and conservative when EA intensity is low. 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 suggested, patients with malignancy or respiratory infections should be monitored more closely with a more conservative treatment. By contrast, current treatment protocols used in hospitals are generic, recommending that treatment be tailored based on the intensity or duration of EA (e.g., more aggressive treatment for status epilepticus), but providing no 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.

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 anti-epileptic drugs (AEDs) 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 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, as well as 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 thus far remained unclear whether current treatment approaches are helping or hurting patients.

We address 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 or sedatives, such as phenytoin and pentobarbital, that themselves may worsen outcome. Because prior studies did not disentangle the potential harmful effects of EAs and seizures from anti-seizure drugs, the field remains worried but uncertain. Another key innovation is our application of advanced causal inference for observational data. In the prior studies cited above, multivariate regression is used to adjust for confounders. The nature of observational data and multivariate regression (model misspecification) have made it impossible to establish a causal link between seizures or EAs 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 powerful 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.

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 those periodic discharge patterns, the morphology features (such as seizure with/without triphasic waves), and the spatial extent of the EAs due to lack of high quality human labels. On the other hand, the automatic EA annotator, based on a trained convolutional neural network is imperfect, although being close to the imperfect pairwise agreements among experts for the six normal/EA/seizure patterns ([Ge et al., 2021](#)). To further reduce this uncertainty, we grouped these EA patterns into binary EA (seizure/GPD/LPD/LRDA) vs. non-EA categories (GRDA/normal/other/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 formulation, such as a context sensitive half-life for propofol ([Hughes et al., 1992](#)).

In summary, the results present a data-driven statistical causal inference approach to quantify the harm

of EA in ICU. We not only confirm that EA burdens are indeed harmful for patient outcome, but careful analysis illustrates that there exists important subgroup of patients that are more affected by EA burdens. Based on this, a future direction is to learn an interpretable optimal treatment policy for EA burden patients to improve patient outcomes.

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. AV, CR and HP were supported by NIH R01EB025021. AV was supported by NSF DMS 2046880.

References

- Benbadis, S. R. (2006). Introduction to sleep electroencephalography. Sleep: A comprehensive handbook.
- Blackwell, M. (2014). A selection bias approach to sensitivity analysis for causal effects. Political Analysis, 22(2):169–182.
- Chong DJ, H. L. (2005). Which EEG patterns warrant treatment in the critically ill? reviewing the evidence for treatment of periodic epileptiform discharges and related patterns. J Clin Neurophysiol, 22:79–91.
- 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., Agar-

- wal, 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.
- Diamond, A. and Sekhon, J. S. (2013). Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies. *The Review of Economics and Statistics*, 95(3):932–945.
- Ganesan, S. L. and Hahn, C. D. (2019). Electrographic seizure burden and outcomes following pediatric status epilepticus. *Epilepsy & Behavior*, 101:106409.
- Garoud1, 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*.
- 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. (2021). Deep active learning for interictal ictal injury continuum EEG patterns. *Journal of Neuroscience Methods*, 351:108966.
- 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.

Imbens, G. W. and Rubin, D. B. (2015). Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction. Cambridge University Press.

JA, K., EJ, B., AC, W., AJ, C., KJ, S., SS, Z. S. C., and MB, W. (2018). Epileptiform activity in traumatic brain injury predicts post-traumatic epilepsy. Annals of neurology, 83(4):858–62.

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.

Lalgudi Ganesan S, H. C. (2019). Electrographic seizure burden and outcomes following pediatric status epilepticus. Epilepsy Behav.

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.

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.

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.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.

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.

Stuart, E. A. (2010). Matching Methods for Causal Inference: A Review and a Look Forward. *Statistical Science*, 25(1):1 – 21.

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). zsx139.

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.

Wang, T., Morucci, M., Awan, M. U., Liu, Y., Roy, S., Rudin, C., and Volfovsky, A. (2021). Flame: A fast large-scale almost matching exactly approach to causal inference. Journal of Machine Learning Research, 22 (2021) 1-4, 22:1–4.

Zafar, S. e. a. (2018). Effect of epileptiform abnormality burden on neurologic outcome and antiepileptic drug management after subarachnoid hemorrhage. Clin Neurophysiol, pages 2219–2227.

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)
<hr/>	
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%)

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

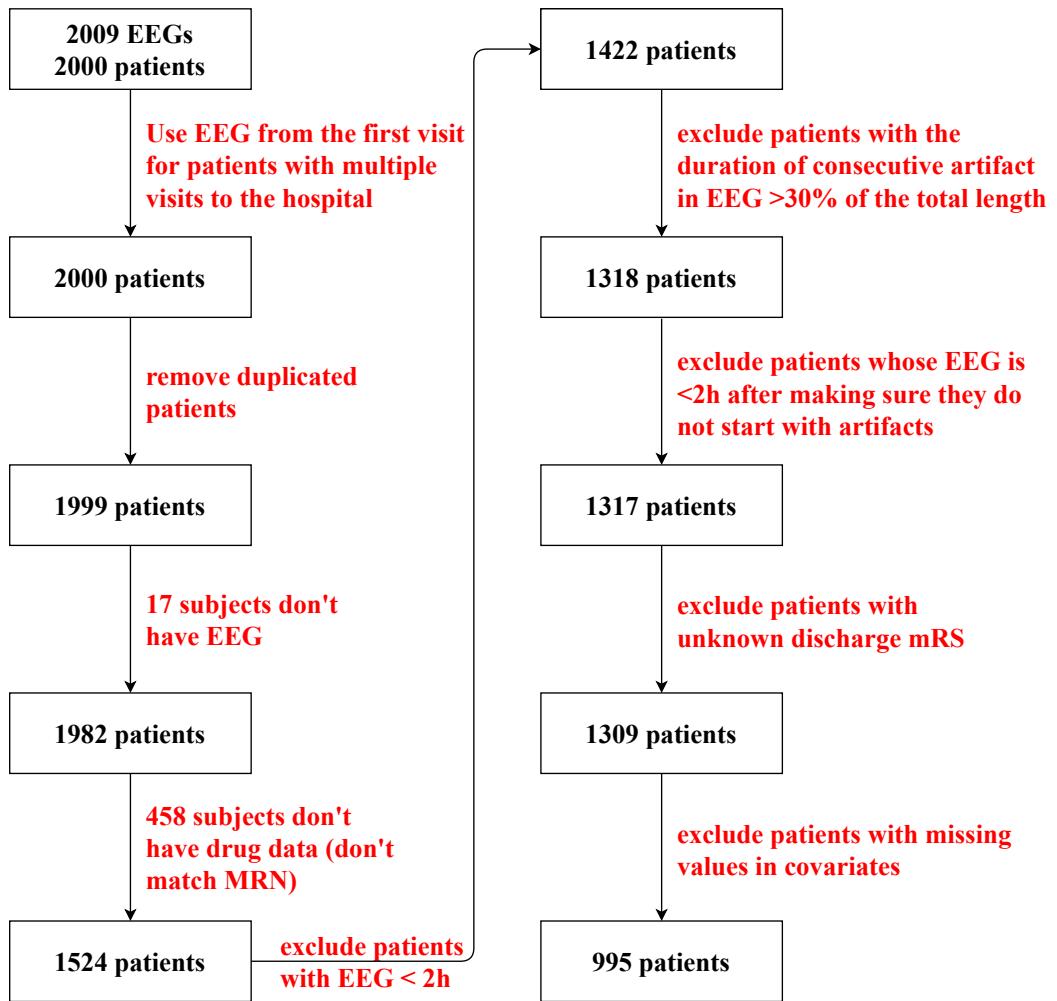


Figure 7: Data flowchart showing the preprocessing of patients

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_{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_{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 ≤ 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 2) 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., 2017); a Hidden Markov model was added to smooth the outputs (Ge et al., 2021). 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 8). 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.

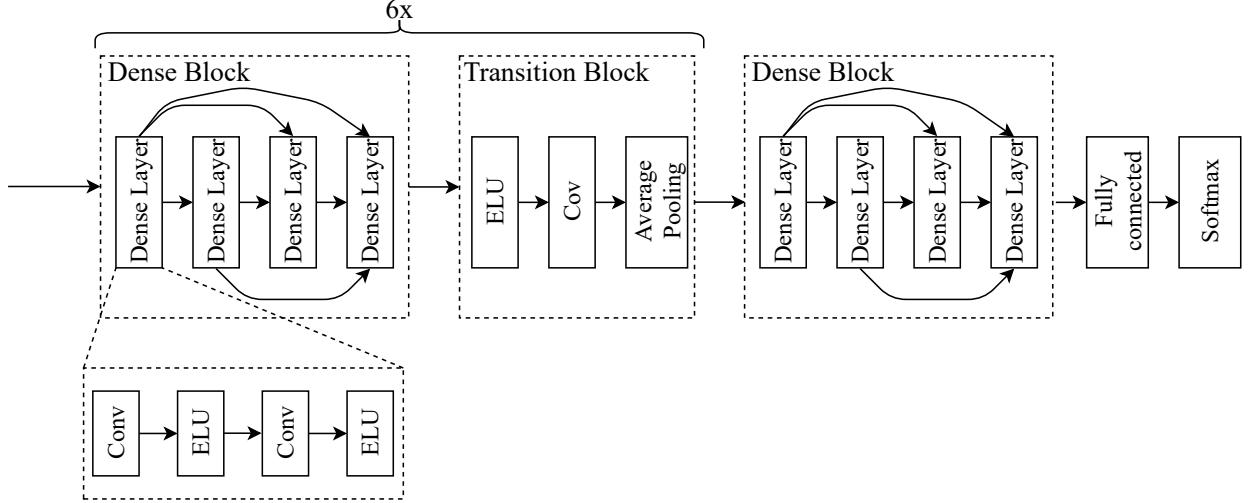


Figure 8: Structure of the DenseNet for automatic EA labeling.

After fitting, it was observed that DenseNet’s classifications were much more volatile than the original

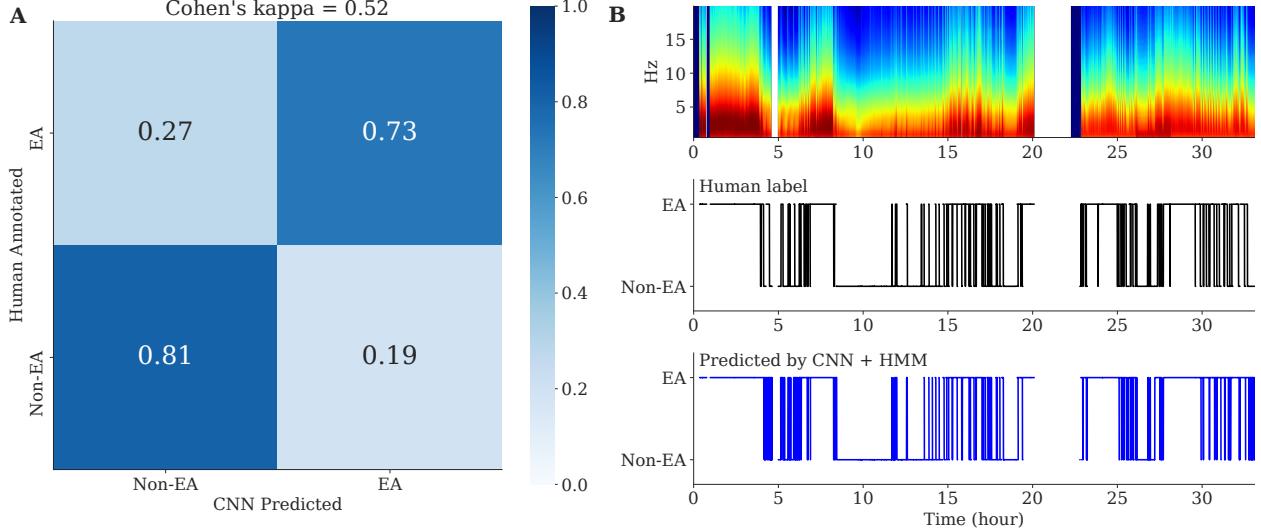


Figure 9: (A) Confusion matrix for the CNN prediction vs. human annotation, where each row represents the fraction of 2-second segments classified into EA (seizure/GPD/LPD/LRDA) or Non-EA (GRDA/other/artifact). The overall Cohen’s kappa is 0.52. (B) The top panel shows the spectrogram of the EEG signal 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.

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% (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 9) 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_{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 10):

- The potential outcome under mild EA burden ($\mathbb{E}[Y([0.0, \rho_1), 0)]$) is mildly sensitive to changes in ρ_2 which is expected. Further, we observe that the gradient of the same with respect to ρ_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 ρ_1 and its gradient with respect to ρ_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 lesser than the point estimates of $\mathbb{E}[Y([\rho_2], 1.0)]$ indicating that the causal effect of EA burden, τ (defined in equation 2) is greater than zero for the possible ranges of ρ_1 and ρ_2 .

Appendix D Missingness Pattern

Based on the flowchart in Figure 7, we compared the mRS at discharge for people with different missing conditions as in Figure 11. We used Mann-Whitney U test (nonparametric t-test) to compare mRS medians, since they do not follow normal distribution.

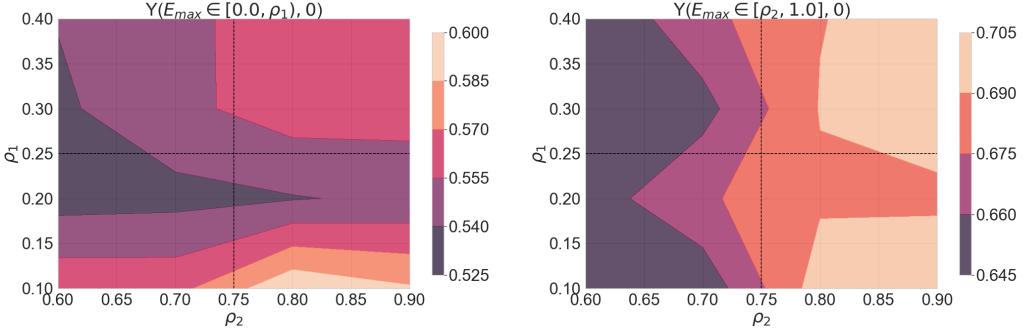


Figure 10: Sensitivity to quantization of EA burden into four levels. ρ_1 is the boundary between mild and moderate EA burden and ρ_2 is the boundary between severe and very severe EA burden. The contour plot shows estimated average potential outcomes - $Y([0, \rho_1], 0)$ and $Y([\rho_2, 1], 0)$ - for a range of ρ_1 and ρ_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.

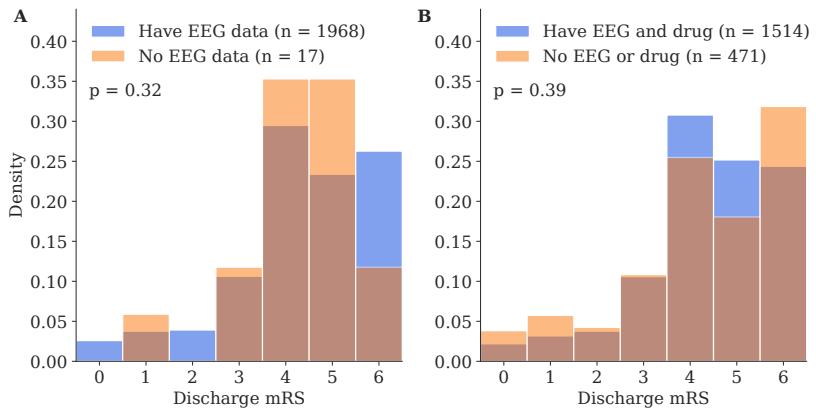


Figure 11: (A) The histogram of patients' mRS at discharge (possible values are 0,1,2,3,4,5,6). The two subsets that are compared are patients who have EEG data ($n = 1968$) vs. patients who do not have EEG data ($n = 17$). To make the subsets comparison, y-axis shows the density instead of the count. The p-value is from 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 patients' mRS at discharge when they have EEG data in the dataset vs. do not have EEG data in the dataset are not significantly different; similarly, the medians when they have both EEG and drug data vs. do not have EEG or drug data are also not significantly different. Therefore the missingness pattern can be considered as not influencing our results.

Appendix E Robustness to causal assumptions

In providing our estimate of average potential outcome, our causal approach makes several important assumptions including: 1) preadmission 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 \overline{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 preadmission covariates and PD parameters

Previously, it was posited that preadmission 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 12 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 preadmission 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.

Average Potential Outcomes under differing assumptions

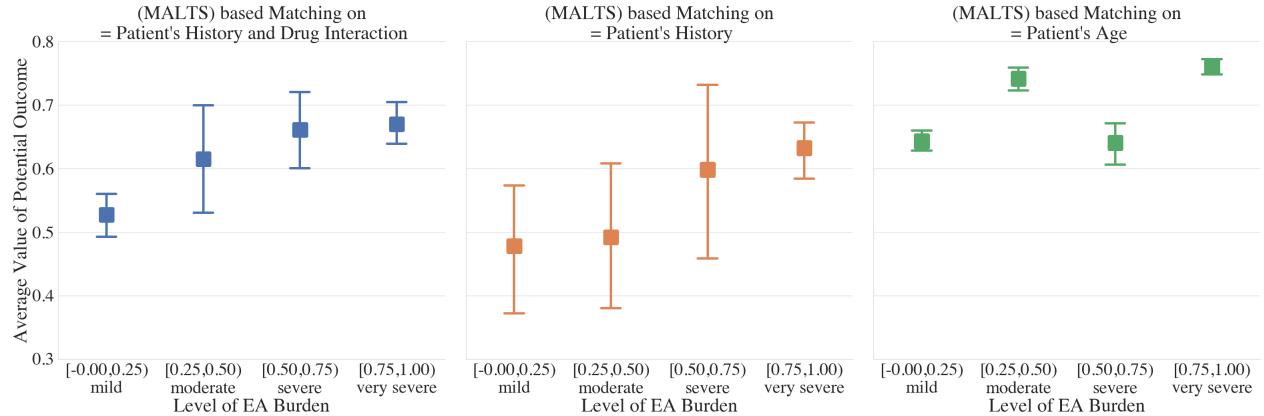


Figure 12: Estimated average potential outcome for different levels of E_{\max} by matching on (left) all pre-admission covariates and PD parameters, (middle) all pre-admission covariates, and (right) only age of the patients.

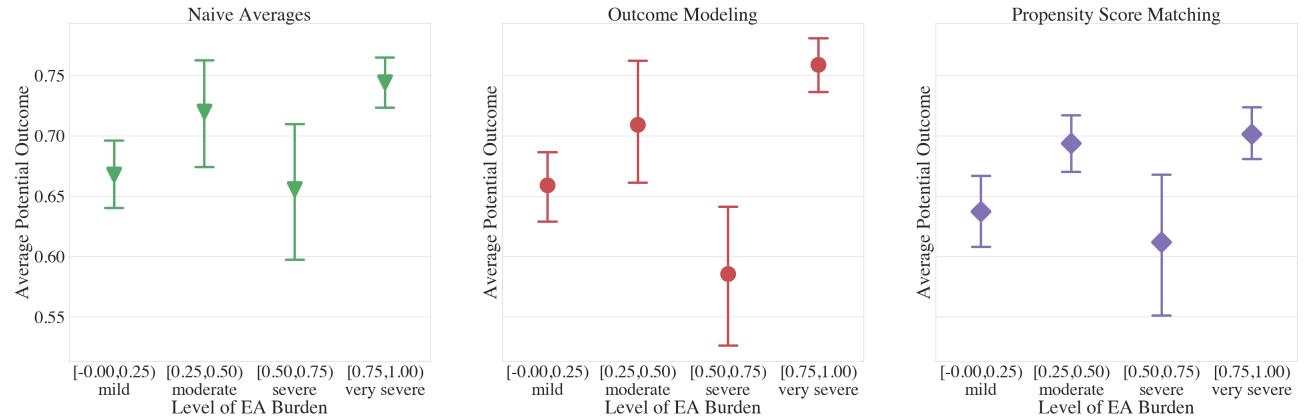


Figure 13: Estimated average potential outcomes computed using Left) Naive Average approach, Middle) Outcome modeling approach, and Right) Propensity Score matching.

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, a "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 report as the left-most figure in Figure 13. 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 13, 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 preadmission 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 13, 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 a 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 12 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 ψ . 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 ψ 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 correction to the observed outcome and re-estimate the average potential outcomes:

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

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 ψ

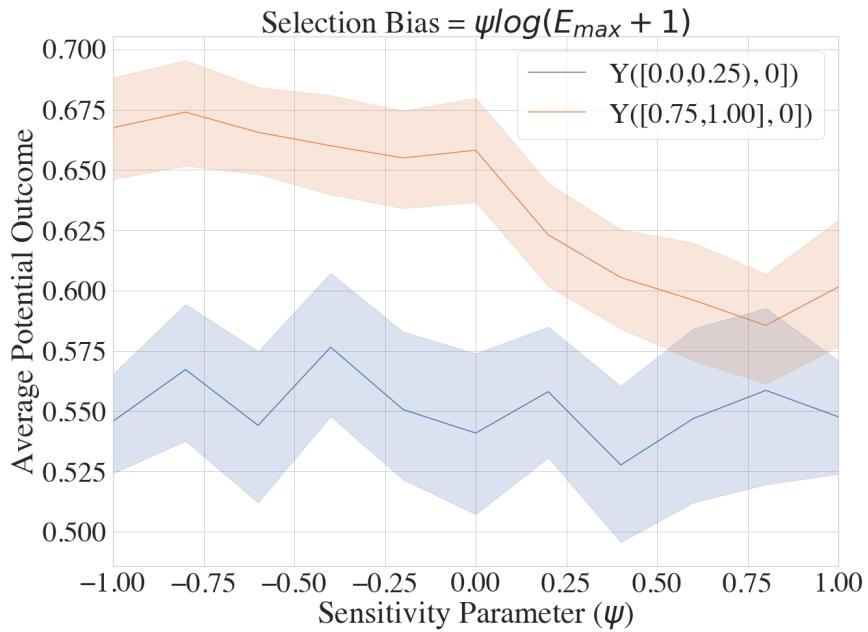


Figure 14: Sensitivity to unobserved confounding The results show that even at very high levels of selection bias, the effect of EA burden is not lost, indicating a degree of robustness in our results.

between $[-1,1]$. We further find that our inference is statistically significant for a wide range of ψ : $-1.0 \leq \psi \leq 0.50$. The sensitivity highlights that the conclusions from our study and analysis are insensitivity to high levels of unobserved confounding.