Aus der Klinik für Psychiatrie und Psychotherapie
Klinik der Ludwig-Maximilians-Universität München
Direktor: Prof. Dr. med. Peter G. Falkai

Using machine learning to predict individual severity estimates of alcohol withdrawal syndrome in patients with alcohol dependence

Dissertation
zum Erwerb des Doktorgrades der Medizin
an der Medizinischen Fakultät der
Ludwig-Maximilians-Universität zu München

Vorgelegt von
Gerrit Burkhardt
aus Freiburg im Breisgau

2021

Mit Genehmigung der Medizinischen Fakultät
der Universität München

Berichterstatter:	Prof. Dr. med. Oliver Pogarell
Mitberichterstatter:	Prof. Dr. Peter Brieger apl. Prof. Dr. Markus Backmund
Mitbetreuung durch die promovierte Mitarbeiterin:	Dr. med. Kristina Adorjan
Dekan:	Prof. Dr. med. dent. Reinhard Hickel
Tag der mündlichen Prüfung:	08.07.2021

Für Klaus Burkhardt

Publikation

Teile dieser Dissertation wurden unter dem Titel *A machine learning approach to risk assessment for alcohol withdrawal syndrome* in der Fachzeitschrift *European Neuropsychopharmacology* veröffentlicht (Burkhardt et al., 2020).

Danksagung

Hiermit möchte ich mich sehr herzlich bei den folgenden Personen bedanken, die mich in den letzten Jahren unterstützt und diese Dissertation somit ermöglicht haben:

Dr. Kristina Adorjan – *für die inzwischen langjährige engagierte und kompetente Betreuung, den zur Verfügung gestellten TU Datensatz sowie für die gelegentlich notwendigen, motivierenden Nachfragen bezüglich des Stands dieser Arbeit.*

Dr. Dominic B. Dwyer – *for the expertly teaching on complex machine learning, scientific writing and research in general as well as extensive, always helpful feedback.*

Prof. Dr. med. Oliver Pogarell – *für die Vermittlung und geduldige Betreuung dieses anspruchsvollen und interessanten Forschungsvorhabens.*

Sina Brückner-Amin – *für ein äußerst hilfreiches Lektorat.*

Antje Burkhardt und Karl-Heinz Raach – *für die bedingungslose Unterstützung meiner beruflichen Ausbildung und das Vertrauen in allen Lebenssituationen.*

CONTENT

Acronyms

ALT	Alanine aminotransferase
AST	Aspartate aminotransferase
AUC	Are-under-the-curve
AUD	Alcohol use disorder
AUDIT(-PC)	Alcohol Use Disorders Identification Test (-Piccinelli Consumption version)
AWS	Alcohol withdrawal syndrome
BAC	Balanced accuracy
BDNF	Brain derived neurotrophic factor
BrAC	Breath alcohol concentration
BlAC	Blood alcohol concentration
CDT	Carbohydrate-deficient transferrin
CIWA-A(r)	Clinical Institute Withdrawal Assessment for Alcohol (revised)
CV	Cross validation
DT	Delirium tremens
EDDP	2-ethylidene-1,5-dimethyl-3,3-diphenylpyrrolidine
EEG	Electroencephalography
FN	False negatives
FP	False positives
GABA	Gamma-aminobutyric acid
GGT	Gamma glutamyltransferase
GLM	General linear model
ICD-10	International Statistical Classification of Diseases and Related Health Problems, 10th edition
ICU	Intensive care unit
LARS (10/11)	Luebeck Alcohol Withdrawal Risk Scale (10/11 item versions)
LMU	Department of Psychiatry and Psychotherapy, University Hospital, Ludwig Maximilian University of Munich
LR	Logistic regression
LSD	Lysergic acid diethylamide
MCV	Mean corpuscular volume
ML	Machine learning

MSAWS	Moderate to severe alcohol withdrawal syndrome
NPV	Negative predictive value
OOCV	Out-of-sample cross validation
PAWSS	Prediction of Alcohol Withdrawal Severity Scale
PPV	Positive predictive value
PSI	Prognostic summary index
RCT	Randomized-controlled trial
SD	Standard deviation
Sens	Sensitivity
SSAGA	Semi-Structured Assessment for the Genetics of Alcoholism
Spec	Specificity
SVM	Support vector machine
TSA	Total Severity Assessment
TN	True negatives
TP	True positives
TU	Department of Clinical Toxicology of the Technical University of Munich
WS	Withdrawal seizures

List of Tables

List of Figures

1 Summary

1.1 Abstract

Despite its high prevalence in diverse clinical settings, treatment of alcohol withdrawal syndrome (AWS) is mainly based on subjective clinical opinion. Without reliable predictors of potential harmful AWS outcomes at the individual patient's level, decisions like provision of pharmacotherapy rely on resource-intensive in-patient monitoring. By contrast, an accurate risk prognosis would enable timely preemptive treatment, open up possibilities for safe out-patient care and lead to a more efficient use of health care resources.

The aim of this project was to develop such tools using clinical and patient-reported information easily attainable at patient's admission. To this end, a machine learning framework incorporating nested cross-validation, ensemble learning, and external validation was developed to retrieve accurate, generalizable prediction models for three meaningful AWS outcomes: (1) Separating mild and more severe AWS as defined by the established AWS scale, and directly identifying patients at risk of (2) delirium tremens as well as (3) withdrawal seizures. Based on 121 sociodemographic, clinical and laboratory-based variables, that were retrieved retrospectively from the patients' charts, this classification paradigm was used to build predictive models in two cohorts of AWS patients at major detoxification wards in Munich (Ludwig-Maximilian-Universität München, n=389; Technische Universität München, n=805).

Moderate to severe AWS cases were predicted with significant balanced accuracy (BAC) in both cohorts (LMU, BAC = 69.4%; TU, BAC = 55.9%). A post-hoc association between the models' poor outcome predictions and higher clomethiazole doses further added to their clinical validity. While delirium tremens cases were accurately identified in the TU cohort (BAC = 75%), the framework yielded no significant model for withdrawal seizures. Variable importance analyses revealed that predictive patterns highly varied between both treatment sites and withdrawal outcomes. Besides several previously described variables (most notably, low platelet count and cerebral brain lesions), several new predictors were identified (history of blood pressure abnormalities, positive urine-based benzodiazepine screening and years of schooling), emphasizing the utility of data-driven, hypothesis-free prediction approaches. Due to limitations of the datasets as well as site-specific patient characteristics, the models

did not generalize across treatment sites, highlighting the need to conduct strict validation procedures before implementing prediction tools in clinical care.

In conclusion, this dissertation provides evidence on the utility of machine learning methods to enable personalized risk predictions for AWS severity. More specifically, nested-cross validation and ensemble learning could be used to ensure generalizable, clinically applicable predictions in future prospective research based on multi-center collaboration.

1.2 Deutsche Zusammenfassung

Die prädiktive Einschätzung der Ausprägung von Entzugssymptomen bei Patient*innen mit Alkoholabhängigkeit beruht trotz jahrzehntelanger wissenschaftlicher Bemühungen weiterhin auf subjektiver klinischer Einschätzung. Entgiftungsbehandlungen werden daher weltweit vorwiegend im stationären Rahmen durchgeführt, um eine engmaschige klinische Überwachung zu gewährleisten. Da über 90 % der Entzugssyndrome mit lediglich milder vegetativer Symptomatik verlaufen, bindet dieses Vorgehen wertvolle Ressourcen. Datenbasierte Prädiktionstools könnten einen wichtigen Beitrag in Richtung einer individualisierten, akkuraten und verlässlichen Verlaufsbeurteilung leisten. Diese würde sichere ambulante Behandlungskonzepte, prophylaktische medikamentöse Behandlungen von Risikopatient*innen, sowie innovative Behandlungsforschung basierend auf stratifizierten Risikogruppen ermöglichen.

Das Ziel dieser Arbeit war die Entwicklung solcher prädiktiven Tools für Patient*innen mit Alkoholentzugssyndrom (AES). Hierfür wurde ein innovatives Machine Learning Paradigma unter Verwendung von strikten Validierungsmethoden (Nested Cross-Validation und Out-of-Sample External Validation) verwendet, um generalisierbare, akkurate Prädiktionsmodelle für drei bedeutsame klinische Endpunkte des AES zu entwickeln: (1) die Klassifikation von milden in Abgrenzung zu moderat bis schwer ausgeprägten AES Verläufen, definiert nach einer hierfür etablierten klinischen Skala (AES Skala), sowie die direkte Identifikation der Komplikationen (2) Delirium tremens (DT) sowie von (3) zerebralen Entzugsanfällen (WS). Dieses Paradigma wurde unter Verwendung von 121 retrospektiv erfassten klinischen, laborbasierten, sowie soziodemographischen Variablen auf 1194 Patient*innen mit Alkoholabhängigkeit an zwei großen Entgiftungsstationen in München angewandt (Ludwig-Maximilian-Universität München, n=389; Technische Universität München, n=805).
Moderate bis schwere AES Verläufe konnten an beiden Behandlungszentren mit einer signifikanten Genauigkeit (balanced accuracy [BAC]) prädiziert werden (LMU, BAC = 69.4%; TU, BAC = 55.9%). In einer post-hoc Analyse war die Prädiktion moderater bis schwerer Verläufe zudem mit höheren kumulativen Clomethiazol-Dosen assoziiert, was für die klinische Validität der Modelle spricht. Während DT in der TU Kohorte mit

einer hohen Genauigkeit (BAC = 75%) identifiziert werden konnte, war die Prädiktion von Entzugsanfällen nicht erfolgreich.

Eine explorative Analyse konnte zeigen, dass die prädiktive Bedeutsamkeit einzelner Variable sowohl zwischen den Behandlungszentren als auch den einzelnen Endpunkten deutlich variierte. Neben mehreren bereits in früheren wissenschaftlichen Arbeiten beschriebenen prädiktiv wertvollen Variablen (insbesondere einer durchschnittlich niedrigeren Thrombozytenzahl im Blut sowie von strukturellen zerebralen Läsionen) konnten hierbei mehrere neue Prädiktoren identifiziert werden (Blutdruckauffälligkeiten in der Vorgeschichte; positives Urinscreening auf Benzodiazepine; Anzahl der Schuljahre). Diese Ergebnisse unterstreichen den Wert von datenbasierten, hypothesen-freien Prädiktionsansätzen. Aufgrund von Limitationen des retrospektiven Datensatzes, wie der fehlenden zentrumsübergreifenden Verfügbarkeit einiger Variablen, sowie klinischen Besonderheiten der beiden Kohorten, ließen sich die Modelle am jeweils anderen Behandlungszentrum nicht validieren. Dieses Ergebnis unterstreicht die Notwendigkeit, die Generalisierbarkeit von Prädiktionsergebnissen adäquat zu testen, bevor hierauf basierende Tools für die klinische Praxis empfohlen werden. Solche Methoden wurden im Rahmen dieser Arbeit erstmalig in einem Forschungsprojekt zum AES verwendet.

Zusammenfassend, zeigen die Ergebnisse dieser Dissertation erstmalig einen Nutzen von Machine Learning Ansätzen zur individualisierten Risikoprädiktion schwerer AES Verläufe an. Das hierbei verwendete cross-validierte Machine Learning Paradigma wäre ein mögliches Analyseverfahren, um in künftigen prospektiven Multi-Center-Studien verlässliche Prädikationsergebnisse mit hohem klinischen Anwendungspotential zu erreichen.

2 Introduction

Alcohol withdrawal syndrome (AWS) constitutes a common manifestation of physical alcohol dependence, encountered in patients across diverse clinical treatment settings who stop or reduce their alcohol consumption (Hall & Zador, 1996). While the majority of patients with AWS will only develop mild vegetative symptoms, that can nonetheless cause significant distress, complications like delirium tremens (DT) and withdrawal seizures (WS) require timely recognition and care (Schuckit, 2014; Victor & Adams, 1953). Despite numerous research efforts to find predictors of such adverse outcomes, there are currently no reliable, objective markers that allow risk stratification at an individual patient's admission (Goodson et al., 2014; National Institute for Health and Care Excellence (NICE), 2010). As a result, clinicians either rely on close symptom assessment to offer post-hoc medication as needed, usually within costly in-patients settings, or they preemptively administer high-dose pharmacological treatment without means to estimate actual demands (Mayo-Smith, 1997). In order to offer patients more efficient treatment alternatives, tailored to their specific needs, tools to accurately predict individual AWS trajectories are urgently needed.

In this introduction, I first discuss which clinical questions arise while treating AWS patients and how they might benefit from accurate disease course prediction. To this end, I outline the development of a unified concept of alcohol withdrawal syndrome and describe how clinicians tend to approach AWS patients diagnostically. Furthermore, I describe state-of-the-art treatment approaches to AWS and discuss their efficiency and safety. Secondly, I summarize the current state of research efforts seeking prediction of AWS severity and explore if and how these have been translated into clinical practice. This entails a discussion on common methodological difficulties and shortfalls of previous research. Thirdly, I discuss Machine Learning (ML) as a possible alternative framework to achieve prediction outcomes, meaningful on a single-patient level. Finally, I describe the aims of this project, namely, using ML approaches to predict individual severity estimates of AWS in patients with alcohol dependence.

2.1 Why AWS severity prediction matters

2.1.1 Current concept of alcohol withdrawal and diagnostic approaches

Although physical and psychological symptoms typically seen in heavy drinkers have been described in Arabic and Western medical literature for centuries (Pearson, 1813; Sutton, 1813), our current understanding of AWS as a spectrum of frequently occurring symptoms following cessation or reduction of alcohol consumption in alcohol-dependent patients was shaped in the middle of the last century (Porcel & Schutta, 2015). Victor and Adam's thorough description of symptom trajectories in 226 male patients admitted to the Boston city hospital for alcohol-related illness in 1953 revealed four "clinical states"—tremulous, hallucinatory, epileptic and delirious—which the authors linked "not only upon the effects of prolonged exposure to alcohol, but temporarily, on abstinence from the drug" (Victor & Adams, 1953).This etiological attribution was later affirmed by an interventional study that induced withdrawal symptoms in opioid-dependent patients by applying controlled amounts of alcohol (Isbell et al., 1955). Due to the development of effective sedative drugs in the late 1950s and 1960s (Kaim et al., 1969; Thomas & Freedman, 1964), only few studies on medication-naïve patients have been conducted since (Whitfield et al., 1978), mainly due to ethical consideration (Hall & Zador, 1996). Therefore, most recent reviews and clinical recommendations (Hall & Zador, 1996; Mann et al., 2016; National Institute for Health and Care Excellence (NICE), 2011; Schuckit, 2014) describe a "natural history" (Foy et al., 1997) of alcohol withdrawal that closely follows the clinical states outlined in Victor and Adam's work conceptually:

Corresponding to the "tremulous state", most patients develop somatic (tremors, sweating, heart rate and blood pressure increases, increases in body temperature and breathing rate, nausea and vomiting) and mental symptoms (anxiety, agitation, sleep disturbances), that are partly explained by autonomic hyperactivity revealing readjustment to the prolonged effects of alcohol intoxication on the brain (Hall & Zador, 1996; Littleton, 1998). Typically, these symptoms occur six to eight hours after the cessation or reduction of repeated, usually high-dose alcohol consumption, peak in severity during day one to three, and subside till day five to seven after the last alcohol intake (Mann et al., 2016; National Institute for Health and Care Excellence (NICE), 2010; Schuckit, 2014; Victor & Adams, 1953). Though potentially distressing manifestations, that warrant pharmacological treatment, do frequently occur in hospital

settings (Eyer, Schuster, et al., 2011; Foy & Kay, 1995), AWS cases with solely autonomic symptoms are mostly classified as mild to moderate withdrawal (Maldonado et al., 2014; Mayo-Smith et al., 2004; Schuckit, 2014). In contrast, the “delirious state” describes a potentially life-threatening manifestation of alcohol withdrawal: Autonomic symptoms of greater intensity that cooccur with fluctuating neurocognitive deficits like decreased attention and disorientation, delusions, as well as perceptional disturbances, typically starting between 48 to 72 hours after alcohol cessation, are usually subsumed under the 19th-century-term “delirium tremens” (Hall & Zador, 1996; Schuckit, 2014; Sutton, 1813). DT rates in AWS inpatients have been reported in the range of 3 to 20% (Eyer, Schuster, et al., 2011; Ferguson et al., 1996; Salottolo et al., 2017; Soravia et al., 2018). The generally lower incidence in more recent studies (Eyer, Schuster, et al., 2011; Soravia et al., 2018) has been primarily attributed to more rigorous pharmacological treatment (Schuckit, 2014). DT can result in increased mortality due to hyperthermia, cardiac arrhythmias, worsening of medical diseases or complications of physical restraint (Khan et al., 2008; Salottolo et al., 2017). Furthermore, DT has been associated with more frequent admissions to intensive care units (ICU), longer ICU treatment duration and longer treatment duration in general (Salottolo et al., 2017; Wright et al., 2006). Therefore, it is considered the most severe or complicated manifestation of AWS (Mayo-Smith et al., 2004; Schuckit, 2014), which frequently requires in-patient treatment and pharmacological intervention (Schmidt et al., 2016). Lastly, corresponding to the “epileptic state”, up to 10% of AWS patients may experience single or recurrent generalized tonic-clonic seizures that usually occur within 48 hours after alcohol cessation (Hillbom et al., 2003; Hughes, 2009; Victor & Brausch, 1967). These withdrawal seizures are most likely caused by central nervous system hyperexcitability due to adaptive changes in neurotransmitter homeostasis that subside during abstinence, often occur before patient admission and show low mortality rates (Hughes, 2009). However, more hazardous seizure etiologies that are frequently associated with alcohol-use have to be excluded to prevent complications like status epilepticus and brain damage (Hillbom et al., 2003). Retrospectively, cases which are first considered withdrawal-related can be attributed to other etiologies (e.g. head injury, idiopathic, cerebrovascular lesions, toxicity) in over 50% of cases (Rathlev et al., 2000). An adequate diagnostic workup via laboratory and imaging-based evaluation is therefore necessary, especially in cases of first-time seizure (Hillbom et al., 2003).

Considering these potential adverse outcomes, efficient diagnostic tools that differentiate between uncomplicated and complicated trajectories of AWS are mandatory to ensure adequate patient care. Following a categorical approach, the International Classification of Diseases and Related Health Problems, 10th Edition (World Health Organization, 2004), currently used as a classification system to operationalize diagnoses in Germany, allows for separate coding of withdrawal states depending on the occurrence of DT and WS (Table 2.1). The validity and clinical utility of such broad disease categories has been questioned repeatedly across psychiatry, since potentially complex pathophysiological underpinnings and individual disease trajectories are disregarded (Jablensky, 2016). Alcohol withdrawal researchers have therefore recommended to incorporate standardized clinical assessment scales into operationalized classifications of AWS (Sellers et al., 1991). Such assessment scales have been developed and used in clinical practice to guide treatment decisions based on AWS severity: The Total Severity Assessment (TSA) comprised 32 items correlated with clinical judgement and withdrawal severity in an observational cluster analysis of 100 male alcohol depended patients (Gross et al., 1973). It was later refined to the 15-item Clinical Institute Withdrawal Assessment for Alcohol (CIWA-A) Scale (Shaw et al., 1981), by excluding variables that could not be used for half-hourly scoring (e.g. sleep disturbances). The scale was applied in non-pharmacological (Naranjo et al., 1983) and pharmacological (Sellers et al., 1983) withdrawal treatment studies. Sullivan et. al validated a revised 10-item version, the CIWA-Ar (Sullivan et al., 1989) that is still widely used to assess withdrawal severity in clinical and research contexts (Sachdeva et al., 2014; Schuckit, 2014). Another 10-item derivative of the CIWA-A, the Alcohol Withdrawal Scale (AWS scale) (Wetterling et al., 1997), that offers sub-scales for somatic and mental symptoms, is more frequently used for severity assessment in German-speaking countries (Eyer, Schreckenberg, et al., 2011; Eyer, Schuster, et al., 2011; Soravia et al., 2018). Several other scales have been reported in single studies, but not applied more widely (Mennecier et al., 2008; Williams, 2001). While most standardized scales have been tested for interrater-reliability (Williams, 2001) and certainly allow a more accurate assessment of individual withdrawal trajectories than broad ICD-10 diagnosis, they also require sufficient staff that is trained in its application (Sullivan et al., 1989). Moreover, some authors suggest that they do not sufficiently capture negative affective symptoms that often occur during AWS and may last up to several weeks (Heilig et al., 2010). Nonetheless, since no accurate biomarkers are

available in the field, assessment scales are endorsed as the main diagnostic tools to guide treatment decision during the withdrawal course (Mann et al., 2016; National Institute for Health and Care Excellence (NICE), 2010), as will be described in the following section.

Table 2.1. ICD-10 criteria for Alcohol withdrawal states, based on: (World Health Organization, 2004)

Alcohol withdrawal state (F10.3)
Recent termination or reduction of alcohol consumption after repeated, prolonged use of high doses. At least three of the following symptoms: - tremor of the tongue, eyelids, or the outstretched hands - sweating - nausea, retching, or vomiting - tachycardia or hypertension - psychomotor agitation - headache - insomnia - malaise or weakness - transient visual, tactile, or auditory hallucinations or illusions - generalized tonic-clonic seizures Symptoms not explained by medical disorder unrelated to alcohol use or other mental disorder. **Further specified: uncomplicated (F13.30) / with convulsions (F13.31)**
Alcohol withdrawal delirium (F10.4)
Alcohol withdrawal state (F10.3) and criteria for delirium (F05.-) are fulfilled: - clouding of consciousness with attentional deficits - disturbance of cognition, manifest by memory deficits and disorientation - psychomotoric disturbances - disturbance of sleep or the sleep-wake cycle - rapid onset and fluctuations of these symptoms over the course of the day **Further specified: without convulsions (F10.40) / with convulsions (F10.41)**

2.1.2 Challenges of withdrawal treatment

AWS patients may present in various treatment settings like psychiatric wards (Wetterling et al., 1994), general medical departments (Jaeger et al., 2001), specialized detoxification units (Eyer, Schreckenberg, et al., 2011), surgical wards (Maldonado et al., 2015), intensive care units (Lukan et al., 2002), outpatient-clinics (Whitfield et al., 1978) or the correctional health care system (Fiscella et al., 2004). They may contact health-services for planned withdrawal treatment in context of a known AUD (Soravia et al., 2018) or be forced into unplanned withdrawal after experiencing trauma (Holt et al., 1980) or severe somatic illness (Wojnar, Bizoń, et al., 1999). Such treatment settings are characterized by varying access to medical resources, different staff training and availability, diverging treatment approaches and—crucially—heterogenous patient collectives (National Institute for Health and Care Excellence (NICE), 2010, 2011). The development of effective and comparatively safe sedative drugs, most notably benzodiazepines, has replaced treatments like lumbal punctures, hydrotherapy, insulin coma therapy and paraldehyde sedation since the 1960s (T. A. Stern et al., 2010). But while numerous randomized-controlled trials (RCTs) on a variety of drugs for different outcomes of AWS have been published, low methodological quality pervades as a main impediment to evidence-based treatment recommendations (Amato et al., 2011; Moskowitz et al., 1983). Following a survey in the United States that revealed heterogenous treatment practices amongst clinicians (Saitz et al., 1995), a first evidence-based guideline was published in 1997 that defines still valid treatment principles (Mayo-Smith, 1997) that have been incorporated into more recent guidelines (Mann et al., 2016; National Institute for Health and Care Excellence (NICE), 2011). Goals of AWS treatment are minimization of withdrawal symptoms, promotion of patient dignity and comfort, prevention of complications like WS and DT, prevention of relapse, as well as the transition to further AUD treatment (Mann et al., 2016; Mayo-Smith, 1997; National Institute for Health and Care Excellence (NICE), 2010, 2011). To achieve these goals, clinicians have to answer the following questions:

a) *What treatment setting is best suited for the individual patient?*

The only randomized controlled study comparing an inpatient with an outpatient treatment setting for AWS patients showed, that while more patients completed the

costly and lengthy inpatient program, both treatment safety and follow-up outcomes after six months did not differ between groups (Hayashida et al., 1989). Further uncontrolled studies suggest that both home-based (Bartu & Saunders, 1994; Stockwell et al., 1991) and specialized out-patient (Soyka & Horak, 2004) settings are feasible and safe alternatives for patients with mild withdrawal severity. From a public-health perspective such treatment settings certainly would be desirable to reduce high in-patient treatment costs (Hayashida et al., 1989; National Institute for Health and Care Excellence (NICE), 2011; Wright et al., 2006). They might even show beneficial effects on long-term treatment outcomes (Soyka & Horak, 2004). Unfortunately, there are no studies available that offer clinicians objective eligibility criteria for different treatment settings (National Institute for Health and Care Excellence (NICE), 2011). While in cases of acute somatic illness or physical trauma in-patient treatment is usually warranted, in other cases clinicians have to decide if a patient can be treated in an out-patient program, can be referred to a planned detoxification program at a later point in time or should be admitted to inpatient treatment right away. Current expert opinions state that, with high-quality evidence missing, these decisions should be individually based on criteria like patient age, physical and cognitive abilities, and available social support (National Institute for Health and Care Excellence (NICE), 2011).

b) What non-pharmacological treatment should be provided?

As mentioned above, most AWS patients will develop only mild autonomous withdrawal symptoms (Victor & Adams, 1953), that can be potentially managed if adequate staff attendance, monitoring, hydration and a well-lit environment is ensured (Naranjo et al., 1983; Whitfield et al., 1978). Since only one controlled study on non-pharmacological treatment is available (Naranjo et al., 1983), it seems unclear if this approach translates to modern patients' expectations, if one considers important factors like patient dignity and comfort.

c) *Does an individual patient require pharmacological treatment and if, which dosing scheme should be used?*

Three dosing approaches are usually proposed for pharmacological withdrawal treatment: In fixed-dose regimes a standard dose of a sedative medication is administered that is then tapered off during subsequent days (Sachdeva et al., 2014). Front-loading refers to a regime where a long-acting sedative medication is administered in high dose at the beginning of withdrawal, followed by rescue medication if needed (Maldonado et al., 2012). In symptom-triggered therapy withdrawal severity is assessed via standardized scales such as the CIWA-Ar or AWS scale by attendant nurses or doctors in regular intervals. If withdrawal severity exceeds predefined cut-off ratings, sedative medication is provided to the patient ad-hoc (Holleck et al., 2019). Several studies compared the efficiency between fixed dose and loading dose (Day et al., 2004; Jauhar, 2000; Manikant et al., 1993; Wasilewski et al., 1996), symptom-triggered therapy and loading dose (Maldonado et al., 2012), symptom-triggered therapy and fixed dose (Daeppen et al., 2002; Elholm et al., 2011; Lange-Asschenfeldt et al., 2003; Sachdeva et al., 2014; Saitz et al., 1994; Weaver et al., 2006), as well as symptom-triggered therapy and a variably defined “treatment-as-usual” (Jaeger et al., 2001; J P Reoux & Miller, 2000; Soravia et al., 2018), mainly with benzodiazepines or clomethiazole as sedative agents. Symptom-triggered therapy is recommended by most treatment guidelines, due to beneficial effects on treatment duration and total benzodiazepine doses administered (Mayo-Smith, 1997; National Institute for Health and Care Excellence (NICE), 2010). While a recent meta-analysis confirmed these advantages, it could not show a decrease in mortality or occurrence of DT or WS and observed, that most results were achieved in low-risk patient cohorts, therefore might not transfer to many AWS treatment settings (Holleck et al., 2019). Indeed, a recent study in ICU patients suggested that front-loading regimes with focus on the first 24 hours after admission might be beneficial in preventing adverse outcomes of severe AWS (J. A. Lee et al., 2019). Moreover, symptom-triggered therapy requires frequent monitoring by specialized staff that is trained in standardized assessment, which may limit its transferability to treatment settings without respective resources. To this date, no studies have been conducted that compare different treatment regimens based on patient characteristics or predictive ratings.

d) If pharmacological treatment is required, which medication should be administered?

Benzodiazepines are the most widely prescribed pharmacological agents for AWS (Mayo-Smith, 1997). Besides having been tested in multiple RCTs (Amato et al., 2010; Mayo-Smith, 1997), they offer a treatment rationale based on neurobiological underpinnings: Frequent alcohol consumption is hypothesized to cause adaptive changes in neurotransmitter homeostasis that lead to a decrease in activity of the inhibitory neurotransmitter gamma-aminobutyric acid (GABA) (Littleton, 1998; Petty et al., 1993) and increased activity of the excitatory neurotransmitter glutamate (Hermann et al., 2012; Tsai et al., 1995) causing withdrawal symptoms. Due to their modulating effects on the GABA-receptor, benzodiazepines are hypothesized to counteract these acute imbalances and therefore ameliorate withdrawal symptoms (Amato et al., 2010). Nonetheless, compared to other possible AWS medications and placebo, benzodiazepines did solely perform superior in prevention of withdrawal seizures in a more recent meta-analysis, showing no benefits on withdrawal severity, DT development or safety (Amato et al., 2010, 2011). Emphasizing the methodological heterogeneity of underlying studies, the authors concluded that more studies are needed to establish the efficacy and safety of benzodiazepine treatment (Amato et al., 2010). Others warned, that AWS prophylaxis with benzodiazepines in patients who may only develop mild withdrawal symptoms could cause unnecessary side effects like sedation, falls, and paradoxical delirium (Maldonado et al., 2014). Several other sedative agents like anti-convulsant medication (Eyer, Schreckenberg, et al., 2011; Minozzi et al., 2010), clomethiazole (Eyer, Schreckenberg, et al., 2011; Eyer, Schuster, et al., 2011) and antipsychotics (National Institute for Health and Care Excellence (NICE), 2010) are currently used in clinical care either as main or adjacent treatment. In ICU contexts, propofol or dexmedetomidine have been suggested as treatments for patients unresponsive to benzodiazepines, although results were unsatisfactory (Vanderweide et al., 2016; Wong et al., 2015). Other agents like Baclofen (Cooney et al., 2019; Liu & Wang, 2019), gamma-hydroxybutyrate (Leone et al., 2010) and nitrous oxid (Gillman et al., 2007) have been studied, but are currently not recommended by clinical guidelines due to insufficient evidence of their efficiency (Mann et al., 2016; National Institute for Health and Care Excellence (NICE), 2010). Considering the increasing range of possible pharmacological agents for treatment of AWS, an

objective selection based on patients' characteristics and individually expected disease courses is urgently needed.

As has been outlined above, AWS treatment currently relies on several decisions that require an estimate on how severe an individual patient's disease course will likely unfold. Key decisions—choosing an adequate treatment setting, initiating prophylactic pharmacological treatment to prevent DT or WS—have to be made shortly after initial contact. Without precise tools to predict further withdrawal development, clinicians are dependent on subjective judgement to weigh an overall low prior risk of severe AWS development against the potentially hazardous consequences of untreated DT and WS. This entails several ethical considerations: A cautious approach with low-threshold admissions to inpatient care might ensure patient safety, but disproportionately binds resources not required for most cases of mild AWS. Moreover, this focus on potential withdrawal development might unnecessarily delay referral to further addiction treatment. Even if a patient is admitted to inpatient treatment, DT often develops before pharmacological treatment is initiated (Foy et al., 1997). Without means to identify cases at risk of DT, clinicians either rely on frequent monitoring via assessment scales or have to treat all patients preemptively, even when most patients do not require pharmacotherapy. To summarize, adequate tools that predict further withdrawal trajectories would certainly benefit treatment decision making (Eyer, Schuster, et al., 2011; Maldonado et al., 2014; National Institute for Health and Care Excellence (NICE), 2011). In the next chapter I will outline previous research efforts to develop means to assess patients for risk of adverse AWS outcomes.

2.2 Current state of AWS prediction research

The need to predict AWS severity to optimize treatment decisions has motivated research efforts for decades (Shaw et al., 1981). Since then, numerous, mostly retrospective studies on AWS prediction have been published in regular intervals (e.g.: Benson et al., 2019; Eyer et al., 2011; Kraemer et al., 2003; Mennecier et al., 2008; Ramos, 2013; Wetterling et al., 1994; Wright et al., 2006). A 2014 meta-analysis identified 226 epidemiological studies on AWS of which 43 studies reported risk factors for severe withdrawal (Goodson et al., 2014). With few exceptions (Khan et al., 2008), most studies have focused on three outcomes of AWS: severity during the withdrawal

course as measured by a standardized assessment scale (Kraemer et al., 2003; Mennecier et al., 2008), the occurrence of DT (Berggren et al., 2009; Ferguson et al., 1996; Monte et al., 2009; Palmstierna, 2001) or the occurrence of WS (Hillemacher et al., 2012; Morton et al., 1994). Few studies have considered multiple outcomes like DT and WS (Eyer, Schuster, et al., 2011). In the following sections I give an overview on frequently reported independent risk factors for all three outcomes. Subsequently, I consider studies that introduce predictive tools suggested for clinical use. Finally, I investigate how and if these research efforts have been translated into clinical practice and which difficulties need to be addressed by further research efforts.

2.2.1 Single risk factors

a) Previous AWS history

Based on the much cited kindling hypothesis, which suggests that AWS severity increases as a function of the number of previous detoxifications due to long term adaptions in central nervous system (CNS) excitability (Ballenger & Post, 1978; Becker, 1998; Booth & Blow, 1993; Gonzalez et al., 2001; Lechtenberg & Worner, 1991), several studies have examined if previous withdrawal experiences influence the likelihood of severe AWS in the index episode. Considering severity as defined by a standardized assessment, Malcolm et al. reported a slower decline in CIWA-Ar score ratings in patients with multiple previous detoxifications (Malcolm et al., 2000). Kraemer et al. observed an increased risk of severe AWS as defined by the CIWA-Ar score for patients with a self-reported history of DT and/or two or more alcohol treatments (Kraemer et al., 2003). In contrast, AWS severity was not associated with prior detoxifications or self-reported history of DT when assessed by the AWS scale (Wetterling, 2001) or Cushman score (Mennecier et al., 2008). Likewise, a greater number of detoxification (Booth & Blow, 1993; Lechtenberg & Worner, 1991, 1992) and a history of previous seizures (Berggren et al., 2009; Morton et al., 1994) have been suggested as risk factors for incident withdrawal seizures, while other studies could not find such associations (Eyer, Schuster, et al., 2011; Rathlev et al., 2000; Wojnar, Bizon, et al., 1999). For delirium tremens, both a history of DT (Berggren et al., 2009; Fiellin et al., 1998; J. H. Lee et al., 2005; Palmstierna, 2001; Wright et al., 2006) as well as a history of WS (Fiellin et al., 2002; Palmstierna, 2001) were correlated with higher risk of incident delirium tremens. Again, several studies could

not find an increased risk for prior withdrawal events for the same outcome (Eyer, Schuster, et al., 2011; Ferguson et al., 1996). The only meta-analysis in the field—excluding several studies due to lack of adequate control groups, insufficient statistical reporting or lack of standardized AWS definitions (Findley et al., 2010; Hillemacher et al., 2012; Lukan et al., 2002; Palmstierna, 2001)—reported that both a history of DT or WS increased the risk of a respective event in the incident episode (Goodson et al., 2014).

b) Degree of intoxication at admission

Blood or breath alcohol concentrations are routinely collected, easily attainable measurements in most treatment facilities. Vinson et al. reported a linear increased risk of AWS severity, as measured by clinical assessment and total amount of withdrawal medication in the first 48 days, for patients in both psychiatric and general medicine departments depending on higher breath alcohol concentrations (BrAC) at admission (Vinson & Menezes, 1991). Palmstierna et al. reported an increased risk of incident DT for blood alcohol concentrations (BlAC) greater than one gram per liter body fluid at admission, but only if autonomous withdrawal symptoms were present (Palmstierna, 2001). In trauma patients a BlAC of more than 200 mg/dL was suggested as a risk factor of DT (Lukan et al., 2002). Negative findings, that did not yield an association between alcohol intoxication at admission and withdrawal outcome, were reported for severe AWS assessed by the CIWA-Ar scale (Kraemer et al., 2003), for DT (Eyer, Schuster, et al., 2011; Fiellin et al., 1998) and WS (Eyer, Schuster, et al., 2011). Furthermore, Rathlev et al. found lower rates of withdrawal seizures in patients with a blood alcohol level greater then 100 mg per deciliter (Rathlev et al., 2000). Combining the few available studies in the mentioned meta-analysis did not yield significant results (Goodson et al., 2014).

c) Laboratory assessment at admission

Several laboratory parameters have been reported as risk factors of withdrawal outcomes: A retrospective study by Berggren et al. first observed thrombocytopenia (defined as a platelet count below 150 × 109 cells per liter) as more commonly occurring in DT and WS patients (Berggren et al., 2009). Summarizing several studies that reported platelet-count at admission (Berggren et al., 2009; Eyer, Schuster, et al.,

2011; Huang et al., 2011; Monte et al., 2009), a meta-analysis found lower platelet-values in DT and WS patients (Goodson et al., 2014). Electrolyte abnormalities, especially hypokalemia (Wadstein & Skude, 1978), have been repeatedly proposed as predictors of DT development (Berggren et al., 2009; Eyer, Schuster, et al., 2011; Wetterling et al., 1994). In meta-analysis lower potassium levels were found as predictive of both DT and WS, while sodium and chloride were either non-significant or only available in single studies (Goodson et al., 2014). Increases in laboratory markers of risky alcohol consumption, like the liver enzymes alanine aminotransferase (ALT), aspartate aminotransferase (AST), and gamma glutamyltransferase (GGT) as well as mean corpuscular volume (MCV) and carbohydrate-deficient transferrin (CDT) have been reported as risk factors of severe AWS development defined by assessment scales (Mennecier et al., 2008; Wetterling et al., 1994), incident withdrawal seizures (Bråthen et al., 2000) and DT (Berggren et al., 2009). The aforementioned meta-analysis found higher ALT to be predictive of general severe AWS and higher GGT of withdrawal seizures (Goodson et al., 2014). Several small exploratory studies reported further laboratory values, that are usually not incorporated in standard laboratory panels, like elevated homocysteine levels as predictors of WS (Bleich et al., 2006; Hillemacher et al., 2012) and differentiating levels of serum brain derived neurotrophic factor (BDNF) between DT patients, AWS patients without DT and healthy controls (Huang et al., 2011).

d) Vital parameters at admission

Similar to laboratory testing, vital parameters are routinely taken at a patient's admission and then used to monitor autonomous symptoms during the withdrawal course (Sullivan et al., 1989; Wetterling et al., 1997). Palmstierna et al. and Lee et al. both reported heart rate greater then respectively 120 and 100 beats per minute as predictive of DT (J. H. Lee et al., 2005; Palmstierna, 2001). While Monte et al. and Fiellin et al. reported systolic blood pressures greater than 150 mmHg and 145 mmHg as risk factors of DT (Fiellin et al., 2002; Monte et al., 2009), Berggren et al. and Ferguson et al. found relatively lower systolic blood pressure values in DT patients (Berggren et al., 2009; Ferguson et al., 1996). Furthermore, Monte et. al reported a body temperature greater than 38 degree Celsius (Monte et al., 2009) and Ferguson an average increased respiratory rate as predictive of DT (Ferguson et al., 1996).

Collectively, neither systolic or diastolic blood pressure or heart rate were predictive of severe withdrawal in meta-analysis (Goodson et al., 2014).

e) Sociodemographic characteristics and alcohol disorder history

Several sociodemographic risk factors have been discussed in previous research: While Lukan et al., Blondell et al., and Salottolo et al. all reported patient age of greater then respectively 40, 45 or 55 years as risk factors of DT development in trauma patients (Blondell et al., 2004; Lukan et al., 2002; Salottolo et al., 2017), no correlations between age and maximum AWS scale ratings (Wetterling, 2001), maximum CIWA-Ar ratings (Kraemer et al., 2003) or withdrawal seizures (Eyer, Schuster, et al., 2011) could be found. Meta-analysis did not find significant results for DT, WS or assessment scale ratings (Goodson et al., 2014). No predictive gender effects have been reported in the overall male weighted study populations and male gender did not increase risk of severe AWS in a meta-analysis (Goodson et al., 2014). Several studies explicitly reported that the duration of life-time alcohol consumption and the daily amount of alcohol consumption did not increase the severity of AWS (Kraemer et al., 2003; Lechtenberg & Worner, 1991; Wetterling, 2001). Correspondingly, the aforementioned meta-analysis did not find significant effects of the duration of alcohol abuse, the age of onset of alcohol abuse or the daily intake of alcohol on DT, WS or assessment scale ratings (Goodson et al., 2014).

f) Comorbidities

AWS often occurs in patients that are hospitalized for medical diseases other than AUD or for alcohol-related diseases (Wojnar, Bizoń, et al., 1999). Also, prevalence of medical comorbidities is elevated in patients with psychiatric disorders compared to the general population (Walker et al., 2015). Several studies have examined the influence of somatic disease on DT development: Ferguson et al. showed that DT development was more likely in internal-medicine patients with acute somatic disease, including pneumonia, alcoholic pancreatitis, alcoholic hepatitis, upper gastrointestinal bleeding, sepsis, pyelonephritis, dehydration, and renal failure (Ferguson et al., 1996). Lee et al. reported higher rates of acute somatic illness (not further specified) for internal-medicine DT patients, that did however yield no significant predictive results in multivariate analysis (J. H. Lee et al., 2005). Wojnar et al. suggested a longer duration and greater severity of DT in patients with pneumonia, coronary heart disease,

alcohol liver disease, and anemia (Wojnar, Bizoń, et al., 1999). While several studies included both diagnosis of liver and pancreatic diseases as variables in their prediction studies, these did not significantly predict withdrawal severity (Eyer, Schuster, et al., 2011; Mennecier et al., 2008; Monte et al., 2009; Wetterling et al., 1994). This was confirmed in meta-analysis (Goodson et al., 2014). Since AWS is frequently encountered in physical trauma patients (Holt et al., 1980), several studies explored possible risk factors of this specific population (Blondell et al., 2004; Findley et al., 2010; Lukan et al., 2002; Salottolo et al., 2017). Regarding trauma mechanisms, severe head injury (Salottolo et al., 2017) and burns (Lukan et al., 2002) were positively, and motor vehicle accidents (Lukan et al., 2002) negatively correlated with DT development. Single studies reported further somatic risk factors like ataxia and polyneuropathia (Wetterling et al., 1994) and current infectious diseases like pneumonia or urinary tract infections (Palmstierna, 2001) as well as diagnostic findings like structural CNS lesions (Eyer, Schuster, et al., 2011). Considering psychiatric comorbidities, a history of non-medical benzodiazepine use (Kraemer et al., 2003; Schuckit et al., 1995) and other sedative psychotropic agents (Morton et al., 1994) have been associated with severe AWS.

2.2.2 Prediction tools

Beyond identifying correlations between single variables and outcomes of AWS severity, attempts have been made to combine multiple predictors into clinically viable rating tools for more accurate risk assessment. Methodologically, three approaches can be distinguished and are further discussed below: The use of regression models to find multivariate prediction models, the evaluation of preexisting assessment scales for risk prediction and the development of rating scales based on findings from literature.

A common approach in previous studies has been to develop prediction models by identifying possible independent risk factors of severe AWS via univariate testing, which are then evaluated in a multivariate regression model: Ferguson et al. applied stepwise logistic regression in a retrospective sample of 200 internal-medicine in-patients to detect risk factors of DT (Ferguson et al., 1996). While the authors reported the application of bootstrap validation, they did not further specify how this method was used. The combination of two variables, longer duration since last drink and

concomitant acute medical illness, yielded a risk of 54% for DT in the study cohort. Palmstierna et al. used a similar approach to identify five binary risk factors of DT in a cohort of 334 inpatients of whom 23 were discharged with a DSM-IV DT diagnosis (Palmstierna, 2001), though they did not report any performance measures of their regression model. In a retrospective study on 822 in-patients that developed AWS severe enough to require pharmacological treatment, Eyer et al. used a stepwise multivariate linear regression framework without prior selection based on univariate testing to separately identify 46 DT and 61 WS cases (Eyer, Schuster, et al., 2011). The parameters which showed significance in multivariate testing were then used to construct nomograms that allow risk calculation for both outcomes, yielding probabilities between 5 and 80% for WS and 0.1 and 70% for DT.

Several authors have evaluated if the clinical use of already established assessment scales for either alcohol-use disorder diagnosis or AWS severity assessment could be further extended to AWS risk prediction: The 10-item Alcohol Use Disorders Identification Test (AUDIT), a screening test for potentially harmful alcohol consumption (Saunders et al., 1993), has been tested for prediction capabilities in three studies: Based on simple t-test group differences, Reoux et al. reported high prediction capabilities (sensitivity 0.982%, specificity of 0.28%) for a total AUDIT score ≥ 27, identifying patients who developed a CIWA-Ar score ≥ 9 and therefore required symptom-triggered medication among 118 alcohol dependent patients (Joseph P. Reoux et al., 2002). Dolman et al. reported similarly performance measures, especially if AUDIT ratings were combined with laboratory parameters such as liver enzymes, for prediction of 17 AWS patients in a sample of 874 medical in-patients of whom 98 were labeled as alcohol dependent (Dolman & Hawkes, 2005). Both studies did not report, if AUDIT and CIWA-Ar assessments were conducted under blinded conditions (Dolman & Hawkes, 2005; Joseph P. Reoux et al., 2002). In a retrospective case-control study of a non-intensive care surgical and medical ward cohort, Pecoraro et al. tested AUDIT-Piccinelli Consumption version's (AUDIT-PC) ability to differentiate between 223 patients with an ICD-9 discharge diagnoses of AWS and 466 randomly selected non-AWS patients (Pecoraro et al., 2014). Applying a hierarchical logistic regression analysis, they reported a 91.0% sensitivity and 89.7% specificity for an AUDIT-PC score of greater than or equal four in identifying AWS patients. Control patients showed a mean AUDIT-PC score of 1.1, implying overall low alcohol

consumption. Considering the CIWA assessment scale, Foy et al. observed higher CIWA ratings before the onset of withdrawal complications like DT and WS in 203 preselected patients with high daily alcohol consumption and alcohol-related problems at a general hospital and therefore attributed a “predictive value” to CIWA assessment, though they did not report on analysis intervals (Foy et al., 1988). In a multivariate regression analysis on 284 patients admitted to an alcohol detoxification unit, Kramer et al. identified a CIWA-Ar rating ≥10 at admission as correlated with subsequent severe AWS development (Kraemer et al., 2003).

A further approach has been to construct risk assessment scales based on findings from literature: Wetterling et al. compiled 22 variables—including information on previous alcohol use, drinking patterns, clinical symptoms at admission and laboratory values—which the authors identified as easily attainable risk factors of different AWS severity outcomes based on an unsystematic literature review into the Luebeck Alcohol Withdrawal Risk Scale (LARS) (Wetterling et al., 2006). The scale was then applied to 100 psychiatric in-patients treated with a symptom-triggered detoxification scheme with the AWS scale as severity assessment. A cut-off value of the LARS that yielded optimal performance metrics (100% sensitivity and 88% specificity) in separating patients with mild to moderate AWS and patients with severe AWS as defined by the maximum AWS scale rating during the withdrawal course was then defined post hoc. Based on a test of internal consistency, the scale was further reduced to 11-item (LARS11) and, excluding chloride measurement due to clinical reasons, 10 item (LARS10) versions. Following a similar approach, Maldonado et al. developed the Prediction of Alcohol Withdrawal Severity Scale (PAWSS) for in-patients with somatic disease based on prior literature findings (Maldonado et al., 2014). They conducted a systematic literature review to identify studies that reported on possible risk factors of moderate to severe AWS outcomes and then, without specifying objective criteria for their decision, chose 10 clinical variables which are assessed if a patient fulfills the threshold criteria of alcohol consumption within 30 days before and/or a positive blood alcohol concentration at admission. In a pilot study, 68 general-medicine in-patients were assessed with the PAWSS, of whom 51 did not fulfill the threshold criteria and thus, were not assessed further. After detoxification of the remaining 17 patients was completed applying a symptom-triggered treatment regime, the authors retrospectively determined the withdrawal course, classifying four cases

as moderate to severe AWS, defined as either a respective CIWA-Ar rating or unspecified clinical judgement. A cut-off value of four on the PAWSS was then chosen post-hoc as optimal separator between the four moderate to severe cases and all 64 other patients, yielding performance measures of 100%. In a further study, the same authors tested the PAWSS in a sample of 403 in-patients on general medicine and surgery wards of whom 29 were judged at-risk of severe AWS, defined as a CIWA-Ar rating ≥ 15 or a respective DSM-IV diagnosis (Maldonado et al., 2015). This resulted in a sensitivity of 93.1% and specificity of 99.5%. Unfortunately, the authors did not report how many patients developed AWS in the whole sample, how CIWA-Ar ratings were distributed between groups, if there was an overlap between positive CIWA-Ar items and symptoms due to medical conditions other than AWS (e.g. nausea, headaches, sweating) and if cases of delirium tremens or withdrawal seizures occurred. Judging by the reported prevalence of 1.7% for alcohol use disorder, which lies far below the general population average (Grant et al., 2015), the study cohort seems to consist of low risk patients.

2.2.3 Translation into clinical practice

Despite the research efforts outlined above, tools for AWS prediction have not been integrated into clinical routine. Several underlying reasons have been recognized and are discussed in the field (Fiellin et al., 1998, 2015; Saitz, 2018):

Mirroring the contexts in which AWS patients are commonly treated, studies on AWS risk factors have been conducted in diverse populations within psychiatric (Wetterling et al., 1994), trauma (Lukan et al., 2002), or general medicine facilities (Maldonado et al., 2014). Consequently, patient characteristics like medical and psychiatric comorbidities, sociodemographic measures, alcohol use history including previous withdrawal experiences and incidence of severe withdrawal outcomes greatly vary between studies (Eyer, Schuster, et al., 2011; Lukan et al., 2002; Maldonado et al., 2015; Salottolo et al., 2017; Wetterling et al., 2006). Different treatment settings are also likely to differ in AUD-specific knowledge and staff training, assessment routines, allocated resources and, importantly, AWS treatment strategies. Importantly, the attended clinicians are intrinsically interested in different predictive questions: screening all admissions to a general hospital for possible severe AWS (Maldonado et al., 2014, 2015) might require different predictive models than separating more or less pronounced AWS cases in patients specifically admitted for detoxification (Eyer,

Schuster, et al., 2011; Wetterling et al., 1994). Further impeding comparability, a wide range of assessment methods and AWS definitions have been applied across studies. Due to their retrospective nature, some studies have retracted ICD diagnosis from patient records as their main severity outcome (Ferguson et al., 1996), although their accuracy in displaying disease trajectories seems questionable (K. J. O'Malley et al., 2005). While using standardized severity ratings as outcomes allow more nuanced outcome estimates, preference for a specific scale varies between countries (Mennecier et al., 2008; Sullivan et al., 1989; Wetterling et al., 1997). Moreover, while these scales have been mostly developed in specialized detoxification settings excluding severely medical-ill patients (Sullivan et al., 1989; Wetterling et al., 2006), they are readily applied to such patients, ignoring overlap between AWS and symptoms of concomitant diseases (Maldonado et al., 2015; Salottolo et al., 2017). Others have based severity estimation solely on unspecified clinical judgement (Wojnar, Bizoń, et al., 1999) or include an option to do so (Maldonado et al., 2015).
Besides outcome assessment, patients were treated with diverse medication strategies and pharmacological agents across studies, including application of various benzodiazepines (Kraemer et al., 2003; Mennecier et al., 2008), clomethiazole (Berggren et al., 2009; Eyer, Schuster, et al., 2011), or adjacent anticonvulsants (Eyer, Schuster, et al., 2011; Hillemacher et al., 2012), that are likely to influence the further withdrawal course with varying efficiency (Amato et al., 2011).

The heterogeneity in study populations and study designs has not been addressed by adequate methodological approaches: Most studies have applied frequentist univariate and multivariate methods, like correlation tests and regression, to identify independent or combined risk factors for severe AWS outcomes (Ferguson et al., 1996; Fiellin et al., 2002; Kraemer et al., 2003; Mennecier et al., 2008; Palmstierna, 2001). To my knowledge, only one study applied bootstrap techniques to ensure internal validity of their results, but does not report how exactly these methods were implemented or how these influenced the study results (Ferguson et al., 1996). Studies comparing populations at different treatment sites or across medical specialties to ensure external validity are missing in the field. Under the term "replication crisis" (Ioannidis, 2005) this overreliance on frequentist p-value testing without methods to ensure reproducibility of the retrieved results has been discussed as a major reason for failing translation of research findings into clinical practice (Munafò et al., 2017).

Without validation, predictive models will unlikely generalize to patients outside the initial study cohort, since they rely on noise and peculiarities in those datasets (Siontis et al., 2015). Expectedly, as has been shown above, previous results in AWS prediction have been largely inconsistent. While a recent meta-analysis offers a warranted overview of reported findings, the authors emphasize the shortcomings of previous studies and encourage further, methodologically more rigorous research rather than recommending the use of the retrieved risk factors in clinical practice (Goodson et al., 2014). Nonetheless, other guidelines and reviews do not refrain from clinical recommendations: The recently updated NICE guideline on alcohol-related complications, intended for practitioners in the British health system, emphasizes the low-quality evidence of AWS prediction but still suggests to consider several risk factors that have been reported in a single study (National Institute for Health and Care Excellence (NICE), 2010; Palmstierna, 2001). Both the LARS as well as the PAWSS have been endorsed by their respective developers for immediate clinical use (Junghanns & Wetterling, 2017; Maldonado et al., 2015). Consequently, a high-impact clinical review has recently simulated the application of these tools for treatment of a AWS patient (E. Wood et al., 2018). Others have warned against such clinical use and stressed the importance of further research efforts applying adequate methods to ensure generalizability (Saitz, 2018).

2.3 Machine learning as a framework for prediction

ML is the study of computational methods designed to solve data-based problems without explicit programming (T. Mitchell, 1997). It can be situated in the broader context of an algorithmic modeling culture questioning the utility of traditional statistics that heavily rely on prior assumptions on given datasets (Breiman, 2001). Instead, ML algorithms contain parameters that are optimized via instance-based learning to directly model optimal input-output relationships based on a given dataset (Hastie et al., 2009). Since such processes are assumed to mirror aspects of human learning, ML is usually considered to be a subfield of artificial intelligence (Topol, 2019). While early applications were restricted by limitations in computing power (Rosenblatt, 1958), ML researchers have since developed a vast array of computational feasible algorithms that allow accurate modeling of possibly non-linear patterns in multidimensional datasets (Kotsiantis, 2007). Since these algorithms can be highly

sensitive to noise and idiosyncratic sample characteristics, referred to as overfitting in ML terminology, methods to ensure generalizability to previously unseen observations are an integral part of state-of-the-art analysis pipelines (Varoquaux et al., 2017). This focus on internal and external validity is considered to be a major advantage to classical statistics, thus contributing to innovations across diverse fields like manufacturing, financial modeling, education and science (Dwyer, Falkai, et al., 2018; Jordan & Mitchell, 2015). Recently, ML approaches yielded promising results for various health-care tasks like x-ray evaluation (Rajpurkar et al., 2017), skin cancer classification (Esteva et al., 2017) or the development of new antibiotic drugs (Stokes et al., 2020). Some diagnostic applications, driven by AI-technology, have recently been approved by the U.S. Food and Drug Administration (Abràmoff et al., 2018).

In psychiatry, ML has been promoted as a crucial tool to translate research efforts into precise, clinical-meaningful applications in areas such as diagnosis, prognosis, treatment enhancement, as well as biomarker prediction (Dwyer, Falkai, et al., 2018). A common framework applied in studies across various psychiatric disorders, including addiction research, is to classify future outcomes based on multidomain baseline variables: For example, Whelan et al. used a wide range of information on genetics, demographics, patient history, cognitive test performance, personality traits as well as structural and functional imaging to predict current and future binge drinking behavior in a large sample of adolescents (n=692) via a combination of regularized logistic regression and elastic net algorithms (Whelan et al., 2014). Importantly, they used nested cross-validation as a means to ensure generalizability (see below) and externally validated their result in a separate dataset (Whelan et al., 2014). While imaging modalities are frequently included in ML prediction studies to possibly detect objective biomarkers (Arbabshirani et al., 2017), such models are currently not translatable to most clinical settings, where comparable imaging technology is not available. Therefore, other researchers have focused on clinical parameters that are easier to obtain: Koutsouleris et al. used a battery of sociodemographic variables as well as multiple, frequently used clinical assessment scales to predict treatment outcomes in a cohort of patients with first-episode psychosis (Koutsouleris et al., 2016). Importantly, they then independently tested a further model with the obtained 10 most predictive variables to generate a shorter, clinically feasible prediction tool (Koutsouleris et al., 2016). Since highly accurate prediction models may be useless if they address questions that can be efficiently assessed by clinical judgement (Wiens

et al., 2019), several studies directly compared their models' performance to clinical ratings, showing superior estimates for suicide risk prediction (Tran et al., 2014) and functional outcome prediction in patients with high-risk for psychosis or recent-onset depression (Koutsouleris et al., 2018).

While ML research efforts are increasing across psychiatry (Janssen et al., 2018), a recent systematic review could identify only three ML studies on alcohol use disorder or alcohol consumption (Mak et al., 2019). Only one study in AWS research has applied ML methods—a random forest algorithm with bootstrapping validation—to identify risk factors of withdrawal seizures (Hillemacher et al., 2012). Since the authors did not observe any seizures in their patient cohort, they instead used prior seizures as the outcome, which undermines the very idea of a prediction (Hillemacher et al., 2012). This hiatus represents a promising opportunity to address the major limitations in previous AWS risk research discussed above: After decades of focus mainly on single risk factor identification, a data-driven ML analysis could exploratively determine multivariate predictive patterns that increase accuracy for meaningful AWS outcomes and generate better understanding about disease trajectories in specific populations. Instead of relying on prior research findings, ML feature selection methods could then be used to objectively built and test clinically feasible prediction tools. Finally, stringent methods to optimize and test generalizability to unseen patients could enable clinical implementation.

2.4 Aims

The primary aim (1) of this dissertation was to develop generalizable and accurate predictive models within a machine learning framework to identify patients at risk of developing severe alcohol withdrawal. To ensure clinical utility, I focused on severity outcomes that are both already implemented in clinical care as well as relevant for differential clinical decision making. Therefore, specific models were built to predict (a) moderate to severe alcohol withdrawal as assessed via the AWS scale, (b) cases of delirium tremens, and (c) cases of withdrawal seizures. Following a hypothesis-free, data-driven approach, all clinical, sociodemographic and laboratory variables available across patients' admission were included in the analysis pipeline. Using a cross-over design, specific models were built in retrospective cohorts of two treatment sites—a psychiatric detoxification ward (Department of Psychiatry and Psychotherapy,

University Hospital, Ludwig Maximilian University of Munich (LMU)) and a specialized toxicology unit (Department of Clinical Toxicology of the Technical University of Munich (TU)).

The secondary aims were to compare differences in predictive accuracy and variable importance across (2) different withdrawal outcomes as well as (3) the two treatment sites. For latter end, separate models were built in both cohorts utilizing only variables available at both treatment sites. These were then tested at the respective other treatment site to generate estimates of external validity.

The following hypotheses were formulated:

(1) Machine learning analysis enables accurate risk predictions for three outcomes of AWS severity within a framework ensuring internal validity.
(2) Predictive performance will vary across withdrawal outcomes with specific variable patterns contributing to each model's performance.
(3) Predictive performance will decrease across treatment sites due to site-specific characteristics in external validation testing, but still yield accurate results.

3 Methods

3.1 Study Cohort

Retrospective data was gathered from in-patient cohorts of two major detoxification units at two university hospitals in Munich.

For the first sample, patient charts of 445 admissions for AWS treatment at the psychiatric detoxification unit of the LMU between 1995 and 2005 were reviewed. The inclusion criteria encompassed an AWS diagnosed defined by ICD-10 criteria (F10.3) comprising DT (F10.4), patient age 18 years or older, and a daily disease course documentation by the attending hospital staff to ensure sufficient data quality. Exclusion was limited to patients with documented multiple drug use in the hospital chart. Patients that self-reported concomitant substance use in the Semi-Structured Assessment for the Genetics of Alcoholism (SAAGA) questionnaire (Bucholz et al., 1994), but were not recognized as multiple-drug users by the attendant clinic staff were still included in the analysis, since the goal was to predict AWS severity in a naturalistic clinical sample. As standard symptom-triggered treatment either clomethiazole or benzodiazepines (diazepam or oxazepam) was administered. If delusional or hallucinogenic symptoms developed, haloperidol was given as an adjunctive treatment. In cases of marked blood pressure increases, clonidine was added. Patients who had suffered from known seizures in the past or who developed seizures during withdrawal received adjacent treatment with carbamazepine. Prophylaxis for vitamin deficiency with thiamin and a multivitamin supplement was provided routinely. Mild cases of AWS, that did not receive pharmacological treatment, were included to achieve a dataset representative of the clinic's treatment population. 399 patients fulfilled these criteria and entered the analysis.

For the second sample, retrospective data of 2691 patients admitted to the TU between 2000 and 2009 was used. This dataset had been already gathered and fully described in two papers (Eyer, Schreckenberg, et al., 2011; Eyer, Schuster, et al., 2011). Patients with ICD-10 diagnosis of alcohol dependence (F10.2) and either alcohol intoxication (F10.0), AWS (F10.3), or DT (F10.4) were included. Exclusion criteria encompassed co-dependence of other psychotropic substances (illegal drugs, benzodiazepines, or other), mild withdrawal without pharmacological treatment, early discontinued therapy,

and treatment protocol deviations, and incomplete patient chart documentation. Symptom triggered therapy with clomethiazole was used to manage withdrawal symptoms. Similar to the LMU sample, clonidine, haloperidol and thiamin were given as adjunctive treatments. As antiepileptic prophylaxis either carbamazepine or valproate had been administered. Inclusion criteria were fulfilled by 812 patients whose data was then used for further analysis.

Due to the retrospective anonymous datasets, the analysis of both cohorts was exempt from evaluation by the respective institutional ethics committees (LMU and TU).

3.2 Variable battery

3.2.1 Data collection

Following a data-driven approach, all variables available at the day of patients' admission were extracted from the charts without any statistical or clinical preconditions. In the TU sample the selection was limited, since the dataset had been already gathered for previous studies.

In the LMU sample patients' age and gender were included as sociodemographic variables. Self-reported daily alcohol consumption was converted to the pure alcohol mass (grams of ethanol) with following equation:

> *Pure alcohol mass (g) = volume (ml) x alcohol by volume (%) x volumetric mass density (constant of 0.8 g/ml) with g=gram, ml=milliliter*

All available laboratory data collected on the day of admission were included, encompassing electrolytes (sodium, potassium, chloride, calcium), liver and bile enzymes (AST, ALT, GGT, alkaline phosphatase, bilirubin total, glutamate dehydrogenase), creatine kinase, lactate dehydrogenase, full blood count and thyroid-stimulating hormone. Furthermore, urine screening tests for opiates, benzodiazepines, cannabinoids, barbiturates, lysergic acid diethylamide (LSD), methadone, 2-ethylidene-1,5-dimethyl-3,3-diphenylpyrrolidine (EDDP), cocaine, and amphetamines were included as dichotomous variables. The creatine concentration in urine was included as a measure of urine test compliance. BrAC had been measured at admission using a breathalyzer. Since all patients had received electroencephalography (EEG) shortly after admission, the respective investigator's

assessment remarks were included categorized in the following groups: excessive beta activity, theta-waves, vigilance disruption, slow basic activity, low potential, and sign of cerebral excitability. Furthermore, 90 sociodemographic and alcohol-related items of the SAAGA (see above) were included, which patients had completed shortly after admission. The questionnaire encompasses self-reported data on education, employment status, medical history, family medical history, previous alcohol patterns, previous detoxification treatment, history of withdrawal symptoms, complications of long-term alcohol use including psychological problems, and utilization of treatment offers. It has been previously used to characterize a large population of AWS patients (Schuckit et al., 1995).

In the TU sample age and gender, several laboratory measures (sodium, potassium, GGT, white blood cell and platelet count, blood alcohol levels), vital parameters at admission, reason for admission, prior withdrawal history (previous withdrawal, history of withdrawal seizures, history of DT) and medical comorbidities (structural CNS lesions, liver status) were available. Deviating from a previous investigation on the TU sample (Eyer, Schuster, et al., 2011), I excluded one variable (day after admission with the highest AWS scale rating), since it is not available at admission and would thus introduce information on the further withdrawal course into the prediction model.

All data was gathered in a spreadsheet and then imported into a MATLAB (R2015a) data-matrix for further analysis.

3.2.2 Data preparation

Applying an initial filtering procedure, all features with ≥ 25% missing values, features centered in ≥ 95% of cases on one value, and patients with ≥ 25% missing features were removed. This heuristic approach has been used previously to remove variables that are potentially difficult to obtain, variables that are highly unlikely to inform the ML algorithms due to low variance as well as patients that might have received inadequate clinical assessment (Koutsouleris et al., 2016).

In the LMU sample a total of 18 features were excluded due to ≥ 95% of values being the same, including certain EEG changes (theta waves, low potential, signs of excitability), several urine drug screening measures (opiates, cannabinoids, barbiturates, LSD, methadone, EDDP, cocaine, amphetamine), and multiple SAAGA

items (history of encephalitis, history of meningitis, history of stroke, history of myocardial damage, combining alcohol with opioids, two unspecified alcohol treatment items). Eight SSAGA features were excluded due to ≥ 25% missing values across participants (age of first episode of seizures/unconsciousness/falling to the ground/amnesia after reducing/stopping drinking, age of first time that alcohol was combined with medication or drugs, age of first alcohol dependence treatment including self-help groups, five unspecified alcohol treatment items). Furthermore, 10 patients were removed due to ≥ 25% missing values. This resulted in a dataset containing 389 patients and 109 variables used for the ML pipeline (LMU discovery).

In the TU sample, the variable “pancreatitis” was excluded due to ≥ 95% of values being the same, three variables were removed because ≥ 25% values were missing across participants (positive urine-based benzodiazepine screening, duration of alcohol dependence in years, daily alcohol consumption in years), and a total of 7 subjects were excluded due to ≥ 25% missing values across all variables. Thus, a dataset of 805 patients and 21 variables was obtained for further analysis (TU discovery).

Additionally, separate datasets containing the nine variables shared across cohorts were built to allow for external validation. Applying the same filtering procedures described above, data on 396 LMU patients (LMU validation) and 797 TU patients (TU validation) could thus be included.

3.2.3 Baseline analysis

Independent-sample t-tests were used for continuous variables and Fisher’s chi-square test for categorical variables to test for group-differences between the LMU and TU validation datasets. Significance levels were defined at $p = 0.01$.

3.3 Prediction outcomes

The 11-item AWS scale, validated in psychiatric in-patients (Wetterling et al., 1997), was chosen to assess AWS symptom severity. Following the cut-off values established during the scale’s development (Wetterling et al., 1997), withdrawal cases were classified as mild (total score ≤ 5) or moderate to severe (total score 6-33) depending

on the highest total score during the whole in-patient detoxification. These risk groups have been previously used in randomized, controlled AWS treatment studies (Soravia et al., 2018) and reflect a degree of clinical decision making. While the maximum AWS score was documented for all TU patients, it was retrospectively established in the LMU sample from chart information by two investigators (Kristina Adorjan, thesis supervisor; Gerrit Burkhardt). WS and DT were defined as separate, binary variables and based on explicit diagnosis of the attending medical doctors.

A clinically motivated classification approach was chosen to guide decision-making at an individual patient's admission: a) moderate to severe AWS (MSAWS), including all cases of WS and/or DT in addition to maximum AWS scale ratings, was predicted to allow for a general risk stratification at admission; b) cases of DT and c) cases of WS were separately predicted to identify patients possibly benefiting from more frequent monitoring and specific medication (e.g. anticonvulsants for WS, antipsychotics for DT). Both DT and WS were only predicted in the TU dataset due to low occurrence in the LMU sample.

3.4 ML pipeline

In this chapter, each component of the ML analysis is first described separately, before I outline their integration into an automated ML analysis pipeline to build the predictive models. The ML setup was first developed in the LMU discovery sample to predict MSAWS and then applied to each outcome in the TU discovery sample. For external validation, separate MSAWS models were built in both validation samples and then tested across cohorts (see below). All ML analysis steps were computed via the open-source software package NeuroMiner (version 0.992; http://www.pronia.eu/neurominer) running with MATLAB (R2015a). An overview of the ML pipeline is depicted in Figure 3.1.

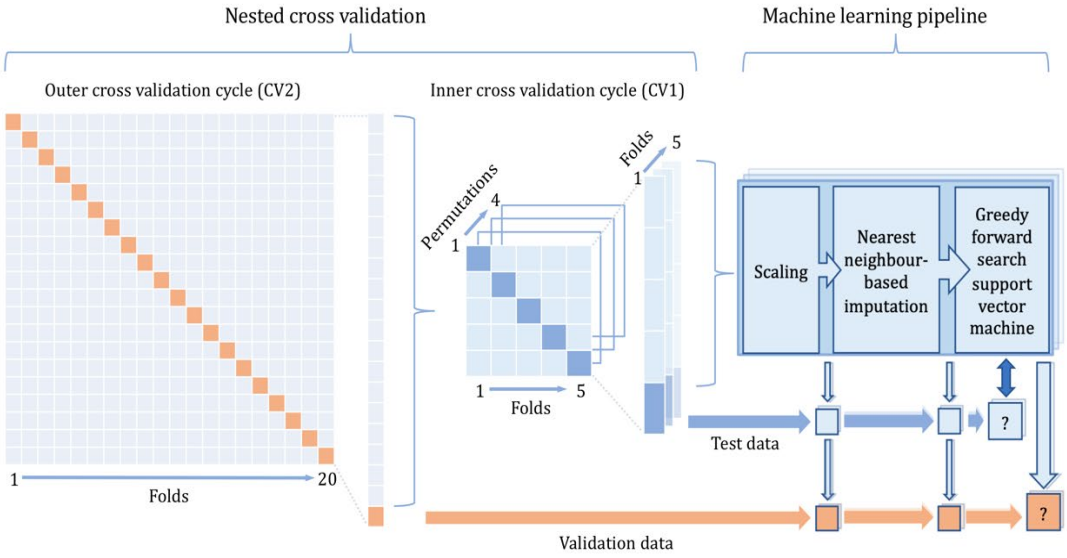

Figure 3.1. Machine learning pipeline. In the inner cross-validation cycle (CV1), classifier performance, as defined by left-out CV1 test data, was optimized via hyperparameter training. In the outer 20-fold cross-validation cycle (CV2), left-out CV2 test data was used to estimate the models' generalizability to yet unseen patients.

3.4.1 Nested cross-validation setup

The translation of predictive tools into clinical care requires means to optimize and test a predictive model's generalizability (Stone, 1974). Generalizability is defined as the extent to which such a model accurately classifies previously unseen patients (Varoquaux et al., 2017). This test mimics the real-world situation in which a clinician has to choose optimal treatment approaches or venture a disease prognosis based on available patient information (Dwyer, Falkai, et al., 2018). Preferably, resampling techniques like cross-validation (CV) are used that separate datasets into instances used for training the model and instances used for testing its performance (Varoquaux et al., 2017). In *k*-fold CV the dataset is split into *k* folds of a predefined size. Leaving one fold out, the model is then trained on the remaining *k*-1 folds, after which its performance is evaluated on the left-out data. This process is usually repeated for every *k*-fold to increase both variability in the training data as well as the number of test subjects (Hastie et al., 2009). This process can be further enhanced to allow for both testing a model's predictive performance as well as optimizing its generalizability: To this end, I used repeated nested CV, currently considered the gold-standard technique for generalizability evaluation (Filzmoser et al., 2009; Ruschhaupt et al.,

2004). In nested CV, an inner CV loop (CV1) is used to adjust the algorithms' hyperparameters towards optimal predictive performance on held-out CV1 test subjects. To assess generalizability, the learned model is then applied to test subjects in an outer CV loop (CV2), that have not been used in the CV1 training cycle and are thus independent of the parameter optimization. To implement the CV setup, a number of k-folds has to be chosen. While statistical criteria (Hastie et al., 2009) and commonly used sizes (Breiman, 2001) have been proposed, the optimal choice may depend on the sample size, number of variables as well as the algorithms and feature selection techniques (see below) applied (Dwyer, Falkai, et al., 2018). I therefore chose a setup that has been successfully used in a clinical dataset of similar size within a comparable ML setup (Koutsouleris et al., 2016): In the CV2 loop, 20 training and validation folds were randomly compiled. Each CV2 training fold was then iteratively separated into five CV1 training and test sets. These were subsequently permuted four times to induce sample variance.

3.4.2 Preprocessing

A major advantage of nested CV is the possibility to also separate any preprocessing steps that are required prior to statistical analysis between the training and test subjects: If procedures like imputation are applied before the dataset is divided, information about the whole dataset is likely to be included and therefore learned by the algorithms in the training phase—a bias usually termed "information leakage" (Dwyer, Falkai, et al., 2018). The following preprocessing steps were carried out on each of the 20 (5 folds x 4 permutations) subsets in the CV1 loop: First, variables with no variance were excluded, since they will likely not benefit classification. Secondly, the training data was scaled (0 to 1). Variable scaling is applied in multivariate analysis since variables in greater numeric ranges tend to dominate small range variables, which may alter the outcome of machine learning algorithms (Keun et al., 2003; Koutsouleris et al., 2016). In addition, complex numerical calculations that require increased computational capacity are thus avoided (Keun et al., 2003). Thirdly, missing values were substituted via nearest neighbour-based imputation (Jönsson & Wohlin, 2004; Troyanskaya et al., 2001): For each case in the dataset with a missing value, all cases that provided a measure for this value were identified. The similarity of these cases to the case with a missing value was then determined via a measure of

geometrical distance (for continuous variables the Euclidean distance was used, applying the “distance2” function from the Large Margins Nearest-Neighbours toolbox in MATLAB; for categorical variables the dichotomous Hamming distance was applied via the “pdist2” from the Statistics and Machine Learning Toolbox in MATLAB). The median of the five most similar—and therefore nearest—neighbours was used to fill the missing value. This imputation method was repeatedly applied to all missing values in the dataset, always using the primary, non-imputed dataset as a source for neighbour identification. Finally, since I was primarily interested in building accurate prediction models with all information available at a patient’s admission and without any prior hypothesis regarding variable importance, I did not regress out nuisance covariates like age and gender.

3.4.3 Choice of Algorithm

A crucial feature of ML is the deployment of sophisticated algorithms that can be modified to optimally perform statistical tasks like regression, classification or clustering (Kotsiantis, 2007). In contrast to classical statistical methods like linear or logistic regression, which fit simple functions to a given dataset, ML algorithms allow the modification of regularization parameters (also called hyperparameters) to optimize accuracy and generalizability of learned functions (T. Mitchell, 1997). There is a wide range of algorithms available (e.g. random forests, neural networks, regularized regression), which differ in their ability to model complex relationships as well as their options for modification (Hastie et al., 2009). For the classification tasks at hand I chose a linear Support Vector Machine algorithm (SVM; LIBSVM 3.12; http://www.csie.ntu.edu.tw/~cjlin/libsvm)(Noble, 2006), which has been shown as stable and efficient in datasets with high collinearity and noise (Cortes & Vapnik, 1995). Based on early multivariate algorithms, SVMs have been developed to more accurately classify subjects to groups while maximizing generalizability (Cortes & Vapnik, 1995): This is achieved by only using cases at the outer boundaries of an outcome’s distribution—so called support-vectors—to maximize a margin between those cases and a hyperplane that optimally separates cases of different outcomes. The size of this margin, controlled by the C hyperparameter, determines the degree to which cases are allowed to be misclassified. While a so-called hard margin classifier may lead to the correct classification of all cases, it will likely overfit the dataset and thus perform

worse when used to classify unseen cases. In comparison, a soft margin may achieve a worse performance in the training phase, but possibly generalizes better to new cases. Optimal hyperparameter selections for a given dataset can be learned from a heuristically range of possible values (in my analysis a sequence from $C=2^{-6}$ to $C=2^{6}$) as described below.

In a supplementary analysis, the SVM-based models were compared to logistic regression (LR), tested in the identical CV framework, since LR has been a frequent choice in previous AWS prediction studies (Eyer, Schuster, et al., 2011). The performance of both algorithms was compared using the McNemar's test (McNemar, 1947).

3.4.4 Feature Selection and Training

In multivariate analyses of datasets with a high-dimensional variable space, a mismatch between the number of variables relative to the number of cases can lead to a decrease in accuracy and generalizability—referred to as the "curse of dimensionality" (Bellman, 1957). Feature selection encompasses methods to reduce the variable space of a given dataset in order to increase predictive performance (Guyon, 2003). While this can theoretically be achieved by preselection based on prior hypotheses (e.g. an expert opinion or the results of previous studies), ML approaches tend to integrate feature selection either as part of the preprocessing pipeline or the classifier's training (Guyon, 2003).

Following the latter approach, a greedy forward search wrapper (Inza et al., 2004) was applied on the preprocessed CV1 training data to identify a subset of variables in the variable pool that maximizes prognostic performance. This wrapper followed a simple "hill-climbing logic" (Guyon, 2003): In each CV1 training sample, the predictive value of each variable was evaluated using the linear SVM. The variable that achieved the best performance (defined as the highest balanced accuracy (BAC= sensitivity + specificity) / 2 (Brodersen et al., 2010)) on the held out CV1 test subjects was extracted. The same procedure was then reiterated over the remaining variable pool to select the second-best performing variable, which was also extracted. This process was reiterated until 80% of the original variables were chosen. The subset was then used to optimize a hyperplane between most similar good- and poor-outcome cases

(=the support vectors). Repeating the whole selection process across the predefined sequence of C hyperparameters, optimal parameters for each CV1 fold were established. In an additional weighting step, these optimal parameters were multiplied with the inverse ration of the training group sizes to adjust for uneven outcome distributions and resulting classification bias (Yang et al., 2005). Finally, all CV1 models in a given partition were retrained using the best weighted hyperparameter settings across all models (Koutsouleris et al., 2016).

3.4.5 Model Evaluation

The separate training in each of the 20 CV1 partitions (5 folds x 4 permutations) resulted in 20 predictive models that were then combined into a so-called ensemble classifier. Ensemble theory is based on the central limit theory put forward by Galton (Galton, 1886) and assumes that a set of diverse classifiers is likely to yield more generalizable prediction results, analogous to a team of experts from different domain knowledge (Polikar, 2006). Following this approach, the generated SVM ensemble predicted each CV2 test subjects group membership by the following decision rule:

$$f(x_{CV2}) = \text{sign}(\sum_{i=1,j=1}^{k_{CV1}=5,n_{CV1}=4} f_{i,j}(x_{CV2})/(k_{CV1} * n_{CV1})),$$

where $f_{i,j}(x_{CV2})$ is the average decision value of a CV1 ensemble for a given CV2 test subject and k / n are the numbers of CV folds / permutations at the CV1 level. Comparing all resulting CV2 predictions with the actual observed outcome, the ensemble performance on yet unseen cases was calculated. This yielded both performance metrics commonly reported only in ML research (e.g. BAC) as well as several classical statistical estimates (e.g. sensitivity, specificity, area under the curve (AUC); Table 4.2). Additionally, the prognostic summary index (PSI = PPV+NPV-1; 1/PSI measures how many patients need to be assessed to correctly predict a single outcome) was provided as an estimate of certainty based on prior knowledge of disease prevalence (Linn & Grunau, 2006).

In order to establish the statistical significance of each model's predictive performance I conducted a permutation analysis (Gaonkar & Davatzikos, 2013; Golland & Fischl, 2003): Therefore, 100 random permutations of the outcome labels were calculated. Retraining all classification models for each permutation, using the same ML setup and

feature subsets obtained in the non-permuted analyses, a random ensemble prediction for each CV2 test subject was retrieved and thus a null distribution of out-of-training classification performance generated. The significance of the non-permuted models was then evaluated by calculating the number of predictions where the permuted out-of-training BAC was higher or equal to the non-permuted BAC divided by the number of permutations. Significance was determined at $\alpha=0.01$.

In two supplementary analyses I further explored specific aspects of each model: Firstly, I calculated the median balanced accuracy and respective standard deviations across the 20 CV2 folds for all discovery models to gain insight into performance variability. Secondly, to explore possible age and gender effects on the predictive performance of all significant models via t-test (age) or chi squared (gender) analysis between correctly versus incorrectly classified patients.

3.5 External validation

In order to test the generalizability of the ML classifiers across treatment sites, an external validation analysis was conducted. Since no cases of DT and only five cases of WS were observed in the LMU sample, I focused on predicting the MSAWS outcome. In an out-of-sample cross-validation (OOCV) setup separate models were trained based on all shared variables in both datasets (LMU and TU validation samples), following the described ML setup. For external validation these learned models were then applied without further modification to the CV2 test subjects of the respective other dataset. Thus, the performance of a model trained at one treatment site was directly tested in an independently collected dataset, excluding any information leakage between sites.

3.6 Clinical Post-hoc validation

In an additional post-hoc analysis, I explored how each significant model's outcome predictions were related to pharmacological AWS treatment decisions in order to better understand their potential clinical relevance. Thus, linear regression models (GLM) were calculated with each models' mean decision scores as independent variable and the cumulative clomethiazole dose (defined as the total amount of clomethiazole a patient received during detoxification) as dependent variable. Clomethiazole was the most common AWS medication used at both treatment sites. Cases that did not follow this regimen were thus excluded from the analysis.

4 Results

4.1 Baseline characteristics

The baseline characteristics of all discovery and validation datasets are reported in Table 4.1.

148 (38%) LMU discovery patients were rated as MSAWS and 12 (3%) WS were documented. No cases of DT were reported. The TU discovery sample consisted of 674 (84%) MSAWS, 59 (7%) WS, and 45 (6%) DT cases. Due to the data preparation criteria, discovery and validation datasets at each site differed in case count, but this did not change the proportion of MSAWS cases. Univariate testing of group differences between the external validation samples revealed the following significant results (Table 4.1, indicated in red): The TU sample included more MSAWS cases, higher age, more cases with self-reported previous WS, lower mean measurements for sodium and platelets, and higher mean measurements for potassium, GGT, and BrAC. Non-significant differences were found for patients' gender and white blood cell count. The MSAWS outcome distributions for both datasets are shown in Figure 4.1.

Table 4.1. Prediction targets and full variable battery of the discovery and validation samples

		Discovery samples		**External validation samples**		
		LMU discovery (n=389)	TU discovery (n=805)	LMU validation (n=396)	TU validation (n=797)	t and χ^2tests, (p-value)[1]
OUTCOME						
	Mild AWS	241 (62%)	131 (16%)	247 (62%)	129 (16%)	..
	MSAWS[2]	148 (38%)	674 (84%)	149 (38%)	668 (84%)	261.5 (<0.001)
	Delirium tremens	..	45 (6%)	..	..	..
	Withdrawal seizures	12 (3%)	59 (7%)	..	..	..
VARIABLES						
Sociodemographic						
	Mean age, years	43.1 (9.7)	45.2 (10.4)	43.1 (9.8)	45.3 (10.4)	3.4 (<0.001)
	Sex, female	81 (21%)	216 (27%)	83 (21%)	216 (27%)	5.3 (0.021)

Electroencephalogram					
Excessive beta activity	81 (21%)	..	..	..	..
Vigilance disruption	49 (13%)	..	..	..	..
Slow basic activity	23 (6%)	..	..	..	..
Clinical chemistry on admission					
Sodium, mmol/l	141 (3)	139 (5)	141 (3)	139 (5)	-6.06 (<0.001)
Potassium, mmol/l	4.1 (0.4)	4.2 (0.5)	4.1 (0.4)	4.2 (0.5)	3.68 (<0.001)
Chloride, mmol/l	103 (5)	..	..	..	..
Calcium, mmol/l	2.40 (0.15)	..	..	..	..
Aspartate Aminotransferase, mU/ml	62 (86)	..	..	..	..
Alanine Aminotransferase, mU/ml	48 (49)	..	..	..	..
Gamma Glutamyltransferase, mU/ml	222 (427)	417 (634)	220 (423)	417 (635)	5.56 (<0.001)
Alkaline Phosphatase, mU/ml	121 (69)	..	..	..	..
Bilirubin Total, mg/dl	0.82 (1.17)	..	..	..	..
Glutamate Dehydrogenase, mU/ml	24 (67)	..	..	..	..
Creatine Kinase, mU/ml	127 (253)	..	..	..	..
Lactate Dehydrogenase, mU/ml	214 (82)	..	..	..	..
White Blood Cell Count, cells/nl	7.6 (2.6)	7.4 (2.9)	7.6 (2.6)	7.4 (2.9)	-1.12 (0.261)
Neutrophils, %	63.1 (10.7)	..	..	..	..
Lymphocytes, %	25.2 (9.1)	..	..	..	..
Monocytes, %	9.1 (3.3)	..	..	..	..
Eosinophils, %	1.8 (1.5)	..	..	..	..
Basophils, %	0.9 (0.6)	..	..	..	..
Red Blood Cell Count, cells/pl	4.54 (0.52)	..	..	..	..
Hemoglobin, g/dl	14.8 (1.7)	..	..	..	..
Hematocrit, %	44.1 (4.6)	..	..	..	..
Mean Corpuscular Volume, fl	97.3 (5.3)	..	..	..	..
Mean Corpuscular Hemoglobin, pg	32.8 (2.1)	..	..	..	..
Mean Corpuscular Hemoglobin Concentration, g/dl	33.8 (1.0)	..	..	..	..
Platelet count, cells/nl	222 (80)	187 (94)	223 (81)	187 (94)	-6.43 (<0.001)
Thyroid-Stimulating Hormone, uU/ml	8.74 (101.50)	..	..	..	..
Creatinine, mg/dl	..	0.7 (0.4)	..	..	..
Blood Urea Nitrogen, mg/dl	..	10 (5)	..	..	..
Creatinine Urine, mg/dl	156 (285)	..	..	..	..
Benzodiazepines positive in urine	81 (21%)	..	..	..	..
Ethanol in serum, g/l	..	2.4 (1.7)	..	..	..
Breath alcohol concentration, ‰	0.96 (1.06)	..	0.96 (1.06)	1.93 (1.36)[3]	12.33 (<0.001)

Patient-reported data					
Approximated daily alcohol intake, g	244 (130)	..	..	..	..
Body height, cm	176 (8)	..	..	..	..
Body mass, kg	76 (13)	..	..	..	..
Number of first degree family member with alcohol dependency	102 (26%)	..	..	..	..
Number of second degree family member with alcohol dependency	63 (16%)	..	..	..	..
Years in school (not specified), years	13 (4)	..	..	..	..
Graduation (not specified)	357 (92%)	..	..	..	..
Unemployment during the last 12 months	73 (19%)	..	..	..	..
Employment at the moment	187 (48%)	..	..	..	..
History of ...					
... high cholesterol	105 (27%)	..	..	..	..
... high or low blood pressure	184 (47%)	..	..	..	..
... migraine	51 (13%)	..	..	..	..
... head injury/concussion/seizure/unconciousness >5min/enzephalitis/meningitis/stroke	221 (57%)	..	..	..	..
... head injury	112 (29%)	..	..	..	..
... concussion	139 (36%)	..	..	..	..
... withdrawal seizure	71 (18%)	336 (42%)	70 (18%)	334 (42%)	69.4 (<0.001)
... unconciousness >5 min	76 (20%)	..	..	..	..
... vascular disease	33 (8%)	..	..	..	..
... cardiac disease	47 (12%)	..	..	..	..
... hepatic disease	167(43%)	..	..	..	..
... thyroid disease	43 (11%)	..	..	..	..
... asthma	32 (8%)	..	..	..	..
Number of treatments in psychiatric hospital/ detoxification unit	23 (17)	..	..	..	..
Age of onset (alcohol dependency), years	30 (9)	..	..	..	..
Age of first drink, years	15 (5)	..	..	..	..
First period of drinking regularily (at least once/month) for at least 6 months, age, years	21 (8)	..	..	..	..
First time being drunk (slurred speech, unsteady gait), age, years	18 (6)	..	..	..	..
Drunkeness more than once before age of 15 years	61 (16%)	..	..	..	..
Approximated highest daily alcohol intake ever, g	536 (341)	..	..	..	..
Symptoms caused by alcoholic beverage:					
… flush	169 (43%)	..	..	..	..
… hives	41 (11%)	..	..	..	..

… drowsiness	294 (76%)	..	..	..	..
… nausea/vomiting	214 (55%)	..	..	..	..
… headaches, feeling of heaviness/hammering	127 (33%)	..	..	..	..
… rapid heartbeat	113 (29%)	..	..	..	..
Longest period of abstinency since drinking regularily, months	22 (52)	..	..	..	..
Ability to stop drinking at any time	106 (27%)	..	..	..	..
History of withdrawal symptoms:					
… tremor	196 (50%)	..	..	..	..
… insomnia	145 (37%)	..	..	..	..
… depressions	113 (29%)	..	..	..	..
… sweating	190 (49%)	..	..	..	..
… heart hurry	85 (22%)	..	..	..	..
… nausea/vomiting	114 (29%)	..	..	..	..
… weakness	119 (31%)	..	..	..	..
… headaches	89 (23%)	..	..	..	..
… optical/acoustic hallucinations	34 (9%)	..	..	..	..
Age of first occurrence of withdrawal symptoms, years	36 (10)	..	..	..	..
Number of episodes with withdrawal symptoms	23 (34)	..	..	..	..
Alcohol intake to stop withdrawal symptoms (3 or more times)	282 (72%)	..	..	..	..
Medication/ drug use to stop withdrawal symptoms	98 (25%)	..	..	..	..
Withdrawal related seizures/unconsciousness/falling to the ground/amnesia	78 (20%)	..	..	..	..
Long-term complications of alcohol use:					
… liver diseases/jaundice (before)	80 (21%)	..	..	..	..
… intestinal diseases or blood vomiting	46 (12%)	..	..	..	..
… pancreatitis	21 (5%)	..	..	..	..
… tingling sensations or numbness for hours	29 (7%)	..	..	..	..
… memory problems (also not drinking related)	64 (16%)	..	..	..	..
… other problems	23 (6%)	..	..	..	..
Combination of alcohol with medication/drugs (3 or more times) although knowing this could be dangerous	132 (34%)	..	..	..	..
Combination of alcohol with...					
... cannabis	39 (10%)	..	..	..	..
... benzodiazepines	44 (11%)	..	..	..	..

... antidepressants/ anticonvulsiva/ antipsychotics/ z-substances/ clomethiazole	29 (7%)	..	..	..	..
... cocaine, MDMA, amphetamines, LSD	40 (10%)	..	..	..	..
Psychological/ emotional problems caused by drinking:					
… feeling of depression or lacking interest in things (≥ 24 h, impairing normal behavior)	213 (55%)	..	..	..	..
… restless, impatient, nervous (impairing normal behavior)	205 (53%)	..	..	..	..
… unable to think clearly (≥ 24 h, impairing normal behavior)	163 (42%)	..	..	..	..
… paranoid, other people appearing strange (≥ 24 h, impairing normal relationships)	67 (17%)	..	..	..	..
… hearing, seeing, smelling non-existing things	72 (19%)	..	..	..	..
Continued drinking despite knowing that drinking could cause psychological/emotional problems	282 (72%)	..	..	..	..
Consultation of professional therapist because of psychological/ emotional problems	204 (52%)	..	..	..	..
Previous treatment for drinking behavior	267 (69%)	..	..	..	..
Treatment for drinking behavior by...					
... AA or other self-help group	161 (41%)	..	..	..	..
...out-patient alcohol program	80 (21%)	..	..	..	..
...other out-patient treatment	46 (12%)	..	..	..	..
...in-patient alcohol program	213 (55%)	..	..	..	..
...in-patient treatment because of medical complications	63 (16%)	..	..	..	..
Previous withdrawal	..	575 (71%)	..	..	..
History of delirium	..	111 (14%)	..	..	..
Structural cerebral lesions	..	41 (5%)	..	..	..
Preexisting comorbid conditions					
Liver cirrhosis	..	99 (12%)	..	..	..
Liver Cirrhosis: Child-Pugh A/B/C	..	0.15 (0.4)	..	..	..
Others (at admission)					
Heart rate, bpm	..	97 (18)	..	..	..
Systolic blood pressure, mmHg	..	140 (20)	..	..	..
Ventricular extrasystoles (ecg)	..	119 (15%)	..	..	..
Cause of admittance: intoxication	..	309 (38%)	..	..	..
Cause of admittance: seizure	..	125 (16%)	..	..	..

Notes: Values stated as % or mean (SD). [1]External validation data was compared between samples using T-test and χ^2. [2]MSAWS = Moderate to severe AWS-scores, [3]Breath alcohol concentration = Blood alcohol concentration multiplied by the inverse blood/breath ratio (2100:1).

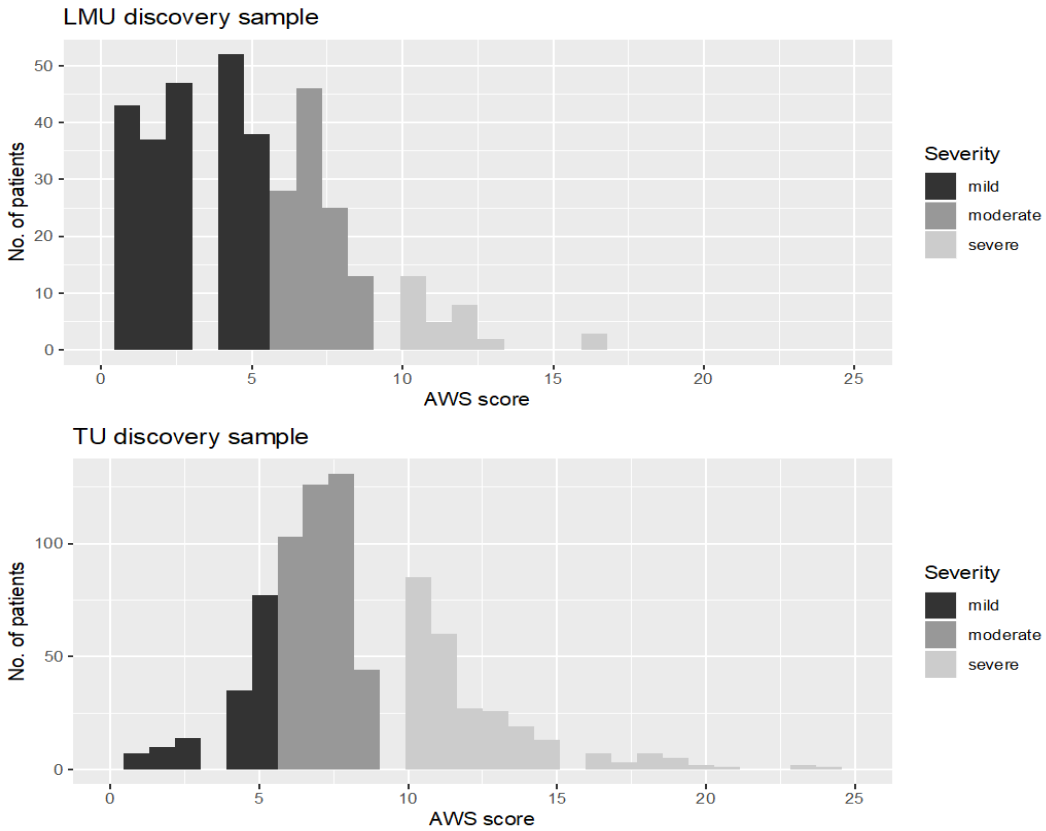

Figure 4.1. AWS score distribution. Maximum score on the Alcohol Withdrawal Syndrome scale (AWS) during withdrawal treatment in both study cohorts. Risk groups are defined as mild (total score ≤ 5), moderate (total score 6-9), and severe (total score ≥ 10)

4.2 Main support vector machine classifiers

Table 4.2. Predictive performance of the main discovery SVM classifiers

	TN	TP	FP	FN	Sens	Spec	BAC	PPV	NPV	PSI	AUC
LMU discovery:											
MSAWS*	170	101	71	47	68.2	70.5	69.4	58.7	78.3	37.1	0.75
TU discovery:											
MSAWS*	77	358	54	316	53.1	58.8	55.9	86.9	19.6	6.5	0.59
DT*	515	37	245	8	82.2	67.8	75	13.1	98.5	11.6	0.75
WS	495	24	251	35	40.7	66.4	53.5	8.7	93.4	2.1	0.54

Notes: * Significance as defined via permutation testing: p<0.01. Sens, Spec, BAC, PPV, NPV, PSI and AUC in %. Abbreviations: True negatives (TN), true positives (TP), false positives (FP), false negatives (FN), sensitivity (Sens), specificity (Spec), balanced accuracy (BAC), positive predictive value (PPV), negative predictive value (NPV), prognostic summary index (PSI), area-under-the-curve (AUC)

4.2.1 LMU discovery sample

Trained and tested on the LMU discovery dataset, the SVM classifier predicted MSAWS cases with a balanced accuracy of 69.4%. Permutation testing established this result as statistically significant ($p < 0.01$). Considering the prognostic summary index, an additional predictive certainty of 37.1% was achieved. Sensitivity and specificity were 68.2% and 70.5% respectively (Figure 4.3). Receiver operator characteristics are depicted in Figure 4.2. Further statistics are reported in Table 4.2. A feature selection probability of more than 50%, indicating high predictive value, was observed for 5 poor-outcome predictors (high BrAC, positive urine-based benzodiazepine screening, low platelets, history of blood pressure abnormalities, years of schooling; Figure 4.4).

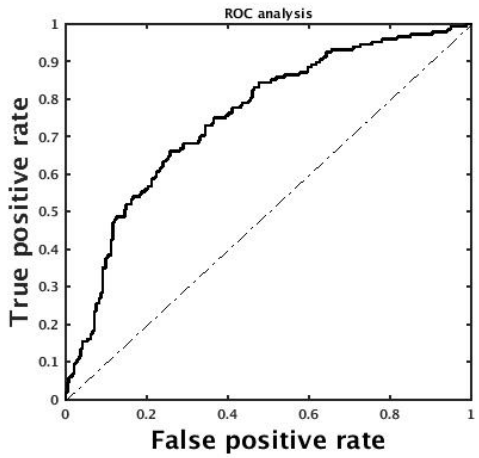

Figure 4.2. Area under the receiver operator curve for the LMU discovery MSAWS model

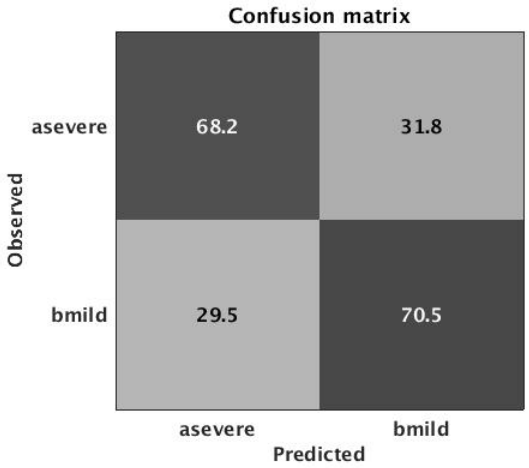

Figure 4.3. Confusion matrix for the LMU discovery MSAWS model

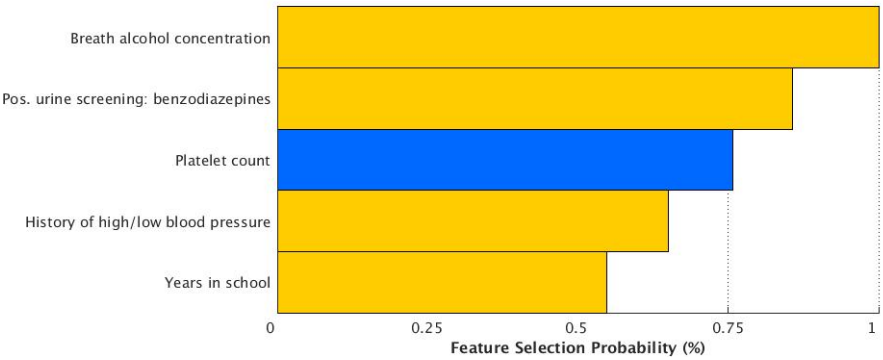

Figure 4.4. Feature selection probability for the LMU discovery MSAWS model

4.2.2 TU discovery sample

4.2.2.1 Moderate to severe AWS (MSAWS)

The TU discovery classifier predicted MSAWS with a balanced accuracy of 55.9%, corresponding to a prognostic summary index of 6.5%. Permutation testing still showed significance ($p < 0.01$). Sensitivity was 53.1% and specificity was 58.8% (Figure 4.6). Additional statistics can be found in Table 4.2. Receiver operator characteristics are depicted in Figure 4.5. A feature selection probability of > 50% was observed for three variables (history of DT, low platelets, high heart rate at admission; Figure 4.7).

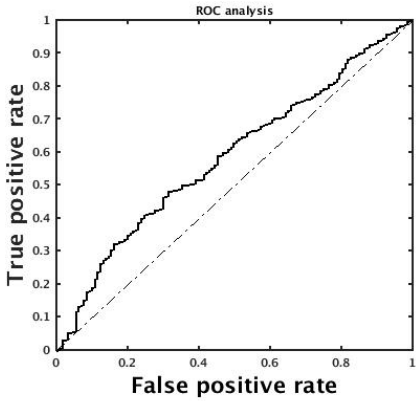

Figure 4.5. Area under the receiver operator curve for the TU discovery MSAWS model

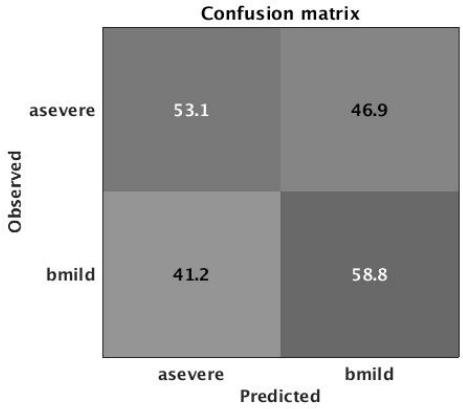

Figure 4.6. Confusion matrix for the TU discovery MSAWS model

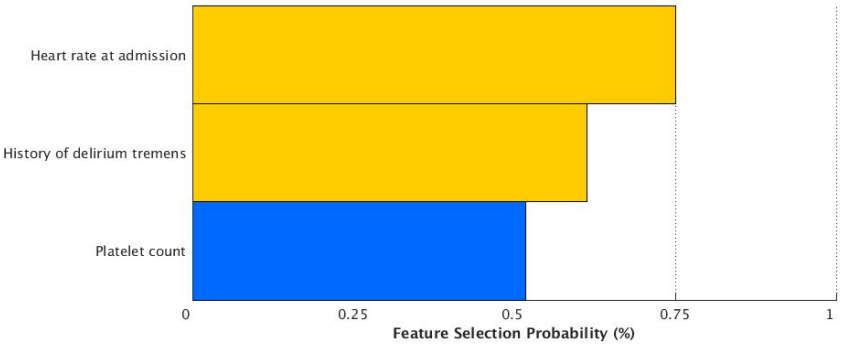

Figure 4.7. Feature selection probability for the TU discovery MSAWS model

4.2.2.2 Delirium tremens (DT)

Delirium tremens cases were predicted with a balanced accuracy of 75% by the SVM classifier, corresponding to a prognostic summary index of 11.6%. The result was significant in in permutation testing ($p<0.01$). Sensitivity was 82.2% and specificity was 67.8% (Figure 4.9). Receiver operator rates are depicted in Figure 4.8 and further statistics are reported in Table 4.2. High feature selection probability (>50%) was observed for 3 poor-outcome predictors (low platelets, age, structural CNS lesions; Figure 4.10).

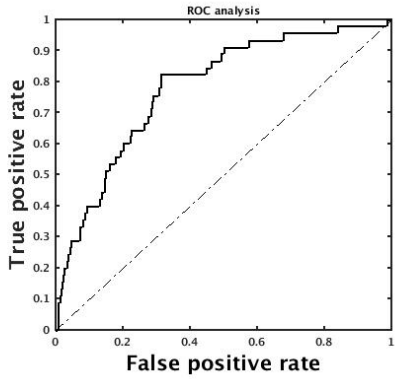

Figure 4.8. Area under the receiver operator curve for the TU discovery DT model

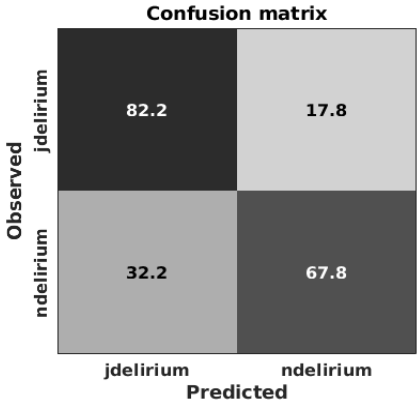

Figure 4.9. Confusion matrix for the TU discovery DT model

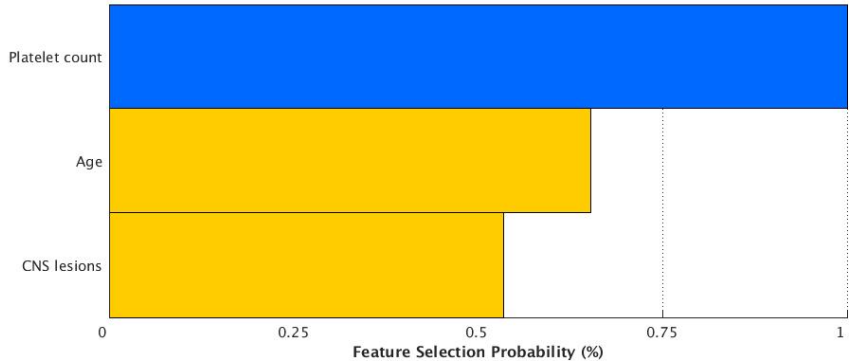

Figure 4.10. Feature selection probability for the TU discovery DT model

4.2.2.3 Withdrawal seizures (WS)

The TU WS model showed a balanced accuracy of 53.5%, established as non-significant in permutation testing (p=0.1). Sensitivity was 40.7% and specificity was 66.4% (Figure 4.12). Receiver operator rates are depicted in Figure 4.11 and further statistics are reported in Table 4.2. Feature selection probability of > 50% was observed for 4 bad-outcome predictors (male gender, liver cirrhosis diagnosis, high Child-Pugh classification rating, history of withdrawal seizures; Figure 4.13).

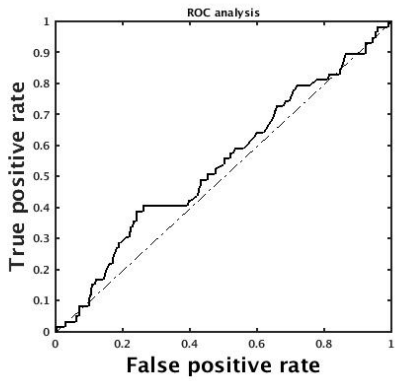

Figure 4.11. Area under the receiver operator curve for TU discovery WS model

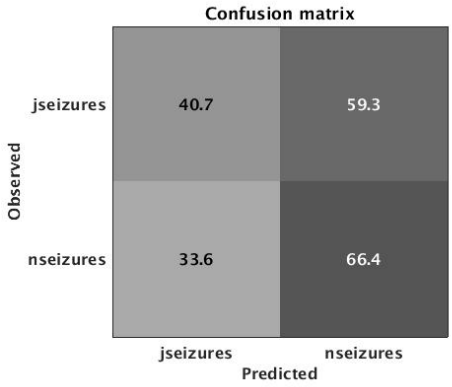

Figure 4.12. Confusion matrix for the TU discovery WS model

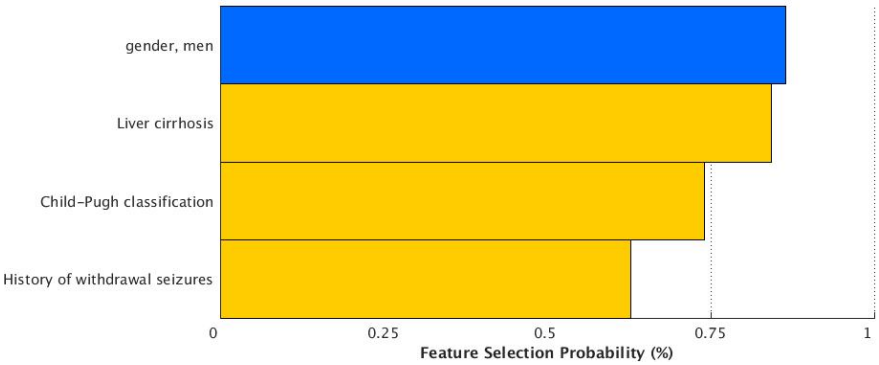

Figure 4.13. Feature selection probability for the TU discovery WS model

4.3 External validation analysis

In out-of-sample validation, the SVM classifiers did not find generalizable solutions: Trained and tested on LMU validation, MSAWS was predicted with a significant balanced accuracy (BAC = 66.3%, p<0.01), but showed a BAC of only 48.1% when presented with the TU validation data. For TU validation, the reduced variable set resulted in a SVM model without significant predictions (BAC = 53%, p=0.18) that, expectedly, did result in out-of-sample predictions below chance (BAC = 38.5%).

Table 4.3. Predictive performance of the external validation SVM classifiers

	TN	TP	FP	FN	Sens	Spec	BAC	PPV	NPV	PSI	AUC
LMU validation:											
MSAWS*	75	203	74	44	82.2	50.3	66.3	73.3	63.0	36.3	0.69
OOCV	73	265	56	403	39.7	56.6	48.1	82.6	15.3	-2.1	0.47
TU validation:											
MSAWS	58	408	71	260	61.1	45.0	53.0	85.2	18.2	3.4	0.53
OOCV	61	89	88	158	36.0	40.9	38.5	50.3	27.9	-21.9	0.34

Notes: * Significance as defined via permutation testing: p<0.01. Sens, Spec, BAC, PPV, NPV, PSI and AUC in %. Abbreviations: True negatives (TN), true positives (TP), false positives (FP), false negatives (FN), sensitivity (Sens), specificity (Spec), balanced accuracy (BAC), positive predictive value (PPV), negative predictive value (NPV), prognostic summary index (PSI), area-under-the-curve (AUC).

4.4 Post-hoc analysis

For treatment of AWS symptoms, 102 patients in the LMU discovery dataset and 800 patients in the TU discovery dataset had received pharmacotherapy with clomethiazole. Entering the significant SVM discovery models' mean decision scores as independent variable and the patients' cumulative clomethiazole demand as dependent variable, significant linear relationships were found for the LMU (F(1,100)=15.66, R2=0.135, p<0.001, Figure 4.14) and TU (F(1,798)=13.78, R2=0.017, p<0.001; Figure 4.15) MSAWS models. A respective linear association was not found for the TU DT classifier (F(1,798)=5.8, R2=0.007, p=0.016; Figure 4.16).

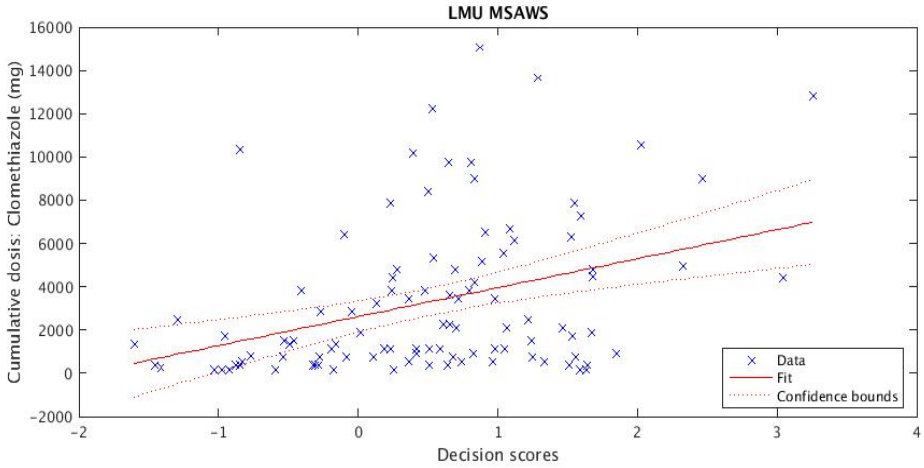

Figure 4.14. Post-hoc analysis for the LMU discovery MSAWS model

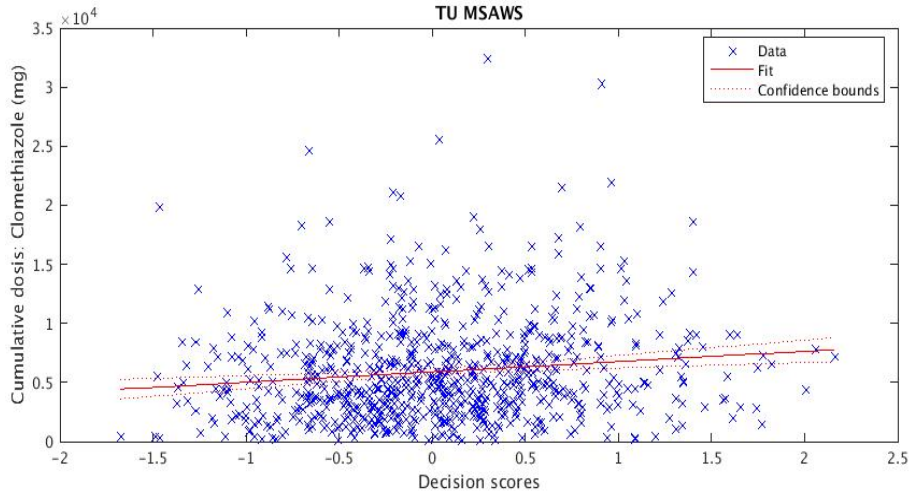

Figure 4.15. Post-hoc analysis for the TU discovery MSAWS model

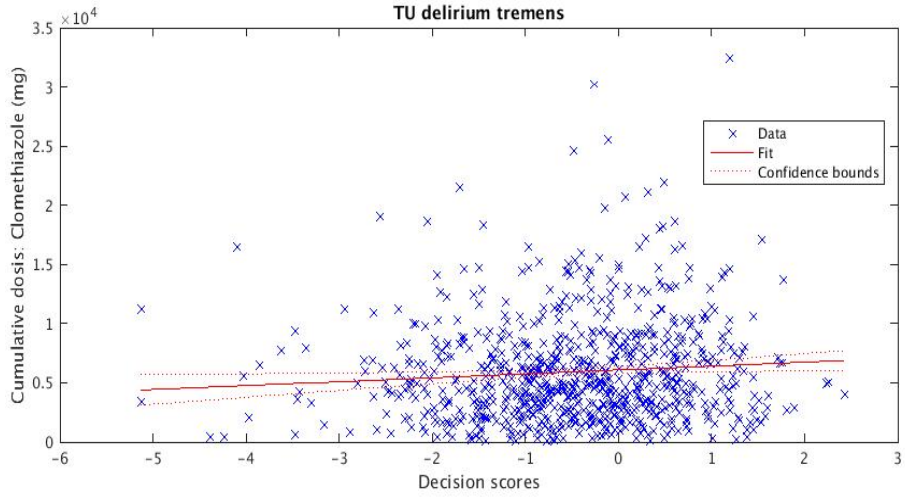

Figure 4.16. Post-hoc analysis for the TU discovery DT model

4.5 Supplementary analyses

4.5.1 Logistic regression classifiers

Table 4.4. Predictive performance of the logistic regression classifiers

	TN	TP	FP	FN	Sens	Spec	BAC
LMU discovery:							
MSAWS*	196	77	56	71	52	81.3	66.7
TU discovery:							
MSAWS	0	674	131	0	100	0	50
DT	760	0	0	45	0	100	50
WS	746	0	0	59	0	100	50
LMU validation:							
MSAWS*	65	200	84	47	81	43.6	62.3
OOCV	68	292	61	376	43.7	52.7	48.2
TU validation:							
MSAWS	0	668	129	0	100	0	50
OOCV	0	274	149	0	100	0	50

Notes: * Significance as defined via permutation testing: $p<0.01$. Sens, Spec, BAC, PPV, NPV, PSI and AUC in %. Abbreviations: True negatives (TN), true positives (TP), false positives (FP), false negatives (FN), sensitivity (Sens), specificity (Spec), balanced accuracy (BAC), out-of-sample-cross validation (OOCV).

4.5.1.1 LMU sample

In the LMU sample, cross-validated LR resulted in a non-significant BAC-loss of 3.9% for MSAWS prediction, reaching significance in permutation testing (Table 4.4). Sensitivity was 52% and specificity was 81.3% (Figure 4.19). Receiver operator rates are depicted in Figure 4.18 and further statistics are reported in Table 4.4. In addition to BrAC, platelet count, and positive urine-based benzodiazepine screening, that were selected by both SVM and LR models, the latter also chose high beta electroencephalographic activity, high glutamate dehydrogenase, and tingling sensations or numbness as long-term complications of alcohol use with feature selection probability of > 50% (Figure 4.18). Similar to the main SVM analysis the LR

validation model still reached significance with a BAC-loss of 4.4% but did not generalize to the TU cohort (Table 4.4).

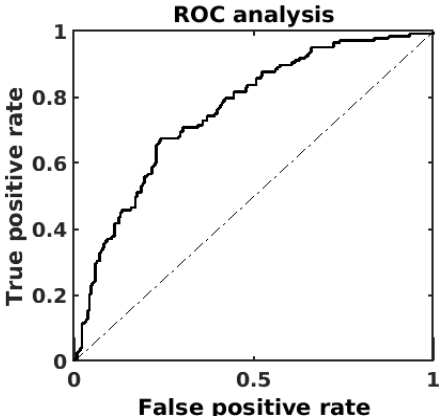

Figure 4.17. Area under the receiver operator curve for the LMU logistic regression MSAWS model

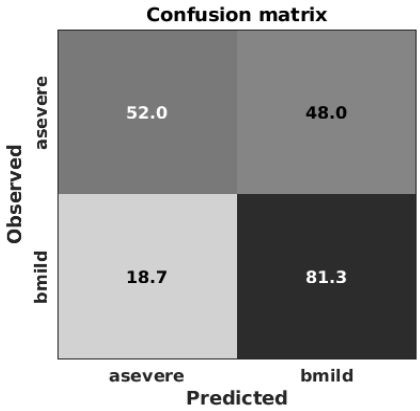

Figure 4.18. Confusion matrix for the LMU logistic regression MSAWS model

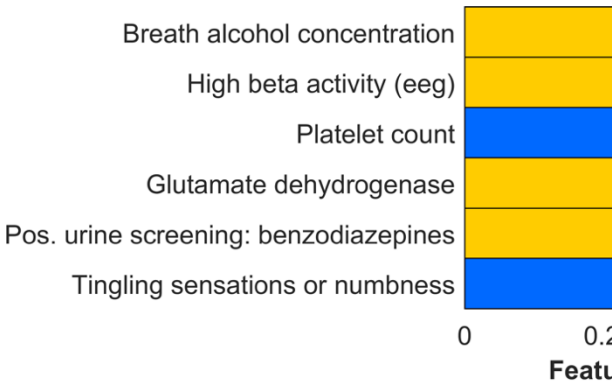

Figure 4.19. Feature selection probability for the LMU logistic regression MSAWS model

4.5.1.2 TU sample

In the TU discovery and validation datasets, the LR classifiers did not yield significant predictions, with a BAC at chance level for all outcomes (Table 4.4).

4.5.2 BAC variation

Median balanced accuracy across CV2 folds showed considerable standard deviations in both LMU and TU discovery models (Table 4.5), indicating that model performance was driven via a combination of diverse classifiers. Distribution differences between the point estimates of MSAWS, WS, and DT were detected in the TU discovery sample (post-hoc t-test between DT and WS as well as DT and MSAWS with p <0.01).

Table 4.5. Balanced accuracy across CV2 folds

	BAC in % (SD)
LMU discovery:	
MSAWS	68.5 (12.2)
TU discovery:	
MSAWS	56.6 (10.5)
DT	78.9 (12.9)*
WS	55.4 (12.7)

Notes: Since the BAC for each individual fold is calculated solely based on the models built in the respective training partition, the median BAC slightly differs from the overall BAC based on all ensemble models.*Distribution differences between DT and WS as well as DT and MSAWS for TU discovery significant (t-test, p <0.01). Abbreviations: Balanced accuracy (BAC), standard deviation (SD).

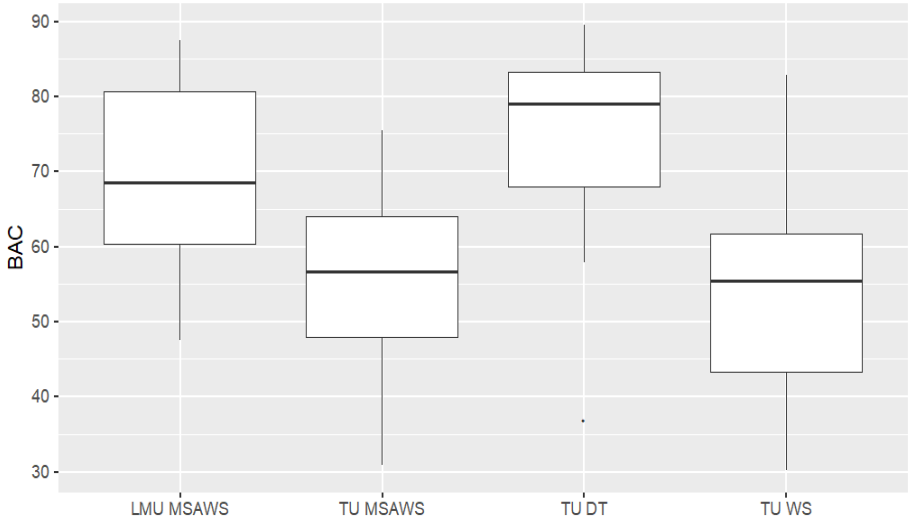

Figure 4.20. Performance variance across CV2 folds. Balanced accuracy (BAC) variance across the 20 folds in the outer cross validation loop (CV2) for all discovery models

4.5.3 Age and Gender effects on classification performance

Cases who were correctly classified as developing delirium tremens (DT) were significantly younger (mean age (SD)=43.6 (10.5)) compared to incorrectly classified cases (mean age (SD)=48.7(9.3)). Similar differences were not found for the MSAWS predictions at either site.

Table 4.6. Age and gender effects on prediction (true versus false classifications)

		t/ χ^2	p value
LMU discovery - MSAWS			
	Age	-0.65	0.51
	Sex	0.87	0.35
TU discovery - MSAWS			
	Age	0.68	0.5
	Sex	1.45	0.23
TU discovery – DT			
	Age	-6.59	<0.001*
	Sex	2.85	0.09
TU discovery – WS			
	Age	1.35	0.18
	Sex	6.43	0.01*

Notes: Students t-test (t), chi-square test (χ^2).

5 Discussion

Within the scope of this dissertation, I present evidence endorsing state-of-the-art machine learning as a promising framework to predict individual severity estimates of alcohol withdrawal syndrome in patients with alcohol dependence. A comprehensive battery of multidomain variables, routinely collected at patients' admission, was successfully used to predict cases of moderate to severe AWS as well as delirium tremens. Furthermore, important insights into how multivariate predictive patterns differ for distinct AWS severity estimates and patient populations were obtained, that could help to better understand the current state of AWS prediction research, potential methodological shortcomings, and promising future directions.

In the following sections I discuss model performance for each withdrawal outcome in the context of previous research efforts in the field. Following, I investigate similarities and differences in variable selection across different withdrawal outcomes and explore possible underlying clinical and neurobiological explanations. Considering the implemented external validation analysis, I subsequently highlight model generalizability as a main methodological imperative towards clinical translation. After discussing possible limitations of this dissertation, I propose future directions aimed at translatable risk prediction in AWS prediction research and summarize my results.

5.1 ML as a framework for accurate risk predictions

5.1.1 Predicting moderate to severe AWS

The MSAWS models were built to objectively support a clinically useful stratification task: Separating patients who will only develop mild withdrawal symptoms from patients who will develop moderate to severe symptoms, the latter potentially requiring more intense further monitoring and treatment (Mayo-Smith, 1997). For both study cohorts—a psychiatric sample showing mostly mild AWS trajectories (LMU) and a toxicology sample including more severe cases (TU)—separate models yielded significant predictions with gains in accuracy compared to pre-test outcome probabilities in the respective patient population. The predictive accuracy of the LMU model was comparable to results in other ML studies using similar analysis frameworks

for classification tasks (Chekroud et al., 2016; Koutsouleris et al., 2016). For example, Chekroud et al. predicted treatment outcomes (clinical remission) in patients with depression using patient-reported variables with accuracy of 64.6%, sensitivity of 62.8% and specificity of 66.2% (Chekroud et al., 2016). Koutsouleris et al. employed nested cross-validation to predict poor Global Assessment of Functioning (GAF) ratings in first-episode psychosis with balanced accuracy of 75.0% at week 4 and 73.8% at week 52 after treatment initiation (Koutsouleris et al., 2016).

In AWS research, two groups have developed tools for risk prediction using symptom assessment scales as severity estimates (Maldonado et al., 2014, 2015; Wetterling et al., 2006). In a psychiatric sample of 100 alcohol-dependent patients, Wetterling et al. predicted severe ratings on the AWS scale using a predefined set of 22 variables (LARS) that was later reduced to 11 variables (LARS11) based on a test of internal consistency (Wetterling et al., 2006). While the authors reported performance measures that markedly exceed the MSAWS models' results—a sensitivity of 95% and a specificity of 92.5% for the LARS11—these have to be scrutinized considering a major limitation: The scale's optimal cut-off value separating mild to moderate and severe withdrawal cases was defined post-hoc, but not tested in an independent sample. Indeed, the applied label "prospective study" (Wetterling et al., 2006) is misleading, since it does not refer to the scale's evaluation but to the structured gathering of the dataset that was later used for its development. A further clinical reason could possibly impair the scale's internal validity: Several input variables—tremor, sweating, tachycardia, sleep disturbances—coincide with the rating scale used as the severity outcome (Wetterling et al., 1997). Therefore, especially since no temporal information on individual withdrawal trajectories is provided, these variables may function simply as measurements of withdrawal severity limiting the scales' actual predictive power. Considering these limitations, it seems questionable wether the reported performance measures, retrieved in this rather small, unbalanced dataset, will translate to future patient cohorts. Moreover, the comparability of the LARS and the MSAWS models is impeded by a different clinical focus: While the LARS has been proposed as a tool to predict severe withdrawal cases, the MSAWS models were built including patients with moderate withdrawal symptoms. The latter cut-off was chosen since it potentially could indicate the need for pharmacological treatment, as has been proposed by the authors of the AWS scale (Wetterling et al., 1997).

Similarly, the PAWSS was developed to identify patients at risk of moderate to severe withdrawal in a cohort of general medicine and surgery patients (Maldonado et al., 2014). Analogous to the LARS the authors also constructed the scale based on prior literature findings, but prospectively tested it in an independent sample (Maldonado et al., 2015). The resulting performance measures a sensitivity of 93.1% and specificity of 99.5%, implying high accuracy, although two crucial omissions in the study's reporting impede the assessment of what was actually predicted: Firstly, the authors defined moderate to severe alcohol withdrawal as either a respective CIWA-Ar rating or a DSM-IV diagnosis, but did not differentiate between these definitions in their analysis. Since the DSM-IV diagnosis of alcohol withdrawal does not include a detailed severity assessment, it would be necessary to investigate the distribution of diagnoses more closely across the sample. However secondly, neither the distribution of CIWA-Ar ratings, the distribution of PAWSS ratings across the whole sample, nor the incidence rate of AWS cases has been reported. Moreover, the CIWA-Ar is not yet validated as an assessment scale in general medicine and surgical patients. Since AWS highly overlaps with symptoms seen in infectious or cardiovascular diseases and more detailed information on symptom trajectories across the whole study sample was not regarded, it is impossible to assess if the PAWSS identifies patients at risk of severe withdrawal, withdrawal in general, or patients with prior alcohol use that are treated for diseases associated with autonomous hyperactivity. In addition, the study has been conducted in a setting with a low AUD prevalence of only 1.7% compared to the general population (Grant et al., 2015). Hence their findings should not be generalized without further validation to treatment settings with high AUD probability, like the LMU and TU cohorts.

While these previous efforts rely on group-based analyses that may be useful for exploratory inference, they do not reversely ensure accurate prediction of individual subject outcomes, as is the goal of precision medicine (Insel & Cuthbert, 2015; Janssen et al., 2018; Walter et al., 2019). In contrast, the MSAWS models were developed via a framework that applies strict validation procedures, namely nested-cross validation, to generate estimates meaningful at the single-subject level (Dwyer, Falkai, et al., 2018). In this framework, performance is not driven by a single statistical model but a combination of classifiers varying in accuracy across the cross-validation folds and hypothetically converging on a more accurate estimate which better

represents sample diversity (Polikar, 2006). This expected variation across CV2 folds was indeed observed in a supplementary analysis (Figure 4.21). Furthermore, the overall performance estimates are generated through testing in subsamples independent of the model generation procedure. This approach is expected to yield results with lower accuracy that are nonetheless more likely to ensure generalizability to yet unseen patients (Varoquaux et al., 2017). While the main LMU SVM model outperformed the respective logistic regression model in accuracy, the difference was non-significant. Since the latter model was also learned in the same analysis pipeline, this further underscores the value of the CV-based ensemble framework, while, at least in the LMU sample, the choice of algorithm might have been less relevant.

Compared to the LMU MSAWS model, reaching a clear increase in prognostic certainty, the respective TU model resulted in significant, but clinically not useful balanced accuracy. Before I explore potential underlying reasons for this pronounced difference, mainly limitations of the available dataset and site-specific characteristics, I outline how the models predicting DT and WS performed in the same dataset.

5.1.2 Predicting Delirium tremens

Due to its potentially adverse outcomes, DT is often described as the most severe complication of AWS (Schuckit, 2014). Since it usually develops with latency (Mann et al., 2016; Victor & Adams, 1953), accurate prediction at admission could potentially improve its timely recognition and enable preemptive treatment. The respective TU model predicted cases of DT with a significant BAC of 75%. This result is interesting as it is comparable to a previously published, non-validated prediction model based on the same cohort, that predicted DT with a c-index of 0.81 (Eyer, Schuster, et al., 2011). Eyer et al. entered 27 sociodemographic and clinical variables into a stepwise logistic regression analysis to detect risk factors of DT, which were then combined into nomograms to allow for visually intuitive risk estimation (Eyer, Schuster, et al., 2011). Convincingly, such nomograms do not only allow for a binary group prediction, that being a patient is at risk of DT or not, but offer a range of certainty. Three risk factors contributed to Eyer's model: low potassium, low platelet count, and documented structural CNS lesions. While all three variables were included in the TU model, only platelet count and structural CNS lesions were frequently selected by the algorithms as predictors of DT. Despite the findings of a recent meta-analysis supporting low potassium as a risk factor of DT (Goodson et al., 2014) based on four studies

(Berggren et al., 2009; Eyer, Schuster, et al., 2011; Monte et al., 2009; Wetterling et al., 1994), this result questions if potassium levels can contribute to generalizable predictions. If not, it would lead to a profoundly lower predictive performance of the respective nomogram (Eyer, Schuster, et al., 2011). Two comparable studies have used stepwise logistic regression frameworks to identify risk factors of DT: Palmstierna et al. reported five binary variables associated with DT that they recommend to include in risk assessments (Palmstierna, 2001). Since they do not report any diagnostic performance measures, like sensitivity and specificity estimates (Palmstierna, 2001), the relevance and comparability of these results to the TU DT model cannot be assessed. Ferguson et al. propose a combination of two variables that indicate a risk of 54% for DT—just above chance (Ferguson et al., 1996).

The comparably strong validated accuracy of the TU DT model may have been achieved by the following two methodological decisions: Firstly, since only 6% of patients developed DT, I accounted for this imbalance by using a weighted SVM algorithm. SVM algorithms are well known for penalizing misclassifications of small sized classes in unbalanced datasets, which leads to decreased accuracy (He & Garcia, 2009; Yang et al., 2005). Such effects have also been observed for logistic regression frameworks (Maalouf et al., 2018). As expected by prevalence rates in historic AWS studies prior to benzodiazepine treatment (Victor & Adams, 1953), DT rates in prediction studies are only rarely higher than 20% (Ferguson et al., 1996). Nonetheless I am not aware of any studies considering potential effects of this imbalance or using respective methods to optimize predictive performance. Secondly, the SVM algorithms clearly outperformed a logistic regression analysis in the same cross-validated framework. Indeed, logistic regression did not result in predictions above chance. This discrepancy seems especially noteworthy since LR has been used in many DT prediction studies without further validation (Eyer, Schuster, et al., 2011; Ferguson et al., 1996; Palmstierna, 2001). Since SVMs are advantageous in highly collinear and noisy datasets (Cortes & Vapnik, 1995), this could implicate that DT cases are indeed difficult to demarcate from other withdrawal trajectories without comparable multivariate methods. Consequently, such methods seem promising in complementing clinical judgement that might not be sensitive to complex multivariate dependencies.

A supplementary finding, specific to the TU DT model, was that cases correctly classified as developing delirium tremens (DT) were significantly younger (mean age (SD)=43.6 (10.5)) compared to incorrectly classified cases (mean age (SD)=48.7(9.3)). As discussed below, prediction was mainly driven by three variables (low platelets, older age, history of brain lesions). Thus, while older age predicts DT, there could be an upper limit to its maximal contribution to predictive performance. Therefore, patients could potentially benefit from models separately built in different age groups in future studies. Still, the model's high sensitivity indicates a strong potential as a screening test to identify patients at risk of DT if further validated in independent samples.

5.1.3 Predicting Withdrawal seizures

Unfortunately, the TU model built to predict WS did not yield significant results. This is a surprising finding, since a respective model by Eyer et al. in the same dataset was reported to enable predictions up to 80% certainty with a c-index of 0.73 (Eyer, Schuster, et al., 2011). While this can be seen as further evidence that previous AWS research did not adequately consider means to ensure external validity, there is also a specific difference between both studies that could help explain this shortfall. The variable most strongly contributing to the seizure nomogram is the timepoint in hours after admission at which the highest score of the AWS scale was observed, the "apex of withdrawal severity" (Eyer, Schuster, et al., 2011). This discovery matches previous observations that withdrawal seizures often occur either before patients' admissions or in the first 48 hours of withdrawal treatment (Hillbom et al., 2003; Hughes, 2009). Therefore, there might be an early period of increased vulnerability to withdrawal seizures in the course of AWS. An early peak in withdrawal symptoms after admission could indicate that a patient has already passed this critical phase and thus has only a low probability of developing WS. Conversely, patients which have not yet reached the peak of withdrawal severity at admission would be at high risk of WS development. While this interesting aspect should be further explored to better understand different withdrawal trajectories, it also signifies a crucial limitation of the proposed nomogram: The climax of the AWS score introduces longitudinal information that is not yet known at a patient's admission into the prediction model. Since my aim was to build a predictive model that would potentially benefit clinical decisions shortly after admission, I did not include the variable into the TU analyses. Evidently, the expected loss of accuracy could not be compensated for by the remaining variable battery.

Other previous studies have mainly focused on single variables associated with withdrawal seizures which cannot be used to derive risk estimates meaningful to future patients (Berggren et al., 2009; Rathlev et al., 2000; Wojnar, Bizon, et al., 1999), with few exceptions: Morton et al. reported a discriminant function model of five clinical variables (previous number of withdrawal seizures, concurrent use of psychotropic substances, history of head injury, low sodium levels during the first 48 hours after admission, elevated pulse rates during the first 48 hours of admission) to show 96% efficiency in predicting cases without (n=28) and 84% efficiency in predicting cases with (n=12) alcohol withdrawal seizures in a small, retrospective dataset (Morton et al., 1994). Unfortunately, they do not further elaborate on the applied statistical methods and do not report any performance measures. Thus, it remains unclear what estimate the authors term as efficiency. The withdrawal seizures in the sample occurred with a mean 174 hours after admission, which lies surprisingly far off the usually expected timeframe of 24 to 48 hours (Hillbom et al., 2003). Two variables retrieved during the first 48 hours were included into the model (sodium level and pulse rate). Furthermore, the seizure patients were treated with markedly higher mean diazepine-equivalent doses (373 mg compared to 68 mg) suggesting considerably different withdrawal courses between groups. Whether these results would translate to other psychiatric populations remains questionable.

Hillemacher et al. used a random forest algorithm with bootstrap validation that reached an area-under-the-curve estimate of 0.824 based on six variables (Hillemacher et al., 2012). Besides number of daily smoked cigarettes, patient age, number of previous withdrawals, and BAC on admission, these included two laboratory variables—prolactin and homocysteine—that are usually not part of routine admission assessments. The analysis was conducted in 200 patients that were prospectively followed but did not develop any withdrawal seizures. To compensate for this, the authors used variables on index admission as input and previous withdrawal seizures as output variables. This approach, not relating to future events, can hardly be referred to as prediction. Questionably, the authors still stated their result as “widely transferable” findings (Hillemacher et al., 2012).

In summary, the applied ML framework led to accurate results in identifying patients at risk of MSAWS in a setting with predominantly mild withdrawal cases, and DT in more

severely ill withdrawal patients. Cases of WS could not be predicted based on the available variable battery. Furthermore, MSAWS was predicted with only low accuracy in the TU dataset. While some variation in predictive accuracy was hypothesized, the marked differences between outcomes were not expected.
Considering my first hypothesis, it can be concluded that accurate risk predictions based on the applied ML approaches are feasible but might not be suitable for all withdrawal outcomes. I thus further explore reasons for these differences in predictive performance in the following sections.

5.2 Exploring underlying variable patterns

Predictive performance largely depends on the available variables' ability to inform the posed learning problem in a given dataset (Hastie et al., 2009). In order to gain further insight into the performance differences between models, I therefore explore which variables were frequently selected by each model. Previous studies have used the feature selection probability, the percentage of times that each variable was selected by the ensemble classifiers across all CV training partitions, as a proxy for variable importance (Dwyer, Cabral, et al., 2018; Koutsouleris et al., 2016). Variables that are selected by >50% of the classifiers are then interpreted to meaningfully contribute to the models overall decision (Dwyer, Cabral, et al., 2018; Koutsouleris et al., 2016). Since these variables are learned in a multivariate framework, they do not signify and cannot be directly compared to classical univariate analysis (Gaonkar & Davatzikos, 2013). While this impedes direct quantitative comparison to previous research in the field, I narratively investigate how single variables match reported risk factors and hypothesized pathophysiological underpinnings of AWS nonetheless.

Several variables frequently selected by the MSAWS models, have been previously—but not consistently—described as risk factors of severe withdrawal development:
Vinson et al. reported a high breath alcohol concentration as predictor of severe clinical AWS ratings and high medication doses in a sample of both psychiatric and medical inpatients (Vinson & Menezes, 1991). Although BrAC is routinely assessed at admission in clinical settings, this finding has to my knowledge not been replicated in further studies. Palmstierna et al. found high blood alcohol concentrations predictive of delirium tremens, but only when other withdrawal symptoms were present

(Palmstierna, 2001). Such development of withdrawal symptoms during alcohol intoxication might indicate higher levels of tolerance seen in heavy drinkers, which could hypothetically then increase severe withdrawal risk. However, markers of high-dosage consumption or long-time AUD have not shown predictive utility in previous studies (Goodson et al., 2014; Kraemer et al., 2003; Lechtenberg & Worner, 1992; Wetterling, 2001) and similar variables in the LMU variable battery (age of onset alcohol dependency, approximated daily alcohol intake) were not frequently selected by the model's algorithms. Furthermore, Palmstierna et al.'s findings could not be confirmed by other multivariate analysis (Eyer, Schuster, et al., 2011), meta-analysis (Goodson et al., 2014), or my TU DT model. Interestingly, high alcohol concentration was selected only by the LMU MSAWS but not the respective TU model. Since patients can be expected to stop alcohol consumption at varying timepoints before admission, some mild AWS cases might have passed the peak of withdrawal severity before admission and therefore shown low prior probability of high AWS scoring. Such cases were likely excluded in the TU dataset which did not include any patients without need for withdrawal medication. Furthermore, while mild symptoms of withdrawal (e.g. increased autonomic activity and agitation) typically develop only a few hours after reduced alcohol consumption, DT manifests with latency (Victor & Adams, 1953). Our finding therefore potentially indicates that alcohol levels differentiate only between the earlier, milder manifestations of AWS.

Two further previously described variables were only available in the TU dataset: Firstly, a history of DT as predictor of severe clinical AWS ratings has been previously described by Kramer et al. (Kraemer et al., 2003). Such previous withdrawal experiences have been hypothesized to induce neuronal changes leading to low-threshold neuronal excitability which might then increase the risk for subsequent, more severe withdrawal episodes (Ballenger & Post, 1978). This much cited kindling hypothesis originated in electrophysiological studies and animal research (Becker, 1998; Gonzalez et al., 2001). As outlined in the introduction, successive clinical studies have shown inconsistent and conflicting results regarding the predictive value of different variables as indicators of previous withdrawal experience. The aforementioned meta-analysis found a previous DT predictive only of subsequent DT episodes (Goodson et al., 2014), which could not be confirmed by the TU DT model. Other indicators of recurrence available in my dataset, such as previous withdrawal

episodes, previous withdrawal treatments, number of episodes with withdrawal symptoms, repeated alcohol intake to stop withdrawal symptoms, and previous withdrawal seizures showed only low feature selection probability across all models. In summary, these negative findings are indicative that the kindling hypothesis might not be directly transferable to human withdrawal experience. More specifically, since a history of DT was only selected by the MSAWS model with low accuracy, even if such neuronal adaptions do occur, related anamnestic variables might offer only limited predictive value in specific populations. Such idiosyncratic signatures could hardly be expected to generalize across patient cohorts or treatment outcomes.

Secondly, an increased heart rate at admission has been previously described only as a dichotomous predictor of DT with varying cut-off frequencies (J. H. Lee et al., 2005; Palmstierna, 2001). Tachycardia is a typical autonomous nervous system reaction in alcohol withdrawal and might warrant intensive monitoring and treatment in severe cases (Khan et al., 2008). It is included as an item in all established severity assessment scales (Sullivan et al., 1989; Wetterling et al., 1997). Patients who are admitted with increased heart rates are therefore potentially already experiencing acute withdrawal. From an explorative perspective this highlights a major limitation of most clinical withdrawal studies: Without exact information about the timeline of withdrawal, meaning the development of potential withdrawal symptoms after termination or reduction of subject-specific alcohol consumption, as well as reliable information on drinking history, causal inferences about withdrawal trajectories are impeded. Clinically, such variance in baseline withdrawal manifestations can be expected, since patients are often admitted unplanned or stop alcohol consumption at different timepoints prior to planned admission (National Institute for Health and Care Excellence (NICE), 2011). From a prediction perspective a more relevant question is, how much the identification of adverse outcomes in the later disease course can benefit from the diagnosis of an increased heart rate. Here, the low selection frequency of the variable by the DT model indicates that its predictive value is limited to early withdrawal manifestations.

Besides AWS, tachycardia is a frequently occurring, general symptom seen in various conditions like cardiovascular and infectious diseases (Brugada et al., 2019). An increased heart rate at admission could therefore also indicate an underlying medical condition that might predispose for severe withdrawal development. Correlations between somatic diseases and adverse AWS outcomes, like DT development and

longer hospital stay, have been proposed by several authors (Ferguson et al., 1996; Wojnar, Bizoń, et al., 1999). While the TU dataset only included established liver cirrhosis as a disease variable, which did not show predictive value across models, the LMU dataset contained more extensive patient-reported information about the presence of various comorbid conditions like neurological, cardiovascular, hepatic, thyroid, and asthmatic disease. The sole selection of a history of blood pressure abnormalities by the LMU MSAWS model narrows the focus on cardiovascular disease as a relevant comorbidity, which could especially agree with the TU heart rate selection. Due to both variables not being available across datasets, the inclusion of more extensive information on cardiovascular status promises an exploration of a possible link between a potential predisposition to more pronounced autonomous reactions in AWS patients with underlying cardiovascular disease in future studies.

A history of blood pressure abnormalities is one of three newly identified predictors of severe withdrawal, that have to my knowledge not been reported in preceding research: A positive urine-based benzodiazepine screening was frequently selected by the LMU MSAWS model, while not available in the TU dataset. Rates of concomitant benzodiazepine addiction in patients with AUD are high (Compton et al., 2007). Furthermore, both substances show largely overlapping withdrawal syndroms (Busto et al., 1986). Without including any patients that received a formal ICD-10 diagnosis of benzodiazepine addiction in the LMU cohort, a positive urine screening could indicate two separate situations: firstly, co-dependency of benzodiazepines might remain undetected by the primary care team and said patients then develop both benzodiazepine and alcohol withdrawal simultaneously, with possibly more severe symptom development. Secondly, though not in line with usual admission routine, some patients could have received first benzodiazepine doses even before laboratory testing due to severe symptom presentation at admission. In conclusion, a positive urine-based benzodiazepine screening might not be specific of AWS but reflects common clinical circumstances where accurate information is not always readily attainable. Therefore, it should be included as a variable in a prediction tool seeking clinical implementation.

Another novel finding is the selection of years of schooling by the LMU MSAWS model. While this variable should be interpreted cautiously as no further detailed information on the patients' education was available, it could lead to a novel hypothesis that could

be explored in further longitudinal research: Epidemiological studies in diverse conditions like neurodegenerative disease (Y. Stern, 2013), psychotic disorders (Barnett et al., 2006), and brain trauma (Kesler et al., 2003) have suggested that educational and occupational attainment influences the cognitive reserve of patients, thus leading to decreased susceptibility to age- and disease-related brain changes and attenuated illness courses. In AWS such adaptions could be hypothesized to guard against pronounced symptom development. Importantly, the variable highlights that the comprehensive, unbiased inclusion of developmental and other sociodemographic information might be leading to the identification of variables with unexpected predictive value. Sociodemographic information beyond age, gender, and alcohol consumption history is scarcely included in AWS prediction studies (Goodson et al., 2014). In contrast to prediction approaches strongly relying on prior literature findings (Maldonado et al., 2014; Wetterling et al., 2006), the three newly identified MSAWS predictor variables encourage and justify the use of techniques such as the applied wrapper-based feature selection in high-dimensional datasets. Indeed, these results are in line with work promoting theory agnostic (Huys et al., 2016), data-driven models for knowledge generation (Breiman, 2001).

Variable importance analyses of the TU DT model showed that only three variables were frequently selected by the algorithms. Except low platelet count, they differ from the selections of the MSAWS models: The selection of CNS lesions suggests a morphologically-based vulnerability for DT. This was first shown in the aforementioned study in the same dataset (Eyer, Schuster, et al., 2011), but has not yet been replicated. In contrast to Eyer et al.'s study the same variable did not contribute to significant predictions of WS. These inconsistent results can be attributed to the cross-validation setup, which indicates that CNS lesions contribute to generalizable predictions specifically for the DT outcome. The variable's predictive value potentially corresponds to the cognitive reserve hypothesis, since education might forestall cognitive deficits that can be observed in patients with cerebral white matter lesions (Dufouil et al., 2003).

Only a low platelet count was persistently utilized as a poor-outcome predictor by all significant models across both treatment outcomes—MSAWS and DT—as well as treatment sites. Thrombocytopenia in patients with AUD might result from various

pathophysiological mechanisms: While alcohol is assumed to have direct toxic effects on the bone marrow, resulting in decreased cell generation (Berggren et al., 2009; Latvala et al., 2004), and the induction of platelet apoptosis (Zhao et al., 2017), low platelets might also be attributed to chronic liver disease (O. Mitchell et al., 2016). Occurrence of liver cirrhosis has been more frequently observed in patients with preferred continuous alcohol consumption opposed to binge drinking patterns (Barrio et al., 2004). Diagnosis of liver cirrhosis as well as laboratory liver enzyme measurements were included in LMU and TU datasets but did not inform the models' predictions. As suggested by prior research (Barrio et al., 2004; Goodson et al., 2014) low platelet count could therefore indeed indicate mainly intermittent alcohol consumption.

In line with meta-analytic results (Goodson et al., 2014), the finding that MSAWS and DT models only share platelet count as a poor-outcome predictor directly indicates that both diagnosis categories reflect specific, underlying pathomechanisms, with a history of cerebral lesions as a possible specific differentiator. This would give reason for treating both outcomes as separate entities in further prediction studies. Since the WS model did not yield significant results, I won't speculate about possible predictive variables.

Lastly, with variables being available across multiple domains, it may be helpful to also highlight information not benefitting predictions: Besides the limited evidence on prior withdrawal experience and somatic comorbidity, no evidence of increased AWS risk in patients with a positive family history was found, which could have indicated a genetic risk for more severe AWS development. Furthermore, EEG measurements shortly after admission did not inform MSAWS prediction. Importantly, this does not reduce its value in differential diagnosis of seizure etiologies in the later AWS course (Hillbom et al., 2003).

5.3 Site specific differences and external validation

A main criticism on previous AWS research and the most likely reason for the absence of viable, evidence-based clinical tools beyond symptom assessment scales has been the lack of attempts to validate results across treatment settings (Goodson et al., 2014; Saitz, 2018). The need for such validation is emphasized by the finding that the MSAWS models did not generalize across LMU and TU cohorts. To my knowledge this has indeed been the first attempt for an external validation analysis in AWS research, hence an exploration of potential underlying reasons for the negative results could possibly benefit future research.

As expected, reducing the variable battery to the 9 shared variables led to a decrease in classification performance in both datasets. Surprisingly, while out of the variables selected most frequently by the respective discovery model only platelet count and breath alcohol concentration could be included, the reduced LMU model showed a modest BAC loss of 3.1% that still reached statistical significance. Hence, overall prediction performance was likely driven by a multivariate pattern with likely complex relationships between variables. While such complex relationships might limit the interpretability of single contributing variables, a model not reliant on single variables could be more easily adjusted to specific treatment settings where the attainment of certain variables may not be feasible. Unfortunately, the model's predictions resulted in a balanced accuracy below chance when applied to the TU cohort in the out-of-sample cross-validation analysis. In the respective TU model only CNS lesions could not be included out of the discovery model's most frequently selected variables. Still, based on the already low performance of the discovery model, the BAC loss of 2.9% led to a performance not reaching statistical significance in permutation testing. Expectedly, the out-of-sample testing did not yield accurate predictions as well.

Several reasons for the lack of generalizability across treatment sites are conceivable: Reflecting a limitation of the retrospective data gathering process, several important variables that might have contributed to predictive performance (e.g. years in school, history of DT, heart rate at admission, urine-based benzodiazepines screening) were not available for the external validation analysis. Furthermore, univariate comparison of the underlying variables between both datasets revealed significant differences for most variables as well as the predicted outcome. The TU dataset consisted of

markedly more moderate to severe withdrawal cases (see Figure 4.1). Partly, this distinction could be explained by differences in patient referral due to the closer linkage of the TU's toxicology unit to the general medicine department, which provides access to better monitoring and care of withdrawal cases suffering from more intense medical comorbidity, including acute somatic complications of harmful alcohol use. Correspondingly, the univariate baseline differences fit the notion of the TU sample as a higher severity cohort, including significantly older patients with higher mean gamma-GT and potassium measures, higher alcohol concentration at admission, and more frequent history of withdrawal seizures. The finding that mean platelet count was significantly lower in the LMU sample while low platelet count was selected by all models as a bad outcome predictor, can be attributed to its wider distribution in the TU sample (reflected in a higher standard deviation).

Based on these baseline characteristics patients would then have entered procedures at two separate hospitals, with potentially different approaches to clinical care. A level of standardization between cohorts is certainly provided through the implementation of symptom-triggered medication regimes (Mayo-Smith, 1997) with well-established pharmacological agents (Amato et al., 2011; Eyer, Schuster, et al., 2011) in both clinics—excluding more experimental treatments sometimes used in intensive-care and trauma settings (Vanderweide et al., 2016; Wong et al., 2015). However, information on other context-specific factors like non-pharmacological treatment protocols, AWS specific training of attending staff, and treatment facilities were not available. While both treatment settings are located in the same city, decreasing the probability of major cultural and sociodemographic divergences, the influence of treatment settings has only rarely been studied in previous research (Naranjo et al., 1983; Whitfield et al., 1978) and might have modulated withdrawal trajectories after admission.

In summary, the lack of generalizability of MSAWS models offers direct evidence that AWS prediction models built in single-clinic cohorts cannot be easily transferred to other treatment settings without prior validation. Importantly, the cross-over design of this study adds an important incentive for future AWS research: Prior attempts to find more robust predictors of AWS severity focused on summarizing results across diverse settings and outcomes via meta-analysis (Goodson et al., 2014) or systematic review (Maldonado et al., 2014) to derive generally valid assertions. By contrast, the marked

differences of the MSAWS models in both predictive performance as well as variable importance across treatment sites could be seen as evidence that a more feasible road towards clinically useful prediction tools would be the development of models specifically adjusted to circumscribed populations. The external validation's negative results therefore support the appraisal of recent treatment guidelines that do not recommend any AWS prediction tools published to date (Department of Defense, 2017; Mann et al., 2016; National Institute for Health and Care Excellence (NICE), 2010, 2011), opposed to recently published, high-impact endorsements of non-validated predictive models (A. M. Wood et al., 2018).

5.4 Clinical utility

Even a robust computational model, reaching both generalizable as well as highly accurate predictions, might not necessarily be clinical useful unless its predictions are connected to modifiable treatment decisions that could then improve patient outcomes. This requirement is especially relevant considering recent advances in ML algorithm development, like deep learning (Esteva et al., 2019), leading to a surge in studies reporting highly accurate results but often do not benefit clinical care (Wiens et al., 2019). To clinically validate the prediction models, I examined how their classification decisions relate to the pharmacological treatment received during withdrawal. For the MSAWS models, a regression analysis revealed significant linear relationships between the decision score (a measure indicating the certainty of a model's predictive appraisal) and the cumulative clomethiazole demand during AWS treatment. In both samples symptom-triggered pharmacotherapy was provided (Mayo-Smith, 1997), using the AWS scale as measure of AWS severity to guide treatment. Hence, the relationship between the classifiers' decision and provided treatment doses indicate that the MSAWS models predict clinically meaningful outcomes. Despite its post-hoc nature, this result suggests that assessments based on ML predictions could potentially guide treatment decision currently relying on clinical judgement (National Institute for Health and Care Excellence (NICE), 2010, 2011).

For the DT model, a comparable post-hoc association between its decision scores and clomethiazole demand was not found. This negative result can be attributed to a ceiling effect due to generally high dose clomethiazole application in the cohort. Considering

a recent meta-analysis on symptom-triggered AWS treatment (Holleck et al., 2019) highlighting a shortage of controlled studies in high severity AWS cases as well as studies indicating limits in efficacy of both symptom-triggered therapy (J. A. Lee et al., 2019) as well as benzodiazepine treatment (Vanderweide et al., 2016) in ICU patients, this finding agrees with the notion that DT patients should be considered to be a subgroup with specific challenges to clinical care (Schuckit, 2014). While innovations in AWS pharmacotherapy are thus warranted and clinical studies on various pharmacologically agents are regularly published (Cooney et al., 2019; Eyer, Schreckenberg, et al., 2011; Gillman et al., 2007; Leone et al., 2010; Wong et al., 2015), clinical care in general seems to have much improved since the first systematic reports on AWS patients (Victor & Adams, 1953): In both LMU and TU cohorts—totaling 1194 AWS patients including 822 cases of moderate to severe withdrawal, 45 DT cases, and 71 withdrawal seizures—no fatalities have been reported. Low mortality is certainly a primary goal of AWS treatment but should not let other treatment outcomes, like patient expectations regarding comfort and autonomy, unrecognized. Psychological impact and mental health outcomes following delirium, especially symptoms of post-traumatic stress, have recently received increased attention (Bolton et al., 2019; G. O'Malley et al., 2008), but have not been studied in AWS patients. Furthermore, adverse side effects of generously administered sedative medication might outweigh its benefits in cohorts with predominantly mild cases (Amato et al., 2010, 2011; Maldonado et al., 2014). These issues imply strong imperatives for further optimizing clinical care for AWS patients. Besides their promising predictive accuracies, the models built within the scope of this study point out two further clinically relevant aspects that could help to specify where predictive models could be integrated into such endeavors: Firstly, the still significant BAC of the reduced 10 variable MSAWS model in the LMU cohort indicates that few easily attainable variables could be sufficient for accurate outcome predictions in similar populations, thus considerably facilitating the implementation into clinical practice. However, without further validation of the reduced models, this result should be evaluated with caution. Secondly, the high sensitivity of the TU DT model signifies potential as a screening tool, which could be used to stratify patients into low and high-risk groups and then allocate costly clinical resources like hospital beds and intensified monitoring capabilities based on predictable individual needs. Such proceedings could also enable evidence-based

referral to safe out-patient treatment (National Institute for Health and Care Excellence (NICE), 2010).

In summary, the ML models' association with pharmacological treatment decision, their predictive power based on a few clinical variables, and their useful screening capabilities support the clinical utility of further validated models derived from ML methodology. After disclosing several limitations of this project, I discuss how these results could function as incentives for future research in the field more specifically.

5.5 Limitations

This project has certain inherent limitations considering its dataset, analysis, and results. First, being based on retrospective data, it was not possible to exclude systematic biases resulting from inaccurate documentation. Specifically, since AWS-scores had to be retrospectively derived, individual estimates potentially do not agree with nuanced in situ appraisal. Moreover, the AWS scale includes breath rate at admission as a time-consuming, non-automated clinical measurement (Wetterling et al., 1997), which in clinical experience is often evaluated diligently only in severe cases. Such inconsistencies could potentially have let to underrated AWS scores, especially in the less severe LMU sample.

Secondly, since the available data was limited to psychiatric and internal-medicine in-patients, latter treated in a specialized detoxification unit, other cohorts with considerable AWS incidence like surgical wards, intensive care units, out-patient clinics or correctional facilities (Fiscella et al., 2004; Lukan et al., 2002; Maldonado et al., 2015; Whitfield et al., 1978) were not considered in this project. As the external validation analysis shows, such settings might also require different predictive models adjusted to distinct patient characteristics, clinical procedures and other setting-specific features. Since the LMU and TU datasets were both acquired in German hospitals, regional distinctions and differing cultural norms, potentially influencing treatment approaches, were not regarded and could further prevent generalizability across countries.

Thirdly, the range of possible predictors as well as further meaningful AWS treatment outcomes was restricted by the content of the utilized patient charts. Regarding input variables, more nuanced information on prior addictive behavior like number of former

withdrawal episodes or drinking patterns would have been desirable. While the inclusion of standardized questionnaires and structured interviews to assess potential psychiatric co-morbidities would have been interesting from an explanatory viewpoint, such assessments are often time consuming (First et al., 2015) and thus not suitable for a timely evaluation at admission. This is also true for MR-based imaging techniques like spectroscopic measurements of brain metabolites, which have shown promising results towards a better understanding of the neurobiological underpinnings of AWS (Hermann et al., 2012) but are not available outside of research facilities. Regarding output data, long-term outcomes like relapse rates and adherence to subsequent addiction treatment would have been informative, both to further validate the LMU and TU models as well as being possible predictive targets by itself.

Fourthly, due to restricted overlap between available variables across the treatment sites, an external validation of the full predictor set of the main discovery models was not possible. The failed validation attempt in the reduced variable sets certainly is the major limitation in regard to clinical translation, but also has been informative in better understanding shortfalls of previous research.

Finally, while the MSAWS models' association with pharmacological treatment signify a degree of clinical validation, a direct comparison to clinical judgement (e.g. the appraisal of the attending psychiatrist regarding outcome prediction) would have been desirable to further judge their potential to complement decision-making. However, 31% of patients in this study developed only mild withdrawal symptoms, although all patients were treated as in-patients. This observation could be seen as implicit evidence for the need to further improve such evaluations.

5.6 Future directions

Despite these limitations the results of this project indicate several promising avenues for future research aiming at optimizing clinical care for AWS patients.

The capability of ML to generate accurate risk predictions in large datasets by utilizing a broad spectrum of potential predictors across diverse domains seems especially suitable for these purposes. Due to open-source software packages like NeuroMiner (https://www.pronia.eu/neurominer), mlr3 (https://github.com/mlr-org/mlr3) or tensorflow (https://www.tensorflow.org), the reproducible implementation of even complex ML tools no longer demands extensive computational or engineering

knowledge, but can be applied by researchers across diverse scientific disciplines. Even so, a main prerequisite is the availability of large datasets to enable both model generation and external validation. Fortunately, large AWS datasets have been repeatedly compiled across medical specialties and different geographical regions (Lukan et al., 2002; Schuckit et al., 1995; Wojnar, Bizoń, et al., 1999) and could be shared via open-source databases. This would also allow the application of deep-learning algorithms that can achieve highly accurate results but require big data (Beam & Kohane, 2018; Esteva et al., 2019). Nevertheless, overreliance on previously gathered datasets would not solve issues like heterogenous definitions of AWS outcomes, non-standardized assessments, or diverse treatment protocols across cohorts (Goodson et al., 2014). Indeed, assembling representative, high-quality data is a substantial part of ML model development (Rajkomar et al., 2019) and essential in forestalling prediction biases, like the amplification of already existing socioeconomic disparities in the health-care system (Gianfrancesco et al., 2018).

Therefore, future collaborations between multiple treatment centers starting at the project development phase are highly warranted to address previous limitations of single-center studies as well as incorporating modern demands for reproducible science (Ioannidis, 2005; Wiens et al., 2019). Considering AWS, this procedure could help to harmonize datasets across treatment settings via standardized gathering of patient information and treatment endpoints, including meaningful goals developed in collaboration with stakeholders, like clinical experts, public-health administrators, and patient representatives. Models built on such datasets could be prospectively evaluated against clinicians' ratings in multiple settings and used for randomized-controlled trials that test their ability to optimize relevant clinical decision, for instance choice of treatment setting or pharmacological treatment, henceforth establishing a potential for clinical implementation. Finally, as has been recently shown by a large clustering study on psychosis patients (Dwyer et al., 2020), data-driven research based on multi-center collaboration could function as a promising approach to characterize individual disease trajectories, potentially leading to more targeted treatment approaches.

5.7 Conclusion

The analyses conducted within the scope of this research contribute to moving alcohol withdrawal research towards unbiased, data-driven personalized medicine approaches that include robust tests of generalizability.

The primary aim was to identify patients at risk of severe AWS, defined by three clinically meaningful outcomes, via a machine learning framework incorporating tests of internal validity. Based on easily attainable clinical, sociodemographic and laboratory variables available at patients' admission, this approach generated significant prediction models for separating cases of mild and moderate to severe AWS in a cohort of predominantly mild withdrawal cases as well as for identifying patients at risk of DT in a more severely ill sample. Other than hypothesized, it was not successful in generating cross-validated predictions for the third outcome, withdrawal seizures. In the case of the MSAWS models the prediction results could be connected to pharmacological treatment, adding to its clinical validity and potential to inform clinical decisions.

Considering the secondary aims, predictive patterns highly varied across withdrawal outcomes with a low platelet count as the only variable contributing to adverse outcome predictions across models. This expected variation strengthens the notion of DT as a specific complication of AWS with unique neurobiological underpinnings—highlighting structural CNS lesions as a potential morphological risk-factor. Apart from that, several new predictive variables were found (self-reported previous blood pressure abnormalities, positive urine-based benzodiazepine screening, and years of schooling), emphasizing the value of data-driven, hypothesis-free prediction approaches.

Unexpectedly, in an out-of-sample external validation analysis, separate models using only variables available in both datasets did not achieve significant predictions across samples. While this result impedes the translation of the newly developed ML models into clinical use, it also points to important limitations of previous research and offers strong incentives guiding future research towards multi-center collaboration.

Overall, this dissertation provides evidence promoting ensemble learning within a nested cross-validation setup as a potential approach to enable reliable risk prediction for AWS severity outcomes in future prospective studies. Such models seem promising in enabling further data-driven research on important treatment decisions, like pharmacological treatment and choice of setting, that, at the moment, exclusively rely on resource-intensive monitoring and clinical judgement.

6 References

Abràmoff, M. D., Lavin, P. T., Birch, M., Shah, N., & Folk, J. C. (2018). Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *Npj Digital Medicine*, *1*(1), 1–8. https://doi.org/10.1038/s41746-018-0040-6

Amato, L., Minozzi, S., & Davoli, M. (2011). Efficacy and safety of pharmacological interventions for the treatment of the Alcohol Withdrawal Syndrome. *The Cochrane Database of Systematic Reviews*, *6*, CD008537. https://doi.org/10.1002/14651858.CD008537.pub2

Amato, L., Minozzi, S., Vecchi, S., & Davoli, M. (2010). Benzodiazepines for alcohol withdrawal. *Cochrane Database of Systematic Reviews*. https://doi.org/10.1002/14651858.CD005063.pub3

Arbabshirani, M. R., Plis, S., Sui, J., & Calhoun, V. D. (2017). Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. *NeuroImage*, *145*, 137–165. https://doi.org/10.1016/j.neuroimage.2016.02.079

Ballenger, J. C., & Post, R. M. (1978). Kindling as a model for alcohol withdrawal syndromes. *British Journal of Psychiatry*, *133*(7), 1–14. https://doi.org/10.1192/bjp.133.1.1

Barnett, J. H., Salmond, C. H., Jones, P. B., & Sahakian, B. J. (2006). Cognitive reserve in neuropsychiatry. In *Psychological Medicine* (Vol. 36, Issue 8, pp. 1053–1064). https://doi.org/10.1017/S0033291706007501

Barrio, E., Tomé, S., Rodríguez, I., Gude, F., Sánchez-Leira, J., Pérez-Becerra, E., & González-Quintela, A. (2004). Liver Disease in Heavy Drinkers with and without Alcohol Withdrawal Syndrome. *Alcoholism: Clinical and Experimental Research*, *28*(1), 131–136. https://doi.org/10.1097/01.ALC.0000106301.39746.EB

Bartu, A., & Saunders, W. (1994). Domiciliary detoxification: a cost effective alternative to inpatient treatment. *The Australian Journal of Advanced Nursing : A Quarterly Publication of the Royal Australian Nursing Federation*, *11*(4), 12–18. http://www.ncbi.nlm.nih.gov/pubmed/7980884

Beam, A. L., & Kohane, I. S. (2018). Big data and machine learning in health care. In *JAMA - Journal of the American Medical Association* (Vol. 319, Issue 13, pp. 1317–1318). American Medical Association. https://doi.org/10.1001/jama.2017.18391

Becker, H. C. (1998). Kindling in alcohol withdrawal. *Alcohol Health and Research World*, *22*(1), 25–33.

Bellman, R. (1957). *Dynamic Programming: A Rand Corporation Research Study*. Princeton Unversity Press. https://b-ok.cc/book/445606/ae145e

Benson, G., McCallum, J., & Roberts, N. J. (2019). Risk Factors for Severe Alcohol Withdrawal Syndrome in an Acute Hospital Population. *Journal of Addictions Nursing*, *30*(3), 159–168. https://doi.org/10.1097/JAN.0000000000000287

Berggren, U., Fahlke, C., Berglund, K. J., Blennow, K., Zetterberg, H., & Balldin, J. (2009). Thrombocytopenia in early alcohol withdrawal is associated with development of delirium tremens or seizures. *Alcohol and Alcoholism*, *44*(4), 382–386. https://doi.org/10.1093/alcalc/agp012

Bleich, S., Bayerlein, K., Hillemacher, T., Degner, D., Kornhuber, J., & Frieling, H. (2006). An assessment of the potential value of elevated homocysteine in predicting alcohol-withdrawal seizures. *Epilepsia*, *47*(5), 934–938. https://doi.org/10.1111/j.1528-1167.2006.00560.x

Blondell, R. D., Powell, G. E., Dodds, H. N., Looney, S. W., & Lukan, J. K. (2004). Admission characteristics of trauma patients in whom delirium develops.

American Journal of Surgery, *187*(3), 332–337. https://doi.org/10.1016/j.amjsurg.2003.12.027
Bolton, C., Thilges, S., Lane, C., Lowe, J., & Mumby, P. (2019). Post-traumatic Stress Disorder Following Acute Delirium. *Journal of Clinical Psychology in Medical Settings*, 1–9. https://doi.org/10.1007/s10880-019-09689-1
Booth, B. M., & Blow, F. C. (1993). The kindling hypothesis: further evidence from a U.S. national study of alcoholic men. *Alcohol and Alcoholism (Oxford, Oxfordshire)*, *28*(5), 593–598. http://www.ncbi.nlm.nih.gov/pubmed/8274184
Bråthen, G., Bjerve, K. S., Brodtkorb, E., & Bovim, G. (2000). Validity of carbohydrate deficient transferrin and other markers as diagnostic aids in the detection of alcohol related seizures. *Journal of Neurology Neurosurgery and Psychiatry*, *68*(3), 342–348. https://doi.org/10.1136/jnnp.68.3.342
Breiman, L. (2001). Statistical modeling: The two cultures. *Statistical Science*, *16*(3), 199–215. https://doi.org/10.1214/ss/1009213726
Brodersen, K. H., Ong, C. S., Stephan, K. E., & Buhmann, J. M. (2010). The balanced accuracy and its posterior distribution. *Proceedings - International Conference on Pattern Recognition*. https://doi.org/10.1109/ICPR.2010.764
Brugada, J., Katritsis, D. G., Arbelo, E., Arribas, F., Bax, J. J., Blomström-Lundqvist, C., Calkins, H., Corrado, D., Deftereos, S. G., Diller, G.-P., Gomez-Doblas, J. J., Gorenek, B., Grace, A., Ho, S. Y., Kaski, J.-C., Kuck, K.-H., Lambiase, P. D., Sacher, F., Sarquella-Brugada, G., … Zaza, A. (2019). 2019 ESC Guidelines for the management of patients with supraventricular tachycardia. *European Heart Journal*, *41*(5), 655–720. https://doi.org/10.1093/eurheartj/ehz467
Bucholz, K. K., Cadoret, R., Cloninger, C. R., Dinwiddie, S. H., Hesselbrock, V. M., Nurnberger, J. I., Reich, T., Schmidt, I., & Schuckit, M. A. (1994). A new, semi-structured psychiatric interview for use in genetic linkage studies: A report on the reliability of the SSAGA. *Journal of Studies on Alcohol*, *55*(2), 149–158. https://doi.org/10.15288/jsa.1994.55.149
Burkhardt, G., Adorjan, K., Kambeitz, J., Kambeitz-Ilankovic, L., Falkai, P., Eyer, F., Koller, G., Pogarell, O., Koutsouleris, N., & Dwyer, D. B. (2020). A machine learning approach to risk assessment for alcohol withdrawal syndrome. *European Neuropsychopharmacology*. https://doi.org/https://doi.org/10.1016/j.euroneuro.2020.03.016
Busto, U., Sellers, E. M., Naranjo, C. A., Cappell, H., Sanchez-Craig, M., & Sykora, K. (1986). Withdrawal Reaction after Long-Term Therapeutic Use of Benzodiazepines. *New England Journal of Medicine*, *315*(14), 854–859. https://doi.org/10.1056/NEJM198610023151403
Chekroud, A. M., Zotti, R. J., Shehzad, Z., Gueorguieva, R., Johnson, M. K., Trivedi, M. H., Cannon, T. D., Krystal, J. H., & Corlett, P. R. (2016). Cross-trial prediction of treatment outcome in depression: A machine learning approach. *The Lancet Psychiatry*, *3*(3), 243–250. https://doi.org/10.1016/S2215-0366(15)00471-X
Compton, W. M., Thomas, Y. F., Stinson, F. S., & Grant, B. F. (2007). Prevalence, correlates, disability, and comorbidity of DSM-IV drug abuse and dependence in the United States: Results from the national epidemiologic survey on alcohol and related conditions. *Archives of General Psychiatry*, *64*(5), 566–576. https://doi.org/10.1001/archpsyc.64.5.566
Cooney, G., Heydtmann, M., & Smith, I. D. (2019). Baclofen and the alcohol withdrawal syndrome-a short review. In *Frontiers in Psychiatry* (Vol. 10, Issue JAN). Frontiers Media S.A. https://doi.org/10.3389/fpsyt.2018.00773
Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*, *20*(3), 273–297. https://doi.org/10.1023/A:1022627411411

Daeppen, J. B., Gache, P., Landry, U., Sekera, E., Schweizer, V., Gloor, S., & Yersin, B. (2002). Symptom-triggered vs fixed-schedule doses of benzodiazepine for alcohol withdrawal: A randomized treatment trial. *Archives of Internal Medicine*, *162*(10), 1117–1121. https://doi.org/10.1001/archinte.162.10.1117

Day, E. J., Patel, J., & Georgiou, G. (2004). Evaluation of a sympton-triggered front-loading detoxification technique for alcohol dependence: A pilot study. *Psychiatric Bulletin*, *28*(11), 407–410. https://doi.org/10.1192/pb.28.11.407

Department of Defense, F. G. A. D. of V. A. and V. H. A. (2017). VA/DoD Clinical Practice Guideline for the Management of Substance Use Disorders. *Assessment*, *January 2017*, 1–169. https://doi.org/10.1108/00346659810224208

Dolman, J. M., & Hawkes, N. D. (2005). Combining the AUDIT questionnaire and biochemical markers to assess alcohol use and risk of alcohol withdrawal in medical inpatients. *Alcohol and Alcoholism*, *40*(6), 515–519. https://doi.org/10.1093/alcalc/agh189

Dufouil, C., Alpérovitch, A., & Tzourio, C. (2003). Influence of education on the relationship between white matter lesions and cognition. *Neurology*, *60*(5), 831–836. https://doi.org/10.1212/01.WNL.0000049456.33231.96

Dwyer, D. B., Cabral, C., Kambeitz-Ilankovic, L., Sanfelici, R., Kambeitz, J., Calhoun, V., Falkai, P., Pantelis, C., Meisenzahl, E., & Koutsouleris, N. (2018). Brain Subtyping Enhances The Neuroanatomical Discrimination of Schizophrenia. *Schizophrenia Bulletin*, *44*(5), 1060–1069. https://doi.org/10.1093/schbul/sby008

Dwyer, D. B., Falkai, P., & Koutsouleris, N. (2018). Machine Learning Approaches for Clinical Psychology and Psychiatry. *Annual Review of Clinical Psychology*, *14*(1), 91–118. https://doi.org/10.1146/annurev-clinpsy-032816-045037

Dwyer, D. B., Kalman, J. L., Budde, M., Kambeitz, J., Ruef, A., Antonucci, L. A., Kambeitz-Ilankovic, L., Hasan, A., Kondofersky, I., Anderson-Schmidt, H., Gade, K., Reich-Erkelenz, D., Adorjan, K., Senner, F., Schaupp, S., Andlauer, T. F. M., Comes, A. L., Schulte, E. C., Klöhn-Saghatolislam, F., … Koutsouleris, N. (2020). An Investigation of Psychosis Subgroups with Prognostic Validation and Exploration of Genetic Underpinnings: The PsyCourse Study. *JAMA Psychiatry*, *77*(5), 523–533. https://doi.org/10.1001/jamapsychiatry.2019.4910

Elholm, B., Larsen, K., Hornnes, N., Zierau, F., & Becker, U. (2011). Alcohol withdrawal syndrome: Symptom-triggered versus fixed-schedule treatment in an outpatient setting. *Alcohol and Alcoholism*, *46*(3), 318–323. https://doi.org/10.1093/alcalc/agr020

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, *542*(7639), 115–118. https://doi.org/10.1038/nature21056

Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S., & Dean, J. (2019). A guide to deep learning in healthcare. In *Nature Medicine* (Vol. 25, Issue 1, pp. 24–29). Nature Publishing Group. https://doi.org/10.1038/s41591-018-0316-z

Eyer, F., Schreckenberg, M., Hecht, D., Adorjan, K., Schuster, T., Felgenhauer, N., Pfab, R., Strubel, T., & Zilker, T. (2011). Carbamazepine and valproate as adjuncts in the treatment of alcohol withdrawal syndrome: A retrospective cohort study. *Alcohol and Alcoholism*, *46*(2), 177–184. https://doi.org/10.1093/alcalc/agr005

Eyer, F., Schuster, T., Felgenhauer, N., Pfab, R., Strubel, T., Saugel, B., & Zilker, T. (2011). Risk assessment of moderate to severe alcohol withdrawal-predictors for seizures and delirium tremens in the course of withdrawal. *Alcohol and*

Alcoholism, *46*(4), 427–433. https://doi.org/10.1093/alcalc/agr053
Ferguson, J. A., Suelzer, C. J., Eckert, G. J., Zhou, X. H., & Dittus, R. S. (1996). Risk factors for delirium tremens development. *Journal of General Internal Medicine*, *11*(7), 410–414. https://doi.org/10.1007/BF02600188
Fiellin, D. A., Fiellin, D. A., O'Connor, P. G., Holmboe, E. S., Horwitz, R. I., Fiellin, D. A., Holmboe, E. S., & Horwitz, R. I. (2002). Risk for delirium tremens in patients with alcohol withdrawal syndrome. *Substance Abuse*, *23*(2), 83–94. https://doi.org/10.1080/08897070209511478
Fiellin, D. A., Samet, J. H., & O'Connor, P. G. (1998). Reducing bias in observational research on alcohol withdrawal syndrome. *Substance Abuse*, *19*(1), 23–31. https://doi.org/10.1080/08897079809511370
Fiellin, D. A., Samet, J. H., & O'Connor, P. G. (2015). Addressing Bias in Observational Studies of Alcohol Withdrawal Syndrome: A Call to the Field. *Alcoholism: Clinical and Experimental Research*, *39*(2), 390–390. https://doi.org/10.1111/acer.12634
Filzmoser, P., Liebmann, B., & Varmuza, K. (2009). Repeated double cross validation. *Journal of Chemometrics*, *23*(4), 160–171. https://doi.org/10.1002/cem.1225
Findley, J. K., Park, L. T., Siefert, C. J., Chiou, G. J., Lancaster, R. T., Demoya, M., Gervasini, A., & Velmahos, G. C. (2010). Two routine blood tests-mean corpuscular volume and aspartate aminotransferase-as predictors of delirium tremens in trauma patients. *Journal of Trauma - Injury, Infection and Critical Care*, *69*(1), 199–201. https://doi.org/10.1097/TA.0b013e3181bee583
First, M. B., Williams, J. B. W., Karg, R. S., & Spitzer, R. L. (2015). Structured clinical interview for DSM-5 research version. *American Psychiatric Association, Washington D.C.*
Fiscella, K., Pless, N., Meldrum, S., & Fiscella, P. (2004). Alcohol and opiate withdrawal in US jails. *American Journal of Public Health*, *94*(9), 1522–1524. https://doi.org/10.2105/AJPH.94.9.1522
Foy, A., & Kay, J. (1995). The incidence of alcohol-related problems and the risk of alcohol withdrawal in a general hospital population. *Drug and Alcohol Review*, *14*(1), 49–54. https://doi.org/10.1080/09595239500185051
Foy, A., Kay, J., & Taylor, A. (1997). The course of alcohol withdrawal in a general hospital. *QJM - Monthly Journal of the Association of Physicians*, *90*(4), 253–261. https://doi.org/10.1093/qjmed/90.8.552
Foy, A., March, S., & Drinkwater, V. (1988). Use of an Objective Clinical Scale in the Assessment and Management of Alcohol Withdrawal in a Large General Hospital. *Alcoholism: Clinical and Experimental Research*, *12*(3), 360–364. https://doi.org/10.1111/j.1530-0277.1988.tb00208.x
Galton, F. (1886). Regression Towards Mediocrity in Hereditary Stature. *The Journal of the Anthropological Institute of Great Britain and Ireland*. https://doi.org/10.2307/2841583
Gaonkar, B., & Davatzikos, C. (2013). Analytic estimation of statistical significance maps for support vector machine based multi-variate image analysis and classification. *NeuroImage*, *78*, 270–283. https://doi.org/10.1016/j.neuroimage.2013.03.066
Gianfrancesco, M. A., Tamang, S., Yazdany, J., & Schmajuk, G. (2018). Potential Biases in Machine Learning Algorithms Using Electronic Health Record Data. In *JAMA Internal Medicine* (Vol. 178, Issue 11, pp. 1544–1547). American Medical Association. https://doi.org/10.1001/jamainternmed.2018.3763
Gillman, M. A., Lichtigfeld, F., & Young, T. (2007). Psychotropic analgesic nitrous

oxide for alcoholic withdrawal states. *Cochrane Database of Systematic Reviews*, *2*. https://doi.org/10.1002/14651858.CD005190.pub2
Golland, P., & Fischl, B. (2003). Permutation tests for classification: Towards statistical significance in image-based studies. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/978-3-540-45087-0_28
Gonzalez, L. P., Veatch, L. M., Ticku, M. K., & Becker, H. C. (2001). Alcohol Withdrawal Kindling: Mechanisms and Implications for Treatment. *Alcoholism: Clinical and Experimental Research*, *25*(s1), 197S-201S. https://doi.org/10.1111/j.1530-0277.2001.tb02396.x
Goodson, C. M., Clark, B. J., & Douglas, I. S. (2014). Predictors of Severe Alcohol Withdrawal Syndrome: A Systematic Review and Meta-Analysis. *Alcoholism: Clinical and Experimental Research*, *38*(10), 2664–2677. https://doi.org/10.1111/acer.12529
Grant, B. F., Goldstein, R. B., Saha, T. D., Patricia Chou, S., Jung, J., Zhang, H., Pickering, R. P., June Ruan, W., Smith, S. M., Huang, B., & Hasin, D. S. (2015). Epidemiology of DSM-5 alcohol use disorder results from the national epidemiologic survey on alcohol and related conditions III. *JAMA Psychiatry*, *72*(8), 757–766. https://doi.org/10.1001/jamapsychiatry.2015.0584
Gross, M. M., Lewis, E., & Nagarajan, M. (1973). *An Improved Quantitative System for Assessing the Acute Alcoholic Psychoses and Related States (TSA and SSA)* (pp. 365–376). Springer, Boston, MA. https://doi.org/10.1007/978-1-4684-3234-3_24
Guyon, I. (2003). *An introduction to variable and feature selection*. The Journal of Machine Learning Research. https://dl.acm.org/doi/10.5555/944919.944968
Hall, W., & Zador, D. (1996). The alcohol withdrawal syndrome. *Lancet*, *349*(9069), 1897–1900. https://doi.org/10.1016/S0140-6736(97)04572-8
Hastie, T., Tibshirani, R., & Friedman, J. H. (Jerome H. . (2009). *The elements of statistical learning : data mining, inference, and prediction*. Springer Verlag New York.
Hayashida, M., Alterman, A. I., Mclellan, A. T., O'brien, C. P., Purtill, J. J., Volpicelli, J. R., Raphaelson, A. H., & Hall, C. P. (1989). Comparative Effectiveness and Costs of Inpatient and Outpatient Detoxification of Patients with Mild-to-Moderate Alcohol Withdrawal Syndrome. *New England Journal of Medicine*, *320*(6), 358–365. https://doi.org/10.1056/NEJM198902093200605
He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, *21*(9), 1263–1284. https://doi.org/10.1109/TKDE.2008.239
Heilig, M., Egli, M., Crabbe, J. C., & Becker, H. C. (2010). Acute withdrawal, protracted abstinence and negative affect in alcoholism: are they linked? *Addiction Biology*, *15*(2), 169–184. https://doi.org/10.1111/j.1369-1600.2009.00194.x
Hermann, D., Weber-Fahr, W., Sartorius, A., Hoerst, M., Frischknecht, U., Tunc-Skarka, N., Perreau-Lenz, S., Hansson, A. C., Krumm, B., Kiefer, F., Spanagel, R., Mann, K., Ende, G., & Sommer, W. H. (2012). Translational magnetic resonance spectroscopy reveals excessive central glutamate levels during alcohol withdrawal in humans and rats. *Biological Psychiatry*, *71*(11), 1015–1021. https://doi.org/10.1016/j.biopsych.2011.07.034
Hillbom, M., Pieninkeroinen, I., & Leone, M. (2003). Seizures in Alcohol-Dependent Patients: Epidemiology, Pathophysiology and Management. In *CNS Drugs* (Vol. 17, Issue 14, pp. 1013–1030). https://doi.org/10.2165/00023210-200317140-

00002
Hillemacher, T., Frieling, H., Wilhelm, J., Heberlein, A., Karagülle, D., Bleich, S., Lenz, B., & Kornhuber, J. (2012). Indicators for elevated risk factors for alcohol-withdrawal seizures: An analysis using a random forest algorithm. *Journal of Neural Transmission*, *119*(11), 1449–1453. https://doi.org/10.1007/s00702-012-0825-8
Holleck, J. L., Merchant, N., & Gunderson, C. G. (2019). Symptom-Triggered Therapy for Alcohol Withdrawal Syndrome: a Systematic Review and Meta-analysis of Randomized Controlled Trials. *Journal of General Internal Medicine*. https://doi.org/10.1007/s11606-019-04899-7
Holt, S., Stewart, I. C., Dixon, J. M. J., Elton, R. A., Taylor, T. V., & Little, K. (1980). Alcohol and the emergency service patient. *British Medical Journal*, *281*(6241), 638–640. https://doi.org/10.1136/bmj.281.6241.638
Huang, M. C., Chen, C. H., Liu, H. C., Chen, C. C., Ho, C. C., & Leu, S. J. (2011). Differential Patterns of Serum Brain-Derived Neurotrophic Factor Levels in Alcoholic Patients With and Without Delirium Tremens During Acute Withdrawal. *Alcoholism: Clinical and Experimental Research*, *35*(1), 126–131. https://doi.org/10.1111/j.1530-0277.2010.01329.x
Hughes, J. R. (2009). Alcohol withdrawal seizures. *Epilepsy and Behavior*, *15*(2), 92–97. https://doi.org/10.1016/j.yebeh.2009.02.037
Huys, Q. J. M., Maia, T. V., & Frank, M. J. (2016). Computational psychiatry as a bridge from neuroscience to clinical applications. In *Nature Neuroscience* (Vol. 19, Issue 3, pp. 404–413). Nature Publishing Group. https://doi.org/10.1038/nn.4238
Insel, T. R., & Cuthbert, B. N. (2015). Brain disorders? Precisely. *Science*. https://doi.org/10.1126/science.aab2358
Inza, I., Larrañaga, P., Blanco, R., & Cerrolaza, A. J. (2004). Filter versus wrapper gene selection approaches in DNA microarray domains. *Artificial Intelligence in Medicine*. https://doi.org/10.1016/j.artmed.2004.01.007
Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLoS Medicine*, *2*(8), e124. https://doi.org/10.1371/journal.pmed.0020124
Isbell, H., Fraser, H. F., Wikler, A., Belleville, R. E., & Eisenman, A. J. (1955). An experimental study of the etiology of rum fits and delirium tremens. *Quarterly Journal of Studies on Alcohol*, *16*(1), 1–33.
Jablensky, A. (2016). Psychiatric classifications: Validity and utility. *World Psychiatry*, *15*(1), 26–31. https://doi.org/10.1002/wps.20284
Jaeger, T. M., Lohr, R. H., & Shane Pankratz, V. (2001). Symptom-triggered therapy for alcohol withdrawal syndrome in medical inpatients. *Mayo Clinic Proceedings*, *76*(7), 695–701. https://doi.org/10.4065/76.7.695
Janssen, R. J., Mourão-Miranda, J., & Schnack, H. G. (2018). Making Individual Prognoses in Psychiatry Using Neuroimaging and Machine Learning. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *3*(9), 798–808. https://doi.org/10.1016/j.bpsc.2018.04.004
Jauhar, P. (2000). Is daily single dosage of diazepam as effective as chlordiazepoxide in divided doses in alcohol withdrawal - a pilot study. *Alcohol and Alcoholism*, *35*(2), 212–214. https://doi.org/10.1093/alcalc/35.2.212
Jönsson, P., & Wohlin, C. (2004). An evaluation of k-nearest neighbour imputation using Likert data. *Proceedings - International Software Metrics Symposium*, 108–118. https://doi.org/10.1109/METRIC.2004.1357895
Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. In *Science* (Vol. 349, Issue 6245, pp. 255–260). American

Association for the Advancement of Science. https://doi.org/10.1126/science.aaa8415
Junghanns, K., & Wetterling, T. (2017). Der komplizierte Alkoholentzug: Grand-Mal-Anfälle, Delir und Wernicke-Enzephalopathie. *Fortschritte Der Neurologie Psychiatrie*, *85*(3), 163–177. https://doi.org/10.1055/s-0043-103052
Kaim, S. C., Klett, C. J., & Rothfeld, B. (1969). Treatment of the acute alcohol withdrawal state: a comparison of four drugs. *The American Journal of Psychiatry*, *125*(12), 1640–1646. https://doi.org/10.1176/ajp.125.12.1640
Kesler, S. R., Adams, H. F., Blasey, C. M., & Bigler, E. D. (2003). Premorbid intellectual functioning, education, and brain size in traumatic brain injury: An investigation of the cognitive reserve hypothesis. *Applied Neuropsychology*, *10*(3), 153–162. https://doi.org/10.1207/S15324826AN1003_04
Keun, H. C., Ebbels, T. M. D., Antti, H., Bollard, M. E., Beckonert, O., Holmes, E., Lindon, J. C., & Nicholson, J. K. (2003). Improved analysis of multivariate data by variable stability scaling: application to NMR-based metabolic profiling. *Analytica Chimica Acta*, *490*(1–2), 265–276. https://doi.org/10.1016/S0003-2670(03)00094-1
Khan, A., Levy, P., DeHorn, S., Miller, W., & Compton, S. (2008). Predictors of mortality in patients with delirium tremens. *Academic Emergency Medicine*, *15*(8), 788–790. https://doi.org/10.1111/j.1553-2712.2008.00187.x
Kotsiantis, S. B. (2007). Supervised machine learning: A review of classification techniques. *Informatica (Ljubljana)*, *31*(3), 249–268. https://doi.org/10.31449/inf.v31i3.148
Koutsouleris, N., Kahn, R. S., Chekroud, A. M., Leucht, S., Falkai, P., Wobrock, T., Derks, E. M., Fleischhacker, W. W., & Hasan, A. (2016). Multisite prediction of 4-week and 52-week treatment outcomes in patients with first-episode psychosis: a machine learning approach. *The Lancet Psychiatry*, *3*(10), 935–946. https://doi.org/10.1016/S2215-0366(16)30171-7
Koutsouleris, N., Kambeitz-Ilankovic, L., Ruhrmann, S., Rosen, M., Ruef, A., Dwyer, D. B., Paolini, M., Chisholm, K., Kambeitz, J., Haidl, T., Schmidt, A., Gillam, J., Schultze-Lutter, F., Falkai, P., Reiser, M., Riecher-Rössler, A., Upthegrove, R., Hietala, J., Salokangas, R. K. R., … Borgwardt, S. (2018). Prediction Models of Functional Outcomes for Individuals in the Clinical High-Risk State for Psychosis or with Recent-Onset Depression: A Multimodal, Multisite Machine Learning Analysis. *JAMA Psychiatry*, *75*(11), 1156–1172. https://doi.org/10.1001/jamapsychiatry.2018.2165
Kraemer, K. L., Mayo-Smith, M. F., Calkins, D. R., & Mayo-Smith, M. F. (2003). Independent clinical correlates of severe alcohol withdrawal. *Substance Abuse*, *24*(4), 197–209. https://doi.org/10.1080/08897070309511551
Lange-Asschenfeldt, C., Müller, M. J., Szegedi, A., Anghelescu, I., Klawe, C., & Wetzel, H. (2003). Symptom-triggered versus standard chlormethiazole treatment of inpatient alcohol withdrawal: Clinical implications from a chart analysis. *European Addiction Research*, *9*(1), 1–7. https://doi.org/10.1159/000067729
Latvala, J., Parkkila, S., & Niemelä, O. (2004). Excess alcohol consumption is common in patients with cytopenia: studies in blood and bone marrow cells. *Alcoholism, Clinical and Experimental Research*, *28*(4), 619–624. https://doi.org/10.1097/01.alc.0000122766.54544.3b
Lechtenberg, R., & Worner, T. M. (1991). Relative kindling effect of detoxification and non-detoxification admissions in alcoholics. *Alcohol and Alcoholism*, *26*(2), 221–225. https://doi.org/10.1093/oxfordjournals.alcalc.a045104

Lechtenberg, R., & Worner, T. M. (1992). Total ethanol consumption as a seizure risk factor in alcoholics. *Acta Neurologica Scandinavica*, *85*(2), 90–94. https://doi.org/10.1111/j.1600-0404.1992.tb04004.x

Lee, J. A., Duby, J. J., & Cocanour, C. S. (2019). Effect of early and focused benzodiazepine therapy on length of stay in severe alcohol withdrawal syndrome. *Clinical Toxicology*, *57*(7), 624–627. https://doi.org/10.1080/15563650.2018.1542701

Lee, J. H., Jang, M. K., Lee, J. Y., Kim, S. M., Kim, K. H., Park, J. Y., Lee, J. H., Kim, H. Y., & Yoo, J. Y. (2005). Clinical predictors for delirium tremens in alcohol dependence. *Journal of Gastroenterology and Hepatology (Australia)*, *20*(12), 1833–1837. https://doi.org/10.1111/j.1440-1746.2005.03932.x

Leone, M. A., Vigna-Taglianti, F., Avanzi, G., Brambilla, R., & Faggiano, F. (2010). Gamma-hydroxybutyrate (GHB) for treatment of alcohol withdrawal and prevention of relapses. *The Cochrane Database of Systematic Reviews*, *2*, CD006266. https://doi.org/10.1002/14651858.CD006266.pub2

Linn, S., & Grunau, P. D. (2006). New patient-oriented summary measure of net total gain in certainty for dichotomous diagnostic tests. *Epidemiologic Perspectives and Innovations*. https://doi.org/10.1186/1742-5573-3-11

Littleton, J. (1998). Neurochemical mechanisms underlying alcohol withdrawal. *Alcohol Health and Research World*, *22*(1), 13–24.

Liu, J., & Wang, L. N. (2019). Baclofen for alcohol withdrawal. *Cochrane Database of Systematic Reviews*, *2019*(11). https://doi.org/10.1002/14651858.CD008502.pub6

Lukan, J. K., Reed, D. N., Looney, S. W., Spain, D. A., & Blondell, R. D. (2002). Risk Factors for Delirium Tremens in Trauma Patients. *Journal of Trauma*, *53*(5), 901–906. https://doi.org/10.1097/00005373-200211000-00015

Maalouf, M., Homouz, D., & Trafalis, T. B. (2018). Logistic regression in large rare events and imbalanced data: A performance comparison of prior correction and weighting methods. *Computational Intelligence*, *34*(1), 161–174. https://doi.org/10.1111/coin.12123

Mak, K. K., Lee, K., & Park, C. (2019). Applications of machine learning in addiction studies: A systematic review. *Psychiatry Research*, *275*(March), 53–60. https://doi.org/10.1016/j.psychres.2019.03.001

Malcolm, R., Roberts, J. S., Wang, W., Myrick, H., & Anton, R. F. (2000). Multiple previous detoxifications are associated with less responsive treatment and heavier drinking during an index outpatient detoxification. *Alcohol*, *22*(3), 159–164. https://doi.org/10.1016/S0741-8329(00)00114-2

Maldonado, J. R., Nguyen, L. H., Schader, E. M., & Brooks, J. O. (2012). Benzodiazepine loading versus symptom-triggered treatment of alcohol withdrawal: A prospective, randomized clinical trial. *General Hospital Psychiatry*, *34*(6), 611–617. https://doi.org/10.1016/j.genhosppsych.2012.06.016

Maldonado, J. R., Sher, Y., Ashouri, J. F., Hills-Evans, K., Swendsen, H., Lolak, S., & Miller, A. C. (2014). The “Prediction of Alcohol Withdrawal Severity Scale” (PAWSS): Systematic literature review and pilot study of a new scale for the prediction of complicated alcohol withdrawal syndrome. *Alcohol*, *48*(4), 375–390. https://doi.org/10.1016/j.alcohol.2014.01.004

Maldonado, J. R., Sher, Y., Das, S., Hills-Evans, K., Frenklach, A., Lolak, S., Talley, R., & Neri, E. (2015). Prospective validation study of the prediction of alcohol withdrawal severity scale (PAWSS) in medically ill inpatients: A new scale for the prediction of complicated alcohol withdrawal syndrome. *Alcohol and Alcoholism*, *50*(5), 509–518. https://doi.org/10.1093/alcalc/agv043

Manikant, S., Tripathi, B. M., & Chavan, B. S. (1993). Loading dose diazepam therapy for alcohol withdrawal state. *The Indian Journal of Medical Research*, *98*, 170–173. http://www.ncbi.nlm.nih.gov/pubmed/8262577

Mann, K., Batra, A., & Hoch, E. (2016). *S3-Leitlinie "Screening, Diagnose und Behandlung alkoholbezogener Störungen." [S3 Guideline "Screening, diagnosis and treatment of alcohol use disorders"]. AWMF, Register Nr. 076-001*. *001*(076). https://doi.org/https://doi.org/10.1024/0939-5911/a000464

Mayo-Smith, M. F. (1997). Pharmacological management of alcohol withdrawal. A meta-analysis and evidence-based practice guideline. American Society of Addiction Medicine Working Group on Pharmacological Management of Alcohol Withdrawal. *JAMA: The Journal of the American Medical Association*, *278*(2), 144–151. https://doi.org/10.1001/jama.278.2.144

Mayo-Smith, M. F., Beecher, L. H., Fischer, T. L., Gorelick, D. A., Guillaume, J. L., Hill, A., Jara, G., Kasser, C., Melbourne, J., & for the Working Group on the Management of Alcohol Withdrawal Delirium American Society of Addiction Medicine, P. G. C. (2004). Management of Alcohol Withdrawal Delirium: An Evidence-Based Practice Guideline. *Archives of Internal Medicine*, *164*(13), 1405–1412. https://doi.org/10.1001/archinte.164.13.1405

McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*. https://doi.org/10.1007/BF02295996

Mennecier, D., Thomas, M., Arvers, P., Corberand, D., Sinayoko, L., Bonnefoy, S., Harnois, F., & Thiolet, C. (2008). Factors predictive of complicated or severe alcohol withdrawal in alcohol dependent inpatients. *Gastroenterologie Clinique et Biologique*, *32*(8–9), 792–797. https://doi.org/10.1016/j.gcb.2008.06.004

Minozzi, S., Amato, L., Vecchi, S., & Davoli, M. (2010). Anticonvulsants for alcohol withdrawal. *The Cochrane Database of Systematic Reviews*, *3*, CD005064. https://doi.org/10.1002/14651858.CD005064.pub3

Mitchell, O., Feldman, D., Diakow, M., & Sigal, S. (2016). The pathophysiology of thrombocytopenia in chronic liver disease. *Hepatic Medicine: Evidence and Research*, *8*, 39. https://doi.org/10.2147/hmer.s74612

Mitchell, T. (1997). *Machine Learning*. McGraw-Hill Education Ltd.

Monte, R., Rabuñal, R., Casariego, E., Bal, M., & Pértega, S. (2009). Risk factors for delirium tremens in patients with alcohol withdrawal syndrome in a hospital setting. *European Journal of Internal Medicine*, *20*(7), 690–694. https://doi.org/10.1016/j.ejim.2009.07.008

Morton, W. A., Laird, L. K., Crane, D. F., Partovi, N., & Frye, L. H. (1994). A prediction model for identifying alcohol withdrawal seizures. *American Journal of Drug and Alcohol Abuse*, *20*(1), 75–86. https://doi.org/10.3109/00952999409084058

Moskowitz, G., Chalmers, T. C., Sacks, H. S., Fagerstrom, R. M., & Smith, H. (1983). Deficiencies of Clinical Trials of Alcohol Withdrawal. *Alcoholism: Clinical and Experimental Research*, *7*(1), 42–46. https://doi.org/10.1111/j.1530-0277.1983.tb05409.x

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie Du Sert, N., Simonsohn, U., Wagenmakers, E. J., Ware, J. J., & Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*(1), 1–9. https://doi.org/10.1038/s41562-016-0021

Naranjo, C. A., Sellers, E. M., Chater, K., Iversen, P., Roach, C., & Sykora, K. (1983). Nonpharmacologic intervention in acute alcohol withdrawal. *Clinical Pharmacology and Therapeutics*, *34*(2), 214–219. https://doi.org/10.1038/clpt.1983.155

National Institute for Health and Care Excellence (NICE). (2010). *Alcohol-use disorders: diagnosis and management of physical complications. Clinical Guideline 100*.
National Institute for Health and Care Excellence (NICE). (2011). *Alcohol-use disorders: diagnosis, assessment and management of harmful drinking and alcohol dependence. Clinical guideline 115* (Vol. 5, Issue October). https://www.nice.org.uk/guidance/cg115
Noble, W. S. (2006). What is a support vector machine? *Nature Biotechnology*, *24*(12), 1565–1567. https://doi.org/10.1038/nbt1206-1565
O'Malley, G., Leonard, M., Meagher, D., & O'Keeffe, S. T. (2008). The delirium experience: A review. In *Journal of Psychosomatic Research* (Vol. 65, Issue 3, pp. 223–228). Elsevier. https://doi.org/10.1016/j.jpsychores.2008.05.017
O'Malley, K. J., Cook, K. F., Price, M. D., Wildes, K. R., Hurdle, J. F., & Ashton, C. M. (2005). Measuring diagnoses: ICD code accuracy. In *Health Services Research* (Vol. 40, Issue 5 II, pp. 1620–1639). https://doi.org/10.1111/j.1475-6773.2005.00444.x
Palmstierna, T. (2001). A model for predicting alcohol withdrawal delirium. *Psychiatric Services*, *52*(6), 820–823. https://doi.org/10.1176/appi.ps.52.6.820
Pearson, S. B. (1813). Observations on brain fever. *Edingburgh Medical and Surgical Journal*, *9*(November 1973), 326–332.
Pecoraro, A., Ewen, E., Horton, T., Mooney, R., Kolm, P., McGraw, P., & Woody, G. (2014). Using the AUDIT-PC to predict alcohol withdrawal in hospitalized patients. *Journal of General Internal Medicine*, *29*(1), 34–40. https://doi.org/10.1007/s11606-013-2551-9
Petty, F., Fulton, M., Moeller, F. G., Kramer, G., Wilson, L., Fraser, K., & Isbell, P. (1993). Plasma gamma-aminobutyric acid (GABA) is low in alcoholics. *Psychopharmacology Bulletin*, *29*(2), 277–281. http://www.ncbi.nlm.nih.gov/pubmed/8290677
Polikar, R. (2006). Ensemble Based Systems in Decision Making. *Ieee Circuits And Systems Magazine*, 21–45.
Porcel, F. J. R., & Schutta, H. S. (2015). From Antiquity to the N-Methyl-D-Aspartate Receptor: A History of Delirium Tremens. *Journal of the History of the Neurosciences*, *24*(4), 378–395. https://doi.org/10.1080/0964704X.2015.1034515
Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. In *New England Journal of Medicine* (Vol. 380, Issue 14, pp. 1347–1358). Massachussetts Medical Society. https://doi.org/10.1056/NEJMra1814259
Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., Ding, D., Bagul, A., Langlotz, C., Shpanskaya, K., Lungren, M. P., & Ng, A. Y. (2017). *CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning*. http://arxiv.org/abs/1711.05225
Ramos. (2013). Predictors of Severity of Alcohol Withdrawal in Hospitalized Patients. *Journal of Clinical Medicine Research*, *5*(5), 376–380. https://doi.org/10.4021/jocmr1514w
Rathlev, N. K., Ulrich, A., Fish, S. S., & D'Onofrio, G. (2000). Clinical characteristics as predictors of recurrent alcohol-related seizures. *Academic Emergency Medicine*, *7*(8), 886–891. https://doi.org/10.1111/j.1553-2712.2000.tb02066.x
Reoux, J P, & Miller, K. (2000). Routine hospital alcohol detoxification practice compared to symptom triggered management with an Objective Withdrawal Scale (CIWA-Ar). *The American Journal on Addictions / American Academy of Psychiatrists in Alcoholism and Addictions*.

Reoux, Joseph P., Malte, C. A., Kivlahan, D. R., & Saxon, A. J. (2002). The alcohol use disorders identification test (AUDIT) predicts alcohol withdrawal symptoms during inpatient detoxification. *Journal of Addictive Diseases*, *21*(4), 81–91. https://doi.org/10.1300/J069v21n04_08

Rosenblatt, F. (1958). The Perceptron: A Probabilistic Model for Information Storage and Organization in The Brain. *PSYCHOLOGICAL REVIEW*, 65--386. https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.588.3775

Ruschhaupt, M., Huber, W., & Poustka, Annemarie, U. M. (2004). Statistical Applications in Genetics and Molecular Biology A Compendium to Ensure Computational Reproducibility in High-Dimensional Classification Tasks A Compendium to Ensure Computational Reproducibility in High-Dimensional Classification Tasks ∗. *Statistical Applications in Genetics and Molecular Biology*, *3*.

Sachdeva, A., Chandra, M., & Deshpande, S. N. (2014). A comparative study of fixed tapering dose regimen versus symptom-triggered regimen of lorazepam for alcohol detoxification. *Alcohol and Alcoholism*, *49*(3), 287–291. https://doi.org/10.1093/alcalc/agt181

Saitz, R. (2018). Medications for alcohol use disorder and predicting severe withdrawal. *JAMA - Journal of the American Medical Association*, *320*(8), 766–768. https://doi.org/10.1001/jama.2018.10061

Saitz, R., Friedman, L. S., & Mayo-Smith, M. F. (1995). Alcohol withdrawal: a nationwide survey of inpatient treatment practices. *Journal of General Internal Medicine*, *10*(9), 479–487. https://doi.org/10.1007/bf02602395

Saitz, R., Mayo Smith, M. F., Roberts, M. S., Redmond, H. A., Bernard, D. R., & Calkins, D. R. (1994). Individualized Treatment for Alcohol Withdrawal: A Randomized Double-blind Controlled Trial. *JAMA: The Journal of the American Medical Association*, *272*(7), 519–523. https://doi.org/10.1001/jama.1994.03520070039035

Salottolo, K., Mcguire, E., Mains, C. W., Van Doorn, E. C., & Bar-Or, D. (2017). Occurrence, predictors, and prognosis of alcohol withdrawal syndrome and delirium tremens following traumatic injury. *Critical Care Medicine*, *45*(5), 867–874. https://doi.org/10.1097/CCM.0000000000002371

Saunders, J. B., Aasland, O. G., Babor, T. F., De La Fuente, J. R., & Grant, M. (1993). Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption-II. *Addiction*, *88*(6), 791–804. https://doi.org/10.1111/j.1360-0443.1993.tb02093.x

Schmidt, K. J., Doshi, M. R., Holzhausen, J. M., Natavio, A., Cadiz, M., & Winegardner, J. E. (2016). Treatment of Severe Alcohol Withdrawal. In *Annals of Pharmacotherapy* (Vol. 50, Issue 5, pp. 389–401). SAGE Publications Inc. https://doi.org/10.1177/1060028016629161

Schuckit, M. A. (2014). Recognition and management of withdrawal delirium (delirium tremens). *New England Journal of Medicine*, *371*(22), 2109–2113. https://doi.org/10.1056/NEJMra1407298

Schuckit, M. A., Tipp, J. E., Reich, T., Hesselbrock, V. M., & Bucholz, K. K. (1995). The histories of withdrawal convulsions and delirium tremens in 1648 alcohol dependent subjects. *Addiction*, *90*(10), 1335–1347. https://doi.org/10.1046/j.1360-0443.1995.901013355.x

Sellers, E. M., Naranjo, C. A., Harrison, M., Devenyi, P., Roach, C., & Sykora, K. (1983). Diazepam loading: simplified treatment of alcohol withdrawal. *Clinical Pharmacology and Therapeutics*, *34*(6), 822–826.

https://doi.org/10.1038/clpt.1983.256
Sellers, E. M., Sullivan, J. T., Somer, G., & Sykora, K. (1991). Characterization of DSM-III-R Criteria for Uncomplicated Alcohol Withdrawal Provides an Empirical Basis for DSM-IV. *Archives of General Psychiatry*, *48*(5), 442–447. https://doi.org/10.1001/archpsyc.1991.01810290054011
Shaw, J. M., Kolesar, G. S., Sellers, E. M., Kaplan, H. L., & Sandor, P. (1981). Development of optimal treatment tactics for alcohol withdrawal. I. Assessment and effectiveness of supportive care. *Journal of Clinical Psychopharmacology*, *1*(6), 382–389. https://doi.org/10.1097/00004714-198111000-00006
Siontis, G. C. M., Tzoulaki, I., Castaldi, P. J., & Ioannidis, J. P. A. (2015). External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. *Journal of Clinical Epidemiology*, *68*(1), 25–34. https://doi.org/10.1016/j.jclinepi.2014.09.007
Soravia, L. M., Wopfner, A., Pfiffner, L., Bétrisey, S., & Moggi, F. (2018). Symptom-triggered detoxification using the alcohol-withdrawal-scale reduces risks and healthcare costs. *Alcohol and Alcoholism*, *53*(1), 71–77. https://doi.org/10.1093/alcalc/agx080
Soyka, M., & Horak, M. (2004). Outpatient alcohol detoxification: Implementation efficacy and outcome effectiveness of a model project. *European Addiction Research*, *10*(4), 180–187. https://doi.org/10.1159/000079840
Stern, T. A., Gross, A. F., Stern, T. W., Nejad, S. H., & Maldonado, J. R. (2010). Current approaches to the recognition and treatment of alcohol withdrawal and delirium tremens: “Old wine in new bottles” or “new wine in old bottles.” *Primary Care Companion to the Journal of Clinical Psychiatry*, *12*(3), 1–14. https://doi.org/10.4088/PCC.10r00991ecr
Stern, Y. (2013). Cognitive reserve in ageing. *Lancet Neurol.*, *11*(11), 1006–1012. https://doi.org/10.1016/S1474-4422(12)70191-6.Cognitive
Stockwell, T., Bolt, L., Milner, I., Russell, G., Bolderston, H., & Pugh, P. (1991). Home detoxification from alcohol: Its safety and efficacy in comparison with inpatient care. *Alcohol and Alcoholism*, *26*(5–6), 645–650. https://doi.org/10.1093/oxfordjournals.alcalc.a045169
Stokes, J. M., Yang, K., Swanson, K., Jin, W., Cubillos-Ruiz, A., Donghia, N. M., MacNair, C. R., French, S., Carfrae, L. A., Bloom-Ackerman, Z., Tran, V. M., Chiappino-Pepe, A., Badran, A. H., Andrews, I. W., Chory, E. J., Church, G. M., Brown, E. D., Jaakkola, T. S., Barzilay, R., & Collins, J. J. (2020). A Deep Learning Approach to Antibiotic Discovery. *Cell*, *180*(4), 688-702.e13. https://doi.org/10.1016/j.cell.2020.01.021
Stone, M. (1974). Cross-Validatory Choice and Assessment of Statistical Predictions. Journal of the Royal Statistical Societ. *Journal of the Royal Statistical Society*, *36*(2), 111–147. https://doi.org/10.2307/2984809
Sullivan, J. T., Sykora, K., Schneiderman, J., Naranjo, C. A., & Se, E. M. (1989). Assessment of Alcohol Withdrawal: the revised clinical institute withdrawal assessment for alcohol scale (CIWA-Ar). *British Journal of Addiction*, *84*(11), 1353–1357. https://doi.org/10.1111/j.1360-0443.1989.tb00737.x
Sutton, T. M. D. (1813). *Tracts on Delirium Tremens, on Peritonitis, and on some other Inflammatory Affections, and on the Gout*. *1970*(0), 1900–1970.
Thomas, D. W., & Freedman, D. X. (1964). Treatment of the Alcohol Withdrawal Syndrome: Comparison of Promazine and Paraldehyde. *JAMA*, *188*(3), 316–318. https://doi.org/10.1001/jama.1964.03060290120044
Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*, *25*(1), 44–56.

https://doi.org/10.1038/s41591-018-0300-7
Tran, T., Luo, W., Phung, D., Harvey, R., Berk, M., Kennedy, R. L., & Venkatesh, S. (2014). Risk stratification using data from electronic medical records better predicts suicide risks than clinician assessments. *BMC Psychiatry*, *14*(1), 76. https://doi.org/10.1186/1471-244X-14-76
Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., Botstein, D., & Altman, R. B. (2001). Missing value estimation methods for DNA microarrays. *Bioinformatics*, *17*(6), 520–525. https://doi.org/10.1093/bioinformatics/17.6.520
Tsai, G., Gastfriend, D. R., & Coyle, J. T. (1995). The glutamatergic basis of human alcoholism. In *American Journal of Psychiatry* (Vol. 152, Issue 3, pp. 332–340). American Psychiatric Association. https://doi.org/10.1176/ajp.152.3.332
Vanderweide, L. A., Foster, C. J., Maclaren, R., Kiser, T. H., Fish, D. N., & Mueller, S. W. (2016). Evaluation of Early Dexmedetomidine Addition to the Standard of Care for Severe Alcohol Withdrawal in the ICU. *Journal of Intensive Care Medicine*, *31*(3), 198–204. https://doi.org/10.1177/0885066614554908
Varoquaux, G., Raamana, P. R., Engemann, D. A., Hoyos-Idrobo, A., Schwartz, Y., & Thirion, B. (2017). Assessing and tuning brain decoders: Cross-validation, caveats, and guidelines. *NeuroImage*, *145*, 166–179. https://doi.org/https://doi.org/10.1016/j.neuroimage.2016.10.038
Victor, M., & Adams, R. D. (1953). The effect of alcohol on the nervous system. *Research Publications - Association for Research in Nervous and Mental Disease*, *32*, 526–573.
Victor, M., & Brausch, C. (1967). The Role of Abstinence in the Genesis of Alcoholic Epilepsy. *Epilepsia*, *8*(1), 1–20. https://doi.org/10.1111/j.1528-1157.1967.tb03815.x
Vinson, D. C., & Menezes, M. (1991). Admission alcohol level: a predictor of the course of alcohol withdrawal. *The Journal of Family Practice*, *33*(2), 161–167. http://www.ncbi.nlm.nih.gov/pubmed/1653816
Wadstein, J., & Skude, G. (1978). Does hypokalaemia precede delirium tremens? *Lancet (London, England)*, *2*(8089), 549–550. https://doi.org/10.1016/s0140-6736(78)92883-0
Walker, E. R., McGee, R. E., & Druss, B. G. (2015). Mortality in mental disorders and global disease burden implications a systematic review and meta-analysis. *JAMA Psychiatry*, *72*(4), 334–341. https://doi.org/10.1001/jamapsychiatry.2014.2502
Walter, M., Alizadeh, S., Jamalabadi, H., Lueken, U., Dannlowski, U., Walter, H., Olbrich, S., Colic, L., Kambeitz, J., Koutsouleris, N., Hahn, T., & Dwyer, D. B. (2019). Translational machine learning for psychiatric neuroimaging. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *91*, 113–121. https://doi.org/10.1016/j.pnpbp.2018.09.014
Wasilewski, D., Matsumoto, H., Kur, E., Dziklinska, A., E., W., Stencka, K., Skalski, M., Chaba, P., & Szelenberger, W. (1996). Assessment of diazepam loading dose therapy of delirium tremens. *Alcohol and Alcoholism*, *31*(3), 273–278. https://doi.org/10.1093/oxfordjournals.alcalc.a008147
Weaver, M. F., Hoffman, H. J., Johnson, R. E., & Mauck, K. (2006). Alcohol withdrawal pharmacotherapy for inpatients with medical comorbidity. *Journal of Addictive Diseases*, *25*(2), 17–24. https://doi.org/10.1300/j069v25n02_03
Wetterling, T. (2001). The severity of alcohol withdrawal is not age dependent. *Alcohol and Alcoholism*, *36*(1), 75–78. https://doi.org/10.1093/alcalc/36.1.75
Wetterling, T., Kanitz, R. -D, Veltrup, C., & Driessen, M. (1994). Clinical Predictors of

Alcohol Withdrawal Delirium. *Alcoholism: Clinical and Experimental Research*, *18*(5), 1100–1102. https://doi.org/10.1111/j.1530-0277.1994.tb00087.x
Wetterling, T., Kanitz, R. D., Besters, B., Fischer, D., Zerfass, B., John, U., Spranger, H., & Driessen, M. (1997). A new rating scale for the assessment of the alcohol-withdrawal syndrome (AWS scale). *Alcohol and Alcoholism*, *32*(6), 753–760. https://doi.org/10.1093/oxfordjournals.alcalc.a008326
Wetterling, T., Weber, B., Depfenhart, M., Schneider, B., & Junghanns, K. (2006). Development of a rating scale to predict the severity of alcohol withdrawal syndrome. *Alcohol and Alcoholism*, *41*(6), 611–615. https://doi.org/10.1093/alcalc/agl068
Whelan, R., Watts, R., Orr, C. A., Althoff, R. R., Artiges, E., Banaschewski, T., Barker, G. J., Bokde, A. L. W., Büchel, C., Carvalho, F. M., Conrod, P. J., Flor, H., Fauth-Bühler, M., Frouin, V., Gallinat, J., Gan, G., Gowland, P., Heinz, A., Ittermann, B., … Garavan, H. (2014). Neuropsychosocial profiles of current and future adolescent alcohol misusers. *Nature*, *512*(7513), 185–189. https://doi.org/10.1038/nature13402
Whitfield, C. L., Thompson, G., Lamb, A., Spencer, V., Pfeifer, M., & Browning-Ferrando, M. (1978). Detoxification of 1,024 Alcoholic Patients Without Psychoactive Drugs. *JAMA*, *239*(14), 1409–1410. https://doi.org/10.1001/jama.1978.03280410051018
Wiens, J., Saria, S., Sendak, M., Ghassemi, M., Liu, V. X., Doshi-Velez, F., Jung, K., Heller, K., Kale, D., Saeed, M., Ossorio, P. N., Thadaney-Israni, S., & Goldenberg, A. (2019). Do no harm: a roadmap for responsible machine learning for health care. *Nature Medicine*, *25*(9), 1337–1340. https://doi.org/10.1038/s41591-019-0548-6
Williams, D. (2001). A comparison of rating scales for the alcohol-withdrawal syndrome. *Alcohol and Alcoholism*, *36*(2), 104–108. https://doi.org/10.1093/alcalc/36.2.104
Wojnar, M., Bizon, Z., & Wasilewski, D. (1999). Assessment of the Role of Kindling in the Pathogenesis of Alcohol Withdrawal Seizures and Delirium Tremens. *Alcoholism: Clinical and Experimental Research*, *23*(2), 204–206. https://doi.org/10.1111/j.1530-0277.1999.tb04100.x
Wojnar, M., Bizoń, Z., & Wasilewski, D. (1999). The role of somatic disorders and physical injury in the development and course of alcohol withdrawal delirium. *Alcoholism: Clinical and Experimental Research*, *23*(2), 209–213. https://doi.org/10.1111/j.1530-0277.1999.tb04101.x
Wong, A., Benedict, N. J., Lohr, B. R., Pizon, A. F., & Kane-Gill, S. L. (2015). Management of benzodiazepine-resistant alcohol withdrawal across a healthcare system: Benzodiazepine dose-escalation with or without propofol. *Drug and Alcohol Dependence*, *154*, 296–299. https://doi.org/10.1016/j.drugalcdep.2015.07.005
Wood, A. M., Kaptoge, S., Butterworth, A., Nietert, P. J., Warnakula, S., Bolton, T., Paige, E., Paul, D. S., Sweeting, M., Burgess, S., Bell, S., Astle, W., Stevens, D., Koulman, A., Selmer, R. M., Verschuren, M., Sato, S., Njølstad, I., Woodward, M., … Thompson, S. (2018). Risk thresholds for alcohol consumption: combined analysis of individual-participant data for 599 912 current drinkers in 83 prospective studies. *The Lancet*, *391*(10129), 1513–1523. https://doi.org/10.1016/S0140-6736(18)30134-X
Wood, E., Albarqouni, L., Tkachuk, S., Green, C. J., Ahamad, K., Nolan, S., McLean, M., & Klimas, J. (2018). Will this hospitalized patient develop severe alcohol withdrawal syndrome? The rational clinical examination systematic review.

JAMA - Journal of the American Medical Association, *320*(8), 825–833. https://doi.org/10.1001/jama.2018.10574

World Health Organization. (2004). *ICD-10 : international statistical classification of diseases and related health problems : tenth revision* (2nd ed). World Health Organization. https://apps.who.int/iris/handle/10665/42980

Wright, T., Myrick, H., Henderson, S., Peters, H., & Malcolm, R. (2006). Risk factors for delirium tremens: A retrospective chart review. *American Journal on Addictions*, *15*(3), 213–219. https://doi.org/10.1080/10550490600625798

Yang, X., Song, Q., & Cao, A. (2005). Weighted support vector machine for data classification. *Proceedings of the International Joint Conference on Neural Networks*, *2*(August), 859–864. https://doi.org/10.1109/ijcnn.2005.1555965

Zhao, L., Liu, J., He, C., Yan, R., Zhou, K., Cui, Q., Meng, X., Li, X., Zhang, Y., Nie, Y., Zhang, Y., Hu, R., Liu, Y., Zhao, L., Chen, M., Xiao, W., Tian, J., Zhao, Y., Cao, L., … Dai, K. (2017). Protein kinase A determines platelet life span and survival by regulating apoptosis. *Journal of Clinical Investigation*, *127*(12), 4338–4351. https://doi.org/10.1172/JCI95109

7 Curriculum vitae

Vor- und Nachname: Gerrit Burkhardt

Geburtsdatum: 24.02.1993 in Freiburg im Breisgau

Berufliche Tätigkeit

seit 01.07.2018	Arzt in Weiterbildung an der Klinik für Psychiatrie und Psychotherapie, LMU Klinikum München - diagnoseübergreifende Privatstation B1a - German Center for Brain Stimulation

Forschung

Seit 11/2019	Studienarzt am German Center for Brain Stimulation (GCBS), Leitung Prof. Dr. med. Frank Padberg
Seit 04/2019	Stellvertretender Prüfer, Multimodal Prevention of First Psychotic Episode, Esprit B1 (EudraCT No.: 2014-003076-22)
Seit 02/2019	Drug Monitor, AMSP-Projekt (Institut für Arzneimittelsicherheit in der Psychiatrie AMSP e.V.)
Seit 12/2016	Promotion: *Using machine learning to predict individual severity estimates of alcohol withdrawal syndrome in patients with alcohol dependence*. Betreuung: Prof. Dr. med. Oliver Pogarell, Dr. med. Kristina Adorjan; Klinik für Psychiatrie u. Psychotherapie, LMU Klinikum München

Publikationen

Burkhardt G, Adorjan K, Kambeitz J, Kambeitz-Ilankovic L, Falkai F, Eyer F, Koller G, Pogarell O, Koutsouleris N, Dwyer DB „A machine learning approach to risk assessment for alcohol withdrawal syndrome“, European Neuropsychopharmacology 2020, DOI: 10.1016/j.euroneuro.2020.03.016, PMID: 32418843.

Ausbildung

11.06.2018	Approbation; Regierung von Oberbayern
2011 – 2018	Studium der Humanmedizin: Ludwig-Maximilian-Universität München - 3. Abschnitt der ärztlichen Prüfung (03./04.Mai 2018, Note 1,7) - Wahlfächer: Medizinethik und Psychiatrie
2003 – 2010	Schulische Laufbahn: Rotteck-Gymnasium Freiburg im Breisgau (Abiturnote 1,4)

Weitere Praktische Erfahrungen

Klinische Praktika während des Medizinstudiums:

2017/2018	Klinik und Poliklinik für Psychiatrie und Psychotherapie, LMU Klinikum
2017	Klinikum Starnberg (Innere Medizin)
2017	Klinik für Allgemeine, Unfall-, Hand- und Plastische Chirurgie – IS, LMU Klinikum
2016	Kreiskrankenhaus St. Elisabeth, Dillingen (Innere Medizin und Anästhesie)
2015	Padhar Hospital, Dist. Betul (M.P.), Indien (Chirurgie und Psychiatrie)
2014	Diakoniekrankenhaus Freiburg (Innere Medizin)

Teilnahme „DGPPN Summer School Forensische Psychiatrie und Psychotherapie“ (06/2016, Göttingen)

Sprachkenntnisse	Englisch – verhandlungssicher
Sonstige Kenntnisse	Grundlegende Programmierkenntnisse in Matlab, R, und Python (Statistik, Machine Learning)

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Promotionsbüro
Medizinische Fakultät

MMRS

Eidesstattliche Versicherung

Burkhardt, Gerrit

Name, Vorname

Ich erkläre hiermit an Eides statt,

dass ich die vorliegende Dissertation mit dem Titel

Using machine learning to predict individual severity estimates of alcohol withdrawal syndrome in patients with alcohol dependence

selbständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

Ich erkläre des Weiteren, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

München, 09.07.2021

Ort, Datum

Gerrit Burkhardt

Unterschrift Doktorandin bzw. Doktorand