The Ability of Forensic Psychiatric Patients with Substance Use Disorder to Learn Neurofeedback

Sandra Fielenbach\textsuperscript{a,b}, Franc C.L. Donkers\textsuperscript{c}, Marinus Spreen\textsuperscript{a}, and Stefan Bogaerts\textsuperscript{b,d}

\textsuperscript{a}Research Department, FPC Dr. S. Van Mesdag, Groningen, Netherlands; \textsuperscript{b}Department of Developmental Psychology, Tilburg University, Tilburg, Netherlands; \textsuperscript{c}Department of Cognitive Neuroscience, Maastricht University, Maastricht, Netherlands; \textsuperscript{d}Forensic Psychiatric Center de Kijvelanden, Portugal, Netherlands

ABSTRACT
Despite the increasing use of neurofeedback in clinical psychology, it is rarely used in forensic psychiatric settings. This study investigated whether forensic psychiatric patients ($n=19$) diagnosed with substance use disorder were able to learn to control EEG-activity based on a sensorimotor rhythm/theta neurofeedback protocol. Criteria for qualifying patients as responders were established and scores on impulsivity measures and changes in level of craving over time were assessed. Results indicated that one in five patients was able to consistently change the targeted frequency bands. All patients improved on self-reported impulsivity measures and levels of craving, but only levels of craving were associated with responding to neurofeedback treatment. Patients were more able to up-train the sensorimotor rhythm magnitude than to down-train theta magnitude. Although these results are encouraging for some forensic patients, it is important to assess which patients will respond positively to the training and which will not. This requires more research.

KEYWORDS
Impulsivity; substance use disorders; neurofeedback; offenders; EEG-learning

Introduction
Neurofeedback training is increasingly used in clinical psychology as a noninvasive treatment method. With neurofeedback training, patients learn to regulate aberrant electroencephalographic (EEG) activity assumed to underlie the manifestation of clinical and behavioral symptoms of various disorders. Since neurofeedback training aims directly at changing basic neurophysiological brain functioning, it is less dependent on direct patient-therapist interaction than traditional psychotherapeutic interventions (Cashner, 2013). Patients’ motivation and compliance for psychotherapy tends to be especially low in forensic psychiatric populations (O’Brien & Daffern, 2017; Ogloff, Wong, & Greenwood, 1990). Neurofeedback training might provide a promising alternative treatment option for this patient population.

Forensic patients are often diagnosed with externalizing disorders characterized by lack of inhibitory control, such as Cluster B personality disorders, schizophrenia, and attention-deficit/hyperactivity disorder (ADHD). In addition, about 70% of the forensic patients have a comorbid substance use disorder (Van Nieuwenhuizen et al., 2011). In substance use disorder (SUD), lack of inhibitory control is especially pronounced. Drug-taking can be seen as a loss over control of drug-intake, despite the fact that people are usually aware of the negative consequences of their drug-use. Individuals with SUD show significantly higher score on various impulsivity measures, regardless of type of drug addiction (Fillmore & Rush, 2006; Nielsen et al., 2012). They show impaired response inhibition, which has often-times already been observed in childhood and early adulthood (Hawkins, Catalano, & Miller, 1992). Individuals who report higher levels of impulsivity also report more frequent and more severe symptoms of SUD, such as the level of experienced drug craving. Studies in patients with methamphetamine (Tziortzis, Mahoney, Kalechstein, Newton, & De La Garza, 2011) or cocaine addiction (Moeller et al., 2001), as well as alcohol dependency (Joos et al., 2013), have shown that patients who score higher on impulsivity report more severe levels of craving.

For forensic psychiatric patients, the combination of SUD and comorbid major mental disorders also has a negative impact on treatment (Van Nieuwenhuizen et al., 2011), as high levels of impulsivity increase chances for relapse in substance abuse and treatment dropout (Van der Veeken, Lucieer, & Bogaerts, 2016). Most
likely, chronic substance abuse results in neurocognitive and neurophysiological changes, causing a structural state of reduced inhibitory control and high levels of impulsivity (Jentsch, & Taylor, 1999; Lyvers, 2000). Neurofeedback protocols aimed at enhancing the sensorimotor rhythm (SMR; 12–15 Hz) and reducing slower waves such as theta (3.5–7.5 Hz) have shown promising results in reducing levels of impulsivity in ADHD (Fuchs, Birbaumer, Lutzenberger, Gruzelier, & Kaiser, 2003). A reduction in levels of impulsivity through neurofeedback training could possibly also have a positive effect on SUD, as both impulsivity and SUD are characterized by a lack of inhibitory control (Tomko, Bountress, & Gray, 2016).

To our knowledge, neurofeedback training for forensic psychiatric patient populations with multiple externalizing disorders has not been performed so far. A possible explanation might be that co-morbidity is often seen as a contra-indication to include patients in Randomized Controlled Trials (RCTs) (Janssen et al., 2017; Mayer et al., 2016; Mayer, Wyckoff, Schulz, & Strehl, 2012; Mohammadi, Malmir, Khaledhi, & Aminiorani, 2015; Moreno-Garcia, Delgado-Pardo, Camacho-Vara de Rey, Meneres-Sancho, & Servera-Barcelo, 2015). Although neurofeedback training is increasingly applied in clinical populations, its effectiveness is often solely described in terms of a reduction in behavioral and clinical symptoms (e.g., Bink, Van Nieuwenhuizen, Popma, Bongers, & Van Boxtel, 2015). Results describing changes in deviant EEG-patterns due to neurofeedback training are often not reported.

A necessary first step in showing that clinical improvements are actually linked to successful regulation of cortical activity is to demonstrate that learning of EEG-regulation has occurred during neurofeedback training (Gruzelier, Egner, & Vernon, 2006; Zuberer, Brandeis, & Drechsler, 2015). Previous research shows mixed results regarding the specific effects of neurofeedback training (e.g., Cortese et al., 2016; Zuberer et al., 2015), regardless of the applied neurofeedback protocol or the disorder for which the training was employed. Oftentimes, participants seem unable to learn to regulate the targeted cortical activity over the course of neurofeedback training. As an example, in a study among children diagnosed with ADHD, approximately 50% of participants were classified as so-called non-responders (Doehnert, Brandeis, Straub, Steinhausen, & Drechsler, 2008). Even in studies with non-clinical participants, responder rates tend to be quite low. Weber, Köberl, Frank, and Doppelmayr (2011) trained non-clinical participants to increase SMR frequency through neurofeedback training. They found that only 43–54% of participants were able to consistently increase their SMR-frequency.

Studies describing learning progress during neurofeedback training have also reported different patterns of participants regarding the adaptation of learning strategies, with large inter-individual variability in learning performance (Drechsler et al., 2007; Leins et al., 2007; Strehl et al., 2006). Several studies (e.g., Bakhshayesh, Hänsch, Wyschkon, Rezai, & Esser, 2011; Lubar, Swartwood, Swartwood, & O’Donnell, 1995) report that participants showed positive learning curves at the beginning of a theta/beta neurofeedback training, followed by a stagnation in learning progress, again followed by a second learning curve at the end of the training sessions. This was similar to a Slow Cortical Potential (SCP) training protocol in a study by Mayer et al. (2016). The participants in this study showed a positive learning curve in the first few sessions, but then deteriorated and needed time to adapt strategies to deliberately control cortical activity. Inconsistent learning curves also add to the debate about the appropriate number of sessions required to see improvements, as studies vary significantly in the number of sessions (e.g., from 10 in schizophrenic patients (Gruzelier, Hardman, Wild, & Zaman, 1999) to 35–50 in ADHD (Gevensleben et al., 2009; Heinrich, Gevensleben, Freisleder, Moll, & Rothenberger, 2004)). It is plausible that the complexity of comorbidity can complicate the learning process to regulate and control cortical activity.

This raises the question of (pre)conditions of neurofeedback training that must be met before starting an intervention among forensic psychiatric patients. Zoefel, Huster, and Herrmann (2011) have established three criteria that in their view any neurofeedback training should adhere to, in order to be valid: (a) trainability: neurofeedback training should lead to effects in the trained frequency band; (b) independence: the training should not affect other frequency bands; and (c) interpretability: frequency bands should be associated with certain cognitive functions to increase the probability of reliable behavioral effects.

The current study focuses on the trainability and interpretability aspect of neurofeedback training in forensic psychiatric patients with SUD and other co-morbidities, such as personality disorders, ADHD and/or schizophrenia. It will investigate: (a) to what extent this patient group is able to learn to regulate neurophysiological activity through a SMR/theta-neurofeedback training; (b) to what extent changes in frequency bands are related to changes in their levels of impulsivity; and (c) to what extent a reduction in levels of impulsivity is related to a reduction in their
SUD-related behavior, such as reduced levels of craving for substances and actual drug intake.

Methods

This study is part of a RCT investigating the effects of neurofeedback training for impulsivity in a forensic psychiatric population with SUD (Fielenbach, Donkers, Spreen, & Bogaerts, 2017). Results of the RCT will be reported elsewhere. In this study, only outcomes for those patients who received neurofeedback training will be reported. The study was conducted according to the principles of the Declaration of Helsinki (version 59, Seoul, October 2008) and in accordance with the Medical Research Involving Human Subjects Act. It has been approved by the medical ethical council of Brabant, the Netherlands (study number NL46390.008.13).

Participants

Participants were 26 forensic psychiatric patients residing in a maximum-secured inpatient treatment facility situated in Groningen, the Netherlands. Patients in this treatment facility are convicted for a crime with a minimum penalty of at least four years according to Dutch jurisdiction. These patients are held to be only partially responsible for their behavior due to mental illness and are admitted to a forensic psychiatric center by order of the state (called Ter Beschikking Stelling (TBS); Van Nieuwenhuizen et al., 2011). All patients suffer from at least one DSM-IV-TR disorder (American Psychiatric Association, 2000). Inclusion criteria for participants in this study consisted of at least one diagnosis of SUD according to DSM-IV-TR, and positive drug testing at the treatment facility in the past 24 months before the start of inclusion. Exclusion criteria were acute psychosis, acquired/congenital neurological brain disorders (e.g., epilepsy), and visual and/or auditory impairments, which would severely influence a patient’s ability to follow neurofeedback training.

Patients were allowed to take medication during the study and clinical supervisors were asked to report pharmacological changes during the course of the training. Patients received 20 neurofeedback sessions, lasting 40 minutes at a time, 2 times a week. Neurofeedback training was added to treatment-as-usual (TAU). TAU was different for every patient, but typically consisted of cognitive behavioral therapy, non-verbal therapy (such as music therapy or psychomotor therapy), and behavioral skills training.

Before the start of the training, patients participated in pre-training measurements. After the last session, the same instruments were assessed again as post-treatment measurements.

Sample characteristics

Descriptive statistics are given in Table 1. Out of 26 patients, 19 patients completed all sessions of

<table>
<thead>
<tr>
<th>Table 1. Sample characteristics (n = 19).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample</td>
</tr>
<tr>
<td>Age Mean (SD)</td>
</tr>
<tr>
<td>Range</td>
</tr>
<tr>
<td>IQ Range</td>
</tr>
<tr>
<td>Months in treatment</td>
</tr>
<tr>
<td>Mean (SD)</td>
</tr>
<tr>
<td>Range</td>
</tr>
<tr>
<td>Number of Axis I and II disorders Mean (SD)</td>
</tr>
<tr>
<td>Diagnoses</td>
</tr>
<tr>
<td>Pervasive Developmental Disorder*1</td>
</tr>
<tr>
<td>Attention-deficit hyperactivity disorder (ADHD)*2</td>
</tr>
<tr>
<td>Schizophrenia or other psychotic disorder</td>
</tr>
<tr>
<td>Mood and Anxiety disorder</td>
</tr>
<tr>
<td>Cognitive impairment</td>
</tr>
<tr>
<td>Pedophilia</td>
</tr>
<tr>
<td>Other</td>
</tr>
<tr>
<td>Substance use disorders</td>
</tr>
<tr>
<td>Alcohol</td>
</tr>
<tr>
<td>Cannabis</td>
</tr>
<tr>
<td>Amphetamines</td>
</tr>
<tr>
<td>Opioids</td>
</tr>
<tr>
<td>Cocaine</td>
</tr>
<tr>
<td>Diverse substances</td>
</tr>
<tr>
<td>Other</td>
</tr>
</tbody>
</table>

*1Pervasive developmental disorder: Autism, Aspergers disorder, developmental disorder not otherwise specified.
*2ADHD: All types of attention-deficit disorder; Index offense: In case of more than one index offense, the most serious one is reported, based on Van Nieuwenhuizen et al. (2011).
neurofeedback training. Patients dropped out due to transfers to different treatment facilities \( (n = 1) \) or lack of motivation for the training \( (n = 6) \). Patients who did not complete the training were excluded from further analysis. Although neurofeedback sessions were originally planned to take place 2 times a week resulting in a duration of 10 weeks of training per patient, due to the special setting in which this study took place, the intervention lasted for an average of 16.8 weeks \( (\text{range} = 11–25, \ SD = 4.86) \). Interruptions were due to patients not feeling well enough to complete a training session, aggressive incidents which resulted in temporary separation/placement on a specialized crisis unit, unplanned furlough of patients, and lack of motivation to attend training sessions. The mean number of months in treatment by the start of neurofeedback was 95.3 months \( (\text{range} = 22–247, \ SD = 61.32) \). The large standard deviation was due to one patient who had already spent 247 months in treatment. Patients had an average of 4.6 \( (\text{range} = 2–8, \ SD = 1.64) \) DSM-IV-TR axis I and axis II disorders. The most common diagnosis on axis I was schizophrenia \( (n = 10) \), and Antisocial Personality Disorder \( (n = 7) \), and Personality Disorder Not Otherwise specified (PDNOS) \( (n = 7) \) on axis II.

**Measurements**

**Eeg**

For pre-and post-training measurements, participants received a 5-min resting state 21-channel eyes closed EEG measurement with Nexus-32 hardware and Biotrace + software (Mind Media BV). The EEG was collected from 19 standard 10/20 positions (Herbert & Jasper, 1958), and the right and left mastoid with a sampling rate of 512 samples per second. The left mastoid served as the online reference. Flat type electrodes were placed above and below the left eye and at the outer canthi of each eye to be able to correct for vertical and horizontal eye movements. EEG magnitude across delta \( (0.5–3.5 \text{ Hz}) \), theta \( (3.5–7.5 \text{ Hz}) \), alpha \( (7.5–12 \text{ Hz}) \), beta \( (12–20 \text{ Hz}) \), SMR \( (12–15 \text{ Hz}) \), high beta \( (20–32 \text{ Hz}) \), and gamma \( (32–49 \text{ Hz}) \) frequency bands was assessed.

**Barratt impulsiveness scale-11 (BIS-11)**

The Dutch version of the BIS-11 (Lijffijt & Barratt, 2005), is a self-report questionnaire designed to measure the behavioral and personality construct of impulsivity. It consists of 30 items and is scored on a four-point Likert scale ranging from 1 (rarely/never) to 4 (almost always/always). The total score can be subdivided in three second-order factors: attentional, motor, and nonplanning. The BIS-11 has been shown to be an internally consistent measure of impulsivity among inmate populations (Cronbach’s \( \alpha = .80 \) (Patton, Stanford, & Barratt, 1995).

**Modified desire for alcohol questionnaire—short form (DAQ-SF)**

The Dutch version of the DAQ-SF (Franken, Rosso, & Van Honk, 2003) is a self-report questionnaire measuring the craving for alcohol at the moment of assessment. The short form of the DAQ consists of 14 items scored on a 7-point Likert scale, ranging from 1 (strongly disagree) to 7 (strongly agree). The DAQ-SF has shown to be a reliable measure to assess craving in a substance-dependent population (Cronbach’s \( \alpha = .70 \) (Courtney et al., 2013). All questions in the original version are designed to measure craving for alcohol, however, as the aim of this study was to measure craving for drugs in general, the word alcohol has been replaced by the word drugs. Patients were instructed to focus on their preferred drug of choice and indicate the level of experienced craving on the questionnaire. An extra written instruction was given, indicating that “drugs” can refer to alcohol as well as soft and hard drugs.

**Substance abuse**

To score substance abuse, the item “substance abuse” on the risk assessment scale Historische, Klinische, Toekomst-Revised (HKT-R) (Historical, Clinical, Future-Revised) was used (Spreen, Brand, Horst, & Bogaertsths, 2014). The HKT-R is a clinical risk assessment instrument which was validated in a nationwide population of forensic psychiatric patients (Bogaerts, Spreen, Ter Horst, & Gerlsma, 2018). This item is scored on a 5-point scale ranging from 0–4. The items are scored as follows: 0-no drug use whatsoever, 1-the patient did not test positive for drug use, but did not cooperate with drug testing, 2-the patient had one positive drug testing, and might have also failed to cooperate with drug testing, 3-the patient tested positive for drug use at least twice, but did not refuse to cooperate with drug testing, and 4-the patient tested positive at least twice and also refused to undergo drug testing. The questionnaire differentiates between type of drug (“Soft drugs”, “Hard drugs”, “Alcohol”, and “Other”), and the item is scored on a five-point scale for each type of drug. The highest score on any of the different type of drugs is considered the final score. To score this item, results of regular drug testing as part of treatment facility policy
were used. Positive drug testing is operationalized as any analysis testing positive for either marijuana, and/or psychostimulant/opioid drugs, and/or alcohol. Corresponding with treatment facility policy, non-prescribed medication that is used as recreational drug consumption such as inhaled Methylphenidate (e.g., Ritalin or Concerta), was also scored as a positive drug test, as well as refusal to undergo drug-testing. This item was scored weekly for the ten weeks prior to pre-training measurements, and for ten weeks after post-training measurements.

**Cued go/no-go reaction time task**

The cued Go/No-Go reaction time task is a measure of impulse control. It is a continuous performance task measuring the ability to inhibit a prepotent response (Fillmore, 2003). Participants sit in front of a computer, where blue and green squares are presented in five rounds with a short break in between. A total of 250 targets appear on the screen. Participants are asked to react as fast as possible to a green square but are instructed to inhibit a response when a blue square appears. A go or no-go cue appears before the target, indicating the likelihood of a green or blue target. The likelihood of a correct target is manipulated so that in 80% of the time the cue provides correct information and in 20% an incorrect cue is presented. Cues are presented with a fixed stimulus-onset interval (SOA) of 100, 200, 300, or 400 ms. Outcome measures are the number of commission errors, reflecting the failure to inhibit responses to no-go targets. The cued Go/No-Go reaction time task has been shown to be a valid measure of impulse control in a substance abusing population (Fillmore, 2003).

**Intervention: Neurofeedback**

A standard SMR-enhancement protocol was used, where SMR (12–15 Hz) was up-trained and theta (3.4–7.5 Hz) was down-trained (i.e., inhibited). If excess high beta (20–32 Hz) or delta (0.5–3.5 Hz) was detected, these frequency bands were inhibited as well, with a maximum of three frequency bands being trained in each session. Feedback training was performed on the EEG signal recorded from electrode position Cz against a right ear mastoid reference. Neurofeedback was applied as implemented in the Brainmarker software engine (BrainMarker Device, Brainmarker BV Gulpen). Every training session consisted of a number of feedback rounds, in which patients had to learn to control simple video-games by increasing and inhibiting the EEG signal in the desired frequency bands. Each round of video-game neurofeedback training lasted 60 seconds at a time with a short break in between rounds. Besides videogame based feedback, movie-based feedback was also given. Here, participants had to keep the monitor screen free from black curtains appearing over the displayed movie, resembling training parameter activity. Movie-based feedback lasted 90 seconds per round. Patients received positive feedback once a frequency band was maintained above or below a threshold for 80% of the time, depending on the frequency band that was enhanced or inhibited. Feedback thresholds were adjusted manually, based on how successful a patient was in regulating EEG activity. During treatment sessions, patients were continuously encouraged to try their best and engage in the training. Positive verbal reinforcement was used whenever patients met feedback thresholds.

**Data processing**

As for the analysis of neurofeedback sessions, the last session (session 20) was excluded from the analysis. This session was usually combined with post-training measurements and therefore shorter in duration than the other sessions. Hence, data from sessions 1–19 were used for final analysis.

For each session of neurofeedback, standardized values for all frequency bands were calculated for every round of training. Absolute magnitude values at Cz were calculated per round of neurofeedback within the Brainmarker software and subsequently exported into SPSS. Magnitude values for frequency bands delta (0.5–3.5 Hz), theta (3.5–7.5 Hz), alpha (7.5–12 Hz), SMR (12–15 Hz), beta1 (12–18 Hz), beta2 (18–22 Hz), and gamma (22–30 Hz) were calculated. As the neurofeedback protocol consisted of up-training SMR magnitude while down-training theta magnitude, subsequent analysis focused on SMR and theta magnitude changes across sessions. Training rounds in which magnitude values differed more than two standard deviations from the mean were excluded, as these rounds were most likely influenced by artifacts due to eye blinking and/or movement. The resulting artifact free data was averaged, resulting in a mean and median magnitude per frequency band for each session.

Next, the total number of neurofeedback sessions was divided into time periods (TP). As Weber et al. (2011) argue, due to the high individual variability, no single session median value should be used for analysis, but median magnitudes across three consecutive
sessions can be calculated as a so-called time period (e.g., TP2 is the median value of session two, three and four). As the analysis consisted of 19 sessions, 17 TP’s were created. The median of SMR and theta magnitude of each TP was calculated. Before beginning of the actual training, a one-minute baseline EEG measurement with eyes-closed and eyes-open was performed. However, as cognitive demands required by neurofeedback tasks are quite different from an eyes-open or eyes-closed resting state, it can be argued that these measures do not reflect patients’ baseline median frequency magnitudes adequately. Therefore, TP1 (consisting of the median value of sessions one, two and three) was chosen to serve as the baseline.

Criteria for establishing neurofeedback responders vs. non-responders
This study adapted part of the criteria set out in the study by Weber et al. (2011) for the definition of (non-) responders. They argued that, for responders: (a) EEG magnitudes should change in the desired direction during all training sessions, resulting in a positive mean magnitude change across all sessions; and (b) the increase of mean percentage of EEG magnitude during the training should exceed 8% by the end of training as compared to the baseline state. The increase of 8% or higher given by Weber et al. (2011) was based on their clinical experience, where less than half of the subjects were able to gain a 10% increase in the investigated frequency band. However, since Weber’s et al. developed these criteria based on non-clinical participants, it can be expected that forensic patients have more difficulties in learning to regulate cortical activity and may not achieve magnitude changes in the desired direction during every single session. Therefore, these criteria were somewhat adjusted to fit the population of this study. The definition of neurofeedback responders for this study was as follows.

1. Mean magnitudes of theta and SMR should change in the desired direction for 60% of all training sessions. In this study, 19 sessions of neurofeedback training were used for analysis, resulting in a minimum of 11.4 sessions (we will use 11 sessions). As the neurofeedback protocol applied consisted of downtraining theta while simultaneously uptraining SMR frequency, in order to be qualified as neurofeedback-responders, patients had to show a change in both frequency bands during at least 11 sessions. To test for possible differences between frequency bands (e.g., that patients find the regulation of one frequency band easier than regulation of the other), number of sessions in which only SMR magnitude increased were investigated separately, as were training sessions in which only theta magnitude decreased.

2. Overall, participants had to show an average change in EEG magnitude of 8% in the desired direction (increase of SMR/decrease of theta) by the end of the training as compared to baseline. This criterion proposed by Weber et al. (2011) was not adjusted for the current patient population, as it can be argued that an average change in EEG magnitude of less than 8% might not be clinically relevant anymore.

In order for patients to be qualified as neurofeedback-responders, both criteria had to be met. Next, percentage increase/decrease of median values for each TP relative to baseline was calculated, as well as average increase/decrease over all time periods.

Statistical analysis
To test for changes between pre- and post-training, a Wilcoxon signed-ranks test was performed for behavioral measures DAQ-SF, BIS-11, and Cued Go/No-Go reaction time task.

Multiple linear regression analyses were performed to investigate whether successful regulation of frequency bands could predict changes in the dependent variables BIS-11, DAQ-SF, and commission errors on the cued Go/No-Go reaction time task. For each behavioral measure, two regression models were computed. In the first model, the post-training score was predicted by the pretraining score to determine the percentage of variance of the post-training score that was already explained by the pretraining score. In the second model, the successful regulation of SMR and theta magnitude for 11 or more sessions, the average increase of SMR magnitude and the average decrease of theta magnitude were added subsequently and evaluated by a partial F-test to statistically decide whether this addition contributed significantly to the increase in explained variance.

To test for changes in actual drug use, scoring of the item “substance abuse” of the HKT-R was averaged for the weeks prior to training and the weeks after the training. A paired-sample t-test was performed to test for changes in drug use.
Spearman’s rho correlations were performed to assess whether the number of times patients successfully achieved up or down-regulation in the desired direction was correlated with the height of average increase or decrease in the frequency band. Spearman correlations were also calculated for outcomes on BIS-11, DAQ-SF, amount of commission errors on the Cued Go/No-Go reaction time task, and performance during neurofeedback.

All data were analyzed with SPSS version 22 (IBM Corp).

**Results**

**Responder vs. non-responder**

Results with regard to responding vs non-responding to neurofeedback training were analyzed in multiple ways. Results are given in Table 2.

1. The first criterion for being a neurofeedback-responder stated that the patient showed magnitude changes in the desired direction of frequency bands for at least 60% of the 19 sessions, resulting in magnitude changes during a minimum of 11 sessions. When inspecting sessions in which patients successfully decreased theta magnitude while also simultaneously increasing SMR magnitude, only 4 out of 19 patients (21%) were able to do so for 11 or more sessions. So, according to criterion one, only 4 out of 19 patients could be qualified as a neurofeedback-responder.

To test for differences between frequency bands, sessions in which only theta magnitude changed or only SMR magnitude changed, were also investigated. When inspecting mean decrease in theta magnitude, 7 out of 19 (37%) patients were able to decrease their theta magnitude during 11 sessions or more. For increases in SMR magnitude, 12 out of 19 (63%) patients managed to consistently increase SMR magnitude in 11 or more sessions.

2. The second criterion for establishing neurofeedback-responding stated that, next to an magnitude change in the desired direction for 11 or more sessions, the average magnitude change had to exceed 8%. When inspecting magnitude change in the SMR frequency over all sessions, all 4 responders showing magnitude changes in the desired direction achieved an average increase in SMR magnitude of 8% or higher (range 10%–26%). For theta, 3 of 4 responders showed an average magnitude decrease of 8% or more (range 8%–11%).

Average increase in SMR magnitude of 8% or higher (range 13%–19%) could also be observed for patients who did not manage to simultaneously decrease theta magnitude for 11 or more sessions, but who still managed to increase only SMR magnitude for 11 or more times \( (n = 7) \), except for two patients. For theta magnitude, this was the case for only two patients.

Patients that could not be classified as a responder seemed unable to learn to regulate the targeted EEG-frequencies in the required direction. About a third of the patients (6 out of 19 patients (32%)) showed an increase in magnitude where a decrease was desired, and/or vice versa.

<table>
<thead>
<tr>
<th>Patient</th>
<th>Number of times SMR up</th>
<th>Number of times theta down</th>
<th>Number of times SMR up AND theta down</th>
<th>Average change in SMR in %</th>
<th>Average change in theta in %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>6</td>
<td>0</td>
<td>-4.8</td>
<td>1.5</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>1</td>
<td>0</td>
<td>2.3</td>
<td>5.5</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-5.3</td>
<td>5.1</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>-2.9</td>
<td>2.5</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-2.0</td>
<td>12.9</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>16</td>
<td>0</td>
<td>-24.0</td>
<td>-12.1</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-1.8</td>
<td>10.3</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>18.4</td>
<td>13.7</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>-3.1</td>
<td>8.0</td>
</tr>
<tr>
<td>10</td>
<td>17</td>
<td>6</td>
<td>5</td>
<td>13.5</td>
<td>4.1</td>
</tr>
<tr>
<td>11</td>
<td>16</td>
<td>14</td>
<td>10</td>
<td>4.7</td>
<td>-11.9</td>
</tr>
<tr>
<td>12</td>
<td>12</td>
<td>11</td>
<td>6</td>
<td>2.3</td>
<td>-1.3</td>
</tr>
<tr>
<td>13</td>
<td>17</td>
<td>8</td>
<td>8</td>
<td>19.1</td>
<td>-.73</td>
</tr>
<tr>
<td>14</td>
<td>17</td>
<td>6</td>
<td>5</td>
<td>15.4</td>
<td>1.8</td>
</tr>
<tr>
<td>15*</td>
<td>17</td>
<td>16</td>
<td>16</td>
<td>12.2</td>
<td>-8.5</td>
</tr>
<tr>
<td>16*</td>
<td>15</td>
<td>16</td>
<td>14</td>
<td>15.3</td>
<td>-11.4</td>
</tr>
<tr>
<td>17*</td>
<td>17</td>
<td>15</td>
<td>15</td>
<td>9.5</td>
<td>-8.7</td>
</tr>
<tr>
<td>18*</td>
<td>17</td>
<td>15</td>
<td>14</td>
<td>26.0</td>
<td>-6.7</td>
</tr>
<tr>
<td>19</td>
<td>17</td>
<td>5</td>
<td>4</td>
<td>19.0</td>
<td>1.4</td>
</tr>
</tbody>
</table>

*Patients marked with a * are classified as a responder.
Figures 1 and 2 show the mean increase in SMR and mean decrease in theta magnitude for responders and non-responders over all 17 TPs. Responders showed a steeper learning curve of SMR magnitude increase than non-responders. For theta magnitude, only responders showed a stepwise decrease in theta magnitude while theta magnitude of non-responders fluctuated but remained flat over all 17 TPs.

Spearman’s rho correlations between number of times patients successfully achieved up or down-regulation in the desired direction and height of average increase or decrease in the frequency band was significant at the $\alpha < 0.01$ level in a single frequency (either theta or SMR) (theta: $r = -0.972$, $p < .01$; SMR: $r = 0.924$, $p < .01$), as well as for the number of times SMR and theta were successfully regulated simultaneously (number of times SMR and theta were regulated simultaneously and average success SMR $r = 0.619$, $p < .005$, number of times SMR and theta were regulated simultaneously and average success theta $r = -0.697$, $p < .001$) for the total patient group.

**Behavioral measures**

A Wilcoxon Signed-ranks test indicated that results on the BIS-11 and DAQ-SF were significantly lower post-training (BIS-11: $Z = -2.2$, $p < 0.05$, $r = -0.5$; DAQ-SF: $Z = 1.982$, $p < 0.05$, $r = 0.45$) for the whole group of patients. Commission errors on the Cued Go/No-Go task did not show significant changes between pre and post-training ($Z = -0.6$, $p > 0.05$, $r = -0.14$).

A paired sample t-test showed that there was a significant difference ($p < 0.05$) in degree of positive drug testing post-training ($M = 0.29$, $SD = 0.43$), indicating a decrease in positive drug testing post-training. Separate multiple linear regression analyses were performed to investigate results on DAQ-SF, BIS-11, and number of commission errors post-training based on
various responding criteria. Only significant results are reported; see Table 3 for results.

Discussion

To the best of our knowledge, this is the first study that employed a SMR/theta frequency neurofeedback protocol in forensic psychiatric patients with SUD and co-morbidities. It was investigated whether this complex group of patients is actually able to consistently regulate SMR and theta frequency band activity during a 20 session neurofeedback training and whether changes in these frequency bands were related to changes in level of impulsivity, craving, and actual drug use.

Patients were categorized as responders when they showed both a successful up-regulation of SMR magnitude and a successful down-regulation of theta magnitude during at least 11 out of 19 sessions and showed an average increase/decrease in the desired direction of 8% or higher. Despite the fact that personality measures of impulsivity, level of craving for addictive substances and degree of positive drug testing were significantly lower post-training than pre-training on a group level, only 4 out of 19 patients (21%) could be categorized as a responder to the neurofeedback training. The low number of patients who achieved a responder status was mainly due to the low number of patients who managed to consistently decrease theta magnitude, as the number of times patients successfully achieved up-training of SMR magnitude was much higher. Subsequent analysis showed that the ability to consistently train frequency bands in the desired direction was not related to scores on impulsivity measures post-training. Levels of craving post-training could partially be explained by whether patients could be categorized as responders, but were only related to the increase in SMR frequency and not to a decrease in theta frequency.

Zuberer et al. (2015) differentiate between “EEG-learning” (comparable to the trainability criterion by Zoefel et al., 2011) and “EEG-training response”. “EEG-learning” refers to the improvement in a targeted cortical training parameter in the desired direction, whereas EEG-training response refers to any change in neurophysiological parameters due to neurofeedback training. The results of the current study showed that although all patients showed an EEG-training response, only a fifth of patients showed EEG-learning. This can be considered quite a low number. The difficulties patients seem to experience in downregulating theta as opposed to upregulating SMR activity have been observed in studies with other populations as well. Doppelmayr and Weber (2011) showed that non-clinical individuals were able to increase SMR frequency over the course of training, but failed to decrease the theta/beta ratio. Janssen et al. (2017) employed a theta/beta neurofeedback protocol in children with ADHD, and found theta to remain unchanged during the course of training when inspecting results on a group level, whereas a linear increase was observed for beta activity. However, when investigating individual learning curves in the study by Janssen et al. (2017), the number of participants that could be qualified as responders was much higher than that in the current study: 39% of participants could be qualified as responders with regard to theta, and 53% of participants could be qualified as responders with regard to beta. Research on which patients will be able to benefit from neurofeedback training in terms of EEG-training response is still in its infancy. Therefore, no adequate comparisons can be made with regard to whether 21% of participants is a good representation of EEG-learning abilities in the current patient population. Janssen et al. (2017) also investigated individual learning curves of participants and found that 18% of participants displayed a change of theta in the opposite direction over the course of training whereas only 8% of patients showed a change of beta in the opposite direction. In the current study sample, the number of participants showing learning

Table 3. Multiple regression with results of DAQ-SF post-treatment as dependent variable in model 1, and the average increase of SMR magnitude as predictor in model 2 (n = 19).

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized coefficients</th>
<th>Standardized coefficients</th>
<th>T</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>SE B</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>(Constant)</td>
<td>12.54</td>
<td>8.65</td>
<td>1.45</td>
</tr>
<tr>
<td>2</td>
<td>DAQ-SF T0</td>
<td>1.51</td>
<td>.20</td>
<td>.53</td>
</tr>
<tr>
<td></td>
<td>DAQ-SF T0</td>
<td>19.27</td>
<td>8.13</td>
<td>2.25</td>
</tr>
<tr>
<td></td>
<td>Average increase SMR</td>
<td>-2.53</td>
<td>.23</td>
<td>-2.29</td>
</tr>
</tbody>
</table>

$R^2$ Model 1 = .28.

$R^2$ Model 2 = .46; $F_{part} = .04$.

Note: The t-test in model 2 is identical to the partial F-test, as there is only one variable added as compared to model 1.
curves in the opposite of the intended direction was also much higher: 58% of patients actually increased their theta magnitude as opposed to decreasing it, whereas 36% patients decreased SMR as opposed to increasing it. With only one in five patients being able to achieve EEG-learning, it is questionable whether this intervention will actually be able to lead to clinically significant improvements for a sufficient number of patients.

Several studies have suggested that training outcomes might depend on the number of frequency bands trained, with a negative correlation between the number of trained frequency bands and training success (Rogala et al., 2016). Especially when EEG-frequency bands are adjoining and effects in one frequency might be susceptible to training effects in another frequency band, this could pose a problem. However, SMR and theta frequency bands seem sufficiently far separated from one another to prevent that up-regulation in one band is cancelled out by the down-regulation of the other. It might be possible, however, that for this specific patient population, the training of two frequency bands simultaneously is just too difficult. Future studies with this patient population might benefit from neurofeedback training that is focused solely on enhancing SMR frequency, since improvements on craving measures in this study were related to an increase in SMR magnitude specifically.

Another possible explanation for the low number of responders might be the fact that patients were allowed to continue using prescription medication during the course of the study. It can be considered unethical to ask patients to stop taking medication for the sake of an intervention for which efficacy is not yet established. However, to date, the effects of medication on the trainability of EEG-frequency bands are unclear. It is possible that the effects of medication might ‘overrule’ training effects of neurofeedback. Previous research has shown that stimulant medication can produce a normalization of relative power in the theta band frequency in the resting-state EEG of patients with ADHD (Clarke, Barry, Bond, McCarthy, & Selikowitz, 2002). It is possible that patients with this type of medication might not be able to further normalize theta frequency through neurofeedback. Nonetheless, even if stimulant medication prevents patients from further decreasing their theta-frequency power, additional research is needed to investigate why more than half of the patients in the current study increased their theta frequency when the neurofeedback protocol was aimed at decreasing it.

Future studies need to assess the specific effects of neurofeedback training on the modulation of other EEG-frequency bands in forensic patients, as well as the necessary number of sessions to achieve optimal clinical results. It is possible that more patients would have been classified as responders if more than 20 training sessions had been applied. Strehl et al. (2006) argue that it is not a necessity for participants to show a positive learning curve over each individual session, as some participants might not find an optimal strategy until the end of the training. At the same time, it is of great importance to be able to predict as early as possible in the process which patients will be able to benefit from neurofeedback and which will not. Especially for vulnerable patient populations like forensic psychiatric patients, who are already difficult to engage in therapy as it is, it is important to be able to decide as quickly as possible whether they are likely to benefit from a neurofeedback training in order to not burden them with a therapy modality that they might not be able to benefit from.

**Limitations**

This study did not investigate within-session learning curves of patients. This makes it difficult to interpret results in terms of whether patients were actually able to regulate frequency bands within each session, and to draw conclusions on whether other mechanisms are involved that lead to the consolidation of cortical changes across sessions. Furthermore, it is important to note that, although 21% of patients did show neurofeedback training responses in the desired direction and, as a group, improved on self-report measures of impulsivity and levels of craving, the lack of a sham neurofeedback condition makes it difficult to rule out the possibility of a placebo effect. Furthermore, although participants were continuously encouraged to try their best and engage in neurofeedback training, it cannot be ruled out that some patients were not trying as hard as others. This is a general problem with neurofeedback training. There is no way to be absolutely certain that a patient really does focus on the training, or is just pretending to do so. In the current study, patients in the treatment facility did not receive more privileges due to participating in the study, and it can be argued that it is too hard and possibly too boring to just sit and stare at the monitor for 20 neurofeedback-sessions. Seven patients dropped out during the course of the study, of which six patients dropped out due to lack of motivation. However, given that patients did receive a financial
reward for participation, it is possible that some patients pretended to engage in the training in order to receive a financial compensation. This could be tackled by the use of a sham neurofeedback condition in future studies.

Another limitation is the use of a modified version of the DAQ-SF. While the DAQ-SF itself has good reliability, the modification made to the questionnaire may have influenced the validity and reliability of the questionnaire to some extent.

Another limitation concerns the way the HKT-item “substance abuse” has been assessed. While the questionnaire differentiates between types of drugs (“Soft drugs”, “Hard drugs”, “Alcohol”, and “Other”), the scoring itself does not, as the scoring is based on the highest score given for any of the different types of drugs used. Therefore, the questionnaire is not able to detect changes in type of drug used. In SUD, substance abuse often times is not limited to one specific type of drug, but in many cases concerns polydrug abuse. It can be argued that a patient’s efforts to refrain from using hard drugs such as cocaine is a huge step in the recovery process, even though certain type(s) of soft drugs such as marijuana are still used. With the questionnaire employed in this study, this type of change could not be assessed.

Conclusion

In conclusion, the trainability aspect of an SMR/theta neurofeedback training for forensic psychiatric offenders with SUD could partially be validated, as 21% of patients were able to regulate cortical activity in the desired direction. This study shows the importance of intervention sensitivity and assessment of responders and non-responders to the applied neurofeedback protocol. Additional research is needed to examine possible placebo effects of neurofeedback training and to establish criteria that can predict within a few neurofeedback sessions which patients will likely be able to benefit from neurofeedback and which patients will most likely not benefit from this type of intervention.

References


