AUS DEM LEHRSTUHL
FÜR PSYCHIATRIE UND PSYCHOTHERAPIE
PROF. DR. RAINER RUPPRECHT
DER FAKULTÄT FÜR MEDIZIN
DER UNIVERSITÄT REGENSBURG

NARCOLEPSY VS. IDIOPATHIC HYPERSOMNIA: DIFFERENTIATING GROUPS USING CLUSTER ANALYSIS

Inaugural – Dissertation
zur Erlangung des Doktorgrades
der Medizin

der
Fakultät für Medizin
der Universität Regensburg

vorgelegt von
Emanuel Sitka

2018

AUS DEM LEHRSTUHL
FÜR PSYCHIATRIE UND PSYCHOTHERAPIE
PROF. DR. RAINER RUPPRECHT
DER FAKULTÄT FÜR MEDIZIN
DER UNIVERSITÄT REGENSBURG

NARCOLEPSY VS. IDIOPATHIC HYPERSOMNIA: DIFFERENTIATING GROUPS USING CLUSTER ANALYSIS

Inaugural – Dissertation
zur Erlangung des Doktorgrades
der Medizin

der
Fakultät für Medizin
der Universität Regensburg

vorgelegt von
Emanuel Sitka

2018

Dekan:	Prof. Dr. Dr. Torsten E. Reichert
1. Berichterstatter:	PD Dr. Roland Popp
2. Berichterstatter:	Prof. Dr. Michael Arzt
Tag der mündlichen Prüfung:	22.08.2018

Table of contents

Zusammenfassung

Das Ziel dieser Promotionsarbeit ist es, einige Beiträge zu dem Thema der Differentialdiagnose der Narkolepsie zu leisten. Gemäß der dritten Edition der ICSD (International Classification of Sleep Disorders) ist die Narkolepsie in zwei Untergruppen, Typ 1 und Typ 2, unterteilt. Während Narkolepsie Typ 1 pathophysiologisch auf den Untergang Hypocretin-freisetzender Neuronen im Hypothalamus zurückzuführen ist und klinisch häufig durch das Auftreten von Kataplexie leicht zu identifizieren ist, beruht die Diagnose der Narkolepsie Typ 2 fast ausschließlich auf den Ergebnissen des Multiplen Schlaflatenztests (MSLT). Dadurch gestaltet sich die differentialdiagnostische Abgrenzung der Narkolepsie Typ 2 von der Idiopathischen Hypersomnie als sehr schwierig.

Im Rahmen dieser Doktorarbeit werden verschiedene statistische Methoden eingesetzt, um diese Probleme und ihren Zusammenhang zu den bestehenden diagnostischen Kriterien näher zu beleuchten. Für diese Fragestellung wird ein Datensatz des Schlafmedizinischen Zentrums Regensburg mit insgesamt 141 Fällen von Narkolepsie oder Idiopathischer Hypersomnie herangezogen. Ferner stehen 73 MSLT-Messungen von gesunden Kontrollpersonen zur Verfügung.

Zunächst wird im Rahmen einer linearen Regressionsanalyse die Korrelation zwischen der Häufigkeit von SOREM (sleep onset REM) – Episoden und der MSLT-Schlaflatenz genauer untersucht. Hier stellt sich ein negativer affin-linearer Zusammenhang für beide Narkolepsie-Typen und gesunde Kontrollen heraus. Für den kürzlich von Pizza et al. vorgeschlagenen Parameter Delta, der die Zeit vom ersten Einschlafen bis zum konsolidierten Einschlafen misst, ist sowohl für Narkolepsie als auch für Idiopathische Hypersomnie eine schwache positive Korrelation zur üblichen Einschlaflatenz zu verzeichnen.

Die im Anschluss durchgeführte Hauptkomponentenanalyse erfüllt zwei wesentliche Funktionen im Rahmen der Zielsetzungen dieser Doktorarbeit. Einerseits zeigt die resultierende dominante Hauptkomponente, dass die typischen MSLT- und Polysomnographie-Parameter, die zur Beschreibung und Unterscheidung von Narkolepsie und Idiopathischer Hypersomnie angewendet werden, in der Tat eine wichtige beschreibende Achse darstellen, an der sich die größten Unterschiede im Datensatz aufschlüsseln lassen. Andererseits dienen die erhaltenen

Hauptkomponenten als Grundlage, um algorithmisch einen geeigneten Variablensatz für die Clusteranalysen zu gewinnen.

Für die Clusteranalysen selbst werden drei verschiedene Methoden angewandt. Entsprechend ihrer unterschiedlichen Konzepte zeigen sich deutliche Unterschiede zwischen den erhaltenen Clusterlösungen. Dennoch fallen einige Gemeinsamkeiten auf. So finden alle Clusteralgorithmen stets zwei klinisch relevante Cluster, wobei alle Methoden ähnlich zusammengesetzte Clusterpaare identifizieren. Insbesondere fällt auf, dass die Narkolepsie Typ 2-Fälle algorithmisch in weitgehend konsistenter Weise teilweise der Narkolepsie Typ 1 und teilweise der Idiopathischen Hypersomnie zugeordnet werden. Verschiedene Erklärungen für diese Beobachtung werden angeboten. Einerseits kann auf Phänomene wie einem verzögerten Einsetzen von Kataplexie bei Narkolepsie hingewiesen werden, andererseits können methodische Schwächen des MSLT die Trennschärfe zwischen Narkolepsie Typ 2 und Idiopathischer Hypersomnie reduzieren.

Zuletzt wird ein Interpretationsansatz weiter ausgeführt, der entsprechend den Ergebnissen der Clusteranalysen zwei statt drei diagnostische Gruppen für den Datensatz vorschlägt, wobei diese beiden Gruppen entstehen, indem die Narkolepsie Typ 2-Fälle in der beobachteten Weise der Narkolepsie (ohne Subtypen) bzw. der Idiopathischen Hypersomnie zugeschrieben werden. Im direkten Vergleich zeigt sich diese alternative diagnostische Einteilung gut hinsichtlich der Clustervariablen nachvollziehbar. Diese Arbeit endet mit dem Vorschlag eines aus der dominanten Hauptkomponente abgeleiteten diagnostischen Scores, der die Differentialdiagnose insbesondere dieser neuen Kategorien zu erleichtern scheint.

Narcolepsy vs. idiopathic hypersomnia: Differentiating groups using cluster analysis

1. Introduction

Narcolepsy is a disabling sleep disorder, which – despite its relative rarity – has always attracted the interest of many sleep medicine researchers. Much progress has been made in understanding the etiology and pathophysiology of this condition. In case of narcolepsy type 1 a specific loss of certain neurons in the hypothalamus has been found to be the morphologic correlate of this disease[1]. However, until today, no curative treatment option is available, but the majority of patients show significant improvement under the recommended medication.

Some cases of narcolepsy are easily diagnosed due to very suggestive MSLT (multiple sleep latency test) results or a clear-cut history of cataplexy. For a significant number of patients, however, diagnosis is a much more challenging task. One of the main issues in these situations is the distinction between patients with narcolepsy type 2, who by definition do not show cataplexy, and patients with idiopathic hypersomnia (IH). Until today, despite the recent update of the ICSD (International Classification of Sleep Disorders), the definition and diagnostic criteria of these conditions are heavily discussed. Additionally, patients subsumed under the diagnosis of IH show high clinical variance and heterogeneity[2, 3]. Furthermore, in terms of clinical presentation, a distinction between narcolepsy and IH is often not possible. Due to the absence of biomarkers, the diagnosis and differential diagnoses are based on the results of the MSLT. Since the MSLT is known to have several methodical weaknesses, scientists and clinicians remain in an unsatisfactory situation.

This thesis addresses some of the issues listed above. Using linear regression analysis, the correlation between different MSLT parameters is examined. Emphasis is put on two main aspects: On the one hand, the correlation between the SOREM (sleep onset REM) count and the mean sleep latency is investigated. On the other hand, focus lies on the sustained sleep latency and its possible diagnostic purpose, for example via the parameter Delta which has recently been suggested by Pizza[4].

The core of this thesis consists of a principal component analysis and subsequent cluster analyses. Using these descriptive statistical tools, the diagnostic and differential diagnostic value of several sleep medical parameters is addressed. Furthermore, the cluster analysis results serve as foundation for the discussion, whether the current diagnostic entities, i.e. the narcolepsy subtypes and IH, actually form separate clusters and whether the results of this thesis justify the current classification system. Three entirely different cluster algorithms are employed, yielding three different cluster solutions. In order to illustrate and quantify both similarities and differences between these solutions, various methods for cluster validation and interpretation are presented.

In the final chapter of this thesis, all results of the different methodical approaches are summarized. The general discussion further examines the question, whether the conventional classification into three groups as stated in the ICSD-3 should be challenged by the consistent finding of two essential groups reported by the cluster analyses. Eventually, a linear score is introduced as a suggestion on how the differential diagnostic process could be refined.

2. Hypersomnolence and EDS: Definition and diagnostic concepts

According to the ICSD-3, hypersomnolence is the occurrence of excessive sleepiness[5]. The conditions that are discussed in this thesis are usually accompanied by excessive daytime sleepiness (EDS), which is defined as " […] the inability to stay awake and alert during the major waking episodes of the day, resulting in periods of irrepressible need for sleep or unintended lapses into drowsiness or sleep"[5]. On the contrary, the term hypersomnia should only be used to describe conditions that may cause hypersomnolence/excessive sleepiness.

Excessive sleepiness and EDS should be treated as multidimensional concepts[6–8], which cannot be properly detected and quantified by a single diagnostic procedure. Both subjective and objective assessments of EDS are important in the clinical context. In the following, all diagnostic tools that are of further relevance for this thesis will be described.

2.1. The ESS: A subjective measurement of EDS

The Epworth Sleepiness Scale (ESS) is a self-administered questionnaire that was introduced in 1991 by Johns[9]. According to Johns, the MSLT merely allows an estimation of a very specific situational sleep propensity for the MSLT setting without addressing the general condition of EDS/sleep propensity[10]. In contrast to the other diagnostic tools discussed below, the ESS was designed to measure the general sleep propensity by explicitly exploring different situations in which sleepiness might occur.

The ESS consists of eight items, which represent different situations in everyday life (e.g., "watching TV"). For each item, the patient estimates the likeliness to fall asleep or to doze off in the given situation using the numbers 0 to 3. A score of 0 means that the patient never dozes off or falls asleep, whereas 3 represents a high chance of doing so. The individual item scores are added, resulting in a total score ranging between 0 and 24[9].

In the original publication by Johns, the scores of healthy controls ranged from 2 to 10 with an average value of 5,9 , whereas all IH and narcolepsy patients in this study showed ESS scores higher than 12, sometimes reaching values of 20 or more[9]. A normative study regarding the German version of the ESS found an average value of 6,6 and recommended regarding ESS scores higher than 10 as "clinically suspicious" and scores higher than 12 as "clinically relevant"[11].

The ESS has been validated well for different language versions[12–14], for which also the matter of internal consistency has been investigated[12, 15]. Indeed, high levels of consistency have been found, indicating that all items do address the same theoretic construct. Factor analysis showed the existence of one dominant factor which was interpreted as general sleep propensity by Johns[15]. Furthermore, most studies found an acceptable test-retest reliability[12, 13], although a low reliability has been reported in a population being evaluated for sleep-related breathing disorders[16].

Due to its simplicity, the ESS is widely established in clinical practice. However, its diagnostic and differential diagnostic value has been subject of discussion. A brief overview of this topic will be given in section 3.3.3. .

2.2. Objective measurements of EDS

2.2.1. EDS and vigilance tests

Vigilance can be defined as "more careful attention, especially in order to notice possible danger"[17] and is therefore not directly linked to EDS. However, vigilance is needed in many tasks of everyday life, e.g. work or traffic. It is to be expected that patients suffering from severe EDS may also have an impaired vigilance, so it is reasonable to measure vigilance in hypersomnia patients.

In this thesis the term vigilance test will always refer to the Quatember-Maly test. This test, which is included in the test collection Wiener Testystem[18, 19], is a digitalized version of the clock-test, which has initially been designed by Mackworth[20].

In the Quatember-Maly test the patient is instructed to follow a light source which is moving along a circle, like the second hand of a clock. In irregular intervals a wider jump of the light source occurs, which the patient has to report as quickly as possible. As test results, among other parameters, the number of correct reactions, the number of false reactions (i.e. a jump is reported that has not occurred) and the average reaction time for correct reports are obtained.

According to the test manual the number of correct reactions is the most direct measure for visual vigilance, whereas many false reactions indicate that the patient did not comprehend the test instructions or did not take the test serious[21].

2.2.2. EDS evaluation using polysomnography (PSG)

According to the ICSD-3 criteria for IH and narcolepsy, the results of the nocturnal PSG serve two important diagnostic purposes. Mainly, the PSG is needed for the exclusion of other causes of hypersomnolence[22], e.g. sleep related breathing disorders. Furthermore, if a REM sleep episode occurs in the first 15 minutes of the PSG, it is treated as an equivalent to a SOREM (sleep onset REM) episode in the MSLT, including all implications for the (differential) diagnostic process.

However, some publications indicate that there might be PSG parameters that could be of additional use for the diagnosis and differential diagnoses of IH and narcolepsy. These parameters will be discussed further in section 3.3. .

Regarding this thesis, several PSG parameters will be included in the upcoming analyses: Any SOREM episodes during the PSG are taken into account indirectly by the (PSG) REM latency. The sleep efficiency index, which is defined as the fraction of the total time in bed that is spent asleep, is also included. As a measure of sleep quality, the fraction of the total amount of sleep that is spent in sleep stage N3 is considered. Furthermore, the arousal index, which does not differentiate different causes of nocturnal arousals, is used as an estimate of sleep fragmentation.

2.2.3. The MSLT

The MSLT was designed as a tool for objective measurement of sleepiness in a standardized environment. In principle, the MSLT is based on the findings of Rechtschaffen and Kales, who were able to define EEG criteria for the human sleep architecture[23]. Before that, sleepiness could only be measured by observing subjects directly or by employing subjective questionnaires.

According to Arrand et al.[24] in the seventies of the 20th century several studies were performed investigating the sleep behavior of subjects in a 90 minute day[25–27], i.e. 60 minutes of activity and 30 minutes of rest using EEG criteria discovered by Rechtschaffen and Kales. It was found that the subjective sleepiness measured by the SSS (Stanford Sleepiness Scale) showed a strong correlation with the sleep latency of these subjects. Therefore, it was concluded that in a situation like the MSLT test environment, which does not allow any alerting stimuli, the physiological sleep tendency is unmasked and can be objectively measured using these sleep latencies[28].

In 1977, the MSLT was first used to assess sleepiness in an experimental setting[29]. In 1979, the MSLT was used for the first time to detect REM episodes in narcolepsy patients[30].

The guidelines formulated by Carskadon et al. in 1986[31] and more recently by Littner et al.[32] describe in a precise manner how the MSLT should be performed. Table 2.1 summarizes these guidelines.

Table 2.1: Essential guidelines for the multiple sleep latency test (MSLT) according to Carskadon et al.[31] and Littner et al.[32]

General considerations

- Performance of the MSLT on the day following a NPSG, no MSLT if the total sleep time was less than 6 hours
- Sleep diaries 1-2 weeks preceding the MSLT
- Consideration of drug intake; withdrawal two weeks before the MSLT
    - Drugs affecting sleep latency: stimulants, hypnotics, sedatives, antihistamines
    - Drugs affecting REM latency: tricyclic antidepressants, MAO inhibitors, amphetamines

Test settings

- Five nap opportunities in 2-hour intervals, beginning 1,5 to 3 hours after the end of nocturnal sleep
- Quiet bedroom, constant and low light level, constant room temperature
- No ingestion of alcohol or caffeine during the whole day
- Between the naps, patients should be out of bed and prevented from sleeping.

Test procedure

- Recording of EEG (C3-A2, C4-A1, O1-A2, O2-A1 derivations), left and right eye electrooculograms, mental or submental EMG and ECG; bio-calibrations preceding each nap
- Each nap opportunity starts with the instruction "Please lie quietly, assume a comfortable position, keep your eyes closed and try to fall asleep." Then, the lights are turned off.
- Each nap session is terminated after 20 minutes if no sleep occurs. In this case, a sleep latency of 20 minutes is noted.
- If the patient falls asleep within the first 20 minutes, the nap session continues for another 15 minutes starting from the first sleep epoch.

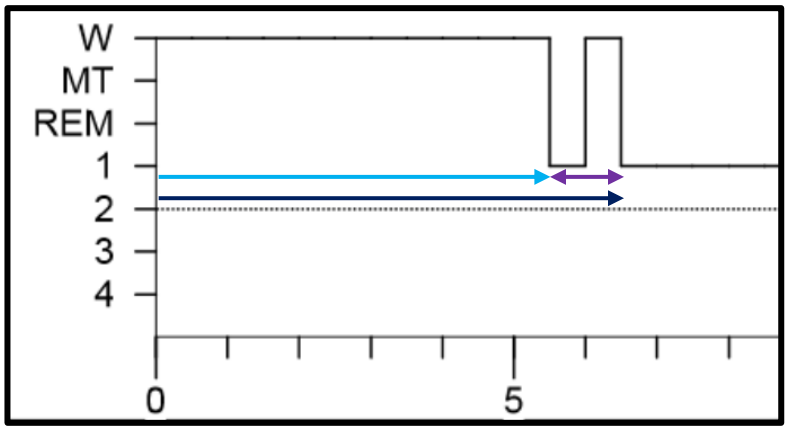
The main diagnostic results of a MSLT test are several different sleep latencies which have been calculated as the average value of the five sleep opportunities. For each sleep opportunity, the sleep latency is defined as the timespan starting from the closing of the eyes until the first episode of a prespecified sleep stage is recorded. Most of the time and in the ICSD-3 criteria for narcolepsy and IH, sleep latency refers to the timespan until the first episode of sleep stage N1 occurs.

Apart from the usual sleep latencies the concept of the sustained sleep latency will become relevant in this thesis. According to Pizza et al.[4] sustained sleep latency is defined as the timespan until unequivocal sleep is reached, which is defined as at least three consecutive periods of sleep or one

period of a sleep stage different than N1. In the same publication the parameter Delta has also been introduced, which is defined as the difference between the sustained sleep latency and the conventional sleep latency to stage N1[4]. Furthermore, the total sleep time in all five sleep opportunities will be considered as an important parameter later on.

Diagram 2.1: Sleep latency, sustained sleep latency and Delta

The arrows below demonstrate the sleep latency to sleep stage N1, the sustained sleep latency and Delta for a single MSLT nap opportunity. As final test results, the average values with respect to all five nap opportunities are considered.



Apart from the sleep latencies, the MSLT also allows an estimation of REM proneness using the SOREM (sleep onset REM) count. SOREM episodes are defined as REM sleep episodes occurring during the first 15 minutes after sleep onset[33]. For the SOREM count the number of sleep opportunities that have shown a SOREM episode are added up. Therefore, the SOREM count value can range from zero to five.

In this thesis, data obtained by a modified version of the MSLT, the MSLT30, will be evaluated. Compared to the standard procedure, the only difference is the rule that all five nap opportunities terminate exactly after 30 minutes, regardless of the observed sleep pattern. The major advantage of this test variant is that no real-time assessment of sleep stages is necessary. Further details regarding the MSLT30 can be found in the publications of Geisler et al.[34, 35].

As it is reflected in the current diagnostic criteria, the MSLT has become an essential tool in the diagnosis of narcolepsy and IH. Richardson et al. found in 1978 that narcolepsy patients tended to fall asleep much earlier than healthy controls[36], which was an early justification for using the MSLT in the diagnostic process. Both Drake et al. and Chen et al. report an "excellent" inter- and intrarater reliability of the MSLT[37, 38]. There are some studies that indicate a problematic test-retest reliability of the MSLT[39, 40], which will be discussed later in further detail.

Caution is needed for the interpretation of MSLT results, since it is only useful when combined with clinical findings or other diagnostic results[32, 41]. It is important to consider the age dependence of both the sleep latency and the SOREM frequency. Geisler et al. reported a quadratic dependence of the former from the age of the subject, reaching its minimum in middle-aged subjects[35], whereas Dauvilliers et al. found an age-dependent decrease in the SOREM frequency[42].

According to the review of Arand et al., the MSLT is thought to measure the physiological sleep tendency in the absence of alerting factors[24]. However, some concern has been raised if this is indeed the case. For example, Harrison et al. discuss a certain group of individuals showing low sleep latencies but no other signs of subjective or objective sleepiness[43]. This phenomenon is called "high 'sleepability' without sleepiness" by the authors, raising the concern that the attribute quantified by the MSLT is merely the sleepability of the patients.

Johns, who initially introduced the ESS into clinical practice, discussed the terms "general vs situational sleep propensity", stating that the MSLT just measures the latter in a very specific situation[44]. A recent critical comment of Mayer et al.[45] on a study by Goldbart et al.[40] discusses the question if the MSLT is indeed a suitable tool for the differential diagnosis of EDS. The authors raise the provocative question if the MSLT, often reporting unclear or inconclusive results, might be the very reason because of which the category narcolepsy without cataplexy/type 2 might have been introduced.

These findings highlight the fact that "pathological MSLT results" by themselves do not justify any diagnosis but have to be interpreted in the context of the clinical symptoms. Further issues regarding subtleties in diagnosis and differential diagnosis of narcolepsy and IH will be discussed below.

3. Narcolepsy and idiopathic hypersomnia

Before discussing the actual statistical methods and results of this thesis, a brief overview of the diseases addressed in this thesis will be given. After that, current issues in their diagnosis and differential diagnosis will be discussed, which will eventually lead to the motivation of the central aims of this thesis.

3.1. Narcolepsy

The term "narcolepsy" derives from the classical Greek words *νάρκη (nárkē)* and *λῆψις (lepsis)* and can be translated to "attack of numbness". It has first been used by the French physician Jean-Baptiste-Édouard Gélineau[46], who has published one of the first articles about narcolepsy in 1881. Four years before that, Karl Friedrich Otto Westphal had published two case reports about patients showing typical symptoms of narcolepsy[47, 48]. More than 130 years later, having learned about SOREM episodes[49], HLA genotypes[50] and hypocretin[51], our understanding has vastly increased. However, until today, narcolepsy can most easily be characterized by the narcoleptic tetrad, consisting of hypnagogic hallucinations, sleep paralysis, excessive daytime sleepiness (EDS) and cataplexy[33].

Table 3.1: The narcoleptic tetrad

Symptoms	**Description/Comment**
Hypnagogic hallucinations	Hallucinations during the transition from wakefulness to sleep
Sleep paralysis	Paralysis for a brief time after awakening
Excessive daytime sleepiness	Often leading to an irresistible urge to go to sleep
Cataplexy	Sudden loss of muscle tone after exposure to emotional triggers; not present in narcolepsy type 2/without cataplexy

Hence, the typical patient suffering from narcolepsy shows a severe urge to go to sleep during daytime, which often cannot be resisted. During the transition from wakefulness to sleep, he encounters hallucinations. Furthermore, having awakened from sleep, these patients often experience an inability to move, which usually vanishes after some minutes. Finally, if these patients are exposed to certain emotional stimuli, a sudden loss of muscle tonus occurs, leading to a sudden fall. However, there are several patients suffering from most of these symptoms but cataplexy, which has led to the definition of two subgroups of narcolepsy: narcolepsy with and

without cataplexy according to the ICSD-2[52], or almost equivalently narcolepsy type 1 and type 2 according to the ICSD-3[5].

3.1.1. Epidemiology

Narcolepsy is a rare sleep medical disease. Its prevalence varies greatly between different ethnics. The highest prevalence of 0,16 % has been observed in Japan[53], whereas in Israel, only 0,0002 % of the population are affected[33, 54]. In European countries narcolepsy has an intermediate prevalence value. According to Akintomide et al.[33] it varies between 0,02 % and 0,05 %, which complies with a prevalence of 0,047 % reported by Ohayon et al.[55]. For Olmstedt County in Minnesota, USA a prevalence of 0,0563 % was reported by Silber et al[56]. The incidence for the latter population was 1,37/100000 per year[56].

The incidence rate is highest in the second life decade, with a mean age of onset of about 24 years as reported by Dauvilliers et al.[57]. However, in this reference two age peaks for narcolepsy onset have been found, one around the age of 14,7 years, the second one at the age of 35. Also, narcolepsy has been found to be more common in men[33]. Another interesting finding is that the narcolepsy onset is strongly seasonal, reaching its maximum in April to July in Beijing in China, as reported by Han et al. .[58] In this article it was also found that the incidence of narcolepsy was – with a delay of 5-7 months – correlated to the occurrence of upper airway infections and the H1N1 pandemic in China in 2009. Also, the month of birth seems to affect the individual risk for developing narcolepsy: Dauvilliers et al. found an odd ratio of 1,45 for a birth in March and 0,63 for persons born in September[59].

Common comorbidities of narcolepsy are PLMS (periodic limb movement in sleep), sleep talking, REM sleep behavior disorders and depression[60].

3.1.2. Hypocretin, HLA DQB1-0602 and the etiology of narcolepsy

In recent years, a convincing body of evidence has been collected suggesting that the primary cause of the narcoleptic symptoms is a hypocretin deficiency in the human brain.

Hypocretins or orexins are neuropeptides that are physiologically produced by a group of neurons in the lateral hypothalamus[60]. As Peyron et al. showed in an immunohistochemical study of rat brains, these neurons have very widespread projections across the brain and are involved in the regulation of food intake, blood pressure, body temperature and also the sleep-wake-cycle[61]. The

latter aspect was more highlighted by the findings of Chemelli et al. in 1999, who demonstrated that orexin knockout mice suffered from symptoms which are very similar to narcolepsy [62]. Furthermore, the results of Lin et al. in 1999 indicated that the canine version of narcolepsy is caused by an altered hypocretin receptor gene [63].

Motivated by these results, Nishino et al. examined the cerebrospinal fluid of nine patients with narcolepsy type 1. In seven out of nine patients, no hypocretin could be detected, whereas all control patients had hypocretin levels above 250 pg/ml in their cerebrospinal fluid (CSF)[51]. Thannickal et al. conducted a post-mortem brain tissue study of four narcolepsy patients (three of which had shown cataplexy) and found a 85-95% reduction in the number of hypocretin releasing neurons in the hypothalamus[1]. Furthermore, gliosis was detected in the corresponding hypothalamic regions, indicating a degenerative process which may have caused the loss of neurons. Additionally, as Thannickal was able to demonstrate in another study, at least some cases of narcolepsy without cataplexy are associated with a more localized and less severe loss of hypocretin neurons, which did not affect neurons in the anterior hypothalamus[64]. Further studies confirmed the high specifity of undetectable low CSF levels for narcolepsy with cataplexy[65, 66], but also pointed out that there are other causes that might explain reduced CSF hypocretin levels such as central nervous system (CNS) inflammation, trauma and Guillain-Barré-Syndrome[33, 67, 68]. Whereas the connection between narcolepsy type 1/with cataplexy and hypocretin deficiency is well established, empiric evidence hints at a more complex situation for narcolepsy type 2/without cataplexy. Krahn et al. reported that considering the total average, narcolepsy without cataplexy patients also show lowered CSF hypocretin levels. These levels however are significantly higher than those in narcolepsy with cataplexy. In this study, patients were also tested for the HLA DQB1*0602 allele, which revealed that only the HLA positive subgroup of narcolepsy without cataplexy had significantly lowered hypocretin levels[66]. Similarly, Mignot et al. found that only few patients diagnosed with narcolepsy without cataplexy had reduced hypocretin levels[69].

This apparent heterogeneity of narcolepsy type 2/without cataplexy was further investigated by Andlauer et al., who found lowered hypocretin levels in 24% of patients in a large collective of narcolepsy without cataplexy patients. They also observed that a delayed onset of cataplexy almost exclusively occurred in this subgroup[70].

Even before these discoveries regarding hypocretin, a lot of evidence had been gathered showing that most narcolepsy patients had a HLA DQB1*0602 genotype. Conversely, being positive for HLA DQB1*0602 turned out to be a major risk factor for developing narcolepsy[71]. Again, a higher correlation was reported for narcolepsy with cataplexy[33, 72]. In patients with narcolepsy without cataplexy HLA DQB1*0602 can only be found in 40-60%[60] and being positive for HLA DQB1*0602 also correlates with reduced CSF hypocretin levels[70].

While according to Aran et al. HLA DQB1*0602 is beneficial in the situation of a septic shock due to a streptococcus infection[73], there is some evidence indicating that precisely this improved protection against an acute streptococcus infection might also be involved in the pathogenesis of narcolepsy: Aran found that anti-streptococcal antibodies are elevated in DBQ1*0602 positive narcolepsy patients[73], whereas Koepsell et al. published data which suggests that in patients having a history of a strep throat before the age of 21, narcolepsy is more than five times more common[74]. The connection to upper airway infection was also discussed by Han et al., who reported a time-delayed correlation between infection frequency and narcolepsy incidence[58]. Other environmental factors have been discussed, such as H1N1 vaccinations, head trauma or exposure to toxic agents[33].

In conclusion, the current knowledge about the etiology and pathogenesis of narcolepsy could be summarized like this: Environmental factors, which have not been certainly identified, lead to an autoimmune reaction, which especially affects hypocretin positive neurons in the hypothalamus[60]. Individuals who are positive for HLA DQB1*0602 are more likely to develop this reaction[75], which eventually leads to the degeneration of the involved neurons. As a result, hypocretin is lacking in many areas of the brains, which causes the typical symptoms of narcolepsy[60].

However, this model is far from comprehensive. There are some patients suffering from narcolepsy with cataplexy, who do not have lowered CSF hypocretin levels[65]. In reality more complex genetic interaction are thought to be responsible for the development of narcolepsy[76].

Various other findings have been published showing other abnormal findings in narcolepsy patients. Several amino acids concentrations seem to be altered in the CSF of narcolepsy patients[77]. The COMT genotype of patients has an impact on the disease severity[78]. Histamine neurons are increased in patients having narcolepsy with cataplexy[79], whereas the histamine CSF concentration is lowered[80]. The hippocampal volume is reduced in narcolepsy patients[81]. Other genetical factors have been identified, for example the T-cell receptor alpha polymorphism[82]. These results indicate

that despite some major breakthroughs there is still no comprehensive pathophysiological model which could explain all cases of narcolepsy, especially those without cataplexy.

It is important to consider that the symptoms of narcolepsy can also be caused by other CNS pathologies. For an example, brain lesions due to tumors, inflammation or ischemia may cause secondary narcolepsy[60]. Interestingly, lesions in the hypothalamus lead to the complete phenotype of narcolepsy, whereas pathologies in non-hypothalamic regions are more likely to cause cataplexy without the other symptoms of the narcolepsy tetrade[83]. Also Niemann-Pick disease of type C and muscular dystrophies can induce symptomatic narcolepsy[60].

3.1.3. Clinical aspects

Before discussing the current diagnostic criteria, a brief overview of the typical symptoms of narcolepsy will be given. Both the variety at which the four core symptoms of narcolepsy may present themselves, and the fact that not in all patients the whole narcoleptic tetrad can be observed[84, 85] contribute to the persisting issues in diagnosis and differential diagnosis.

3.1.3.1. Cataplexy

Cataplexy can be defined as "rapid eye movement (REM) sleep atony occurring at an inopportune moment", as it is stated by Overeem et al.[86]. In principal, all striated muscles can be affected by cataplexy, with the diaphragm being the only exception[60]. The most typical forms of cataplexy are sagging of the jaw and trembling of knees.[87] Using video recording, Rubboli et al. could identify three phases of cataplexy: The initial phase, the falling phase and the atonic phase. In particular, these findings suggest that apart from the negative atonic component of cataplexy there also seem to be positive motoric phenomena, possibly reflecting different aspects of motoric signs of REM sleep.[88]

Common triggers of cataplexy are joy, happiness, surprise and anger, in rare cases attacks of cataplexy are also triggered by sports, sudden noises and tickling[87]. Sturzenegger et al. have also noted that the presence of persons known to the patients increase the frequency of cataplectic attacks[87]. Cataplexy usually lasts from less than one second up to several minutes. Some patients encounter less than one episode per year, whereas others suffer from cataplexy many times each

day.[60] Poryazova et al. have described a special variant of "rebound" cataplexy occurring most often on withdrawal from the antidepressant fluoxetine which acts anticataplectic[89].

While cataplexy is thought to be specific for narcolepsy, cataplexy-like symptoms can also occur in non-narcoleptic individuals. According to Sturzenegger et al. these symptoms are more atypical regarding the affected muscle groups and less pronounced than attacks of clear cut cataplexy. Additionally, they are more frequently observed in patients with hypersomnolence[87].

Knudsen et al. report that a hypocretin deficiency leads to more pronounced cataplectic symptoms[90]. Combined with the findings of Heier et al.[65], who reported a strong association between HLA DQB1*0602, cataplexy and low hypocretin levels, one may conclude that cataplexy is most likely and most severe in "typical narcolepsy patients" and that cataplexy – possibly with a delayed onset – is a direct consequence of the pathophysiological pathway described above.

3.1.3.2. EDS, Hypnagogic hallucinations and sleep paralysis

Compared to cataplexy, the occurrence of excessive daytime sleepiness is much less specific for narcolepsy. On the contrary, a lot of conditions can be responsible for this common symptom[91]. However, there are some features of EDS which help to simplify the differential diagnosis. Typically, EDS resulting from narcolepsy leads to an almost irresistible urge to nap[60, 87]. The resulting naps are usually short and tend to be refreshing for the patients[92]. For some narcolepsy type 2 patients, spontaneous improvement of EDS has been observed[93].

Hypnagogic hallucinations are defined as hallucinations occurring during sleep onset. Despite of being part of the narcolepsy tetrad, they are also not uncommon in healthy individuals[94]. Usually, they are of visual or auditory nature[95], but can also include physical sensations[60].

Sturzenegger et al. report that sleep paralysis occurs in about 50 % of narcolepsy patients[87]. Episodes of sleep paralysis usually last for a few seconds but can also persist for some minutes[60]. Often, sleep paralysis occurs together with hypnagogic hallucinations[95]. Sleep paralysis is also not specific for narcolepsy and may also occur isolated in an otherwise healthy population as a study by Bell et al. indicates[96]. Furthermore, Knudsen et al. pointed out that sleep paralysis and hypnagogic hallucinations are typical properties of both narcolepsy type 1 and type 2, suggesting neuronal pathways that are not affected by hypocretin may be involved in these symptoms[90].

3.1.3.3. Other symptoms of narcolepsy

Another typical symptom of narcolepsy is automatic behavior during daytime, which has been hypothesized to be caused by microsleep episodes[60, 97]. Furthermore, despite the severe EDS of many narcolepsy patients, their night sleep is often disrupted by many awakenings[60].

Akintomide et al. describe various minor symptoms of narcolepsy such as blurry vision and loss of concentration and memory[33]. It has also been observed that narcolepsy patients have an increased BMI on average[98].

Other symptoms of narcolepsy patients may stem from the various comorbidities that have been described. Empirical evidence suggests that narcolepsy is often associated with PLMS[60], REM behavior disorders[99], sleep-related breathing disorders[84] and depression[100].

3.1.4. Diagnosis

According to the ICSD-3 narcolepsy can be diagnosed considering the following aspects:

Anamnesis should reveal some history of an irresistible urge to sleep or, consequently, episodes of unwillingly falling asleep. Then, for narcolepsy type 1 cataplexy is major diagnostic factor, whereas the diagnostic criteria for narcolepsy type 2 list cataplexy as an exclusion criterion. In order to objectively measure the EDS, the MSLT must be used for the definite diagnosis.

The third edition of the ICSD-3 includes another procedure for diagnosing narcolepsy type 1: The measurement of CSF concentration of hypocretin-1[5]. This new criterion takes into account the nowadays widely accepted concept that narcolepsy type 1 is caused by hypocretin deficiency[101]. The CSF hypocretin concentration is measured by using a radioimmunoassay. A hypocretin concentration below 110 ng/ml or below one third of the average concentration in the general population is thought to be highly specific for narcolepsy[5]. However, as Knudsen et al. have remarked, measuring the CSF concentration of hypocretin-1 is of limited use in practice, since hypocretin deficient narcoleptics usually show a very severe phenotype, which allows an easy diagnosis even without referring to CSF hypocretin concentrations[90].

Table 3.2: Diagnostic criteria for narcolepsy type 1 according to the ICSD-3[5]		
Both criteria A and B must be true.		
A	History of daily periods of irresistible need to sleep or actual lapses to sleep for *at least three months.*	
B	At least one statement of the following two has to be true.	
	B1	Occurrence of cataplexy *and* a mean sleep latency *less than 8 minutes* in the MSLT *and at least two* SOREM episodes in the MSLT. One SOREM episode can be replaced by one found in the preceding NPSG.
	B2	The CSF concentration of hypocretin-1 is either below 110 pg/ml or less than a third of the average values obtained from control subjects using the same standardized assay.

Table 3.3: Diagnostic criteria for narcolepsy type 2 according to the ICSD-3[5]	
All criteria A to E must be met.	
A	History of daily periods of irresistible need to sleep or actual lapses to sleep for *at least three months.*
B	A mean sleep latency *less than 8 minutes* in the MSLT *and at least two* SOREM episodes in the MSLT. One SOREM episode can be replaced by one found in the preceding NPSG.
C	Cataplexy is absent.
D	The CSF concentration of hypocretin-1 has not been measured *or* is either above 110 pg/ml or higher than one third of the average values obtained from control subjects using the same standardized assay.
E	The hypersomnolence and MSLT results are not better explained by other causes.

Comparing the two tables, it becomes clear that narcolepsy type 2 is diagnosed by ruling out type 1 and other diseases which may cause similar symptoms. Furthermore, being deprived of both CSF measurement and cataplexy as positive diagnostic criteria, the diagnosis of narcolepsy type 2 heavily relies on clinical assessment. On the contrary, it is quite easy to follow the diagnostic criteria for narcolepsy type 1 and therefore to decide if a patient really has narcolepsy type 1.

The diagnostic uncertainty for narcolepsy type 2 is a major aspect of this thesis, and further complications arise if one has to consider the differential diagnosis of idiopathic hypersomnia.

3.1.5. Treatment and prognosis

Until today no curative treatment option is available for narcolepsy. Therefore, the therapy of narcolepsy aims at minimizing and controlling the main symptoms. There are several pharmacological treatment options that target different core symptoms of narcolepsy. In general, stimulants are used to treat narcoleptic EDS, whereas antidepressants show some effect on cataplexy and sodium oxybate affects both EDS and cataplexy.

Apart from that, several other drugs have also been used: Selegiline, a selective irreversible MAO-B inhibitor, proved some efficacy in the treatment of EDS and cataplexy[102], but is only recommended as an "option" in the treatment of narcolepsy due to possible diet and drug interactions. Pemoline has been a treatment option for narcoleptic EDS[103], but has fallen into disrepute since liver toxicity has been shown to be a rare but potentially lethal side effect[104]. Furthermore, there is class II evidence suggesting that mazindol has a beneficial effect on sleepiness of narcoleptics[105]. Finally, the 5HT2 antagonist ritanserin is also regarded as a treatment option[104], acknowledging the results of two studies which indicate a beneficial effect on subjective sleepiness[106] and sleep quality[107].

Stimulants

The standard drug used for the treatment of EDS in narcolepsy is the stimulant modafinil[60]. While the precise mechanism of action is still unknown, it is believed that modafinil affects neuronal pathways of the neurotransmitters histamine, noradrenalin and dopamine[33]. The efficacy of modafinil, administered at a dose of 200-400 mg/d, has been shown in several studies. Billard et al. provide a summary of the most important results[108], whereas Golicki et al. have conducted a meta-analysis including 1054 patients in total[109]. In the latter, significant improvements are reported regarding EDS measured by the MSLT, the MWT (maintenance of wakefulness test) and the Epworth sleepiness scale. Improvements in the quality of life in terms of the SF-36 have also been observed. There are also some positive results regarding the treatment with armodafinil, the R-enantiomer of the racemate modafinil.[110] It is important to remark that modafinil does not seem to affect frequency or severity of cataplexy. The main side effects of modafinil include headaches, dryness of the mouth, insomnia and nausea[33], whereas only a low potential for abuse has been observed[60].

Other stimulants like amphetamines, methamphetamines and methylphenidate are valid alternatives to modafinil[104], but Littner et al. point out that their benefit-risk-ratio is difficult to estimate due to the small numbers of patients that have been included in the respective clinical trials[103].

Sodium oxybate

In cases where stimulants cannot be used or do not yield a significant effect, sodium oxybate, which is also referred to as GHB (gamma-hydroxy-butyrate), is a treatment option for EDS. GHB is a

natural neurotransmitter which interacts with the GABA-B receptor[108]. A crucial advantage of sodium oxybate is that it improves many symptoms of narcolepsy. As a study of Lammers et al.[111] indicates, GHB reduces the frequency of hypnagogic hallucinations and daytime sleep attacks significantly as well as the subjective daytime sleepiness. Several studies also highlight the significant, dose-dependent effect of a regular intake of GHB on the frequency of cataplectic attacks[112, 113]. GHB also improves sleep quality by reducing night-time awakenings[114].

While it has been shown to be equally effective (compared to modafinil) in the treatment of EDS (measured using the MWT) and to provide an additive effect when combined with modafinil in a multicenter study by Black et al.[115], the drug remains problematic mainly due to its potential for abuse. GHB has been used by athletes for enhancing the release of growth hormone and is also frequently used as a "date rape" drug[108]. Possibly side effects of GHB are nausea, nocturnal enuresis, confusional arousals and headaches.[108]

Antidepressants

Furthermore, several classes of antidepressants (tricyclic antidepressants, SSRI, SSNRI) are also used as anticataplectic medication. Among the TCA, mainly clomipramine has been used to treat cataplexy[108, 116]. Although this is only supported by expert consensus, venlafaxine is often employed to reduce the frequency of cataplectic attacks[105].

Regarding SSRIs, among others, fluoxetine and escitalopram have been shown to improve cataplexy[117–119]. The dosage for anticataplectic therapy may greatly differ from antidepressant doses[120].

Tricyclic antidepressants, SSRIs and venlafaxine are also considered an optional treatment option for hypnagogic hallucinations and sleep paralysis[104].

Possible future treatment options

Recently, pitolisant has been emerging as a new treatment option for narcoleptic EDS[105]. The effectiveness of this inverse H3 antagonist has already been proven in several clinical trials[121]. Other substance are being tested for their effectiveness on the symptoms of narcolepsy, such as JZP-100, which has been shown to significantly reduce EDS (measured by ESS score and MWT latency) in narcolepsy patients[122].

The discovery of the involvement of hypocretin in the pathophysiology of narcolepsy has led to various efforts aiming to correct the hypocretin deficiency and thereby to provide a causal treatment for narcolepsy. Animal models have proven that this is indeed a valid treatment principle[123], but both peripheral and intranasal application of hypocretin have failed to meet the expectations[124]. Regarding other ways of substituting hypocretin, hypocretin cell transplantation using pluripotent stem cells as well as hypocretin gene therapy has been discussed[105].

Scheduled naps

Apart from pharmacological interventions, certain lifestyle recommendations have been made to narcoleptic patients. Since narcolepsy is often characterized by an irresistible urge to sleep, it has been suggested that scheduled naps might reduce sleep pressure and therefore contribute to less frequent and less severe sleep attacks. Rogers et al. as well as Mullington et al. provided some evidence that scheduled naps might be a therapeutic option for narcolepsy[125, 126]. However, Littner et al. state that in most cases, scheduling daytime naps is not sufficiently effective to be used as primary therapy[103]. In conclusion, scheduled naps are widely employed as an addition to pharmacological therapy and have shown to be beneficial especially for patients showing persisting severe daytime sleepiness under stimulant therapy[127].

Prognosis and burden of narcolepsy

For narcolepsy patients pharmacological treatment reduces the disabling symptoms of narcolepsy and improves the health related quality of life[128]. Nevertheless, since no curative treatment is available yet, life-long medication is often required to keep the symptoms at bay. But even in patients receiving medication, the impact of the condition on the quality of life is severe: Daniels et al. report reduced scores in all domains of the SF-36, especially in "physical, energy/vitality, and social functioning"[129]. Narcolepsy with cataplexy is associated with significantly lower scores than narcolepsy without cataplexy[130]. Beck Depression Inventory (BDI) scores indicate that more than half of all narcolepsy patients show signs of depression[129] and Kales et al. found higher levels of psychopathology as a consequence of narcolepsy[131]. Unfortunately, it is not unusual that narcolepsy is diagnosed years after the onset of the first symptoms[84], which further worsens the psychosocial implications of the condition.

3.2. Idiopathic hypersomnia

In many ways idiopathic hypersomnia is associated with the same clinical difficulties as narcolepsy type 2: Until today the pathophysiology of the condition is unknown. Furthermore, due to the absence of a specific symptom like cataplexy for narcolepsy type 1, no reliable diagnosis is possible that is based merely on clinical grounds. Furthermore, no biomarker has been identified yet which could be used to confirm or reject the diagnosis of IH. Finally, as with narcolepsy type 2, diagnosis essentially relies on the MSLT and is therefore prone to several methodical weaknesses.

The recognition of IH as a distinct condition began in 1956, when - more than 70 years after the first description of narcolepsy - Roth reported cases of hypersomnolent patients suffering from sleep drunkenness while wakening up[132]. Roth also noticed that a significant fraction of hypersomnolent patients could be distinguished from typical narcolepsy cases by clinical observation[133]. Dement et al. remarked in 1966 that some patients showing hypersomnolence similar to narcolepsy did neither suffer from cataplexy nor from sleep paralysis[134]. Additionally, no early REM episodes usually occurred in these patients. Berti-Ceroni et al. as well as Passouant et al. reported similar findings in the following years[135, 136]. Many of the typical clinical features of IH like (compared to narcolepsy) longer lasting but unrefreshing naps and prolonged instead of fragmented night sleep were described by Rechtschaffen et al. in 1969[137].

In the 1970s, the continuing research efforts of Roth regarding different kinds of central hypersomnias[138, 139] eventually led to the introduction of the term "idiopathic hypersomnia" in 1976, when Roth published a classification of 642 cases of hypersomnolent patients[140]. Roth also suggested the notion of a polysymptomatic and a monosymptomatic form of IH. The latter was mainly characterized by EDS, which, however, was not as irresistible as in narcolepsy. The former additionally included a prolonged night sleep and great difficulties in awakening in the morning.

This proposed subdivision of IH remained extensively discussed. Bassetti et al., on the one hand, distinguished three different subgroups, which were called "classic", "narcoleptic-like" and "mixed"[2]. On the other hand, Billiard et al. described a complete and an incomplete phenotype of IH, essentially following the original suggestion by Roth[141]. In the ICSD-2, two forms of IH were included: IH with and without long sleep time[52]. Both subgroups could only be diagnosed if the MSLT yielded a sleep latency below 8 minutes. However, a study by Anderson et al. found no specific symptom for either subgroup of IH as described in the ICSD-2 as well as the somewhat

paradox result that several IH patients did not fulfill the 8 minute criterion[142]. Eventually, the diagnostic subgroups of IH were formally abandoned with the introduction of the ICSD-3[101], when a major revision of the diagnostic criteria for the IH was done. As it will be pointed out later the IH subgroups as proposed by Roth are still implicitly acknowledged in the diagnostic criteria.

3.2.1. Epidemiology, etiology and pathophysiology

IH is a rare disease and seems to be even less frequent than narcolepsy. Roth reported a relative prevalence of 47%[140], but more recently lower values were reported by Bassetti[2] (16%) and Dauvilliers[130](5%). Similar to narcolepsy, IH onset usually occurs in adolescents or young adults[2, 143]. According to some studies women show a slightly higher prevalence of IH[2, 144].

Little is known about the etiology and pathophysiology of IH. About 30 % of cases have a positive family anamnesis[142]. Also, a correlation of IH with head traumas, viral illnesses and general anesthesia has been reported[143], but it remains unknown if and how these factors might be involved in the disease onset. It has been shown that IH patients have normal hypocretin CSF levels[145]. It has been suggested that CSF histamine might serve as a biomarker for hypersomnias of central origin[146], but according to the results of Dauvilliers et al., significant differences in CSF histamine levels exist neither between different groups of hypersomnia nor between hypersomnolent patients and controls[147]. Other findings indicated an involvement of GABA-related pathways[148], but again these results could not be reproduced[149]. Hence, until today no neurochemical diagnostic procedure is available which allows the confirmation or exclusion of the diagnosis IH.

One of the most promising recent results was contributed by Lippert et al., who demonstrated that a dysregulation of the circadian clock might be involved in the pathogenesis of IH: More precisely, dermal fibroblasts of IH patients showed a reduced amplitude of the periodical expression of several circadian clock genes[150].

Billiard et al. also mention possible genetical and immunological aspects in the pathogenesis of IH[143], but the current state of knowledge does not allow to deduce a comprehensive pathophysiological model for IH.

3.2.2. Clinical aspects

IH lacks specific clinical features like cataplexy that might simplify the diagnosis. Like other hypersomnias, IH is also characterized by EDS. Typically, this EDS does not cause an urge to go to sleep so irresistible as in narcolepsy, and if naps are taken, they tend to be of a longer duration and less refreshing[143].

Sleep drunkenness, also known as sleep inertia[151], is another typical, but not specific symptom of IH. Roth describes this symptom as inertia in the process of waking up, which is associated by confusion, disorientation and bad motoric coordination. Consequently, sleep drunken patients show severe difficulties in reaching complete wakefulness and tend to fall asleep again for several times[139]. Despite the abandonment of different IH subgroups in the ICSD-3, sleep drunkenness is still used to describe the clinical heterogeneity in IH. In their review from 2016, Dauvilliers et al. distinguish between IH with and without prolonged night sleep. The former group shows sleep drunkenness, EDS resulting in long, unrefreshing naps and more than 10 hours of night-sleep, which usually is of good quality[93]. IH without prolonged night sleep on the other hand, lacks sleep inertia, shows a normal amount of night-sleep and is mainly characterized by EDS, which results in short and refreshing naps[93].

Also, signs of autonomic dysfunction like headaches, palpitations and digestive problems can occur in IH[143]. In contrast to narcolepsy type 1, spontaneous remission of the IH has been reported[2, 141, 152].

3.2.3. Diagnosis

Like narcolepsy type 2, IH is essentially diagnosed by exclusion, which is reflected by the fact that four of the six ICSD-3 criteria listed below are designed to rule out other causes of hypersomnolence. Considering both the MSLT and the NPSG, not more than one SOREM episode is allowed to occur, since otherwise REM-sleep related disorders would be more likely. As it has been discussed before, the ICSD-3 does not explicitly acknowledge two subtypes of IH. Criterion D however takes the undisputable heterogeneity of IH into account: A sleep latency of less than 8 minutes is not required for the diagnosis of IH if more than 11 hours of sleep in 24 hours have been measured by actigraphy or PSG. Hence, IH patients that have been assigned to the former IH with long sleep time subgroup could be diagnosed with IH even without fulfilling the MSLT sleep latency threshold. For patients suffering from IH without long sleep type, who usually do not

complain about the typical sleep drunkenness, a sleep latency below 8 minutes is required as an objectively measurable correlate of the EDS.

Table 3.4: Diagnostic criteria for idiopathic hypersomnia according to the ICSD-3[5]		
Criteria A-F must be true.		
A	History of daily periods of irresistible need to sleep or actual lapses to sleep for *at least three months.*	
B	Cataplexy is absent.	
C	Less than two SOREM episodes in the MSLT - If the REM latency in the preceding NPSG was less than 15 minutes, no SOREM in the MSLT is allowed.	
D	At least one of the following criteria is true.	
	D1	The sleep latency in the MSLT is less than 8 minutes.
	D2	24-hour PSG or wrist actigraphy combined with a sleep log show a 24h sleep time of more than 660 minutes.
E	Insufficient sleep syndrome has been ruled out.	
F	The symptoms are not better explained by another sleep or medical disorder, or by the intake of drugs or medication.	

3.2.4. Treatment and prognosis

Due to its similarity to narcolepsy almost all treatment options for narcolepsy have been "borrowed" for the treatment of IH[143]. Bassetti summarizes that the majority of IH patients improve under treatment with stimulants, but some cases show a better response to antidepressants[2].

For modafinil, which is the first line therapy option for narcolepsy, a large body of evidence suggests its efficacy for EDS in IH[104, 142, 144, 153, 154]. Furthermore, methylphenidate has been shown to significantly improve daytime sleepiness[2, 144]. In treatment resistant cases, amphetamines and pitolisant have been recommended[93]. Sodium oxybate yields an improvement in ESS similar to its effect on narcolepsy patients, and has also demonstrated some improvement in sleep inertia[155], which in general is very hard to treat[93]. Finally, mazindol is an option if other drugs have failed to show significant improvement[156].

As in most cases of IH daytime naps are not refreshing, planned naps are less effective in IH than in narcolepsy[93]. IH has an impairing effect on the quality of life that is comparable to the effect of narcolepsy type 2[157]. However, most patients respond well to treatments with stimulants[142] and prognosis is further improved by the chance of spontaneous remission.

3.3. Challenges in diagnosis and differential diagnosis of narcolepsy and IH

Having introduced the three illnesses that are relevant for this thesis, in this section several issues that are encountered in clinical practice will be highlighted. For each diagnostic tool, current problems that typically arise in diagnosis and differential diagnosis will be discussed.

It should be noted that the summary below is by no means comprehensive. Other hypersomnias of central origin (i.e. the Kleine-Levin-Syndrome, hypersomnia due to a medical condition, hypersomnia due to substance abuse or drug intake and hypersomnia caused by a psychiatric condition[93]) will not be discussed in further detail. Differential diagnostic considerations will be restricted to the comparison between IH and the two subtypes of narcolepsy.

3.3.1. Clinical aspects

Excessive daytime sleepiness (EDS)

Obviously, EDS is not specific for IH or narcolepsy. Indeed, for patients presenting themselves with EDS, several causes more common than hypersomnias of central origin have to be considered. A study in the European population found an overall prevalence of EDS of 15% in the general population. 1,6 % of the subject randomly selected for this study reported having two or more naps each day[55]. Guilleminault lists several important causes for EDS, the most common being insufficient sleep, which may be caused by self-chosen behavior or social necessities[91].

But even if there is no lack of quantity, insufficient sleep quality can deteriorate daytime wakefulness. Typically, night sleep is fragmented due to various disturbing factors. A very common cause of fragmented night sleep are sleep related breathing disorders[91]. Restless leg syndrome and periodic limb movements are other widespread causes of sleep fragmentation. Furthermore, several internal conditions may lead to a disturbed night sleep, such as nocturnal angina, gastrointestinal diseases and urinary dysfunction.[158]

Regarding the comparison between IH and narcolepsy, EDS in narcolepsy is usually described to be more severe, leading to an irresistible urge to nap. These naps are assumed to be short and refreshing, in contrast to IH associated daytime naps, which do not lead to temporary refreshment and are usually longer in duration[92, 143, 152, 159]. This dichotomy may be useful for describing typical patients with IH or narcolepsy, but several results hint at a more complex reality. Vernet et al. have

reported narcolepsy patients, whose naps are unrefreshing and whose night sleep is prolonged like in some patients with IH[160]. On the other hand, in a recent review from 2017 by Dauvilliers et al., IH without prolonged nighttime sleep is also characterized by short and refreshing daytime naps[93]. In conclusion, EDS remains a highly unspecific symptom, and the classically reported differences in nap duration and quality should not be overestimated regarding the diagnostic usefulness.

Other symptoms of narcolepsy

Despite being part of the classic narcoleptic tetrad, sleep paralysis is not a specific symptom of narcolepsy. A study on Afro-American individuals found a prevalence of more than 20 % for at least one episode per month, and a more frequent occurrence was associated with elevated stress levels and the occurrence of panic disorder[96]. Aldrich found that sleep paralysis as well as sleep hallucinations occur equally common in IH and narcolepsy without cataplexy[161]. As, on the other hand, only 50% of all narcolepsy patients show the symptom of sleep paralysis[87], one may conclude that neither its presence nor its absence can be used for diagnostic purposes.

A similar conclusion can be reached for sleep hallucinations that are often associated with sleep paralysis[95], since they are also relatively frequent in the general population[94] and only occur in about two third of all narcoleptic patients[87].

Even the most specific symptom of narcolepsy, cataplexy, can lead to a diagnostic pitfall if cataplexy-like symptoms, that may also occur in non-narcoleptic individuals[162], are not taken into account. Anic-Labat et al. suggest that "real" cataplexy is most easily identified by the typical trigger situations[162] and Sturzenegger et al. have found that cataplexy-like muscle weaknesses are usually less pronounced and often lack the typical jaw sagging in clear-cut cataplectic attacks[87]. Finally, if CSF hypocretin levels are not determined, the possibility remains that narcolepsy type 2 patients have to be relabeled as narcolepsy type 1 after a delayed onset of cataplexy.

Regarding sleep inertia in IH and narcolepsy, not much comparative data is available. Both narcolepsy and OSA are also known to lead to an increased sleep inertia[151]. Sturzenegger et al found a prevalence of 40 % in "non-narcoleptic" hypersomnia compared to 24 % in narcolepsy and 10 % in healthy controls[87]. It should be noted that these differences did not reach the level of significance. Similarly, Martinez-Rodriguez et al. report no significant difference in terms of sleep drunkenness between narcolepsy with cataplexy, narcolepsy without cataplexy and IH[92]. In

conclusion, sleep inertia is neither specific for hypersomnias of central origin nor suitable for differential diagnostic purposes between IH and narcolepsy.

3.3.2. The MSLT

3.3.2.1. (Sustained) sleep latency

According to the current diagnostic criteria, the gold standard tool to assess daytime sleepiness is the MSLT. The ICSD-3 postulates a sleep latency under 8 minutes for the diagnosis of both narcolepsy and IH (when no 24h-PSG or wrist actigraphy is used). Considering the clinical observation, that daytime sleepiness and the urge for daytime naps is usually more imperative in narcoleptics than in IH patients, it seems counterintuitive at first glance to set the diagnostic threshold at the same level for both diseases. Indeed, whereas the 8-minutes threshold for narcolepsy was chosen as a compromise between sensitivity, specifity and predictive values (as it is discussed by, e.g. Aldrich et al.[41]), the cutoff for IH has been set on the same level only for the sake of simplicity[83]**.**

Littner et al. published data of a control population without sleep medical diseases, in which the average sleep latency was 11,6 minutes with an empirical standard deviation (SD) of 5,2 minutes[32]. Assuming normal distribution in this population, more than 15 percent of healthy controls would fulfill the 8 minute-criterion. For the modified version of the MSLT, the MSLT30, Geisler et al. found an quadratic age dependence of sleep latency in normal subjects, with a minimum for middle-aged subjects[42]. Hence especially middle-aged subjects are expected to drop below the 8-minute threshold, possibly leading to false positive diagnosis. Goldbart et al. used data of the Wisconsin Sleep Cohort to demonstrate that a positive MSLT, i.e. at least two SOREMs and a sleep latency under 8 minutes, occurs in 3,4 % of the normal population, with shift work and short sleep being strong positive predictors[40]. Arand et al. conclude, that due to the high standard deviations as well as floor and ceiling effects, the MSLT does not distinguish reliably between sleep disorder patients and healthy controls[24].

Focusing on the sleep latencies of IH and narcolepsy patients, some differences between the diseases have been found. According to Vernet[3], IH patients have an average sleep latency of 7,8 minutes (SD 0,5 minutes), suggesting that quite a significant fraction does not fulfill the 8 minutes criterion. Vernet et al. also distinguished between the two subtypes of IH. Whereas patients with

IH without long sleep time had a low sleep latency of 5,6 minutes (SD 0,3 minutes), in the IH with long sleep time population the mean sleep latency was 9,6 minutes (SD 0,7 minutes). Regarding the narcolepsy subtypes, Sturzenegger et al. found the lowest sleep latencies (mean 1,6 minutes) in narcolepsy patients with confirmed hypocretin deficiency. Slightly higher values were found in patients with clinically confirmed cataplexy, whereas patients having "probably" cataplexy had a mean sleep latency of 3,7 minutes[87]. Šonka et al. report a mean sleep latency of 4,5 minutes for narcolepsy without cataplexy and 2,8 minutes for narcolepsy with cataplexy[163].

In conclusion, the lowest sleep latencies are to be expected from narcolepsy type 1/with cataplexy patients, followed by narcolepsy without cataplexy patients. In total, IH patients also show reduced sleep latencies, which tend to be higher than those in narcolepsy. If one distinguishes the two clinical subtypes of IH, IH patients with long night sleep usually do not fulfill the 8 minutes criterion, but could be diagnosed by confirming a total sleep time of more than 660 minutes per day using PSG or actigraphy. Hence, although in principle the MSLT could be useful in the differential diagnosis of narcolepsy and IH, its actual value in clinical practice is limited by the considerable standard deviations reported for each group.

In order to further enhance the value of sleep latencies as a differential diagnostic tool Pizza et al.[4] suggested considering the sustained sleep latency instead, which is defined as the amount of time until the patient reaches a deeper sleep stage than stage 1 or three consecutive epochs of sleep. Pizza et al. found that the difference between the sustained sleep latency and the conventional sleep latency allows to distinguish IH from both narcolepsy types. Whereas high values of this difference, which is referred to as "Delta" by Pizza et. al, suggest the physiological "waxing and waning" occurring during the transition from wakefulness to sleep in IH patients, low values of Delta reflect a sudden and complete sleep onset, which was found to be typical for narcolepsy[4].

3.3.2.2. SOREM episodes

According to the ICSD-3, the count of SOREM episodes during a MSLT allows the discrimination between the "REM-sleep-disorder" narcolepsy and idiopathic hypersomnia. However, like the sleep latency, the SOREM count can be affected by various aspects. Shift work and insufficient sleep can increase the chance that a healthy individual has two or more SOREM in a MSLT[40]. Other factors that are associated with an increased SOREM count are the intake of non REM-suppressing antidepressants and a positive HLA DQB1*0602 status (which, however, also

increases the risk for developing narcolepsy)[164]. Furthermore, depression can reduce the REM latency[33] and hence produce more REM and even SOREM episodes, although Mignot et al. found no association between depression and the SOREM count[165]. Apart from that there is increasing evidence that REM sleep regulation differs heavily between the sexes with a reported odds ratio of 2,62 for at least two SOREMS in men compared to women[165]. OSAS (obstructive sleep apnea syndrome) is also increasing the occurrence of SOREMs. This effect is more significant if the patients have a more severe nocturnal oxygen desaturation[166]. Great variance can be found in the empirically observed prevalence of two or more SOREMs. Singh et al. report a prevalence of 3,9 % in a population based sample[167]. Allen et al. calculated a prevalence of 13,1 % for men and 5,6 % for woman, whereas 5,9 % of men and 1,1 % of women would meet both the MSLT criteria for narcolepsy type 2[164].

In the review article of Arand et al. the finding of at least two SOREM has been calculated to have a sensitivity of only 0,78 , compared to a rather high specifity of 0,93 for the diagnosis of narcolepsy[24]. Hence, one can expect that most individuals not suffering from narcolepsy will yield less than two SOREMs in the MSLT. Conversely, approximately only three out of four narcolepsy patients will actually fulfill the SOREM criterion when an MSLT is performed.

On average, IH patients have been reported to have 0,37 SOREMs[168]. Equivalently, about a third of IH patients has exactly one SOREM. While the occurrence of one SOREM (compared to none) does not change the ICSD-3 diagnosis of these patients, Bozluolcay et al. have shown that regarding sleep and REM latency there are significant differences between IH patients with one and no SOREM episode in the MSLT. The authors concluded that this "intermediate" group defined by one SOREM in the MSLT might actually be closer to narcolepsy type 2 than to "0-SOREM-IH"[169]. This finding might be explained by the reported limitations regarding test-retest-reliability of the MSLT: Trotti et al. examined a population of IH and narcolepsy without cataplexy patients and found that after having repeated the MSLT, in 31% of all cases diagnosis had to be revised due to a changed SOREM count[39]. Thus, especially in the situation of one or two SOREMs in the MSLT, a second MSLT run might lead to a switch of the diagnosis from IH to narcolepsy type 2 or vice versa.

3.3.3. The Epworth Sleepiness Scale

The Epworth Sleepiness Scale (ESS) is widely employed for estimating the subjective sleepiness of patients. However, its resulting score does not explicitly appear in the ICSD-3 criteria for narcolepsy or IH. Johns reported ESS score ranging from 2 to 10 for healthy controls, 12-24 for idiopathic hypersomnia and 13-23 for narcolepsy[9]. More recent evidence confirms that healthy subjects usually score less than 10 points[11], whereas IH and narcolepsy patients have significantly higher scores[163].

There are inconclusive results regarding differences of the ESS score between the conditions. Pizza et al. report the surprising fact that narcolepsy without cataplexy is associated with lower ESS scores than narcolepsy with cataplexy and IH, whereas the latter two conditions do not significantly differ[4]. One other study found similar ESS scores in narcolepsy type 1 and type 2[170]. Anderson et al. calculated a mean initial ESS score of 16,3 for IH and a significantly higher score of 18,6 for narcolepsy with cataplexy[142] with a standard deviation of 3,3 for each group.

Overall, the ESS allows a reliable distinction between healthy controls and narcolepsy/IH patients. On average, ESS scores of IH patients seem to be slightly below scores for narcolepsy, but the standard deviations for each disease are too high to justify differential diagnostic decisions.

3.3.4. PSG parameters

The nocturnal PSG is important for the diagnosis of narcolepsy and IH in many ways. Obviously, common differential diagnoses like sleep related breathing disorders have to be ruled out as a cause of EDS. Furthermore, SOREMs occurring in the NPSG are effectively added to the SOREMs in the MSLT. Therefore, for the diagnosis of IH, no more than one SOREM, either in the MSLT or the NPSG may be observed. Regarding narcolepsy, if the preceding PSG yields a SOREM, one additional SOREM in the MSLT suffices for the fulfillment of the diagnostic criteria.

On average, 77 % of narcolepsy patients show a SOREM in their NPSG, compared to 1,7% of IH patients[171]. Closely related to this finding are the differences in REM latency, which directly influences the occurrence of SOREMs: Various comparative studies report a shorter REM latency for narcolepsy than for IH[168, 171, 172]. However, IH patients show REM latencies that do not significantly differ from healthy controls[3].

There are other clinical differences regarding night sleep between narcolepsy and IH patients: Night sleep in narcolepsy tends to be shallow, disrupted and fragmented[33, 92], whereas IH patients usually have deep, undisturbed night sleep[172, 172]. This observation is reflected by the differences that have been reported regarding arousals and sleep efficiency.

Martínez-Rodríguez et al. found no significant differences in sleep efficiency and arousal index but nevertheless report a gradient ranging between the conditions. Sleep efficiency is highest, and the arousal index is lowest for IH, whereas the most pathological findings occur in patients with narcolepsy with cataplexy. Intermediate values are observed for narcolepsy without cataplexy[92].

Regarding sleep efficiency, these results are confirmed in most other studies[142, 168, 172, 173]. Furthermore, the sleep efficiency in IH seems to be even higher than in healthy controls[3]. However, contrary results have also been published, indicating that sleep efficiency might be equal[171] or even higher in narcolepsy than in IH[168]. Additionally, the shallow night sleep in narcolepsy also affects the relative duration of the different sleep stages: Typically, sleep stage 1 is increased in narcolepsy, whereas sleep stage 2 shows an decreased duration and the time in sleep stage 3 might be reduced or normal[173]. In comparison, IH patients do not differ from healthy controls regarding sleep stages 1 and 2. The percentage of sleep stage 3 seems to be reduced, but not the absolute time in stage 3[3]. Bassetti et al. found no significant difference for the REM sleep fraction of IH and narcolepsy patients[2].

In conclusion, the clinical differences that have been observed between IH and narcolepsy transform to some degree to differences in the PSG parameters. Their differential diagnostic value has not been investigated in detail yet.

3.3.5. Vigilance tests

Impairment of vigilance is a dimension of sleepiness that cannot be measured by the MSLT[174]. Schulz et al. have summarized the performance results of narcolepsy patients in several neuropsychological tests. They concluded that, whereas narcoleptics usually perform well in short and challenging tasks, significantly worse (compared to healthy controls) results are observed in monotonous tasks that stretch over an extended period of time. The reason for these specific cognitive limitations is assumed to be hypovigilance[175].

Several vigilance tests are available, such as the sustain attention to response test (SART), the steer clear test, the Oxford sleep resistance test (OSLER) , the psychomotor vigilance task (PVT)[174] and the Mackworth Clock Test, whose computerized version has been used for the dataset of this thesis.

One study by van Schie et al. found that vigilance as measured by the SART test is similarly deteriorated in narcolepsy as in other causes of EDS[176]. Findley et al., however, found significant differences between sleep apnea patients and narcoleptics in the Steer Clear test: Compared to control subjects, whose performance did not decrease over time, narcolepsy patients showed a clear linear vigilance decrement. For sleep apnea patients, only a "trend" towards a vigilance decrement was observed[177]. Comparative studies regarding the performance of narcolepsy or IH patients could not be identified.

In conclusion, vigilance is an important dimension of sleep- and wakefulness, whose importance for the diagnosis and differential diagnosis of narcolepsy and IH remains to be explored in further detail.

3.4. Aims of this thesis

After having summarized both the conditions and the diagnostic tools at hand, a perspective has been reached that allows the statement of the central aims of this thesis.

I.	Given the MSLT as an objective tool for measuring sleepiness: To what degree are the different MSLT parameters like sleep latency, SOREM count and the recently suggested parameter Delta redundant? Is it justified to treat them as separate diagnostic measures or are they equally influenced by the hidden common factor "sleepiness"?

II.	Motivated by the shortcomings of the current diagnostic criteria and the resulting entities: Can certain variables, which would allow a clearer classification of hypersomnolent patients in the dataset, be identified? Are there diagnostic parameters that naturally subdivide patients with narcolepsy and IH?

III.	Having identified suitable variables for the classification or clustering of the dataset: What groups do emerge if the given dataset is explored with respect to the cluster variables and how are these groups related to the current diagnostic concepts?".

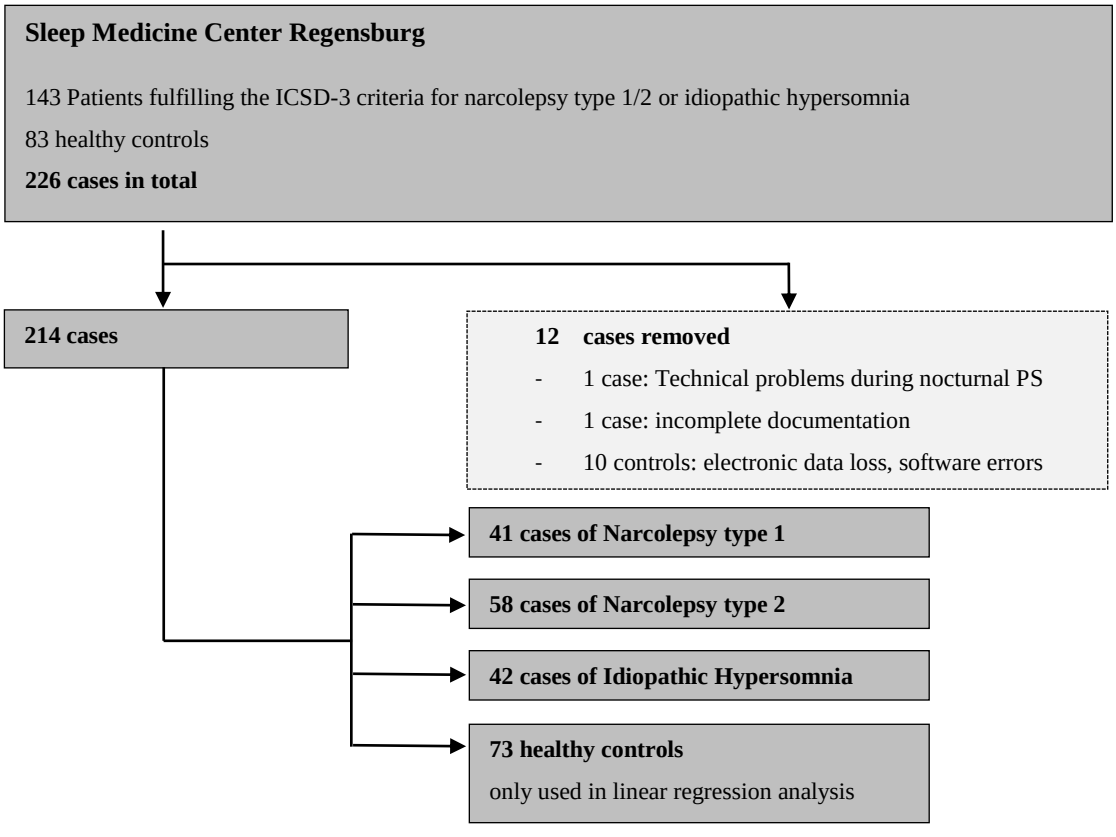
IV.	Reflecting the current diagnostic groups and the algorithmically obtained clusters: Are there ways to improve the diagnostic criteria or to refine the given concepts of narcolepsy and IH ?

These questions will structure the remaining parts of this thesis. First, linear regression analysis will be used to address the correlation between the different MSLT parameters. Next, a set of "important" variables will be determined by calculating the principal components of the dataset. These variables will serve as cluster variables for the central part of this thesis, which comprises three different cluster analyses. The cluster analyses will yield three different partitions of the dataset and allow valuable insights into the structure of the dataset. The thesis will end with a summarizing discussion of all results and the possibility of alternative diagnostic classifications.

4. The Dataset

All patient data was obtained at the Center for Sleep Medicine Regensburg. Each patient gave a written consent to a scientific evaluation of his/her anonymized data. All patients whose MSLT had been performed between January 1st 1996 and December 31st 2015, were considered. Eventually, only patients who strictly fulfilled the diagnostic criteria of the ICSD-3 for narcolepsy type 1, narcolepsy type 2 or idiopathic hypersomnia were included. No 24h-PSG or actigraphy was used for the diagnosis of IH, hence all patients labeled with IH had sleep latencies below 8 minutes. Also, no CSF hypocretin levels were determined, so all narcolepsy type 1 cases had a history of cataplexy. Furthermore, data of 83 healthy individuals who had participated in the normative MSLT study by Geisler et al.[35] were included as a control group.

Diagram 4.1: The dataset



The dataset initially comprised 226 cases. Two patient cases had to be removed due to technical problems during the nocturnal PSG and incomplete documentation of PSG results. Another ten healthy controls were omitted because software errors had caused the loss of important parameter values. In total, a dataset consisting of 141 patients was obtained. In the end, 214 cases (including 73 controls) were included.

Several patients were also suffering from sleep medical comorbidities, with the most frequent diagnoses being OSAS and PLMS. As it explicitly stated in the diagnostic criteria for narcolepsy type 2 and idiopathic hypersomnia, OSAS had to be ruled out as an exclusive explanation for the hypersomnolence. Hence, all patients also diagnosed with OSAS were stable and under sufficient treatment regarding their breathing disorder.

Table 4.1: Frequent comorbidities

OSAS: Obstructive sleep apnoea syndrome; RLS: Restless legs snydrom; PLMS: Periodic limb movement in sleep

	OSAS	RLS	Depression	PLMS	Parasomnias
narcolepsy type 2	7	2	8	8	0
narcolepsy type 1	4	1	2	10	0
idiopathic hypersomnia	4	1	3	2	2
total	15	4	13	20	2

Since only MSLT results had been obtained from the healthy individuals, in the cluster analyses and the preliminary principal component analysis only the aforementioned 141 patients could be included. Table 4.2 lists all variables that will be considered for the remainder of this thesis.

Table 4.2: Considered variables

From healthy controls only MSLT30 results were obtained; measurement units are given in the square brackets. If not specified otherwise, the values of the (unstandardized) values will always be given in these units

MSLT30:		
mean sleep latency to stage 1	[min]	(SL1)
mean sleep latency to stage 2	[min]	(SL2)
mean sustained sleep latency	[min]	(susSL)
SOREM count	[0,..,5]	(#SOREM)
sum of sleep time	[min]	(TST)
Delta [=susSL – SL1]	[min]	(Delta)
Epworth Sleepiness Scale Score	**[0,…,24]**	(ESS)

Vigilance test (Quatember-Maly):		
number of correct reactions		(VigCorr)
number of false reaction		(VigFalse)
mean reaction time in correct reactions [sec]		(VigRT)
Nocturnal 12h-PSG:		
sleep efficiency	[%]	(PSG_SEI)
sleep stage 3 fraction of TST	[%]	(PSG_N3_TST)
REM latency	[minutes]	(PSG_REML)
arousal index	[1/h]	(PSG_AI)
History of cataplexy	**[0 / 1]**	(CATAP)

The tables 4.3 and 4.4 list the descriptive statistics of all considered variables. For the sake of completeness, also age and sex ratio are mentioned here.

Table 4.3: Descriptive statistics for the categorical variables

	narcolepsy type 2 N=58	narcolepsy type 1 N=41	idiopathic hypersomnia N=42	controls N=73
CATAP	0/58	41/41	0/42	0/73
Sex (m/f)	23/30	18/23	10/32	44/29

As it has been mentioned above, a modification of the MSLT, the MSLT30 was used for this thesis. Apart from the changed termination rule, the guidelines for the MSLT in table 2.1 were followed. In most cases, the PSG was conducted in the night preceding the MSLT. In the remaining cases, the PSG took place few days before the MSLT.

Table 4.4: Descriptive statistics for the continuous variables								
	narcolepsy type 2 N=58		narcolepsy type 1 N=41		idiopathic hypersomnia N=42		controls N=73	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Age [years]	29,9	11,0	31,5	11,2	35,4	12,4	42,9	11,3
MSLT								
SL1	4,39	2,21	2,43	1,42	5,45	1,67	13,2	6,96
SL2	7,89	3,34	5,69	3,25	10,7	3,52	20,3	6,64
susSL	4,98	2,49	2,87	1,69	7,01	2,89	15,8	7,36
#SOREM	2,79	0,932	3,95	1,02	0,190	0,397	0,233	0,566
TST	120	14,9	131	11,0	108	17,3	63,3	34,7
Delta	0,585	0,723	0,439	0,974	1,56	1,97	2,64	2,30
ESS	15,6	4,16	16,3	3,10	15,0	3,72		
Vigilance test								
VigCorr	83,8	22,2	87,7	14,6	85,0	17,1		
VigFalse	3,62	4,94	3,98	5,57	2,29	2,85		
VigRT	0,560	0,124	0,557	0,111	0,573	0,118		
PSG								
PSG_SEI	91,7	5,57	86,7	7,31	89,6	6,56		
PSG_N3_TST	15,2	9,58	11,7	8,39	15,3	9,17		
PSG_REML	65,2	40,4	48,9	56,6	95,6	54,4		
PSG_AI	7,30	8,31	9,00	12,5	6,43	7,47		

5. Correlation and linear dependence of MSLT parameters: Sleep latencies and SOREMs

Addressing the first question that has been raised above, this section will further investigate the intercorrelation of three different variables: The sleep latency to stage I (SL1), the SOREM count (#SOREM) and the parameter Delta. Linear regression analysis will serve as the central statistical tool for this task. This chapter starts with a brief methodical overview, which will closely follow the introduction by Dougherty[178]. After that, two different linear models will be introduced and investigated by linear regression.

At the end of this chapter it will have become clear, if and to what degree the different MSLT parameters are correlated to each other and how their linear relationship can be described. These insights will be obtained separately for each diagnostic group, which might support a deeper characterization of narcolepsy and IH based on the typical MSLT results.

5.1. Methodical considerations

Linear regression analysis aims to investigate whether a linear relationship exists between several random variables. More precisely, one assumes that a *dependent* variable Y is related to one or several *explanatory* variables $X_1, X_2, \dots, X_n$ in the following way:

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_nX_n + u.$$

Here, u denotes a disturbance term, which is assumed to be normally distributed with mean value 0. Linear regression analysis uses the information of m (in the situation of this thesis, m is the number of patients/controls that are included in the analysis) observed corresponding values of Y and $X_1, X_2, \dots, X_n$ to estimate the coefficients $b_0, b_1, \dots, b_n$. For each observation, a different disturbance term u affects the value measured for Y, thereby potentially hiding the assumed linear relationship. Mathematically, one obtains the estimates $\beta_0, \beta_1, \dots, \beta_n$ for $b_0, b_1, \dots, b_n$ based on the principle of minimizing the sum of squared residuals (RSS). Here, the difference between the observed value of Y and the predicted value of Y based on the observed values of $X_1, X_2, \dots, X_n$ is defined as *residual* e:

$$e_j = Y_j - \left(\beta_0 + \beta_1 X_{1j} + \beta_1 X_{1j} + \cdots + \beta_n X_{nj}\right),$$

$$RSS = e_1^2 + e_2^2 + \ldots + e_m^2.$$

Estimating the regression coefficients in this manner does not always yield optimal results. However, it can be shown that four conditions, the *Gauss-Markov conditions*, guarantee that the estimates are unbiased and as efficient as possible[178].

These conditions are related to various properties of the disturbance terms $u_1, u_2, \ldots, u_m$. Not all conditions will be discussed in detail, but it is important to acknowledge the most critical one regarding this dataset: The second Gauss-Markov condition postulates that all disturbance terms (corresponding to the different observations) have the same constant variance. This situation is also called *homoscedasticity*. If it is not fulfilled, i.e. in the situation of *heteroscedasticity*, the coefficient estimates remain unbiased nevertheless. However, all error estimates regarding the significance of the obtained results, i.e. confidence intervals, p-values and so on, will be invalid. Hence, in this thesis, the dataset will be tested for the occurrence of heteroscedasticity using the Koenker test[179]. If the test indicates that heteroscedasticity may be present, adjusted error estimates using the heteroscedasticity-consistent standard error estimators by Hayes[180] will be employed. More precisely, the RLM macro that was released in a recent publication by Darlington and Hayes[181] will be used for calculating the adjusted standard errors.

Dummy variables

If linear regression has to be performed on different groups, as it is the case in this thesis, dummy variables can be employed. These artificial explanatory variables take into account possible differences between the diagnostic groups in slope and intercept coefficients that might occur. Therefore, for every group other than the reference group, two dummy variables have to be introduced for each explanatory variable[178]. The main advantage of the usage of dummy variables over just performing separate regression analyses for each subgroup is the additional information regarding the possibly significant differences of the regression coefficients between the distinct groups.

Variable inclusion

Due to the inclusion of dummy variables, every regression analysis in this thesis will be a multivariable regression analysis, i.e. more than one explanatory variable will be assumed. In such

cases, it is very likely that at least some a priori assumed explanatory variables will turn out to not significantly contribute to the model. In practice, this would be reflected by small corresponding regression coefficients and high associated p-values. In order to avoid an unnecessarily complicated linear model, there are several approaches that remove such "insignificant" explanatory variables. Here, the "stepwise" approach was chosen, in which independent variables are included successively and possibly excluded again, if their corresponding p-values are below predefined thresholds[182]. In this case, the SPSS defaults of p=0,05 for inclusion and p=0,1 for exclusion were used.

Reported parameters in linear regression

Several different parameters will be reported for each linear regression. Each regression coefficient will have a corresponding p-value that represents the hypothetical probability that this coefficient is a consequence of pure chance. A low p-value therefore minimizes the danger that the regression coefficient at hand is merely a statistical artifact that has been caused by unlikely coinciding disturbance terms.

Confidence intervals for each coefficient are interval estimators for the true coefficient. That means that with a high probability (which is set as 95 % for this thesis), the "true coefficient" is included in the estimated interval.

Additionally, the F-value will be reported for each model. This value corresponds to the null hypothesis, that the dependent variable is not depending on any of the explanatory values at all (or equivalently, all "real" coefficients b_i are zero except for b_0). The F-value is closely related to the R^2-value, which can be interpreted as the fraction of the dependent variable variance that can be explained by the explanatory variables. Hence, a high F-value is associated to a high R^2-value, indicating that the suggested linear model does indeed explain a lot of the dependent variable variance.

Furthermore, the Durbin-Watson test for autocorrelation will be employed. Autocorrelation describes the correlation between the disturbance terms for each measurement (or in this case, patient). This also leads to an autocorrelation of the residuals, which are directly accessible and can therefore be tested for this phenomenon. The coefficient point estimates are not affected by autocorrelation, but an underestimation of the p-values (and an overestimation of the confidence intervals) would occur[178]. If the Durbin-Watson statistic yields values close to 2, one can conclude

that no significant autocorrelation is present. Savin et al. have provided a table of critical values of the Durbin-Watson test. If the test statistics yield values above the upper limit, the null hypothesis stating that no autocorrelation is present will be accepted[183].

Finally, the tolerances for each regression coefficient will be mentioned. Low levels of tolerance indicate multicollinearity, which means that the given explanatory variable is itself explainable by the remaining explanatory variables. This phenomenon does not deteriorate the actual point estimates of the regression coefficients but increases confidence interval and equivalently the corresponding p-value. The occurrence of a low tolerance (usually, values below 0,2 are interpreted as severe multicollinearity[184]) will be mentioned in the interpretation of the different regression results.

All linear regression analyses were performed using SPSS, Version 23.0.0.0[185] .

5.2. SOREMs and sleep latencies

This first linear regression analysis will investigate the relationship between the sleep latency and the SOREM count in the MSLT. If one agrees that the MSLT sleep latency is an objective measure for sleepiness and the number of SOREM episodes is a correlate of dysregulated REM sleep, the question can be raised whether REM sleep dysregulation (which is assumed to lead to cataplectic attacks) and daytime sleepiness are symptoms that are related to each other in any way. In terms of the dataset variables, this transforms into investigating the correlation between the SOREM frequency and the mean sleep latency.

Despite the discovery of the central role of hypocretin in the pathophysiology of narcolepsy, several theories regarding the narcoleptic symptoms are still being discussed[186]. Although the similarity of cataplexy and REM sleep atony suggests that an impairment of REM sleep regulation might be the primary cause of the narcolepsy phenotype, it has also been considered that controlling instances between sleep and wakefulness or various components of sleep regulation itself might be affected.

Of course, these theories should not be understood as exclusive and significant differences have to be expected between the narcolepsy subtypes, and in particular between narcolepsy and IH. The results of this regression analysis will give some insight into the interdependence of the sleep latency and the occurrence of SOREMs.

5.2.1. Modeling assumptions

In order to justify a linear regression analysis for the issue at hand, one postulates a linear relationship of the following form:

$$SL1 = b_0 + b_1 \cdot SOREM + u$$

However, since there are three different diagnoses to consider (and the additional control group) it is unreasonable to assume that – even if this linear relationship is a valid assumption- all groups have the same regression coefficients. Therefore, as it has been remarked above, dummy variables are employed which model the differences in slope and intercept coefficients between the diagnostic groups.

Additionally, one has to take into account that by definition, narcolepsy patients show at least two SOREMs, whereas IH patients can have a maximum of one SOREM in a MSLT. Therefore, the dummy variables would be artificially correlated with the SOREM count, essentially leading to multicollinearity and unprecise coefficient estimates. Also, with respect to the above model no reasonable direct comparison is possible between IH and narcolepsy since these conditions cover different intervals regarding the SOREM count.

This issue can be resolved by splitting the analysis into two subgroups: One analysis will be restricted to the narcolepsy subtypes; the other analysis will comprise IH patients and healthy controls. Thereby, the unwanted collinearity effects are avoided, and more precise coefficient estimates are possible. Because of this approach, three healthy controls had to be excluded from the analysis due to their increased SOREM count. These cases are listed in Table 5.1.

Table 5.1: Omitted healthy controls

SOREM count	Sleep latency [min]
2	7,5
3	2,1
2	10,3

In conclusion, the following two models will be investigated:

On the one hand

$$SL1 = b_0 + b_1 \cdot SOREM + \delta_{IH} \cdot IH + \lambda_{IH} \cdot SIH + u$$

for the subgroup of IH patients and healthy controls. The two dummy variables are IH and SIH, representing possible differences in the intercept and slope coefficient, respectively. IH and SIH will be zero for controls, whereas in IH patients one defines IH as 1 and SIH equal to the SOREM count of the case. As it is explained in detail by Dougherty[178], this approach effectively treats δ_{IH} as a coefficient regarding differences in the intercept and λ_{IH} as a coefficient which encodes differences in the slope.

On the other hand, the equation

$$SL1 = b_0 + b_1 \cdot SOREM + \delta_{TYPE1} \cdot TYPE1 + \lambda_{TYPE1} \cdot STYPE1 + u,$$

when narcolepsy patients are considered as the dataset. Here, narcolepsy type II patients represent the reference case where both dummy variables are defined as zero. $TYPE1$ and $STYPE1$ are defined analogously to IH and SIH.

5.2.2. Results

5.2.2.1. Controls and IH patients

By stepwise multiple regression, all a priori hypothesized variables entered the final model. Table 5.2 depicts the unstandardized coefficients as well as the 95 % confidence intervals, p-values and tolerances. The order of the independent variables is the order of inclusion into the model.

Table 5.2: Regressions coefficients, confidence intervals and p-values
70 healthy controls and 42 IH patients were included
The 95 %- confidence intervals (CI) were adjusted using Hayes' HC3 algorithm

Coefficients		**CI: lower limit**	**CI: upper limit**	**p-values**
Constant	14,3	12,5	16,1	<0,001
IH	-8,84	-10,7	-6,99	0,005
#SOREM	-5,87	-9,90	-1,84	<0,001
SIH	5,89	1,58	10,2	0,008

A F-value of 30,6 (df=111) and a corrected R^2=0,369 were reported. The Durbin-Watson test did not reveal significant autocorrelation of the residuals (Durbin-Watson statistic 1,87). The lowest tolerance value (0,497) of all explanatory variables was reported for SIH, hence the absence of severe collinearity can be assumed.

Plotting the standardized residuals against the standardized predicted values raised the suspicion of significant heteroscedasticity. This observation was confirmed by a highly significant result of the Koenker test (value=17,8, p<0,001). Hence, the RLM macro was used to obtain heteroscedasticity adjusted standard errors and confidence intervals.

Translating this multiple regression model into two separate models for healthy controls and IH patients, the following two models were obtained:

- For healthy controls:

$$SL1 = 14{,}3\ min - 5{,}87\ min \cdot \#SOREM$$

- For IH patients:

$$SL1 = 5{,}46\ min + 0{,}02\ min \cdot \#SOREM$$

5.2.2.2. Narcolepsy patients

The stepwise multiple regression of the narcolepsy patients did not include the TYPE1 variable into the model. Table 5.3 shows the regression coefficients.

Table 5.3: Regression coefficients, confidence intervals and p-values
41 cases of narcolepsy type 1 and 58 cases of narcolepsy type 2 were included
The confidence intervals and p-values were adjusted using Hayes' HC3 algorithm

Coefficients		**CI: lower limit**	**CI: upper limit**	**p-values**
Constant	5,66	4,12	7,21	<0,001
STYPE1	-0,319	-0,620	-0,017	0,039
#SOREM	-0,478	-1,07	0,118	0,115

A F-Value of F=22,8 (p<0,001) and an adjusted R^2=0,243 were reported. Again, the Durbin-Watson test (statistic: 1,76) gave no hint for significant autocorrelation. The Koenker test (value=14,9, p=0,001) led to the rejection of the assumption of homoscedasticity, so again the RLM macro was used. The coefficient corresponding to #SOREM was kept in the model, since its unadjusted p-value was below the cutoff of 0,1 for the stepwise exclusion. All tolerance values were reported to be higher than 0,5 , hence no severe collinearity had to be assumed.

In conclusion, the following linear models were obtained:

- For narcolepsy type 2:

$$SL1 = 5{,}66\,min - 0{,}478\,min \cdot \#SOREM$$

- For narcolepsy type 1:

$$SL1 = 5{,}66 min - 0{,}797\,min \cdot \#SOREM$$

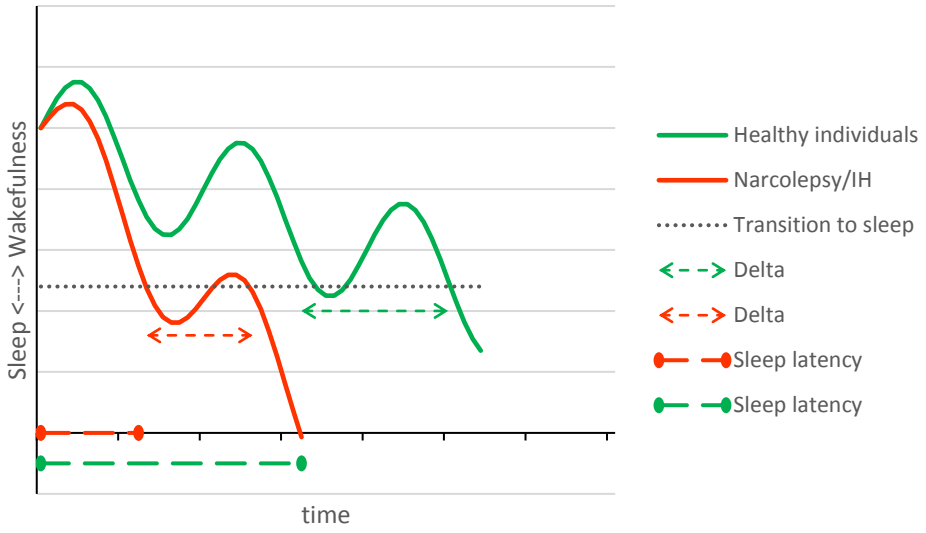
5.3. Delta and sleep latencies

The third linear regression of this thesis will address the possibility that an additional parameter apart from SOREM count and regular sleep latency to stage 1 might have a diagnostic and differential diagnostic purpose. Pizza et al. recently emphasized the possible value of Delta, which is defined as the difference between the sustained sleep latency and the usual sleep latency to stage 1 (SL1). It was pointed out that whereas healthy controls as well as IH patients show the physiological "waxing and waning" during the transition from wakefulness to sleep (hence having normal values of Delta), narcolepsy patients usually have significantly smaller values of Delta, reflecting a quick and direct onset of sleep without intermittent periods of wakefulness[4].

This observation raises the question whether Delta should be regarded as an additional parameter for the diagnosis and differential diagnosis of narcolepsy. To be of additive use in the diagnostic process, Delta should be "independent" from the existing parameters, especially #SOREM and SL1.

In the following linear regression analysis, the diagnostic value of Delta is addressed in a more precise manner. The linear model explained below suggests that Delta is merely a linear function depending on the sleep latency. Hence, low values of Delta could be merely a consequence of a high sleep pressure as represented by low sleep latencies. This approach can be further illustrated by the following graphical considerations.

Diagram 5.1: Motivation for the linear model



As the diagram illustrates a higher sleep pressure might by itself reduce the physiological "waxing and waning" process during the transition to sleep. If this were the case, differences in Delta would mostly be caused by differences in sleep latencies and no additional diagnostic benefit could be gained from Delta. If, in contrast to the explanation suggested above, Delta shows no direct dependence from the sleep latency, it might yield additional information regarding sleep onset in hypersomnolent patients.

5.3.1. Modeling assumptions

The considerations above can be summarized in the following short equation:

$$Delta = b_0 + b_1 \cdot SL1 + u$$

This arguably very simple model proposes a linear relationship between Delta and the sleep latency SL1. However, in order to take into account Pizza's recent results, possible differences regarding intercept and slope coefficients for each diagnostic group should also be included. At this point a subtle methodical issue occurs. Healthy controls have sleep latencies ranging from zero to 30 minutes (the upper limit is due to the MSLT30 protocol), whereas both IH and narcolepsy patients must have sleep latencies under 8 minutes. Therefore, every data point with SL1 greater than 8 minutes will be contributed by a healthy individual. Hence, the regression line for IH and

narcolepsy patients would extrapolate into the situation of sleep latencies above the 8-minutes cutoff, without allowing a reasonable interpretation.

These issues have been solved by splitting the linear regression into two groups again, this time into the patient and the control subgroup. For the healthy controls, one assumes the above model and estimates slope and intercept parameters based on the whole range of zero to 30 minutes. The remaining cases of the dataset will be used in a second linear regression, where again dummy variables will allow a differentiation between the diagnoses. This time, however, one defines IH as "reference group", so only four dummy variables representing both narcolepsy subtypes will appear in the model:

$$Delta = b_0 + b_1 \cdot SL1 + \delta_{N1} \cdot N1 + \lambda_{N1} \cdot SN1 + \delta_{N2} \cdot N2 + \lambda_{N2} \cdot SN2 + u$$

For IH patients, every dummy variable will be zero. The following linear regressions were performed completely analogous to the previous analyses. By setting IH as reference group, the significance of the narcolepsy dummy variables will always reflect the difference between IH and the narcolepsy subgroup at hand.

5.3.2. Results

5.3.2.1. Healthy controls

The following coefficients were obtained:

Table 5.4: Regressions coefficients, confidence intervals and p-values
73 control cases were included
No heteroscedasticity adjustment was needed

Coefficients		**CI: lower limit**	**CI: upper limit**	**p-values**
Constant	2,58	1,42	3,74	<0,001
SL1	0,004	-0,074	0,083	0,909

The correlation coefficient between Delta and SL1 was reported to be 0,014. Consequently, the R^2-value was smaller than 0,001. The Durbin-Watson statistic was 2,05 and therefore not pointing towards significant autocorrelation. The Koenker test statistic yielded 1,43 (p=0,232), so the assumption of homoscedasticity was not refuted. An F-value of 0,013 with a corresponding p-value

of 0,909 led to the conclusion, that the linear model using the coefficients has most likely emerged due to pure chance. In such cases, the constant model should be preferred. It should be noted that both coefficients were computed by minimizing the squared residuals. Therefore, the intercept coefficient is *not* the arithmetical average of all Delta values, which was 2,64 minutes.

5.3.2.2. IH and narcolepsy

The stepwise method excluded both intercept dummy variables, leading to the following model.

Table 5.5: Regressions coefficients, confidence intervals and p-values
All 141 patient cases were included
The confidence intervals and p-values were adjusted using Hayes' HC3 algorithm

Coefficients		**CI: lower limit**	**CI: upper limit**	**p-values**
Constant	0,303	-0,37	0,643	0,08
SL1	0,236	0,099	0,372	0,001
SN2	-0,168	-0,289	-0,048	0,006
SN1	-0,200	-0,333	-0,067	0,004

The Durbin-Watson statistic was 2,13 , indicating no significant autocorrelation. A highly significant F-value of 4,13 (p=0,006) was reported. The corresponding R^2-value was 0,147. All regression coefficients had tolerances above 0,7 , hence collinearity did not severely affect the results. The Koenker test statistic was 12,67 (p=0,005), therefore the heteroscedasticity adjusted errors had to be considered.

Hence, the following equations were obtained for the different diagnoses:

- For IH:

$$Delta = 0{,}303\ min + 0{,}236\ \cdot SL1$$

- For narcolepsy type 1:

$$Delta = 0{,}303\ min + 0{,}036\ \cdot SL1$$

- For narcolepsy type 2:

$$Delta = 0{,}303\ min + 0{,}068\ \cdot SL1$$

5.4. Discussion

5.4.1. Sleep latencies and SOREMs

Diagram 5.2: Regression line for IH patients and healthy controls

See table 5.2 for the regression coefficients.

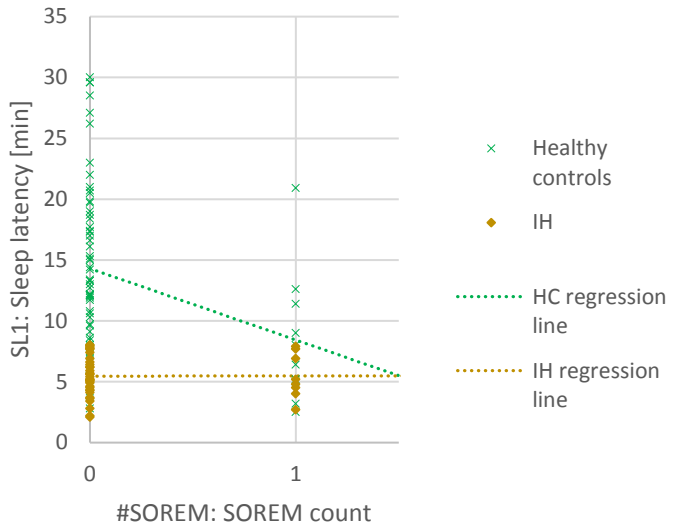


Diagram 5.2 illustrates the regression lines for IH patients and healthy controls. Healthy controls, who do not show any SOREM in the MSLT, have a mean sleep latency of 14,3 minutes, quite significantly above the 8-minute-threshold for IH and narcolepsy. In healthy controls who have a SOREM episode in one of the five MSLT sleep opportunities, an average of 8,84 minutes is found. Regardless of the SOREM count, the average sleep latency for IH patients is about 5,5 minutes.

As all cases of this analysis have at most one SOREM in the MSLT, the extrapolation of the regression lines towards the right end of the diagram is merely an extrapolation. Nevertheless, following the lines to the right highlights a central problem in the diagnosis of narcolepsy and idiopathic hypersomnia: Healthy controls who have one or even more SOREMs in the MSLT, possibly due to shift work, insufficient sleep etc., cannot easily be distinguished from IH and narcolepsy patients, at least not by referring solely to the 8-minute criterion for the mean sleep latency. It is important to remember that three healthy controls were excluded from the analysis due to their SOREM counts. Two of them had sleep latencies below 8 minutes (see table 5.1). This

as well as the extrapolated regression line suggests that for healthy individuals who show two SOREMs in the MSLT, sleep latencies can be expected to drop regularly below 8 minutes.

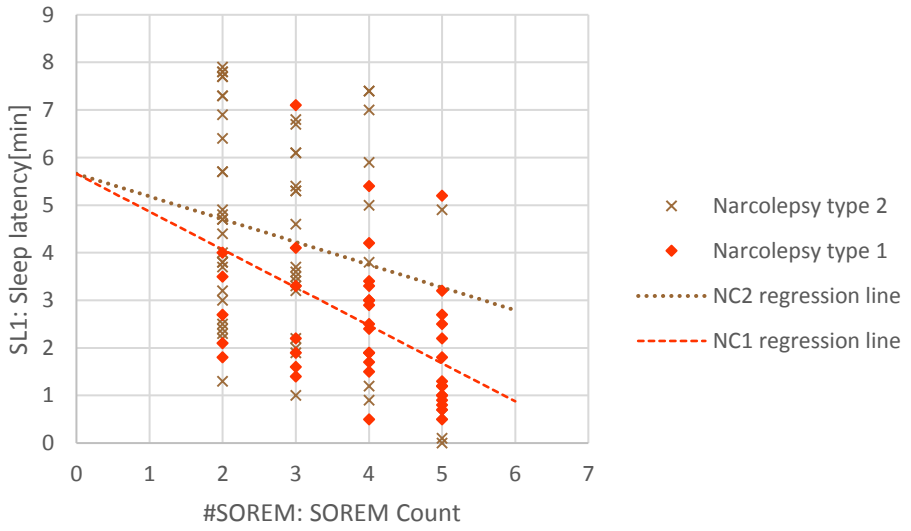
In conclusion, these results confirm that healthy individuals may show MSLT results, which could easily mimic typical MSLT findings for IH or narcolepsy. These observations correspond well to the findings of Allen et al., who reported that 13,1 % of men and 5,6 % of women in the general population show two or more SOREMs in the MSLT. In the same study it was found that 5,9 % of men and 1,1 % of women even fulfill both MSLT criteria that are requested for narcolepsy[164].

Regarding the IH patients one obtains the surprising result that the sleep latency in these patients is almost completely unaffected by the SOREM count. This finding will be further discussed after the regression results for the narcolepsy subgroups have been considered.

The reported corrected R^2-value was 0,369. As Dougherty[178] explains, this value can be interpreted as the fraction of the variance of the dependent variable that is explained by the independent variables. Therefore, only a third of the total variance in sleep latency can be associated with the SOREM count and differences between the diagnostic groups. Furthermore, as the scatterplot above indicates, the precision of the slope coefficient estimates might be compromised by the small number of data points with one SOREM.

Diagram 5.3: Regression line for patients with narcolepsy type 1 or 2

See table 5.3 for the regression coefficients



Looking at the second linear regression (corresponding to table 5.3), one notices an even lower R^2-value of 0,243. This suggests that only a quarter of the total variance in the sleep latency of narcolepsy patients is explained by #SOREM and STYPE1. Of the two dummy variables regarding narcolepsy type 1, only the slope variable proved to be significant enough for inclusion. Hence, one observes an identical intercept of both regression lines.

The negative slope coefficients for both subtypes represent the expected result: Patients with more SOREMs have on average a lower sleep latency. This effect is more pronounced in narcolepsy type 1 patients, who, as diagram 5.3 illustrates, tend to have a more severe narcolepsy phenotype. However, it should be recalled at this point that adjusting the regression analysis for the observed heteroskedasticity led to a p-value higher than 0,1. Therefore, if the assumed linear model would actually be wrong, the observed regression coefficient for #SOREM could still emerge by change, with a probability higher than 10 %.

The two regression analyses have yielded a negative linear relationship between sleep latencies and SOREM count for healthy controls and both narcoleptic subgroups. On the other hand, sleep latencies for IH are almost unaffected by the SOREM count.

It is important to consider that linear regression analysis is based on the calculation of the correlation coefficient. Therefore, in a strict sense, only conclusions regarding the correlation between independent and dependent variables can be drawn. Consequently, there is no way to decide whether an increase in SOREM count *causes* a reduced sleep latency (or/and vice versa) or if the correlation should be explained by a common hidden factor that affects both variables in the observed manner.

The negative linear dependence of sleep latencies from the SOREM count in healthy controls suggests that these two parameters are physiologically connected. It remains unclear whether one of the parameters highlights a primary attribute that consequently influences the other parameter or if SOREM count and sleep latency are affected by a hidden factor.

Considering the fact that an increased SOREM count occurs frequently in insufficient sleep syndrome and shift workers, a simple model would be the assumption of a common factor called “sleepiness”. In healthy controls who have an increased sleepiness, this could cause both the reduced sleep latency and the increased SOREM count.

As it has been mentioned above, several theories about the pathogenesis of narcolepsy exist. The theoretical considerations above somewhat reappraise the different approaches as listed by Dauvilliers et al.[186]. If, at its core, narcolepsy is primary a disease causing increased sleepiness, the decreased sleep latencies and increased SOREM count can be explained as different aspects of hypersomnolence. More severe cases would suffer from greater sleepiness, which then would lead to extremely high SOREM counts and sleep latencies well below the 8-minute threshold. If, on the other hand, narcolepsy is mainly a REM sleep disorder (as the occurrence of cataplexy suggests), one would assume that the decreased sleep latencies are a consequence of the REM sleep dysregulation.

However, the regression results for IH discourage the concept of a common factor "sleepiness" which causes the reduced sleep latencies and increased SOREM counts. At least in IH, hypersomnolence, as measured by the MSLT, is not correlated with the SOREM count. Hence, if one assumes mediating mechanisms between sleepiness (as measured by the MSLT) and REM proneness (as measured by the SOREM count) in the situation of healthy individuals as well as narcolepsy patients, these mechanisms do not seem to be effective in the case of IH.

These heterogeneous results allow the following conclusion. In general, sleep latencies and SOREM count should be treated as independent diagnostic parameters. The highest obtained R^2-value was 0,369, so almost two third of the variance in sleep latencies remain unexplained by the linear models above. However, for healthy controls and narcolepsy highly significant regression coefficients hint to a correlation between SOREM count and sleep latency that should not be neglected. The precise nature of this correlation and the existence of a direct causal relationship cannot be determined by the employed methods.

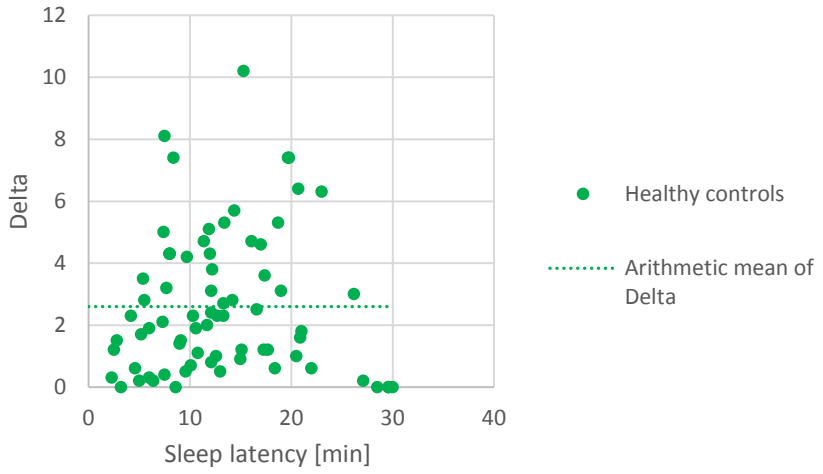
5.4.2. Delta and sleep latencies

The linear regression for the healthy controls is the only linear regression in this thesis which did not reach a significant F-value. The F-statistic is designed to investigate whether the obtained linear model might have emerged only by chance. The null hypothesis corresponding to this statistic is the constant model, where no linear dependence from any regressors exists.

In this case a slope coefficient close to zero was estimated together with an extremely unprecise confidence interval (compared to the coefficient value of 0,04). The p-value suggests, that if no linear relationship would be present, there is a chance of more than 90 % that a coefficient of 0,04

(or with a less extreme absolute value) would be estimated. Hence, in this situation the only sensible decision is to accept the null hypothesis and reject the assumption of a linear dependence between Delta and the sleep latency. Diagram 5.4 further illustrates this result.

Diagram 5.4: No correlation between SL1 and Delta in healthy controls



Correspondingly, the Pearson correlation coefficient is R=0,014, which represents almost perfectly uncorrelated variables.

Regarding the IH and narcolepsy patients a small but significant correlation was found. The highest slope coefficient was obtained for IH (see table 5.5), allowing the interesting observation that an IH patient with a borderline sleep latency of about 8 minutes is expected to have a Delta value of

$$Delta = 0{,}303\,min + 0{,}236\,\cdot SL1 = 2{,}4\,min,$$

which is rather close to the average Delta value in healthy controls (2,64 minutes). Hence in IH patients with unclear or inconclusive MSLT results, one should expect an almost normal Delta value and therefore not use Delta for diagnostic purposes. More clear-cut cases of IH on the other hand (i.e. patients having lower sleep latencies) show Delta values comparable to those of narcolepsy patients. Geometrically, this is reflected by the fact that IH and narcolepsy do only differ with respect to the slope coefficient, but not with respect to the intercept. Hence, differences between the groups regarding Delta diminish with decreasing sleep latency.

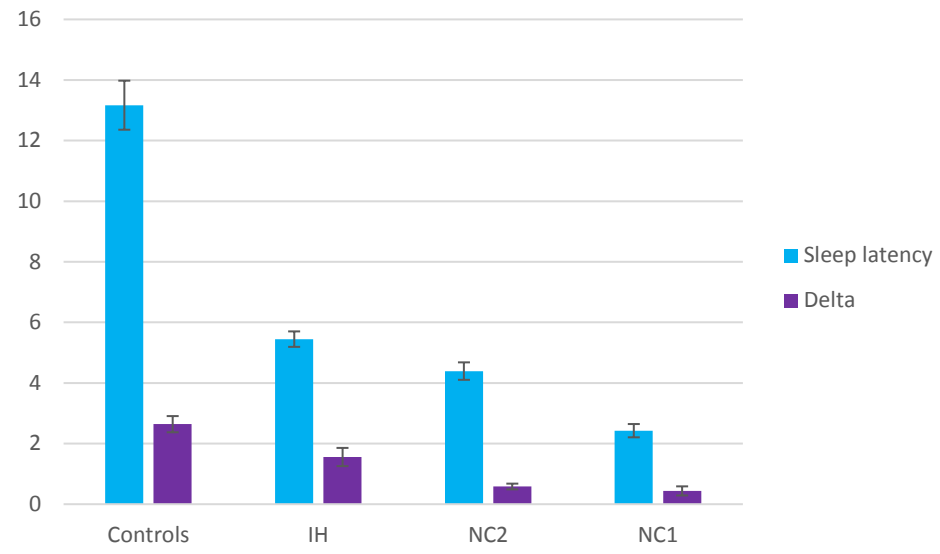
Continuing with the situation for narcolepsy patients, one notices that the slope dummy variables almost cancel out the regression coefficient for SL1 for both narcolepsy subgroups, thus a significant linear dependence between Delta and SL1 remains doubtful for narcolepsy as well. The point estimates for the regression variables suggest that a narcolepsy patient having a borderline sleep latency of 8 minutes will be expected to show a Delta value of less than 0,6 minutes for narcolepsy type 1 and less than 0,9 minutes for narcolepsy type 2.

Therefore, one can draw the following conclusion: In the case of healthy individuals and narcolepsy patients no significant dependency of Delta from SL1 became apparent. The expected Delta values of narcolepsy patients are below 1 minute regardless of their sleep latency, whereas healthy individuals show a Delta of about 2,6 minutes on average, although with great variance. For IH patients however, each additional minute of sleep latency raises Delta about 0,236 minutes, which leads to almost normal Delta values in cases with borderline normal sleep latencies and to Delta values close to those of narcoleptic patients for IH patients with severely reduced sleep latency. Finally, it is important to take the low R^2-value of 0,147 into account. Conversely, more than 85 % of the variance in Delta remains unexplained by linear dependence from sleep latency and diagnosis, suggesting that treating Delta as an independent parameter is justified from the perspective of linear regression analysis.

On average, analogous to the sleep latencies, the Delta values of IH patients lie between those of healthy individuals and those of narcolepsy patients. Thus, one reaches the conclusion that although no high correlation between Delta and the sleep latency could be found, regarding the average values the typical hierarchy “healthy controls > IH > narcolepsy type 2 > narcolepsy type 1”, that has been observed for the sleep latency[163], is also preserved in the parameter Delta (see diagram 5.5).

Diagram 5.5: Delta values and sleep latencies for controls, IH and narcolepsy patients

Error bars: Standard errors of the mean



5.5. Conclusion

Having performed the regression analysis, question I stated in 3.4. has been addressed. Informally speaking, the following results have been obtained.

1. For healthy controls and both narcolepsy subgroups, a significant linear dependence of sleep latencies from the SOREM count has been found. In healthy controls, individuals showing one SOREM in the MSLT have an expected sleep latency that is almost six minutes below healthy controls who have not presented a SOREM in the MSLT. In both cases, the average sleep latency remains above the 8 minutes threshold, although especially some of the healthy individuals with one or more SOREM episodes will drop below 8 minutes. In both narcolepsy subgroups there is a negative linear dependence of the sleep latencies from Delta. For each additional SOREM in the MSLT, sleep latency shows a steeper decline in narcolepsy type 1 than in type 2. Remarkably, no significant linear dependence between SL1 and SOREM count appeared in IH patients. Some interpretations involving the concept of general sleepiness have been suggested to explain this unexpected result.
2. The second regression analysis addressed a possible dependence of Delta from the sleep latency. No strong relationship between Delta and SL1 could be shown for IH patients.

Nevertheless, the obtained coefficients hint at Delta values close to those of narcolepsy patients if sleep latencies are low, but at almost normal values if sleep latencies get close to 8 minutes. In healthy controls, Delta and SL1 proved to be almost perfectly uncorrelated, whereas for both narcolepsy subgroups only a weak positive linear correlation could be observed. The low correlation values that have been reported for this analysis indicate a possible independent value of Delta for the diagnostic and differential diagnostic process.

6. Identifying important variables: Principal component analysis

Apart from anamnesis, the occurrence of cataplexy and the exclusion of other conditions, only the MSLT sleep latency, the MSLT SOREM count and the REM latency in the nocturnal PSG are considered in the current diagnostic criteria of IH and narcolepsy.

Aim II of this thesis therefore addresses the question if any additional parameters might be useful for distinguishing different groups in the given dataset of IH and narcolepsy patients and if the parameters listed above are indeed the most suitable ones for this task.

A priori, two requirements for a set of "important" variables seem reasonable: First, all relevant aspects of the variance in the dataset should be depicted in these variables, or – in other words – no important diagnostic dimension should be ruled out. Second, redundancy should be kept at a minimum. If two variables essentially provide the same information about the dataset, the inclusion of both would overemphasize their (differential) diagnostic impact.

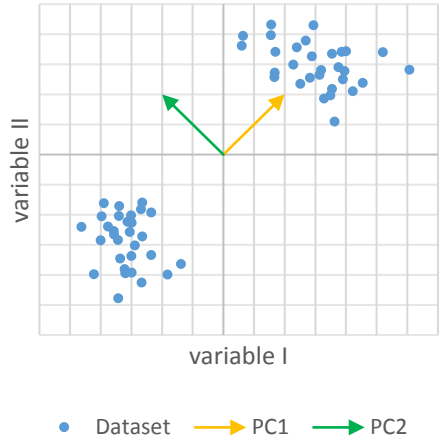
In this chapter a mathematical method called principal component analysis (PCA) will be presented that allows the identification of such suitable variables by transparent algorithmical decisions. PCA will yield a set of principal components (PCs), which essentially comprise the combinations of parameters that can explain most of the variance in the dataset without being too correlated to each other. First, the number of PCs that are needed to essentially grasp the dataset will be determined. Then, the PCs will provide the information needed to erase non-essential variables from the initial set of 14 parameters such that only variables suitable for cluster analysis remain. Due to the mathematical nature of PCA, both requirements discussed above will be considered during this process.

Apart from finding a set of suitable cluster variables, PCA will also contribute to this thesis in another distinct manner. Interpreting the most important PCs will allow additional insights into the structure of the dataset and its most important diagnostic axes.

6.1. PCA: An introduction

In terms of mathematics, principal components are the eigenvectors of the correlation matrix. Intuitively they are linear combinations of the given variables with certain beneficial properties. Diagram 6.1 demonstrates a low-dimensional example.

Diagram 6.1: Principal components: a two-dimensional example



As one can see, neither variable I or II are optimal in the way that they allow an easy identification of the two groups (or clusters) in the dataset. Performing a PCA on this simple dataset will yield two principal components, which are better suited to explain the dataset.

Each principle component has a corresponding "eigenvalue", which reflects the importance or significance of the principal component. One always obtains the same number of principal components as variables have been used as input for the PCA. These principal components are then usually sorted by their eigenvalues, which are always non-negative, in descending order. The PCs corresponding to the highest eigenvalues explain most of the dataset variance: Clearly, in the example of diagram 6.1, PC1 is the most important axis for distinguishing the two clusters. Usually only the first few principal components are considered, which effectively reduces the dimension of the dataset and simplifies further analysis. Therefore, PCA is sometimes referred to as a tool for dimensional reduction.

However, this procedure cannot differentiate between variance due to "real" groups or clusters and variance due to imprecise measurements or just random distribution. Therefore, the results of PCA

should be treated with care and a priori knowledge about the dataset should be used to discuss the validity of the obtained principal components.

As explained before, every PC is defined as a linear combination of the input variables $x_1,\ldots, x_n$:

$$PC = \lambda_1 x_1 + \ \ldots + \lambda_n x_n \ .$$

The real-valued coefficients $\lambda_1, \ldots, \lambda_n$ determine how the given PC can be constructed as a weighed sum of the input variables. These coefficients are often also called *loadings* of the PC with respect to the variable at hand. High absolute coefficient values indicate a high impact of the corresponding variable to the PC. The input variable yielding the highest (absolute) coefficient value will also be called the dominant variable of the given PC.

6.2. The PCA algorithm and the total variance criterion

The PCA was applied to the dataset of all IH or narcolepsy patients (141 individuals in total). All acquired variables were used except susSL. The reason for this exclusion is that susSL is defined as SL1+Delta, hence its inclusion would not contribute additional information to the dataset but only add redundancy. Hence, 14 variables in total were included into the PCA, which consequently yielded 14 PCs.

Furthermore, the algorithm relied on the correlation matrix (which can be found in the appendix) with respect to the 14 input variables. In contrast to the alternative, the covariance matrix, all variables are implicitly standardized. This eliminates unwanted weighing of the variables due to the different measurement units. Otherwise, variables allowing high absolute values with respect to their measurement unit would have an increased impact on the calculation of the PCs.

For the consecutive cluster variable selection, one must agree on the number of needed principal components. As Jolliffe explains in detail, several criteria have been elaborated allowing a standardized way of defining this number. In this thesis, the total variance criterion is used, which states that starting from the first PC as many PCs are included as needed to explain a certain fraction of the total variance in the dataset[187]. Here, a total variance of 80 % was defined as the threshold fraction.

The PCA was performed using the FACTOR syntax command of SPSS 23.0.0.0[182].

6.3. Cluster variable selection

While it is certainly possible to use the first few PCs directly for cluster analysis, a method will be discussed in the following, which relies on the PCs to choose suitable cluster variables. Jolliffe[187] lists several different methods for variable selection. Most methods share a common principle: Replacing each PC with its dominant variable. Highly correlated variables are likely to have similarly high loadings (in absolute terms) on the same PC. Hence, only one dominant variable of each significant PC should be chosen for further analysis, since otherwise there is the risk of including two variables that are highly correlated. Similarly, if one relies on the omitted PCs to select variables that are not to be included in the cluster analysis, it is reasonable to rule out the dominant variables of the most insignificant PCs.

In this thesis, a standard procedure recommended by Jolliffe was chosen, that has been demonstrated to yield sensible variable selections in simulated data[188]. Table 6.1 describes the precise algorithm, which uses the least significant PCs to find variables that can be left out without risking too much loss of information.

Table 6.1: The cluster variable selection algorithm[187]

Given n variables (and therefore n PCs), k cluster variables have to be selected

- Sort the PCs by their eigenvalue in descending order
- VarSet is defined as the set of **all** given n variables
- Repeat the following steps for the **last n-k PCs, in ascending order (starting with the nth PC)**
    - For the selected PC, sort all n variable by the absolute value of their coefficient for this PC, in descending order
    - Identify the first variable which is still included in VarSet
    - Remove this variable from VarSet
    - Select the preceding PC; repeat the steps above
- After n-k PCs considered, k variables remain in VarSet
- Use the remaining k variables in VarSet as cluster variables

6.4. Results

Applying the 80%-variance criterion, the first seven principal components had to be included. For a more detailed look on the first seven PCs, a recommendation by Jolliffe was followed: In each column, "++" indicated the dominant variable of the PC, whereas other important contribution are marked by "+" or "-", depending on the sign of their loading coefficient.

Table 6.2: The first seven principle components
++ : dominant variable, i.e. highest absolute loading
+/- : coefficient with an absolute value of 50% or more compared to the dominant variable;
(+)/(-): absolute value between 50% and 25%

	1	**2**	**3**	**4**	**5**	**6**	**7**
TST	++		(+)		(+)	(-)	
SL2	-						(-)
SL1	-					(+)	-
#SOREM	+				(-)	(+)	
CATAP	+	(+)	(-)	(-)	-		(+)
Delta	-		-	(+)	(-)	(+)	+
PSG_REML	-			-		-	+
VigCorr		++	(+)				
VigRT		-	-		(+)	(-)	(+)
VigFalse		-	-	-		++	
PSG_N3_TST		(-)	++			+	++
ESS			-	++	-	(-)	
PSG_SEI		(-)	+	+	(+)	(+)	
PSG_AI		+	-		++	(+)	(+)
Eigenvalue	4,1	1,9	1,6	1,2	0,9	0,86	0,73
Cumulative explained Variance	28,9 %	42,6 %	54 %	62,6 %	69,1 %	75,2 %	80,4 %

Starting with all 14 input variables, the algorithm for variable selection as it is described in table 6.1 was applied. For the principal components 14 to 10 as well as 8 their dominant variable could be removed. In the case of PC9 the second most important variable (CATAP) was chosen since its dominant variable #SOREM had been removed when considering PC 12.

Hence, after the removal of the variables TST, SL2, #SOREM, VigCorr, VigRT, CATAP and PSG_N3_TST the following variables remained for the cluster analyses: **ESS, PSG_AI, VigFalse, Delta, SL1, PSG_REML, PSG_SEI.**

6.5. Discussion

6.5.1. The selection of cluster variables

Although the variable selection was done in a purely algorithmical manner, the choice of cluster variables seems very sensible. From the group of highly correlated (see table A.1 in the appendix)

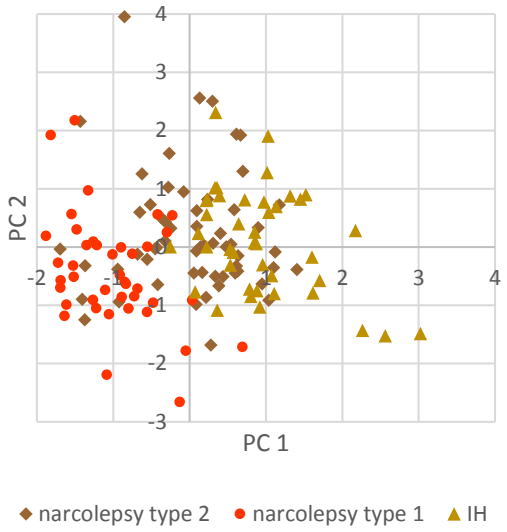
variables SL1, SL2 and TST, only one has been included as a marker of objective sleepiness. VigFalse represents the vigilance test performance of the patients whereas PSG_AI takes the fragmentation of the night sleep into account. Subjective sleepiness is considered by the inclusion of the ESS score. Remarkably, Delta also emerged as an important variable, which highlights its possible further use. Although the MSLT SOREM count is essential for the differential diagnosis between IH and narcolepsy, it has not been selected as cluster variable, but is indirectly represented by the PSG REM latency. Another interesting observation is that despite its diagnostic significance, the occurrence of cataplexy (CATAP) is not included.

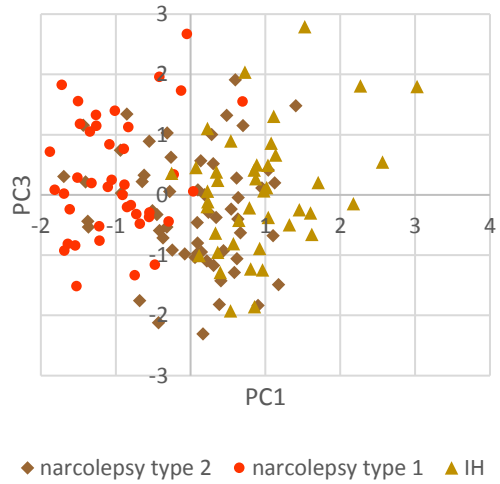
At this point one could argue that a more sensible way of choosing cluster variables is the "arbitrary" way by relying on clinical experience and other cluster analyses in the field of sleep medicine. However, the selection procedure based on PCs can easily be reproduced for other datasets and transformed to other sleep medical questions. Furthermore, being a tool for explorative data analysis, cluster analysis suffers inherently by the fact that several parameters of each analysis must be set by "try and error". By agreeing on objective selection methods, one ensures that at least the variable selection process is reproducible and based on objective calculations, which strengthens the validity of the cluster solutions.

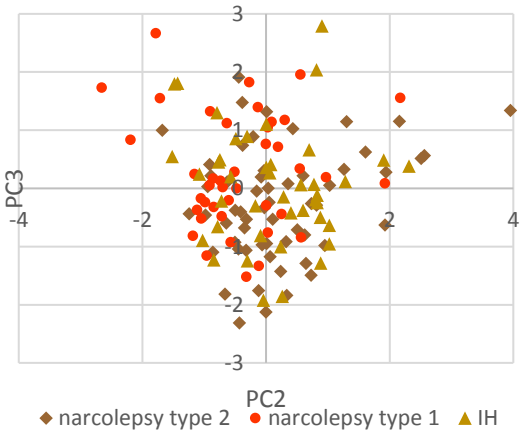
6.5.2. The selected principal components

Seven principal components were needed to explain 80 % of the dataset variance. Before discussing the seven most important PCs in more detail, another main feature of the PCs will be demonstrated: Their ability to reduce the dimension of the dataset with only a minimized loss of information. One can therefore use the first three PCs to plot three two-dimensional projections of the dataset, which are more suitable than the input variables for visualizing the variance in dataset.

Diagram 6.2: The dataset projected onto the first three principal components







Several interesting observations can be made in these diagrams. All diagrams show a homogeneous cloud of data points with several outliers on each side. No obvious clusters become apparent at first sight. Regarding the different diagnoses, PC1 seems to be able to separate narcolepsy type 1 from IH patients, whereas narcolepsy type 2 patients lie in between these groups. This reflects the clinical observation, that narcolepsy type 2 patients usually have a less pronounced phenotype and are much harder to distinguish from IH than narcolepsy type 1 patients. PC2 and PC3 do not show a similar property, neither do they suggest an alternative clear-cut clustering of the data. Judging from these diagrams, diagnosis of IH and narcolepsy seems to happen mainly with respect to the variables incorporated in PC1, although a substantial overlap can be observed. As it will become apparent later, PC1 is indeed closely related to the current diagnostic procedure for narcolepsy and IH.

Additionally, the diagram above does not bolster the present ICSD-3 diagnostic groups, since it suggests that the patients in the dataset form a continuum of cases which cannot be easily separated with respect to the considered parameters.

Hence, from a methodical point of view, no very consistent and precise cluster results should be expected. More likely, the different cluster methods that are presented below will produce quite different results, which might be very vulnerable to the variation of the clustering parameters.

Next, the first seven PCs are discussed in more detail. The dominant variable as well as all other variables that have been marked without brackets in table 6.2 will be listed, allowing a direct interpretation of the significant variable contributions. Here, most PCs will receive a descriptive name which characterizes one of the contrasted phenotypes. It is important to acknowledge that for each typical phenotype, its direct opposite is equally well represented on the “opposite end” of the given PC.

PC 1: IH vs. narcolepsy phenotype

TST	#SOREM	CATAP	SL2	SL1	Delta	PSG_REML
++	+	+	-	-	-	-

As diagram 6.2 indicates, the by far most dominant PC (with a corresponding eigenvalue of 4) manages to split up the different diagnostic subgroups. Looking at the coefficients of PC1, one can

easily see why. On the one end of the PC1 axis one finds patients who quickly fall asleep in the MSLT and have low values of SL1 and Delta. Furthermore, these patients are likely to have many SOREM episodes and have an increased probability of a history of cataplexy. Just as one would expect from patients suffering under cataplectic attacks, the REM latency in the PSG is also reduced. This description fits almost perfectly to the typical narcoleptic symptoms. On the other end of the PC1 axis one consequently finds more "IH like" patients showing less pathological MSLT results, high values of Delta and no sign of a REM sleep dysregulation (high REM latency, no cataplexy).

PC 2: Vigilance test performance

VigCorr	PSG_AI	VigRT	VigFalse
++	+	-	-

This PC is essentially a summarizing score for the vigilance performance. The two extremes in this PC can be described in the following way: Patients having a high number of correct reactions, who were also rather quick, and, on the opposite end, patients showing few correct reactions, a high reaction and a lot of false reactions. Interestingly, PSG_AI also contributes to this PC, hence patients having a high positive score on this PC (and therefore showing a good vigilance test performance) are likely to have a high arousal index in the PSG.

PC 3: Deep sleeper with good vigilance and low subjective sleepiness

PSG_N3_TST	PSG_SEI	VigRT	VigFalse	ESS	PSG_AI	Delta
++	+	-	-	-	-	-

Including the variables marked within brackets above, 10 of 14 variables contributed significantly to this PC, making it hard to find a reasonable interpretation.

As the title of this PC indicates, patients with a high score regarding this PC are efficient sleepers who spend a lot time in sleep stage 3, which may be a consequence of their low arousal index. Furthermore, these patients score themselves low in the ESS, which indicates a low subjective sleepiness and show good reaction and few errors in the vigilance test. Interestingly, these patients

also have low Delta values, although none of the usual MSLT parameters (apart from a weak contribution by TST) contribute significantly to this PC.

Negative values on this PC indicate disrupted sleep associated with a high subjective sleepiness and a bad vigilance test result. High Delta values hint at a slow transition to deeper sleep.

Obviously, this PC does not correspond to any of the current diagnostic concepts. Nevertheless, an eigenvalue of 1,6 suggests that a considerable fraction of the total variance can be projected onto this axis, which does not include the usual MSLT parameters.

PC4: High subjective sleepiness despite efficient sleep

ESS	PSG_SEI	PSG_REML	VigFalse
++	+	-	-

ESS is the dominant variable of PC4. Patients with a high positive score in PC4 have a high sleep efficiency despite their significant subjective sleepiness. Furthermore, their REM latency is decreased, which, however, is not associated with high frequencies of cataplexy ((-) loading for CATAP). PC3 and PC4 highlight the importance of the subjective sleepiness estimation using the ESS.

PC 5: Many nocturnal arousals, low ESS

PSG_AI	CATAP	ESS
++	-	-

Only three variables contribute notably to this PC, which characterizes patients that show many arousals in their night sleep, but at the same time do not report elevated levels of sleepiness.

PC 6

VigFalse	PSG_N3_TST	PSG_REML	SL1	#SOREM	Delta	PSG_SEI	PSG_AI	TST	VigRT	ESS
++	+	-	(+)	(+)	(+)	(+)	(+)	(-)	(-)	(-)

In this case, the weak contributions have been listed as well, since PC6 has significant loadings from 11 out of 14 input variables. In general, interpretation of a PC gets more difficult if more

variable loadings have to be considered[187]. Therefore, no clear characterization could be found for this PC.

PC6 is dominated by false reactions in the vigilance test. Regarding the other two vigilance test parameters, one finds only a low contribution by VigRT. A possible explanation for high positive scores in this PC6 would be the inability to fully grasp the vigilance test instructions, leading to a high number of false reactions. High scores in this PC are also associated with a long sleep time in N3 and a low REM latency.

PC 7: Fast sleep onset, but high Delta values

PSG_N3_TST	Delta	PSG_REML	SL1
++	+	+	-

Like PC3, PC7 is dominated by PSG_N3_TST. Patients scoring high on PC7 are also likely to have high Delta values and a late REM sleep onset. Furthermore, their sleep latency is low. Hence, positive PC7 scores represent patients that fall asleep quickly but need some time to reach sustained sleep. REM sleep is reached comparatively late, but much time is spent in sleep stage III. Apart from the low SL1 values, this PC pronounces typical results for IH patients. The other end of the PC7 axis represents patients, whose Delta and PSG_REML values indicate a narcolepsy-like phenotype, which is however accompanied by an atypically high sleep latency

Summarizing these interpretations, it becomes clear that most PCs of this dataset cannot easily be characterized from a clinical perspective. Nevertheless, the following observations can be made: PC1 comprises all the usual MSLT parameters SL1, SL2, #SOREM, TST and contrasts the typical narcolepsy phenotype against IH. These parameters are almost completely neglected in the following PCs, which however show significant contributions by Delta in the case of PC3 and PC7. This suggests that Delta inherits additional information that is not summarized in the usual MSLT parameters and therefore not completely comprised by PC1.

For the emergence of PC1 as first and therefore most significant PC the following technical aspects should be noted, too. First, the occurrence of cataplexy, the MSLT SOREM count and the PSG REM latency are the three variables that are currently used for differential diagnosis between

narcolepsy subtypes and IH. As the dataset comprises a balanced selection of narcolepsy and IH cases, a lot of variance is to be expected along these variables. Second, by including the highly intercorrelated variables SL1, SL2 and TST (see table A.1 in the appendix) into the PCA, implicitly the sleep duration and, conversely, the sleep latency in the MSLT became a more pronounced aspect of the dataset, since it is directly or indirectly represented in three of 14 dataset dimensions. Therefore, the structure of PC1 is at least partially a consequence of the emphasis that was put on MSLT results in the 14 selected input variables. However, as the description of PC1 illustrates, variables regarding REM sleep characteristics also contribute significantly to the first PC. In the linear regression analyses, a certain correlation between SOREM count and sleep latencies highlighted the relationship between sleep duration/latency and REM sleep occurrence, which is most likely another important reason for the observed structure of PC1.

In PC3 several markers for good sleep quality (a high sleep stage III fraction, a high sleep efficiency and a low arousal index) as well as low ESS scores are linked to low values of Delta, whereas PC7 pronounces phenotypes where the MSLT hints at severe sleepiness that is combined with high Delta values. Therefore, no concise interpretation of the Delta values can be given at this point.

The variable loadings of PC2 were not to be expected a priori. As vigilance tests are not included in the official diagnostic criteria, it is a remarkable finding that the overall vigilance test performance as encoded by PC2 represents the second most important axis along which the dataset variance can be projected. The correlation matrix of all input variables (see table A.1 in the appendix), based on which PCs are obtained as eigenvectors, reveals that all three vigilance test variables are moderately correlated to each other (all correlation coefficients have absolute values above 0,3), whereas only a weak correlation appears with respect to any other input variable (absolute R-values below 0,17). Consequently, PC2 essentially enbodies the variance caused by all vigilance parameters, with no significant effects from any other variable except PSG_AI. Whereas this result suggests the “independence” of vigilance test results from the other sleep medical parameters that have been considered, the usefulness of vigilance test parameters remains to be demonstrated by cluster analysis. One should also keep in mind that, similar to the situation of TST, SL1 and SL2 in PC1, the relative importance of PC2 can also be explained by the decision to include three at least moderately corelated variables which encode different aspects of the vigilance

test. If two or only one vigilance test variables would have been considered for the PCA, the corresponding PC would be less significant.

6.6. Summary

The mathematical method of principal component analysis yielded some insights regarding the relative importance of all variables that have been considered initially in the dataset (see chapter 5). The principal components that have been obtained in this chapter are essentially coordinate axes along which the dataset can be described in the most sensible way. However, as it has been discussed for PC1 and PC2, the effect of the concrete choice of input variables should not be underestimated. The following core statements will serve as a brief recap of the main results of the PCA.

1. According to the distinct structure of the most important PC, MSLT sleep latency, SOREM count, occurrence of cataplexy and PSG REM do indeed explain a large part of the dataset variance when they are used to distinguish the typical findings for IH and narcolepsy. This is a certain justification for the current diagnostic criteria that rely especially on MSLT findings.
2. The second most important PC is essentially an overall vigilance test score. Although this PC (or score) also explains a considerable amount of the dataset variance, it is possible that this finding is just an artificial consequence of including vigilance test results into the PCA in the first place and therefore provides not much additional information regarding the remaining structure of the dataset.
3. Regarding the Delta value that has been suggested by Pizza et al., one notices that Delta contributes significantly to PC3, which does not take into account any other MSLT variable to a relevant degree. This finding supports the suggestion that Delta might indeed comprise additional information regarding the transition from wakefulness to sleep
4. For all but the first two PCs, it is very difficult to find a reasonable interpretation for their component structure.
5. Using a certain cluster variable selection algorithm, the following variables have been chosen and will further be used in cluster analysis: **ESS, PSG_AI, VigFalse, Delta, SL1, PSG_REML, PSG_SEI.** As it has been pointed out above, this is probably a reasonable

selection, since it includes all important concepts for sleepiness that have been discussed above while avoiding unnecessary redundancy.

7. Finding groups in the dataset: Cluster analysis

In this chapter the final statistical analyses will be performed, which will mainly address question III stated in section 3.4. Intuitively, cluster analysis summarizes several different methods which can detect groups in a dataset. The word "group" as well as "cluster" suggests that all members of a group (or equivalently, elements of a cluster) are closely related to each other. On the other hand, elements of separate groups should ideally be unrelated, dissimilar and distant to each other.

These basic considerations reveal a central abstract concept that can be approached in various manners: Before groups in the dataset can be identified, one has to agree on what a group is. This, however, involves a certain definition of similarity/closeness or dissimilarity/distance between the groups for the concrete situation at hand. Three different algorithms will be presented in this thesis that represent different approaches on how a group should be defined and what distance means for elements in the given dataset.

These three different algorithms will yield three different suggestions regarding the optimal subdivision of the dataset, which will be called *cluster solutions*. This leads directly to the question, how the different suggestions should be compared to each other and to the groups/diagnoses defined in the ICSD-3. The process of estimating the quality of a cluster will be referred to as *cluster validation* and will be addressed by statistical and visual measures.

Eventually, the validation of the different cluster solutions as well as the detailed comparison between the solutions will justify a detailed discussion regarding the optimal classification of IH and narcolepsy patients and the implications on current diagnostic concepts.

7.1. Introduction to Cluster analysis

7.1.1. Principles of cluster analysis

According to Tan cluster analysis divides data into several groups, which are also called clusters, that are meaningful or useful[189]. Schendera describes cluster analysis as an objective data classification method and points out that a fundamental implicit assumption is always made when cluster analysis is performed: The existence of meaningful or useful groups that can be discovered by the cluster analysis[190]. In present case of this thesis, the existence of several different reasonable

groups of patients is assumed. These groups are not necessarily identical to the diagnostic groups of the ICSD-3.

Obviously, these assumptions can and should be subject of further discussions. Cluster analysis in general cannot decide whether the dataset structure actually allows reasonable clustering and will therefore always produce some clusters regardless of the properties of the dataset. Hence the user of the cluster analysis must investigate the cluster solutions and possibly refute the obtained clusters as mere mathematical artifacts that bear no meaning or usefulness. In this context cluster validation methods are valuable tools and will be discussed below in further detail.

For the conduction of cluster analysis several technical decisions have to be made. The first decision involves the right notion of distance or similarity. Every definition of distance can in some way be inverted to obtain a notion of similarity and vice versa, as similar elements show a small distance between them and elements that have a large distance between them are expected to be dissimilar. Hence, it suffices to introduce a sensible notion of distance into the given sleep medical dataset. Here, the selection of cluster variables is crucial as they provide the fundamental information based on which the distances between each pair of elements can be calculated. Usually, at the start of a cluster analysis, the distance relations between all element pairs are listed in a distance matrix: The distance between the i-th element and the j-th element can then be found in the i-th row and j-th column.

Having agreed on a distance measure, the actual cluster analysis methods differ greatly in the way how the information stored in the distance matrix is processed to determine the optimal clusters.

Three different methods will be employed: k-means, OPTICS and spectral clustering. The k-means algorithm starts with predetermined suggestions for the cluster centers, which are then iteratively refined. Hereby, dataset elements are always assigned to the cluster whose center is closest to them. As a result k-means clusters tend to be ball-shaped and compact. In the OPTICS method, however, clusters are detected as regions with an increased density of elements. Therefore, outlier elements that are too far away from any other dataset point are unlikely to be assigned to any cluster. Because of that, OPTICS usually yields a set of "noise elements" that could not be assigned to any cluster. Finally, spectral clustering assigns elements, which can be connected via a path consisting of similar neighbors, to the same cluster. This approach allows the formation of more atypically shaped clusters.

One important cluster parameter will be crucial in all three cluster methods: The number of expected clusters. Here, the approach of Šonka et al. is followed, who stated as "null hypothesis" that there are as many clusters in the dataset as there are a priori known diagnostic groups[163]. Therefore, as the ICSD-3 diagnoses are challenged in this thesis, three clusters are expected to emerge. Hence, for all three cluster analyses, we state as a "null hypothesis", that three clusters exist in the dataset, which are identical to the ICSD-3 diagnostic groups.

The number of clusters will directly be specified in the k-means method, whereas in spectral clustering and OPTICS, no direct way of adjusting the cluster number exists. Hence, k-means will yield three clusters regardless of the true structure of the dataset, so only properties and sizes of the three clusters will provide evidence if the null hypothesis should be rejected.

Depending on the concrete clustering algorithm several other parameters will have to specified. Often, the right choice of clustering parameters has enormous impact on the success of the cluster analysis, i.e. in the obtainment of reasonable groups. Usually, only heuristic approaches exist regarding the right parameter choices. Therefore, in this thesis sometimes a try-and-error approach had to be employed, but the final choice of cluster parameters will be justified as well as possible.

7.1.2. Cluster evaluation methods

A multitude of different cluster analysis algorithms are available, which usually provide highly differing cluster solutions when applied to the same dataset. A priori, it is often not possible to predict which algorithm yields the best results for the given task. Hence, tools for *cluster evaluation* are required which allow the estimation of the overall quality of a cluster solution and the comparison between different solutions.

According to Tan et al., cluster evaluation (which is sometimes also referred to as cluster validation) can be divided into two different approaches: *Internal validation* uses the information that is encoded in the cluster variables and hence has been used in the clustering process. Therefore, no external measure of cluster quality is applied to the cluster solutions. *External cluster validation* introduces external information, that has not been used in the cluster analysis itself (such as predefined diagnostic groups, results from other cluster analysis, variables that have not been used in cluster analysis) to assign some measure of quality to each cluster solution[189].

In this thesis two distinct kinds of external information are available: Variables that have not been used in cluster analysis and the preexisting ICSD-3 diagnoses. Both will be taken into account and will serve as measures for external cluster validity. The non-cluster variables will be included in an ANOVA analysis which – by design – checks if the obtained clusters differ regarding these variables. The finding of significant differences would provide a hint that the clustering has been successful and has yielded clusters that differ even in variables that did not enter the actual cluster analysis.

Another external validation will implicitly be performed when, following the three cluster analyses, all solutions will be compared to each other and to the ICSD-3 diagnostic groups using the Rand coefficient. As it will be explained in more detail below, the comparison of an obtained cluster solution to the existing diagnostic groups is a process of external validation. However, using the diagnoses for external validation would in some way collide with the task of this thesis, which is the exploration of the dataset to challenge and possibly improve the current diagnostic groups. Therefore, the ICSD-3 diagnoses are certainly external information, but should not be treated as a measure for cluster validity in this situation.

The most obvious approach for estimating the success and quality of a given cluster solution is describing differences between the obtained clusters regarding all 15 initially acquired variables. Assuming normal distribution for every variable in each cluster, ANOVA analysis is suitable for this task but must be interpreted with caution. The p-values for each variable, which reflect the significance of the differences between the clusters, cannot be treated in the standard way. Since all cluster methods are designed to find groups which differ with respect to the cluster variables, finding such differences in the cluster solution is not a surprising result. Hence, a highly significant p-value of a cluster variable does not by itself reflect a high quality of the cluster solution. P-values of cluster variables will only then be of any use if they show p-values *above* the significance level, therefore indicating *no* significant difference and consequently hinting at a failed clustering process. On the contrary, significant p-values regarding non-cluster variables serve as a measure of external validation and allow a certain insight into the quality of the cluster solution.

However, ANOVA remains a procedure that essentially focuses on the mean values for each variable. For non-convex clusters, both mean values and standard deviations might not provide a suitable characterization of the cluster structure. In such situations, the cluster plots (diagram A.1

to A.3 in the appendix) and other validation methods are essential for a more thorough understanding of the cluster solution.

ANOVA analysis is a parametric statistical test that is based on several assumptions regarding the input variables, with the most prominent one being the normal distribution for each cluster. For most input variables normal distribution is a reasonable assumption. However, VigFalse, VigCorr, ESS and CATAP are measured on discrete scales, rendering the assumption of approximate normal distribution harder to justify. Especially in case of atypical cluster shapes, the assumption of normal distribution might also be violated for other variables. As in this thesis all ANOVA analyses are only used for a first characterization of the cluster solutions and hence serve mostly descriptive purposes, these methodical weaknesses are accepted.

However, for the binary variable CATAP the Fisher-Freeman-Test for independency will be employed. Essentially, the test measures if the frequency of cataplexy is independent from the cluster assignment. High values of the test statistic and corresponding low p-values therefore hint at a significant difference in cataplexy occurrence between the clusters.

It will turn out that the OPTICS clustering procedure only yields two clusters and an additional set of noise points. In order to focus on the comparison between the "real" clusters, a t-test will be performed instead of an ANOVA. Correspondingly, the Fisher-Freeman-Test will also not consider the noise cluster.

As mentioned above, good clusters are characterized by two properties: First, elements belonging to the same clusters are close to each other, and second, elements assigned to different clusters have a large distance between them. Mathematically, both attributes can be formalized and explicitly calculated for each point. These attributes are called *cohesion* and *separation*, respectively. For a sensible characterization of a cluster solution, it seems reasonable to employ a measure that incorporates both aspects. The *silhouette coefficient* is designed to take both cohesion and separation into account. It uses the distance matrix and cluster assignments as input and calculates a silhouette coefficient for each element of the dataset. The arithmetic average of all elements can then serve as an estimate for the overall quality of the cluster solution, whereas the mean of the coefficients restricted to a single cluster allows a comparison of the different clusters of a given cluster solution.

For a third measure of internal validation a recommendation by Tan et al was heeded which suggests plotting the distance matrix of the dataset after the elements have been sorted by their cluster assignment. Ideally, one would expect a matrix of block-diagonal shape with each cluster being represented by a square with clear cut borders[189]. These matrix visualizations allow a quick estimation of the quality of each cluster and are very suitable for comparison of different cluster solutions.

Table 7.1 summarizes all employed measures of internal and external cluster evaluation.

Table 7.1: Measures of internal and external cluster validation	
Internal	**External**
ANOVA: cluster variables Silhouette coefficient Distance matrix plot	ANOVA: non-cluster variables Rand coefficients: comparison to the ICSD-3 diagnoses

7.2. K-means clustering

7.2.1. K-means clustering: iterative center calculation

K-means is one of the oldest and most widely used cluster algorithms. The term "k-means" dates back to MacQueen in 1967[191], but the standard algorithm that will also be applied in this thesis has first been published by Lloyd in 1982[192].

Starting with k initial cluster centers, k-means assigns each element to the closest cluster center. After having assigned all elements in that way, the new cluster centers are calculated as centroids of the clusters that have emerged by the assignments. The centroids are defined as the arithmetical mean value of all elements of the given cluster. Then, the whole procedure is repeated using the newly obtained cluster centers. This algorithm leads to "slowly moving" cluster centers, whose movements tend to stagnate after some iterations. If all cluster centers cease to move further than a predefined threshold, the algorithm stops and the final clusters and cluster centers are reported[189, 190].

One major issue with k-means clustering is the choice of the initial cluster centers, which might severely affect the final cluster solution. One way of minimizing this effect is running the k-means clustering multiple times on the same dataset, each time starting with different randomly selected cluster centers.

In this thesis, the centroids of the ICSD-3 diagnostic groups are supplied as initial cluster centers. This decision is consistent with the agreement on three expected clusters: The algorithm starts with the "null hypothesis" cluster centroids and has the chance to refine these centroids and consequently the initial diagnostic clusters.

It should also be noted that k-means tends to produce clusters of comparable sizes and of convex shape (i.e. "ball-shaped"). If a priori clusters of atypical shape or very different sizes are to be expected, k-means might yield suboptimal results. Furthermore, k-means shows also some weaknesses in detecting clusters of different densities and in handling a dataset containing outliers[189].

7.2.2. Implementation and cluster parameters

The k-means clustering was performed using the built-in implementation QUICK CLUSTER (method "KMEANS(NOUPDATE)") in SPSS v.23.0.0.0. In order to prevent implicit weighing of the variables, a z-standardization of all cluster variables was performed. The cluster number was set to three, whereas as initial centroids the standardized centroids of the given three diagnostic groups were used. A maximum of 100 iterations was allowed, and 0,0001 was set as the threshold distance for the centroid refinement. For both the calculation of the centroids and the assignment of cases to the different centroids, a notion of distance must be prespecified. Here, the common Euclidean distance with respect to the selected cluster variables was used.

Table 7.2: k-means: Standardized centroids of the diagnostic groups
NC 1/2: Narcolepsy type 1/2

	ESS	PSG_AI	VigFalse	Delta	SL1	PSG_REML	PSG_SEI
NC 2	-0,005	-0,025	0,063	-0,182	0,118	-0,082	0,311
NC 1	0,174	0,154	0,140	-0,289	-0,780	-0,391	-0,443
IH	-0,164	-0,116	-0,224	0,533	0,600	0,496	0,003

7.2.3. Results

The cluster algorithm stopped after 12 iterations, producing three clusters of the size 76, 46 and 19, respectively. Table 7.3 shows the final cluster centers.

Table 7.3: k-means: Final cluster centers All variables have been z-standardized							
	ESS	PSG_AI	VigFalse	Delta	SL1	PSG_REML	PSG_SEI
I	-0,226	-0,346	-0,274	-0,179	0,278	0,197	0,492
II	0,346	0,204	0,546	-0,412	-0,765	-0,688	-0,577
III	0,068	0,890	-0,228	1,714	0,739	0,880	-0,569

Diagram A.1 in the appendix visualizes the obtained cluster along the axes of PC1 to PC3.

7.2.4. Evaluation

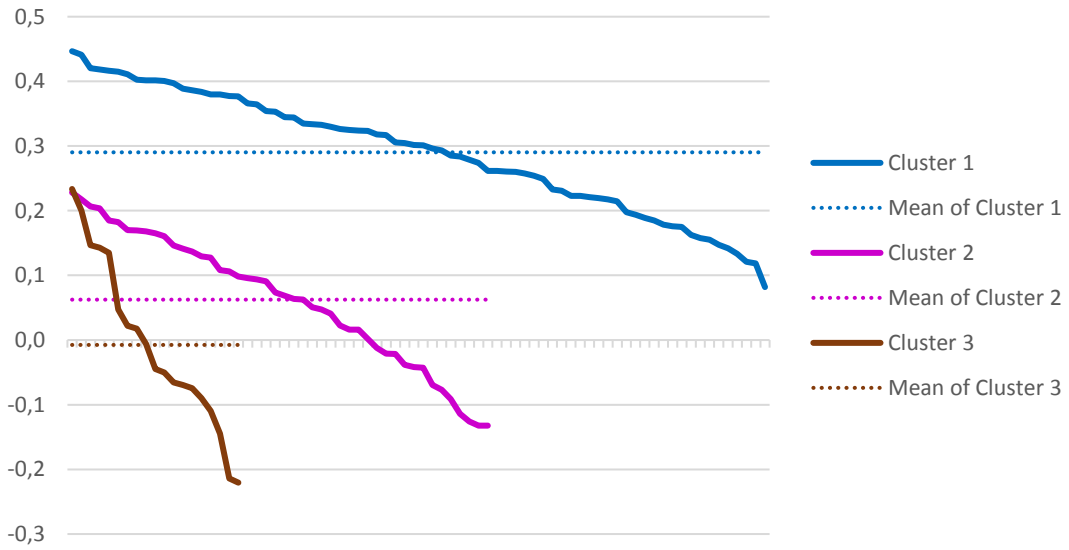
The ANOVA statistics for the cluster variables indicate significant differences between the clusters with respect to all cluster variables. As it has been mentioned before, this fact cannot be used to deduce good cluster quality.

Table 7.4: ANOVA statistics of the k-means cluster solutions: Cluster variables								
	Cluster 1 (N=76)		Cluster 2 (N=46)		Cluster 3 (N=19)		Statistics (df=140)	
	Mean	SD	Mean	SD	Mean	SD	F-value	p-value
ESS	14,8	3,62	16,9	3,90	15,9	3,09	5,01	0,008
PSG_AI	4,25	5,53	9,47	10,1	16,0	13,7	15,8	<0,001
VigFalse	2,05	2,42	5,87	6,62	2,26	3,16	11,8	<0,001
Delta	0,588	0,628	0,272	0,374	3,16	2,37	61,7	<0,001
SL1	4,75	1,91	2,46	1,62	5,75	1,99	30,8	<0,001
PSG_REML	79,9	40,9	33,2	41,3	116	65,6	27,0	<0,001
PSG_SEI	92,9	4,39	85,8	6,93	85,8	7,03	27,5	<0,001

Also, all non-cluster variables (as listed in table A.3 in the appendix) yield a significant F-value, except for VigRT and VigCorr. This finding resembles the comments regarding PC2, which summarized the vigilance test performance. The two-dimensional cluster plots indicate that PC2 is not helpful to distinguish between the k-means clusters, even though VigFalse was included as cluster variable. It can nevertheless be concluded that k-means was able to construct clusters which differ for all cluster variables and even for most of the non-cluster variables.

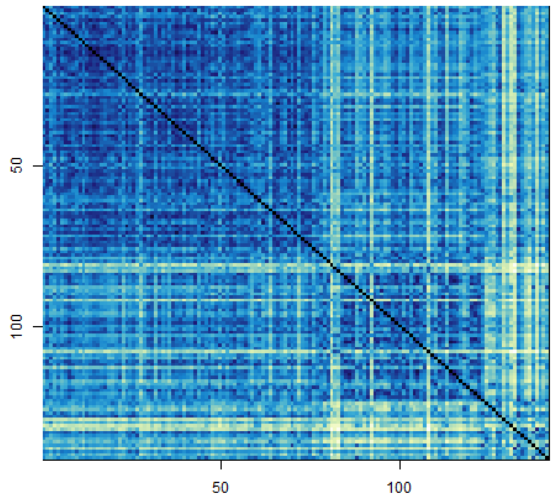
Next, the silhouette coefficient for each element was calculated. Diagram 7.1 shows all coefficients, where the elements have been sorted by cluster assignment and by the coefficient in descending order. Clearly cluster 1 shows the best overall quality, whereas half the elements of cluster three have negative silhouette coefficients, which indicates the residual character of cluster 3. The total average silhouette coefficient was 0,18. Again, the cluster plots might give a hint for this poor overall silhouette coefficient: All three clusters do not seem to be well separated in diagram A.1.

Diagram 7.1: Silhouette coefficients for the k-means cluster solution



Finally, diagram 7.2 shows the distance matrix of all elements of the dataset, ordered by cluster assignments. The Euclidean distance with respect to the standardized cluster variables was used.

Diagram 7.2: Distance matrix plot for the k-means cluster solution



Clusters 1 and 2 are easily identified as the darker squares along the diagonal, whereas the third cluster, which should appear at the lower right end of the diagram is very hard to detect.

In conclusion the k-means cluster algorithm yielded two clusters of acceptable quality and size plus one residual cluster. Section 7.5 will offer a more in-depth interpretation of all cluster solutions.

7.3. OPTICS: a density-based approach

7.3.1. Introduction

As the second cluster algorithm OPTICS will be employed, which is the abbreviation for "Ordering points to identify clustering structure". OPTICS is a density-based algorithm, i.e. clusters are identified as regions showing a high density of elements. OPTICS has been introduced by Ankerst et al. in 1999[193] and can be described as a refinement of the widely-used DBSCAN (Density-Based Spatial Clustering with the Application of Noise) algorithm.

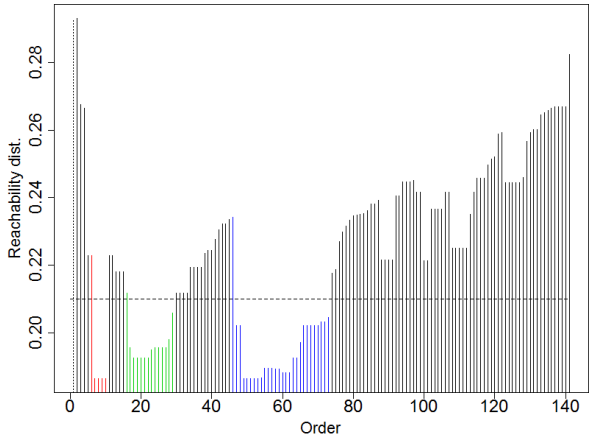
As the name indicates OPTICS does not produce a cluster solution directly, but a certain ordering of all elements in a reachability plot. Intuitively, reachability is the distance of an element to the "core region" of a cluster, i.e. a region showing a high element density. Therefore, low reachability

values suggest that the element at hand can easily be assigned to a cluster. The inspection of the reachability plot allows a quick insight into the general structure of the dataset.

Diagram 7.3 shows an example of a reachability plot. Valleys in the plot indicate potential cluster structures, whereas points above the threshold cannot be assigned to a distinct cluster and are summarized as "noise". It is important to acknowledge that only the reachability plot is a direct result of the OPTICS algorithm. Having obtained this plot, the user is obliged to determine the critical reachability level, depending on the dataset and the aims of the cluster analysis.

Diagram 7.3: An example reachability plot

The horizontal line indicates the critical reachability value 0,21. Other values will yield a different cluster structure. The example was generated by applying OPTICS to the dataset using the angular metric and the parameters eps=20, minPts=7



Clusters may subdivide/merge if the critical reachability is lowered/raised. Too low levels of the critical reachability will mark most of the elements as noise, whereas choosing too high a threshold will comprise the whole dataset into one cluster.

The existence of noise elements is an important characteristic of OPTICS. This means that several cases might lie outside of any cluster detected by OPTICS. From a clinical point of view this translates to the fact that OPTICS will detect "core phenotype groups" in the patient dataset, whereas atypical cases will probably be labeled as noise.

For the sake of comparability to other clustering methods, all noise points of the OPTICS solution will be summarized as a residual cluster, but this artificial cluster should not be compared directly to other clusters in terms of clinical characteristics and cluster quality.

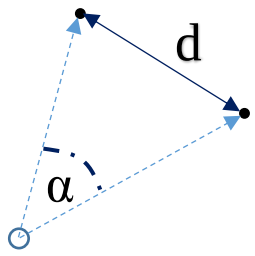
7.3.2. The angular metric and the choice of minPts

Applying the OPTICS method to the dataset, for which the distances between the elements were calculated using the standard Euclidean metric, yielded an ordering of the elements which showed no significant cluster structure, regardless of the chosen reachability threshold (see diagram A.4 in the appendix for the reachability plot).

The reachability plot indicates that the dataset is essentially treated as one big cluster as the algorithm was unable to detect multiple "core regions". One can try to solve this problem by defining alternative measures of distance, which are more suitable for the treatment of this dataset. In this case it was decided to use the angular metric, which is a higher dimensional generalization of the angular metric that is used in astronomy. Repeating OPTICS with respect to this alternative distance function, the reachability plot showed robust signs of inherent cluster structure.

Diagram 7.4: A change of the distance function

Two-dimensional example. Instead of the Euclidean metric **d,** the angular metric **α** is used



7.3.3. Cluster parameters, implementation and results

Two cluster parameters must be specified for OPTICS: minPts and eps. Both determine how "core regions" are detected as regions of increased density. The parameter eps serves as an upper bound in the calculation of "neighborhoods" of elements and does not significantly affect the result if it is not chosen too low[193].

The more critical task is finding a suitable value for minPts, which determines the number of elements that have to be detected in a sufficiently small neighborhood in order to be recognized as a "core region" of a cluster. Hence, smaller minPts values tend to produce a more fragmented reachability plot, which becomes smoother if minPts increases. According to Ankerst et al. usually values between 10 and 20 yield reasonable results[193]. In this situation the reachability plots for minPts values between 5 and 20 were inspected. Diagram 7.5 illustrates the varying shape of the reachability plots with increasing minPts.

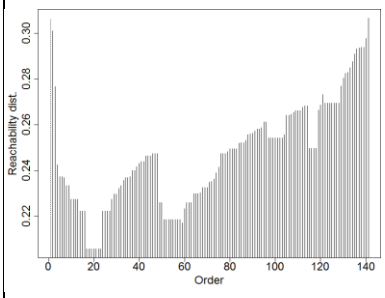
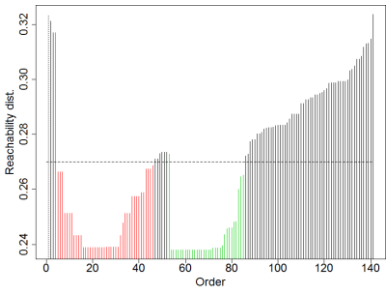
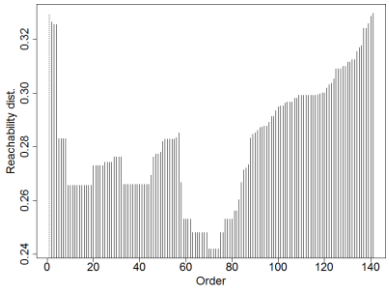
Choosing minPts as 14 leads to an optimal result, with the reachability plot showing two easily identifiable clusters.

The OPTICS clustering was performed using the dbscan package[194] for R (v.1.1-1)[195]. Since a different metric than the Euclidean distance was used, the distance matrix with respect to the angular metric had to be calculated first. The distance matrix was then used as input for the "optics" function of the dbscan package.

Diagram 7.5: OPTICS Reachability plots for different values of minPts

All plots were calculated using the angular metric.

The final choice minPts=14 is shown in the middle. (eps=20, eps_cl=0,27)

minPts=10	minPts=14	minPts=16
		

The cluster parameters were defined as eps = 20 and minPts = 14. After inspecting the reachability plot, a cluster solution using 0,27 as critical reachability value was chosen. Hence, the OPTICS algorithm yielded two clusters with the sizes 43 and 33, respectively and 65 noise points. The projections of the cluster solutions on the first three PCs are depicted in the appendix (see diagram A.2).

7.3.4. Evaluation

The t-test statistics for the cluster variable yielded significant t-values for every variable except VigFalse.

In case of the non-cluster variables (see table A.5 in the appendix), significant t-values were also found except for VigCorr and VigRT. This finding is analogous to the evaluation results for the k-means solutions, which also had no significant differences with respect to these variables.

Concerning the silhouette coefficients, one subtlety must be considered. Just as the clustering process itself, the silhouette coefficient relies on the metric structure of the dataset. Hence two different silhouette coefficients can be calculated for each point, one based on the Euclidean metric (again, referring to the standardized cluster variables) and one based on the angular distance.

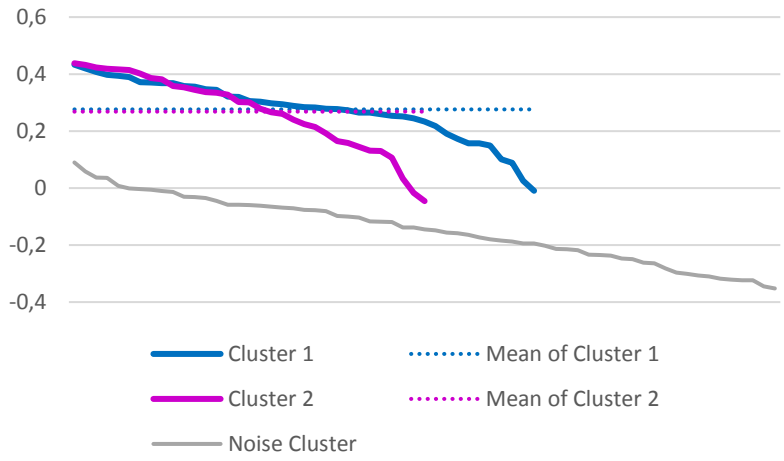
Table 7.5: The t-test for the OPTICS cluster solution: Cluster variables

Mean value and standard deviation for the noise cluster are only listed for the sake of descriptive comparison and not considered in the t-test.

	Noise cluster (N=65)		Cluster 1 (N=43)		Cluster 2 (N=33)		Statistics (df=74)	
	Mean	SD	Mean	SD	Mean	SD	t-value	p-value
ESS	15,5	3,45	13,8	3,83	18,3	2,63	-5,74	<0,001
PSG_AI	9,61	11,3	2,77	3,46	9,68	8,96	-4,63	<0,001
VigFalse	4,35	6,10	2,21	2,63	2,76	2,68	-0,893	0,375
Delta	1,22	1,81	0,640	0,757	0,327	0,403	2,15	0,035
SL1	4,27	1,88	5,59	1,88	1,97	1,24	9,57	<0,001
PSG_REML	92,5	57,3	72,4	32,2	20,4	26,2	7,55	<0,001
PSG_SEI	87,3	7,26	93,7	3,48	88,8	6,40	4,31	<0,001

It should also be noted that since noise elements are summarized into one cluster, low or even negative silhouette coefficients are to be expected for this “artificial” cluster. Diagram 7.6 shows the silhouette coefficients regarding the Euclidean metric, which allows a direct comparison to the other cluster solutions. Additionally, the silhouette coefficient plot using the angular metric can be found in the appendix (see diagram A.5).

Diagram 7.6: Silhouette coefficients for the OPTICS cluster solution using the Euclidean metric.

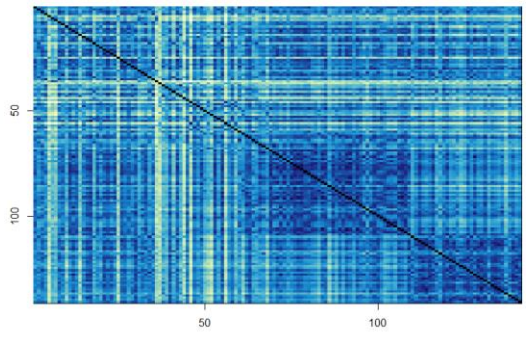
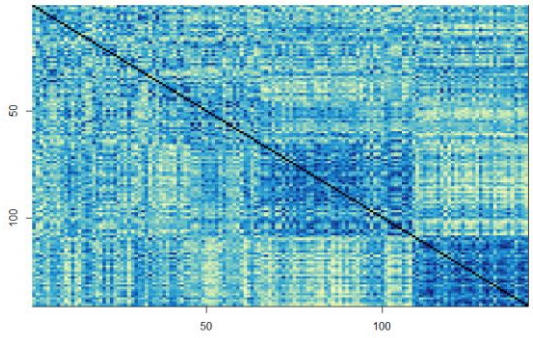


For both metrics, the "real clusters" both have comparable silhouette coefficients about 0,25. Obviously, due to the bad shaped noise cluster, the total silhouette coefficient average is much lower than in the k-means situation. Furthermore, despite the failure of reasonable clustering using the Euclidean metric, both variants of the silhouette coefficients do not significantly differ.

Finally, Diagram 7.7 shows the plotted distance matrices of the OPTICS cluster solution. Here, the matrices with respect to both metrics are reported.

Diagram 7.7: Distance matrix plots for the OPTICS cluster solution

The noise cluster is depicted in the upper left of each diagram

Euclidean distance	Angular distance
	

As to be expected, the noise cluster is almost invisible, whereas both cluster 1 and 2 are reasonably easy to detect for both metrics. Outside the cluster rectangulars, the matrix for the angular distance is brighter, which indicates high distances between points not belonging to the same cluster (in comparison to distances between members of the same cluster). This highlights the finding that only the transformation to the angular distance allowed a successful clustering process.

In conclusion OPTICS was able to detect two rather coherent and well separated clusters, but almost half of the dataset was labeled as noise. The evaluation of the two clusters yielded results that are qualitatively similar to the ones of the k-means cluster solution, but it remains to be discussed whether these good results "have been bought" by the introduction of the large set of noise points.

7.4. Spectral clustering

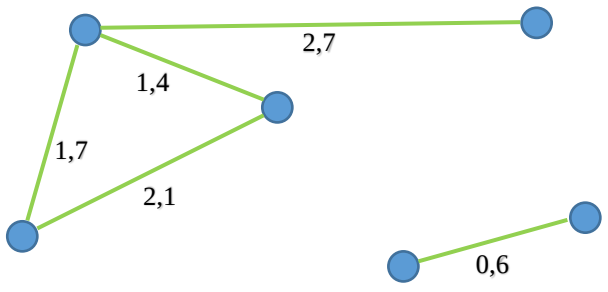
7.4.1. Spectral clustering: a graph theoretical approach

After the iterative centroids refinements of the k-means method and the density-based cluster detection by OPTICS now the third and final cluster method, spectral clustering, will be introduced.

The basic principle in spectral clustering is to reduce the dataset to a graph, i.e. a set of points (or *vertices*) and lines (or *edges*), which connect two different vertices. More precisely, a weighted graph will be used, which means that all edges are labeled with a value. The elements of the dataset will be interpreted as edges of a graph. Each element/vertex will only be connected to its closest neighbor and the corresponding edge weight will encode the distance between these two elements.

Diagram 7.8: A simple example of a weighted graph

The blue dots represent vertices, the green connecting lines are edges. Edges are weighed by the assigned numbers.



Clusters in the original dataset are translated into regions that are highly interconnected by many edges. Between different clusters one expects to find only a few if any edges. Therefore, one should apply mathematical techniques that are able to detect regions that are closely interconnected as clusters. For accomplishing this task, one can rely on the results of a branch of mathematics called graph theory, which served as foundation for the development of spectral clustering.

Formally, spectral clustering aims to reduce the clustering process to the subspace that is generated by the last few eigenvectors of the Laplacian matrix of the adjacency matrix. Ideally, the transformed space is lower dimensional than the initial dataset and shows a structure that allows an easier clustering via conventional methods like k-means.

As a rule of thumb, spectral clustering detects clusters in which two elements can always be connected by a chain of short edges. Therefore, it is possible that clusters of "atypical shape" emerge, being quite different from the convex k-means clusters. Considering this is important, as atypically shaped clusters cannot easily be characterized by mean values and standard deviations of the cluster variables.

From a clinical perspective, two patient cases will be assigned to the same cluster if they can be connected by a chain consisting of other cases, where each case is very similar to the preceding and following case.

7.4.2. Implementation and cluster parameters

In general, spectral clustering consists of several different steps:

- Starting with a given similarity matrix, which contains all similarity relations of elements in the dataset, the dataset is reduced to a suitable graph, which can be represented by its adjacency matrix
- The corresponding *Laplacian matrix* is calculated, which contains the essential pieces of information regarding the weighed edges between the cases
- In order to extract the clusters this information is reduced to the space constructed by the last eigenvectors of the Laplacian matrix
- Using conventional cluster methods like k-means, clusters can easily be detected in this lower dimensional space

Several choices have to be made to adapt the spectral clustering algorithm to the given dataset.

The first step is the reduction of the dataset to a suitable graph. Following the recommendation of Luxburg[196] the *k-nearest-neighbor* method was used. This means that from each case/vertex, k edges are drawn to the k nearest vertices around. This approach ensures that each element is connected to at least k other elements. Implicitly, once again one has to agree on a measure of similarity which fits the dataset. In spectral clustering, a common choice is the Gauß similarity[196]:

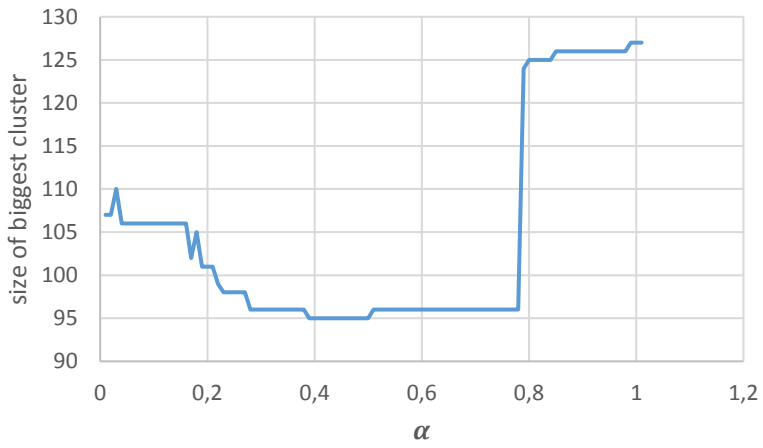
$$s(x_1, x_2) = e^{-\alpha \|x_1 - x_2\|^2}, \alpha > 0.$$

Therefore, the two parameters α and k, have to be chosen in a sensible way. According to Luxburg, the obtained graph should be connected (i.e. any two vertices are connected via a chain of edges) without allowing too high a density of edges. After some initial testing, k was set to two, so each element was connected to its two closest neighbors.

Apart from the choice of k, also finding the right value for α is crucial for a successful clustering. No heuristics exist for this situation. Hence, spectral clustering was tested for different values for α and the size of the biggest cluster was plotted for each solution.

Diagram 7.9: Size of the largest cluster of the spectral cluster solution for different values of α in the Gauß similarity

The clustering of the eigenvector space was done using the k-means algorithm (nstart=50). As the initial cluster centers were chosen randomly, this graph will slightly change if the algorithm is performed again. However, multiple repetitions all yielded very similar graphs.



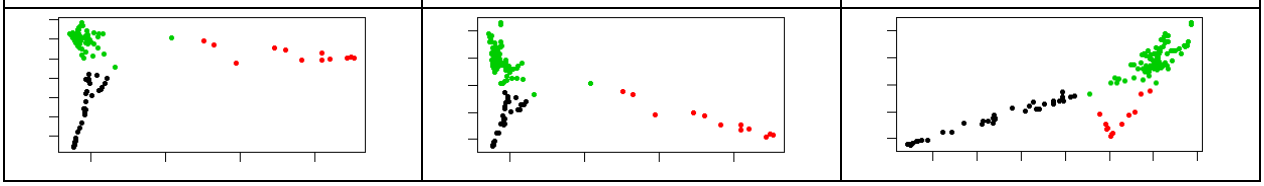
A broad minimum can be observed in the diagram above, so some cluster structure could be detected in the data without being too vulnerable to the variation of α. It was decided to set $\alpha = 0{,}39$, according to the minimum in the plot. This choice ensures that the dominant cluster of the solution is as small as possible.

The number of relevant eigenvectors is usually chosen to be equal to the number of expected clusters. Therefore, the last three eigenvectors were used.

In spectral clustering the clustering of the eigenvector space is usually done using k-means. K-means was also applied in this situation, but since no reasonable cluster centers were available, the *nstart* parameter of the *kmeans* function in R was used and set to 50 (library *stats* v.3.3.3.)[195]. This means that the k-means procedure was repeated 50 times and the best cluster solution was selected. The following diagrams show the clustering results in the eigenvector space. As these diagrams suggest, indeed three clusters can easily be identified in the eigenvector space.

Diagram 7.10: The spectral clustering solution in the "eigenvector space"

EV: Eigenvector; the labels indicate which eigenvector serves as X / Y -axis.



Following the descriptions of Luxburg[196], the R code was manually written to perform the spectral clustering and can be found in table A.9 in the appendix.

Three clusters were obtained which contained 34, 12 and 95 elements, respectively. Diagram A.3 in the appendix shows the clusters projected on the first three principal components.

7.4.3. Evaluation

Due to the special approach of spectral clustering no convex cluster shapes were to be expected. The three clusters have emerged by summarizing elements that are closely connected with respect to the graph simplification of the dataset, in which each element was only connected to its two

closest neighbors. As diagram A.3 demonstrates, the obtained clusters are far more difficult to visually separate than in the k-means or OPTICS solutions.

Table 7.6: ANOVA statistics for the spectral clustering solution

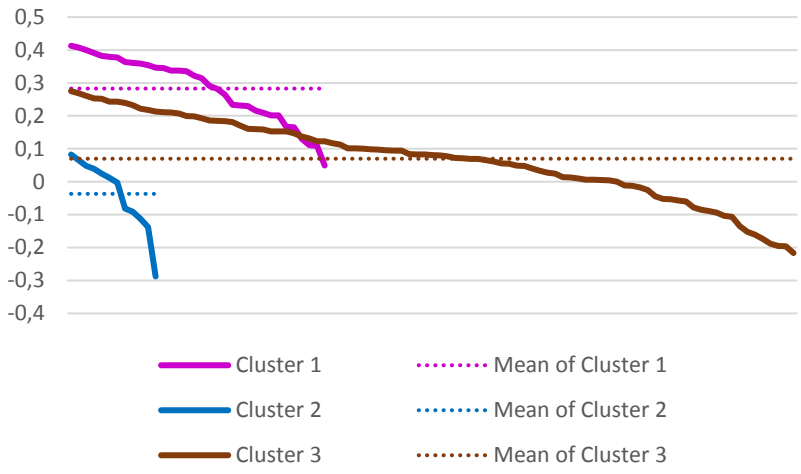
	Cluster 1 (N=34)		Cluster 2 (N=12)		Cluster 3 (N=95)		Statistics (df=140)	
	Mean	SD	Mean	SD	Mean	SD	F-value	p-value
ESS	16,9	3,53	13,5	3,78	15,5	3,71	3,99	0,021
PSG_AI	9,33	8,21	20,3	17,6	5,28	6,88	17,4	<0,001
VigFalse	1,94	1,87	11,3	9,68	2,82	3,33	26,6	<0,001
Delta	0,274	0,403	0,425	0,367	1,08	1,58	5,33	0,006
SL1	1,73	1,01	3,58	2,24	5,07	1,78	50,1	<0,001
PSG_REML	27,4	31,6	40,0	47,9	88,3	49,2	25,1	<0,001
PSG_SEI	89,7	6,59	84,6	5,20	90,2	6,70	3,92	0,022

The ANOVA results regarding the cluster variables nevertheless showed significant differences for all cluster variables. Remarkably, the ANOVA statistics for the non-cluster variables (see tables A.7 and A.8) yielded a result which is closely related to the corresponding statistics for the other cluster solutions: Again, one notices significance for all non-cluster variables except VigCorr and VigRT.

Next, the silhouette coefficients for every element in the dataset were calculated. The Euclidean metric was used as an underlying distance measure to allow a comparison between all three cluster solutions. Furthermore, in the appendix the silhouette coefficients of the elements in the eigenvector space are illustrated (in diagram A.6).

Diagram A.6 in the appendix emphasizes how the spectral clustering approach reduced the high-dimensional dataset to a three-dimensional space, in which three high quality clusters emerged. However, if one considers the silhouette coefficients regarding the Euclidean metric in the original space of the cluster variables, one notices that the spectral cluster solution seems to be of inferior quality compared to the k-means or OPTICS solution. Only cluster 1 shows silhouette coefficients similar to the best clusters in k-means or OPTICS, but the majority of elements is grouped in clusters 2 and 3, which have overall silhouette coefficients close to zero.

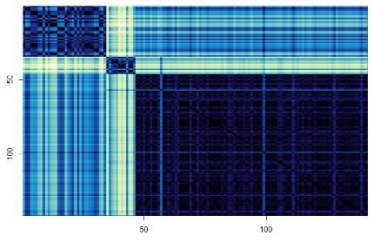
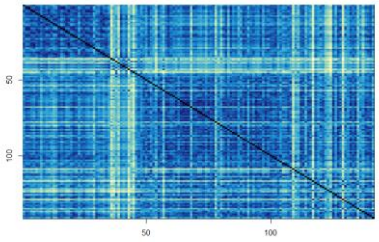
Diagram 7.11: Silhouette coefficients of the spectral clustering solutions with respect to the Euclidean metric



This finding can be explained by the graph theoretical concept of spectral clustering, which allows the formation of non-convex, possibly "intertwined" clusters that are usually characterized by suboptimal separation and cohesion values.

The corresponding matrix plots yield similar results. Diagram 7.12 shows a direct comparison of the distance matrix plots in the eigenvector and cluster variable space.

Diagram 7.12: Distance matrix plots for the spectral clustering solution

Eigenvector space	**Euclidean metric**
	

For the comparison of the different cluster solutions, the right plot should be used. The left plot again highlights the effectiveness of the data transformation in spectral clustering, which leads to the formation of three clusters of seemingly very high quality. On the other hand, regarding the original variable space, spectral clustering has yielded a cluster solution that seems to be inferior to the k-means and OPTICS solution, as only cluster 1 emerges as a clearly visible rectangular.

7.5. Discussion and cluster comparison

7.5.1. The ICSD-3 diagnoses as null hypothesis cluster solution

Before discussing the three different cluster solutions in further detail, the "null hypothesis" cluster solution that has been predefined by the ICSD-3 will be presented briefly. All validation techniques described above will be applied here, allowing a better estimation of the quality of the alternative clusters. The scatterplots of the ICSD-3 clusters projected on the first three PCs are already shown in the discussion of the PCs.

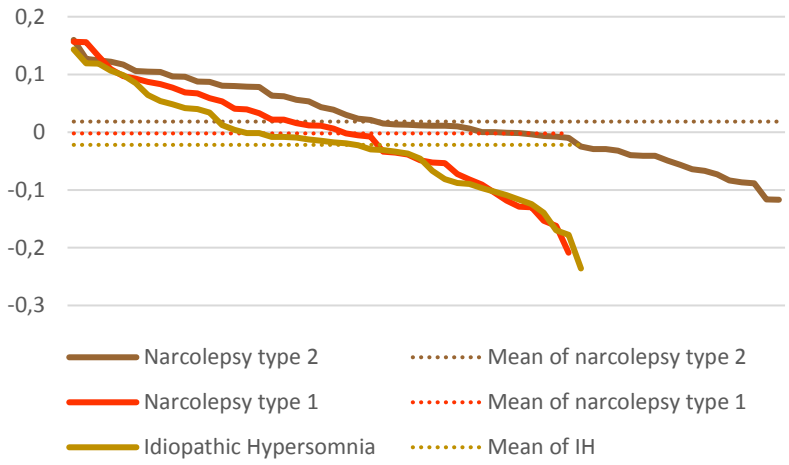
Regarding the cluster variables there are significant differences between the diagnoses for the variables Delta, SL1, PSG_REML, PSG_SEI. In particular, one notices the significantly higher Delta values for IH and the typical hierarchy for the SL1 values with narcolepsy type 2 lying between narcolepsy type 1 and idiopathic hypersomnia, that has also been observed by other authors[163, 172].

Table 7.7: ANOVA statistics for the ICSD-3 Diagnoses: cluster variables

	narcolepsy type 2		narcolepsy type 1		idiopathic hypersomnia		statistics (df=140)	
	Mean	SD	Mean	SD	Mean	SD	F-Value	p-value
ESS	15,6	4,16	16,3	3,10	15,0	3,72	1,19	0,307
PSG_AI	7,30	8,31	9,00	12,5	6,43	7,47	0,785	0,458
VigFalse	3,62	4,94	3,98	5,57	2,29	2,85	1,58	0,210
Delta	0,585	0,723	0,439	0,974	1,56	1,97	9,72	<0,001
SL1	4,39	2,21	2,43	1,42	5,45	1,67	28,5	<0,001
PSG_REML	65,2	40,4	48,9	56,6	95,6	54,4	9,53	<0,001
PSG_SEI	91,7	5,57	86,7	7,31	89,6	6,56	7,47	0,001

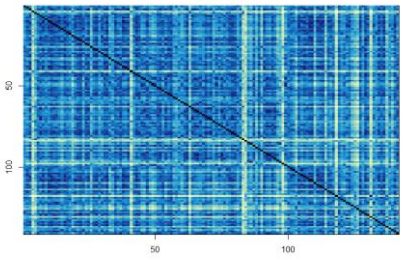
The ANOVA statistics of the non-cluster variables (see table A.10 in the appendix) show significant differences for SL2, TST, susSL and #SOREM. For the SOREM count, this is obviously due to the diagnostic criteria for IH and narcolepsy. Just like for all cluster solutions, no significant F-value was reported for VigCorr and VigRT.

Diagram 7.13: Silhouette coefficients of the ICSD-3 clusters (diagnoses)



The silhouette coefficients as well as the distance matrix plot suggest, that if one follows the ICSD-3 diagnoses, no clusters of high quality are obtained. However, in contrast to the cluster solutions, the ICSD-3 diagnoses have not been designed to yield significant differences regarding the seven cluster variables (that are considered for the calculation of the silhouette coefficient and the distance matrix). Hence, although the diagnostic clusters appear to be inferior to the three obtained cluster solutions, direct comparison is of limited use due to this fact. But if one agrees on the importance of all seven cluster variables, the ICSD-3 diagnoses offer only a suboptimal classification of the dataset.

Diagram 7.14: Distance matrix plot for the ICSD-3 diagnoses
Here, the Euclidean distance with respect to the cluster variables was used.



7.5.2. Rand indices: A first comparison of the cluster solutions

For a first impression of the similarity of the different cluster solutions, the Rand coefficients were calculated. Given two cluster solutions A and B, the Rand coefficient checks if any two elements that share a common cluster/lie in different clusters in cluster solution A also share a common cluster/lie in different clusters in cluster solution B. The more similar two different partitions are, the closer the Rand coefficient will be to 1[189].

Table 7.8: Cluster comparison using Rand coefficients
SPECC: Spectral clustering

	Diagnoses	kmeans	OPTICS	SPECC
Diagnoses	1	0,57	0,56	0,55
kmeans	0,57	1	0,62	0,62
OPTICS	0,56	0,62	1	0,6
SPECC	0,55	0,62	0,6	1

As the Rand indices in table 7.8 suggest, the cluster analyses did not reproduce the diagnostic partition of the dataset, but the three cluster solutions are almost as different from each other as they are different from the diagnostic clusters. Hence, the "null hypothesis" clustering by the ICSD-3 diagnoses cannot be accepted from the perspective of cluster analysis. However, the considerable differences between the three cluster solutions do not support the suggestion of a universally superior partition of the dataset.

7.5.3. Cluster validation results

In matters of the cluster variables, the ANOVA/t-test statistics yielded significant F-values for all variables in all cluster solutions, with the only exception being VigFalse in the OPTICS clustering. As it has been discussed earlier, only the lack of significance would be of real use for the cluster validation, since it would indicate the failure of the corresponding clustering method.

Regarding the non-cluster variables, VigRT and VigCorr never showed significant between-group differences, but these could not be observed between the ICSD-3 diagnoses either (see table A.10 in the appendix). The remaining non-cluster variables significantly differed between the clusters of every cluster solution.

Table 7.9: Cluster size and silhouette coefficients for all cluster solutions

All silhouette coefficients have been calculated using the Euclidean distance measure with respect to the z-standardized cluster variables. "Cluster 3" of the OPTICS solution is the collection of all noise points

	k-means		**OPTICS**		**Spectral clustering**	
	N	Silhouette	N	Silhouette	N	Silhouette
Cluster 1	76	0,290	43	0,22	34	0,28
Cluster 2	46	0,062	33	0,32	12	-0,037
Cluster 3	19	-0,0076	65 (Noise)	-0,043	95	0,070
Total average	141	0,178	141	0,12	141	0,11

In total, the highest average silhouette coefficient can be found in the k-means cluster solution. This could hint at a slight superiority of the k-means cluster solution compared to OPTICS and spectral clustering. Another explanation for this finding is the inherent tendency of k-means to report convex-shaped clusters which are more optimal regarding cohesion and separation than the possibly more atypical cluster shapes in OPTICS and spectral clustering. OPTICS was able to produce two clusters of relatively high quality, but the overall silhouette average is deteriorated by the artificial "noise cluster". Spectral clustering detected the largest single cluster of all cluster solutions, which, however, is characterized by a silhouette coefficient close to zero. Furthermore, only in spectral clustering less than half of all cases were assigned to a cluster with an acceptable overall silhouette coefficient above 0,2. As it has been mentioned above, this is no indication of the principal inferiority of spectral clustering but a consequence of the distinct mathematical approach of this method.

7.6. Characterization and Comparison

Now every cluster solution will be inspected more closely and discussed regarding the core characteristics of each of its clusters. Apart from size and cluster variable values, also the frequency of the different ICSD-3 diagnoses will be taken into account. By doing this, one gets not only a quick impression of the cluster characteristics but may also gain insights into the relationship between IH and narcolepsy type 2. For each considered variable, the average values for every cluster are listed below.

It should be noted that comparing clusters with respect to mean values of their respective elements, implicitly a convex, ball-like shape of the clusters is assumed. Especially for spectral clustering,

such an assumption would be highly unjustified, hence caution is needed for the interpretation of apparent differences.

Diagram 7.15: Comparison of the k-means solution clusters

	N	ESS	PSG_AI	VigFalse	Delta	SL1	PSG_REML	PSG_SEI	SL2	TST	susSL	#SOREM	VigCorr	VigRT	PSG_N3_TST	CATAP
1	76	14,8	4,25	2,05	0,59	4,75	79,9	92,9	8,79	119	5,33	2,03	86,3	0,547	16,9	13/76
2	46	16,9	9,47	5,87	0,27	2,46	33,2	85,8	5,26	132	2,73	3,54	83,3	0,586	11,2	25/46
3	19	15,9	16,0	2,26	3,16	5,75	115,9	85,8	12,2	95	8,91	0,789	88,6	0,572	10,7	3/19

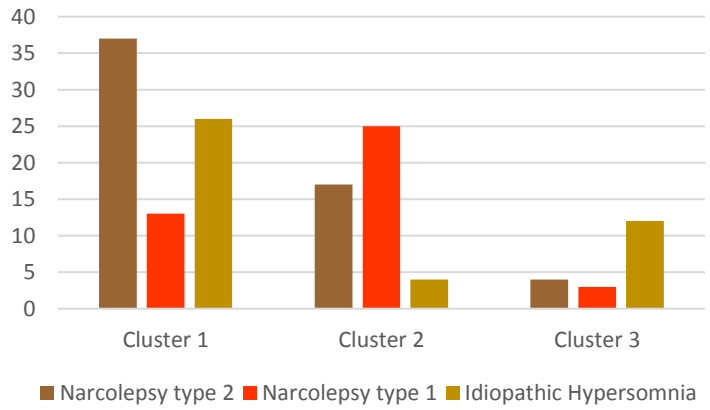


Diagram 7.15. depicts the frequency of the ICSD-3 diagnoses and mean values of all considered variables in the clusters of the k-means solution. Both the silhouette coefficients and the distance matrix plots hint at the residual character of cluster 3. Hence, k-means effectively yielded two clusters and cluster 3 apparently only emerged because the cluster number was prespecified as 3. Therefore, the focus will lie on the descriptions of the clusters 1 and 2. However, one notices that cluster 3 seems to summarize patients showing the highest values of Delta.

- **Cluster 1: Efficient, deep sleepers; IH/narcolepsy type 2:** From the perspective of the ICSD-3 diagnoses, this cluster shows a structure that will also reappear in the other cluster solutions. It summarizes more than half of the total number of both IH and narcolepsy type 2 cases. Delta values are higher than in cluster 2 but much lower than in the residual cluster. Also regarding the SOREM count, the sleep latency and the REM latency, cluster 1 lies in between the clusters 2 and 3. Cluster 1 has the highest average sleep efficiency and the fraction of N3 sleep is significantly higher than in the other clusters, whereas the arousal indices in cluster 1 are – on average – lower than in cluster 2 or 3. Compared to cluster 2 patients of this cluster have slightly better results in the vigilance test.

- **Cluster 2: Typical narcolepsy phenotypes:** Most of the narcolepsy type 1 and almost no IH cases can be found in this cluster. It is characterized by extremely low sleep latencies, more than three SOREMs on average in the MSLT and the shortest REM latencies. Both the high SOREM count and the short REM latencies indicate a dysregulation of REM sleep. These patients also report the highest subjective sleepiness in the ESS and have the poorest performances in the vigilance test.

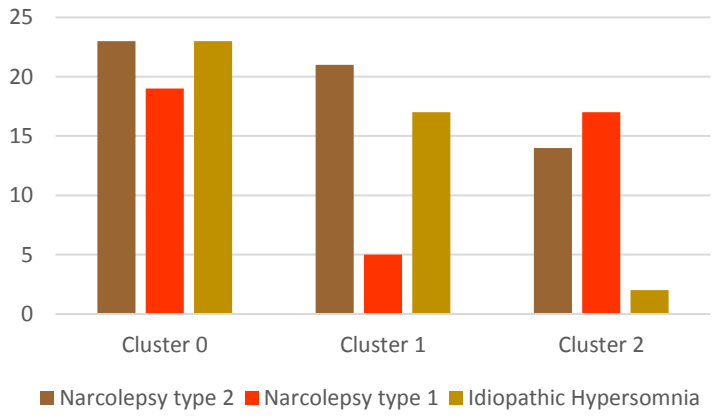
In conclusion, two third of narcolepsy type 2 and one third of narcolepsy type 1 patients have been assigned to cluster 1, which is characterized by higher sleep efficiency, two SOREMS on average in the MSLT, and higher Delta values. Also, the majority of IH cases has been assigned to this cluster. The remaining cases, if one neglects the residual cluster 3, are comprised in cluster 2, which describes a pronounced narcolepsy phenotype.

Hence, k-means was not capable to distinguish the narcolepsy subtypes from each other, nor the narcolepsy subtypes from IH. Narcolepsy patients seem to be distributed along a continuum ranging from the typical presentation of severe narcolepsy (short sleep latencies, many SOREM episodes, low Delta, etc.) to individuals showing higher sleep efficiencies, higher sleep latencies, less SOREMs and a significantly higher REM latency. Narcolepsy type 1 patients tend to appear closer to the former prototype, whereas most of the narcolepsy type 2 patients can be described by the latter characterization and are therefore very difficult to distinguish from IH patients.

It should also be noted, that for the initial cluster centroids, the “centers” of narcolepsy type 2, narcolepsy type 1 and IH were used, respectively. During the k-means procedure, the initial narcolepsy subtypes clusters exchanged about a third of their cases, whereas almost all IH cases were integrated in the initial narcolepsy type 2 cluster. Thus, only a small residual cluster 3 remained.

Diagram 7.16: Comparison of the OPTICS solution clusters

	N	ESS	PSG_AI	VigFalse	Delta	SL1	PSG_REML	PSG_SEI	SL2	TST	susSL	#SOREM	VigCorr	VigRT	PSG_N3_TST	CATAP
0	65	15,5	9,61	4,35	1,22	4,27	92,5	87,3	8,91	116	5,49	2,02	84,1	0,576	13,1	19/65
1	43	13,8	2,77	2,21	0,640	5,59	72,4	93,7	9,52	114	6,23	1,81	86,1	0,552	18,5	5/43
2	33	18,3	9,68	2,76	0,327	1,97	20,4	88,8	4,63	134	2,30	3,73	86,6	0,552	10,9	17/33



The main issue of the OPTICS cluster solution is the noise cluster 0, which comprises almost half of the dataset. In a strict sense, these elements have not been recognized as members of any cluster, but are formally treated as “cluster 0”. As the cluster solution plot in the appendix indicates, many outlying elements contribute to this widespread collection of cases. Therefore, it is not possible to discuss cluster 0 in the standard way. One notices on first sight that all three ICSD-3 diagnoses appear in a similar frequency in cluster 0, which also shows the worst vigilance test results, highest Delta values and highest REM latencies on average.

If one neglects the “noise cluster”, again two clusters are obtained, so both k-means and OPTICS suggest the rejection of the null hypothesis assumption of three clusters in the dataset.

- **Cluster 1: The IH/narcolepsy type 2 cluster: high efficient sleepers without significant REM disorder:** This cluster summarizes the cases showing the highest sleep efficiency and N3 fraction. Due to the contributions by narcolepsy type 2 and IH cases, an average SOREM count of 1,8 is obtained. The average sleep latency lies well under the 8-minute threshold but is significantly higher than in cluster 2. Compared to the noise cluster and cluster 2, individuals assigned to cluster 1 show the lowest subjective sleepiness as it has

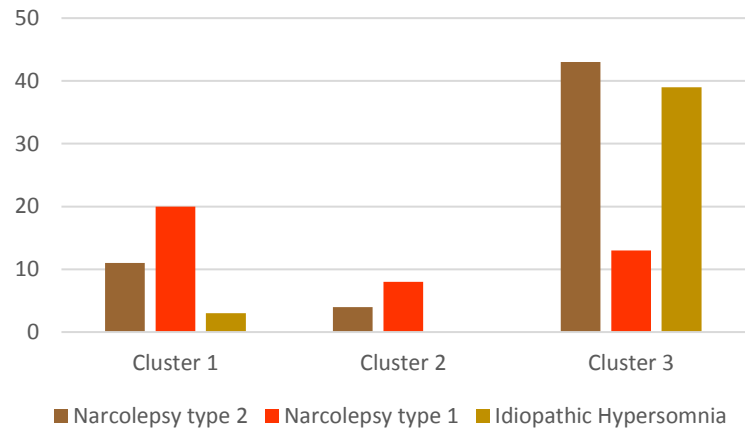
been measured by the ESS (13,8 points on average). Additionally, cluster 1 shows a lower average value for the REM latency and the arousal index.

- **Cluster 2: The narcolepsy phenotype:** Just like cluster 2 in the k-means solution, the most severe cases in terms of sleep latency, SOREM count and REM latency can be found in this cluster. Neglecting cluster 0, this cluster is formed by about a third of all narcolepsy type 2 cases combined with almost all narcolepsy type 1 cases. The average SOREM count is almost four, whereas the average sleep latency is about 2 minutes. Furthermore, one notices that on average, cluster 2 has Delta values about half of those in cluster 1. The individuals assigned to this cluster rated their EDS as 18 points on the ESS on average and show significantly more arousals than cases in cluster 1 did.

The comparison to the k-means cluster solution reveals the close similarity of the respective clusters. In both cases, cluster 1 includes most narcolepsy type 2 and IH cases, which are characterized by high sleep efficiency, relatively high values of Delta and REM latencies significantly above the average of cluster 2. Both in k-means and in OPTICS, the latter consists almost exclusively of narcolepsy cases, with narcolepsy type 1 being the dominating subtype.

Diagram 7.17: Comparison of the clusters of the spectral clustering solution

	N	ESS	PSG_AI	VigFalse	Delta	SL1	PSG_REML	PSG_SEI	SL2	TST	susSL	#SOREM	VigCorr	VigRT	PSG_N3_TST	CATAP
1	34	16,9	9,33	1,94	0,274	1,73	27,4	89,7	4,74	135	2,01	3,79	87,3	0,557	11,6	20/34
2	12	13,5	20,3	11,3	0,425	3,58	40,0	84,6	7,44	124	4,01	3,42	84,1	0,583	11,2	8/12
3	95	15,5	5,28	2,82	1,08	5,07	88,3	90,2	9,37	113	6,15	1,71	84,7	0,563	15,5	13/95



Regarding the distance matrix plots and silhouette coefficients, the spectral clustering procedure has found slightly inferior clusters to those of k-means and OPTICS. Cluster 3 comprises about two third of the dataset, and Cluster 2 is very small and shows poor silhouette coefficients. Therefore, the latter will be treated as a residual cluster and the further descriptions will concentrate on clusters 1 and 3. As a main feature, the residual cluster comprises individuals with a poor performance in the vigilance test. On average, more than 10 false reactions were measured in these cases.

- **Cluster 1: Typical narcolepsy phenotype:** Like cluster 2 in k-means and OPTICS, this cluster mainly consists of narcolepsy type 1 and a fraction of narcolepsy type 2 cases. With respect to cluster and non-cluster variables, one finds values that are comparable to those of cluster 2 in k-means and OPTICS.
- **Cluster 3: Deep, efficient IH/narcolepsy type 2 sleepers:** Again, a cluster has emerged that summarizes the cases showing the highest sleep efficiency and N3 fraction. The average Delta value is about three times the value of cluster 1. Cluster 3 may be larger than the corresponding clusters 1 in k-means and OPTICS but is still very closely related to them in terms of SOREM count, REM latencies and MSLT sleep latencies.

Again, one notices the characteristic compositions of the two relevant clusters that bear close resemblance to the corresponding clusters obtained by k-means and OPTICS. The calculation of the three cluster solutions was based on very different mathematical techniques, but nevertheless a common pattern is shared by all solutions. Each clustering procedure yielded one cluster that is dominated by narcolepsy type 1 patients and characterized by diagnostic findings that are typical for severe cases of narcolepsy. Almost no IH patients can be found in these clusters. Furthermore, every cluster analysis found a second cluster, in which IH and narcolepsy type 2 appear in similar frequencies, accompanied by only a few narcolepsy type 1 cases.

The consistent emergence of two relevant clusters supports the rejection of the null hypothesis, which stated the existence of three important clusters in analogy to the ICSD-3 diagnoses.

7.7. The narcolepsy subtypes: A critical remark based on the cluster results

From the perspective of the null hypothesis, three clusters were expected to emerge, reflecting the three ICSD-3 diagnoses. As it has become apparent in section 7.6., all three diagnoses are split up

in the process of cluster analysis. Regardless of the chosen method, cases sharing the same diagnoses end up in different clusters. Table 7.10. shows, in what ratio each diagnosis was distributed among the two major clusters of the different cluster solutions.

Table 7.10: Separation of the diagnostic groups in the cluster solutions
The residual/noise cluster is not considered; all values are in percentage

	k-means	**OPTICS**	**Spectral clustering**	**Average**
Narcolepsy type 1	65,8 – 34,2	77,3 – 22,7	60,6 – 39,4	67,9 – 32,1
Narcolepsy type 2	68,5 – 31,5	60,0 – 40,0	79,6 – 20,4	69,4 – 30,6
IH	86,7 – 13,3	89,5 – 10,5	92,9 – 7,1	89,7 – 10,3

Neglecting the cases that are comprised in the residual/noise clusters, every cluster analysis assigned more than 85 % of the IH cases to the same cluster. In comparison, this is only true for about two third of the narcolepsy type 1 or type 2 cases.

Hence, despite the lack of a concise pathophysiological concept, IH was separated the least by the cluster algorithms. Narcolepsy type 1 is far less consistently kept in a single cluster, but this is at least partially a consequence of the omission of cataplexy in the cluster variable selection process. Additionally, the diagnosis of narcolepsy type 1 could be consolidated by the measurement of CSF hypocretin.

These considerations highlight narcolepsy type 2 as the most questionable diagnostic group, which also shows intermediate values regarding most variables (see table 7.7). Of course, the simplest explanation for the cluster results is assuming that narcolepsy type 2 indeed exists as a reasonable diagnostic group, which – due to its intermediate position between narcolepsy type 1 and IH – tends to be assigned to the former or latter by cluster algorithms.

However, two important factors could also contribute to the perceived heterogeneity of narcolepsy type 2. One aspect might be the poor test-retest reliability of the MSLT[40]. There is a study by Trotti et al., that focuses on the specific situation of distinguishing between narcolepsy type 2 (without cataplexy) and IH. In a dataset of 36 patients with narcolepsy type 2 and IH, performing a second MSLT led to a change of diagnosis due to a changed SOREM count in 31 % of all cases[39]. Hence, at least some cases of narcolepsy type 2 should probably diagnosed as IH, but showed two or more SOREMs in the MSLT "by chance". On the other hand, the differentiation of narcolepsy type 2 from type 1 still depends mostly on the occurrence of cataplexy, since the invasive measurement

of CSF hypocretin are usually avoided. Sturzenegger et al. report that cataplexy may have its onset some years after the onset of hypersomnolence[87], whereas Rye et al. have identified the late onset of cataplexy as an important factor that might cause a delayed diagnosis of narcolepsy with cataplexy[197]. Therefore, narcolepsy type 1 patients will in many cases be labeled as narcolepsy type 2 until their first episode of cataplexy.

In conclusion, due to its intermediate position between narcolepsy type 1 and IH, narcolepsy type 2 poses many problems in diagnosis and differential diagnosis. A poor test-retest-reliability of the MSLT as well as a delayed cataplexy onset in narcolepsy type 1 can at least partially explain why narcolepsy type 2 was grouped together with IH or narcolepsy type 1 by the cluster algorithms instead of forming its own cluster. In the end, it cannot be decided whether narcolepsy type 2 is indeed a reasonable diagnostic concept or if it is merely a consequence of the shortcomings of the current diagnostic tools and concepts as it has been suggested by Mayer et al.[45].

7.8. Conclusion

In this chapter, cluster analyses were performed to identify groups in the given dataset that agree with respect to the selected cluster variables. Three different approaches for detecting clusters were considered. The k-means algorithm defined clusters as a set of elements that are closer to their cluster center than to the centers of the other clusters. OPTICS detected core regions of cluster as regions showing an increased density of elements. Finally, spectral clustering assigned elements, which could be connected by a "path" consisting of sequentially similar elements, to the same cluster.

The different concepts behind the cluster analyses eventually led to clusters, which significantly differ from the "diagnostic clusters" suggested by the ICSD-3, but which are also not closely related to each other. Nevertheless, all three cluster solutions turned out to be superior to the ICSD-3 diagnostic groups (regarding the detection of differences in the cluster variables), although no cluster solution emerged as a clear, optimal solution.

Despite their dissimilarity all three cluster solutions agree in several important properties.

- K-means, OPTICS and spectral clustering all yielded two significant clusters, thereby discouraging the acceptance of the previously stated null hypothesis (i.e., the existence of

three important clusters in the dataset). However, it should be considered that whereas k-means and spectral clustering found only a small residual third cluster, almost half of all cases were assigned to the noise cluster by OPTICS.

- The three corresponding pairs of relevant clusters, regardless of their differences in absolute size, were remarkably similar with respect to the following aspects:
    - One cluster of each pair always summarizes at least 85 % of the IH cases (cases in the residual/noise cluster are neglected) together with more than half of all narcolepsy type 2 cases. These cases are always characterized by a higher sleep efficiency, higher values of Delta, higher latencies to REM sleep and lower ESS scores than those that occur in the other non-residual cluster.
    - The other cluster of each pair consists of the majority of narcolepsy type 1 cases and a considerable fraction of narcolepsy type 2 cases. Only very few IH cases are found in these clusters. A significant REM sleep dysfunction of their members seems to be present as on average more than three SOREMs occur in these patients. Additionally, members of these clusters rate their subjective sleepiness highest in the ESS, show the lowest sleep latencies and have more fragmented night sleep than cases from the other cluster.

In conclusion, the consistent finding of the two clusters challenge the diagnostic groups defined by the ICSD-3. Instead of the three diagnoses narcolepsy type 1, narcolepsy type 2 and IH, the concept of only two subgroups might be more appropriate. In particular, the results of this chapter suggest that whereas IH and narcolepsy type 1 can be easily distinguished from each other, narcolepsy type 2 cases could not be efficiently separated from cases of the other diagnostic groups.

Section 7.6 further addresses this perceived heterogeneity of narcolepsy type 2. One explanation for this finding would be that at least some narcolepsy type 2 cases are actually patients suffering from IH, but whose MSLT showed two or more SOREMS “by chance”. Another subgroup of narcolepsy type 2 patients might suffer from (probably hypocretin deficient) narcolepsy type 1 but are misdiagnosed as narcolepsy type 2 due to a delayed onset of cataplexy. It cannot be decided with certainty whether narcolepsy type 2 is indeed an “intermediate” condition between narcolepsy

type 1 and IH or an artificial collection of cases that should actually be labeled as IH or narcolepsy type 1.

Addressing question III that has been stated in section 3.4., one can conclude that all three cluster analyses yielded two “essential” clusters, that mostly agree in their basis properties. The frequencies of the ICSD-3 diagnoses in the clusters revealed a close similarity to IH for some narcolepsy type 2 cases, whereas other narcolepsy type 2 cases were assigned to the cluster dominated by narcolepsy type 1 cases. Regarding question IV, this indicates that the current diagnostic groups might not be a reasonable reflection of the true dataset structure. In chapter 8, it will further be discussed, if and to what degree the algorithmically obtained clusters might justify alternative classifications.

8. General Discussion

Having performed all statistical analyses, it is now time to revisit the major topics of this thesis considering all results that have been obtained. First, MSLT parameters will briefly be discussed again, taking into account the PCA and cluster analysis results. After that, the observations regarding vigilance test results will be summarized. Motivated by the consistent findings of two relevant clusters, a conclusion will be reached regarding the optimal classification of the cases in the dataset at hand with the critical question being whether a certain subdivision into two groups might be more suitable than the ICSD-3 diagnoses.

8.1. Sleep latencies, SOREMs and Delta

The MSLT sleep latencies as well as the SOREM count are essential for diagnosing IH and narcolepsy. This fact is well reflected by the components of the dominant principal component, which comprises the sleep latencies and SOREM count in the expected way.

The occurrence of the SOREM count in the principal component should not be surprising, as it is the main criterion for the differential diagnosis between IH and the narcolepsy subtypes. The SOREM count ranges from 0 in IH patients to 5 in very severe narcolepsy cases. Furthermore, a solid correlation between the SOREM count and SL1 has been shown for both narcolepsy subgroups in linear regression analysis. This correlation between SL1 and the SOREM count serves also as an explanation for the appearance of SL1 in the dominant principal component. As the 8-minute-threshold applies for all diagnoses that are considered in this thesis, it was not clear a priori that SL1 would account for much variance in the dataset.

Apart from the correlation between the SOREM count and SL1, differences in SL1 between the diagnoses (also between the narcolepsy subtypes) should be considered. The ANOVA analysis of the ICSD-3 groups as well as several other authors showed that there is a typical hierarchy regarding the sleep latencies[163, 172]. Compared to narcolepsy, IH patients tend to have slightly higher sleep latencies, and narcolepsy type 2 patients usually show sleep latencies ranging between those of IH and narcolepsy type 1 patients.

Also, the Delta parameter is a significant contributor to the dominant PC. This observation may at least partially be explained by the positive correlation between Delta and the sleep latencies. The R^2-value of the corresponding linear regression analysis was 0,147, hence more than 85% of the

variance in the Delta variable cannot be explained by a (linear) dependence from the sleep latency and by differences between the diagnoses (that are represented by the dummy variables). This is well reflected by the fact that Delta also significantly contributes to PC3 and PC7. The former shows no important contributions by other MSLT parameters, indicating that Delta indeed holds some additional diagnostic information. However, the low fraction of explained variance in linear regression analysis raises doubts if Delta can indeed be of differential diagnostic use, as most of its variance cannot be tracked back to varying diagnoses.

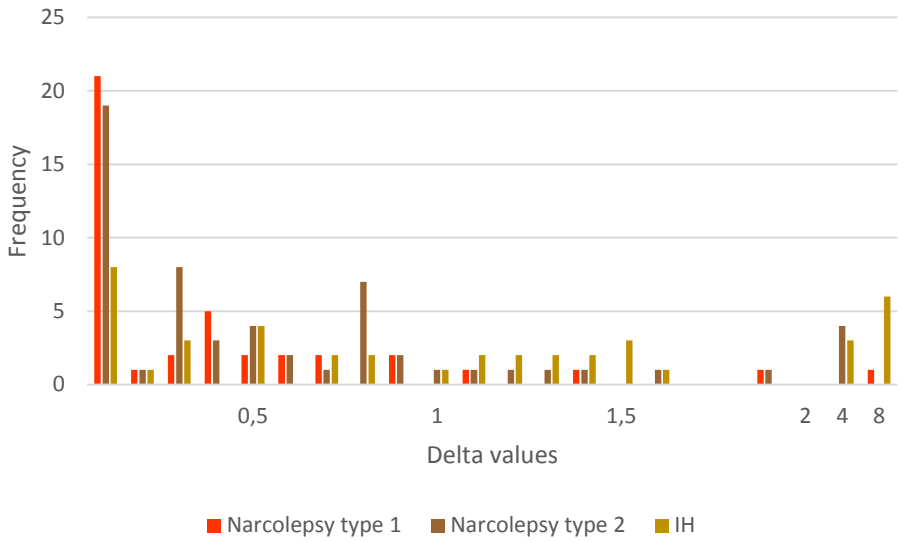
Indirectly, SL1 was chosen as a cluster variable representing PC1 by the variable selection algorithm. Being one of the dominant variables of PC1, the inclusion of SL1 in the cluster analysis ensured that most of the information encoded in PC1 was considered by the cluster algorithms. This leads to a successful clustering process with respect to the axis spanned by PC1, which can best be observed in the k-means cluster plots in the appendix (diagram A.1). Consistently, for each cluster analysis the two important clusters are arranged along the PC1 axis, which essentially means that they contrast cases showing low sleep latency, high SOREM counts, low REM latencies and likely cataplectic episodes with cases presenting the opposite phenotype.

In general, the emergence of PC1 as a principal component summarizing essential and commonly used MSLT parameters suggest a key role of these parameters in the differential diagnosis of IH and the narcolepsy subtypes. In the current diagnostic criteria, which do not consider sleep latency for the purpose of differential diagnosis, this finding is only reflected by the role of the SOREM count in differentiating narcolepsy from IH. Here, a major limitation of the results obtained in this thesis becomes apparent. As no healthy controls could be considered in the PCA and cluster analysis, the interpretations above are limited to differential diagnostic considerations. Nevertheless, the conclusion can be reached that the commonly used MSLT parameters are important tools to characterize cases of IH and narcolepsy. This fact is supported by the significant differences between the ICSD-3 diagnoses and by the consistent patterns regarding these variables, that have been observed in all cluster solutions.

All cluster algorithms proved also to be successful with respect to Delta. The descriptive statistics regarding the ICSD-3 diagnostic groups revealed that on average, narcolepsy patients show Delta values about 0,5 minutes, whereas for IH patients, a mean Delta of 1,5 minutes was found. On the other hand, the corresponding R^2-value indicated that more than 85 % of the variance in Delta

cannot be explained by sleep latency or – more importantly – by the different diagnoses. The following diagram may help to solve this apparent contradiction.

Diagram 8.1: Frequency of different Delta values for the ICSD-3 diagnoses



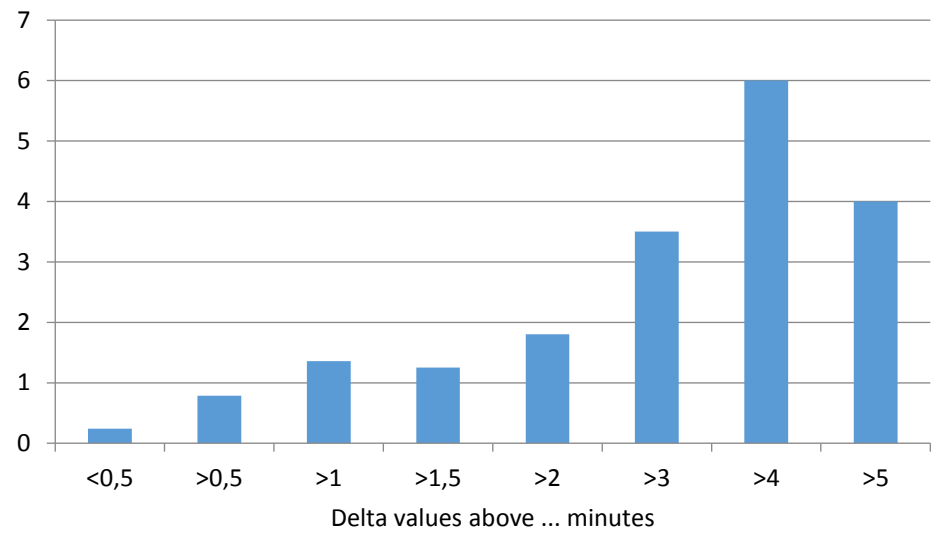
As the diagram indicates, most patients have Delta values below one minute. The significant differences in the mean Delta values are caused by the differing frequencies of outlier cases with extreme high values of Delta. Diagram 8.1 clearly illustrates that most of these outlying cases can be found in the IH group.

Similarly, in the k-means cluster solution, both the members of the narcolepsy-dominated and the "IH-narcolepsy type 2" cluster show low Delta values in most cases. But since more outliers have been assigned to the latter, there is a significant difference in the mean values between the clusters (see diagram A.7 in the appendix).

In conclusion, these diagrams demonstrate that for most cases the Delta parameter is not of differential diagnostic use. However, for increasing Delta values, the diagnosis of narcolepsy becomes less and less likely. Diagram 8.2 depicts the odds ratio for an IH diagnosis for Delta values above a certain threshold. Clearly, the finding of a Delta value above three minutes renders IH as a far more likely diagnosis than narcolepsy.

Diagram 8.2: Odds ratios for the diagnosis of IH versus diagnosis of narcolepsy

Odds ratio for the diagnosis of IH against the diagnosis of narcolepsy of any subtype



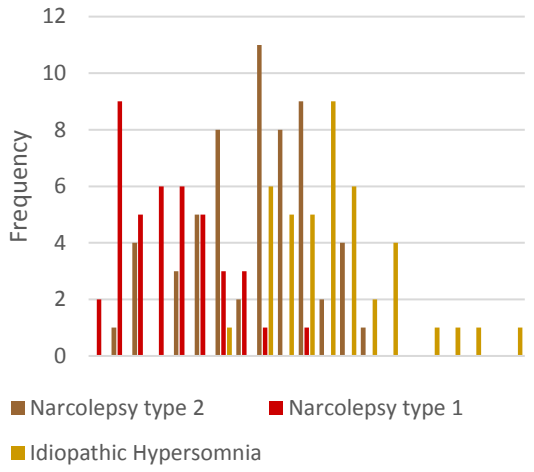
8.2. Vigilance test results

It has been described above that the impairment of vigilance is another aspect of sleepiness that is not covered by the MSLT. By including the results of the Quatember-Maly vigilance test into the PCA and consequently into the cluster analysis, the former was taken into consideration for the detection of clusters. Two essential observations regarding the vigilance test results could be made. First, the second principle component, PC2, is essentially a score for the overall performance in the Quatember-Maly test. It contrasts patients showing a higher than average number of right reactions, a lower than average number of false reactions and relatively low reaction times with cases in which a poor overall performance was observed. Interestingly, this principal component shows only one additional significant contribution and associates a good test performance with an increased nocturnal arousal index.

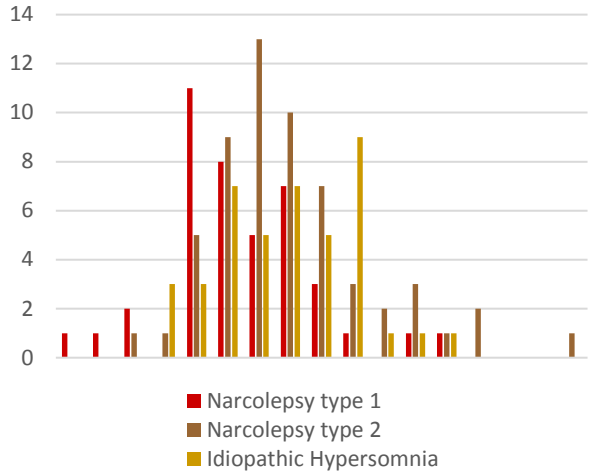
The visual impression of the ICSD-3 groups plotted along the first three principal components suggested that whereas some order of the diagnostic groups along PC1 became apparent, the different diagnoses could not be distinguished along PC2, as all three groups show comparable scores on PC2.

Diagram 8.3: Frequency distributions for the two first principal components

PC 1



PC 2



VigFalse was chosen as cluster variable by the selection algorithms, therefore representing a significant fraction of PC2 in the cluster analysis. This leads to the second important observation: No clustering algorithm yielded significant differences with respect to the non-cluster variables VigRT and VigCorr, and the OPTICS algorithm even failed to produce clusters which significantly differ regarding VigFalse.

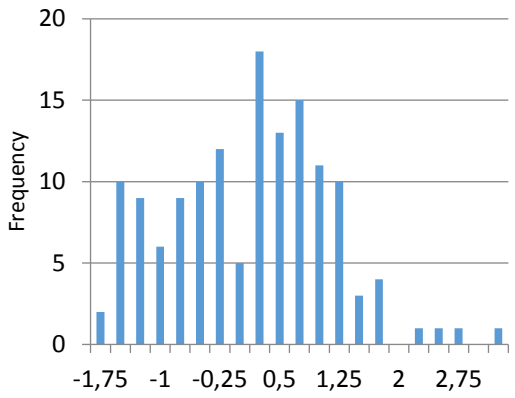
Hence, despite forming the second most important principal component, the vigilance test results were not considered accordingly by the cluster algorithms. Especially the failure of successful clustering with respect to VigFalse in OPTICS highlights this fact. This raises the question why VigFalse has been selected as a cluster variable despite its apparent uselessness for cluster analysis.

From a methodical point of view, a reasonable explanation can be offered. Three of 14 input variables for the PCA were vigilance test parameters, which show a moderate correlation between each other and significantly less correlation with respect to the other variables. Therefore, PCA identified an overall vigilance test score as an important principle component, which could explain a reasonable amount of variance in the vigilance test results. Indirectly, this leads to the selection of VigFalse as a cluster variable.

Diagram 8.4: Frequency distribution of PC 1 and 2

All IH and narcolepsy cases are considered, the diagnoses are not differentiated

PC1



PC2

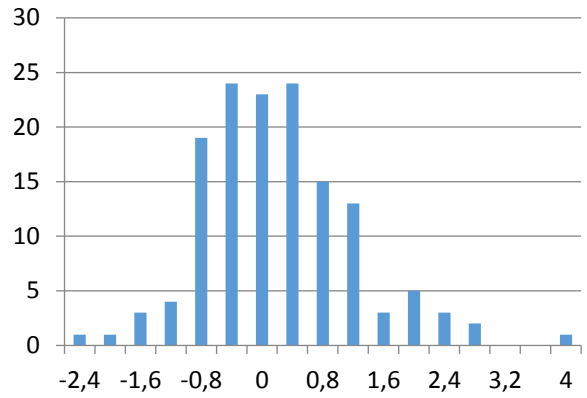


Diagram 8.4 offers distribution plots for the first two principal components. Whereas PC1 indicates a certain possibility for reasonable clustering as more than one peak can be observed, PC2 seems to be almost normally distributed. If there are any reasonable subgroups to be identified in the dataset, no apparent separation of these groups can be observed regarding PC2. The low correlation coefficients regarding variables that are unrelated to the vigilance test indicate that by considering vigilance test results as useful variables for the PCA, three additional dimensions have been added to the dataset that appear to be more or less unrelated to the "remaining structure" of the dataset. According to this interpretation, the emergence of PC2 as an overall vigilance score as well as the selection of VigFalse for the purpose of clustering is merely a mathematical consequence of including these variables in the first place.

Referring to the introductory section for vigilance tests, the argument can be concluded as follows: Despite being an important dimension of sleepiness, impairment of vigilance as measured by the Quatember-Maly test yields no additional information regarding the classification of the dataset at hand. As the results of this thesis suggest, vigilance test results are too unrelated to the established diagnostic parameters and mainly acted as "additional layer of statistical noise" which subsequently leads to the inclusion of VigFalse as a cluster variable. According to the cluster results, the Quatember-Maly vigilance test does not distinguish properly between IH and narcolepsy subtypes and does not provide variables along which clusters could effectively be identified and distinguished from each other.

8.3. Groups in the dataset: How many are there?

As null hypothesis for the cluster analyses, the existence of three reasonable clusters which correspond with the ICSD-3 diagnoses has been postulated. Consistently, despite considerable differences between the algorithms, every cluster solution showed two sensible clusters accompanied by a residual or noise cluster. A closer look at the cluster solutions revealed, that in every case the narcolepsy type 2 cases had been split up in a similar manner. One considerable fraction was assigned to an IH dominated cluster, whereas most of the remaining cases where summarized together with the majority of narcolepsy type 1 cases.

Hence, all cluster solutions hint at a two-group classification of the dataset instead of the three diagnostic groups described in the ICSD-3. The described structure of each cluster pair suggests that the narcolepsy type 2 cases in the dataset form a heterogenous group, which is therefore easily split up during the clustering process.

In this section, the properties of the narcolepsy type 2 subgroup will be investigated in further detail. As it has been discussed above, IH and narcolepsy type 1 are far more easily distinguishable from each other. Because of that, reaching a conclusion regarding the narcolepsy type 2 subgroups will allow a final statement addressing the structure of the whole dataset: If indeed narcolepsy type 2 is a reasonable diagnostic concept despite its apparent heterogeneity, the ICSD-3 groups with narcolepsy type 2 being the "intermediate" group between IH and narcolepsy type 1 should be regarded as reasonable concepts. However, if one agrees that the results of this thesis do not support treating all narcolepsy type 2 cases as one diagnostic entity, a subdivision of the dataset into two diagnostic groups as suggested by all cluster solutions might be the more appropriate approach.

Each cluster analysis divided the narcolepsy type 2 group in a similar fashion. The larger group was always assigned to the majority of IH cases, whereas the remaining cases were summarized together with most narcolepsy type 1 patients. Hence, if one neglects the residual/noise clusters, each cluster analysis offered a different subdivision of the narcolepsy type 2 cases into two subgroups.

In order to investigate how the different pairs of narcolepsy type 2 subdivisions are related to each other, the Rand index was used again. For each pairwise comparison, cases that had been assigned to a residual or the noise cluster were omitted.

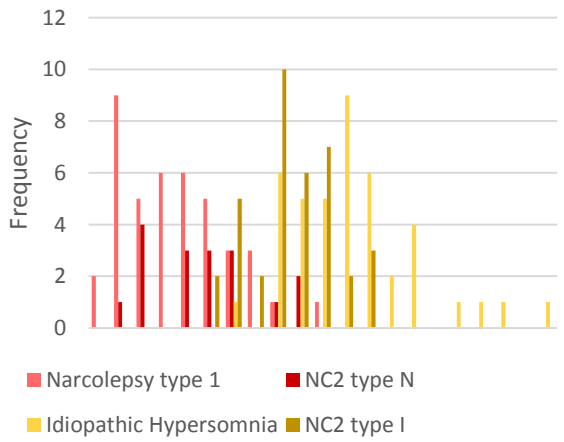
Table 8.1: Comparison of the different narcolepsy type 2 subdivisions using the adjusted Rand index The number of compared cases are shown in brackets						
	k-means		OPTICS		spectral clustering	
k-means	1	(54)	0,84	(35)	0,76	(51)
OPTICS	0,84	(35)	1	(35)	0,71	(35)
Spectral clustering	0,76	(51)	0,71	(35)	1	(54)

Table 8.1 reveals that the three subdivisions of narcolepsy type 2 are closely related to each other: For most pairs of narcolepsy type 2 cases, the cluster algorithms agree on whether they should be assigned to the same cluster or to two different clusters. This observation justifies that the further analysis is conducted using only the k-means solution.

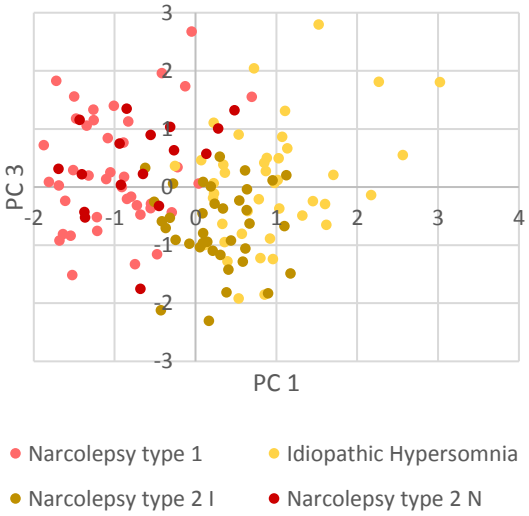
K-means assigned 37 cases of narcolepsy type 2 to the IH dominated clusters. This subgroup of cases will be referred to as “Narcolepsy type 2 I”, whereas the 17 cases that were grouped together with narcolepsy type 1 will be called “narcolepsy type 2 N”.

Diagram 8.5: The frequency distribution along PC1 and the scatterplot of the ICSD-3 diagnoses considering the narcolepsy type 2 subgroups

Distribution along PC1



Scatterplot



The scatterplots indicate that the narcolepsy type 2 subgroups are indeed embedded in surrounding cases of IH and narcolepsy type 1, respectively. The distribution plot reveals that when it comes to PC1, the subgroups are well separated from each other, but hard to distinguish from the dominant

condition in their respective cluster. This observation is confirmed by the descriptive statistics below. It should be noted that although a t-test has been performed, it should be interpreted in a descriptive manner, analogous to the cluster solutions.

Table 8.2: Descriptive statistics for cluster variables considering the proposed narcolepsy type 2 subgroups

The t-test found significant differences between the narcolepsy type 2 subgroups for all variables except ESS (at significance level 0,05). Between NC1 (narcolepsy type 1) and narcolepsy type 2 N, no significant differences were found. Regarding IH and narcolepsy type 2 I, significant differences were only detected for Delta and PSG_SEI.

Diagnosis	**NC1**	**narcolepsy type 2 N**		**narcolepsy type 2 I**		**IH**
Number of cases	**41**	**17**		**37**		**42**
	Mean	Mean	SD	Mean	SD	Mean
ESS	16,3	17,1	4,75	14,9	3,81	15,0
PSG_AI	9,00	11,3	10,4	4,44	5,07	6,43
VigFalse	3,98	7,12	7,36	2,16	2,43	2,29
Delta	0,439	0,200	0,324	0,643	0,726	1,56
SL1	2,43	2,63	2,14	4,93	1,78	5,45
PSG_REML	48,9	33,2	38,5	78,2	32,7	95,6
PSG_SEI	86,7	88,3	6,18	94,0	4,02	89,6

Table 8.2. and A.11 in the appendix provide the mean and standard deviation values for cluster and non- cluster variables, respectively.

Apart from the ESS-score, the two narcolepsy type 2 subgroups show significant differences for every cluster variable. Narcolepsy type 2 N does not significantly differ from narcolepsy type 1 for any cluster variable. Apart from Delta, significant differences between narcolepsy type 2 I and IH exist only for the sleep efficiency index, which has the highest average value for narcolepsy type 2 I cases.

Narcolepsy type 1 and narcolepsy type 2 N both show average sleep latencies below three minutes in contrast to mean sleep latencies about 5 minutes for IH and narcolepsy type 2 I. Regarding the REM latencies in the PSG, narcolepsy type 2 N and narcolepsy type 1 have comparably low mean values, whereas narcolepsy type 2 I and IH both have average values above 70 minutes.

In conclusion, the way narcolepsy type 2 has been split up by the k-means algorithm appears reasonable. Narcolepsy type 2 shows considerable overlap with both narcolepsy type 1 and IH in the PC scatterplots. This unsatisfactory situation can be resolved by accepting the existence of two subgroups of narcolepsy type 2, that are closely related to narcolepsy type 1 and IH, respectively.

On the same time, the introduced subgroups show considerable differences between each other, which resemble the typical reported differences between IH and narcolepsy type 1/with cataplexy.

Again, caution is necessary for the interpretation of these findings. The narcolepsy type 2 subgroups have been defined based on their assignment to different clusters of the k-means solution. Being a cluster analysis procedure, k-means yielded clusters that differ with respect to the provided cluster variables. Therefore, the characteristics of these subgroups and in particular their differences to each other as reported in table 8.2 can also be explained purely as a consequence of their very definition.

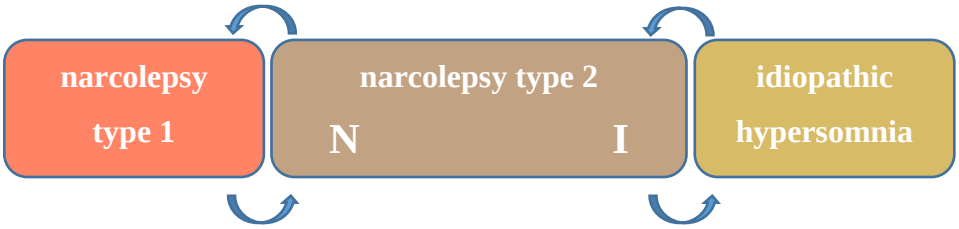
Only by identifying pathophysiological pathways and biomarkers for narcolepsy type 2 and IH, the true relationship between these conditions can be explored. Whereas almost all narcolepsy type 1 patients show significantly lowered levels of CSF hypocretin, this has been reported to be true only for a minor fraction of narcolepsy type 2 patients[152]. Therefore, if another common pathophysiology other than hypocretin deficiency were found for IH as well as for the remaining cases of narcolepsy type 2, this would strongly support a "two groups solution" instead of the current 3 diagnoses.

Until then, the fact that all cluster algorithms have split up narcolepsy type 2 in a comparable manner can be interpreted by two different approaches.

First, one could agree with the assumption that narcolepsy type 2 is a unique nosologic entitiy that shows a considerable heterogeneity. Using the current diagnostic criteria for IH and narcolepsy type 1, the existence of a considerable number of cases that are located "in between" is undebatable. At the current stage, these cases are mostly comprised by the label narcolepsy type 2. For many members of this group, diagnosis can easily change as it is indicated by diagram 8.6 if a second MSLT reveals a different SOREM number or if cataplexy has a delayed onset. Both phenomena have been described in clinical studies and are directly reflected by the two subgroups of narcolepsy type 2 that have been investigated above and which show a close relationship to either IH or narcolepsy type 1.

Diagram 8.6: The ICSD-3 diagnoses and the narcolepsy type 2 subgroups

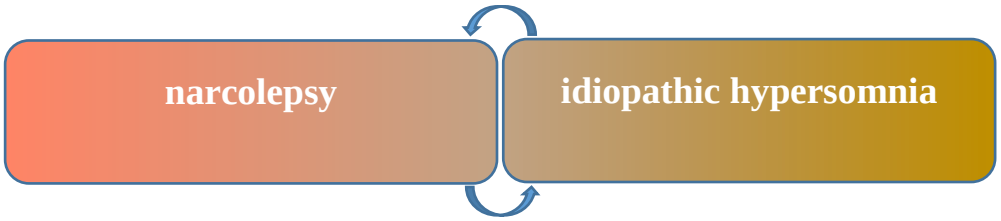
The blue arrows indicate differential diagnostic uncertainties



A more speculative approach would lead to the conclusion that the perceived heterogeneity of narcolepsy type 2 is an artificial one, as narcolepsy type 2 is no distinct diagnostic group but consists of cases that should better be diagnosed with "narcolepsy" (as an undivided diagnostic group) or IH. Continuing this line of thought, these misdiagnoses have led to the impression that narcolepsy type 2 exists, although the heterogeneity of this group suggests the opposite.

Diagram 8.7: The two-clusters classification that is suggested by the cluster solutions

The blue arrows indicate differential diagnostic uncertainties



These considerations correspond well to a study by Andlauer et al.[70], which solely focuses on patients diagnosed with narcolepsy without cataplexy. In this paper, three subgroups of narcolepsy without cataplexy were distinguished by the differing CSF hypocretin concentrations of their members. In the subgroup of low CSF levels, all patients were HLA positive, indicating a close relationship to narcolepsy with cataplexy / type 1. Furthermore, hypocretin deficient patients were far more likely to develop cataplexy than patients with intermediate or normal CSF hypocretin levels. About one fourth of patients in this study were hypocretin deficient, a fraction that is comparable to the 17 of 58 narcolepsy type 2 patients that have been assigned to the N subtype which also showed a high similarity to narcolepsy type 1. In conclusion, the results of Andlauer et

al. suggest a subclassification of narcolepsy type 2 / without cataplexy derived from pathophysiological considerations that matches the algorithmically motivated subdivision proposed above surprisingly well.

8.4. IH or narcolepsy (type 2): differential diagnosis

In the introduction of this thesis, the problem of the differential diagnosis between IH and narcolepsy type 2 was discussed. The last section contributed some insights into the nature of this problem: The narcolepsy type 2 I subgroup shows a high similarity to IH, such that only the MSLT SOREM count can distinguish the otherwise very closely related cases of IH and narcolepsy type 2 I.

If one follows the second approach that has been discussed above and illustrated by Diagram 8.7, this problem might be ill-posed, as a separation of cases that should be diagnosed with the same condition is tried. Nevertheless, for the moment, the validity of the three ICSD-3 diagnoses will be assumed, whereas at the end of this section, the alternative diagnostic groups will be considered.

Hence, all results of this thesis that are related to the issue of differential diagnosis will be summarized at this point. Pizza et al. suggested using Delta as an additional MSLT parameter to improve differential diagnosis between IH and narcolepsy[4]. From the perspective of the mean values, both narcolepsy subtypes have significantly lower values than IH. However, as section 8.1 illustrated, this is merely a consequence of the higher number of outliers in IH. Most cases in the dataset have Delta values below one minute. Only high Delta values, which make the diagnosis IH far more likely than narcolepsy, are of limited differential diagnostic use.

Furthermore, the night sleep of IH and narcolepsy has been reported to show some significant differences. According to Martinez-Rodríguez et al., the night sleep of narcolepsy patients is more fragmented than for IH patients[92]. Another look at the descriptive statistics in table 8.2 reveals that the – on average – lowest arousal indices occurred for the patients labeled with narcolepsy type 2 I, which were not significantly different from IH patients. Hence, more variability seems to exist between different narcolepsy type 2 patients than there is between narcolepsy type 2 I and IH patients.

A similar observation can be made for the fraction of N3 sleep. PSG_N3_TST yielded the highest average value for narcolepsy type 2 I, that does, however, not significantly differ from the mean

value for IH. On the other hand, significantly lower values were found for narcolepsy type 1 and type 2 N.

The t-test found significant differences regarding the average values for SL2 and TST, but as the distribution plots in the appendix indicate, a substantial overlap discourages their use for differential diagnostic purposes (see diagram A.8 in the appendix).

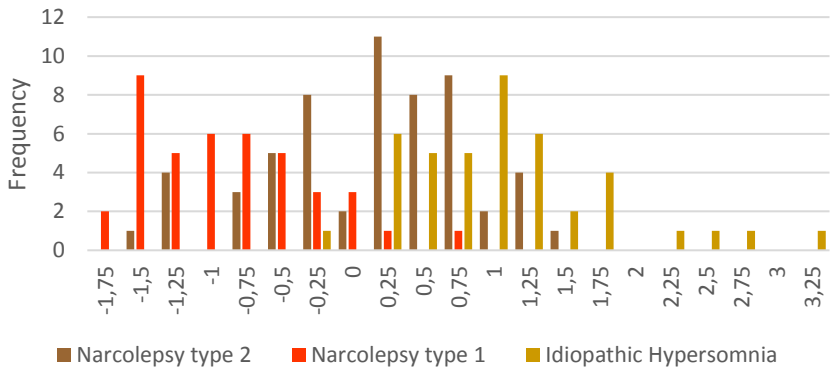
Hence one remains with the SOREM count as remaining and current diagnostic criterion, whose shortcomings have already been discussed above. In conclusion, the analyses performed in this thesis did not yield results that could help to improve differential diagnosis between IH and narcolepsy (type 2).

As the results of this thesis did not reveal any suggestions for the refinement of the differential diagnostic process in the situation of the conventional diagnostic groups, another perspective will be employed from now on. Hence, for the remaining discussion, the narcolepsy type 2 N subgroup and narcolepsy type 1 will be summarized as “narcolepsy +”, whereas narcolepsy type 2 I will be merged with IH, forming the” idiopathic hypersomnia +” group.

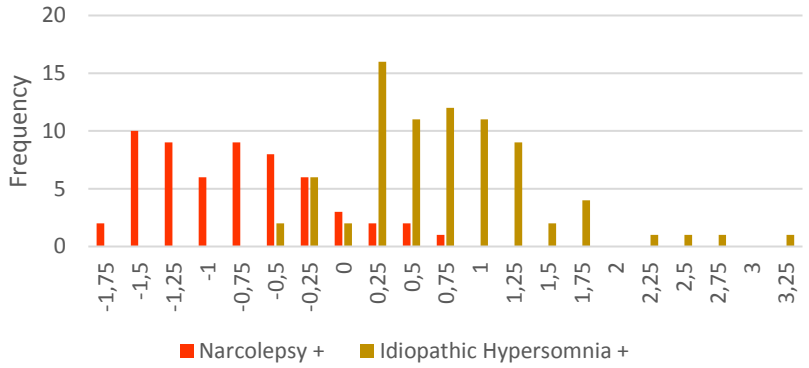
Diagram 8.8 indicates that narcolepsy + and IH + are well separated along the PC1 axis. Regarding PC2 and PC3, no such clear separation becomes visible (see diagrams A.9 and A.10 in the appendix). Hence, the essential information necessary to distinguish between narcolepsy + and idiopathic hypersomnia + is encoded in PC1. This insight allows the suggestion of a diagnostic score that considers all relevant components of PC1.

Diagram 8.8: Frequency distribution along the PC1 axis

ICSD-3 diagnoses



Narcolepsy + vs. idiopathic hypersomnia +



As it has been explained above, principal components are linear combinations of the input variables. Table A.2 in the appendix lists all coefficients for all principal components, but with respect to the z-standardized variables.

Hence, by the following calculation, the derivation of a diagnostic score for the unstandardized variables is possible. Assuming that V_i denotes the i-th unstandardized input variable with the mean value μ_i and standard deviation σ_i, ZV_i its z-standardized version and $\propto_i$ its coefficient for PC1, one obtains:

$$PC1 = \propto_1 ZV_1 + \ldots + \propto_{14} ZV_{14} = \sum_i \propto_i ZV_i = \sum_i \propto_i \frac{V_i - \mu_i}{\sigma_i} = \sum_i \frac{\propto_i}{\sigma_i} V_i + C,$$

where C is a constant that does not depend on any variable. The corresponding coefficients for the unstandardized variables are therefore obtained by dividing the coefficients in table A.1 by the standard deviation of the corresponding variable.

For the diagnostic score, only the seven variables that made significant contributions to PC1 were considered. Furthermore, the coefficients were multiplied by 100 to obtain more practical values.

Eventually, the following linear score N was obtained:

$$N = 11{,}4\ \#SOREM - 10{,}5\ Delta - 0{,}26\ PSG\_REML - 9{,}48\ SL1 - 5{,}42\ SL2 + 1{,}3\ TST + 34{,}6\ CATAP$$

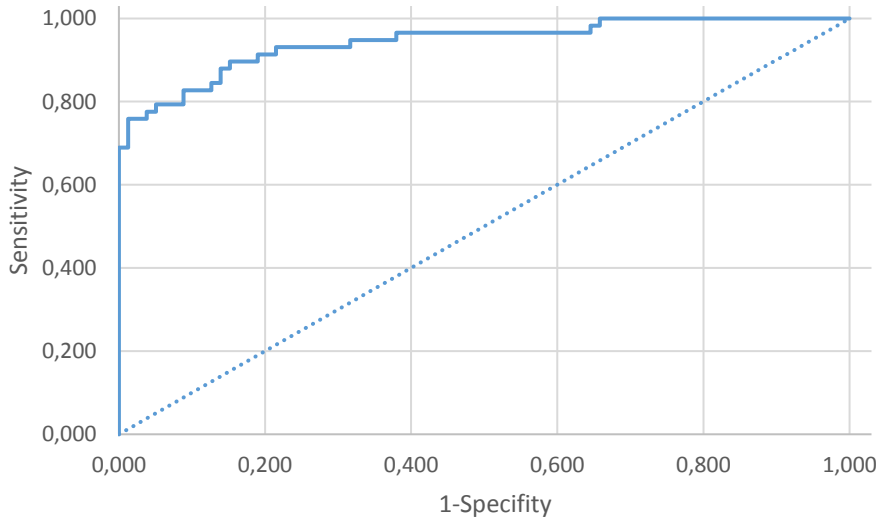
Similar to PC1, the score contrasts cases showing many SOREM episode, short sleep latencies, short REM latencies and low Delta values. Therefore, high score-values are expected to indicate that the given case should be assigned to narcolepsy +.

Effectively, the occurrence of cataplexy adds 34,6 points to the score, which could be counterbalanced by 12 minutes less in the TST or a sleep latency that is prolonged by 4 minutes. Therefore, regardless of the optimal cut-off value, cataplexy does not directly lead to the diagnosis of narcolepsy (+). However, due to the typical results of narcolepsy patients regarding the sleep latencies, PSG REM latencies and Delta, cataplectic patients are expected to score a high total value in the N-score. Furthermore, the occurrence of cataplexy-like symptoms in IH has to be considered. IH patients, whose symptoms might inaccurately be labeled as cataplexy would usually still yield a low N-score. Essentially, the N-score averages all results regarding the considered variables in a balanced way, thereby reducing the risk of a misdiagnosis caused by an atypical result in a single variable. Also, taking into account the possibility of a delayed cataplexy onset, the relative impact of cataplexy in the total score result should not be too high.

For finding the right cut-off values that are suitable for distinguishing narcolepsy + from IH +, the ROC curve was calculated.

Diagram 8.9: ROC curve for distinguishing narcolepsy + from idiopathic hypersomnia +

Values above the cut-off were treated as positive for Narcolepsy +; the dotted line indicates the diagonal



By introducing the alternative diagnostic groups narcolepsy+ and idiopathic hypersomnia+, the differential diagnostic challenge shifted from distinguishing narcolepsy type 2 from IH to the differentiating the new groups from each other. Hence, to allow direct comparison, Diagram A.11 in the appendix depicts the corresponding ROC curve for the latter case. Furthermore, table 8.3 enlists critical cut-off values for the N-score for both situations.

Table 8.3: Critical cut-off values for the N-score

Values above the cut-off are treated as positive for narcolepsy+ / narcolepsy type 2

	narcolepsy+ vs idiopathic hypersomnia +		**narcolepsy type 2 vs. idiopathic hypersomnia**	
Sensitivity	Cut-off value	Specifity	Cut-off value	Specifity
95 %	≥ 51,0	62 %	≥ -15,1	40,5 %
90 %	≥ 77,4	81 %	≥ 14,1	57,1 %
80 %	≥ 110,5	91,1 %	≥ 29,7	66,7 %

It becomes apparent that the N-score performs better in differentiating narcolepsy+ from idiopathic hypersomnia+ compared to the conventional task of differential diagnosis between narcolepsy type 2 and IH. The N-score was derived from the dominant principal component, but nevertheless fails to allow a clear separation between the latter two conditions. This highlights once more the existing differential diagnostic problem, even if the major characteristics of the dataset are considered.

As a critical remark, it should be remembered that the N-score has been directly derived from the properties of PC1. It has been discussed above that the structure of PC1 is also a consequence of including several closely related variables like SL1, SL2 and TST. Indirectly, this has also affected the coefficients of the N-score, so an overestimation of the importance of several MSLT variables cannot be ruled out. Furthermore, it should be considered that all cluster solutions emerged in an algorithmical way, with the cluster variables determined by the principal component analysis being the only input parameters. Hence, all cluster solutions and consequently also the constructs narcolepsy + and idiopathic hypersomnia + are mathematically designed to contrast the differences with respect to the most significant variables.

By considering additional parameters apart from the usual sleep latency, the SOREM count and the occurrence of cataplexy, diagnosis becomes less prone to measurement issues like the test-retest variance regarding the MSLT SOREM count. The linear coefficients in the N-score guarantee, that all included variables are considered in a balanced way, allowing more stable diagnostic estimates. For an example, a narcolepsy patient could, by chance, show only one SOREM in the MSLT, but since the other parameters would probably still yield values that are characteristic for narcolepsy, a diagnosis based on the N-score could remain valid. However, further clinical tests would be necessary to assess the potential of linear scores like the N-score.

9. Conclusion

Three different methodical approaches were chosen for this thesis. First, linear regression analysis demonstrated a considerable correlation between sleep latency and SOREM frequency. This correlation was also reflected in the dominant principal component found by PCA, which allowed an objective choice of variables suitable for cluster analysis. The different cluster solutions did not show any clear similarity to each other. Nevertheless, several cluster features could be observed consistently regardless of the employed cluster method. A summarizing consideration of these results eventually led to the suggestion of alternative diagnostic groups and of a linear score that might prove to be suitable for their differential diagnosis.

Several parts of this thesis are speculative, as it is to be expected by its explorative design. For the most part, no significance levels could be provided to back the interpretations of the results, especially regarding those in the general discussion. As it has briefly been mentioned above, the final judgement of the problems discussed in this thesis is to be expected from future pathophysiological insights. The central interpretations of this thesis could easily be refuted if a common pathophysiological pathway for all narcolepsy type 2 patients would be discovered, thereby rendering the suggested subgroups I and N as well as the suggested diagnostic labels narcolepsy + and idiopathic hypersomnia + irrational.

Diagnostic considerations should have therapeutic consequences. Until today, treatment of both narcolepsy and IH is symptomatic and the most essential substances, stimulants, have been proven to be effective for both conditions. Therefore, even if results like the ones of this thesis eventually led to another refinement of diagnostic entities and criteria, no direct clinical benefit would arise from this.

This situation would drastically change if treatment options based on proven pathophysiological knowledge would enter clinical practice, for example drugs for the substitution of hypocretin. If, according to the interpretations in this thesis, indeed only a fraction of narcolepsy type 2 patients should be labeled as narcolepsy, clinical scores like the N-score could become highly relevant for determining without invasive diagnostics which patients suffering from hypersomnia could benefit from a condition-specific treatment.

Finally, several limitations of this thesis should be considered. First, only 141 cases of narcolepsy and IH could be included. Consequently, different subgroups of this dataset became insufficiently small. This especially affects the suggested narcolepsy type 2 subgroups, with the smaller one only comprising 17 cases. Several patients of the dataset had also a diagnosis for a sleep related breathing disorder. Although each case was only included after reaching a stable, sufficient treatment, the analyses above could have been adjusted to control for possible confounding effects of these conditions. Another issue is the lack of a control group for the PCA and cluster analysis. Considering a group of healthy individuals for both would have allowed to also consider purely diagnostic issues, i.e. distinguishing healthy individuals from IH and narcolepsy. Based on the results of this thesis, two interesting topics would be the following: 1) Does the inclusion of a healthy control group change the basic characteristics of the dominant principal component? 2) Which values of the N-score are to be expected for healthy controls and are they well separated from value regimes typical for IH or narcolepsy?

As a final limitation, the speculative nature of this explorative thesis should be pointed out once more. Several interesting hypotheses can be deduced from the results of this thesis, which could be tested using well established methods of inferential statistics, for example regarding the CSF hypocretin levels in the proposed narcolepsy type 2 subtypes or regarding N-score values in another dataset of narcolepsy or IH patients.

Appendix

1. Tables

Table A.1: Correlation matrix of the PCA input variables

	SL1	SL2	TST	Delta	#SOREM	ESS	VigCorr	VigFalse	VigRT	PSG_SEI	PSG_N3_TST	PSG_REML	PSG_AI	CATAP
SL1	**1,00**	0,76	-0,81	0,28	-0,53	-0,12	-0,02	-0,05	0,00	0,09	0,19	0,37	-0,15	-0,50
SL2	0,76	**1,00**	-0,76	0,46	-0,57	-0,15	0,14	-0,17	-0,08	0,00	-0,01	0,38	0,04	-0,40
TST	-0,81	-0,76	**1,00**	-0,67	0,57	0,11	-0,06	0,09	0,05	0,02	-0,09	-0,39	-0,04	0,41
Delta	0,28	0,46	-0,67	**1,00**	-0,40	0,07	0,09	-0,10	-0,03	0,01	-0,07	0,19	0,14	-0,19
#SOREM	-0,53	-0,57	0,57	-0,40	**1,00**	0,18	-0,06	0,17	-0,04	-0,01	-0,13	-0,45	0,05	0,60
ESS	-0,12	-0,15	0,11	0,07	0,18	**1,00**	-0,17	-0,01	0,18	0,05	-0,16	-0,21	-0,02	0,11
VigCorr	-0,02	0,14	-0,06	0,09	-0,06	-0,17	**1,00**	-0,33	-0,49	-0,13	-0,03	-0,04	0,17	0,08
VigFalse	-0,05	-0,17	0,09	-0,10	0,17	-0,01	-0,33	**1,00**	0,32	-0,12	0,00	-0,10	-0,09	0,09
VigRT	0,00	-0,08	0,05	-0,03	-0,04	0,18	-0,49	0,32	**1,00**	-0,02	-0,05	0,00	0,00	-0,03
PSG_SEI	0,09	0,00	0,02	0,01	-0,01	0,05	-0,13	-0,12	-0,02	**1,00**	0,25	-0,07	-0,18	-0,28
PSG_N3_TST	0,19	-0,01	-0,09	-0,07	-0,13	-0,16	-0,03	0,00	-0,05	0,25	**1,00**	0,14	-0,29	-0,18
PSG_REML	0,37	0,38	-0,39	0,19	-0,45	-0,21	-0,04	-0,10	0,00	-0,07	0,14	**1,00**	-0,14	-0,25
PSG_AI	-0,15	0,04	-0,04	0,14	0,05	-0,02	0,17	-0,09	0,00	-0,18	-0,29	-0,14	**1,00**	0,10
CATAP	-0,50	-0,40	0,41	-0,19	0,60	0,11	0,08	0,09	-0,03	-0,28	-0,18	-0,25	0,10	**1,00**

Table A.2: Coefficients of the principal components
The bold line separates the first seven PCs from the remaining PCs, see section 6.2 .

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
SL1	0,207	0,084	0,002	-0,021	0,133	0,193	-0,403	0,139	0,463	0,031	0,041	0,159	-0,845	2,394
SL2	0,210	-0,066	0,092	0,005	0,066	0,098	-0,281	-0,063	0,328	-0,140	0,209	-0,534	1,646	-0,400
TST	-0,220	0,025	-0,139	-0,026	-0,170	-0,251	-0,053	0,045	-0,162	-0,002	0,171	-0,357	0,802	2,835
Delta	0,143	-0,097	0,238	0,220	0,186	0,213	0,590	-0,329	-0,472	-0,234	-0,216	0,303	0,271	1,472
#SOREM	-0,195	-0,030	-0,002	0,039	0,212	0,280	-0,080	-0,295	0,493	0,137	0,021	10,239	0,605	0,026
ESS	-0,054	0,086	0,214	0,536	0,473	-0,234	-0,039	0,438	-0,018	0,531	0,265	-0,157	0,035	-0,126
VigCorr	0,024	-0,410	-0,161	-0,023	0,069	0,137	-0,037	0,365	-0,282	-0,232	0,958	0,353	-0,163	-0,042
VigFalse	-0,051	0,268	0,198	-0,321	0,041	0,586	-0,113	-0,069	-0,459	0,482	0,465	-0,232	0,050	-0,040
VigRT	-0,020	0,337	0,300	-0,019	-0,224	-0,166	0,211	0,279	0,212	-0,763	0,527	0,374	0,020	-0,126
PSG_SEI	0,022	0,166	-0,296	0,500	-0,275	0,194	0,075	-0,587	0,113	0,002	0,546	-0,286	-0,269	-0,211
PSG_N3_TST	0,046	0,169	-0,386	-0,074	0,132	0,295	0,642	0,589	0,248	0,126	-0,132	-0,078	0,371	0,019
PSG_REML	0,137	0,049	-0,045	-0,324	0,135	-0,575	0,342	-0,396	0,125	0,552	0,514	0,286	0,015	0,123
PSG_AI	-0,015	-0,249	0,293	0,057	-0,619	0,148	0,314	0,179	0,426	0,562	0,052	-0,135	-0,093	0,279
CATAP	-0,158	-0,143	0,117	-0,154	0,445	0,112	0,286	-0,280	0,474	-0,336	0,188	-0,929	-0,505	0,091

Table A.3: ANOVA statistics of the k-means cluster solution: Non-cluster variables								
	Cluster 1 (N=76)		Cluster 2 (N=46)		Cluster 3 (N=19)		Statistics	
	Mean	SD	Mean	SD	Mean	SD	F-value (df=140)	p-value
SL2	8,79	3,25	5,26	2,71	12,2	3,74	36,2	<0,001
TST	119	11,7	132	10,4	94,6	18,7	59,9	<0,001
susSL	5,33	2,07	2,73	1,78	8,92	3,03	58,8	<0,001
#SOREM	2,03	1,51	3,54	1,52	0,789	0,976	28,4	<0,001
VigCorr	86,3	17,6	82,3	20,7	88,6	17,6	0,986	0,376
VigRT	0,547	0,108	0,586	0,129	0,572	0,125	1,64	0,197
PSG_N3_TST	16,9	9,10	11,2	8,54	10,7	8,10	7,66	0,001

Table A.4: Frequency of cataplexy for the k-means clusters			
Fisher-Freeman-Halton statistic: 19,908 (p<0,001)			
Cluster	1	2	3
No cataplexy	63	21	16
Cataplexy	13	25	3

Table A.5: The t-test for the OPTICS cluster solution: Non-cluster variables Mean value and standard deviation for the noise cluster are only listed for the sake of descriptive comparison and not considered in the t-test.								
	Noise cluster (N=65)		Cluster 1 (N=43)		Cluster 2 (N=33)		Statistics (df=74)	
	Mean	SD	Mean	SD	Mean	SD	t-value	p-value
SL2	8,91	3,87	9,52	3,39	4,63	1,99	7,37	<0,001
TST	116	18,5	114	12,5	134	8,98	-7,82	<0,001
susSL	5,49	2,96	6,23	2,28	2,30	1,41	8,68	<0,001
#SOREM	2,02	1,60	1,81	1,53	3,73	1,44	-5,54	<0,001
VigCorr	84,1	19,3	86,1	18,9	86,6	17,5	-0,139	0,89
VigRT	0,576	0,132	0,552	0,115	0,552	0,0887	0,009	0,993
PSG_N3_TST	13,1	8,97	18,5	8,77	10,9	8,35	3,84	<0,001

Table A.6: Frequency of cataplexy for the OPTICS clusters		
Chi-square statistic: 14,4 (p<0,001)		
Cluster	1	2
No cataplexy	38	16
Cataplexy	5	17

Table A.7: ANOVA statistics of the spectral clustering solution: Non-cluster variables								
	Cluster 1 (N=34)		Cluster 2 (N=12)		Cluster 3 (N=95)		Statistics	
	Mean	SD	Mean	SD	Mean	SD	F-value	p-value
SL2	4,74	2,62	7,44	4,22	9,37	3,47	24,1	<0,001
TST	135	8,91	124	16,0	113	15,5	30,1	<0,001
susSL	2,01	1,25	4,01	2,40	6,15	2,54	42,2	<0,001
#SOREM	3,79	1,49	3,42	1,08	1,71	1,46	29,9	<0,001
VigCorr	87,3	17,4	84,1	15,8	84,7	19,6	0,261	0,771
VigRT	0,557	0,102	0,583	0,126	0,563	0,123	0,211	0,810
PSG_N3_TST	11,6	9,06	11,2	6,77	15,4	9,30	3,10	0,048

Table A.8: Frequency of cataplexy in the spectral cluster solution			
Fisher-Freeman-Halton statistics: 32,562 (p<0,001)			
	1	2	3
No cataplexy	14	4	82
Cataplexy	20	8	13

Table A.9: R-code for spectral clustering

The Gauß similarity function

```
# x1, x2, are Vectors that represent different patient cases in the seven-dimensional space spanned by the cluster variables
GaussSimilarity <- function(x1, x2, alpha) {
        exp(- alpha * norm(as.matrix(x1-x2), type="F")^2)
}
```

Calculation of the similarity matrix

```
# As a similarity function, GaussSimilarity with respect to the standardized cluster variables is used.
Similarity_Matrix <- function(data, similarity)
{
        N <- nrow(data)
        S <- matrix(rep(NA,N^2), ncol=N)
                for(i in 1:N)   {
                        for(j in 1:N)   {
                                S[i,j] <- similarity(data[i,], data[j,])
                                        }
                                }
        S
}
```

Calculation of the adjacency matrix (k-nearest neighbor)

```
ADJACENCY_MATRIX_KNN <- FUNCTION(SIM_MATRIX, N.NEIGHBORS=2) {
        N <- LENGTH(SIM_MATRIX [,1])

        IF (N.NEIGHBORS >= N)
        {
                # FULLY CONNECTED
                A <- S
        } ELSE
        {
                A <- MATRIX(REP(0,N^2), NCOL=N)
                        FOR(I IN 1:N)
                        {       # FOR EACH LINE
                                # ONLY CONNECT TO THOSE POINTS WITH HIGH SIMILARITY
                                # EACH VERTEX IS MOST SIMILAR TO HIMSELF, HENCE START AT SECOND BEST!
                                BEST.SIMILARITIES <- SORT(SIM_MATRIX [I,],
                                DECREASING=TRUE)[2:(N.NEIGHBOORS+1)]
                                FOR (S IN BEST.SIMILARITIES)
                                {
                                J <- WHICH(SIM_MATRIX [I,] == S)
                                A[I,J] <- S[I,J]
                                A[J,I] <- S[I,J]        # TO MAKE AN UNDIRECTED GRAPH, IE, THE MATRIX BECOMES
                                                        # SYMMETRIC
                                }
                        }
        }
  A
}
```

The spectral clustering process

```
# PRELIMINARY CALCULATION: SUPPOSE "CASES" IS THE MATRIX WHERE EACH ROW REPRESENTS A CASE AND EACH COLUMN A CLUSTER
VARIABLE
SIM<- SIMILARITY_MATRIX(CASES, GAUSSSIMILARITY)
ADJ_MAT<- ADJACENCY_MATRIX_KNN(SIM, 2)

# CALCULATE THE (NORMALIZED) LAPLACIAN MATRIX L
        D<- DIAG(APPLY(ADJ_MAT, 1, SUM))
        L<- DIAG(1,LENGTH(ADJ_MAT[,1])-SOLVE(D)%*%A
# CALCULATE EIGENVECTORS OF THE LAPLACIAN MATRIX
        EIGVEC<-EIGEN(L)
#OBTAIN SPECTRAL SPACE, I.E. THE LAST THREE EIGENVECTORS
        Z<- EIGVEC$VECTORS[,(NCOL(EIGVEC $VECTORS)-2):NCOL(EIGVEC$VECTORS)]
```

```
#PERFORM K-MEANS CLUSTERING ON Z
        KM<-KMEANS(Z, CENTERS=3,NSTART=50)
```

Table A.10: ANOVA statistics for the ICSD-3 diagnoses: Non-cluster variables

SD: Standard deviation

	narcolepsy type 2		narcolepsy type 1		idiopathic hypersomnia		Statistics	
	Mean	SD	Mean	SD	Mean	SD	F-value	p-value
SL2	7,89	3,34	5,69	3,25	10,7	3,52	23,4	<0,001
TST	120	14,9	131	11,0	108	17,3	24,3	<0,001
susSL	4,98	2,49	2,87	1,69	7,01	2,89	30,3	<0,001
#SOREM	2,79	,932	3,95	1,02	0,190	0,397	220	<0,001
VigCorr	83,8	22,2	87,7	14,6	85,0	17,1	0,512	0,600
VigRT	0,560	0,124	0,557	0,111	0,573	0,118	0,215	0,807
PSG_N3_TST	15,2	9,58	11,7	8,39	15,3	9,17	2,19	0,116

Table A.11: Descriptive statistics for non-cluster variables considering the proposed narcolepsy type 2 subgroups

The t-test showed significant differences between the narcolepsy type 2 subgroups for all variables except VigCorr and VigRT,

Compared to narcolepsy type 1, narcolepsy type 2 N differed only significantly regarding #SOREM

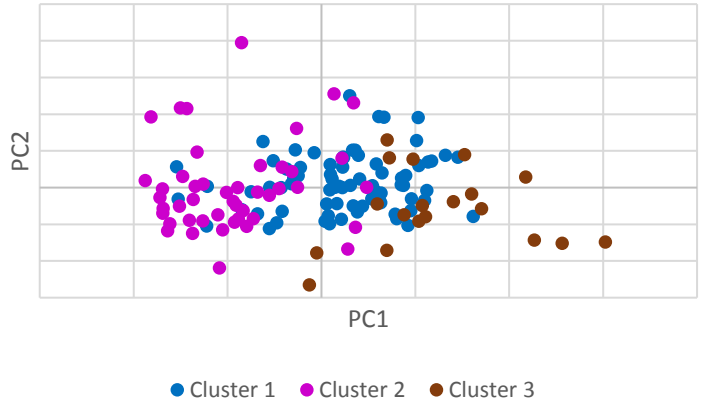
A significant difference between narcolepsy type 2 I and IH was found for: SL2, TST, susSL and #SOREM

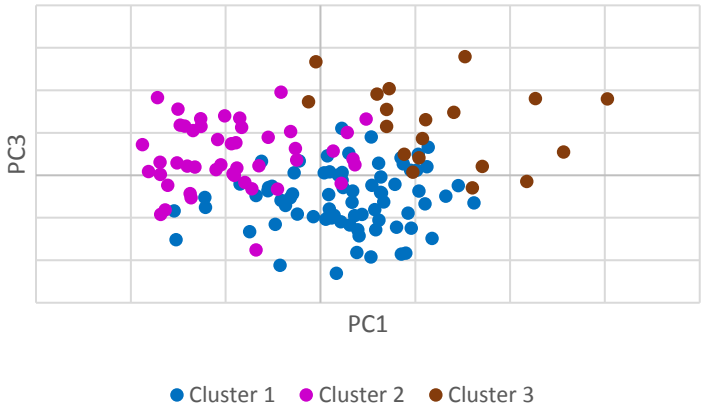
SD: Standard deviation

	NC1	Narcolepsy type 2 N		Narcolepsy type 2 I		IH
	Mean	Mean	SD	Mean	SD	Mean
SL2	5,69	5,67	3,82	8,50	2,34	10,7
TST	131	132	12,8	117	11,4	108
susSL	2,87	2,83	2,28	5,58	1,84	7,01
#SOREM	3,95	3,29	1,10	2,65	0,789	0,190
VigCorr	87,7	81,4	25,6	86,1	19,6	85,0
VigRT	0,557	0,597	0,147	0,540	0,112	0,573
PSG_N3_TST	11,7	10,4	8,15	18,2	9,43	15,3

2. Diagrams

Diagram A.1: The k-means cluster solution, projected onto the first principal components





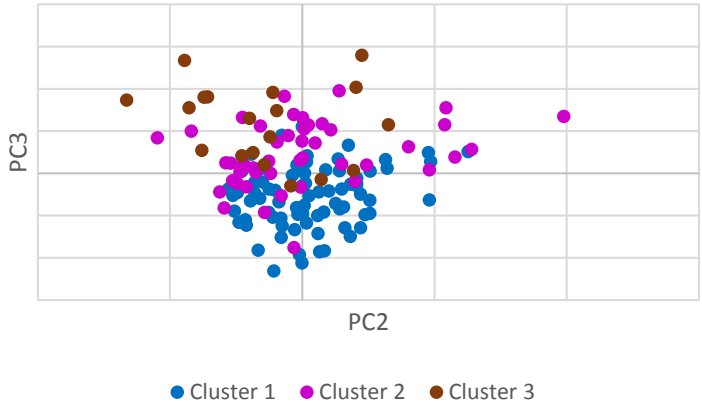
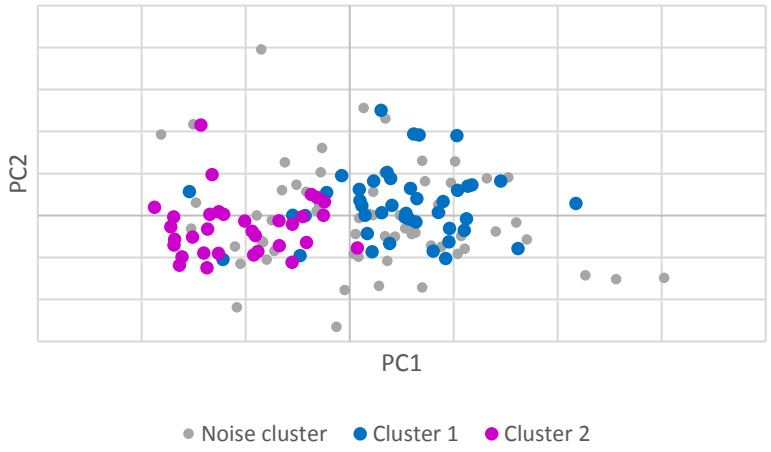
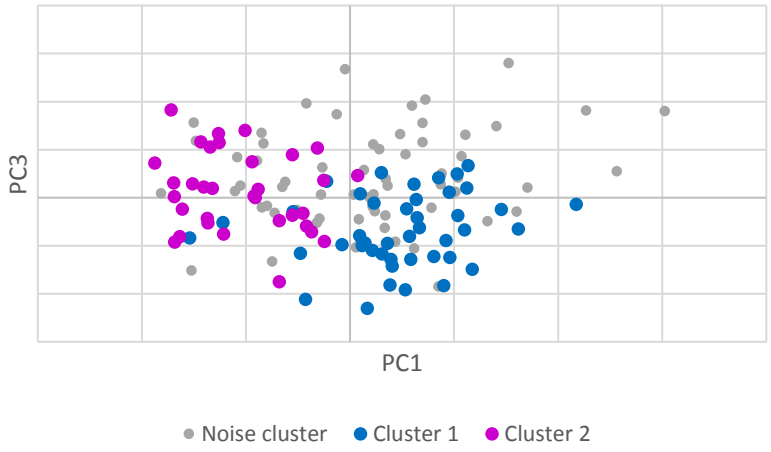


Diagram A.2: The OPTICS cluster solution, projected onto the first principal components





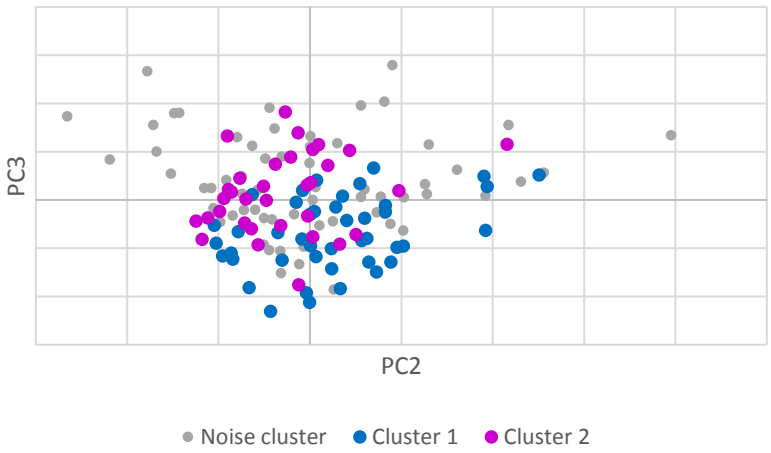
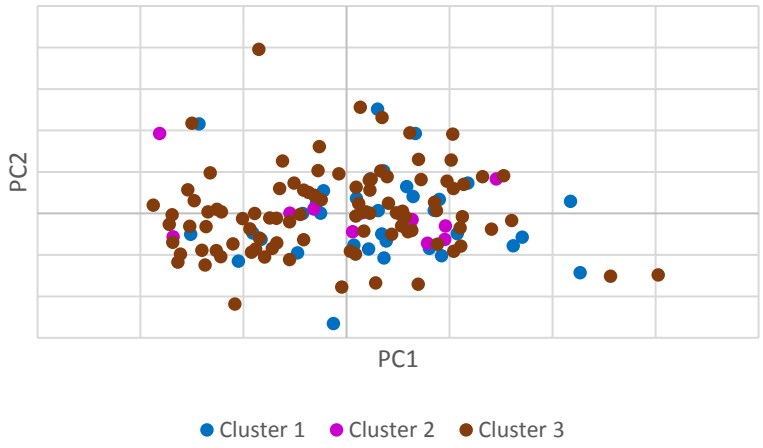
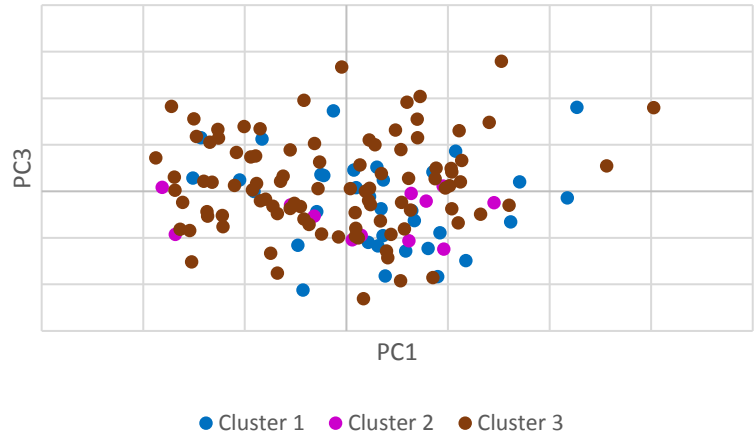


Diagram A.3: The spectral clustering solution, projected onto the first principal components





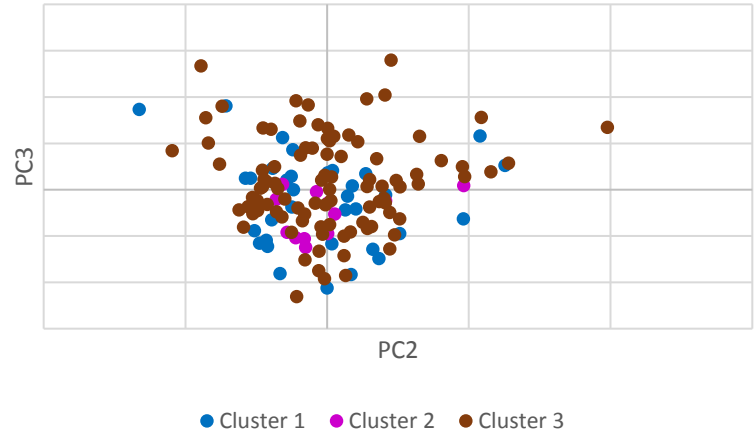


Diagram A.4: No cluster structure detectable by OPTICS using the standard Euclidean metric

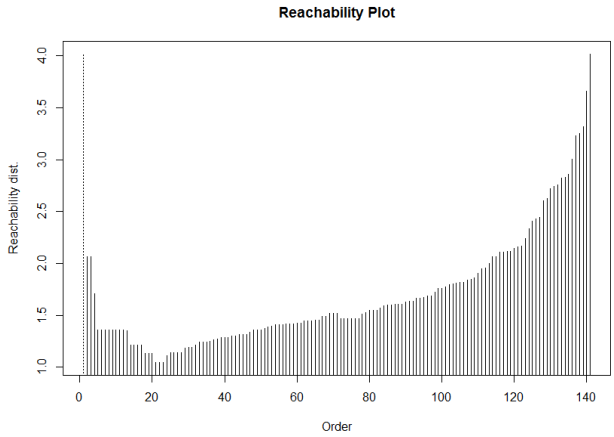


Diagram A.5: Silhouette coefficients for the OPTICS cluster solution using the angular distance

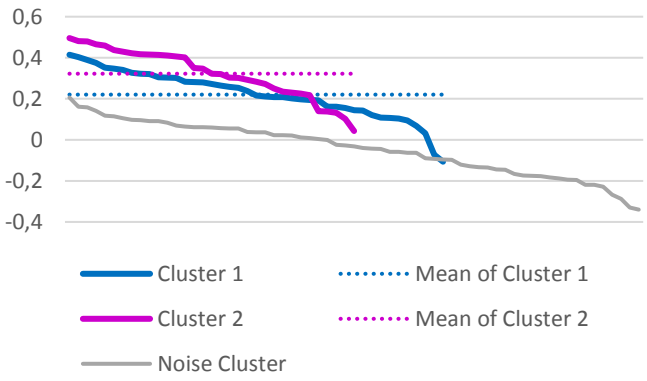


Diagram A.6: Silhouette coefficients of the spectral clustering solutions in the eigenvector space

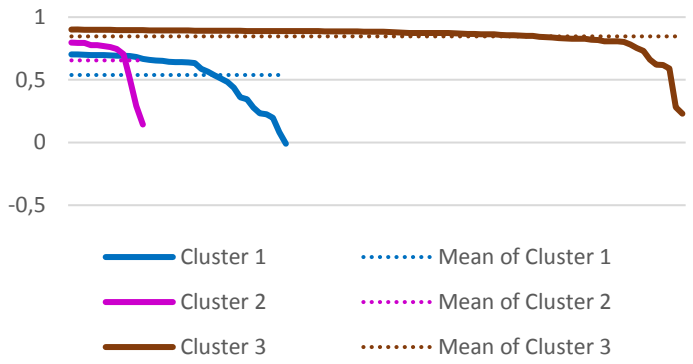


Diagram A.7: Frequency of different Delta values in the k-means clusters 1 and 2

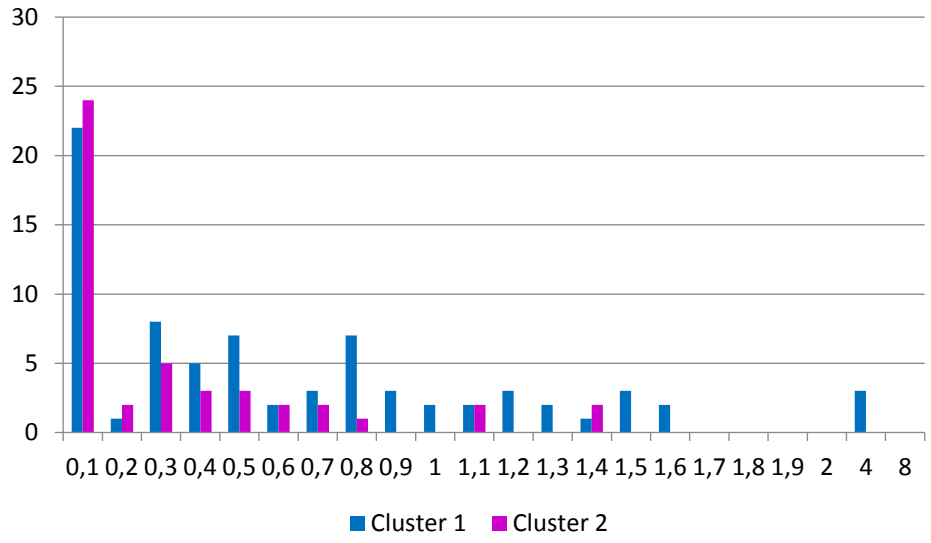
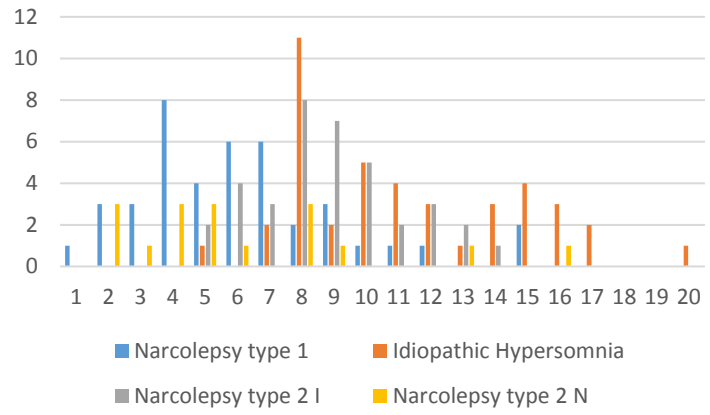


Diagram A.8: Frequency distributions for SL2 and TST

For narcolepsy type 2, the two subgroups proposed in chapter 8 are considered

SL2



TST

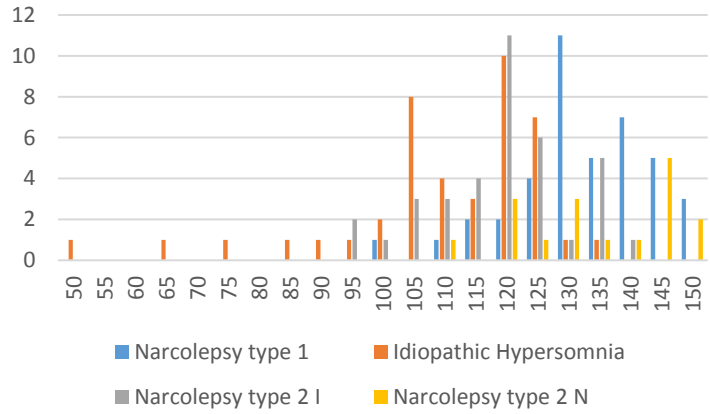


Diagram A.9: Distribution plot of narcolepsy + and idiopathic hypersomnia + along PC2

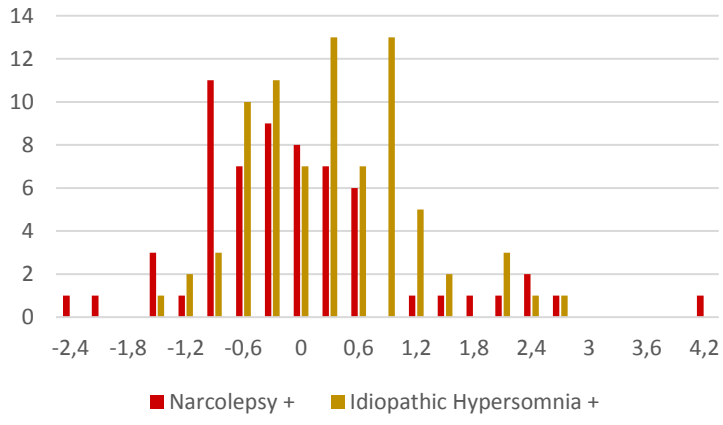


Diagram A.10: Distribution plot of narcolepsy + and idiopathic hypersomnia + along PC3

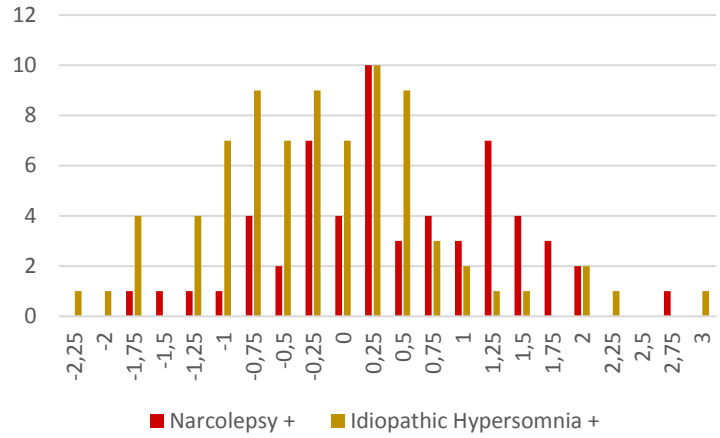
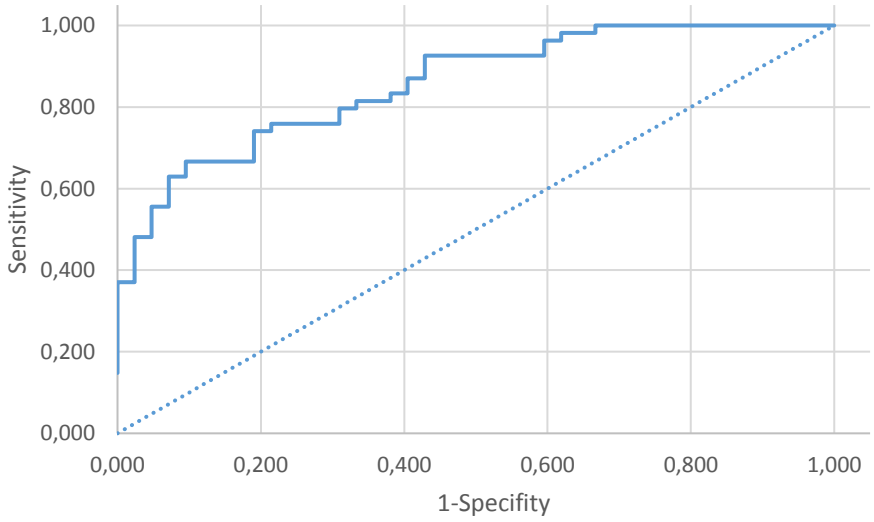


Diagram A.11: ROC curve for distinguishing narcolepsy type 2 from idiopathic hypersomnia using the N-score

Values above the cut-off were treated as positive for Narcolepsy type; the dotted line indicates the diagonal



List of tables

List of diagrams

Index of abbreviations

#SOREM	number of SOREM episodes in the MSLT
CATAP	occurrence of cataplexy, encoded as binary variable
CNS	central nervous system
CSF	cerebrospinal fluid
Delta	susSL – SL1
EDS	Excessive Daytime Sleepiness
ESS	Epworth Sleepiness Scale
ICSD	International Classification of Sleep Disorders
IH	idiopathic hypersomnia
MSLT	Multiple Sleep Latency Test
MWT	maintenance of wakefulness test
OSAS	obstructive sleep apnea syndrome
PC	principal component
PCA	principal component analysis
PLMS	periodic limb movement in sleep
PSG	polysomnography
PSG_AI	nocturnal polysomnography: arousal index
PSG_N3_TST	nocturnal polysomnography: fraction of total sleep time that is spent in sleep stage N3
PSG_REML	nocturnal polysomnography: REM latency
REM	rapid eye movement

RLS	restless legs syndrome
SD	standard deviation
SL1	MSLT: sleep latency to sleep stage N1
SL2	MSLT: sleep latency to sleep stage N2
SOREM	sleep onset REM episode
susSL	sustained sleep latency: Timespan until three consecutive periods of sleep stage N1 or at least one episode of another sleep stage is reached
TST	MSLT: Total sleep time
VigFalse	Quatember-Maly vigilance test: False reactions
VigCorr	Quatember-Maly vigilance test: Correct reaction
VigRT	Quatember-Maly vigilance test: Mean reaction time in correct reactions

References

1. Thannickal TC, Moore RY, Nienhuis R, Ramanathan L, Gulyani S, Aldrich M et al. Reduced Number of Hypocretin Neurons in Human Narcolepsy. Neuron 2000; 27(3):469–74. doi:10.1016/S0896-6273(00)00058-1.

2. Bassetti C, Aldrich MS. Idiopathic hypersomnia. A series of 42 patients. Brain 1997; 120(8):1423–35.

3. Vernet C, Arnulf I. Idiopathic hypersomnia with and without long sleep time: a controlled series of 75 patients. Sleep 2009; 32(6):753–9.

4. Pizza F, Vandi S, Detto S, Poli F, Franceschini C, Montagna P et al. Different sleep onset criteria at the multiple sleep latency test (MSLT): an additional marker to differentiate central nervous system (CNS) hypersomnias. J Sleep Res 2011 [cited 2018 Feb 3]; 20(1 Pt 2):250–6. doi:10.1111/j.1365-2869.2009.00808.x.

5. International classification of sleep disorders. 3. ed. Darien, Ill.: American Acad. of Sleep Medicine; 2014.

6. Kim H, Young T. Subjective Daytime Sleepiness: Dimensions and Correlates in the General Population. Sleep 2005; 28(5):625–34. doi:10.1093/sleep/28.5.625.

7. Martinez D, Lumertz MS, Lenz MdCS. Dimensions of sleepiness and their correlations with sleep-disordered breathing in mild sleep apnea. J. bras. pneumol. 2009; 35(6):507–14. doi:10.1590/S1806-37132009000600003.

8. McClelland LE, Pilcher JJ. Assessing subjective sleepiness during a night of sleep deprivation: examining the internal state and behavioral dimensions of sleepiness. Behav Med 2007; 33(1):17–26. doi:10.3200/BMED.33.1.17-28.

9. Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. Sleep 1991; 14(6):540–5.

10. Johns MW. Sleepiness in different situations measured by the Epworth Sleepiness Scale. Sleep 1994; 17(8):703–10.

11. Sauter C, Popp R, Danker-Hopfe H, Büttner A, Wilhelm B, Binder R et al. Normative values of the German Epworth Sleepiness Scale. Somnologie 2007; 11(4):272–8. doi:10.1007/s11818-007-0322-8.

12. Chen N-H, Johns MW, Li H-Y, Chu C-C, Liang S-C, Shu Y-H et al. Validation of a Chinese version of the Epworth sleepiness scale. Qual Life Res [cited 2018 Feb 3]; 11(8):817–21. Available from: URL: http://link.springer.com/content/pdf/10.1023%2FA%3A1020818417949.pdf. doi:10.1023/A:1020818417949.

13. Izci B, Ardic S, Firat H, Sahin A, Altinors M, Karacan I. Reliability and validity studies of the Turkish version of the Epworth Sleepiness Scale. Sleep Breath 2008; 12(2):161–8. doi:10.1007/s11325-007-0145-7.

14. Bloch KE, Schoch OD, Zhang JN, Russi EW. German Version of the Epworth Sleepiness Scale. Respiration 1999; 66(5):440–7. doi:10.1159/000029408.

15. Johns MW. Reliability and factor analysis of the Epworth Sleepiness Scale. Sleep 1992; 15(4):376–81.

16. Nguyen AT, Baltzan MA, Small D, Wolkove N, Guillon S, Palayew M. Clinical reproducibility of the Epworth Sleepiness Scale. J Clin Sleep Med 2006; 2(2):170–4.

17. Meaning of "vigilance" in the English Dictionary [cited 2018 Feb 3]. Available from: URL: https://dictionary.cambridge.org/dictionary/english/vigilance.

18. Stuck BA, Maurer JT, Schredl M, Weeß H-G. Praxis der Schlafmedizin: Schlafstörungen bei Erwachsenen und Kindern Diagnostik, Differenzialdiagnostik und Therapie: Springer-Verlag; 2013.

19. Kipman U, Fritz A. Psychologische Diagnostik von Aufmerksamkeits-und Konzentrationsfähigkeit im Kindergarten-und Schulalter: Österreichisches Zentrum für Begabtenförderung und Begabungsforschung; 2014.

20. Mackworth NH. Researches on the Measurement of Human Performance. Researches on the Measurement of Human Performance 1950; (268).

21. SCHUHFRIED - VIGIL [cited 2018 Feb 3]. Available from: URL: https://www.schuhfried.at/test/VIGIL.

22. Kushida CA, Littner MR, Morgenthaler T, Alessi CA, Bailey D, Coleman J et al. Practice Parameters for the Indications for Polysomnography and Related Procedures: An Update for 2005. Sleep 2005; 28(4):499–523. Available from: URL: https://academic.oup.com/sleep/article-pdf/28/4/499/13662731/sleep-28-4-499.pdf. doi:10.1093/sleep/28.4.499.

23. Rechtschaffen A, Kales A. A manual of standardized terminology, techniques and scoring for sleep stages in human adults.: Brain Information Service. Brain Research Institute, University of California 1968.

24. Arand D, Bonnet M, Hurwitz T, Mitler M, Rosa R, Sangal RB. The clinical use of the MSLT and MWT. Sleep 2005; 28(1):123–44.

25. Dement W, Sidoric K, Simmons J, Lentz R, Carpenter S, Laughlin E et al. Life on basic rest-activity cycle (BRAC)- Sleep Studies of a ninety minute day 1972; 9(1).

26. Carskadon MA, Dement WC. Sleep studies on a 90-minute day. Electroencephalography and Clinical Neurophysiology 1975; 39(2):145–55. doi:10.1016/0013-4694(75)90004-8.

27. Carskadon MA, Dement WC. Sleepiness and Sleep State on a 90-Min Schedule. Psychophysiology 1977; 14(2):127–33. doi:10.1111/j.1469-8986.1977.tb03362.x.

28. Carskadon MA, Dement WC. The multiple sleep latency test: what does it measure? Sleep 1982; 5 Suppl 2:S67-72.

29. Carskadon MA, Dement WC. Sleep tendency: an objective measure of sleep loss. Sleep Res 1977; 6(200):940.

30. Mitler MM, van den Hoed J, Carskadon M, Richardson G, Park R, Guilleminault C et al. REM sleep episodes during the multiple sleep latency test in narcoleptic patients. Electroencephalography and Clinical Neurophysiology 1979; 46(4):479–81.

31. Carskadon MA, Dement WC, Mitler MM, Roth T, Westbrook PR, Keenan S. Guidelines for the multiple sleep latency test (MSLT): a standard measure of sleepiness. Sleep 1986; 9(4):519–24.

32. Littner, Kushida C, Wise M, Davila DG, Morgenthaler T, Lee-Chiong T et al. Practice parameters for clinical use of the multiple sleep latency test and the maintenance of wakefulness test. Sleep 2005; 28(1):113–21.

33. Akintomide G, Rickards. Narcolepsy: A review. NDT 2011:507. doi:10.2147/NDT.S23624.

34. Geisler P, Crönlein T, Tracik F, Zulley J. The MSLT-30: an 'easy-to-use'variant of the MSLT with improved diagnostic value. J Sleep Res 1998; 7(Suppl 2):100.

35. Geisler P, Tracik F, Crönlein T, Fulda S, Wichniak A, Popp R et al. The influence of age and sex on sleep latency in the MSLT-30--a normative study. Sleep 2006; 29(5):687–92.

36. Richardson GS, Carskadon MA, Flagg W, van den Hoed J, Dement WC, Mitler MM. Excessive daytime sleepiness in man: Multiple sleep latency measurement in narcoleptic and control subjects. Electroencephalography and Clinical Neurophysiology 1978; 45(5):621–7.

37. Drake CL, Rice MF, Roehrs TA, Rosenthal L, Guido P, Roth T. Scoring reliability of the multiple sleep latency test in a clinical population. Sleep 2000; 23(7):911–3.

38. Chen L, Ho CK, Lam VK, Fong SY, Li AM, Lam SP et al. Interrater and Intrarater Reliability in Multiple Sleep Latency Test. Journal of Clinical Neurophysiology 2008; 25(4):218–21. doi:10.1097/WNP.0b013e31817f36a6.

39. Trotti LM, Staab BA, Rye DB. Test-Retest Reliability of the Multiple Sleep Latency Test in Narcolepsy without Cataplexy and Idiopathic Hypersomnia. JCSM 2013. doi:10.5664/jcsm.2922.

40. Goldbart A, Peppard P, Finn L, Ruoff CM, Barnet J, Young T et al. Narcolepsy and Predictors of Positive MSLTs in the Wisconsin Sleep Cohort. Sleep 2014. doi:10.5665/sleep.3758.

41. Aldrich MS, Chervin RD, Malow BA. Value of the multiple sleep latency test (MSLT) for the diagnosis of narcolepsy. Sleep 1997; 20(8):620–9.

42. Dauvilliers Y, Gosselin A, Paquet J, Touchon J, Billiard M, Montplaisir J. Effect of age on MSLT results in patients with narcolepsy-cataplexy. Neurology 2004; 62(1):46–50. doi:10.1212/01.WNL.0000101725.34089.1E.

43. Harrison Y, Horne JA. "High sleepability without sleepiness". The ability to fall asleep rapidly without other signs of sleepiness. Neurophysiologie Clinique/Clinical Neurophysiology 1996; 26(1):15–20. doi:10.1016/0987-7053(96)81530-9.

44. Johns MW. Sensitivity and specificity of the multiple sleep latency test (MSLT), the maintenance of wakefulness test and the Epworth sleepiness scale: Failure of the MSLT as a gold standard. J Sleep Res 2000; 9(1):5–11. doi:10.1046/j.1365-2869.2000.00177.x.

45. Mayer G, Lammers GJ. The MSLT: More objections than benefits as a diagnostic gold standard? Sleep 2014; 37(6):1027.

46. Gélineau E. De la narcolepsie: impr. Tessier frères; 1881.

47. Westphal C. Eigentümliche mit einschlafen verbundene anfälle. Arch Psychiatr Nervenkr 1877; 7:631–5.

48. Schenck CH, Bassetti CL, Arnulf I, Mignot E. English Translations Of The First Clinical Reports On Narcolepsy And Cataplexy By Westphal And Gélineau In The Late 19th Century, With Commentary. J Clin Sleep Med 2007; 3(3):301–11.

49. Vogel G. Studies in Psychophysiology of Dreams: III. The Dream of Narcolepsy. Arch Gen Psychiatry 1960; 3(4):421–8. Available from: URL: http://archpsyc.jamanetwork.com/data/journals/psych/12068/archpsyc_3_4_011.pdf. doi:10.1001/archpsyc.1960.01710040091011.

50. Juji T, Satake M, Honda Y, Doi Y. HLA antigens in Japanese patients with narcolepsy. All the patients were DR2 positive. Tissue Antigens 1984; 24(5):316–9.

51. Nishino S, Ripley B, Overeem S, Lammers GJ, Mignot E. Hypocretin (orexin) deficiency in human narcolepsy. The Lancet 2000; 355(9197):39–40. doi:10.1016/S0140-6736(99)05582-8.

52. Medicine AAoS. ICSD-2–International classification of sleep disorders, 2. Aufl. Diagnostic and coding manual. American Academy of Sleep Medicine, Westchester 2005.

53. Tashiro T, Kanbayashi T, Iijima S, Hishikawa Y. An epidemiological study on prevalence of narcolepsy in Japanese. J Sleep Res 1992; 1(Suppl):228.

54. Wilner A, Steinman L, Lavie P, Peled R, Friedmann A, Brautbar C. Narcolepsy-cataplexy in Israeli Jews is associated exclusively with the HLA DR2 haplotype. A study at the serological and genomic level. Hum Immunol 1988; 21(1):15–22.

55. Ohayon MM, Priest RG, Zulley J, Smirne S, Paiva T. Prevalence of narcolepsy symptomatology and diagnosis in the European general population. Neurology 2002; 58(12):1826–33.

56. Silber MH, Le Krahn, Olson EJ, Pankratz VS. The epidemiology of narcolepsy in Olmsted County, Minnesota: a population-based study. Sleep 2002; 25(2):197–202.

57. Dauvilliers Y, Montplaisir J, Molinari N, Carlander B, Ondze B, Besset A et al. Age at onset of narcolepsy in two large populations of patients in France and Quebec. Neurology 2001; 57(11):2029–33.

58. Han F, Lin L, Warby SC, Faraco J, Li J, Dong SX et al. Narcolepsy onset is seasonal and increased following the 2009 H1N1 pandemic in china. Ann Neurol 2011; 70(3):410–7. Available from: URL: http://onlinelibrary.wiley.com/doi/10.1002/ana.22587/full. doi:10.1002/ana.22587.

59. Dauvilliers Y, Carlander B, Molinari N, Desautels A, Okun M, Tafti M et al. Month of birth as a risk factor for narcolepsy. Sleep 2003; 26(6):663–6.

60. Dauvilliers Y, Arnulf I, Mignot E. Narcolepsy with cataplexy. The Lancet 10; 369(9560):499–511. Available from: URL: http://www.sciencedirect.com/science/article/pii/S0140673607602372. doi:10.1016/S0140-6736(07)60237-2.

61. Peyron C, Tighe DK, Pol, Anthony N. van den, Lecea Ld, Heller HC, Sutcliffe JG et al. Neurons Containing Hypocretin (Orexin) Project to Multiple Neuronal Systems. J. Neurosci. 1998; 18(23):9996–10015. Available from: URL: http://www.jneurosci.org/content/18/23/9996.full.

62. Chemelli RM, Willie JT, Sinton CM, Elmquist JK, Scammell T, Lee C et al. Narcolepsy in orexin Knockout Mice. Cell 1999; 98(4):437–51. doi:10.1016/S0092-8674(00)81973-X.

63. Lin L, Faraco J, Li R, Kadotani H, Rogers W, Lin X et al. The Sleep Disorder Canine Narcolepsy Is Caused by a Mutation in the Hypocretin (Orexin) Receptor 2 Gene. Cell 1999; 98(3):365–76. doi:10.1016/S0092-8674(00)81965-0.

64. Thannickal TC, Nienhuis R, Siegel JM. Localized loss of hypocretin (orexin) cells in narcolepsy without cataplexy. Sleep 2009; 32(8):993–8.

65. Heier MS, Evsiukova T, Vilming S, Gjerstad MD, Schrader H, Gautvik K. CSF hypocretin-1 levels and clinical profiles in narcolepsy and idiopathic CNS hypersomnia in Norway. Sleep 2007; 30(8):969–73.

66. Krahn LE, Pankratz VS, Oliver L, Boeve BF, Silber MH. Hypocretin (orexin) levels in cerebrospinal fluid of patients with narcolepsy: relationship to cataplexy and HLA DQB1* 0602 status. Sleep 2002; 25(7):733–6.

67. Nishino S, Kanbayashi T, Fujiki N, Uchino M, Ripley B, Watanabe M et al. CSF hypocretin levels in Guillain-Barre syndrome and other inflammatory neuropathies. Neurology 2003; 61(6):823–5.

68. Bourgin P, Zeitzer JM, Mignot E. CSF hypocretin-1 assessment in sleep and neurological disorders. Lancet Neurol 2008; 7(7):649–62. doi:10.1016/S1474-4422(08)70140-6.

69. Mignot E, Lammers GJ, Ripley B, Okun M, Nevsimalova S, Overeem S et al. The Role of Cerebrospinal Fluid Hypocretin Measurement in the Diagnosis of Narcolepsy and Other Hypersomnias. Arch Neurol 2002; 59(10):1553. doi:10.1001/archneur.59.10.1553.

70. Andlauer O, Moore H, Hong S-C, Dauvilliers Y, Kanbayashi T, Nishino S et al. Predictors of Hypocretin (Orexin) Deficiency in Narcolepsy Without Cataplexy. Sleep 2012. doi:10.5665/sleep.2080.

71. Picchioni D, Mignot EJ, Harsh JR. The month-of-birth pattern in narcolepsy is moderated by cataplexy severity and may be independent of HLA-DQB1* 0602. Sleep 2004; 27(8):1471–5.

72. Mignot E, Hayduk R, Black J, Grumet FC, Guilleminault C. HLA DQB1*0602 is associated with cataplexy in 509 narcoleptic patients. Sleep 1997; 20(11):1012–20.

73. Aran A, Lin L, Nevsimalova S, Plazzi G, Hong SC, Weiner K et al. Elevated anti-streptococcal antibodies in patients with recent narcolepsy onset. Sleep 2009; 32(8):979–83.

74. Koepsell TD, Longstreth WT, Ton TGN. Medical exposures in youth and the frequency of narcolepsy with cataplexy: a population-based case-control study in genetically predisposed people. J Sleep Res 2010; 19(1 Pt 1):80–6. doi:10.1111/j.1365-2869.2009.00756.x.

75. Jacob L. Narcolepsy is a sleep disorder associated with polymorphisms in genes involved in immune response 2014.

76. Mignot E, Lin L, Rogers W, Honda Y, Qiu X, Lin X et al. Complex HLA-DR and -DQ interactions confer risk of narcolepsy-cataplexy in three ethnic groups. Am J Hum Genet 2001; 68(3):686–99.

77. Parkes JD, Fenton G, Struthers G, Curzon G, Kantamaneni BD, Buxton BH et al. Narcolepsy and Cataplexy. Clinical Features, Treatment and Cerebrospinal Fluid Findings. QJM 1974;

43(4):525–36. Available from: URL: http://qjmed.oxfordjournals.org/content/qjmed/43/4/525.full.pdf.

78. Dauvilliers Y, Neidhart E, Lecendreux M, Billiard M, Tafti M. MAO-A and COMT polymorphisms and gene effects in narcolepsy. Molecular psychiatry 2001; 6(4):367–72.

79. John J, Thannickal TC, McGregor R, Ramanathan L, Ohtsu H, Nishino S et al. Greatly increased numbers of histamine cells in human narcolepsy with cataplexy. Ann Neurol 2013; 74(6):786–93. doi:10.1002/ana.23968.

80. Nishino S, Sakurai E, Nevsimalova S, Yoshida Y, Watanabe T, Yanai K et al. Decreased CSF histamine in narcolepsy with and without low CSF hypocretin-1 in comparison to healthy controls. Sleep 2009; 32(2):175.

81. Němcová V, Krásenský J, Kemlink D, Petrovický P, Vaněčková M, Seidl Z et al. Hippocampal but not amygdalar volume loss in narcolepsy with cataplexy. Neuroendocrinology Letters 2015; 36(7).

82. Hallmayer J, Faraco J, Lin L, Hesselson S, Winkelmann J, Kawashima M et al. Narcolepsy is strongly associated with the T-cell receptor alpha locus. Nature genetics 2009; 41(6):708–11.

83. Billiard M. Diagnosis of narcolepsy and idiopathic hypersomnia. An update based on the International Classification of Sleep Disorders, 2nd edition. Sleep Medicine Reviews 2007; 11(5):377–88. Available from: URL: http://www.sciencedirect.com/science/article/pii/S1087079207000500. doi:10.1016/j.smrv.2007.04.001.

84. Frauscher B, Ehrmann L, Mitterling T, Gabelia D, Gschliesser V, Brandauer E et al. Delayed diagnosis, range of severity, and multiple sleep comorbidities: A clinical and polysomnographic analysis of 100 patients of the innsbruck narcolepsy cohort. J Clin Sleep Med 2013; 9(8):805–12. doi:10.5664/jcsm.2926.

85. Green PM, Stillman MJ. Narcolepsy. Signs, symptoms, differential diagnosis, and management. Arch Fam Med 1998; 7(5):472–8.

86. Overeem S, Reijntjes R, Huyser W, Lammers GJ, van Dijk JG. Corticospinal excitability during laughter: implications for cataplexy and the comparison with REM sleep atonia. J Sleep Res 2004; 13(3):257–64. doi:10.1111/j.1365-2869.2004.00411.x.

87. Sturzenegger C, Bassetti CL. The clinical spectrum of narcolepsy with cataplexy: a reappraisal. J Sleep Res 2004; 13(4):395–406. doi:10.1111/j.1365-2869.2004.00422.x.

88. Rubboli G, d'Orsi G, Zaniboni A, Gardella E, Zamagni M, Rizzi R et al. A video-polygraphic analysis of the cataplectic attack. Clinical Neurophysiology 2000; 111:S120-S128. doi:10.1016/S1388-2457(00)00412-0.

89. Poryazova R, Siccoli M, Werth E, Bassetti CL. Unusually prolonged rebound cataplexy after withdrawal of fluoxetine. Neurology 2005; 65(6):967–8. Available from: URL: http://www.neurology.org/content/65/6/967.full. doi:10.1212/01.wnl.0000175978.61048.62.

90. Knudsen S, Jennum PJ, Alving J, Sheikh SP, Gammeltoft S. Validation of the ICSD-2 Criteria for CSF Hypocretin-1 Measurements in the Diagnosis of Narcolepsy in the Danish Population. Sleep 2010; 33(2):169–76.

91. Guilleminault C, Brooks SN. Excessive daytime sleepinessA challenge for the practising neurologist. Brain 2001; 124(8):1482–91. Available from: URL: https://academic.oup.com/brain/article-pdf/124/8/1482/1298052/1241482.pdf. doi:10.1093/brain/124.8.1482.

92. Martínez-Rodríguez JE, Iranzo A, Casamitjana R, Graus F, Santamaria J. [Comparative analysis of patients with narcolepsy-cataplexy, narcolepsy without cataplexy and idiopathic hypersomnia]. Medicina clínica 2007; 128(10):361–4.

93. Dauvilliers Y, Barateau L. Narcolepsy and Other Central Hypersomnias. Continuum (Minneap Minn) 2017; 23(4, Sleep Neurology):989–1004. doi:10.1212/CON.0000000000000492.

94. Ohayon MM, Priest RG, Caulet M, Guilleminault C. Hypnagogic and hypnopompic hallucinations: Pathological phenomena? The British Journal of Psychiatry 1996; 169(4):459–67. doi:10.1192/bjp.169.4.459.

95. Peacock J, Benca RM. Narcolepsy: Clinical features, co-morbidities & treatment. Indian Journal of Medical Research 2010; 131(2):338.

96. Bell CC, Dixie-Bell DD, Thompson B. Further studies on the prevalence of isolated sleep paralysis in black subjects. J Natl Med Assoc 1986; 78(7):649–59.

97. Zorick FJ, Salis PJ, Roth T, Kramer M. Narcolepsy and automatic behavior: a case report. J Clin Psychiatry 1979; 40(4):194–7.

98. Okun ML, Lin L, Pelin Z, Hong S, Mignot E. Clinical aspects of narcolepsy-cataplexy across ethnic groups. Sleep 2002; 25(1):27–35.

99. Nightingale S, Orgill JC, Ebrahim IO, Lacy SF de, Agrawal S, Williams AJ. The association between narcolepsy and REM behavior disorder (RBD). Sleep Med 2005; 6(3):253–8. doi:10.1016/j.sleep.2004.11.007.

100. Roth B, Nevsímalová S. Depression in narcolepsy and hypersomnia. Activitas nervosa superior 1977; 19 Suppl 2:362–3.

101. Berro LF, Tufik SB, Tufik S. A journey through narcolepsy diagnosis: From ICSD 1 to ICSD 3. Sleep Science 2014; 7(1):3–4. doi:10.1016/j.slsci.2014.07.025.

102. Mayer G, Ewert Meier K, Hephata K. Selegeline hydrochloride treatment in narcolepsy. A double-blind, placebo-controlled study. Clin Neuropharmacol 1995; 18(4):306–19.

103. Littner M, Johnson SF, McCall WV, Anderson WM, Davila D, Hartse SK et al. Practice parameters for the treatment of narcolepsy: an update for 2000. Sleep 2001; 24(4):451–66.

104. Morgenthaler TI, Kapur VK, Brown T, Swick TJ, Alessi C, Aurora RN et al. Practice Parameters for the Treatment of Narcolepsy and other Hypersomnias of Central Origin An American Academy of Sleep Medicine Report: An American Academy of Sleep Medicine Report. Sleep 2007; 30(12):1705–11.

105. Barateau L, Lopez R, Dauvilliers Y. Treatment Options for Narcolepsy. CNS Drugs 2016; 30(5):369–79. Available from: URL: http://link.springer.com/content/pdf/10.1007%2Fs40263-016-0337-4.pdf. doi:10.1007/s40263-016-0337-4.

106. Lammers GJ, Arends J, Declerck AC, Kamphuisen HAC, Scbouwink G, Troost J. Ritanserin, A 5-HT2 Receptor Blocker, as Add-on Treatment in Narcolepsy. Sleep 1991; 14(2):109–15. doi:10.1093/sleep/14.2.130.

107. Mayer G. Ritanserin improves sleep quality in narcolepsy. Pharmacopsychiatry 2003; 36(4):150–5. doi:10.1055/s-2003-41200.

108. Billiard M, Bassetti C, Dauvilliers Y, Dolenc-Groselj L, Lammers GJ, Mayer G et al. EFNS guidelines on management of narcolepsy. Eur J Neurol 2006; 13(10):1035–48. doi:10.1111/j.1468-1331.2006.01473.x.

109. Golicki D, Bala MM, Niewada M, Wierzbicka A. Modafinil for narcolepsy: systematic review and meta-analysis. Medical Science Monitor 2010; 16(8):RA177-RA186.

110. Harsh, JR, Hayduk R, Rosenberg R, Wesnes KA, Walsh JK, Arora S et al. The efficacy and safety of armodafinil as treatment for adults with excessive sleepiness associated with narcolepsy. Curr Med Res Opin 2006; 22(4):761–74. doi:10.1185/030079906X100050.

111. Gj L, J A, Ac D, Md F, G S, J T. Gammahydroxybutyrate and narcolepsy: a double-blind placebo-controlled study. Sleep 1993; 16(3):216–20.

112. US Xyrem® Multicenter Study Group. Sodium oxybate demonstrates long-term efficacy for the treatment of cataplexy in patients with narcolepsy. Sleep Med 2004; 5(2):119–23. doi:10.1016/j.sleep.2003.11.002.

113. US Xyrem® Multicenter Study Group. Further evidence supporting the use of sodium oxybate for the treatment of cataplexy: a double-blind, placebo-controlled study in 228 patients. Sleep Med 2005; 6(5):415–21. doi:10.1016/j.sleep.2005.03.010.

114. US Xyrem® Multicenter Study Group. A randomized, double blind, placebo-controlled multicenter trial comparing the effects of three doses of orally administered sodium oxybate with placebo for the treatment of narcolepsy. Sleep 2002; 25(1):42–9.

115. Black J, Houghton WC. Sodium oxybate improves excessive daytime sleepiness in narcolepsy. Sleep 2006; 29(7):939–46.

116. Parkes JD, Baraitser M, Marsden CD, Asselman P. Natural history, symptoms and treatment of the narcoleptic syndrome. Acta Neurologica Scandinavica 1975; 52(5):337–53.

117. Langdon N, Shindler J, Parkes JD, Bandak S. Fluoxetine in the Treatment Of Cataplexy. Sleep 1986; 9(2):371–3. doi:10.1093/sleep/9.2.371.

118. Frey J, Darbonne C. Fluoxetine suppresses human cataplexy A pilot study. Neurology 1994; 44(4):707. Available from: URL: http://www.neurology.org/content/44/4/707.full.pdf. doi:10.1212/WNL.44.4.707.

119. Sonka K, Kemlink D, Pretl M. Cataplexy treated with escitalopram--clinical experience. Neuro Endocrinol Lett 2006; 27(1-2):174–6.

120. Lopez R, Dauvilliers Y. Pharmacotherapy options for cataplexy. Expert Opin Pharmacother 2013; 14(7):895–903. doi:10.1517/14656566.2013.783021.

121. Calik MW. Update on the treatment of narcolepsy: clinical efficacy of pitolisant. Nat Sci Sleep 2017; 9:127–33. doi:10.2147/NSS.S103462.

122. Bogan RK, Feldman N, Emsellem HA, Rosenberg R, Lu Y, Bream G et al. Effect of oral JZP-110 (ADX-N05) treatment on wakefulness and sleepiness in adults with narcolepsy. Sleep Med 2015; 16(9):1102–8. doi:10.1016/j.sleep.2015.05.013.

123. Mieda M, Willie JT, Hara J, Sinton CM, Sakurai T, Yanagisawa M. Orexin peptides prevent cataplexy and improve wakefulness in an orexin neuron-ablated model of narcolepsy in mice. Proc Natl Acad Sci U S A 2004; 101(13):4649–54. doi:10.1073/pnas.0400590101.

124. Weinhold SL, Seeck-Hirschner M, Nowak A, Hallschmid M, Göder R, Baier PC. The effect of intranasal orexin-A (hypocretin-1) on sleep, wakefulness and attention in narcolepsy with cataplexy. Behav Brain Res 2014; 262:8–13. doi:10.1016/j.bbr.2013.12.045.

125. Rogers AE, Aldrich MS. The effect of regularly scheduled naps on sleep attacks and excessive daytime sleepiness associated with narcolepsy. Nurs Res 1993; 42(2):111–7.

126. Mullington J, Broughton R. Scheduled naps in the management of daytime sleepiness in narcolepsy-cataplexy. Sleep 1993; 16(5):444–56.

127. Aldrich MS. A comparison of three different sleep schedules for reducing daytime sleepiness in narcolepsy. Sleep 2001; 24(4):385.

128. Beusterien KM, Rogers AE, Walsleben JA, Emsellem HA, Reblando JA, Wang L et al. Health-related quality of life effects of modafinil for treatment of narcolepsy. Sleep 1999; 22(6):757–65.

129. Daniels E, King MA, Smith IE, Shneerson JM. Health-related quality of life in narcolepsy. J Sleep Res 2001; 10(1):75–81.

130. Dauvilliers Y, Paquereau J, Bastuji H, Drouot X, Weil J-S, Viot-Blanc V. Psychological health in central hypersomnias: the French Harmony study. J Neurol Neurosurg Psychiatry 2009; 80(6):636–41. doi:10.1136/jnnp.2008.161588.

131. Kales A. Narcolepsy-Cataplexy: II. Psychosocial Consequences and Associated Psychopathology. Arch Neurol 1982; 39(3):169. doi:10.1001/archneur.1982.00510150039009.

132. Roth B. Sleep drunkenness and sleep paralysis. Ceskoslovenska neurologie 1956; 19(1):48–58.

133. Roth B. Narcolepsy and hypersomnia from the aspect of physiology of sleep. Statni zdravotnické nakladatelsvi: Prague 1957:331.

134. Dement W, Rechtschaffen A, Gulevich G. The nature of the narcoleptic sleep attack. Neurology 1966; 16(1):18. Available from: URL: http://www.neurology.org/content/16/1/18.full.pdf. doi:10.1212/WNL.16.1.18.

135. Berti-Ceroni G, Coccagna G, Gambi D, Lugaresi E. Considerazioni clinico poligrafiche sull narcolessia essenziole "a somno lento". Sist Nerv 1967; 19(81):e9.

136. Passouant P, Popoviciu L, Velok G, Baldy-Moulinier M. Etude polygraphique des narcolepsies au cours du nycthémère. Rev Neurol 1968; 118:431–41.

137. Rechtschaffen A, Dement WC. Narcolepsy and hypersomnia. In: Sleep: Physiology and pathology: JB Lippincott, Philadelphia; 1969. p. 119–30.

138. Roth B, Broughton RJ, Schierlová M. Narcolepsy and hypersomnia: Karger Basel; 1980.

139. Roth B, Nevsimalova S, Rechtschaffen A. Hypersomnia with sleep drunkenness. Archives of general psychiatry 1972; 26(5):456–62.

140. Roth B. Narcolepsy and hypersomnia: Review and classification of 642 personally observed cases. Schweizer Archiv für Neurologie, Neurochirurgie und Psychiatrie 1976.

141. Billiard M, Merle C, Carlander B, Ondze B, Alvarez D, Besset A. Idiopathic hypersomnia. Psychiatry and clinical neurosciences 1998; 52(2):125–9. doi:10.1111/j.1440-1819.1998.tb00987.x.

142. Anderson KN, Pilsworth S, Sharples LD, Smith IE, Shneerson JM. Idiopathic hypersomnia: a study of 77 cases. Sleep 2007; 30(10):1274–81.

143. Billiard M, Sonka K. Idiopathic hypersomnia. Sleep Medicine Reviews 2015; 29:23–33. doi:10.1016/j.smrv.2015.08.007.

144. Ali M, Auger RR, Slocumb NL, Morgenthaler TI. Idiopathic hypersomnia: clinical features and response to treatment. Journal of clinical sleep medicine: JCSM: official publication of the American Academy of Sleep Medicine 2009; 5(6):562.

145. Kanbayashi T, Inoue Y, Chiba S, Aizawa R, Saito Y, Tsukamoto H et al. CSF hypocretin-1 (orexin-A) concentrations in narcolepsy with and without cataplexy and idiopathic hypersomnia. J Sleep Res 2002; 11(1):91–3. doi:10.1046/j.1365-2869.2002.00284.x.

146. Kanbayashi T, Kodama T, Kondo H. CSF histamine contents in narcolepsy, idiopathic hypersomnia and obstructive sleep apnea syndrome. Sleep 2009; 32(2):181–7.

147. Dauvilliers Y, Delallée N, Jaussent I, Scholz S, Bayard S, Croyal M et al. Normal cerebrospinal fluid histamine and tele-methylhistamine levels in hypersomnia conditions. Sleep 2012; 35(10):1359–66. doi:10.5665/sleep.2114.

148. Rye DB, Bliwise DL, Parker K, Trotti LM, Saini P, Fairley J et al. Modulation of vigilance in the primary hypersomnias by endogenous enhancement of GABAA receptors. Sci Transl Med 2012; 4(161):161ra151. doi:10.1126/scitranslmed.3004685.

149. Dauvilliers Y, Evangelista E, Lopez R, Barateau L, Jaussent I, Cens T et al. Absence of γ-aminobutyric acid-a receptor potentiation in central hypersomnolence disorders. Ann Neurol 2016; 80(2):259–68. doi:10.1002/ana.24710.

150. Lippert J, Halfter H, Heidbreder A, Röhr D, Gess B, Boentert M et al. Altered dynamics in the circadian oscillation of clock genes in dermal fibroblasts of patients suffering from idiopathic hypersomnia. PLoS ONE 2014; 9(1):e85255. doi:10.1371/journal.pone.0085255.

151. Tassi P, Muzet A. Sleep inertia. Sleep Medicine Reviews 2000; 4(4):341–53. doi:10.1053/smrv.2000.0098.

152. Khan Z, Trotti LM. Central Disorders of Hypersomnolence. Chest 2015; 148(1):262–73. doi:10.1378/chest.14-1304.

153. Lavault S, Dauvilliers Y, Drouot X, Leu-Semenescu S, Golmard JL, Lecendreux M et al. Benefit and risk of modafinil in idiopathic hypersomnia vs. narcolepsy with cataplexy. Sleep Med 2011; 12(6):550–6. doi:10.1016/j.sleep.2011.03.010.

154. Mayer G, Benes H, Young P, Bitterlich M, Rodenbeck A. Modafinil in the treatment of idiopathic hypersomnia without long sleep time--a randomized, double-blind, placebo-controlled study. J Sleep Res 2015; 24(1):74–81. doi:10.1111/jsr.12201.

155. Leu-Semenescu S, Louis P, Arnulf I. Benefits and risk of sodium oxybate in idiopathic hypersomnia versus narcolepsy type 1: a chart review. Sleep Med 2016; 17:38–44. doi:10.1016/j.sleep.2015.10.005.

156. Nittur N, Konofal E, Dauvilliers Y, Franco P, Leu-Semenescu S, Cock VCD et al. Mazindol in narcolepsy and idiopathic and symptomatic hypersomnia refractory to stimulants: a long-term chart review. Sleep Med 2013; 14(1):30–6. doi:10.1016/j.sleep.2012.07.008.

157. Ozaki A, Inoue Y, Nakajima T, Hayashida K, Honda M, Komada Y et al. Health-related quality of life among drug-naive patients with narcolepsy with cataplexy, narcolepsy without cataplexy, and idiopathic hypersomnia without long sleep time. Journal of clinical sleep medicine: JCSM: official publication of the American Academy of Sleep Medicine 2008; 4(6):572.

158. Chokroverty S. Sleep disturbances in other medical disorders. Sleep disorders medicine: basic science, technical considerations and clinical aspects. 2nd ed. Boston: Butterworth–Heinemann 1999:587–617.

159. Baumann CR, Mignot E, Lammers GJ, Overeem S, Arnulf I, Rye D et al. Challenges in diagnosing narcolepsy without cataplexy: a consensus statement. Sleep 2014; 37(6):1035–42. doi:10.5665/sleep.3756.

160. Vernet C, Arnulf I. Narcolepsy with long sleep time: a specific entity. Sleep 2009; 32(9):1229–35.

161. Aldrich MS. The clinical spectrum of narcolepsy and idiopathic hypersomnia. Neurology 1996; 46(2):393–401.

162. Anic-Labat S, Guilleminault C, Kraemer HC, Meehan J, Arrigoni J, Mignot E. Validation of a cataplexy questionnaire in 983 sleep-disorders patients. Sleep 1999; 22(1):77–87.

163. Šonka K, Šusta M, Billiard M. Narcolepsy with and without cataplexy, idiopathic hypersomnia with and without long sleep time: a cluster analysis. Sleep Med 2015 [cited 2015 Aug 11]; 16(2):225–31. Available from: URL: 25576137. doi:10.1016/j.sleep.2014.09.016.

164. Allen RP. When, if ever, can we use REM-onset naps on the MSLT for the diagnosis of narcolepsy? Sleep Med 2006; 7(8):657–9.

165. Mignot E, Lin L, Finn L, Lopes C, Pluff K, Sundstrom ML et al. Correlates of sleep-onset REM periods during the Multiple Sleep Latency Test in community adults. Brain 2006; 129(Pt 6):1609–23. doi:10.1093/brain/awl079.

166. Chervin RD, Aldrich MS. Sleep onset REM periods during multiple sleep latency tests in patients evaluated for sleep apnea. American journal of respiratory and critical care medicine 2000; 161(2):426–31.

167. Singh M, Drake CL, Roth T. The prevalence of multiple sleep-onset REM periods in a population-based sample. Sleep 2006; 29(7):890–5.

168. Hong S, Kim T, Joo S, Jeong J, Han J. Comparisons of clinical and polysomnographic findings between narcolepsy without cataplexy and idiopathic hypersomnia. Sleep Med 2013; 14:e153. doi:10.1016/j.sleep.2013.11.350.

169. Bozluolcay M, Nalbantoglu M, Benbir Senel G, Karadeniz D. What Does One Sleep-Onset REM Period—During Either Nocturnal Polysomnography or Multiple Sleep Latency Test—Mean in Differential Diagnosis of Central Hypersomnias? Journal of Clinical Neurophysiology 2015; 32(4):364–8. doi:10.1097/WNP.0000000000000192.

170. Takei Y, Komada Y, Namba K, Sasai T, Nakamura M, Sugiura T et al. Differences in findings of nocturnal polysomnography and multiple sleep latency test between narcolepsy and idiopathic hypersomnia. Clinical Neurophysiology 2012; 123(1):137–41. doi:10.1016/j.clinph.2011.05.024.

171. Erdem M, Bolu A, Ünlü AG, Alper M, Yetkin S. Comparison of Polysomnography and Multiple Sleep Latency Test Findings in Subjects with Narcolepsy and İdiopathic Hypersomnia. Noro Psikiyatr Ars 2013; 50(3):252–5. doi:10.4274/npa.y6477.

172. Baker TL, Guilleminault C, Nino-Murcia G, Dement WC. Comparative polysomnographic study of narcolepsy and idiopathic central nervous system hypersomnia. Sleep 1985; 9(1 Pt 2):232–42.

173. Plazzi G, Serra L, Ferri R. Nocturnal aspects of narcolepsy with cataplexy. Sleep Medicine Reviews 2008; 12(2):109–28. doi:10.1016/j.smrv.2007.08.010.

174. Mathis J, Hess CW. Sleepiness and vigilance tests. Swiss Med Wkly 2009; 139(15-16):214–9.

175. Schulz H, Wilde-Frenz J. The disturbance of cognitive processes in narcolepsy. J Sleep Res 1995; 4(1):10–4. doi:10.1111/j.1365-2869.1995.tb00144.x.

176. van Schie MK, Thijs RD, Fronczek R, Middelkoop HA, Lammers GJ, van Dijk JG. Sustained attention to response task (SART) shows impaired vigilance in a spectrum of disorders of excessive daytime sleepiness. J Sleep Res 2012; 21(4):390–5. doi:10.1111/j.1365-2869.2011.00979.x.

177. Findley LJ, Suratt PM, Dinges DF. Time-on-task decrements in "steer clear" performance of patients with sleep apnea and narcolepsy. Sleep 1999; 22(6):804–9.

178. Dougherty C. Introduction to econometrics: oxford university press, usa; 2007.

179. Koenker R. A note on studentizing a test for heteroscedasticity. Journal of Econometrics 1981; 17(1):107–12. doi:10.1016/0304-4076(81)90062-2.

180. Hayes AF, Cai L. Using heteroskedasticity-consistent standard error estimators in OLS regression: An introduction and software implementation. Behavior research methods 2007; 39(4):709–22.

181. Darlington RB, Hayes AF. Regression analysis and linear models: Concepts, application, and implementation. New York: The Guilford Press; 2017.

182. IBM SPSS Statistics 23 Command Syntax Reference [cited 2018 Feb 11]. Available from: URL:
ftp://public.dhe.ibm.com/software/analytics/spss/documentation/statistics/23.0/en/client/Manuals/IBM_SPSS_Statistics_Command_Syntax_Reference.pdf.

183. Savin NE, White KJ. The Durbin-Watson Test for Serial Correlation with Extreme Sample Sizes or Many Regressors. Econometrica 1977; 45(8):1989. doi:10.2307/1914122.

184. Tabachnick BG, Fidell LS. Using multivariate statistics. 4. ed., [Nachdr.]. Boston, Mass. [u.a.]: Allyn and Bacon; 2005.

185. IBM Corp. IBM SPSS statistics for windows, version 23.0. Armonk, NY: IBM Corp 2015.

186. Dauvilliers Y, Billiard M, Montplaisir J. Clinical aspects and pathophysiology of narcolepsy. Clinical Neurophysiology 2003; 114(11):2000–17.

187. Jolliffe I. Principal component analysis: Springer; 2002.

188. Jolliffe IT. Discarding Variables in a Principal Component Analysis. I: Artificial Data. Applied Statistics 1972; 21(2):160. doi:10.2307/2346488.

189. Tan P-N, Steinbach M, Kumar V. Introduction to data mining. 1st ed. Boston: Pearson Addison Wesley; 2005.

190. Schendera CFG. Clusteranalyse mit SPSS: Mit Faktorenanalyse: Oldenbourg Verlag; 2010.

191. MacQueen J. Some methods for classification and analysis of multivariate observations. In: : The Regents of the University of California; 1967 Available from: URL: https://projecteuclid.org/download/pdf_1/euclid.bsmsp/1200512992.

192. Lloyd S. Least squares quantization in PCM. IEEE Trans. Inform. Theory 1982; 28(2):129–37. doi:10.1109/TIT.1982.1056489.

193. Mihael Ankerst, Markus M. Breunig, Hans-peter Kriegel, Jörg Sander. OPTICS: Ordering Points To Identify the Clustering Structure; 1999.

194. Hahsler M, Piekenbrock M. dbscan: Density Based Clustering of Applications with Noise (DBSCAN) and Related Algorithms; 2017. Available from: URL: https://CRAN.R-project.org/package=dbscan.

195. Team RC. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2017. Available from: URL: https://www.R-project.org/.

196. Luxburg U von. A tutorial on spectral clustering. Stat Comput 2007; 17(4):395–416. doi:10.1007/s11222-007-9033-z.

197. Rye DB, Dihenia B, Weissman JD, Epstein CM, Bliwise DL. Presentation of narcolepsy after 40. Neurology 1998; 50(2):459–65.

Acknowledgements

First and foremost, I want to express my deep gratitude to Prof. Thomas Wetter and Dr. Roland Popp for choosing a topic for my thesis which matches both my interests and my methodical skills. Their ongoing support has been exemplary and has helped me to overcome various obstacles and pitfalls.

Furthermore, I want to thank Dr. Peter Geisler for his valuable contributions regarding the subtle difficulties in current sleep medical practice.

Special thanks go to Oliver for enriching my process of thought with immensely helpful remarks.

I want to thank my family for their unconditional support, for being my quiet haven in times of restlessness.

Finally, I am immensely grateful to Verena, who has endured all surges and drops in motivation and who has shown incredible patience with both my enthusiasm and doubts.

Curriculum vitae

Personal

Name:	Emanuel Sitka
Address:	Bruckdorfer Str. 9 93161 Sinzing
Place and date of birth:	Kelheim in Germany, 07.06.1988
Citizenship:	German
Email:	Emanuel.Sitka@medbo.de

Education

09/97 – 06/07	Donaugymnasium Kelheim	[Final grade: 1,1]
10/2010 – 12/2016	University of Regensburg State examination in medicine	[Final grade: 1,0]
10/2008 – 03/2012	Technical University of Munich / University of Regensburg Bachelor of Science in mathematics Bachelor thesis: *Stabilitätsanalyse von Infektionsmodellen*	[Final grade: 1,1]
04/2012 - 03/2014	University of Regensburg Master of Science in mathematics Master thesis: *Modeling tumor growth: A mixture model with mass exchange*	[Final grade: 1,0]

Scholarships

03/2012 – 12/2016	German Academic Scholarship Foundation (Studienstiftung des deutschen Volkes)

Career

01/2017 – 06/2017	Klinikum Fichtelgebirge, Selb resident in internal medicine

Since 09/2017	medbo: Medizinische Einrichtungen des Bezirks Oberpfalz resident in psychiatry and psychotherapy

Publications

Holzamer, A., Sitka, E., Hengstenberg, C., Schmid, C., Debl, K., Maier, L., Camboni D., Husser O., Hilker, M. (2015). Multislice computed tomography-based prediction of the implantation plane in transcatheter aortic valve implantation: determination of the line of perpendicularity and the implanter's views. *European Journal of Cardio-Thoracic Surgery*, ezv095.

Garcke, H., Lam, K. F., Sitka, E., Styles, V. (2016). A Cahn–Hilliard–Darcy model for tumour growth with chemotaxis and active transport. *Mathematical Models and Methods in Applied Sciences*, *26*(06), 1095-1148.

Garcke H., Lam K. F., Nürnberg R., Sitka E. (2018) A multiphase Cahn–Hilliard–Darcy model for tumour growth with necrosis. *Mathematical Models and Methods in Applied Sciences*, 28(03), 525–577.