

SleepData - Sleep Disorders Clinical Platform

Insomnia Population Characterization

Tiago André Ferreira Castanheira

Thesis to obtain the Master of Science Degree in

Biomedical Engineering

Supervisors: Prof. Doutor Mário Jorge Costa Gaspar da Silva

Prof. Doutor Bruno Emanuel da Graça Martins

Examination Committee

Chairperson: Prof. Doutora Ana Luísa Nobre Fred

Supervisor: Prof. Doutor Mário Jorge Costa Gaspar da Silva

Members of the Committee: Prof. Doutor Alberto Abad Gareta

October 2017

Acknowledgements

I would like to thank my supervisors Professor Mário Gaspar da Silva and Professor Bruno Graça Martins, for guiding me throughout this thesis with their helpful feedback, as well as Dr. Teresa Paiva, for letting me use her clinic, CENC, as if my second home and for her crucial insights regarding sleep medicine. I couldn't mention CENC without also thanking Cátia Reis, as she was always willing to provide a helping hand.

A special thank you to my college and partner in this project, Bárbara Pinho, for her continuous support and inspiring drive and without whom, this experience would have been much less enjoyable.

I would also like to thank my family, my brother for his encouragement and always positive attitude and my mother for always supporting me, and to mention my friends Cátia Pereira, Diogo Pais, Hugo Farinha, Mariana Fazendeiro, Nuno Pedro, Pedro Silva and Vasco Silva for helping to lighten the mood, whenever necessary, and for always being such good friends.

Finally, I want to thank my employers Nuno Homem and Pedro Sequeira for being flexible with my work schedule and for even letting me work on this thesis during business hours.

Resumo

SleepData é uma plataforma de dados de distúrbios do sono que engloba informação de actigrafia, polissonografia, anamnese e exames laboratoriais. Foi construída sobre o moderno *stack* MEAN (MongoDB, Express.js, Angular.js e Node.js) e a estrutura da base de dados segue a do HL7 FHIR, que é *open source* e inclui standards para comunicação entre plataformas. Ontologias médicas internacionais (LOINC, SNOMED-CT e as do FHIR) foram usadas para codificar cada uma das variáveis. SleepData possui interfaces para introduzir dados provenientes de cada uma das fontes mencionadas, capacidade para *upload* de ficheiros não processados e ferramentas para visualização de dados. Estas últimas incluem painéis para visualizar dados de cada doente e gráficos/estatísticas sobre sub-conjuntos da população. Usando estas ferramentas, foi feita uma caracterização da população da Insónia (N=100) – a maioria dos pacientes é mulher (64,9%) e mais velha que a população em geral. Em relação a cinco queixas cognitivas características da Insónia, 76% afirmam sofrer de mais que uma, sendo que apenas 6% não sofre de nenhuma. Para além disso, 90% é comórbida, sendo ansiedade e depressão as mais comuns (afetam 51% e 24%, respetivamente). A média do tempo total de sono (TTS), 6h, é inferior á recomendada e 54,9% tem uma opinião pessimista sobre ele, sendo o TTS reportado apenas 4h30min. O início do sono é atrasado, com 29,3% a ter latências superiores a 30 minutos e 60% com eficiências do sono deficientes (inferiores a 85%). Foram ainda calculadas as pontuações médias dos questionários PSQI, ISI e Glasgow – 12,64; 18 e 15,77, respetivamente. A análise do SCL90 revelou que somatização, comportamento obsessivo-compulsivo, depressão e ansiedade são as dimensões dos problemas psicológicos mais afetadas.

Abstract

SleepData is a sleep disorders clinical platform that encompasses data from actigraphy, polysomnography, anamnesis and lab exams. It was built over the modern software stack MEAN (MongoDB, Express.js, Angular.js and Node.js) and the database structure follows HL7 FHIR's structure, which is open source and includes standards for inter-platform communication. International medical ontologies (LOINC, SNOMED-CT and FHIR's) were employed to code each variable. SleepData has interfaces to input data from the aforementioned sources, capability to upload raw files and data visualization tools. These include dashboards to visualize patient data and graphs/stats regarding subsets of the population. Using the aforementioned tools, Insomnia population characterization was done (N=100) – most patients are female (64.9%) and older than the general population. Regarding five of the cognitive complaints characteristic of Insomnia, 76% report having more than one and just 6% don't experience any. Furthermore, 90% are comorbid, being anxiety and depression the most common (51% and 24%, respectively). The average total sleep time (TST), 6h, is lower than the recommended and 54.9% have a pessimistic opinion about it, being the average perceived TST 4h30min. Sleep onset is delayed, with 29.3% having latencies over 30 minutes and 60% having sub-par sleep efficiencies (under 85%). For the sleep questionnaires PSQI, ISI and Glasgow, average scores were calculated – 12.64, 18 and 15.77, respectively. SCL90 analysis revealed that somatization, obsessive-compulsive behavior, depression and anxiety are the most accentuated dimensions of the psychological problems.

Palavras Chave

Keywords

Palavras Chave

Plataforma de dados de distúrbios do sono

Sistemas de informação clínica

Insónia

Keywords

Sleep disorders clinical platform

Health informatics technology

Insomnia

Contents

1	Introduction	1
1.1	Objectives	2
1.2	Contributions	2
1.3	Methodology	4
1.4	Structure of the Document	5
2	Sleep and Insomnia	7
2.1	Sleep Phases	7
2.2	Main Sleep Disorders	9
2.3	Insomnia	10
2.3.1	Insomnia Types	11
2.3.1.1	Chronic Insomnia	12
2.3.1.2	Short-Term Insomnia	13
2.3.2	Risk Factors	13
2.3.3	Associated Comorbidities	14
2.4	Sleep Disorders Diagnosis Process	15
2.4.1	Actigraphy	16
2.4.2	Melatonin Exams	17
2.4.3	Polysomnography	18
2.4.4	Anamnesis	20
2.4.5	Questionnaires	22
2.5	Overview	23

3	Health Informatics Technology	25
3.1	MEAN Stack	25
3.1.1	MongoDB	26
3.1.2	Node.js	26
3.1.2.1	Express	26
3.1.2.2	Passport	27
3.2	Health Information Resources and Standards	28
3.2.1	HL7 FHIR	28
3.2.2	Health Terminologies	29
3.3	Overview	31
4	SleepData	33
4.1	Database Structure	33
4.2	Coding Systems	37
4.3	Access Control	39
4.4	Compliance with Security Regulations	40
4.5	Functionalities	41
4.5.1	Landing Page and Home Pages	41
4.5.2	Manual Data Input Interfaces	44
4.5.3	Sleep Questionnaires	45
4.5.4	Exam Upload and Processing	46
4.5.5	Clinical Data Visualization	47
4.5.5.1	Patient Data Dashboard	47
4.5.5.2	Population Statistics	49
4.6	Overview	51

5	Insomnia Population Characterization	53
5.1	Insomnia Patient Data Dashboard	53
5.2	Insomnia Population Statistics	57
5.2.1	Demographic Data	57
5.2.2	Anamnesis	58
5.2.3	Polysomnography Data	59
5.2.4	Sleep Questionnaires	61
5.2.4.1	Pittsburgh Sleep Quality Index	62
5.2.4.2	Insomnia Severity Index	62
5.2.4.3	Glasgow Sleep Effort Scale	63
5.2.4.4	Symptom Checklist-90	64
5.3	Overview	65
6	Conclusions and Future Work	67
	Bibliography	73
A	List of Parameters by Collection	75
B	Clinical Data Dashboards	81
C	Available graphs in Population Statistics	87

List of Figures

2.1	Sleep cycles on a night.	9
2.2	Melatonin blood-concentration levels during a day.	18
3.1	Observation Resource JSON from HL7's FHIR.	29
4.1	SleepData database collections.	34
4.2	Patient document (JSON), based on the FHIR's <i>Patient</i> Resource.	35
4.3	Component from a document based on the <i>Observation</i> Resource.	36
4.4	Diagram of the SleepData website structure.	42
4.5	SleepData's landing page.	43
4.6	Homepage of a CENC professional user.	44
4.7	Clinical notes form.	45
4.8	Excerpt of the SleepData's Symptom Checklist 90 (SCL-90) questionnaire.	46
4.9	Screenshot of the Delayed Sleep Phase Disorder Dashboard.	48
4.10	Color coding in the Insomnia Patient Data Dashboard.	48
4.11	General Information section of the Population's Statistics feature.	49
4.12	Pareto chart from the Clinical Notes section of the population's statistics.	50
4.13	"ISI" section of the population's statistics.	51
5.1	Clinical Notes section of the Insomnia Patient Data Dashboard.	54
5.2	PSG and PSG Report sections of the Insomnia Patient Data Dashboard.	54
5.3	Questionnaire section of the Insomnia Patient Data Dashboard.	55
5.4	Additional subsection on the Insomnia Patient Data Dashboard.	56

5.5	Age and gender distribution of the Insomnia patients.	58
5.6	Prevalence of day-time complaints and distribution of comorbid disorders.	59
5.7	Distribution of total sleep time and sleep latencies derived from PSG.	60
5.8	Distribution of the perceived total sleep times.	61
5.9	Differences between perceived TST and actual TST, measured through PSG.	61
5.10	Distribution of PSQI scores.	62
5.11	Distribution of ISI scores.	63
5.12	Distribution of Glasgow scores.	63
5.13	SCL90 questionnaire symptoms dimension scores.	65

List of Tables

2.1	Summary of the characteristics of Slow-wave and REM Sleep.	8
2.2	Prevalence of Medical Problems in persons with or without Insomnia.	15
4.1	SleepData's Compliance with the CNPD Data Security Guidelines.	41
A.1	List of the SleepData's collections parameters.	75
B.1	Insomnia dashboard parameters, variables and the respective colour code.	81
B.2	DSPD dashboard parameters, variables and the respective colour code.	83
C.1	List of graphs in Population Statistics.	87

Abbreviations

CENC Dra. Teresa Paiva – Centro de Eletroencefalografia e Neurofisiologia Clínica

DLMO Dim light melatonin onset

DSPD Delayed sleep phase disorder

ECG Electrocardiography

EEG Electroencephalography

EMG Electromyography

Epworth Epworth Sleepiness Scale

FHIR Fast Healthcare Interoperability Resources

Glasgow Glasgow Sleep Effort Scale

EHR Electronic health records

ICSD International Classification of Sleep Disorders

ISI Insomnia Severity Index

JSON JavaScript Object Notation

MCTQ Munich Chronotype Questionnaire

MEQ Morningness Eveningness Questionnaire

PSG Polysomnography

PSQI Pittsburgh Sleep Quality Index

SCL90 Symptom Checklist 90

TST Total sleep time

1 Introduction

Every clinic or hospital deals with vast amounts of patient data. To cope with these data, Electronic Health Records (or EHR for short) are getting more complex and complete. They are also growing in popularity. The patients themselves want to actively check, analyze and input their own health data. To add to that, exam and diagnostic technologies are more sophisticated than before, requiring more storage and processing capabilities. In turn this trend allow for improvement on two fronts: on the diagnosis side, it is now possible to apply new and more precise data-driven diagnosis algorithms to aid the clinician on this process; on the research side, there are many more opportunities to use these data to characterize entire populations, something that was much more difficult before the digitalization of the health sector, to derive causes for certain diseases or to determine risk behaviors, just to name a few.

With these benefits in mind, the SleepData Project, which I will cover in this dissertation, was initiated to centralize sleep data, from many sources, under one platform. This could allow for easier integration of the information from multiple diagnosis methods employed on sleep studies and sleep medicine as well as to use powerful analysis tools. A platform of this kind could solve the problems of data storage, analysis and visualization of sleep clinics everywhere and thus the motivation to make SleepData as widely applicable as possible.

SleepData was developed on a partnership with a renowned sleep clinic, "Dra. Teresa Paiva – Centro de Eletroencefalografia e Neurofisiologia Clínica", from now on referred to as CENC. This clinic is a country-wide golden standard in terms of sleep-related disorder diagnosis and treatment and, as such, makes use of multiple diagnosis and treatment equipment. CENC could benefit from the implementation of SleepData, as most of its acquired data is stored either on a paper records or spread throughout different machines and external drives (as with actigraphy and polysomnography exams or melatonin-level lab reports).

With centralized sleep diagnosis, it becomes possible to characterize sleep-related disorders across populations. To showcase this, Insomnia, one of the most common disorders of this kind, is a very interesting candidate. According to the American Academy of Sleep Medicine (2014), its chronic type affects 15-20% of the population (full clinical syndrome) and short-term

insomnia has a one-year prevalence on the 15-20% range. Not only that, but it has severe health and societal implication, as it hinders sleep and its quality. This considered, it is crucial to characterize its population, allowing for the discovery of risk behaviors, aggravating factors or triggering mechanisms. And, on the other hand, also establish guidelines to minimize insomnia incidence and its symptoms.

1.1 Objectives

The objective of this dissertation is to develop an online platform that can handle data produced by any sleep clinic, as well as self reported user data. This involves storing, parsing and handling of medical files, user interfaces for the manual input of information (as on doctor's consultations) and a way to provide access to sleep related questionnaires to anyone. The platform should work world-wide and be as general as possible by supporting multiple languages and by using informatic health standards for data manipulation, disease and diagnosis coding as well as for data storing. There must also be ways to incorporate data from other sleep clinics (even from other countries) and all the data must be quickly visualized and analyzed (both at individual and entire population levels) through a dashboard of medical data. Finally, security and privacy must be assured by preventing informatic attacks and by ensuring that individual patient data is only available to the physician.

In brief, the objectives set for this dissertation were:

- Create SleepData, an online platform for storage, handling, analysis and visualization of sleep data.
- Characterize insomnia based on data in SleepData – developing a dashboard with the most relevant parameters to study the disorder and extract metrics from a sample population, using only the included data visualization tools.

1.2 Contributions

With this thesis, and in line with the motivations and established objectives, SleepData was created and is now a platform ready to receive data from sleep clinics. It allows for the import and storage of medical exam files, as well as manual input of information. With SleepData, patient data can be quickly observed through a clinical dashboard that highlights critical parameters

(for instances values over or under the recommended or triggering factors) and population-wide metrics for a specific sub-set of the population can also be observed. Furthermore, regular users (anyone can create an account) can fill sleep-related questionnaires directly on the platform. These include the Epworth Sleepiness Scale, Munich Chronotype Questionnaire, Pittsburgh Sleep Quality Index, Morningness Eveningness Questionnaire, Glasgow Sleep Effort Scale, Insomnia Severity Index as well as the Symptom Checklist 90. For the first time those questionnaires are all available online, in Portuguese, under a central hub.

The platform is also general as it uses international standards to code the database structure as well as the individual findings, observations and exams.

SleepData is now implemented within CENC and all its data can be centralized for analysis. Data of exams, questionnaires, clinical notes and PSG has been loaded, regarding 100 insomnia patients and over 120 delayed sleep phase disorder patients.

With this work, using only the platform’s available tools, it was possible to characterize the Insomnia population and extract key statistics, namely:

- Most patients are female (64.9%) and older, on average, than the general population (85.4% are over 40 years old).
- Regarding five of the cognitive complaints characteristic of insomnia (reduced attention, reduced executive functions, memory deficit and increased fatigue and irritability), 76% of the patients report having more than one. One of the most severe, reduced executive functions, affects up to 26%. Only 6% don’t experience any of these.
- The vast majority are comorbid (90%), being anxiety and depression the most common, affecting 51% and 24% of the patients, respectively.
- The average total sleep time, about 6 hours, is lower than the recommended value, 7 to 8 hours. 15.2% slept less than 5 hours. Moreover, most patients (54.9%) have a pessimistic opinion about their sleep – only 4.4% are optimistic. The average perceived total sleep time is 4h30min, much lower than the actual value.
- Sleep is short and sleep onset delayed, with 29.3% having sleep latencies over 30 minutes. This reflects on 60% of them having sub-par sleep efficiencies (under 85%).
- Average scores for PSQI, ISI and Glasgow were computed – 12.64, 18 and 15.77, respectively. SCL90 scores were also analyzed, revealing that somatization, obsessive-compulsive

behaviour, depression and anxiety are the most accentuated dimensions of the psychological problems (when compared to a healthy population).

1.3 Methodology

The development of SleepData is the result of a collaborative effort with Pinho (2017), who developed and used the platform to study another disorder, delayed sleep phase disorder. To understand the SleepData requirements, CENC allowed us to access their facilities and diagnosis equipment, interview their workers and study the usual workflow on a sleep clinic. The main tasks I conducted to develop SleepData included:

1. Review of the literature on sleep disorders, covering their main characteristics, impacts, diagnosis methods and equipment.
2. Analysis of the workflow of a sleep clinic (with CENC as reference) with hands-on observation of the different steps associated with diagnosis and treatment at sleep clinics. This included the determination of the different data sources throughout the diagnostic/treatment process, for each patient, and assessment of how each one is stored now and how it would be on the platform.
3. Characterization of the technology used at CENC. Namely machines and diagnostic methods employed. For each, determination of the output files' types as well as handling/pre-processing needs.
4. Assessment of the best software stack to base the development of the platform, including database, server and programming languages.
5. Identification of the security and legal requirements necessary for a operating platform of this kind.
6. Design of the database using health standard resources to ensure interoperability.
7. Development of the platform with three user profiles – an administrator profile, capable of assigning permissions to the different users, a clinician profile that can access patient data as well as input data and a patient profile that can access only some of their own data and some of the platform statistics.

8. Creation of user interfaces for medical data input as well as sleep questionnaires, with continuous validation with doctors and sleep technicians to ensure that all their needs are met.
9. Creation of a clinical dashboard that allows not only for a patient data summary and visualization as well as population-wide characterization, categorized by disorder. This proceeded with iterative improvements with the feedback from doctors and researchers at CENC.
10. Characterization of a Insomnia population using only the platform's resources. Revealing common characteristics of the population, highlighting relevant risk factors and diagnostic cues.

The study of the CENC's workflow involved the observation of parts of the clinical process, for some patients, and conversations with the different technicians that intervene in the process. With their needs in mind, all the database structure was defined and parameters set and mapped to the most appropriate health informatic resources.

1.4 Structure of the Document

This dissertation follows the following structure: the next chapter, Chapter 2, introduces sleep related concepts, including Insomnia characterization aspects, and then details how sleep clinics operate, with emphasis on how diagnosis is conducted, with what equipment and resources; Chapter 3 details the health informatics technology that supports this work, both the informatic tools used to develop the platform (server, file handling, security and user interfaces) as well as health data standards and coding systems; Chapter 4 presents the platform developed, including structure, security and features; Chapter 5 describes the insomnia study that used the tools available on the platform to characterize the Insomnia population; Lastly, Chapter 6 summarizes the key contributions of this work and discusses the possibilities of how it will grow (with further usage and development) in the future.

Sleep and Insomnia



Sleep is one of the main bodily functions, being essential to the maintenance of mood, memory and cognitive performance (Phillips and Gelula, 2006). Because of its undisputed importance its study, by philosophers and later by scientists, goes back to Antiquity. However, it was only in the 19th century that Neurologists embraced this field. In the early 20th century, EEG was invented, allowing a deeper understanding of the brain activity during sleep versus its wakefulness state (Paiva and Anderson, 2014a). This changed the paradigm of sleep study from the analysis of sleep as dream analysis (psychology) to the neurologists' perspective of trying to understand the physiological underlying mechanisms (Palagini and Rosenlicht, 2011). Nowadays the sleep, as field of neurology, employs many medical diagnosis equipment and techniques in addition to EEGs, including EMGs, actigraphies or ECGs to properly characterize the sleep phases and its mechanisms.

This chapter introduces some of the key aspects of sleep; later, I'll address some of its disorders, with major focus on insomnia and lastly the sleep clinics' methods of diagnosis.

2.1 Sleep Phases

During each night, a person goes through two alternate stages of sleep. These are the so called rapid eye movement (REM) sleep and non-rapid eye movement (NREM) sleep. The deepest part of NREM sleep is usually called slow-wave sleep. This name is due to the fact that during this phase the brain waves are very intense but with very low frequency. This phase happens first, on the first hour of sleep, and is responsible for the deep and restful sleep.

As the name implies, during REM sleep the eyes tend to move rapidly, even though the person is asleep. REM sleep usually happens for short periods of time (5-30 minutes), about every 90 minutes and it amounts, in total, to about a quarter of the total sleep time. In this phase the sleep restfulness characteristics aren't so present and is when most vivid dreams happen (Guyton and Hall, 2006). Table 2.1 summarizes some of the characteristics of the REM and Slow-wave sleep phases.

Table 2.1: Summary of the characteristics of Slow-wave and REM Sleep based on Guyton and Hall (2006).

Slow-wave Sleep	REM Sleep
Very restful	Not so restful
Decreases in blood pressure, respiratory and basal metabolic rate (10 to 30 %)	Irregular heart and respiratory rates
No recalling of dreams. No bodily muscle activity	Possibility of active dreams associated with bodily muscle movements
Total duration comprehends most of the night	Comprehends about a quarter of the total night

Since 2007, the American Academy of Sleep Medicine (AASM) defines three different phases within the NREM sleep. These are the N1, N2 and N3 phases. According to Schulz (2008) this changes from the previous N1 through N4 as phases 3 and 4 were combined.

The N1 phase of NREM (or 1-NREM) is a transition period between the wake state and the sleep. As such, the sleep is still light and the person can be easily awake. The EEG exam shows the disappearing of alfa waves (7 to 14Hz range) and the surging of theta waves (4 to 7Hz range). Both can appear simultaneously and alternating. Some slow eye movement can be detected as well with an electrooculogram (Paiva and Silva, 2014).

On the N2 phase the theta waves are predominant, with moderate amplitude, and with characteristic intrusions of activity on the 12 to 14 Hz range, the so-called K complexes (Paiva and Penzel, 2011). Cardiac and respiratory frequencies are lower (average value) than during the wake state.

Lastly, the N3 phase relates to the slow-wave sleep and its the deepest part of the sleep. Delta waves (0.5 to 2 Hz) should cover more than 20% of each sleep period. As mentioned before, there is a reduction of average cardiac frequency as well as lower arterial pressure. The core temperature decreases slightly. Oxygen consumption by the brain is also reduced.

On the other hand, REM sleep is characterized by a very low-amplitude EEG record, with no predominant frequency range.

REM and NREM (N1, N2 and N3) phases happen several times during a night. Some micro-awakenings are also common. Figure 2.1, shows an example of a night of sleep and the rotations between the different sleep phases.

When we first fall asleep we do a full NREM phase before the REM Phase. After this, REM and NREM cycles follow. As the night progresses and the body is more rested the amount of deep sleep (N3) decreases and the lighter sleep increases, according to Guyton and Hall (2006).

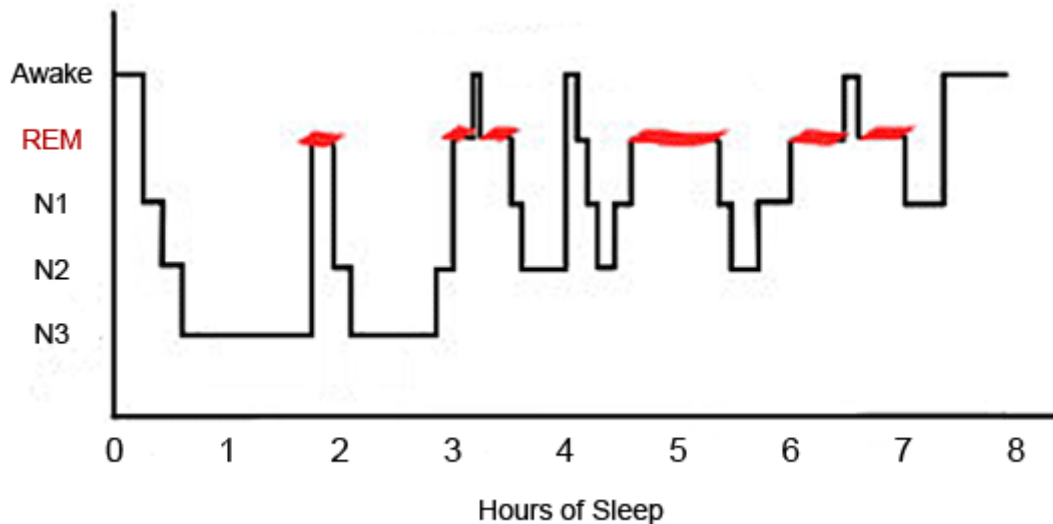


Figure 2.1: Example of the sleep cycles on an night (Copied from the website page "The Stages of Sleep: NREM Sleep and REM Sleep" – <https://sleephabits.net/stages-of-sleep>)

2.2 Main Sleep Disorders

Sleep is a very complex process of which we still don't know much about and the sleeping brain remains one of the biggest puzzles to neuroscientists. According to Chokroverty (2014), this is due to the complex interactions between the sympathetic, parasympathetic and somatic nervous systems that lead to distinct physiological changes on each person. The intricateness and diversity of the systems that surround the sleep process allow for many openings for physiological malfunctions and thus leading to the possibility of many sleep related disorders.

Sleep disorders can have many adverse consequences not only to the quality of life of the patient as well as related expenses due to loss of productivity, medical treatments, sick leave and many others (Phillips and Gelula, 2006). As such it is crucial to better understand these disorders to both diagnose them quickly and effectively as well as to develop better treatments.

The *International Classification of Sleep Disorders* (ICSD3), proposed by the American Academy of Sleep Medicine (2014) categorizes them on seven major groups.

1. Insomnia
2. Sleep Related Breathing Disorders
3. Central Disorders of Hypersomnolence
4. Circadian Rhythm Sleep-Wake Disorders

5. Parasomnias
6. Sleep Related Movement Disorders
7. Other Sleep Disorders

Insomnia includes its sub types, Chronic Insomnia, Short-Term Insomnia and Other Insomnia. Sleep Related Breathing Disorders encompass obstructive sleep apneas, central sleep apneas, hyperventilating disorders and sleep related hypoxemia. Central Disorders of Hypersomnolence are mainly narcolepsies, idiopathic hypersomnia, due to medical disorders, medication or associated with psychiatric disorders and Kleine-Levin Syndrome. Circadian Rhythm Sleep-Wake Disorders include delayed sleep-wake phase disorder, advanced sleep-wake phase disorder, irregular sleep-wake rhythm disorder, non 24-hour sleep-wake rhythm disorder, shift work disorder, jet lag disorder and circadian sleep-wake disorder not Otherwise specified. Broadly parasomnias can be NREM-related, REM-related or others. Finally, sleep related movement disorders comprise the restless legs syndrome, periodic limb movement disorder, among others. The "Other" classification tries to encompass all the sleep disorders that don't follow under any other category. This may be due to the fact that it overlaps multiple categories or that there is just not enough data to determine other diagnosis.

As it became apparent from this introduction to the categories of sleep disorders, there are copious amounts of distinct disorders. Even though the platform developed within the scope of this work was designed to handle information to characterize all of them (to more or less extent) the focus of my analysis, starting from the next section, is Insomnia, with all its many different causes, symptoms and categories.

2.3 Insomnia

According to American Academy of Sleep Medicine (2014), insomnia is defined as *a persistent difficulty with sleep initiation, duration, consolidation, or quality that occurs despite adequate opportunity and circumstances for sleep, and results in some form of daytime impairment*. If no daytime impairments are present it is not considered insomnia.

Adults with insomnia usually report having difficulties initiating and maintaining sleep which leads to lengthy periods of wakefulness during the night, reduction of total sleep time and a feeling of overall poor sleep quality. Children that suffer from this disorder may strongly resist going to bed at night and struggle to sleep independently.

During the day, symptoms include decreased mood and cognitive capacity, fatigue and irritability. On chronic patients this can, of course, lead to reduced quality of life as it interferes with the social life and working capabilities. In children, it can lead to poor school performance. More severe forms of insomnia can lead to an increased risk of car/work accidents and psychiatric and cardiovascular disorders.

Besides causing so many impairments on everyday life, one of the factors that make the insomnia so relevant is its enormous prevalence. According to Shneerson (2005) it is, in fact, the most common sleep disorder, as most adults suffer from it at some point in their lives. Multiple studies show just how common it is. Morin et al. (2011b), a study of 2000 18-year olds in Canada showed that 13.4% had Insomnia with up to 40.2% showing at least one of the symptoms. Not only is the prevalence high already but there are indications that it is rising (in developed countries). A 10-year (from 2000 to 2010) insomnia prevalence study, Pallesen et al. (2014), in Norway, shown its relative increase of more than 15%, from 13.1% to 15.2%, over this period. Ohayon and Paiva (2005), a study over 1858 Portuguese indicates that our country's situation is similar, with results that don't stray away from these – 12.1% of them shown difficulty of initiating sleep, 21.0% had difficulty maintaining sleep and 9.8% found their sleep not totally restorative, with 28.1% of the total sample having one of these Insomnia symptoms at least 3 nights per week.

2.3.1 Insomnia Types

On the brief introduction to sleep disorders, above, I mentioned that insomnia included some sub types, such as Chronic Insomnia, Short-Term Insomnia and Other Insomnia (American Academy of Sleep Medicine, 2014). Different classifications were used prior to this edition – it is still common to hear about primary or secondary insomnia. Secondary insomnia indicates to the fact that it was caused by another *primary* disorder (being it either of psychiatric or medical nature or even from substance abuse). As symptoms and features from both previous types overlapped and sometimes treating the first primary disorder didn't address insomnia that classification was dropped.

The 2nd Edition of the ICSD, from the American Academy of Sleep Medicine (2005) then sub-categorized the so called primary insomnia into psycho-physiological insomnia, idiopathic insomnia, inadequate sleep hygiene and paradoxal insomnias. Once again, these categories were deprecated since barely any real-life patients met the diagnosis criteria for just one sub type.

And thus, the currently accepted diagnostic categories for insomnia are: chronic insomnia

disorder, short-term insomnia disorder and other insomnia disorder (ICSD3), much broader (and defensive) categorization system.

2.3.1.1 Chronic Insomnia

Chronic Insomnia is characterized by complaints of sleep initialization and maintenance (with associated daytime impairments) that last for a period that is bigger than a specific threshold, with a certain minimum weekly frequency, and is associated with clinical morbidity outcomes. The detailed criteria, copied from the ICSD3 (American Academy of Sleep Medicine, 2014), are on the following list (Criteria A through F must be met):

A - The patient reports, or the patient's parent or caregiver observes, one or more of the following:

- 1 - Difficulty initiating sleep
- 2 - Difficulty maintaining sleep
- 3 - Waking up earlier than desired
- 4 - Resistance to going to bed on appropriate schedule
- 5 - Difficulty sleeping without parent or caregiver intervention

B - The patient reports, or the patient's parent or caregiver observes, one or more of the following related to the nighttime sleep difficulty:

- 1 - Fatigue/malaise
- 2 - Attention, concentration or memory impairment
- 3 - Impaired social, family, occupational or academic performance
- 4 - Mood disturbance/irritability
- 5 - Daytime sleepiness
- 6 - Behavioral problems (e.g. hyperactivity, impulsivity, aggression)
- 7 - Reduced motivation/energy/initiative
- 8 - Proneness for errors/accidents
- 9 - Concerns about or dissatisfaction with sleep

C - The reported sleep/wake complaints cannot be explained purely by inadequate opportunity (i.e. enough time is allotted for sleep) or inadequate circumstances (i.e. the environment is safe, dark, quiet and comfortable) for sleep.

- D - The sleep disturbance and associated daytime symptoms occur at least three times per week
- E - The sleep disturbance and associated daytime symptoms have been present for at least 3 months
- F - The sleep/wake difficulty is not explained more clearly by another sleep disorder

Going through the list, the main criteria for diagnosis are made clear – continuous and persistent difficulties starting and maintaining sleep that leads to poor sleep quality and has repercussions on the day time activities.

The amount/type of sleep-related complaints that can be considered enough for the diagnosis of insomnia are, of course, very subjective. Nonetheless, the ICSD3 defines some guidelines to help defining this threshold – periods of sleep wakefulness over 20 minutes (after sleep onset) for children and young adults and over 30 for adults and elders and termination of sleep more than 30 minutes before the desired.

As mentioned before, the primary and secondary classifications for insomnia (and their sub categories) are currently considered inadequate. However there is another "pseudo category" that is quite useful even though evidence to support its relevance is limited – *Insomnia with objectively short sleep duration*, i.e. less than 6 hours per night.

According to the American Academy of Sleep Medicine (2014) the prevalence of chronic insomnia (as described here) is about 10%.

2.3.1.2 Short-Term Insomnia

The other type of insomnia, is the short-term insomnia. As with chronic insomnia, it is characterized by complaints of difficulties with sleep initialization and maintenance that lead to daytime impairments. Unlike the previous its duration is under 3 months and/or it does not occur over 3 times per week. This type of insomnia is usually easily related to a triggering event.

According to the ICSD3 the one-year prevalence of short-term insomnia is on the 15% to 20% range.

2.3.2 Risk Factors

A risk factor is a variable that is associated with increased prevalence of insomnia (in this case). Following the structure of the Principles and Practices of Sleep Medicine, Kryger et al.

(2011), we can sub-categorize these as Static Risk Factors and Modifiable Risk Factors.

The main static risk factors are gender, age, ethnicity and genetics. Generally speaking, women are more likely to report insomnia than man. One of the explanations for this focus on the characteristic fluctuating hormone levels (e.g. melatonin and cortisol) that interfere with sleep onset and offset. According to Paiva and Pinto (2014), other aspects are the pregnancies, which lead to awakenings during the night and the reduced quality of sleep inherent with having a baby child. In terms of age as a risk factor, insomnia prevalence increases linearly with age, to a maximum of 50% on elders. However, if controlled for important comorbidities, age is no longer a significant risk. In terms of ethnicity, insomnia is more severe in African Americans, more prevalent on Caucasian elders (vs African) but less prevalent on middle-aged people of this ethnicity. In terms of genetics and heredity it is reasonable to assume that there is a genetic risk for developing insomnia but still little is known and the data existent so far is of low quality (Kryger et al., 2011).

The modifiable risk factors are hyper arousal, stress and life style, transient insomnia and comorbidities (medical and psychiatric). These refer to characteristics of life that are, or can be, temporary. Stressful life, unemployment, grief, accidents, etc. have obviously a significant impact on the quality of sleep and are a risk for insomnia. A transient insomnia also greatly increases the risk for developing chronic insomnia. Comorbidities are also very common with insomnia, being both a risk factor for insomnia and being insomnia a risk factor for some disorders. A more in-depth analysis of this is next, on subsection 2.3.3.

2.3.3 Associated Comorbidities

Alongside insomnia there are numerous times other disorders (medical or psychiatric). This is called comorbidity (comorbid insomnia). However, it is important to remember the difficulty of distinguishing between insomnia as a symptom of another disorder or insomnia as its own, separate disorder, since often the first type originates the second.

A study by Ford and Kamerow (1989) involving 7954 people revealed that 40% of insomnia patients had a psychiatric disorder (versus 16.4% on the group without sleep complaints). Sarsour et al. (2010) calculated the odds ratio of having at least one psychiatric diagnosis for severe insomnia (5.04) and moderate insomnia (2.63) when compared to those with no insomnia.

From the psychiatric problems co-existing with insomnia one of the most common is depression. A study by Soehner et al. (2014) focused on the analysis of a population with a major depressive episode (MDE) on the past year, according to the Diagnostic and Statistic Manual

of Mental Disorders (DSM), or a major depressive disorder (MDD) according to the National Comorbidity Survey, USA (NCS-R). The results showed that from the MDE 59.1% had insomnia only and 27.7% had insomnia and hypersomnia symptoms and from the MDD group 57.4% had insomnia only and 25.6% insomnia and hypersomnia symptoms.

Budhiraja et al. (2011), in a study with 3282 people from Detroit, determined the odds ratio of having insomnia and particular disorders (versus not having the disorders) – OR = 1.6 for heart disease, OR = 1.5 for hypertension, 1.4 for diabetes, 2.1 for stomach ulcers, 1.8 for arthritis, 1.8 for migraine, 1.6 for asthma, 1.9 for COPD, 2.0 for neurological problems and 1.7 for menstrual problems.

The book Principles and Practices of Sleep Medicine, by Kryger et al. (2011), has a great summary of the prevalence of other medical problems in persons with and without insomnia (Table 2.2 presents an adapted version).

Table 2.2: Prevalence of Medical Problems in persons with or without Insomnia, based on Kryger et al. (2011)

Medical Problem	Prevalence of Medical Problem (%)	
	People with Insomnia	People without Insomnia
Heart Disease	21.9	9.5
Cancer	8.8	4.2
Hypertension	43.1	18.7
Neurologic disease	7.3	1.2
Breathing problems	24.8	5.7
Urinary problems	19.7	9.5
Diabetes	13.1	5.0
Chronic pain	50.4	18.2
Gastrointestinal	33.6	9.2
Any medical problem	86.1	48.4

This reveals that almost every person with insomnia has some type of medical problem (86.1%), with some specific problems having a huge relative increase in prevalence from one population to the other. Breathing problems increase (relatively) by 335% from the regular population to the insomnia one. Neurological problems see an increase of 508% and gastrointestinal ones 265%.

2.4 Sleep Disorders Diagnosis Process

To properly understand the requirements of storing and processing of data and handling of files on a sleep clinic it is crucial to firstly understand the fundamental principles of sleep

disorders' diagnosis. Due to the number of different disorders and factors that need to be checked for a proper diagnosis, multiple diagnosis methods and equipment are employed. The following subsections address each type of method including clinical relevance of the gathered data, equipment used, types of files and their analysis' methods as well as crucial parameters for insomnia characterization that can be derived from each one.

2.4.1 Actigraphy

Actigraphy is a non-invasive method of monitoring temperature, movement and light exposure. The actigraph unit is a small watch-like gadget that is strapped to the wrist of the patient and records data for a few days/weeks at the time, being posteriorly connected to a hub and the data exported. The sensors it may include vary from brand to brand but some common setups are: a piezoelectric accelerometer, two temperature meters (one for the skin and other for the ambient/air) and a RGB light sensor. A considerable part of the unit's housing is due to the battery, that can usually last for upwards of a month, and the waterproofing (most actigraphs can be submerged as they comply with the IP67 standard).

CENC (Dra. Teresa Paiva – Centro de Eletroencefalografia e Neurofisiologia Clínica) has multiple actigraph units, from both Condor and Philips. Condor's equipments are of the ActTrust¹ model and the Philips ones are Actiwatch 2². Each equipment has a raw output has a time-series of values for activity, light (red, green and blue components) and temperatures as well as some key statistics about the equipment (ID, battery voltage, memory used, etc.) and patient (name, gender and description). ActTrust exports this data as a text file and Actiwatch 2 as a *Microsoft Office Excel* (xlsx) file. Regarding their size, each actigraphy exam is only a few hundred kilobytes.

The brand-specific analysis software compiles a report from this data, with some key information. In fact, the raw data is ever barely used since it is not processed and the values are not directly comparable. Algorithms that detect awakenings, sleep onset and other events are not disclosed and vary from manufacturer to manufacturer. The reports made by Condor's ActStudio are Portable Document Format (pdf) files and *Microsoft Office Word* (docx) files for the Philips' Actiware Software.

Besides patient information (age, name, gender) and recording period information, the report includes recording time summary statistics, day-by-day summaries, time-series with light,

¹Condor ActTrust - <http://www.condorinst.com.br/en/acttrust/>

²Philips Actiwatch 2 - <http://www.actigraphy.com/solutions/actiwatch/actiwatch2.html>

activity and temperatures as well as hardware information.

The summary statistics refer to minimums, maximums and averages of different parameters, during the entirety of the recording time. These offer a good indication of the quality of sleep and sleep profile of the patient at a glance. The information in this category includes bed times, get up times, time in bed, total sleep time, onset latency (difference between bed time and beginning of sleep), WASO (wake time after sleep onset) and number of awakenings in a night. This same information is afterwards displayed on a day by day basis, instead of just the global minimum, maximum e average.

The accuracy of this information is sometimes low since the algorithms that estimate, for example, sleep onset (based on motion and temperature), can be flawed. According to Hauri and Wisbey (1992), the inaccuracy is increased, as expected, on patients suffering from movement disorders while asleep. While usually no diagnosis is made just from this exam it is still very useful to monitor differences between groups, sleep pattern variations over time or the effects of behavioral or treatment interventions (Lee-Chiong, 2006). The inaccuracy can also be mitigated by long recording periods, visual inspection of the data and by comparison of the patient's sleep diary (if applicable).

In terms of Insomnia diagnosis, the actigraphy can provide some important insight. Insufficient light exposure during the morning can be a risk factor, as well as excessive artificial light during the night and right before going to sleep. The actigraph can distinguish between natural and artificial light since the latter has more B (blue) component of the RGB light. Another valuable information is the number and duration of naps during the day that may lead to increased sleep onset latency. Sometimes, insomnia complaints can be ignored or considered psychological simply by comparing the description of the nights (by the patient) to the actigraphy record – a perceived sleep onset of hours could be, in reality, just a few minutes. Exaggerated perception of the difficulties to initiate sleep is common within insomnia patients.

2.4.2 Melatonin Exams

Melatonin is an hormone produced by the pineal gland, a small endocrine gland located in the epithalamus, near the center of the brain, and between the two cerebral hemispheres. During the day melatonin production by the pineal gland is halted but during the night, after sunset, it begins, triggered by the suprachiasmatic nucleus (region of the brain in the hypothalamus, situated directly above the optic chiasm and responsible for controlling circadian rhythms). This process usually occurs at around 9 p.m. and the blood melatonin levels stay high for about 12

hours. Daytime levels of melatonin are barely detectable¹.

The melatonin exam takes place on a dim room where saliva is collected from the patient every hour, for a few hours (usually from 8 p.m. to 3 a.m.). The results are compiled on a written report where the most important data is the pairs melatonin level/time at measurement.

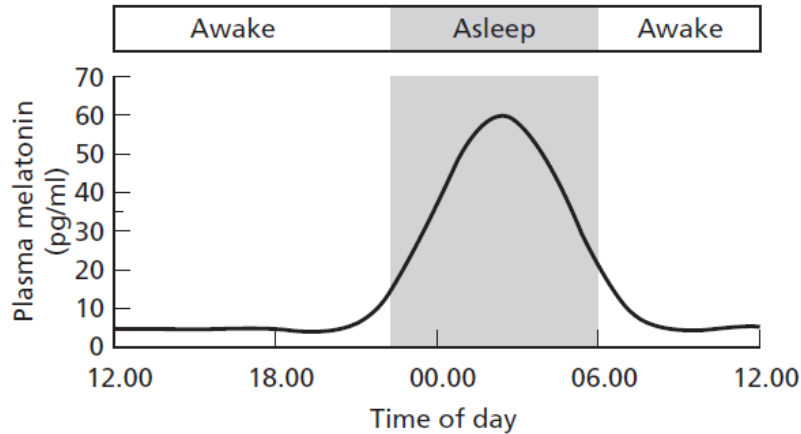


Figure 2.2: Melatonin blood-concentration levels during a day. Adapted from Shneerson (2005).

With these values, DLMO (dim light melatonin onset) can be estimated, i.e. the process of increased melatonin production after darkness and before sleep (see Figure 2.2). This is crucial to access circadian rhythm disorders with which melatonin production doesn't occur at normal hours. One example is delayed sleep-wake phase disorder (DSPD), a chronic deregulation of the circadian rhythm where sleep patterns are shifted, namely sleep onset, rendering the patient unable to fall asleep early (e.g. before 3 a.m.).

2.4.3 Polysomnography

Polysomnography (PSG) is a combination of recordings of biophysical parameters during the sleep and is one of the most complete (and complex) sleep-related exams there are. According to Paiva and Pinto (2014), it is the golden standard for objective evaluation of insomnia. PSG encompass Electroencephalography (EEG), Electromyography (EMG), Electrocardiography (ECG) and Oximetry measures. CENC's PSG setup also include video recording.

In terms of equipment, PSG consists of a central hub, with an amplifier, to which is connected every lead coming from the electrodes (that are part of the setup of the EEG, EMG and ECG). The lead configurations can be adapted, depending on which type of recording the doctor intends

¹Melatonin and Sleep article from the National Sleep Foundation – <https://sleepfoundation.org/sleep-topics/melatonin-and-sleep>

and on what disorder is being studied. As mentioned before, CENC's PSG setup also includes cameras, which are synchronized with the other exams and allow to investigate the physiological (electrical) signals resulting from visually detectable events, such as periodic leg movements, myoclonus or awakenings.

At CENC there are five different types of PSGs, from distinct brands or models. These are – Embla and Nicolet from Natus Medical Incorporated, SOMNOscreen from SOMNOmedics and Alice 5 from Philips. There are also a few Emblettas (portable units), also from Natus. The SOMNOmedics' PSG is also referred to as Domino (same name as its analysis software).

Each of these has a very different output, with varying file types, number and sizes. The following list summarizes the outputs for each PSG unit:

Alice has three main components. The first one are the time-series files, of unspecified type, and are a few hundred per patient. The second part contains the montage information (on a couple of proprietary type files) and the options used within the program as XML files. The last part are disk image files with the video recordings.

Domino (SOMNOscreen) time series are stored as a text-based DAT file and there are a few dozens of dnl type files that store the acquisition settings like FFT (fast fourier transform) parameters, filter parameters, user data and others.

Embla is composed of one EBM file is created per measure (for example activity on each electrode, heart rate, heart activity variability, O2 pressure and many others). Other than these, there is a WMV File with the video recording.

Nicolet time-series are all stored under one ".e" file with a AVI file for the video recording.

Regarding their size, each PSG exam, with its numerous files, can be of up to a few gigabytes, mainly due to the additional video recording. As it is clear from the list above, each manufacturer has its unique style of exports (there are even differences within the same brand, from model to model) and the file types, apart from the video recordings, are not easily opened and its information analyzed. Each PSG's data can only be visualized using the corresponding proprietary software.

From each polysomnography exam, a report (word file) is written by the sleep technicians and validated by the doctor. These reports compile information of three types:

- Objective data, directly observed with the included PSG software.

- Subjective data, derived from the objective data.
- Hand-written notes (by the technician) about the patient's over-night stay.

The objective data encompasses parameters like sleep onset time, wake up time, sleep latency (and with those total sleep time and sleep efficiency are calculated), the percentage of the sleep time spent on N1, N2 and N3, number of sleep cycles, Apnea-hypopnea index, Oxygen desaturation index, Minimum oxygen saturation, duration of the snoring period, periodic sleep movements and micro awakenings. By examining this information and the time-series directly some subjective information is derived like the existence of abnormalities on the deep sleep or REM time profiles, analysis of alpha or beta wave intrusions (EEG) and period limb movement disorders. From the night diaries, some of the most important information are the time at lights on and off, the perceived sleep duration (helps to determine the accuracy of the perceived total sleep time) and quality, number of trips to the bathroom and the number of vocalizations. On CENC sometimes the SIT (Suggested Immobilization Test) is also performed and its result written here.

It is evident that PSG is a very vast source of both numerical objective data as well as subjective information. The patient's perception of sleep duration can be confronted with the actual duration (calculated with the PSG). An overly pessimistic (more than 30 minutes of difference) perception is a good indicator of insomnia. Total sleep time (TST) and sleep efficiency (quotient between total time in bed and TST) are also good indicators, being TST less than 5 hours very alarming as well as efficiency under 85%. Latency over 30 minutes usually also indicates the presence of insomnia. The EEG analysis is also crucial as alpha or beta intrusions can also be indicators. Not only can PSG help confirm a Insomnia diagnosis but it can also determine its underlying cause. Some common causes, that are caught with an exam of this type are sleep apneas, restless legs syndrome, arrhythmias or periodic sleep movements.

2.4.4 Anamnesis

Anamnesis is the process of interviewing a patient to retrieve clinically relevant data. On sleep medicine this is still a very relevant part of the diagnosis process. On the insomnia context, Paiva and Pinto (2014) do an in depth guide for the anamnestic process, dividing it in subcategories. Those are:

- Circumstances of beginning
- Specific anamnesis for sleep problems
- Severity and frequency of complaints
- Day to day consequences
- Aggravation factors
- Improvement factors
- Sleep hygiene
- Circadian rhythms
- Family and personal history
- Current insomnia treatments
- Other treatments
- Stress factors

Within the *circumstances of beginning* category, it is relevant to establish the age when the symptoms started, how and why (check if there are any triggering factors), as well as the progress (slow and gradual or sudden) and duration of the complaints.

When considering the sleep problems, it matters to know attitudes, behaviours and symptoms right before sleep (namely working/staying on the computer, seeing TV on the bed, going to sleep listening to music, doing late-night exercise or drinking coffee/tea or having heavy meals) as well as the characteristics of the sleep environment (if it is on a sofa or on a bed, alone or not, with lighting or not, temperature, noise, etc.). Still on this topic one key parameters is the average total sleep time (TST), the perceived necessity of sleep time for a completely restoring sleep, and the total time in bed. From these values, sleep latency and efficiency (that should be over 85%) can be calculated. It is common to have exaggerated sleep expectations like nine hours a night and many people try to go to sleep extra early to ensure enough sleep time, rendering sleep latency even higher (over 30 minutes is considered abnormal). Also important are the sleep interruptions, their number, duration and when they occur as well as to check for the reasons why (some common ones are hunger, trips to the bathroom or partner's snoring noise). After the interruptions, it is crucial to know if there is a great effort to go back to sleep and if the patient is scared of not being able to fall asleep again. Lastly it is accessed the overall mood when waking up (for example tired, unresponsive, with headaches or dry mouth).

In terms of severity and frequency of complaints it must be checked how many times they happen per week (usually over three is alarming) and if there are specific nights when they happen (before meetings, trips, earlier rise time). After this the consequences for the day-to-day life are accessed (fatigue, daytime sleepiness, naps, cognition impairments, accidents or bad humor).

Aggravation factors usually encompass next-day responsibilities (usually Sunday nights), preoccupations or noises and improvement factors, the opposite (for example sleeping on another

environment, with no worries).

Patient's sleep hygiene is also very important since frequently they have habits that jeopardize sleep quality (such as falling asleep on the couch before going to bed, doing naps on the afternoon, drinking coffee or other stimulants or even alcohol).

Circadian rhythms must also be characterized, being key parameters the hours of sleep and waking up. These are accessed for week-days, weekends, holidays and shifts (if applicable).

Existence of current insomnia treatments must also be checked since this type of patient is hypersensitive to sleeping drugs and are often over medicated. Treatments to target other disorders can also have impacts on insomnia or even trigger it (like thyroid stimulants, narcotic analgesics with codeine, beta blockers, diuretics, etc.).

Finally, psychological or social stress factors such as too much work (either too much responsibility, work hours, tight schedules and pressure), unemployment, financial problems, family conflicts or marital problems, problems with the children, etc. should be determined.

2.4.5 Questionnaires

Alongside anamnesis, standardized questionnaires are also used to access key aspects of the patient's habits, complaints, rhythms, etc. According to Paiva and Pinto (2014), on insomnia diagnosis the most common are the Pittsburg Sleep Quality Index (PSQI), the Insomnia Severity Index (ISI) and the Glasgow Effort Scale.

The PSQI questionnaire was developed at the university of Pittsburg and its goal is to access general sleep quality over the course of 1 month with questions about sleep times and TSTs, sleep latencies, wake up times, reasons for not sleeping and its frequency (per week) and questions related with daytime impairments. This questionnaire has been translated to 56 languages¹, with CENC using the Portuguese version.

ISI, as the name implies, tries to quantify insomnia severity and impact using a scale. Points are awarded for each option, on each question (0 to 4). A final score is calculated giving an indication of the severity of the insomnia – [0,7] range is normal; [8,14] a little alarming; [15,21] quite severe; over 21 very severe and alarming. The questions target difficulties in falling asleep, keeping asleep or waking too early as well as sleep satisfaction and problems with sleep. Studies

¹PSQI Language Translation List, Sleep and Chronobiology Center (University of Pittsburg) – <http://www.sleep.pitt.edu/research/ewExternalFiles/PSQI%20Language%20Translation%20List.pdf>

show it is a very reliable questionnaire with an Cronbach α (estimator of the reliability of a test, calculated by a function of the number of items, the average covariance between item-pairs, and the variance of the total score) of 0.9 for the community and 0.91 for clinical samples (Morin et al., 2011a).

Lastly, the Glasgow Sleep Effort Scale measures sleep effort, i.e. attempt to force and induce sleep voluntarily. A Portuguese review (Meia-Via et al., 2016) of this Scale found high internal consistency Cronbach $\alpha = 0.79$ and considered the scale a measure of sleep behavior suitable for research and clinical purposes.

According to Paiva and Penzel (2011), other important questionnaires on sleep disorder diagnosis and characterization are the Epworth Sleepiness Scale (measuring daytime sleepiness and can also be used on insomnia studies), Munich Chronotype Questionnaire – MCTQ (questions about sleep schedules for weekdays, weekends and holidays) and Morningness-Eveningness Questionnaire – MEQ (that accesses peak alertness in evening or afternoon).

Besides sleep-related questionnaires some other, more general, questionnaires are employed. One example of these is the Symptom Checklist 90 (SCL90) that consists in a set of 90 questions regarding psychological problems and how much each one affects the patient’s life. As psychological problems are often comorbid or causes for sleep disorders, it is important to access this information to characterize, for instance, an insomnia patient. From the 90 questions, hence the name, 9 sub sets are defined, each one regarding a different dimension of the symptoms. These are somatization, obsessive-compulsive, interpersonal sensitivity, depression, anxiety, hostility, phobic anxiety, paranoid ideation and psychoticism. An extra subset is also considered, that includes additional information about appetite changes, sleep, death thoughts and guilt feelings.

CENC employs all of these questionnaires (and others), which have been translated to Portuguese to fit its target demographic (Portuguese).

2.5 Overview

At the end of this chapter, it is now clear the sheer number of existing sleep disorders there are, with different diagnosis and treatment methods. To worsen the case, many of them are hugely prevalent (like insomnia) and very difficult to classify due to the many different possible triggering factors and symptoms. In sleep clinics, to aid on the diagnosis process, multiple sources of information are employed, from exams and actigraphies to questionnaires, night diaries and direct anamnesis. However, each one of them is totally different from the other

and some, like the polysomnography, very complex. This renders difficult the task of having, readily available, all data sources for one specific patient or to do group analysis from the data. SleepData can help the clinicians and researchers overcome this problem, as it integrates sleep related data under one platform, with quick analysis capability, for both single patients (with a medical information dashboard) or whole populations.

Health Informatics Technology



To develop the SleepData platform some key technologies and informatics resources had to be explored and applied. Over the course of this chapter, I'll cover the basic programming languages and systems (MEAN Stack) used to develop the platform, the main health informatics resources used (HL7 FHIR) as well as some key health informatics standards (SNOMED-CT and LOINC). For each, the advantages and disadvantages against popular alternatives will be analyzed.

3.1 MEAN Stack

The MEAN Stack¹ is a software bundle much like the popular LAMP (Lawton, 2005). While LAMP used a **L**inux operating system, **A**pache to configure the server, **M**ySQL as a relational database and **P**HP as the programming language to connect front end and the database, the new MEAN Stack uses **M**ongoDB as the database, **E**xpress to run the server, **A**ngular to produce a dynamic front end environment and **N**ode to do the server side processing. One of the main advantages of the MEAN stack over LAMP is the possibility of using JavaScript on both client side and server side – data is stored on the database as JSON (JavaScript Object Notation), which is easily processed on the server and then passed onto the client side, where it can be used directly. All programming is based in JavaScript, eliminating the difficulty of switching from PHP to HTML/JavaScript, typical of LAMP. Another up-side of MEAN is the event-driven nature of the Node.js server. Over the next chapters each one of the MEAN's technologies will be addressed (apart from the Angular.js which was not used on this project), covering fundamentals, advantages and disadvantages.

¹The MEAN Stack – <http://mean.io/>

3.1.1 MongoDB

MEAN applications use MongoDB¹ as the database server. This is a non-relational database which organizes data not as tables, but as collections that hold documents. Each document is the core element of this type of database and includes a JSON structure that can have a specific schema or not (any string can be inserted as a part of a document as needed). New fields or information can be added on future revisions of the platform, without having to redo the database schema as is the case with relational databases. On a sleep platform such as this one, this means that more data sources can be added easily, which is an large advantage. MongoDB also assigns a unique ID (.id) to each document making it unequivocally identifiable. However, the main advantage to SleepData lies in its easy integration with Node.js.

On the other hand, unlike rigid relational databases, referential integrity is non-existent with MongoDB. Union operations or cascade deleting, for example, can only be done externally to the database server.

3.1.2 Node.js

Node.js² is a server-side JavaScript runtime built on Chrome's V8 JavaScript engine. It uses an event-based system which allows it to have asynchronous execution of its functions. Unlike, for example, PHP, its functions are non-blocking, executing in parallel (do not block the execution of the next one, before the first one is complete). This efficiency and versatility lead to the adoption of Node.js by many big companies such as LinkedIn, Netflix or PayPal.

A great part of Node.js' versatility is due to its package manager, NPM³, which makes installing, updating and uninstalling libraries extremely easily. With NPM there are more than 475 000 packages available, with 2.4 thousand million downloads (2.4 billion, US) per week and over 7 million users monthly.

3.1.2.1 Express

The last element to the MEAN Stack is the Express or Express.js framework for Node.js. The SleepData's web server was created with Express. The main advantage, besides being integrated with the Node.js' ecosystem with a very large user base and documentation, is its

¹MongoDB – <https://www.mongodb.com/>

²Node.js – <https://nodejs.org/>

³NPM javascript package manager – <https://www.npmjs.com/>

ease of use. The following simple example creates the server, returns 'Hello World' on the '/' handle and serves all the static content on the public folder.

```
var express = require('express');
var app= express();

app.get('/', function (req, res) {
  res.send('Hello World!');
});

app.use(express.static('public'));
```

Being Express.js an extension of Node.js, server hosting and processing are all coded with the same language. This constitutes an advantage over LAMP as it requires Apache, which uses a totally different configuration language, to host the website and then PHP to do the processing.

3.1.2.2 Passport

Passport¹ is an user authentication middleware for Node.js which is used by SleepData to handle user authentication. Passport supports multiple authentication *strategies*, which can all be easily implemented. There are options like signing in through Facebook, Twitter, Google or with a local username and password (which was the case). On the local strategy, after defining the database where to store the user's data and the document schema (MongoDB), the server-side handling of the authentication is simply (copied from the SleepData's server code):

```
app.post('/login', passport.authenticate('local-login', {
  successRedirect: '/homepage',
  failureRedirect: '/badlogin'
}));

app.post('/signup', passport.authenticate('local-signup', {
  successRedirect: '/success',
  failureRedirect: '/badsignup',
}));
```

¹Passport for Node.js – <http://passportjs.org/>

When signing up, Passport automatically hashes the user's password, before storing it on MongoDB. Upon log in, the input password is again hashed and compared with the stored one. This provides an extra layer to password security.

3.2 Health Information Resources and Standards

Health Information Platforms have unique requirements when compared to other, general purpose platforms. In SleepData, data must be specially protected and anonymized, easily shareable, when needed, and unequivocally identifiable. For data to be easily shared while keeping the same exact meaning there is a need for two things: a defined, common information schema (preferably adopted internationally) and a coding system that allows to identify each finding or measure, without ambiguous meanings. The tools chosen to cope with these two necessities are covered over the next two chapters, respectively.

3.2.1 HL7 FHIR

HL7 or Health Level 7¹ is a nonprofit organization that focus on building frameworks and standards for sharing of electronic health information. The Level 7 naming refers to the OSI's model level 7 or application level. OSI (Open Systems Interconnection) model standardizes communication functions without caring about the underlying technologies and structures.

According to their website, HL7 is already supported by more than 1600 members in more than 50 countries. Its vast adoption makes it one of the most compelling options. Others, like openEHR² (open standard for storage and exchange of health data), are less adopted. HL7's focus is not only on the structure of data but also on the protocols to exchange it between systems.

The most recent (since 2004) HL7 standard for medical data structure and transfer is FHIR³ (Fast Healthcare Interoperability Resources), pronounced as "fire". Its resources consist on a collection of data models that define parameters, constraints and relationships between elements. FHIR's resources can come in different sections, depending on their function within health-care applications – clinical (the content of a clinical record, like observations, medication, care plans and others), identification (defines the supporting entities involved, such as patient and

¹Health Level 7 – <http://www.hl7.org/index.cfm>

²openEHR – <http://www.openehr.org/home>

³HL7's FHIR – <http://hl7.org/fhir/>

practitioners, organizations and groups, locations and buildings and devices), workflow (encounters, appointments, tasks, processes), financial (coverage, claims, payment notices), conformance (used to define specifications and structures such as value sets, concepts maps or for testing and development such as implementation guides and test scripts) and finally infrastructure (general resources such as questionnaires, lists and media). As of March 21st, 2017, the Release 3 of FHIR (R3) was published, changing the identification and workflow sections to the new Base section, adding some management resources. The conformance section changed to Foundation, including security resources and moving testing resources and finally, the infrastructure section changed to the new Specialized section. As of R3 the sections are now: Foundation, Base, Clinical, Financial and Specialized.

FHIR resources are available as structure diagrams, UML Diagram or XML, JSON or Turtle Templates. On SleepData, diagrams were used to understand the structure and then JSON to define the database's documents. Having JSON structures available is one of the advantages of FHIR as those are easily integrated on a MEAN Stack based platform. In terms of exchange protocols, FHIR provides a REST API which is also JSON based. OpenEHR has a similar API but it still on a proposal state.

FHIR's structures like Clinical Observations (Observation Resource, Clinical Section, Diagnostics Subsection) need a coding system of health terminologies to uniquely identify the measure/observation being done (example in Figure 3.1).

```

"code": {
  "coding": [
    {
      "system": "http://loinc.org",
      "code": "9279-1",
      "display": "Respiratory rate"
    }
  ],

```

Figure 3.1: Part of the Observation Resource JSON (respiratory rate observation) from HL7's FHIR

The different health terminologies and their coding methods, that are used to identify observations, parameters, measures and exams will be addressed on the next section.

3.2.2 Health Terminologies

In sleep medicine, multiple exams and diagnosis methods are used, which produce many more medical parameters. The expressions used on day-to-day practice can be dubious, and

may only be understood within context. On a platform that intends to centralize all types of medical data (from exam types and characteristics to medical observations and measures) this is unacceptable as it degrades the value of the information due to the loss of accuracy of its meaning. So, to avoid the polysemy inherent with this field, Health Terminologies must be employed. These map each medical term or meaning to a unique identification string that can be used worldwide and thus reducing subjectivity. On SleepData, health terminologies are needed on two categories – firstly, laboratory exam’s results such as DLMO or vitamin D exams and secondly, general anamnesis (medical observation) terms and parameters gathered by equipment like actigraphs, ECG or EEG.

LOINC¹(Logical Observation Identifiers Names and Codes) is an international standard for coding of clinical observations and laboratorial measurements and, as such, covers the first mentioned category. In SleepData it covers measures like the melatonin levels – *Melatonin [Mass/volume] in Saliva (oral fluid)* or vitamin D levels – *Vitamin D [Mass/volume] in Serum or Plasma*. Its importance as a standard of health terminologies and its advantage over other terminologies comes from its reach – users in 166 countries with 20 display languages (based on the official website). As of the update of 23rd of June 2017, LOINC had 84 868 terms, with 7000 edited from the previous version and 2814 deprecated. According to Figueiral (2015), other terminologies, such as ICD-10 (International Statistical Classification of Diseases and Related Health Problems) or ICPC-2 (International Classification of Primary Care), are also widely used but have different main purposes. For instance, ICD focuses more on disorder classification while ICPC is best used for primary care. However, the most appealing factor for the adoption of this system within this work is its great compatibility with SNOMED CT (which will be addressed later).

LOINC provides a search engine² (Search LOINC) that queries the nomenclatures database. The values returned have 6 principal parts – Component (Analyte), Property, Time, System (Specimen), Scale and Method. LOINC’s example for ”manual count of white blood cells in cerebral spinal fluid specimen” subdivides, using these 6 parts, into: Leukocytes, Number Concentration, Point in time, Cerebral Spinal Fluid, Quantitative and Manual Count, respectively. In other words, leukocytes are the analyte, what is measured are their number concentration (which is a quantitative scale), counted manually and only once (a single point in time), from a sample of cerebral spinal fluid.

¹Loinc – <https://loinc.org/>

²Search LOINC – <https://search.loinc.org/>

The second international coding system for health terminologies is SNOMED CT¹. As mentioned, it encompasses LOINC and its scope matches the second type of category of data needed for SleepData – unlike LOINC, its scope isn’t so restricted to laboratorial measurements and clinical observations but instead tries to cover all health-related topics. This includes medical procedures, medical observations and parameters, population characteristics and many others. In SleepData, it codes most parameters from clinical notes, actigraphy and polysomnography. SNOMED’s repertoire is also massive with 326 734 active concepts as of the January 2017 release². According to the SNOMED CT (2016) guide, it is currently available in US English, UK English, Spanish, Danish and Swedish, with partial translations into Canadian French and Lithuanian (with other translations planed).

Based on the same guide, SNOMED’s terminology is based on 3 key components – Concepts, Descriptions and Relationships. Concepts identify (through a unique, numeric and machine-readable identifier) a unique clinical meaning. The descriptions are human readable and can be either FSN (Fully Specified Name) or Synonyms. FSN is a unique description for the concept, while synonyms can be used as display values for the concept (multiple can exist for a single concept). Relationships describe the association between concepts and can be attribute relationships or subtype relationships. An attribute relationship associates a concept with its characteristics. For example, an abscess of the heart can have an ”associated morphology” that is an general abscess and have an ”finding site” which is the heart structure. In theses cases, associated morphology and finding sites are attribute relationships. The subtype relationships, as the name implies, connects each concept with its subcategories. For example, the Ulcer of the Foot concept is a subtype of Disorder of the Foot.

Much like LOINC, SNOMED has a search engine³ (SNOMED CT Browser) that allows quick queries of the concepts library, with custom filters and search options.

3.3 Overview

The data sources and how to analyze them were determined (Chapter 2) and now data structure and terminology standards are needed to allow both, inter-platform connectability and data consistency as well as reduced term ambiguity. HL7’s FHIR addresses the first point, as it provides data structure elements that are specially designed to map medical-related data.

¹SNOMED CT – <http://www.snomed.org/snomed-ct>

²SNOMED Worldwide – <http://www.snomed.org/snomed-ct>

³SNOMED CT Browser – <http://browser.ihtsdotools.org/>

These are "Resources", namely of the "Clinical" type. Its intuitive nature, with schema diagrams and vast documentation, allied with database structural elements defined in JSON (ideal for a MEAN Stack-based platform) makes it a great solution. The need for health terminologies is covered by the pair LOINC/SNOMED CT – these complementary coding systems are both widely used, vast and provide websites with built-in search engines that facilitate the mapping process. These terminologies, even though broad and extensive, may not be sufficient for a platform like SleepData. Some parameters may not be of medical nature (and thus not covered by LOINC/SNOMED CT) but are still very relevant to diagnose sleep disorders or to better adjust treatments and need to be classified unequivocally. Examples are ethnicity, religion or education and the alternative coding systems used to address them will be covered in the next chapter. The ability to add different coding systems to a same parameter, an advantage of using MongoDB and FHIR, leads to a notation that can be quite complex. A simple finding such as "total sleep time (minutes)" can generate more than a dozen JSON fields, as display values, codes, interpretations, and other accessory fields can be registered.

Nonetheless, with these tools as a starting point, SleepData can be built as a useful platform to gather and provide medical data as well as to do powerful data analysis. All the details about the SleepData Platform and its features are covered in the next chapter.

4

SleepData

SleepData is an online platform for gathering, storing, analyzing and visualizing sleep related data. By having a centralized platform, more in-depth data analysis can be performed, with benefits both on the research (with patient or population characterization) and diagnostic fronts. The platform is available at www.sleepdata.inesc-id.pt and is hosted by INESC-ID (Instituto de Engenharia de Sistemas e Computadores – Investigação e Desenvolvimento).

This chapter describes the platform, starting with the database and coding systems employed, followed by concerns regarding data privacy and security (including users' roles and security regulations) and ending with a summary of all features.

4.1 Database Structure

SleepData is built on two MongoDB databases, one for users (to handle login information and access control permissions) and another for the rest of the information. Both of these are organized by collections, which can be of one out of three categories (conceptual categories, not physically different) – administrative, objective measures and reports. In the diagram of Figure 4.1, it is possible to see the different SleepData collections, and to which category they belong.

The collections that belong to the administrative category contain documents with administrative information that may not be useful, at least directly, in terms of clinical data for diagnosis or research. These include the User, Patient, Clinic, Professional and Professional Role collections. As their names imply, they store information about a specific entity or person and not sleep-related data. For instance, the User collection has a document per user, which contains basic contact info and personal information, such as name, language, country and address, gender, etc. Besides that, it also includes the user's unique user ID as well as the user's patient or professional IDs (if the user is also a clinic's patient or employee).

The Patient collection has a document per patient and includes the same information as User, with added fields for a tutor (applicable when the patient is a minor) and the patient's general practitioner, as well as the clinic identification. Unlike the user information, which is

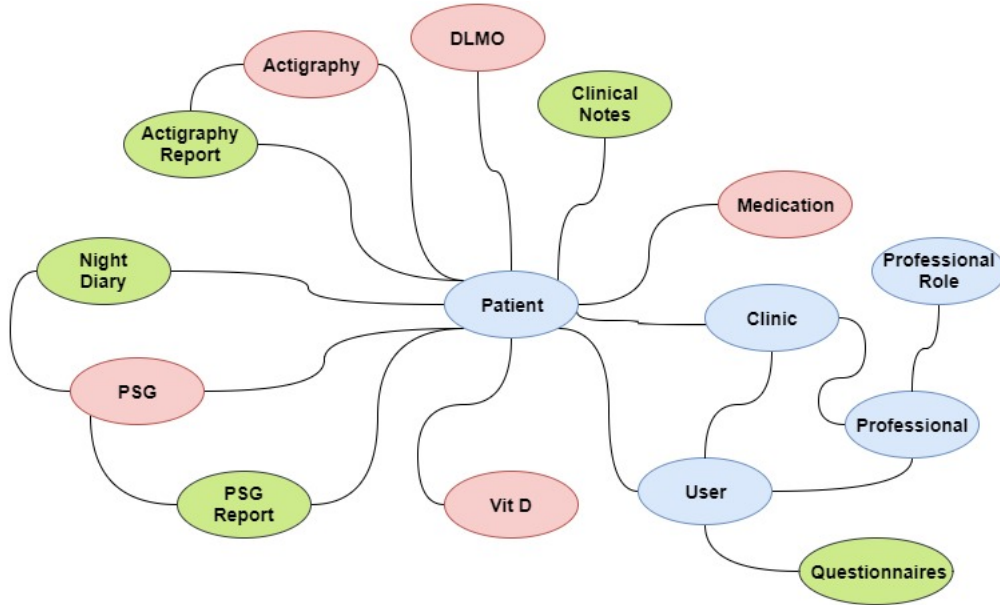


Figure 4.1: Database’s collections, colored by category – administrative in blue; objective measures and exams in red; reports and subjective observations in green. Note that the ”Questionnaires” element is, in fact, a collection per different questionnaire.

self-reported and can be edited at will, patient’s data can only be added by a technician or practitioner. The Clinic collection stores the identification of each clinic available on SleepData. Finally, the Professional collection has information on the collaborators of a specific clinic and includes additional information about their qualifications.

The documents within the Professional Role collection assign a practitioner or other employee to its job, location/office and availability (work schedule and holidays). Both, jobs and medical specialties were coded using HL7 FHIR’s own coding system (more information about the auxiliary coding systems employed is available on the section 4.2). SleepData’s documents are defined using FHIR’s resources. In this case, the Patient and User collections are based on the FHIR resource *Patient* and the Professional and Professional Role collections are based on *Practitioner* and *PractitionerRole*, respectively. The Professional collection is not limited to just medical practitioners. Instead, it supports other employees, as necessary (for instance a IT technician or a system administrator). An example snippet of a Patient collection’s document is shown on Figure 4.2.

Other type of collections map objective findings from exams. On sleep medicine multiple exams are used, with actigraphy, melatonin measures, polysomnography and vitamin D now supported in SleepData. Under this category, the patient’s medication is also included, since it is an objective finding. Each one of these also has its corresponding collection and document structure, based on FHIR’s resources. For each exam, the needed fields were accessed based

```

▼ object {12}
  resourceType : Patient
  active : ☒ true
  ▼ name [1]
    ▼ 0 {4}
      complete : João Márcio Ferreira
      title : Engo
      given : João
      surname : Márcio Ferreira
      photo : /documentos/fotos17/jmf.png
  ▼ maritalStatus [1]
    ▼ 0 {2}
      system : CENC
      code : single
  ▼ telecom [2]
    ▼ 0 {4}
      system : phone
      use : home
      value : 965842157
      rank : 1

▼ 1 {4}
  system : email
  value : jmf@gmail.com
  use : work
  rank : 2
  gender : Male
  nationality : PT
  birthDate : 1974-12-25
  deceasedBoolean : ☐ false
  ▼ address [1]
    ▼ 0 {7}
      use : home
      type : billing
      text : Rua de Camões, 85, 1o esq. Lisboa
      city : Lisboa
      country : PT
      district : Lisboa
      postalCode : 2244-030
  ▼ communication [1]
    ▼ 0 {2}
      language : PT
      preferred : ☒ true

```

Figure 4.2: A portion of an example Patient document (JSON), based on the FHIR’s Patient Resource.

on the literature, consulting multiple CENC’s patients exams and validated by Dra. Teresa Paiva, a renowned doctor on sleep medicine, advisor of this dissertation, owner of CENC and founder of the Portuguese Association of Sleep. A compilation list of all the different parameters and findings stored in each collection can be found in Appendix A (it also includes what FHIR Resources were used).

The mentioned exams were structured after the *Observation* FHIR resource and the medication collection after *Medication*. The *Observation* resource, besides administrative data (such as the date, exam status, subject and performer IDs), consists on a compilation of components, each one of those storing a medical variable. Each component includes a set of three fields per coding system used. Each set has the identification of the coding system used, the code of the medical variable on that particular coding system and a display value. The component also includes the value of the variable itself, as well as a field for the variable’s interpretation and reason for data being absent (if applicable). An example of a component, belonging to a Clinical Notes document is illustrated on Figure 4.3, showing SleepData’s flexibility as multiple coding system are used on the same observation. Even more can be added or removed, as necessary.

Lastly, there are collections that store subjective information, either self-reported, gathered

```

▼ 1 {3}
  ▼ coding [2]
    ▼ 0 {3}
      system : CENC
      code : family_conflicts
      display : Family Conflicts
    ▼ 1 {3}
      system : http://snomed.info/sct
      code : 81935006
      display : Family Conflicts
  valueBoolean : ☒ true
  interpretation : Frequent conflict with the brother.

```

Figure 4.3: Example of a component, belonging to a document (JSON) from the Clinical Notes collection (based on FHIR’s *Observation* resource).

during anamnesis/observation (questionnaires, clinical notes and night diaries, respectively) or derived from objective exams (actigraphy and polysomnography reports). All of these collections were also based on FHIR’s *Observation*, besides the questionnaires that were based on *Questionnaire*. The *Questionnaire* Resource is composed of multiple items, each one with text, code and value (used to calculate scores). For each questionnaire there is a different structure, stored on a different collection. As such, the collections are Epworth, Glasgow, MCTQ, ISI, MEQ, PSQI and SCL90, each one with the name of its corresponding questionnaire.

Now that all document types are described, it is possible to introduce some of the relationships among different concepts.

The *User* is the central element, as he’s the one who does all the actions and interacts with the other elements. As mentioned, a User can be a Professional or a Patient. A User can also fill Questionnaires, that can be of any of the just mentioned kinds. Both Users and Patients can belong to a Clinic. A User also has Permissions, which determine its access to data and to functionality (more details on this topic on section 4.3).

The *Patient* is the main link for all the data sources, since they all refer back to him. A Patient can take Medication or do exams. The exams, in turn, can be PSGs, Actigraphies, Clinical Notes, DLMO or Vitamin D. Each one of these exams can have some or all of 3 components, namely Raw Files, Objective Data and Reports. Raw Files are uploaded files and are saved without being altered. Objective Data can be either produced through parsing of a Raw File or introduced manually. Finally, Reports are compilations of subjective findings, introduced

manually by the practitioner.

4.2 Coding Systems

On a platform such as SleepData, which handles diagnosis information and is applicable to real-life patients, it is absolutely critical to avoid misunderstandings with ambiguous terms. Having international codes to map each finding or parameter also allows the platform to be more general and its data more easily sharable and comparable. SleepData adopts LOINC and SNOMED CT, which were used to code laboratorial and exam parameters as well as medical findings. Nonetheless, there are still other coding systems required for parameters of domains not strictly medical. One example of this is the ethnic group of Patients. Even though they may not constitute clinical variables they may provide crucial insight into the lifestyle and characteristics of the patient, which is relevant to the practitioner.

The coding of ethnicities presented some challenges, firstly because it is a sensitive topic and not a common category to code, and secondly because it is very subjective. It is difficult to distinguish where one ethnicity ends and another begins and because of the commonness of people with traits from multiple ethnicities. To add to that, the primary case of study with the platform used the Portuguese population, for which most international standards (usually focused on the USA) are inadequate.

The solution was to code the most common and most clinically relevant (based on Dr. Teresa Paiva's expert opinion) ethnicities with SNOMED's terms from both the Racial Group and Ethnic Group categories. The categories chosen are the ones listed below (from SNOMED CT).

- Afro-Caucasian (ethnic group)
- Arabs (ethnic group)
- Asian (racial group)
- Black (racial group)
- Caucasian (ethnic group)
- European (ethnic group)
- Gypsies (ethnic group)
- Hispanic (racial group)
- Indian (racial group)

Besides ethnicity, more non-medical observations were coded, including religion, education level and job/occupation. On the Professional Role collection, the roles and medical specialties

of the practitioners were coded as was the alcohol consumption observation on the Clinical Notes collection.

It is important to remember that, much like with the medical observations, these codes can be replaced and others can be added from SNOMED CT or even from other ontologies – an advantage of using MongoDB to structure the database.

To model religions, the FHIR's religious coding system¹ was used – "HL7 v3 Code System ReligiousAffiliation". This code system includes all the main religions of the Portuguese population, including Atheism, Agnosticism, Christian (non-Catholic, non-specific), Non-Roman Catholic, Roman Catholic Church, Islam, Jehovah's Witnesses, Judaism, Church of God, Evangelical, Baptist and Protestant. The full list includes 82 religions and has the advantage of being directly available on JSON and XML formats.

Education levels were coded with ISCED, the International Standard Classification of Education (UNESCO - Institute for Statistics, 2011). These codes are presented in layers, being more specific from one layer to the next. For example, on the code 746 – Long first degree (at least 5 years) – the 7 refers to the category 7, "Master or Equivalent Level", of the first layer, the digit 4 refers to the category 4, "Academic", of the second layer and the last digit, 6, refers to the "Master or Equivalent Level". The different categories from the first layer, from 0 to 8 are "Early childhood education", "Primary education", "Lower secondary education", "Upper secondary education", "Post-secondary non-tertiary education", "Short-cycle tertiary education", "Bachelor's or equivalent level", "Master's or equivalent level" and "Doctoral or equivalent level", respectively. The first digit 9 is reserved to "Not Elsewhere Classified".

ISCO-08 (International Standard Classification of Occupations) from the International Labour Office (2008) was used, as the name implies, as a coding scheme for jobs and occupations. Like ISCED, ISCO follows a layered approach, with each sequential digit representing the category within each layer. The first digit from the code represents one of the 10 groups, respectively: "Armed Forces Occupations", "Managers", "Professionals", "Technicians and associate professionals", "Clerical support workers", "Service and sales workers", "Skilled agricultural, forestry and fishery workers", "Craft and related trades workers", "Plant and machine operators, and assemblers" and "Elementary occupations".

On the Professional Role collection, the professional's "occupation" also need to be coded, but a classification like ISCO-08 brings unnecessary complexity to the few different jobs of

¹FHIR's Religions coding system – <https://www.hl7.org/fhir/v3/ReligiousAffiliation/cs.html>

interest to a clinic. As such, to map each role, FHIR's Practitioner Role coding system¹ was employed. These codes were designed to be used with the *PractitionerRole* resource (which is the base for the Professional Role collection) and its simplicity provides an advantage over ISCO-08. If needed, however, both can be used simultaneously. FHIR's classifications only have 6 codes, namely doctor, nurse, pharmacist, researcher, teacher/educator and ICT professional, which cover all possibilities on CENC. For a doctor, specialty is coded with SNOMED CT.

Finally, on the Clinical Notes collection the coding from the National Institute of Alcohol Abuse and Alcoholism² (NIAAA) was used to encode alcohol consumption observations. This classification has 3 levels:

Moderate alcohol consumption which includes anything up to 1 drink per day for females and 2 per day for males.

Binge drinking a pattern of drinking that brings alcohol blood concentration to 0.08 g/dL (about 5 drinks in 2 hours).

Heavy alcohol use which consists on binge drinking for more than 4 days on the previous month.

On many of these parameters, for instance the multiple doctor observations, a CENC's own coding system was also adopted (as visible on Figure 4.3). This ensures that commonly used parameters can be quickly found by the "in house" name. The idea is to benefit from the platform's flexibility to add coding systems as needed – one international for sharability and another to maintain consistency with the clinic's practices.

4.3 Access Control

After a user logs in to SleepData the content that is displayed and available to him is personalized, according to who the user is. To achieve this, different types of users were defined. To each user class, a different homepage is offered, with very different capabilities.

The base user, from now on referred to as just "User" or "Regular user", has just access to a minimum set of features. He can see basic statistics of the platform (just like any visitor),

¹FHIR's Practitioner Role coding system – <https://www.hl7.org/fhir/valueset-practitioner-role.html>

²Drinking Levels Defined, National Institute of Alcohol Abuse and Alcoholism – <https://www.niaaa.nih.gov/alcohol-health/overview-alcohol-consumption/moderate-binge-drinking>

like the number of Users, exams and other general data. In terms of data input, a User can fill sleep questionnaires in addition to providing his personal information, such as his name, date of birth or gender.

A user can be specialized into the "Professional User" class. This is usually assigned to a clinic's clinician, technician or researcher. This type of user can register new patients, add reports or exams to a certain patient or access patient's data and population-wide statistics. Each "Professional User" is associated to a clinic in SleepData and cannot access other clinics patients data. Whenever a new clinic is registered on SleepData, a new type of "Professional User" can be assigned, by the administrators, corresponding to the clinicians of that specific clinic. "Professional Users" can also map users to a patient profile from their clinic.

Finally, the last class of users are "Admins" or "Administrators". Administrators can register new Professionals (registry of a employee from a clinic, which is different from the "Professional User" category of user, used for access control) and define new clinics. They can also change any users class and assign them to Clinics, as well as map users to Professionals. Admins can also delete user accounts.

This customization of access based on classes or types of users guarantees that the user experience is personalized and, more importantly, that data is only available to the right person. Screenshots of the graphical interface of each type of users' homepage is available on the section 4.5.1.

4.4 Compliance with Security Regulations

SleepData is designed to comply with the security standards defined by the Deliberação nº.1704/2015 from the Comissão Nacional de Protecção de Dados (2015). This is a guideline of the Portuguese Data Protection Authority (CNPd), that provides a list of procedures that must be followed to ensure security of the medical data, during studies. Table 4.1 has a summary of how SleepData complies to some of the key Items from the guideline.

Besides CNPD's recommendations, some other common security measures were adopted. Firstly, by using Node.js' Passport, all the passwords are kept encrypted, preventing eventual password leaks and making them indecipherable to system administrators. To prevent man-in-the-middle attacks, a SSL certificate was acquired for the `sleepdata.inesc-id.pt` domain. With security concerns addressed, the focus of the next chapter will be the platform's functionalities.

Table 4.1: SleepData’s Compliance with the CNPD Data Security Guidelines.

Item no.	Description	SleepData
44	The system must provide logical separation between personal and medical data	User and patient collections (personal data) are separate from the different exams and clinical notes collections (medical data)
	Access to data must be defined based on different user profiles	Three different user profiles "Regular User", "Professional User" and "Admin", with different access to the data
	User privileges must be kept actualized (updated or removed)	"Administrators" can remove or update user’s type (changing their access)
45	Non-authorized persons must not have access to the data	Log in is mandatory to access data. "User"s without other privileges can only fill questionnaires
48	Restricted physical access to the server	Server hosted on FCCN data center, with physical barriers and 24/7 security
	Backed-up data must only be accessible to the system administrator or other technicians (if bound by professional secret)	Only system administrators have access to the backups. Regular backups are done by FCCN.

4.5 Functionalities

SleepData aims to be a central hub for all data related with sleep medicine by allowing to both, add more data to the platform and to visualize it. For this reason, its features focus on those two points – firstly, we have data input interfaces and its inherent processing and secondly, data visualization tools with graphs and summaries. Over the remainder of this section, the main features will be presented and user’s access to each feature discussed. Figure 4.4 provides a diagram of the structure of the SleepData website. It shows how each of the functionalities is accessed and is useful to better understand how they work.

4.5.1 Landing Page and Home Pages

The first page available when navigating to SleepData, available at `sleepdata.inesc-id.pt`, is the landing page (see Figure 4.5). Users can login or register for the first time, choosing a username and a password. Besides login, general platform statistics are displayed, including number of users and patients as well as the number and type of the several types of exams available on the platform. This data is updated each time the page is loaded.

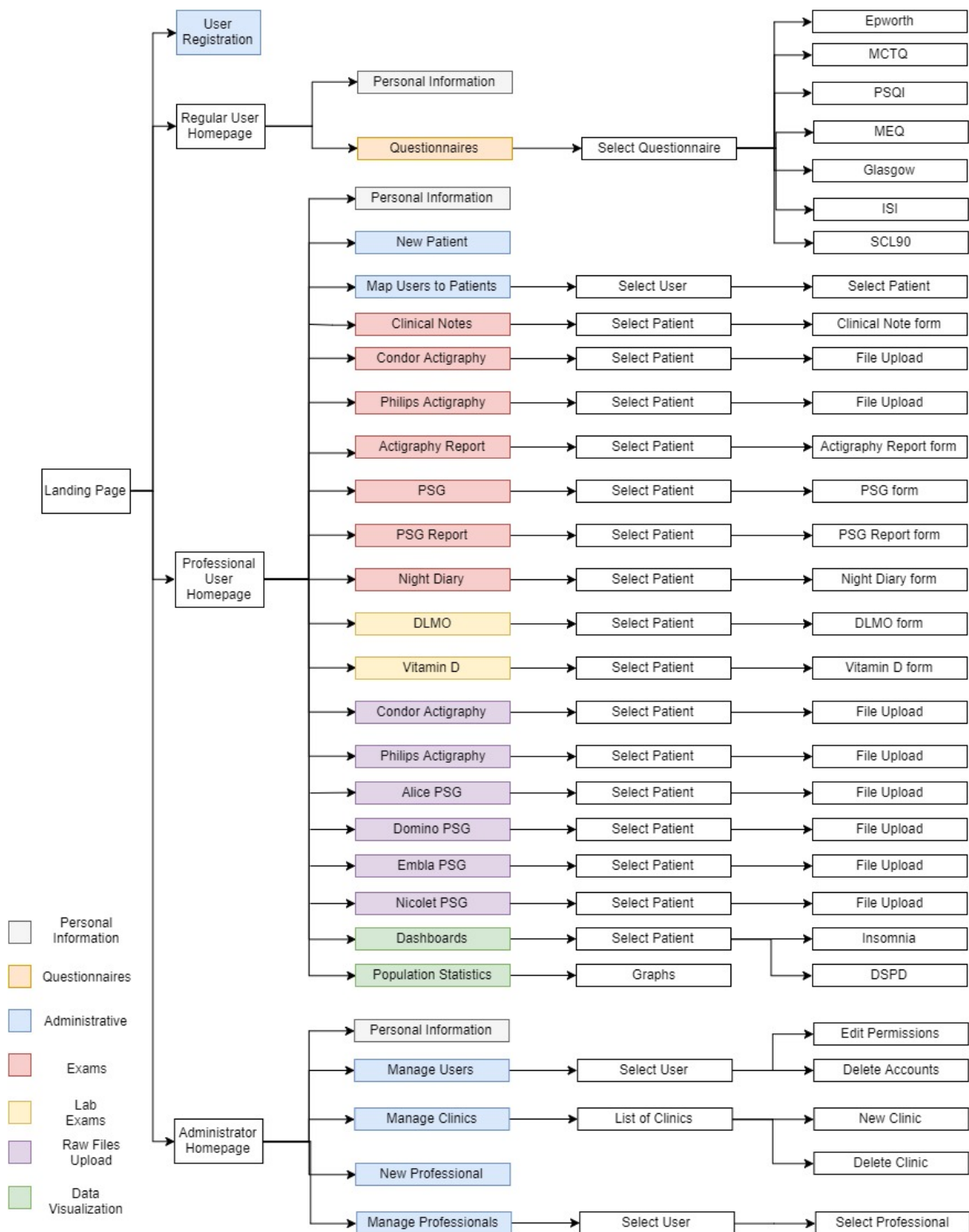


Figure 4.4: Diagram of the SleepData website structure.

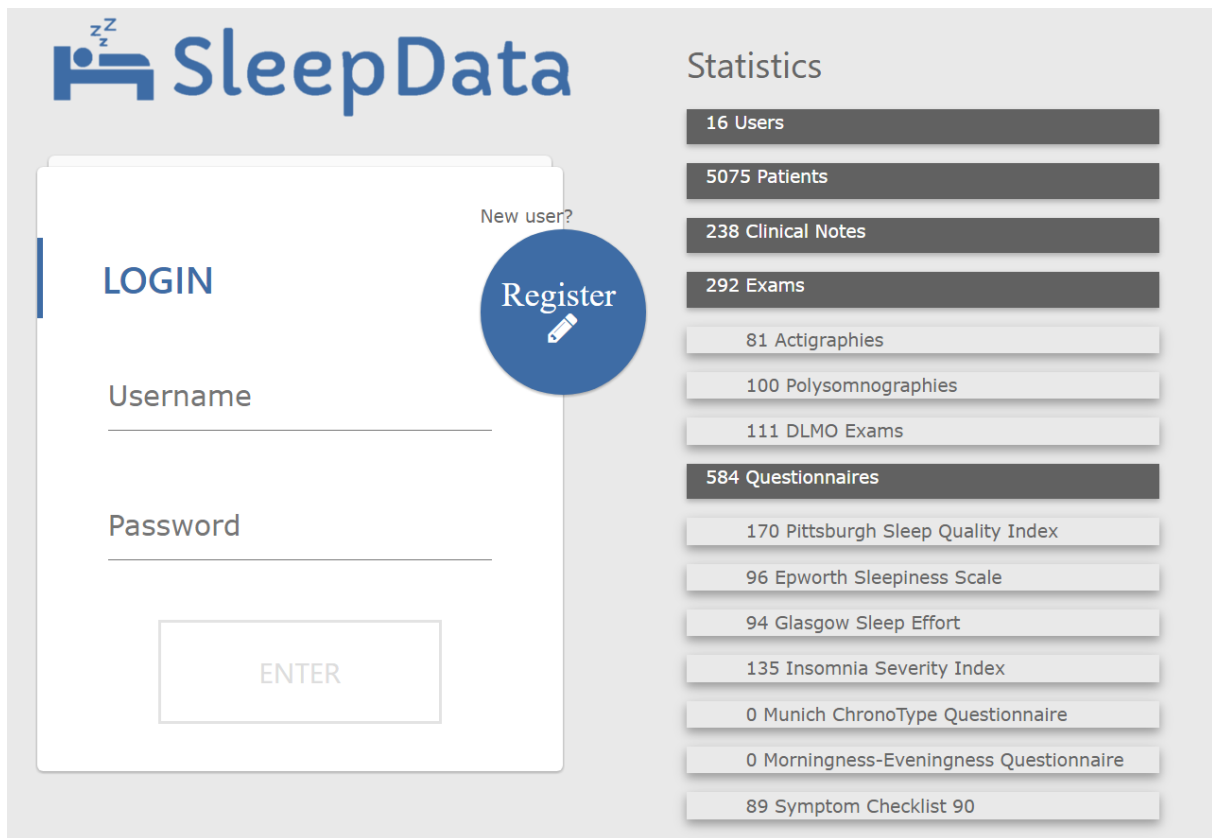


Figure 4.5: SleepData’s landing page – available at `sleepdata.inesc-id.pt`.

After logging in, users are redirected to their homepage. The content of their homepage is dependent on their type so only the appropriate content is available to each user. From all the user types, the ones with more available functionalities are the professional users. Figure 4.6 shows the homepage for a fictitious professional user, Ana Maria, that belongs to the CENC clinic. The homepages for the administrators and regular users follow the same design cues.

Each button leads the user to a specific page related to a different functionality. Regular Users only see two buttons, which allow them to edit their profile information or to fill-in sleep questionnaires. An Administrator, on the other hand, can register clinics employees, manage the user accounts and manage the clinics. Finally, the professional user, which is the user with most functionalities available. Their homepage is divided into categories (Figure 4.6), each one with a different type of functionality:

Administrative functionalities are similar to admins’, but more limited, as professional users can’t delete accounts or edit roles and can’t add professional profiles or register clinics (just register patients and map users to patients).

Exams refer to user interfaces to input data, being either objective data or subjective data.

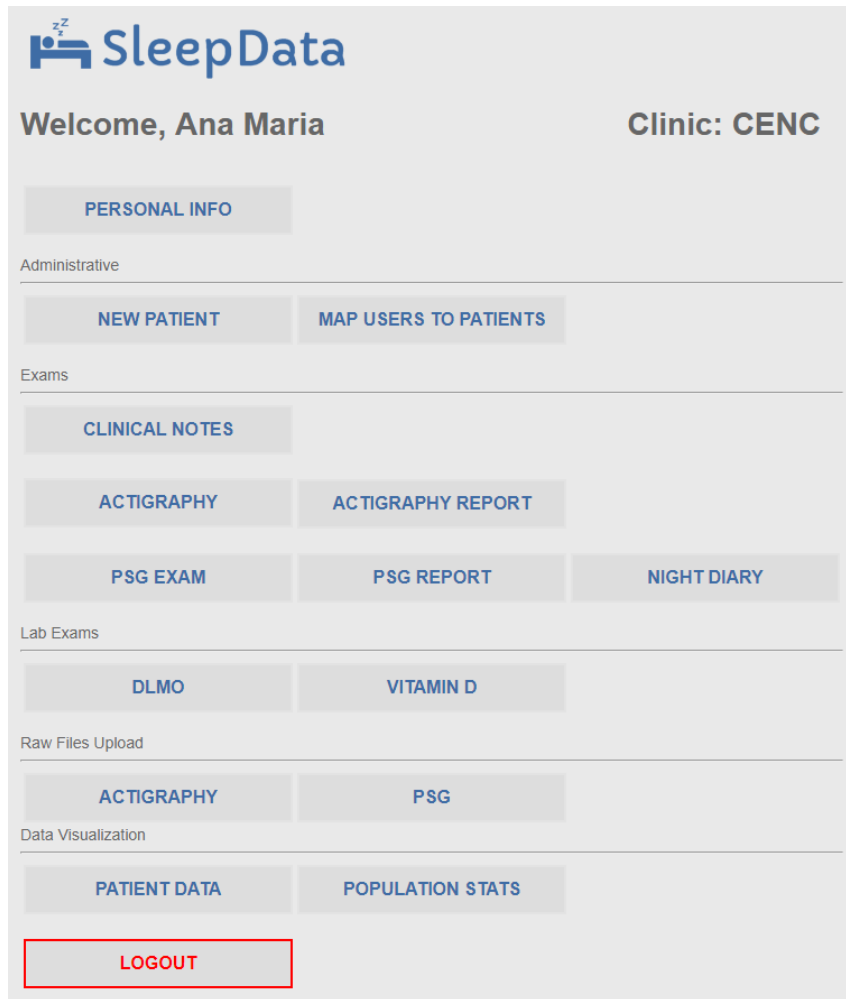


Figure 4.6: Homepage of a CENC professional user. Notice the different sections used to separate the functionalities by theme.

Lab Exams allow the user to input laboratory exam data, namely from melatonin or vitamin D exams.

Raw Files Upload include interfaces to upload files, including actigraphy raw files from Condor or Philips or PSG raw files from Alice, Domino, Embla or Nicolet models.

Data Visualization refers to the graphical reports, being either specific to a patient (in the form of Dashboards) or on a population basis (Population Stats).

4.5.2 Manual Data Input Interfaces

One of the main ways to add data to SleepData is by using its HTML/CSS/JavaScript graphical user interface (GUI). As seen, these forms can be accessed from the homepage and are used to input different types of data. The forms that are used to modify other users' data (such as giving/removing permissions by an admin or addition of clinical notes of a patient by

the practitioner) have an intermediate step – an options table – where said user can be chosen. As mentioned, Figure 4.4 further explains the function of each of these tables.

After choosing an user to add information to, the next form may have some information already filled in, if available on the database. The example of Figure 4.7 shows that, after choosing the patient, his name and ID are automatically added to the form.

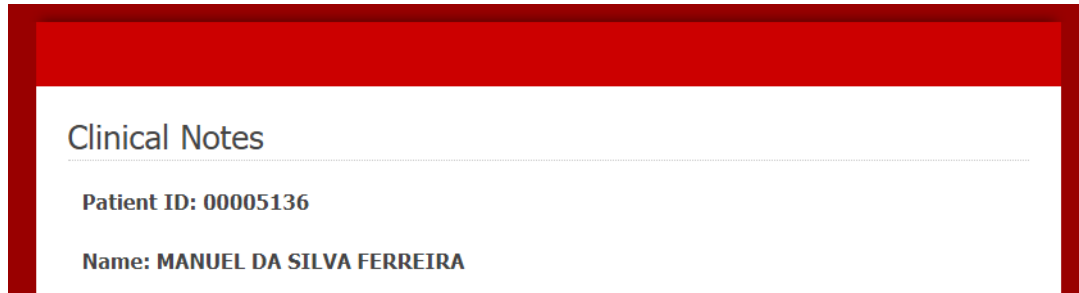
The image shows a web form titled "Clinical Notes" with a red header bar. Below the title, there is a horizontal dotted line. Underneath the line, the text "Patient ID: 00005136" is displayed in bold. Below that, the text "Name: MANUEL DA SILVA FERREIRA" is displayed in bold. The form is enclosed in a red border.

Figure 4.7: Part of the clinical notes form. Note that (as the user was already selected) name and patient’s ID are already filled in.



Regular users only have available forms to edit their personal information and to fill sleep questionnaires. On the other hand, professional users have many other available, including forms to create another patient, add clinical notes, manually insert actigraphy reports, manually insert PSG data (objective parameters only) and PSG report data (subjective), fill DLMO exams as well as vitamin D ones and write night diaries. Admins have access to "administrative" forms to add a new professional, edit permissions or delete users.

4.5.3 Sleep Questionnaires

Besides the forms mentioned on the previous chapter, users also have access to sleep questionnaires. These target a well-known need, from the sleep clinics, to have a centralized, digital hub where users can fill questionnaires (with no need for the doctor’s direct intervention). To adapt to the first clinic to adopt SleepData, CENC (where some patients aren’t proficient in English), these questionnaires were developed in Portuguese, being the implementation of more translations a priority among future improvements. Besides the language, questionnaire’s forms, have different graphical interfaces, with more images, tables and other elements, when compared to the other forms. Those might be needed to illustrate what answer is intended in each question or to highlight the organization that developed the questionnaire, for example. As with all other forms, some information is automatically filled (namely the user’s name, ID and current date). An example of a questionnaire, the Symptom Checklist 90 (SCL-90), is displayed on Figure 4.8.

The information about the relevance of each questionnaire is mentioned on chapter 2.4.5 and,

Questionário SCL-90

Nome: Maria Josefina

CENC ID: 00000001

Data: 20/10/2017

Instruções

Em baixo está uma lista de problemas e queixas comuns. Por favor leia atentamente cada uma delas.

Depois de o fazer, selecione o descritor que melhor caracteriza o quão incomodado foi, por cada um deles, DURANTE A ÚLTIMA SEMANA, INCLUINDO HOJE.

Quão Incomodado foi por:

	Não de todo 0	Um pouco 1	Moderadamente 2	Bastante 3	Extremamente 4
Dor de cabeça	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Nervosismo	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure 4.8: Excerpt of the SleepData’s Symptom Checklist 90 (SCL-90) questionnaire.

as mentioned there, the implemented questionnaires were: Pittsburg Sleep Quality Index (PSQI), Insomnia Severity Index (ISI), Glasgow Effort Scale (“Glasgow” for short), Epworth Sleepiness Scale (“Epworth”), Munich Chronotype Questionnaire (MCTQ), Morningness-Eveningness Questionnaire (MEQ) and Symptom Checklist 90 (SCL-90), on its revised version.

4.5.4 Exam Upload and Processing

Besides manually introducing data, professional users can directly upload exam files. The uploaded files can be of two kinds – raw files (PSG or actigraphy) or actigraphy reports generated by the actigraphs’ analysis software. When uploading these automatically generated reports, the user must specify the actigraph model, as each one has a different export file type – Philips’ word documents or Condor’s PDF files. Those are automatically parsed so relevant information is gathered and wrote to the database. As each file has a predefined structure, each field on the word/pdf file can be mapped, with a script, to the corresponding HL7 FHIR’s structure. The fields that are available for actigraphy (and that were mapped) are available in Appendix A. After this step, the original files (PDF or word) are also saved on the platform, on an organized

manner, so that they can be viewed or analyzed later.

If the user chooses to upload the actigraphy raw files, those will be saved, but are not analyzed (which is not necessary as most relevant data is present on the reports). When sending raw PSG files, users must also choose the corresponding equipment brand/model in order to save the files on separate folders. Currently, no processing is done to those files but later on, scripts can be implemented to parse such files and extract valuable information (note that each PSG brand and model exports on a totally different manner, with different file types, sizes, structures, etc. rendering this task very strenuous). For any file uploaded, the equipment, patient and date are registered.

4.5.5 Clinical Data Visualization

One of the main advantages of having data from multiple sources integrated under the SleepData platform is that the data from each patient can be displayed on a centralized manner, so "the bigger picture" can be quickly accessed by the physician. Not only that but many population-wide statistical measures can be calculated and graphs plotted. This section introduces the SleepData's data visualization features.

4.5.5.1 Patient Data Dashboard

Sleep medicine practitioners face two main issues while looking through patient records. The first one is related to the fact that each exam or consultation produces a separate document (digital or not), which doesn't have any connection with other documents or with the patient. That means that compiling all data sources (for example to prepare an appointment) is a difficult process, that may involve looking through several computers/hard drives, hand written notes and clinical processes. The second issue is related with the sheer amount of data and its variability. Doctors might look for similar parameters on most of their consultations but those can be organized on a unusual manner (on hand-written clinical notes or on a night diary) or might be just too many to quickly find the desired one.

To organize such data, SleepData provides patient data dashboards. As the name implies, these are table-like pages that display crucial information about a patient. The dashboards organize the data by sections, depending on the source, namely: PSGs, PSG Reports, Actigraphies, Actigraphy Reports, Vitamin D exams, DLMO exams, Clinical Notes and Questionnaires. Within the questionnaires section there is a subsection for each one.

Clinicians can have additional dashboards (two were implemented so far) that focus on different aspects and have different key parameters displayed. This way the doctor (any professional user) can quickly visualize all the relevant data on a certain context. The dashboards were developed in cooperation with Dra. Teresa Paiva and based on the guidelines described on her sleep medicine practical guide (Paiva and Penzel, 2011) and overview book (Paiva and Anderson, 2014b). One of the dashboards provides an overview of the key parameters to characterize a insomnia patient (or one where there is a suspicion of such diagnosis) and the other a delayed sleep phase disorder (DSPD) patient. Figure 4.9 shows the DSPD dashboard, namely its actigraphy section.

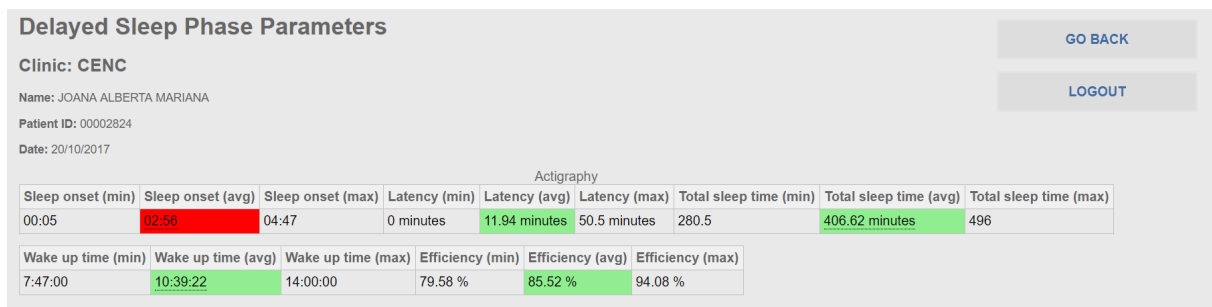


Figure 4.9: Screenshot of part of the Delayed Sleep Phase Disorder dashboard (only the actigraphy section is visible).

To alert the clinician to certain data points, each dashboard has a color coding system to highlight clinical variables. The color change based on the severity – red for very severe, orange for some attention required and green for values within the normal range (see Figure 4.10). This allows the clinician to quickly interpret the data and to do an overall assessment of the patient's condition.

PSG					
Efficiency	Total Sleep Time	Number of cycles	REM (%)	N1 (%)	Latency
98 %	500 minutes	5			5 minutes

PSG					
Efficiency	Total Sleep Time	Number of cycles	REM (%)	N1 (%)	Latency
85 %	350 minutes	4			20 minutes

PSG					
Efficiency	Total Sleep Time	Number of cycles	REM (%)	N1 (%)	Latency
75 %	280 minutes	3			35 minutes

Figure 4.10: Detail of the Insomnia Patient Data Dashboard that refers to PSG data. Notice that the color of the cells change, based on the patient's data.

The list of parameters displayed on each dashboard, alongside the rules used to apply the

colors to each cell, is available in Appendix B.

4.5.5.2 Population Statistics

In addition to summaries for specific patients, professional users also have access to population-wide statistics. These consist on a set of graphs and stats that provide a quick overview over the main characteristics of all of the patients in SleepData. Through the interface, the user can filter the results so only some of the patients are considered. Choosing a filter will update the graphs accordingly. So far, two filters were implemented on SleepData, so Insomnia or DSPD patients can be selected.

After choosing a filter (or none, to select all patients), the graphs and stats are shown, divided into categories. These separate the data based on the corresponding source and are intended to improve usability. The categories are: General information, clinical notes, actigraphy, DLMO, polysomnography, actigraphy, PSG and sleep diary comparison, PSG report, PSQI, ISI, Epworth Sleepiness Scale, Glasgow Sleep Effort Scale and SCL-90. A comprehensive list of all the available graphs of the Population Statistics is available on Table C in Appendix C.

The first section includes demographic data, such as the age and gender distributions of the patients. Figure 4.11 displays a snippet of the population's statistics that includes the general information. On this screenshot, two of the available graph types can be seen – histograms and pie charts. On the platform, all graphs can be hovered with the mouse (or clicked) to show additional information. It is also visible the sample size used for each graph, which is always shown.

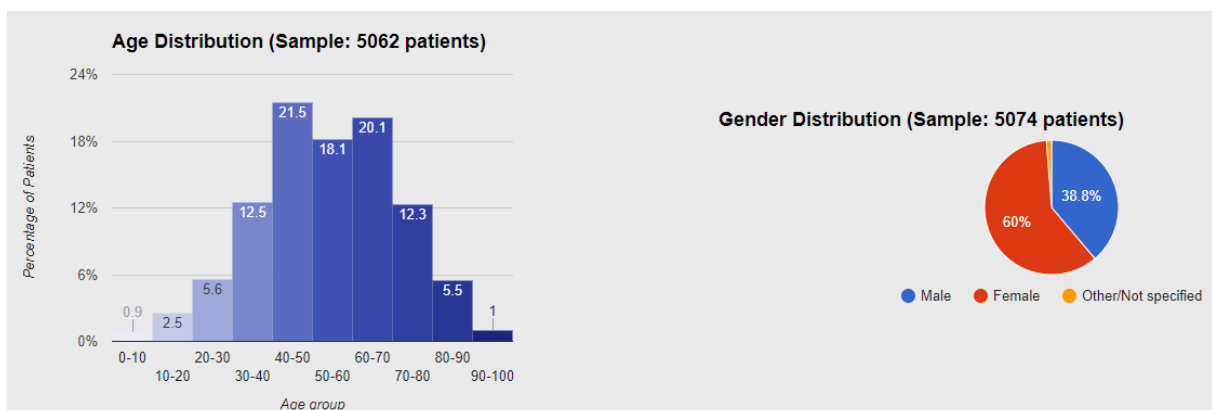


Figure 4.11: Screenshot of the "General Information" section of the Population's Statistics feature. No filter is active so all patients in the platform are considered.

The clinical notes' section includes a overview of the information gathered by the practitioner through anamnesis, which includes histograms of age at first symptoms, traumatic events (car

accidents, family conflicts, etc.) and bad habits (such as smoking, alcohol or drug use). On this section a overview of the patients' associated comorbidities and day-time complaints (originated from the poor sleep quality) is also done, with histograms, bar charts and pareto charts. An example of this last type of chart, in this case regarding number of comorbidities is visible on Figure 4.12.

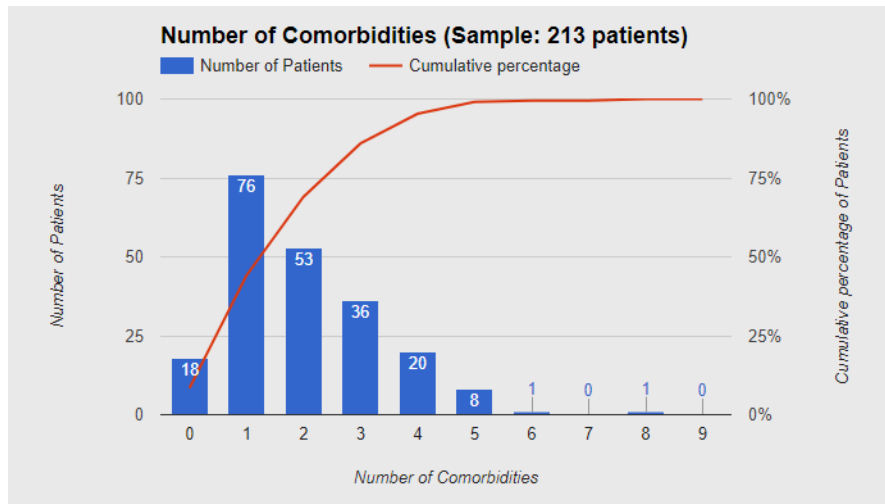


Figure 4.12: Example of a pareto chart from the Clinical Notes section of the population's statistics. It refers to the number of associated comorbidities.

Regarding actigraphy, there are graphs for total sleep time distribution and sleep latency, as well as average bed times and wake up times. On DLMO, the phase angle is shown against the actigraphy values and against the patient's sleep diaries. A reminder that the phase angle stands for the difference between DLMO (dim-light melatonin onset) and sleep onset.

On the PSG's section, as with actigraphy, total sleep times and latencies are analyzed, as well as sleep efficiency. On PSG Report the focus of the analysis is the perceived sleep time and how it compares to the actual elapsed time.

Finally, for each questionnaire, the score's distributions are displayed using boxplots. As mentioned, every graph can be hovered with the mouse for more information. In the case of boxplot, the values for each quartile as well as the average value can be seen. Besides those, there are also pie charts that show the percentage of patients that belong to each bracket of score, similarly to what happens on the Dashboards. Hovering the pie charts highlights how many patients are on each category. An example of the graphs for the ISI questionnaire is shown on Figure 4.13. In this case the score brackets are defined at 8, 14 and 21 points.

Considering all categories, over 30 graphs are available to the professional users (clinicians or researchers), which allows them to gather important statistics for specific sub populations

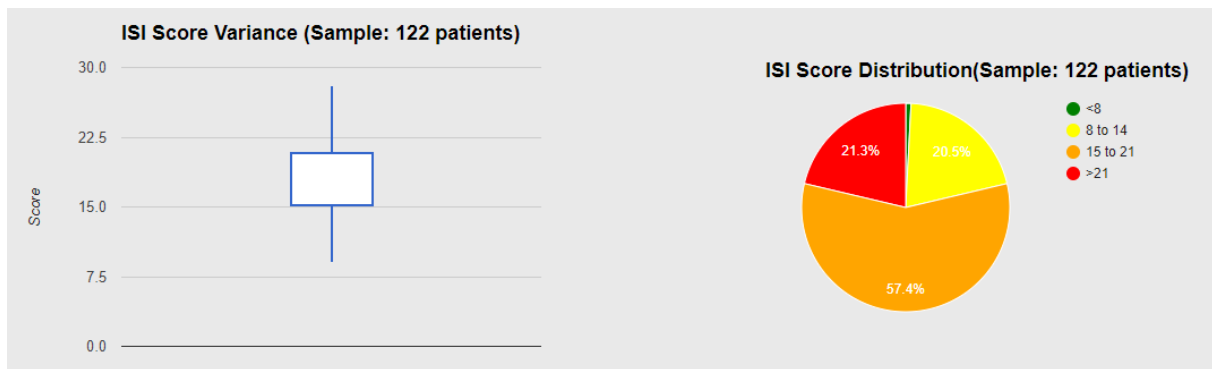


Figure 4.13: "ISI" section of the population's statistics. By hovering the boxplot graph, on the left, users can see the boxplot's details such as the values for each quartile and average score.

within the platform.

4.6 Overview

All of SleepData's details and features were addressed, and it is now possible to quickly assess if the desired specifications for the platform were met. One of the main objectives is that the platform can handle sleep-related medical data from multiple clinics. This implies that more fields may need to be added, to address the need of each clinic's practitioners, and international coding systems are used, so data is consistent across the platform and shareable. As seen, the MongoDB database allied with a FHIR structure, allow the platform to be very flexible in terms of adding fields and coding systems. The latter were chosen from reliable, international sources, so they can be adopted world-wide.

Security was also a concern, as patient data is private and sensitive. To address this topic, a access control system was implemented, with conditional access to functionalities and data, based on the user's type/class. Furthermore, CNPD's security guidelines were followed.

Another requirement, was that users could fill sleep questionnaires and providing clinicians interfaces to introduce data, manually or by uploading files. Both of which are now possible – seven questionnaires were implemented and routines were developed to handle the uploading of PSG and actigraphy files and forms were made so data can be introduced manually.

Finally, data needed to be available to clinicians and researchers so they can use them for either diagnosis or studies and population characterization. These were addressed by the development of patient data dashboards and the population characterization feature. This way, a patient-by-patient data compilation can be seen, as well as a summary of population's statistics. Furthermore, the use of internationally accepted coding schemes makes these statistics general.

To validate the population statistics feature and as a proof of concept of SleepData as a research platform, an Insomnia population characterization was conducted, to be described in the next chapter.

5 Insomnia Population Characterization

SleepData allows for multiple studies to be performed, using its multiple data sources and analysis capabilities. This characterization can be done in two ways: by analyzing each patient, observing its data on an organized manner or by characterizing the population as a whole, finding common traits. The former is addressed with SleepData’s dashboards and the latter by the Population Statistics feature. In this chapter, I’ll address both of these capabilities, applied to insomnia. Firstly, the capabilities and specificities of the Insomnia Patient Data Dashboard as a tool for diagnosis and single-patient characterization and secondly I will characterize a Insomnia Population using the Population Statistics tool. All graphs and tables shown were directly copied from SleepData, as only its tools were used.

5.1 Insomnia Patient Data Dashboard

SleepData’s dashboards have sections that separate the data based on its source. This way, a parameter from a specific exam can be found faster and the data is more organized. On this dashboard there are four main sections – PSG, PSG Report, Clinical Notes and Questionnaires – each one corresponding to a main source of information regarding Insomnia diagnosis and characterization. Besides those, two smaller sections are also available, namely Actigraphy Report and Vitamin D. The remainder of this chapter goes over the importance of each of these data sources and the most crucial variables gathered from each one. I’ll address some of the color codes used, but a comprehensive list of coloring rules is available in the Appendix B.

Clinical Notes (anamnesis) is the main source of information for most sleep disorders and Insomnia is no different. The physician starts by accessing which are the symptoms of the patient, including the circumstances of how they began (when and how). He also tries to infer what were the triggering factors. During this process, the risk factors are also registered. The dashboard’s Clinical Notes section (Figure 5.1) represents just that. The first information shown is the patient’s comorbidities, followed by the symptoms that are characteristic of Insomnia. These include the day-time complaints that are, by definition (American Academy of Sleep Medicine, 2014), necessary for a Insomnia diagnosis and sleep-related fears like fear of not

being able to sleep. After that, the data that refers to the "how" the symptoms started is shown. This includes triggering factors, such as traumatic events or more subtle (although still harmful) lifestyle characteristics such as alcohol consumption, smoke habits or a bad workplace environment, since they are common among Insomnia patients. In the Clinical Notes section of the dashboard, comorbidities and Insomnia-related complaints are highlighted following the color code – comorbidities are red, if they exist, as are the complaints and fears; otherwise, they are green.

Clinical Notes								
Comorbidities	Age at first symptoms	Complaints at wake up	Daytime impairments	Cognitive complaints	Sleep-related fears	Family history	Triggers	Traumatic experience
Depression Anxiety		No	No	No	No	No	No	No
Alcohol consumption	Drug consumption	Smoking habits	Stress	Family Conflicts	Professional underachievement	Works in shifts	Stress at work place	Procrastination
No	No	No	No		No	No	No	No

Figure 5.1: Clinical Notes section of the Insomnia Patient Data Dashboard.

After anamnesis, polysomnography (PSG) and its respective report (developed by the practitioner) are the most important sources of information for evaluation of Insomnia. This exam allows for the determination of key sleep parameters, like sleep efficiency, sleep latency and total sleep time, without the uncertainty associated with self-reported data (sleep questionnaires, for instance). Not only that but more specific sleep-pattern data can be accessed, as REM and N phases cycles and alpha and beta EEG intrusions. The PSG and PSG Report sections of the dashboard reflect this (Figure 5.2). Regarding PSG, the dashboard displays information like total sleep time (TST), sleep efficiency and latency as they are simple, but common, giveaways for Insomnia. Besides those other data is shown, as number of sleep cycles and duration of the REM and N1 phases. According to Paiva and Pinto (2014), both increased N1 sleep and reduced REM sleep (measured as a percentage of TST) are characteristic of Insomnia. Sleep efficiency is marked red if it is under 85% (considered pathological), as are TSTs under 5 hours (300 minutes) and sleep latencies over 30 minutes. Conversely, very high efficiencies, low latencies and long TSTs are positive indicators, and thus marked green. The sleep cycles are also affected by the Insomnia symptoms, reducing in number from the four or more that are common. As such, if the patient only has one cycle, this parameter is marked red or orange if it has two to four.

PSG					
Efficiency	Total Sleep Time	Number of cycles	REM (%)	N1 (%)	Latency
93.7 %	383 minutes	3	10.7	8.1	4 minutes

PSG Report							
Perception of TST	Abnormal deep sleep	Alpha intrusion	Beta intrusion	Sleep Apnea	Restless Legs Syndrom	Periodic Limb Movement Disorder	Cardiac anomalies
Correct	No	Yes	Yes	low or moderate	No	No	No

Figure 5.2: PSG and PSG Report sections of the Insomnia Patient Data Dashboard.

Regarding the PSG Report section, the patient’s perception of TST, deep sleep abnormalities, alpha and beta intrusions are displayed (as they are also indicative of Insomnia). In this section, some of the detected sleep comorbidities, like Restless Legs Syndrome or Periodic Limb Movements are displayed, as are cardiac anomalies. The presence of any of those is marked red as any of them alone might justify the Insomnia symptoms.

Finally, there is the questionnaires section (Figure 5.3). In here, the scores for some of the most relevant questionnaires for Insomnia study are displayed. These are: Pittsburgh Sleep Quality Index (PSQI), Insomnia Severity Index (ISI), Epworth Sleepiness Scale (“Epworth”), Glasgow Sleep Effort Scale (“Glasgow”) and the Symptom Checklist-90 (SCL-90) revised version.

PSQI accesses the general sleep quality of the patient (the lower the score, the worst the perceived sleep quality) across three different scenarios – general days, work days and free days. The total score can give the doctor, at a glance, a estimation of the patient’s overall satisfaction with its sleep. Generally, a total score under 5 would indicate a very displeased patient (and, as such, is marked red). A very bad score on all three scenarios is a good indication towards an Insomnia diagnosis. On the other hand, a bad score on only one of them can suggest that there might just be external causes that explain the bad sleep experience (for example work-related anxiety or lack of sleep hygiene during the weekends). Other key aspect of PSQI is that it asks for the bed and wake up times for each scenario. This can also help to distinguish poor sleep hygiene from an actual Insomnia diagnosis. As such, the dashboard displays trios of bed time, wake up time and total score for each scenario.

PSQI (general)			PSQI (work days)			PSQI (free days)		
Bed time	Wake up time	Score	Bed time	Wake up time	Score	Bed time	Wake up time	Score
01:30	09:00	25	01:15	08:30	26	05:30	13:45	12

Insomnia		
Glasgow	Epworth	Severity
Score	Score	Index
13	11	11

SCL-90-R									
Total score	Somatization score	Obsessive-compulsive score	Interpersonal sensitivity score	Depression score	Anxiety score	Hostility score	Phobic anxiety score	Paranoid ideation score	Psychoticism score
5.62	0.62	0.91	0.67	0.92	0.72	0.44	0.60	0.41	0.33

Figure 5.3: Questionnaire section of the Insomnia Patient Data Dashboard. This section include data regarding the PSQI, Glasgow, Epworth, ISI and SCL-90 questionnaires.

The next questionnaire, ISI, evaluates the severity of the Insomnia symptoms and, as such, is a key questionnaire when there is a suspicion of an Insomnia diagnosis. ISI has four score brackets (up to 7, 8 to 14, 15 to 21 and over 21) and each one gives an indication of the severity of the Insomnia symptoms: no clinically significant insomnia, sub-threshold insomnia, clinical insomnia and severe clinical insomnia, respectively. Each one is colored differently (green, yellow, orange, red) so the doctor has a quick insight into the significance of the total score.

Epworth evaluates just how sleepy the patient is during the day. This characteristic is common among Insomnia patients and one of the possible day-time impairments that define Insomnia, according to American Academy of Sleep Medicine (2014), hence its inclusion on this dashboard. As with ISI, a bracket score system is defined (up to 9, 10 to 12, 13 to 16 and above 17). Each bracket score is colored accordingly, from the red, with scores below 9 to green, above 17.

The Glasgow questionnaire tries to determine just how hard it is to the patient to fall asleep (again, one of the defining features of Insomnia). Unlike the others, there is not a predefined bracket system for the total score. The general rule of thumb is: the higher, the worse (to a maximum of 21).

The last questionnaire, SCL90, is not a sleep questionnaire like the others. Instead, it accesses the severity of psychological problems that patients might have, and how much those affect their everyday life. Its questions are divided into nine sub-categories – somatization, obsessive-compulsive, interpersonal sensitivity, depression, anxiety, hostility, phobic anxiety, paranoid ideation and psychoticism – and each one has a corresponding score. These are used to determine if the patient already has changes in behaviour, which may differentiate a Insomnia patient from other who just slept badly some times (it is part of the definition of Insomnia to have day-time implications, caused by poor sleep quality). Some of the common ones are depression and anxiety. The dashboard displays the score for each category separately, as well as a overall score (as seen on Figure 5.3).

Two additional small sections (Figure 5.4), Actigraphy Report and Vitamin D are also present on the insomnia patient data dashboard. The actigraphy report can be useful to rule out Insomnia and attribute sleep difficulties to poor sleep hygiene. The light sensor on the actigraph can help to determine that the patient is exposed excessively to artificial light before going to bed, for example, or not enough to natural light in the morning. This exam can also determine if the patient naps too frequently (and then complains about its sleep at night). The vitamin D value can indicate to the clinician a need to adjust the patient's prescription, for instance. The lack of this vitamin is sometimes linked with Insomnia.

Actigraphy Report		Vitamin D (ng/mL)
Light exposure	Naps frequently	Value
Unsufficient natural light during the day	Yes	15
Longer exposure in the afternoon than in the morning		
Excessive exposure to artificial light in the evening		

Figure 5.4: Additional subsection on the Insomnia Patient Data Dashboard. It includes information about Actigraphy Report and Vitamin D values.

5.2 Insomnia Population Statistics

The second type of tool for patient characterization offered by SleepData is Population Statistics. To highlight the potential of this tool, and as a proof of concept, I have produced a characterization of the Insomnia population (using only the tools available in SleepData).

To gather data from a minimally relevant group of patients, I created a script to import data that was already available at CENC, in the form of excel spreadsheets. Each spreadsheet column was mapped to the SleepData's structure and loaded to the MongoDB database. In total, data from 100 insomnia patients were imported. Most patients had some clinical notes data (registered through anamnesis), PSG data and the results of four questionnaires: Glasgow, PSQI, ISI and SCL-90.

5.2.1 Demographic Data

The first type of data that is available on SleepData's population statistics feature is the general demographic data that includes statistics regarding the age and gender of the patients. From the insomnia population, 97 patients have age and gender information. Regarding gender, 64.9% are females and 35.1% males (see Figure 5.5). The higher number of women, when compared with men, is expected even though the observed value is considerably higher than some of the literature – a review article by Zhang and Wing (2006), covering 31 studies, calculated the risk ratio as 1.41 for females vs males, which would translate on a 58.51% female population. As discussed on Chapter 2, this may be due, in part, to the fluctuating hormone levels (such as melatonin and cortisol) or even due to pregnancies, which reduce quality of sleep and taking care of the newborn baby afterwards. Considering the data from all the Portuguese population (consulted on PORDATA), in 2015, women were 52.7% of the population which further reinforces the idea of the gender skew within the insomnia population.

Regarding their age, almost all patients are over 40 years old (85.4%), which reflects the documented increase of insomnia incidence with age. For instance, when comparing with the 2015 demographic data, only 56.6% of the Portuguese population was over 40 years old. Considering the insomnia population over 60, it adds up to 36% against 27.2% when considering the entire population.

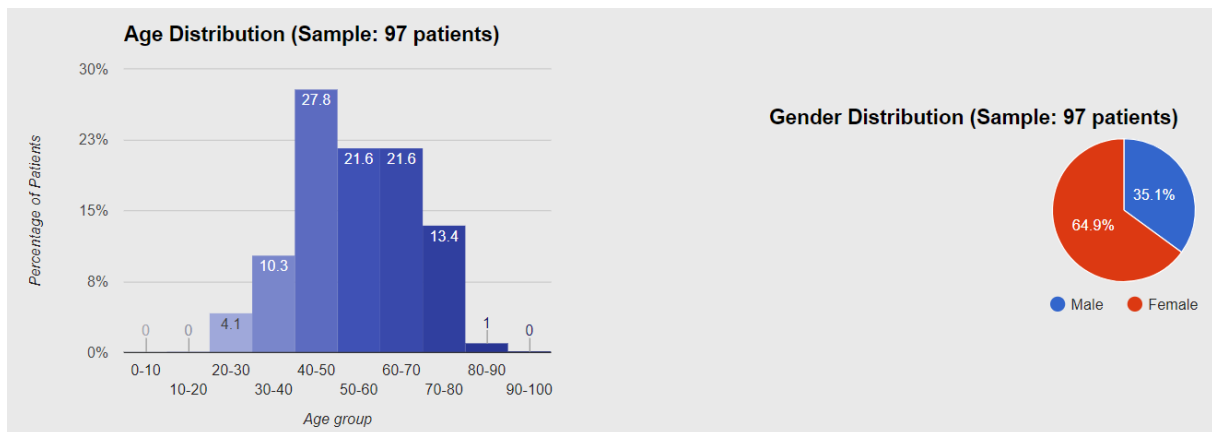


Figure 5.5: Age and gender distribution of the Insomnia patients.

5.2.2 Anamnesis

One of the most relevant data gathered by practitioners in appointments, are the patient's complaints regarding day-time activity, caused by poor sleep quality, which are characteristic of an insomnia patient. Five of the more relevant complaints – attention deficit, the reduction of executive functions, fatigue, irritability and lack of memory – were registered for each of the patient on this data set and the corresponding results are visible in Figure 5.6. Attention deficit is the most common of these complaints, with 72% of the insomnia population reporting it and around half of the patients suffer from fatigue, irritability and memory problems. The most severe of the complaints, reduced executive functions, affects 26% of the patients, which highlights the degree of impairment that Insomnia may provoke. Furthermore, most patients (76%) report having more than one of these five complaints, with only 6% not experiencing any.

Not only do the Insomnia patients experience severe symptoms, but they also have, most of the times, comorbidities. According to the Principles and Practices of Sleep Medicine, by Kryger et al. (2011), 86.1% of the insomnia population has other medical conditions (i.e only 13.9% do not have comorbidities). Within this dataset, only 10% of the population did not have any other disorder, which reinforces that idea (as visible on Figure 5.6). On that same figure, it is visible that 26% of the patients have 2 or more comorbidities. Regarding their distribution, anxiety is the most common comorbidity, followed by depression (affecting 51 and 24% of the population, respectively). Sleep apnea is also a common comorbidity for insomnia – observed in 9% of the patients.

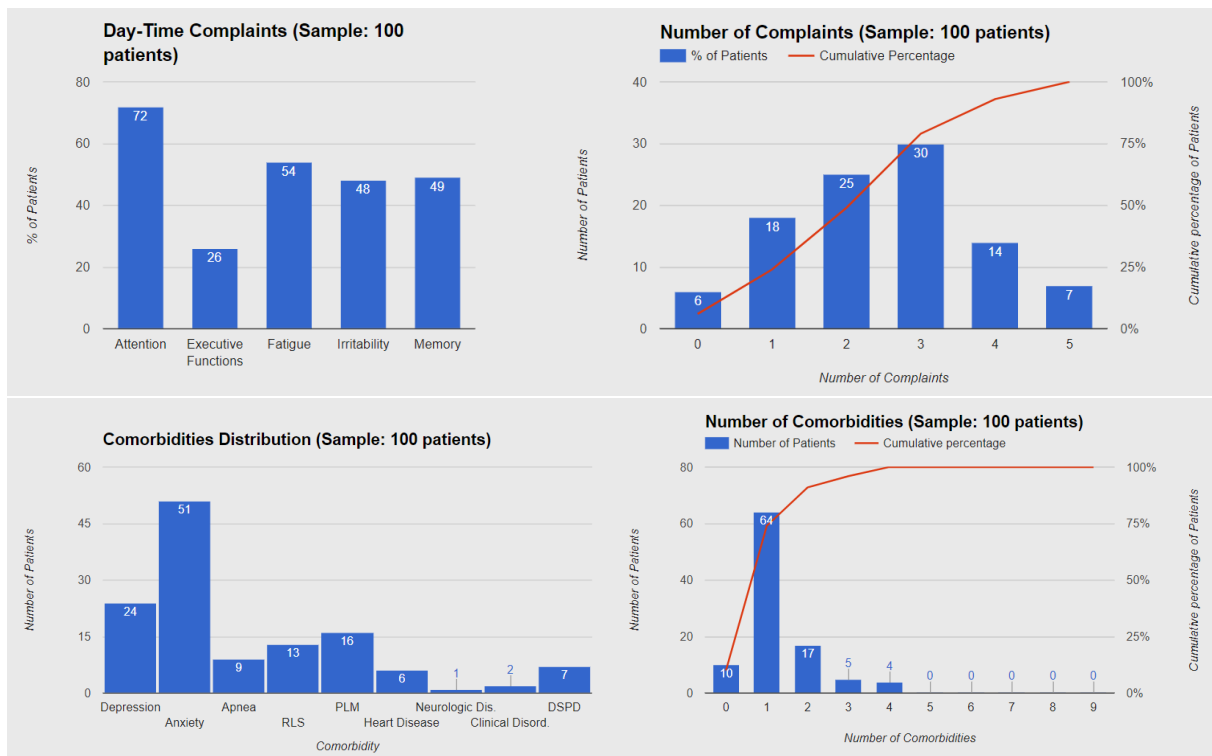


Figure 5.6: Prevalence of day-time complaints caused by poor sleep quality (top) and distribution and prevalence of comorbid disorders (bottom).

5.2.3 Polysomnography Data

Besides the notes from the practitioner, another important source of information is the polysomnography (PSG). From the provided data set, simple, yet valuable statistics like TST and sleep latency were gathered. Unlike self-reported TST and sleep latencies, which can be several hours apart from the real value, these are accurate (even more than by actigraphy). The self-reported sleep latency values are specially inaccurate when dealing with insomnia patients (with inherently long latencies).

Regarding TST (Figure 5.7), the average value (visible on hover on the SleepData's graph) is 360.3 minutes or approximately 6 hours. This is clearly below the value determined by the American Academy of Sleep Medicine (AASM), which is 7-8 hours. The patient on the third quartile (75% of the patients) slept only 409 minutes (6h49min), which is still below the recommended. This little sleep can have severe repercussions in just a few consecutive days. A study by Spiegel et al. (1999), where 11 males were subjected to only 4h of sleep for 6 days, found harmful impacts on carbohydrate metabolism and endocrine function, similar to those cause by ageing, which leads to believe that this kind of sleep debt can lead to increased severity of age-related chronic disorders.

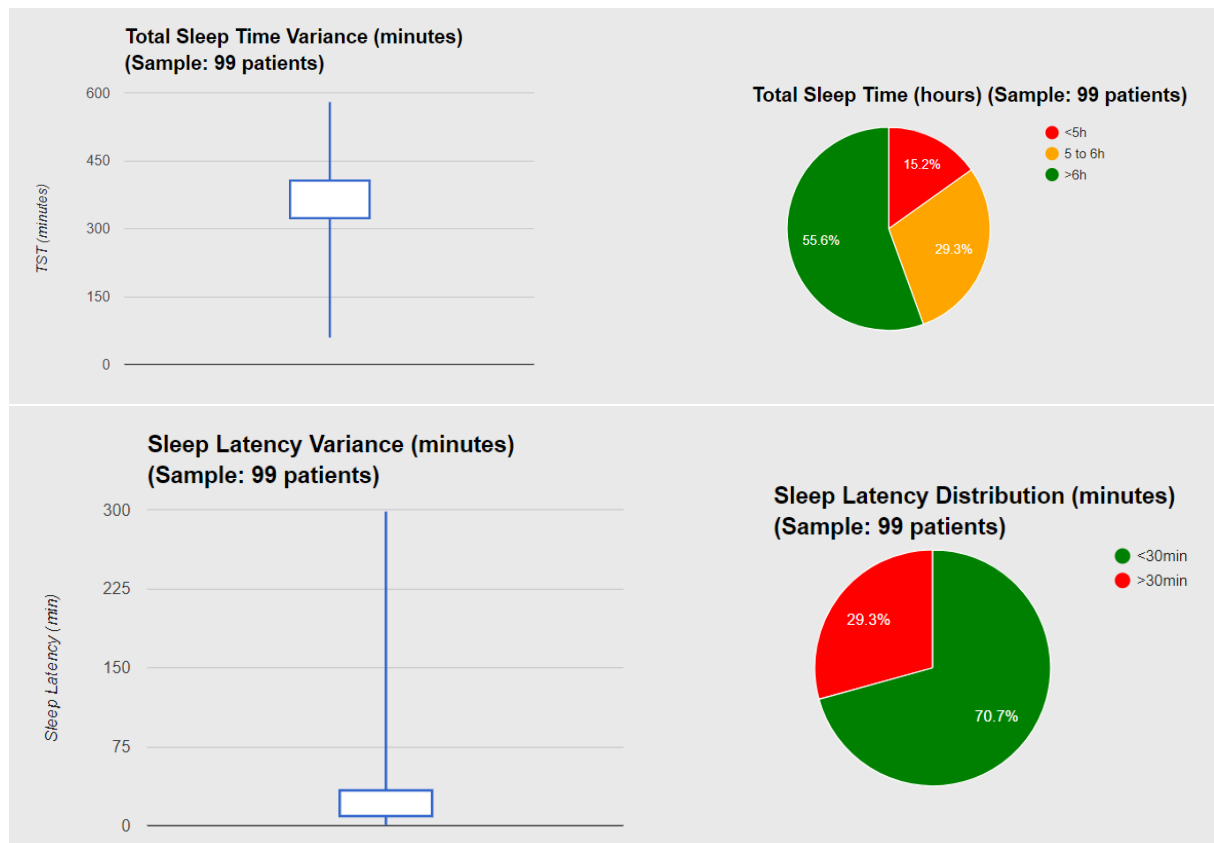


Figure 5.7: Distribution of total sleep time values (top) and sleep latencies (bottom), derived from PSG.

Not only is these patients' sleep short, but also sleep onset delayed. In Figure 5.7 it is visible that 29.3% of them take more than 30 minutes to fall asleep. The fact that the average latency is also about 30 min (29min28s) shows just how high some of the patient's latency is. Regarding sleep efficiency, 60% of patients are under 85% efficiency.

After the PSG exam, which is recorded throughout an entire night, the patient is asked about its opinion regarding the sleep, including how long they thought they slept (i.e. perceived TST). The Figure 5.8 represents that, with graphs of the distribution of the perceived TSTs. The patients' estimates serve as an indication of just how reliable their description of the symptoms and characteristics of their sleep are and serves as a means to distinguish between accurate and exaggerated patients.

Looking at the results, it is clear that Insomnia patients are pessimists, in terms of their perceived sleep quality (i.e they perceive their sleep to be worse than it is). In fact, the average perceived TST is 270 min, against the 360 min that were actually recorded with the PSG exam. Considering patients on the very severe range of measured TST (with less than 5 hours of sleep per night), they are 15.2%, even though more than half of the patients perceive their sleep to be

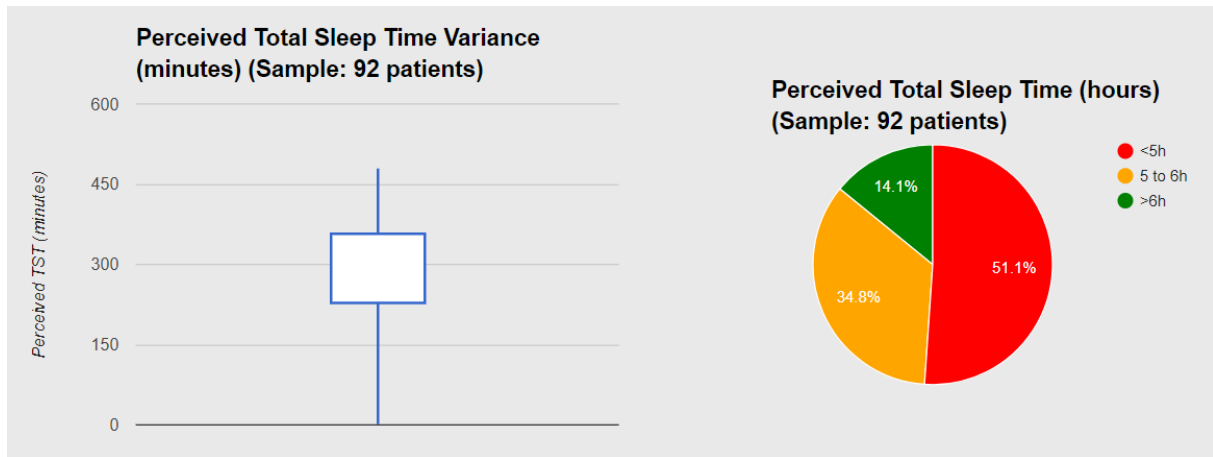


Figure 5.8: Distribution of the perceived total sleep times.

that bad. Considering the difference in values of perceived TST and actual TST for each patient, visible on the Figure 5.9, we can create three categories for the patients: optimists, pessimists and correct. An optimist estimates his TST to be more than 30 min over the actual value and a pessimist, the opposite (perceived TST more than 30 minutes lower than the measured). An estimate is considered correct when within a 30 min range of the actual value. From the studied population, 54.9% of them are pessimists and 40.7% of them were quite accurate on their estimates. This leaves just 4.4% as optimists. These values reinforce the value of PSG as a key exam on the study of Insomnia as most patients are not aware of the severity of their symptoms, and can be excessive on their descriptions.

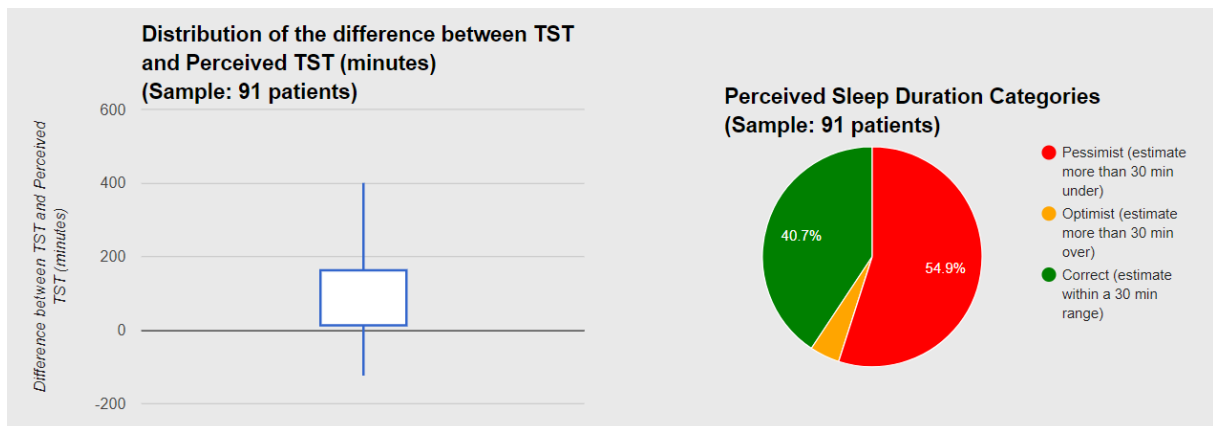


Figure 5.9: Differences between perceived TST and actual TST, measured through PSG.

5.2.4 Sleep Questionnaires

Even though objective data from exams like PSG and practitioner's findings is very important, it is also both expensive and time consuming to acquire. A PSG exam lasts for an entire night and requires that the patient travels to the clinic. Furthermore, its setup is lengthy and

can be uncomfortable for the patient. This way, it is valuable to do analysis with the data that can be reported by the patients themselves. Within this study, most patients filled four questionnaires: Pittsburgh Sleep Quality Index (PSQI), Insomnia Severity Index (ISI), Glasgow Sleep Effort Scale and the Symptom Checklist-90 (SCL90) revised version.

5.2.4.1 Pittsburgh Sleep Quality Index

The PSQI questionnaire is a set of questions regarding sleep quality over a 1-month period. Each question is scored, being the final score a value between 0 and 21, where 0 represents a good sleep quality and 21 the worst. The defined cut-off score for this questionnaire is 5 points. On the Figure 5.10 it is visible that just 2.2% of the patients have scores under 5, which reveals just how degraded is the insomnia patients' sleep quality (or at least their perception of it).

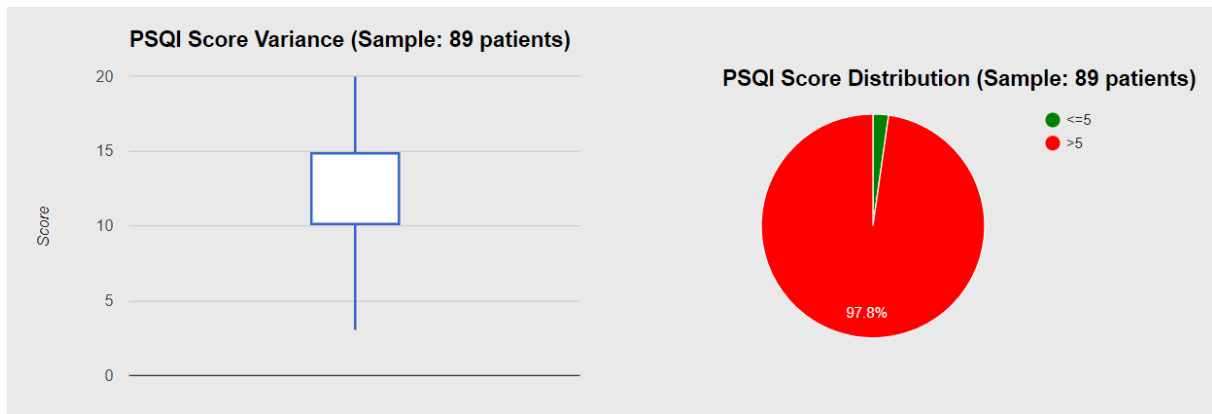


Figure 5.10: Distribution of PSQI scores, on the left, and classification of the patients based on the questionnaire's cut-off points.

The average score is 12.64 which is characteristic of people that are very displeased with their sleep quality. Grandner et al. (2006) found that the average score for PSQI to be 4.07 among a young population (average age of 23) and 3.92 among an older population (average age of 66). Both of these values are inside the 5 points cut-off value and well below the average within this insomnia population.

5.2.4.2 Insomnia Severity Index

The ISI questionnaire is comprised of seven quick questions which intend to do an estimate of the severity of the insomnia symptoms (as the name implies). The questionnaires' cut-off values are defined at 7, 14, 21 and 28, inclusive. These indicate a patient with no clinically significant insomnia, sub-threshold insomnia, clinical insomnia and severe clinical insomnia, respectively.

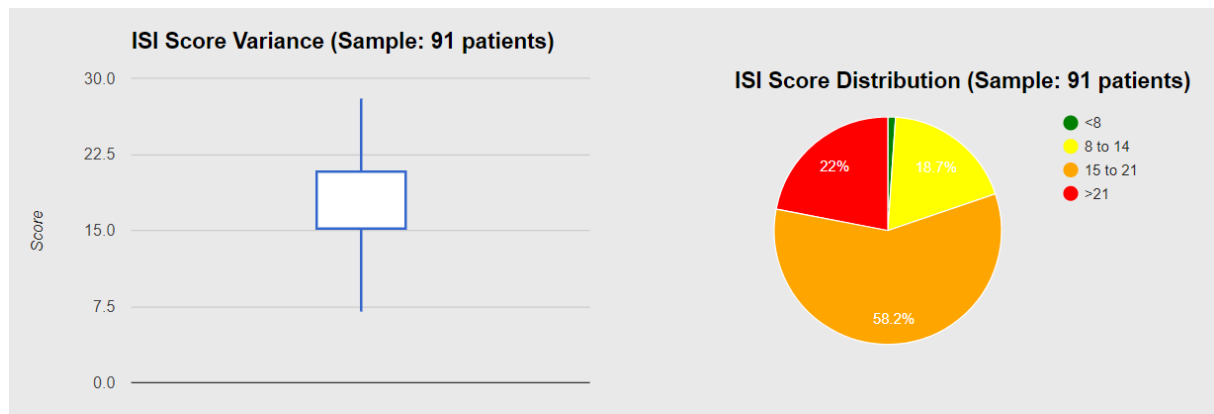


Figure 5.11: Distribution of ISI scores, on the left, and classification of the patients based on the questionnaire's cut-off points.

Figure 5.11 illustrates the distribution of scores amongst the insomnia population. As expected, most patients (98.9%) have scores of 8 or above, consistent with the insomnia diagnosis. 22% have scores higher than 21, which indicates very severe insomnia symptoms. In terms of the average, it is 18, which belongs to the score bracket that indicates clinical insomnia.

5.2.4.3 Glasgow Sleep Effort Scale

The Glasgow Sleep Effort Scale, or "Glasgow" for short, accesses just how hard it is for the patient to sleep. It consists of seven sentences describing a type of difficulty/displeasure regarding sleep and the person has to say if it applies, if neutral or if it does not. Each option has a value (3, 2, 1, respectively), giving rise to a final score between 7 and 21 points. There are no cut-off values defined for this questionnaire – the higher the value, the more difficult it is for the patient to sleep.

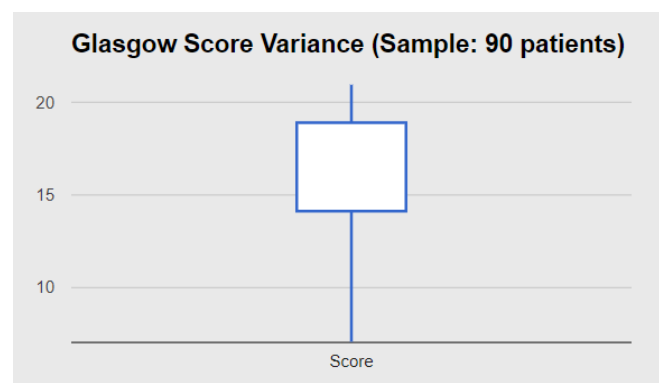


Figure 5.12: Distribution of Glasgow scores.

Among this population, the average question score is 2.25 which mean that the patients, on average, agree with most sentences or, at least, they are neutral. This is expected as it

is, by definition, characteristic of insomnia patients to have difficulties regarding sleep. Figure 5.12 shows how the scores are distributed. A study from Broomfield and Espie (2005), found the average question score to be 1.37, from a sample of 102 normal people, without reports of abnormal sleep. In contrast with the Insomnia population, healthy subjects tend to disagree with most sentences. The study suggests a cut-off value of 9 (total average score), to distinguish normal patients from the insomnia ones. Within SleepData, the average total score was 15.76 and, as such, is according with that suggestion.

5.2.4.4 Symptom Checklist-90

The previously described questionnaires focus on some aspect of sleep, but there are other types of questionnaires that are relevant to characterize an Insomnia population. As mentioned before, the Symptom Checklist-90 (SCL90), revised edition, is one of those as it accesses the severity of psychological problems that patients might have and how much those affect their everyday life. It is composed of 90 questions, divided into 9 sub-categories – somatization, obsessive-compulsive, interpersonal sensitivity, depression, anxiety, hostility, phobic anxiety, paranoid ideation and psychoticism – each one with a corresponding score. On SleepData' Population Statistics, the score for each component is displayed against the average for a healthy population (see Figure 5.13). This average was calculated by considering the results for 302 Portuguese people that did not report any sleep-related complaints.

From the main 9 categories, somatization, obsessive-compulsive, depression and anxiety are the ones where the biggest difference between Insomnia and healthy population is observed. The increased prevalence of anxiety and depression symptoms, in this case self-reported, go in hand with the practitioner's diagnosis as depression and anxiety were two of the most diagnosed comorbidities (affecting 24% and 51%, respectively). The somatization category refers to somatic symptoms, like headaches, dizziness, chest pain and others, as a consequence of or to communicate psychological distress. These are in agreement with the ICSD's definition of insomnia, which requires the patients to have daytime impairments. Finally, the obsessive-compulsive category refers to behaviours like being excessively concerned, cautious, fearing not being able to do or remember certain things. These can be caused by the insufficient total sleep time inherent with insomnia.

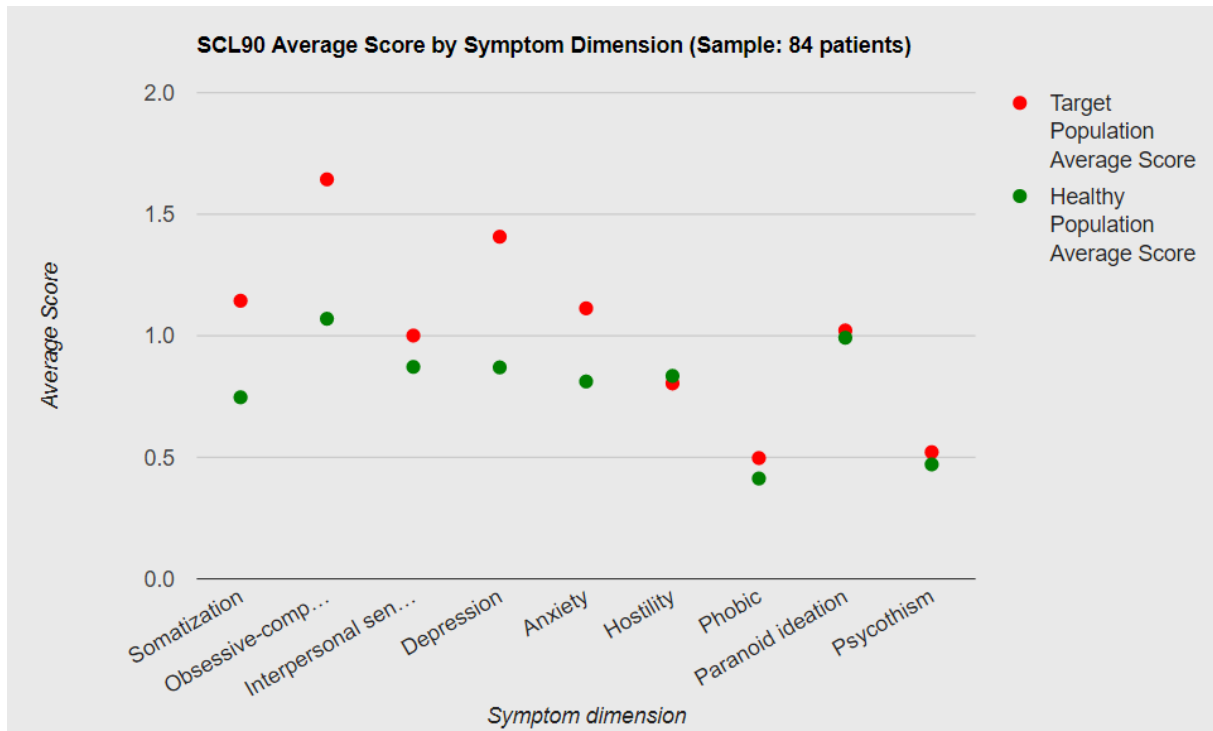


Figure 5.13: Average score for each of the symptom's dimension of the SCL90 questionnaire, versus the score of an healthy population.

5.3 Overview

At the end of the chapter I can discuss the two SleepData data visualization tools, that work symbiotically to characterize insomnia. The insomnia dashboard that allows a clinician to quickly observes the most relevant data regarding a patient, divided by source, and highlighted according to specific criteria and the population statistics, which show a set of charts (also organized by categories) that characterize the entire population as a whole.

Regarding the dashboard, a further improvement would be to allow customization by both, changing fields to be displayed and to change the coloring rules. The fields could also be editable through the dashboard, so slight tweaks could be made to the data as data is studied by the clinician. Nonetheless, its current state covers the data sources that were determined to be the most relevant (polysomnography, anamnesis and questionnaires), displays them by sections and highlights the most relevant (validated by Dr. Teresa Paiva) according to defined rules.

In terms of the population characterization, it highlights several interesting points – predominantly female and older population, with severe consequences caused by the disorder and highly comorbid (with anxiety and depression as the most common). Most of the population experience sub-optimal TSTs, as well as high sleep latencies. This contributes to the general discontent towards sleep quality and accentuates the exaggerated nature of the population, that

describes its own sleep as worse as it is in reality (in terms of perceived TSTs). The results of the questionnaires go in hand with this, as they are all characteristic of a very displeased population. SCL90 corroborates the doctors' diagnosis as the most reported psychological symptoms are anxiety and depression.

However, the population size used in this study is quite small for a characterization of this type, since only 100 patients were considered and some patients did not had complete data sets (missing some questionnaire responses, PSG or other data). A data set from healthy subjects would be ideal for comparison purposes and so more concrete conclusions could be drawn. Despite that, this simple characterization proves the point that SleepData is a capable tool to study populations.

Conclusions and Future Work

The development of the SleepData platform was motivated by the lack of both a central hub to store, visualize and analyze data regarding sleep disorders and a place where anyone could fill sleep questionnaires. As such, the SleepData concept emerged as a solution for both problems.

The first step to achieve these goals was to assess which data sources are most relevant to the users of the platform. Based on relevant literature and using a reputable sleep clinic (CENC) as an example these could be determined – actigraphy, melatonin and vitamin D exams, polysomnography, anamnesis (clinical notes) and the sleep questionnaires Epworth, MCTQ, PSQI, MEQ, Glasgow, ISI and SCL90. For each source, it was determined how the data is acquired and stored and which parameters are the most relevant. From the actigraphy and polysomnography exams, the outputs and file handling methods were studied, for each of the equipment brands/models. Polysomnography proved to be the most complex. Both because of the amount of relevant data acquired as well as due to the its proprietary file types that are variable from brand to brand and even from product to product, within the same brand. To make matters worse, the data files in each patient record are usually very numerous (dozens to few hundreds) and the total file size massive (up to a few gigabytes), mainly due to the additional video recording. Conversely, actigraphy is pre-processed by their respective brand's software, producing a structured report. With this considered, two scripts were developed (for two different brands of equipment) to automatically parse the files uploaded to SleepData and save the relevant parameters.

To aid the visualization of the information from all of these data sources, patient data dashboards were implemented, with rules to highlight certain data to alert the clinician for important parameters, as well as graphs to characterize sub populations in SleepData (called the "Population Statistics" feature).

To map each of the parameters (avoiding ambiguity and improving interoperability with other systems), international, reputable coding systems were needed. The pair SNOMED CT and LOINC was regarded as the best alternative, even though it had shortcomings, specially when regarding general/demographic data, such as religion, ethnicity, occupation or degree of

education. For those, separate coding systems were determined: HL7 v3 Code System ReligiousAffiliation, to code religions, a custom SNOMED terms set for ethnicities/racial groups, ISCO-08 for occupations and ISCED for education. I also concluded that a custom coding system can be necessary for each clinic (CENC's own was created) besides the international ones, so it is easier for the practitioners to identify their variables.

HL7's FHIR was chosen to map the database structure, as it allow to add fields as needed, provides structure diagrams and JSON examples and provides an API for data transfer. In terms of the actual platform development, the MEAN Stack (MongoDB, Express, Angular.js and Node.js) was found to be the most suited. The MongoDB database is perfect for FHIR-based data, with its NoSQL architecture. The Node.js environment is modern, with easy setup of both server and front-end environments (all based on JavaScript) and has a huge library that is free for all users. In fact, throughout all the SleepData's development there was a focus on modern, free, open-source software and tools.

SleepData implements a mandatory access control system to clinical data to ensure data privacy and security. Three classes or types of user were defined: regular users, that can be created by anyone, with a free account, and can only fill questionnaires; professional users that are clinic-specific and can add information of all kinds, including its patients', manage patients and see population statistics; a administrator that can assign roles to regular users, or, on the other hand, remove them or even delete their accounts.

Two populations studies have been conducted with SleepData, one on 145 DSPD patients, by Pinho (2017), and a study on 100 Insomnia patients reported in this dissertation. The insomnia characterization revealed that the population is predominantly female and older, and with severe consequences caused by both the insomnia and comorbidities (with anxiety and depression as the most common). It also became clear that most of the patients slept poorly, in term of total sleep time and sleep latency (i.e take very long to fall asleep). Non surprisingly, this lead to a very displeased population, that exaggerates their symptoms and describe their sleep as worse than it actually is.

SleepData is a platform that can host patient records from multiple clinics, that are independently managed and accessed, offering the necessary tools to visualize the data on both a patient-by-patient or a entire population manner. Being web-based, it has the advantage of being available on any browser, including on mobile devices. This can be an advantage for the clinicians or technicians of the sleep clinics as it allows them to quickly check on patient data from their personal device. This considered, many features can still be improved or even new

ones added.

One of the areas where improvements need to be made is on security. Data should be saved on an encrypted manner and two factor authentication strategies should be implemented to guarantee the security of the patient's data. MongoDB provides a *Encryption at rest* functionality that addresses the first point as it encrypts all data so it is only accessible by parties with the decryption key. This would be essential if SleepData had data on the cloud. The main issue with this solution is that it is only available in the enterprise version of MongoDB, which is paid.

Besides those, improvements can be done regarding additional data sources as many other exams and general data can be adopted on SleepData. These include, for example, genetic data, blood exams or general patient data such as diet, exercise, leisure activities and hobbies. To obtain this kind of data there would have to be integration between SleepData and other equipment, such as wearables, smartphones, respiratory-aid equipment (CPAP or BPAP) or even EHRs and even other health platforms.

To improve adoption and usability, multiple languages could be adopted – a user would select his preference on the homepage which would affect all interfaces, including forms and sleep questionnaires. To also aid adoption, a "data import" feature could be added. On a user-friendly interface users would be able to import excel or SPSS data and then map it to the suggested FHIR's fields on SleepData. This would lead to more adoption from clinics, as they could use the platform more easily, importing data without much hassle. If more fields or codes were to be added there should also be a possibility to do so through the interface.

An interface to add the patient's diagnosis should also be added to SleepData. This way, the population statistics data could be easily filtered by disorder. To code these, the ICSD3 (American Academy of Sleep Medicine, 2014), that only codes sleep disorders, and ICD (World Health Organization, 2010), a more general disorder classification system, should be adopted.

In terms of data visualization, the dashboards could become more dynamic, giving professional users the possibility of creating their own, with the fields they desire (and possibly target other disorders). Clinicians would also be able to edit their patient's records through the dashboard. Regarding the population statistics, similar improvements can still be made – custom filters, designed by the user, so that the displayed graphics adapt to each one of their needs. Still regarding population statistics, there should be an option for the clinics to allow their anonymized patient data to be visible to other clinics. This way a new user type could be defined, that would have access to general statistics across all clinics (this option would have to be thoroughly analyzed from a data privacy point of view).

Finally, a more detailed study, addressing one or more relevant medical questions with a larger patient population and healthy patient records, could be undertaken with the SleepData platform.

Bibliography

- American Academy of Sleep Medicine (2005). *The International Classification of Sleep Disorders, 2nd edition*.
- American Academy of Sleep Medicine (2014). *The International Classification of Sleep Disorders, 3rd edition*.
- Broomfield, N. M. and Espie, C. A. (2005). Towards a valid, reliable measure of sleep effort. *Journal of Sleep Research*, 14(4):401–407.
- Budhiraja, R., Roth, T., Hudgel, D. W., Budhiraja, P., and Drake, C. L. (2011). Prevalence and Polysomnographic Correlates of Insomnia Comorbid with Medical Disorders. *Sleep*, 34(07):859–867.
- Chokroverty, S. (2014). Key points and hot topics in sleep medicine. In *O sono e a medicina do sono*, chapter 2.1, pages 201–210. Editora Manole.
- Comissão Nacional de Protecção de Dados (2015). Deliberação nº.1704/2015.
- Figueiral, M. (2015). CODEGOM : A governance model for clinical ontologies. Master’s thesis, Instituto Superior Técnico.
- Ford and Kamerow (1989). Epidemiologic study of sleep disturbances and psychiatric disorders: An opportunity for prevention? *JAMA*, 262(11):1479–1484.
- Grandner, M. A., Kripke, D. F., YOON, I.-Y., and Youngstedt, S. D. (2006). Criterion validity of the pittsburgh sleep quality index: Investigation in a non-clinical sample. *Sleep and biological rhythms*, 4(2):129–136.
- Guyton, A. C. and Hall, J. E. (2006). *Textbook of Medical Physiology*. Elsevier Ltd, 11th edition.
- Hauri, P. J. and Wisbey, J. (1992). Wrist actigraphy in insomnia. *Sleep*, 15(4):293–301.
- International Labour Office (2008). International Standard Classification of Occupations (ISCO-08).

- Kryger, M. H., Roth, T., and Dement, W. C. (2011). *Sleep Medicine Principles and Practice*. Elsevier.
- Lawton, G. (2005). Lamp lights enterprise development efforts. *Computer*, 38(9):18–20.
- Lee-Chiong, T. (2006). *Sleep: A Comprehensive Handbook*. John Wiley & Sons Inc.
- Meia-Via, M. S., Marques, D. R., Espie, C. A., da Silva, C. F., and Allen Gomes, A. (2016). Psychometric properties of Glasgow Sleep Effort Scale in Portuguese language. *Psychol Assess*, 28(3):e12–18.
- Morin, C. M., Belleville, G., Bélanger, L., and Ivers, H. (2011a). The Insomnia Severity Index: Psychometric Indicators to Detect Insomnia Cases and Evaluate Treatment Response. *Sleep*, pages 18–20.
- Morin, C. M., LeBlanc, M., Bélanger, L., Ivers, H., Mérette, C., and Savard, J. (2011b). Prevalence of insomnia and its treatment in Canada. *Canadian Journal of Psychiatry*, 56(9):540–548.
- Ohayon, M. M. and Paiva, T. (2005). Global sleep dissatisfaction for the assessment of insomnia severity in the general population of Portugal. *Sleep Medicine*, 6(5):435–441.
- Paiva, T. and Anderson, M. L. (2014a). *O sono e a medicina do sono*. Editora Manole.
- Paiva, T. and Anderson, M. L. (2014b). *O sono e a medicina do sono*. Editora Manole.
- Paiva, T. and Penzel, T. (2011). *Centro de medicina do sono - Manual prático*. Lidel.
- Paiva, T. and Pinto, H. R. (2014). Clínica e terapêutica das insónias. In *O sono e a medicina do sono*, chapter 2.5, pages 230–262. Editora Manole.
- Paiva, T. and Silva, F. H. L. (2014). Gênese das atividades do eletroencefalograma de sono: neurofisiologia do eletroencefalograma de sono. In *O sono e a medicina do sono*, chapter 1.7, pages 76–88. Editora Manole.
- Palagini, L. and Rosenlicht, N. (2011). Sleep, dreaming, and mental health: A review of historical and neurobiological perspectives. *Sleep Medicine Reviews*, 15(3):179–186.
- Pallesen, S., Sivertsen, B., Nordhus, I. H., and Bjorvatn, B. (2014). A 10-year trend of insomnia prevalence in the adult Norwegian population. *Sleep Medicine*, 15(2):173–179.
- Phillips, B. and Gelula, R. (2006). Sleep-Wake Cycle: Its Physiology and Impact on Health. *National Sleep Foundation*.

- Pinho, B. (2017). SleepData - Sleep Disorders Clinical Platform: Delayed sleep phase disorder. Master's thesis, Instituto Superior Técnico.
- Sarsour, K., Morin, C. M., Foley, K., Kalsekar, A., and Walsh, J. K. (2010). Association of insomnia severity and comorbid medical and psychiatric disorders in a health plan-based sample: Insomnia severity and comorbidities. *Sleep Medicine*, 11(1):69–74.
- Schulz, H. (2008). Rethinking sleep analysis. *Journal of Clinical Sleep Medicine*, 4(2):99–103.
- Shneerson, J. M. (2005). *Sleep Medicine - A Guide to Sleep and its Disorders*. 2nd edition.
- SNOMED CT (2016). SNOMED CT Starter Guide.
- Soehner, A. M., Kaplan, K. A., and Harvey, A. G. (2014). Prevalence and clinical correlates of co-occurring insomnia and hypersomnia symptoms in depression. *Journal of affective disorders*, 167:93–97.
- Spiegel, K., Leproult, R., and Cauter, E. V. (1999). Impact of sleep debt on metabolic and endocrine function. *The Lancet*, 354(9188):1435 – 1439.
- UNESCO - Institute for Statistics (2011). *International Standard Classification of Education (ISCED)*.
- World Health Organization (2010). *The International Statistical Classification of Diseases and Related Health Problems, 10th revision*.
- Zhang, B. and Wing, Y.-K. (2006). Sex differences in insomnia: a meta-analysis. *Sleep*, 29(1):85–93.

A

List of Parameters by Collection

Table A.1: List of the SleepData's collections parameters and respective FHIR resource.

Patient — <i>Patient</i> Resource
Name
Marital status
Contact information (phone, address and email)
Gender
Nationality
Birthday
Deceased (boolean)
Language preference
Responsible person in case of minors and patients unable to take care of themselves (name, contact and relationship with the patient)
General practitioner (name, identification and institution)
Professional — <i>Practitioner</i> Resource
Name
Contact information (phone, address and email)
Gender
Birthday
Deceased (boolean)
Language preference
Qualification
ProfessionalRole — <i>PractitionerRole</i> Resource
Period of employment
Name and identification number
Organisation

Role
Speciality
Location
Contact information (of workplace)
Schedule
ClinicalNotes — <i>Observation</i> Resource
Number of separations
Family conflicts (boolean)
Car accidents (boolean)
Education level (ISCED2011 levels and codes)
Academic underachievement (boolean)
Occupation (ISCO groups and codes)
Works in shifts (boolean)
Absenteeism (boolean)
Wake Up Time during the work days (minimum, average and maximum values)
Wake Up Time during the weekends (minimum, average and maximum values)
Getting up time during the work days (minimum, average and maximum values)
Getting up time during the weekends (minimum, average and maximum values)
Bed time during the workdays (minimum, average and maximum values)
Bed time during the weekends (minimum, average and maximum values)
Breakfast time (minimum, average and maximum values)
Lunch time (minimum, average and maximum values)
Snack time (minimum, average and maximum values)
Dinner time (minimum, average and maximum values)
Supper time (minimum, average and maximum values)
Skips meals (boolean)
Bad meal schedules (boolean)
Age at first symptoms
Traumatic event (boolean)
Stress and/or depression (boolean)

Triggering factor of the symptoms
Family history
Alcohol consumption (NIAAA drinking levels)
Narcotics usage (boolean and type of usage)
Smoker (boolean, type, quantity and number of years)
Comorbidities
Body weight
Body height
Irregular Schedules (boolean)
Work Schedule (type, beginning and end time, number of hours per day and week, number of days per week)
Long work shifts (boolean)
Stress at the workplace (type)
Physical exercise (boolean, number of days per week, usual time of exercise)
Excessive working hours (boolean)
Travels frequently(boolean)
Light exposure (type if abnormal)
Sedentary lifestyle (boolean)
Too many responsibilities (boolean)
Procrastination (boolean)
Takes homeopathic medicines (boolean)
Sexual orientation
Ethnic group
Religious affiliation
Complaints of reduced cognitive ability (type)
Complaints of daytime impairments (type)
Medication — <i>Medication</i> Resource
Amount, type, usage, manufacturer, form, active ingredient and its amount for each drug
Actigraphy — <i>Observation</i> Resource
Equipment used, including identification number and brand

Number of days used
Total sleep time, in minutes (minimum, average and maximum values)
Bet time (minimum, average and maximum values)
Getting up time (minimum, average and maximum values)
Sleep onset time (minimum, average and maximum values)
Sleep efficiency (minimum, average and maximum values)
Wake-time after sleep onset (minimum, average and maximum values)
Number of awakenings (minimum, average and maximum values)
ActigraphyReport — <i>Observation</i> Resource
Irregular schedules (boolean and type)
Circadian rhythm profile
Reduced motor activity during the day (boolean)
Increased motor activity during the evening (boolean)
Increased motor activity during sleep (boolean)
Sleepiness episodes during the day (boolean)
Light exposure (type if abnormal)
Matching diary and actigraphy schedules (boolean)
Altered total sleep time (type)
PSG — <i>Observation</i> Resource
Setup (EEG configuration)
Total Sleep Time
Time in bed
Sleep Onset Time
Wake Up Time
Sleep Efficiency
REM Latency
N1 Sleep Phase (%TST)
N2 Sleep Phase (%TST)
N3 Sleep Phase (%TST)
N3 Temporal Profile

REM Sleep Phase (%TST)
REM Temporal Profile
Number of sleep cycles
Duration per cycle Last sleep cycle is REM (boolean)
Apnea Hypopnea Index
Oxygen Desaturation Index
Minimum Oxygen Saturation
Snoring (%TST)
Periodic Sleep Movements
Micro-awakening Index
Fragmented Sleep
Suggested Immobilization Test
PSGReport — <i>Observation</i> Resource
Perception of sleep duration
Insomnia Criteria
Reduction of sleep duration
Normal Deep Sleep
Normal REM Sleep
Reduced REM sleep latency
Abnormal Deep Sleep temporal profile
Abnormal REM temporal profile
REM atonia
Alpha wave intrusion
Beta wave intrusion
Cardiac anomalies
Sleep apnea
Periodic Limb Movement Disorder
Abnormal behaviours during the night
NightDiary — <i>Observation</i> Resource
Time at Lights off

Time at Lights on
Patient's perspective of his/hers sleep (perceived sleep time and quality)
Trips to the bathroom
Meals during the night
Vocalizations
Crisis
Urination in bed (boolean)
Slept with someone
use of Continuous Positive Airway Pressure (CPAP)
Malfunctions of the equipment
Suggested immobilisation test result
DLMO
DLMO is not computable (boolean and reason)
Melatonin Level (time and quantity for each measure)
Time at DLMO

B Clinical Data Dashboards

Table B.1: Insomnia dashboard parameters, variables and the respective colour code.

Source	Colour			
	Parameter	Green	Yellow or Orange	Red
PSG	Sleep Efficiency	$\geq 85\%$	-	$<85\%$
	Total Sleep Time (TST)	$>6h$	5 to 6h	$<6h$
	Sleep Onset Latency	≤ 30 minutes	-	>30 minutes
	Number of Cycles	>4	2 to 4	<2
	REM Sleep (%TST)	no colour code		
	N1 Phase (%TST)	no colour code		
PSG Report	Patient's Perception of TST	Correct (30 min range)	Optimistic	Pessimistic
	Abnormal Deep Sleep	False	-	True
	Alpha waves intrusion	False	-	True
	Beta waves intrusion	False	-	True
	Sleep Apnea	No	Light/ Moderate	Severe
	Restless Legs Syndrome	False	-	True
	Periodic Limb Movement Disorder	False	-	True
	Cardiac Anomalies	False	-	True
Actigraphy Report	Exposure to solar light	Good	-	Irregularities
	Frequent napping	False	-	True

Vitamin D Exam	Measures Vit D	>30ng/mL	10 to 30 ng/mL	>10mL
Clinical Notes	Comorbidities	False	-	True
	Age at first symptoms	no colour code		
	Complaints when waking up	False	-	True
	Daytime impairments	False	-	True
	Cognitive complaints	False	-	True
	Sleep-related fears	False	-	True
	Family history	no colour code		
	Triggers	no colour code		
	Traumatic experience	no colour code		
	Alcohol consumption	no colour code		
	Drug consumption	no colour code		
	Smoking habits	no colour code		
	Stress	no colour code		
	Family Conflicts	no colour code		
	Professional underachievement	no colour code		
	Works in shifts	no colour code		
	Stress at workplace	no colour code		
	Procrastination	no colour code		
ISI	Score	≤ 7	Yellow:8-14 Orange:15-21	≥ 22
Epworth Sleepiness Scale	Score	≤ 9	Yellow: 10-12 Orange:13-16	≥ 17
PSQI	Score	< 5	-	> 5
	Bed time	no colour code		
	Wake up time	no colour code		

Glasgow Sleep Effort Scale	Score	no colour code
SCL-90-R	Total score	no colour code
	Somatization score	no colour code
	Obsessive-compulsive score	no colour code
	Interpersonal sensitivity score	no colour code
	Depression score	no colour code
	Anxiety score	no colour code
	Hostility score	no colour code
	Phobic anxiety score	no colour code
	Paranoid ideation score	no colour code
	Psychotism Score	no colour code

Table B.2: DSPD dashboard parameters, variables and the respective colour code.

Source	Colour			
	Parameter	Green	Yellow or Orange	Red
Actigraphy and PSG	Sleep onset hour	Before 2 AM	-	After 2 AM
	Wake up hour	Before 11h00 (AM)	11h00 to 14h59	After 15h00
	Total sleep time	Before 11h00 (AM)	11h00 to 14h59	After 15h00
	Latency	≤ 30 minutes	-	> 30 minutes
	Efficiency	$\geq 85\%$	-	$< 85\%$

Actigraphy Report	Light exposure	-	-	Too much exposure during night/not much during the day
	Matching diary and actigraphy schedules	no colour code		
Clinical Notes	Age at first symptoms	no colour code		
	Works in shifts	no colour code		
	Traumatic experience	no colour code		
	Traumatic car accident	no colour code		
	Family history	no colour code		
	Family conflicts	no colour code		
	Stress at workplace	no colour code		
	Procrastination	no colour code		
	Stress or depression	no colour code		
	Trigger	no colour code		
	Comorbidities	False	-	True
	Fears	False	-	True
	Complaints when waking up	False	-	True
	Daytime impairments	False	-	True
	Cognitive complaints	False	-	True
	Alcohol consumption	no colour code		
	Drug consumption	no colour code		
	Smoking habits	no colour code		
DLMO	Time at DLMO	no colour code		
Epworth Sleepiness Scale	Score	≤ 9	Yellow: 10-12 Orange:13-16	≥ 17
PSQI	Score	-	-	>5

	Bed time	no colour code
	Wake up time	no colour code
MCTQ	All parameters	no colour code
MEQ	Score	no colour code
SCL-90-R	All score	no colour code



Available graphs in Population Statistics

Table C.1: List of the available graphs in SleepData's Population Statistics feature.

General information
Age distribution
Gender distribution
Clinical Notes
Distribution of age at first symptoms
Related features distribution
Triggers distribution
Light exposure distribution
Bad habits distribution
Main comorbidities distribution
Distribution of the total number of comorbidities per patient
Main day-time complaints distribution
Cumulative distribution of complaints per patient
Actigraphy
Total sleep time distribution with variance
Sleep latency distribution with variance
Average bed time distribution
Average wake up time distribution
Sleep efficiency distribution (slots of $\leq 85\%$, $> 85\%$)
DLMO
Sleep onset (actigraphy) - DLMO phase angle distribution
Sleep onset (sleep diary) - DLMO phase angle distribution
Polysomnography
Total sleep time variance (minutes)

Total sleep time distribution (slots of <5 hours, 5 to 6 hours, >6 hours)
Sleep onset variance (minutes)
Sleep onset distribution (slots of ≤ 30 minutes, >30 minutes)
Sleep efficiency distribution (slots of $\leq 85\%$, >85%)
Actigraphy, PSG and Sleep Diary comparison
Average bed time hour
PSG Report
Perceived total sleep time variance (minutes)
Distribution of perceived total sleep time (slots of <5 hours, 5 to 6 hours, > 6 hours)
Distribution of the difference between actual and perceived total sleep time (minutes)
Distribution of the difference between actual and perceived total sleep time (slots of 30 minutes range, over 30 minutes with perceived time higher than actual, over 30 minutes with perceived time lower than actual)
PSQI
Score variance
Score distribution (slots of score ≤ 5 , >5)
ISI
Score variance
Score distribution (slots of score <8, 8-14, 15-21, >21)
Epworth Sleepiness Scale
Score variance
Score distribution (slots of score 10-12, 13-16, >17)
Glasgow Sleep Effort Scale
Score variance
SCL-90
Average of the scores for the 9 symptom dimensions