SleepData - Sleep Disorders Clinical Platform

Insomnia Population Characterization

Tiago André Ferreira Castanheira

October 2017

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.

Key words: Sleep disorders clinical platform, Health informatics technology, Insomnia

1 Introduction

Sleep clinics' way of gathering and analyzing data is still old-fashioned, as most records are paper-based and data is not centralized. Furthermore, multiple data sources are employed, such as actigraphy, polysomnography, clinical notes and lab exams, hindering the process of compiling information. The patient's attitude towards medicine and their own health data is also changing, as they demand to analyze it and input their own. SleepData, a sleep disorders clinical platform, was developed to address these points. The objective is to centralize sleep data, allowing for easier integration of the information and the use of analysis tools. With centralized sleep diagnosis, it becomes possible to characterize sleep-related disorders across populations. To showcase this (using the platform's tools), Insomnia, one of the most common disorders of this kind, is a very interesting candidate. Its chronic type, according to the American Academy of Sleep Medicine (2014), 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.

SleepData was developed on a partnership with a renowned sleep clinic, "Dra. Teresa Paiva – Centro de Eletroencefalografia e Neurofisiologia Clínica" (CENC), that employs multiple diagnosis and treatment equipment. CENC can 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. In turn, it allowed me to access their facilities and diagnosis equipment, interview their workers and study the usual workflow on a sleep clinic. The development of SleepData is the result of a collaborative effort with Pinho (2017), who used the platform to study another disorder, delayed phase sleep disorder (DPSD).

The next sections are organized as follows: Section 2 covers the main concepts regarding sleep, sleep medicine and the sleep disorders' diagnosis process. Section 3 addresses the main health informatic resources used. Section 4 introduces the SleepData platform, including its features. Section 5 describes the characterization of an insomnia population. Finally, Section 6 concludes and mentions future work to be developed.

2 Sleep Medicine and Insomnia

According to the *International Classification of Sleep Disorders* (ICSD3), proposed by the American Academy of Sleep Medicine (2014), sleep disorders can be categorized on 7 major groups – Insomnia, Sleep Related Breathing Disorders, Central Disorders of Hypersomnolence, Circadian Rhythm Sleep-Wake Disorders, Parasomnias, Sleep Related Movement Disorders and Other Sleep Disorders. Insomnia, the focus of this paper, has three types: Chronic, Short-Term and Other Insomnia. Chronic Insomnia is characterized by complaints of sleep initialization and maintenance, with associated daytime impairments such as fatigue, irritability and reduced attention and executive functions, that repeat for at least 3 times a week and last more than 3 months. The sleep complaints can not be explained by external factors or another sleep disorders. Short-term insomnia's definition is similar but the patient's complaints do not occur for at least 3 times a week or the symptoms lasted less than 3 months.

In terms of diagnosis, the main methods employed are actigraphy, melatonin exams, polysomnography (PSG) and anamnesis (regular appointments). Sleep questionnaires, filled by the patients, are also commonly used.

Actigraphy is a non-invasive method of monitoring temperature, movement and light exposure for long periods (days/weeks). CENC has actigraph units from Condor and Philips, two popular brands. Each equipment has a raw output has a time-series of sensors' outputs as well as some key statistics about the equipment and patient (name, gender and description). Condor's equipment export as a txt file and Philips' as a xlsx file, both only a few hundred kilobytes. The brand-specific analysis software compiles a report from this data, being the raw data barely used since it is not processed and the values are not directly comparable. The reports made by Condor's ActStudio are pdf files and docx files for the Philips' Actiware Software. These reports include key statistics as bed times, get up times, time in bed, total sleep time (TST), onset latency (difference between bed time and beginning of sleep), WASO (wake time after sleep onset) and number of awakenings in a night.

Melatonin exams take 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. With these values, DLMO (dim light melatonin onset) can be estimated, i.e. the process of increased melatonin production after darkness and before sleep.

PSG is a combination of recordings of biophysical parameters during sleep and is one of the most complete (and complex) sleep-related exams. 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 recordings, synchronized with the other exams to allow to investigate physiological (electrical) signals resulting from visually detectable events, such as periodic leg movements, myoclonus or awakenings. At CENC there are five different models of PSG equipment, from distinct brands or models. These are – Embla and Nicolet from Natus Medical Incorporated, SOMNOscreen from SOMNOmedics and Alice 5 from Philips. 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 and number. Regarding their size, it can be of up to a few gigabytes, mainly due to the additional video recording. 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 3 types: objective data, directly observed with the included PSG software,subjective data, derived from the objective data and hand-written notes (by the technician) about the patient's over-night stay (also known as a "night-diary"). The objective data encompasses parameters like sleep onset time, wake up time, sleep latency, sleep efficiency, the percentage of the sleep time spent on N1, N2 and N3, number of sleep cycles, blood oxygenation levels, periodic sleep movements and micro awakenings. By examining this information and the time-series some subjective information is derived like the existence of abnormalities on the deep sleep or REM time profiles, alpha or beta wave intrusions (EEG), period limb movement disorders. From the night diaries, some of the most important information are the perceived sleep duration and quality.

In sleep medicine, anamnesis is a very relevant part of the diagnosis process. On the insomnia context, Paiva and Pinto (2014) do a in depth guide for the anamnesic process, dividing it in categories. The most relevant points to address are the circumstances of beginning of the symptoms, the symptoms themselves (including severity and frequency), aggravation and improvement factors, stress factors, sleep hygiene, circadian rhythms, family history and treatments the patient might be doing.

Regarding sleep questionnaires on insomnia diagnosis the most common, according to Paiva and Pinto (2014), 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. 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. Lastly, the Glasgow Sleep Effort Scale measures sleep effort, i.e. attempt to force and induce sleep voluntarily. According to Paiva and Penzel (2011), other important questionnaires on sleep disorder diagnosis 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).

More general questionnaires are also employed as 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. 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.

3 Health Informatics Technology

SleepData was developed over the MEAN Stack[1], a software bundle like the popular LAMP (Lawton, 2005). This stack encompasses **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.

The MongoDB database has some advantages over relational databases, mainly because it allows that more fields are added easily. It is also easily integrated with Node.js. On the other hand, referential integrity is non-existent. Union operations or cascade deleting, for example, can only be done externally to the database server. Node.js[2] 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). The last element to the MEAN Stack is the Express framework for Node.js and its main advantage, besides being integrated with the Node.js' ecosystem, is its ease of use.

MEAN as a whole has the advantage of using JavaScript (JS) on both client side and server side. Data is stored on the database as JSON (JS Object Notation), which is 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/JS, typical of LAMP.

Health Information Platforms have unique requirements when compared to general purpose platforms. In SleepData, data must be secure and anonymized, shareable and unequivocally identifiable. For data to be shared while keeping the same exact meaning there is a need for two things: a defined, common structure (preferably adopted internationally) and a coding system that allows to identify each finding/measure, unequivocally.

The adopted database structure was HL's FHIR. HL7[3]

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

Figure 1: Detail of the Observation Resource JSON (respiratory rate observation) from HL7's FHIR

(Health Level 7) is a nonprofit organization that focus on building frameworks and standards for sharing of electronic health information and is widely supported, which is a big advantage when comparing, for example with alternatives like openEHR[4]. HL7's focus is not only on the structure of data but also on the protocols to exchange it between systems. FHIR[5]((Fast Healthcare Interoperability Resources) is its most recent standard for medical data structure and transfer. 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 healthcare applications – clinical, identification, workflow, financial, conformance and finally infrastructure. The resources are available as structure diagrams, UML Diagram or XML or JSON. On SleepData, diagrams were used to understand the structure and then JSON to define the database's documents. Having JSON structures is one of the advantages of FHIR as those are easily integrated on a MEAN Stack based platform.

These structures need a coding system of health terminologies to uniquely identify each measurement (example on the Figure 1). On SleepData, health terminologies are needed on two main categories – laboratory exam's results, such as DLMO or vitamin D exams and general clinical notes terms and parameters gathered by equipment like actigraphs, ECG or EEG.

LOINC[6] (Logical Observation Identifiers Names and Codes), which is an international standard for coding of clinical observations and laboratorial measurements and, as such, covers the first mentioned category. 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). According to Figueiral (2015), other terminologies, 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 on disorder classification while ICPC is best used for primary care. However, the most appealing factor for the adoption of LOINC within is its great compatibility with SNOMED CT[7]. This other cod-

[1]The MEAN Stack – http://mean.io/

[2]Node.js – https://nodejs.org/

[3]Health Level 7 – http://www.hl7.org/index.cfm

[4]openEHR – http://www.openehr.org/home

[5]HL7's FHIR – http://hl7.org/fhir/

[6]Loinc – https://loinc.org/

[7]SNOMED CT – http://www.snomed.org/snomed-ct

ing system's 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. SNOMED's repertoire is also massive with 326 734 active concepts as of the January 2017 release[8].

4 SleepData Structure and Features

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 on both 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). As seen, its software was based on the MEAN stack. The MongoDB databases, one for users (to handle login information and access control permissions) and another for the rest of the information, are organized by collections, that can be of one out of three categories – administrative, objective measures and reports. In the diagram of Figure 2, it is possible to see the different collections that were used, and to which category they belong.

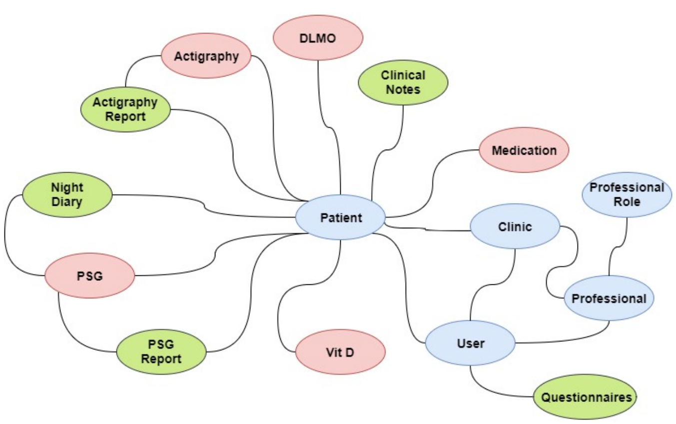

Figure 2: Database's collections, colored by category: administrative in blue; objective measures and exams in red; reports and subjective observations in green. *Questionnaires* element includes a collection per questionnaire.

The collections that belong to the administrative category contain documents with administrative information. For instance, the User collection has a document per user, which contains contact info and personal information, such as name, language, country, address, gender, etc. It also includes the user's unique ID as well as the user's patient or professional IDs (if applicable). Similarly, the Patient collection has a document per patient, with added fields for clinic identification, tutor (for minors) and the patient's general practitioner. The Clinic collection stores 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, office and availability (work schedule and holidays).

SleepData's documents are defined using FHIR's resources. In this case, the Patient and User collections are based on the resource *Patient* and the Professional and Professional Role collections are based on *Practitioner* and *PractitionerRole*, respectively. The Professional collection is not limited to medical practitioners (it supports other employees like IT technicians or a system admins).

Other collections store objective findings from exams. Namely actigraphy, melatonin measures, polysomnography and vitamin D. Under this category, the patient's medication is also included. For each exam, the needed fields were accessed based on the literature, consulting CENC's patient exams and validated by Dra. Teresa Paiva, a renowned doctor on sleep medicine and advisor of this project. The exams were structured after the *Observation* FHIR resource and the medication collection after *Medication*. *Observation*, besides administrative data (such as the date, subject and performer IDs), consists on a compilation of components, storing medical variables. Each component includes a set of three fields per coding system used: identification of the coding system used, the code of the variable on that particular coding system and a display value. The component also includes the value of the measurement, as well as a field for its interpretation and reason for being absent (if it is).

Lastly, there are collections that store subjective information, either self-reported (questionnaires), gathered during anamnesis/observation (clinical notes and night diaries) or derived from objective exams (actigraphy and polysomnography reports). These collections were based on FHIR's *Observation* resource besides the questionnaires that were based on *Questionnaire* that 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. Clinical notes also encompass observations that are not strictly medical (even though they are relevant). Some examples are the patient's religion, profession/occupation and education level. To define religions a FHIR's coding system, *HL7 v3 Code System ReligiousAffiliation*[9], was used. Education levels were coded with ISCED from the UNESCO - Institute for Statistics (2011) and jobs and occupations with ISCO-08

[8] SNOMED Worldwide – `http://www.snomed.org/snomed-ct`

[9] FHIR's Religions coding system – `https://www.hl7.org/fhir/v3/ReligiousAffiliation/cs.html`

from the International Labour Office (2008).

Now that the database structure was addressed, it is possible to introduce the relationships among different elements on SleepData. The *User* is the central element, as he's the one who does all the actions and interacts with the other elements. A User can be a Professional (with Roles) or a Patient. Both Users and Patients can belong to a Clinic. The *Patient* is the main link for all the data sources, since all refer back to him. A Patient can take Medication or do exams. Each one of these exams can have some or all of 3 components, namely Raw Files, Objective Data and Reports.

SleepData implements a mandatory access control system to ensure that the content available to each user is personalized. Three user classes were created:

Regular user with just basic access to the platform, being able to see site statistics like number of Users, exams and other general data (just as any visitor). He can fill sleep questionnaires and basic personal information such as its name, date of birth or gender.

Administrator has administrative functionalities, such as the capability of creating new Professionals, as well as define new Clinics. He can also change the class of users, assign them to Clinics and map users to Professionals. Admins can also delete accounts.

Professional User accounts are attributed to a practitioner, technician or researcher of a clinic (users from a clinic cannot access data from another). He has access to most functionality, except Admins'.

The sleep questionnaires available for the regular user are: PSQI, ISI, Glasgow, Epworth, MCTQ, MEQ and SCL90. Regarding the professional user, Figure 3 displays the homepage of one from CENC (homepages from the other users classes follow the same design cues).

The homepage is divided into categories – Administrative, Exams, Lab Exams, Raw Files Upload and Data Visualization – 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 user classes and can't add professionals or register clinics. Exams and Lab Exams refer to user interfaces to input patient data. When adding data through the form, some information is already filled in, if available on the database. For example, when adding clinical notes, after choosing the patient, his name and ID are automatically added to the form.

The Raw Files Upload category includes interfaces to upload actigraphy or PSG raw files, which are stored on an organized manner with information about equipment, patient and date. Besides raw files, SleepData also handles the actigraphy report files, generated by the analysis software. After uploading these reports, they are automatically parsed so relevant information is gathered and wrote to the database according to the FHIR's structure. The original files are also saved on the platform.

Figure 3: Homepage of a CENC professional user.

The last category, Data Visualization, refers to the graphical reports, either specific to a patient (Dashboards) or on a population basis (Population Statistics). The dashboards are table-like pages that display crucial information about a patient and are organized by sections, depending on the source. Doctors can have multiple dashboards (two were implemented) that focus on different aspects and have different key parameters displayed. This way the doctor can quickly visualize all the relevant data on a certain context. The dashboards were developed in cooperation with Dr. Teresa Paiva and based on the guidelines described on Paiva and Penzel (2011) and Paiva and Anderson (2014). 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. 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 (Figure 4), allowing the clinician to quickly interpret the data and to do an overall assessment of the patient's condition.

In addition to summaries for specific patients, SleepData 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. Through the interface, the user can filter the results so only some of the patients are considered, updating the graphs accordingly. Two filters were implemented so Insomnia or DSPD patients can be selected. After choosing a filter (or none, to select all

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: Detail of the Insomnia Dashboard. Notice that the color of the cells change, based on the patient's data.

patients), the graphs and stats are shown, divided into categories. The categories are: general information, clinical notes, actigraphy, DLMO, PSG, PSG report, PSQI, ISI, Epworth Sleepiness Scale, Glasgow Sleep Effort Scale and SCL-90. The first section includes demographic data, such as the age and gender distributions of the patients. The clinical notes' section includes histograms of age at first symptoms, traumatic events 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. 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. On the PSG's section, 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. 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. The sample size is displayed over each graph. Considering all categories, over 30 graphs were developed.

Regarding security, SleepData was designed to comply to the Deliberação nº.1704/2015 from the Comissão Nacional de Protecção de Dados (2015), a guideline that states procedures that must be followed to ensure security of medical data. The main points that SleepData addresses are the logical separation between medical and administrative data (separate collections are used); use of a mandatory access control system to differentiate access to features; capability of editing user types by the administrators; restricted access to the servers (servers have physical barriers and 24/7 security). Besides these recommendations, other measures were adopted. Firstly, using Node.js' Passport (authentication middleware), all the passwords are encrypted, making them indecipherable to system admins or after password leaks. To prevent man in the middle attacks, SSL certificate is also used.

5 Insomnia Population Characterization

To characterize an insomnia population, two separate approaches were used. The first one was the development of the Insomnia Dashboard as a tool for diagnosis and single-patient characterization. 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. Besides those, two smaller sections are also available, namely Actigraphy Report and Vitamin D.

Clinical Notes (anamnesis) is a common source of information for most sleep disorders, being Insomnia 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 Clinical Notes section (Figure 5) 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 (ICSD3), 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, including triggering factors, such as traumatic events or more subtle (although harmful) lifestyle characteristics such as alcohol consumption, smoke habits or a bad work environment, since they are common in Insomnia. 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.

PSG and its respective report are also important sources of information for evaluation of Insomnia as they objectively determine key sleep parameters like sleep efficiency, sleep latency and total sleep time (TST). Not only that, but more specific sleep-pattern data can be accessed, as REM and N phases cycles and alpha and beta EEG intrusions. Regarding PSG, the dashboard displays information like TST, sleep efficiency and latency as they are simple, but common, giveaways for Insomnia. Besides those, number of sleep cycles and duration of the REM and N1 phases are shown. According to Paiva and Pinto (2014), both increased N1 sleep and reduced REM sleep duration 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 from 4 or more (marked

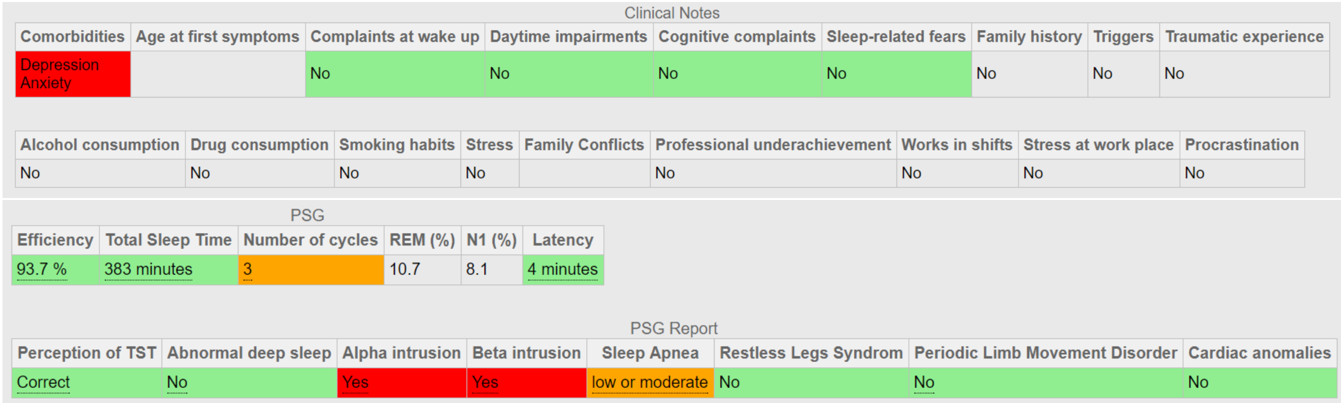

Figure 5: Clinical Notes (top), PSG and PSG Report sections (bottom) of the Insomnia Patient Data Dashboard.

green) to 2 (orange) or even 1 (red). Regarding PSG Report, the patient's perception of TST, deep sleep abnormalities, alpha and beta intrusions are displayed (as they are also indicative of Insomnia). 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.

In the questionnaires section the scores for some of the most relevant questionnaires for Insomnia study are displayed: PSQI, ISI, Epworth, Glasgow and SCL90. 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 an estimation of the patient's overall satisfaction with its sleep. Generally, a total score under 5 indicates 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 but a bad score on only one can suggest that there might be external causes that explain the bad sleep experience (e.g. work-related anxiety or poor sleep hygiene during the weekends). PSQI also asks for the bed and wake up times for each scenario, which can 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.

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 to provide a quick insight about the significance of the score.

Epworth evaluates just how sleepy the patient is during the day, which is characteristic in Insomnia patients and one of the possible day-time impairments that define it. As with ISI, score brackets are defined (up to 9, 10 to 12, 13 to 16 and above 17) and colored accordingly, from the red, with scores below 9 to green, above 17.

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

Finally, SCL90 assesses the severity of psychological problems that patients might have, and how these affect their life. Its divided into 9 sub-categories – somatization, obsessive-compulsive, interpersonal sensitivity, depression, anxiety, hostility, phobic anxiety, paranoid ideation and psychoticism – and each one has a corresponding score. It is used to determine if the patient already has changes in behaviour, caused by the symptoms. The dashboard displays the score for each category separately, as well as an overall score.

Two additional small sections, Actigraphy Report and Vitamin D are also present on this dashboard. The actigraphy report can be useful to rule out Insomnia and attribute sleep difficulties to poor sleep hygiene. This sections show if the patient is exposed excessively to artificial light before going to bed or not enough to natural light in the morning and if the patient naps too frequently. The vitamin D value can indicate to the clinician that it needs to adjust the patient's prescription, for instance. The lack of this vitamin is sometimes linked with Insomnia.

The second approach to characterize insomnia was to study the population as a whole, using the SleepData's Population Statistics. For that, data from a sample of 100 patients was loaded into SleepData. Most patients had some clinical notes data, PSG data and the results of four questionnaires: Glasgow, PSQI, ISI and SCL-90.

Regarding general demographic data, specifically gender, 64.9% are females and 35.1% males. The higher number of women is expected even though the observed value is 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 population constituted on 58.51% by females. This may be due to the fluctuating hormone levels (melatonin and cortisol) or even due to pregnancies. Considering the data from all the Portuguese population (from PORDATA), in 2015, women were 52.7% which

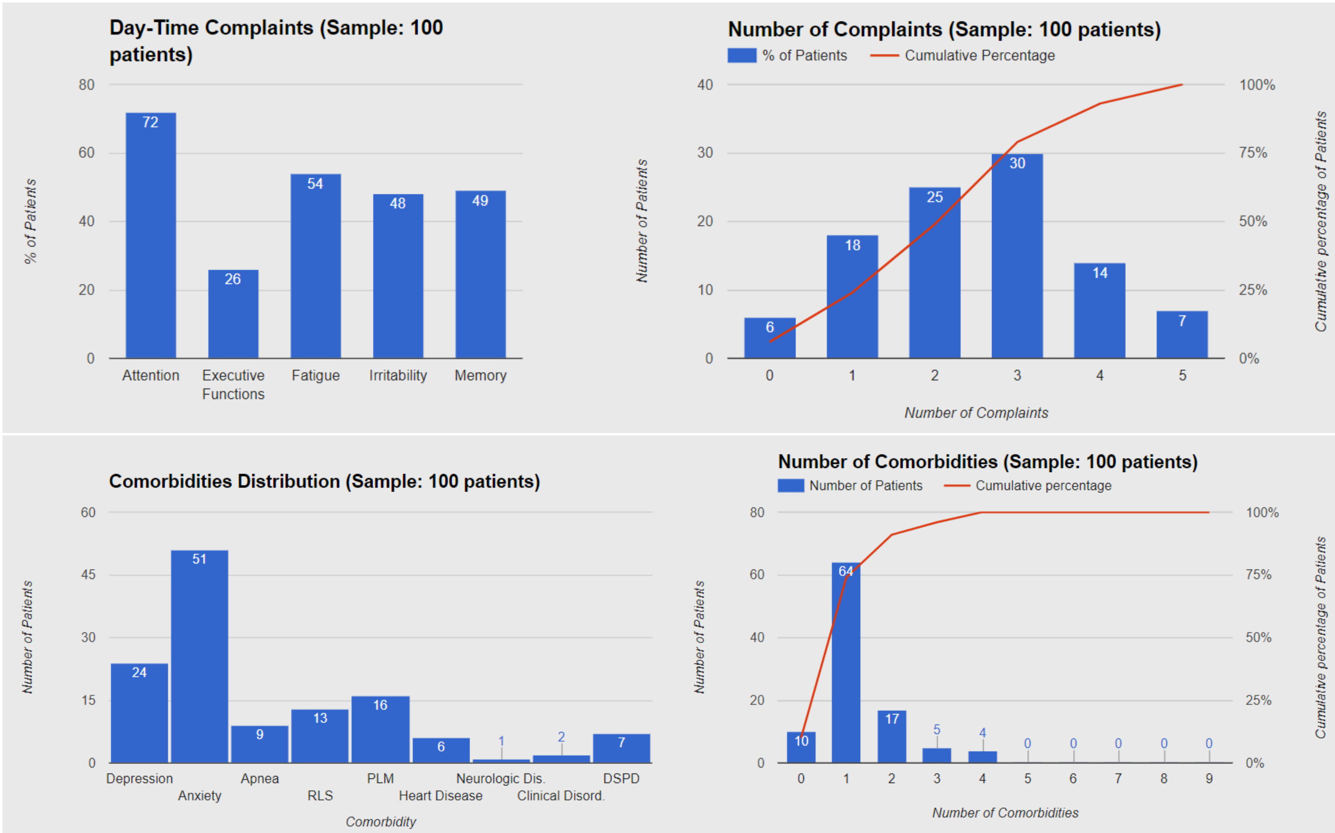

Figure 6: Prevalence of day-time complaints caused by poor sleep quality (top) and comorbid disorders (bottom).

further reinforces the idea of the gender skew within the insomnia population. Regarding their age, most patients are over 40 years old (85.4%) which reflects the documented increase of insomnia incidence with age. When comparing with the 2015 Portuguese population, only 56.6% was over 40 years old. Considering the insomnia population over 60, it adds up to 36% against 27.2% when considering the entire population.

From anamnesis data, one of the most relevant are the patient's complaints regarding day-time activity, caused by poor sleep quality. Five of them – attention deficit, the reduction of executive functions, fatigue, irritability and lack of memory – were available on this data set and the results are visible on the Figure 6. Attention deficit is the most common of these complaints (72%) 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. These patients are also highly comorbid as only 10% of the population did not have any other disorder and 26% have 2 or more. Anxiety is the most common comorbidity, followed by depression (affecting 51 and 24% of the population, respectively). Sleep apnea is also common (observed in 9% of the patients).

Regarding PSG data, the observed average TST is 360.3 minutes or about 6 hours, below the recommended value of 7-8 hours (American Academy of Sleep Medicine). The patient on the third quartile slept only 409 minutes (6h49min), which is still insufficient. 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. Not only is these patients' sleep short, but also sleep onset delayed. In fact, 29.3% take more than 30 minutes to fall asleep, being the average also 30 minutes, which shows just how high some of the patients' latencies are. Regarding sleep efficiency, 60% of patients are under 85% efficiency.

After the PSG exam, the patient is asked about its opinion regarding the sleep, including the perceived TST, and the results show that Insomnia patients are pessimists. The average perceived TST is 270 min, when 360 min was actually recorded with PSG. 15.2% of the patients actually slept less than 5 hours of sleep per night even though more than half of the patients think so. Considering the difference of perceived and actual TST for each patient (Figure 7) 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. An estimate is considered correct when within a 30 min range. From the studied population, 54.9% of them are pes-

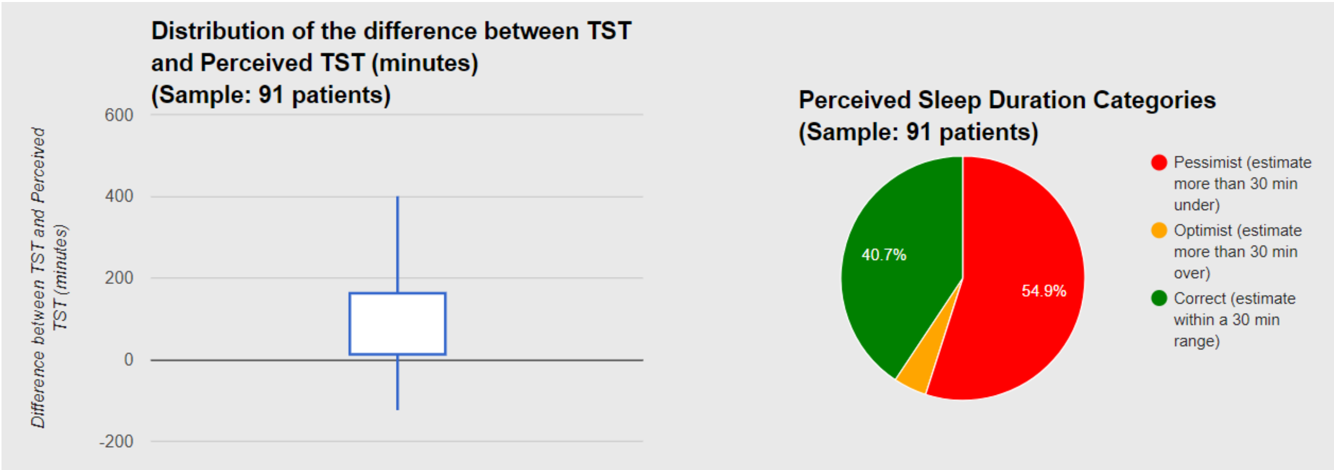

Figure 7: Differences between perceived TST and actual TST, measured through PSG.

simists and 40.7% accurate, which 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.

Regarding the PSQI results, just 2.2% of the patients have scores under 5 (cut-off value). In fact, the average value is 12.64 which reveals just how degraded is the patients' sleep quality (or at least their perception of it). Studies like the one from Grandner et al. (2006) found the average score to be 4.07 among a young population (average age of 23) and 3.92 among an older population (average age of 66). Both are inside the cut-off value and well below the average for this population.

The ISI results are as expected. Most patients (98.9%) have scores of 8 or above, consistent with the insomnia diagnosis. 22% have scores higher than 21, which correspond to very severe insomnia symptoms. The average, 18, belongs to the bracket that indicates clinical insomnia.

Glasgow doesn't have a defined cut-off value. Instead, each question describes a type of difficulty/displease regarding sleep and the person has to say if it applies, if neutral or if it does not (values of 3, 2, 1, respectively). 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. A study from Broomfield and Espie (2005), found the average question score to be 1.37, from a sample of 102 people without reports of abnormal sleep – i.e., healthy subjects tend to disagree with most sentences. The study suggests a cut-off value of 9 (total score), to distinguish normal patients from the insomnia ones. Within SleepData, the average was 15.76, which is according to that suggestion.

Regarding SCL90, SleepData's Population Statistics displays the score of each component against the average for a healthy population, calculated from a sample of 302 Portuguese without sleep 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 those were two of the most diagnosed comorbidities. The somatization category refers to somatic symptoms, like headaches, dizziness and others as a consequence of (or to communicate) psychological distress. The obsessive-compulsive category refers to behaviours like being excessively concerned, cautious or fearing of forgetting. These can be caused by the insufficient total sleep time inherent with insomnia.

6 Conclusions and Future Work

The development of the SleepData platform was motivated by the lack of a 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 was to assess which data sources are most relevant to the users of the platform. For each source, it was determined how the data is acquired, stored and which parameters are the most relevant. Methods were developed to upload and parse (when possible) the exams files. To aid the visualization of the information, patient data dashboards were implemented, with rules to highlight certain data, as well as graphs to characterize sub populations. To map each parameter (avoiding ambiguity and improving interoperability), coding systems were chosen, with focus on modern, widely spread and international standards. The implemented database structure allows that fields are added and followed a standard that also includes APIs for data transfer. In terms of the platform development, a modern software stack was used, which includes a No-SQL database, and allows for easy setup of both server and front-end environments (all based on JavaScript). Throughout all the SleepData's development there was a focus on open-source software and tools. To ensure data privacy and security, Sleep-

Data implements a mandatory access control system with classes of users and was built in compliance with a national guideline for clinical data security.

Two populations studies have been conducted with SleepData, one on DSPD, by Pinho (2017), and a study on Insomnia, reported in this paper. 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. Non surprisingly, this lead to a very displeased population, that exaggerates their symptoms and describe their sleep as worse than it actually is. Average scores for key sleep medicine questionnaires were also calculated.

SleepData can host patient records from multiple clinics, offering tools to visualize the data. Being web-based, it has the advantage of being available on any browser, which can be an advantage for the clinicians that want to quickly check on their patient, from their personal device. This considered, many features can still be improved and new ones added. Security should be a priority, with data saved on an encrypted manner and two factor authentication strategies should be implemented. The first could be achieved through the *Encryption at rest* functionality of the enterprise (paid) version of MongoDB. Besides security, improvements can be done on additional data sources supported in SleepData such as genetic data, blood exams or general patient data such as diet, exercise or hobbies. To do this there would have to be integration with other equipment, such as wearables, smartphones, respiratory-aid equipment or even EHRs and other health platforms.

Multiple languages should be implemented to improve adoption and usability. 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 SleepData. This would lead to more adoption from clinics, as they could import data without 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, allowing the population statistics data to be filtered by any disorder. To code these ICSD3, that only codes sleep disorders, and ICD, a general disorder classification system, should be adopted.

In terms of data visualization, the dashboards could become more dynamic, giving users the possibility of creating their own, with the desired fields. Data should also be editable, from the dashboard. Regarding the population statistics, similar improvements can still be made – custom filters, designed by the user, to adapt to their needs. 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.

References

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.

Castanheira, T. (2017). SleepData - Sleep Disorders Clinical Platform. Insomnia Population Characterization. Master's thesis, Instituto Superior Técnico.

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.

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.

International Labour Office (2008). International Standard Classification of Occupations (ISCO-08).

Lawton, G. (2005). Lamp lights enterprise development efforts. *Computer*, 38(9):18–20.

Paiva, T. and Anderson, M. L. (2014). *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.

Pinho, B. (2017). SleepData - Sleep Disorders Clinical Platform: Delayed sleep phase disorder. Master's thesis, Instituto Superior Técnico.

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.