Aalto University
School of Business

TREATMENT EFFECT ANALYSIS FOR OBSTRUCTIVE SLEEP APNOEA

Master's Thesis
Elina Toivanen
Aalto University School of Business
Information and Service Management
Fall 2022

Aalto University
School of Business

Aalto University, P.O. BOX 11000, 00076 AALTO
www.aalto.fi

Abstract of master's thesis

Author Elina Toivanen

Title of thesis Treatment Effect Analysis for Obstructive Sleep Apnoea

Degree Master of Science in Economics and Business Administration

Degree programme Information and Service Management

Thesis advisor(s) Lina Siltala-Li, Ilkka Leppänen, Tarja Saaresranta

Year of approval 2022 **Number of pages** 93 **Language** English

Purpose of the study

The primary purpose of this thesis is to compare the effectiveness of four Obstructive Sleep Apnoea (OSA) treatment methods. The thesis focuses on the following treatment options: Continuous Positive Air Pressure (CPAP), Oral Appliance (OA), bariatric surgery, and upper airway surgery. The effectiveness of a treatment intervention is measured by 1) the risk of relapse and 2) cumulative relapse recurrence. *Relapse* is defined as the return of the signs of increased OSA severity.

Methodology and data

The data for this thesis consist of Electronic Health Records (EHR) data obtained from Turku University Central Hospital, which contains the information of approximately 24 700 sleep apnoea patients. The patients were followed between the years 2003 and 2019. In addition, the data were collected from patient visits and CPAP machines at home.

Treatment response is studied with a multistate modelling technique that measures how previous events affect the probabilities of future events. The risk of relapse is measured with Markov chain state transition probabilities, and cumulative relapse recurrence is calculated with a Nelson-Aalen procedure.

Results

The study results suggest that for all treatment arms, the risk of relapse is higher in milder OSA states. Consequently, the risk of relapse and the estimated relapse recurrence are lower in patients with severe OSA. The study findings support the previous consensus on the treatment effectiveness of CPAP. In addition, OA performs well in all severity states. Bariatric surgery and upper airway surgery showed poor performance in milder OSA states. Furthermore, the study finds a statistically significant relationship between oxygen desaturation values and OSA severity.

Keywords obstructive sleep apnoea, relapse, electronic health records

Tekijä Elina Toivanen

Työn nimi Analyysi Obstruktiivisen Uniapnean Hoitovasteesta

Tutkinto Kauppatieteiden maisteri

Koulutusohjelma Tieto- ja Palvelujohtaminen

Työn ohjaaja(t) Lina Siltala-Li, Ilkka Leppänen, Tarja Saaresranta

Hyväksymisvuosi 2022 **Sivumäärä** 93 **Kieli** Englanti

Tutkielman tarkoitus

Tämän pro gradu -tutkielman ensisijaisena tarkoituksena on tutkia neljän obstruktiivisen uniapnean hoitokeinon tehokkuutta. Tavoitteena on verrata ylipainehengityshoitoa, uniapneakiskoa, kirurgisia laihdutusleikkauksia sekä ylähengitysteitä avartavia leikkauksia toisiinsa. Hoitokeinojen tehokkuutta mitataan 1) uusiutumisen todennäköisyytenä ja 2) kumulatiivisena uusiutumistodennäköisyytenä.

Metodologia ja aineisto

Tutkielman aineisto koostuu Turun yliopistollisesta keskussairaalasta saadusta potilasrekisteristä, joka pitää sisällään noin 24 700 uniapneapotilaan tiedot. Potilaat ovat olleet seurannassa vuosina 2003–2019. Aineistoa on kerätty sekä potilaskäynneiltä että potilaiden kotona ylipainehengityshoitoihin käytettävistä laitteista.

Hoitovastetta tutkitaan monitilamallin avulla, jossa tarkastellaan miten aiempi tapahtuma vaikuttaa myöhemmän tapahtuman todennäköisyyteen. Mallissa uniapnean uusiutumisen todennäköisyys määritellään mallin alku ja lopputilan väliseksi ennustetodennäköisyydeksi. Lisäksi sairauden progressiivisuutta tutkitaan kumulatiivisen uusiutumisen estimaateilla.

Tulokset

Löydösten mukaan obstruktiivisen uniapnean uusiutumisen riski on korkeimmillaan lievimmissä vaikeusasteissa. Lisäksi vakavimmissa vaikeusasteissa uusiutumisen sekä kumulatiivisen uusiutumisen riski on matala. Tutkimustulokset tukevat asiantuntijoiden yksimielisyyttä CPAP hoidon tehokkuudesta. Lisäksi tulokset osoittavat uniapneakiskon tehokkuuden kaikkien uniapnean vaikeusasteiden hoidossa. Lihavuusleikkaus ja ylähengitysteitä avartavat leikkaukset osoittivat matalaa tehoa lievissä vaikeusasteissa. Tutkielma havaitsi myös tilastollisesti merkittävän yhteyden happisaturaation ja uniapnean vaikeusasteen välillä.

Avainsanat uniapnea, hoitovaste, sähköiset potilastiedot

Table of Contents

1 Introduction 1

1.1 Background and Motivation 1

1.2 Research Problem and Objectives of the Study 2

1.3 Scope of the Thesis 4

1.4 Structure of the Thesis 4

2 Literature Review 5

2.1 Obstructive Sleep Apnoea 5

2.1.1 Disease Progression and Morbidities 5

2.1.2 Diagnosis of the Condition 6

2.1.3 Severity Classification 7

2.1.4 Risk Factors 8

2.2 Treatment Methods 9

2.2.1 Treatment Methods and Indications 9

2.2.2 Previous Research on OSA Treatment Effect 10

2.2.3 Continuous Positive Airway Pressure (CPAP) 11

2.2.4 Oral Appliance (OA) 11

2.2.5 Upper Airway Surgery 12

2.2.6 Bariatric Surgery 12

2.3 Treatment Effect Research in Longitudinal Study Settings 13

2.3.1 The Practice of Treatment Effect Research 13

2.3.2 Longitudinal Studies involving EHR 13

2.3.3 Multistate Models for Studying the Disease Process 14

2.3.4 Key Concepts in Survival Analysis 14

2.3.5 Challenges in EHR Longitudinal Studies 15

2.4 Hypothesis Development for OSA Treatment Effect Analysis 16

3 Research Methodology 19

3.1 Defining the Research Process 19

3.2 Severity Classification 20

3.2.1 Ordinal Response Models 21

3.2.2 Random Effects Account for Patient Clusters 21

3.2.3 The Proportional Odds Assumption 22

3.2.4 CLMM Model Diagnostics 23

3.3 Treatment Analysis 23

3.3.1 Objectives of Treatment Analysis 24
3.3.2 Problem Characterisation 24
3.3.3 Multistate Analysis and Back Transitions 25
3.3.4 Markov Chain Models 26
3.3.5 Obstructive Sleep Apnoea Model 27
3.3.6 Aalen-Johansen Estimator to Compute the Risk of Relapse 28
3.3.7 Nelson-Aalen Estimator to Estimate Cumulative Relapse Recurrence 30
3.3.8 Hypothesis Testing: Differences Between Treatment Arms 31

4 Data 33
4.1 Description of Data 33
4.2 Data Architecture and Tools 34
4.3 Model Variables 34
4.4 Treatment Identifications 36
4.5 The Sample Selection Process 37
4.6 Data Pre-processing 38
4.6.1 Irregular Time-series Data 38
4.6.2 Outlier Detection 39
4.6.3 Data Missingness 39
4.6.4 Multicollinearity 40
4.7 Descriptive Analysis 41
4.7.1 Overview of the Study Sample 41
4.7.2 The Characteristics of The Severity Groups 42
4.7.3 Treatment Types 43
4.7.4 Patient Relapses 44

5 Empirical Findings and Results 46
5.1 Severity Classification 46
5.1.1 Research Hypothesis 1: The Determinants of OSA Severity 47
5.1.2 Reliability of the Results: Ordinal Regression 47
5.2 Treatment Analysis 48
5.2.1 Relapse Probabilities Measure the Risk of Relapse 48
5.2.2 Cumulative Relapse Recurrence and Disease Progressiveness 51
5.2.3 Research Hypothesis 2: Differences in Treatment Outcomes 54
5.2.4 Reliability of the Results: Treatment Effect Analysis 54

6 Conclusions & Discussion 57
6.1 Summary 57

6.2 Severity Classification 58
6.3 Treatment Analysis 58
6.3.1 Summary of OSA Treatment Effect Analysis 59
6.3.2 Comparisons of Outcomes by Treatment Arms 60
6.4 Limitations 62
6.4.1 Intermittent Observations and Missing Data 62
6.4.2 Research Design 63
6.4.3 The Research Sample 64
6.5 Future Research 64
6.6 Ethical Considerations 65
6.7 Conclusions 66
7 References 68

List of Tables

Table 1: The Finnish Severity Classification Guidelines 7
Table 2: OSA Risk Factors 8
Table 3: Treatment Methods and Treatment Indications 9
Table 4: Types of Missingness 16
Table 5: PostgreSQL Tables Utilised for this Thesis 33
Table 6: AHI Thresholds for the Severity Variable 35
Table 7: Variable Descriptions and Inclusion Ranges 35
Table 8: Treatment Variables and Identification Criteria 36
Table 9: The Sample Selection Process 38
Table 10: Correlations Between Variables 40
Table 11: Descriptive Statistics of the Sample Population by Severity Levels 42
Table 12: Results from CLMM on Associations between Patient Biomarker Data and OSA severity 46
Table 13: Probabilities of Relapse, 95% CIs and Number of Patients at Risk. 50
Table 14: Nelson-Aalen Estimates, 95% CIs and Number of Patients at Risk. 53
Table 15: The Markov Assumption Test 55
Table 16: Summary Table: Relapse Probabilities and Cumulative Relapse Estimates 60

List of Figures

Figure 12. Three Types of Relapse Recurrence 3
Figure 3. OSA progression chart. (Adopted from Marin-Oto et al., 2019)..... 6
Figure 4. Scopus Publication Volume on OSA Treatment Methods (1970-2021)..... 10
Figure 5. The Three Bariatric Surgery Types in Finland: (A) Laparoscopic Adjustable Gastric Banding, (B) Sleeve Gastrectomy and (C) Roux-en-Y Gastric Bypass Procedure. (Dixon et al., 2011/ permission granted) 12
Figure 6. Illustration of Study Process..... 19
Figure 7. Four Severity Categories. 20
Figure 8. OSA Progression Diagram. 24
Figure 9. Illustration of Cumulative Relapse Recurrence by Type of Relapse 25
Figure 10. Markov Chain Diagram to Model the Relapsing Events in OSA 26
Figure 11. Transition Diagram of the Relapsing-remitting Process for OSA..... 28
Figure 12. Data Architecture 34
Figure 13. Example of the Per-protocol Approach in a Patient's Timeline..... 37
Figure 14. An Example of the Outlier Detection Procedure. (A boxplot with lower 25th and upper 75th percentiles as box boundaries, the line inside the box is the median, lower and upper error lines are 10th and 90th percentiles, and filled circles indicate data outside the 10th and 90th percentiles.)..... 39
Figure 15. Distribution of Age at Diagnosis by Gender 41
Figure 16. Treatment Distribution Among Patient Sample..... 43
Figure 17. Total Yearly Relapses in Each Treatment Arm 44
Figure 18. Aalen-Johansen Relapse Probabilities in Each Treatment Arm (95% CI)..... 48
Figure 19. Nelson-Aalen Estimates for Relapse Types and Treatment Arms (with 95% CI) 51

Abbreviations

AHI	Apnoea-hypopnoea Index
ASV	Adaptive Servo Ventilation
BMI	Body Mass Index
BP	Blood Pressure
CPAP	Continuous Positive Airway Pressure
CSA	Central Sleep Apnoea
CI	Confidence Interval
CLM	Cumulative Link Model
CLMM	Cumulative Link Mixed Model
DEPS	Depression Questionnaire
EHR	Electronic Health Records
ESS	Epworth Sleepiness Scale Questionnaire
GEE	Generalised Estimating Equations
GLMM	Generalised Linear Mixed Model
G47.3	ICD code for Sleep Apnoea
GLU	Blood Glucose Level
ICD	International Classification of Diseases
LRT	Likelihood Ratio Test
MICE	Multivariate Imputation by Chained Equations
ODI	Oxygen Desaturation Index
OSA	Obstructive Sleep Apnoea
PH	Proportional Hazards
PO	Proportional Odds
PSG	Polysomnography
PMM	Predictive Mean Matching
SaO2	Oxygen Saturation of the Arterial Blood
VSSHP	Hospital District of Southwest Finland
VIF	Variation Inflation Factor

1 Introduction

1.1 Background and Motivation

Obstructive Sleep Apnoea (OSA) is a common disorder among the Finnish population, and it has a rising incidence rate along with its comorbidities (Mattila et al., 2022; Tolonen, 2022). Although general awareness of OSA has increased among the Finnish healthcare professionals and the overall population, the condition remains underdiagnosed (Mandereau-Bruno, 2021; Mattila et al., 2022). Lääkärilehti (Finnish Medical Journal, 2021) recently published a report estimating that approximately 1.46 million Finns (~26% of the population) suffer from mild OSA. In contrast, nearly 850 thousand individuals (~15% of the population) have been diagnosed with a more severe form of the condition (Bachour & Avellan-Hietanen, 2021). These figures emphasise the significance of OSA; therefore, research on *treatment effectiveness* is integral to the further development of medical care.

Almost half of the Finnish lung disease referrals are related to OSA (Bachour & Avellan-Hietanen, 2021). According to the study conducted by Mattila et al. (2022), the annual number of OSA outpatient visits to the Finnish healthcare system increased from 9 700 (in 1996) to 122 000 (in 2018). The expanding number of patients seeking medical care has caused significant resource strains in special healthcare units (Bachour & Avellan-Hietanen, 2021). Furthermore, the process for diagnosis and care is complex; unique patients often require multiple hospitalisations to undergo sleep studies or surgical interventions. The current solution for finding an optimal OSA treatment is a trial-and-error process that often results in treatment failure (Kapur et al., 2017; Pépin et al., 2022). Subsequent treatment failures cause relapsing patients who are then redirected back to the healthcare pipeline for further investigations and new treatment trials. Nevertheless, specific OSA treatments are irreversible surgical interventions, making treatment reallocation an unsustainable approach. The resource requirements for treating OSA are accompanied by a high cost. For example, according to the recent Finnish study conducted by Mattila et al. (2022), the cumulative number of annual work absences due to OSA increased from 1 100 days to 46 000 between 1996 – 2018. Therefore, the economic impact of OSA affects the overall healthcare expenditure and extends to businesses through sick leaves and healthcare costs.

Besides OSA treatment effect research providing an avenue for the healthcare system to optimise its services, the actual impact happens on an individual level. An effective OSA treatment prevents the occurrence and further development of morbidities such as type 2 diabetes, strokes, or heart attacks (Vijayan, 2012; Newman et al., 2005; Gleeson et al., 2022). Aside from physical health problems, the disease can significantly reduce a person's life quality and impact work, family, and social life (Antic et al., 2011; Bergeron et al., 2020). Furthermore, OSA treatment decreases the risk of mortality (Yuan et al., 2015; Guo et al., 2016). Against the backdrop of an increasing OSA patient population, the exposure to the associated risks becomes more evident. Thus, there is a need to capitalise on previous treatment responses by researching OSA treatment outcomes. Furthermore, an enhanced understanding of the OSA treatment effect supports the optimisation of treatment allocation, which reduces queues and upgrades the overall quality of healthcare services.

1.2 Research Problem and Objectives of the Study

Previous observational studies on the OSA treatment effect have often centered on one or two treatment interventions and covered failures in those treatment modalities (Moeller et al., 2021; Martínez-García et al., 2012; Moxness et al., 2014; Marin et al., 2005). However, broader long-term comparative research on OSA treatment effect is scarce (Venema et al., 2020). Therefore, research efforts should be refocused on longer follow-ups instead of short-term comparisons between individual treatment interventions.

The purpose of this study is to compare the long-term effectiveness of four OSA treatment interventions. To determine treatment effectiveness, emphasis is placed on treatment failure or *relapse.* Relapse is defined as a transition to a disease state of increased *severity.* Nevertheless, the definition of OSA severity has continuously evolved, resulting in conflicting research on the assessment of OSA severity (e.g., Edwards et al., 2014; Bakker et al., 2014; Korkalainen et al., 2019). Consequently, as a prerequisite for evaluating patient relapse, this study investigates whether there are differences between the OSA severity levels. Therefore, the first research question covers severity classification:

Research question 1: Are there statistical differences between OSA severity levels?

After statistically validating the severity categories, the following phase of the study considers treatment analysis. Treatment analysis includes four OSA treatment methods:

CPAP, OA, upper airway surgery and bariatric surgery. The sample of this study is divided into four treatment subsamples according to treatment prescriptions. Each of the treatment subsamples represents a distinct *treatment arm*. The following research question covers treatment effect analysis.

***Research question 2*:** Do the different OSA treatment interventions lead to different outcomes?

Research question 2 guides in comparing the four treatment interventions with one another. The outcome of interest is relapse, a transition to a disease state of increased severity. Therefore, this study considers three types of relapses:

- Relapse from 'no OSA' to 'mild OSA'
- Relapse from 'mild OSA' to 'moderate OSA'
- Relapse from 'moderate OSA' to 'severe OSA'

OSA treatment outcomes are measured with *1) the probability of relapse* and *2) cumulative relapse recurrence*. Firstly, the probability of relapse indicates the likelihood of each type of relapse. Secondly, cumulative relapse recurrence estimates the progressiveness of the disease. The three types of relapse recurrence are based on the three relapse types. Relapse recurrence is explained in Figure 12.

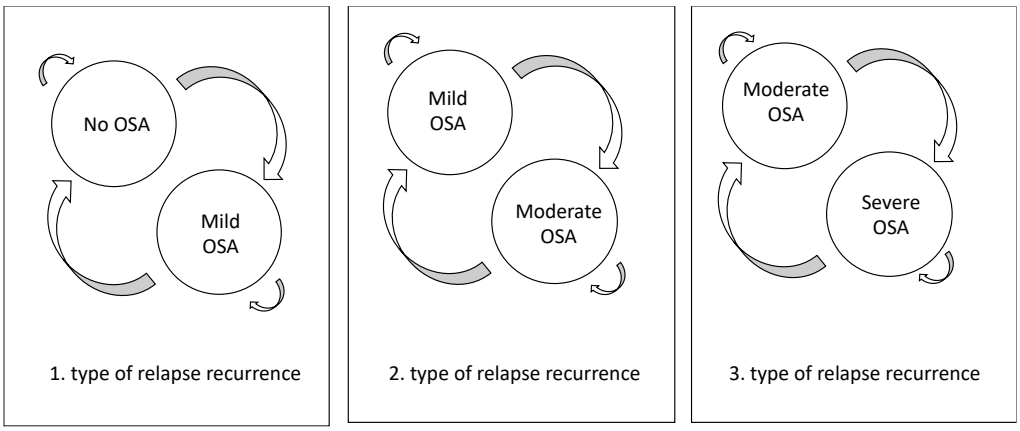


Figure 12. The Three Types of Relapse Recurrence

The interest lies in the cumulative number of expected relapses, assuming each type of relapse is repeatable. In other words, we investigate the number of times we expect relapse

to occur repetitively in each relapse type. A cumulative hazard function then measures the total amount of risk accumulated up to a certain point (Klein, 1991).

1.3 Scope of the Thesis

The scope of this thesis is defined as follows:

- The analysis considers the four official OSA treatment methods stated in the *Finnish Current Care Guidelines (2022):* CPAP, OA, upper airway surgery and bariatric surgery.
- The data of this study contains biological samples from specialised healthcare visits (instead of primary healthcare visits).
- All study subjects were treated within the Hospital District of Southwest Finland.
- All study subjects were diagnosed with OSA according to the *Finnish Current Care Guidelines (2022)*. Patients diagnosed with central sleep apnoea (CSA) are beyond the scope of this study. Additionally, all subjects are over 18 years of age.
- The study cohort consists of nearly 24 700 patients followed between 2003 and 2019.

1.4 Structure of the Thesis

This thesis aims to follow a logical structure. First, chapter 2 covers the literature review, which entails four concrete sections: section 2.1 reviews the previous literature involving OSA, and section 2.2 focuses on the latest research advancements concerning OSA treatment. Subsequently, section 2.3 addresses relevant modelling concepts to EHR data, and section 2.4 presents the hypothesis development of the study. Chapter 3 explains the research methodology. Research methodology is divided into two segments that structure the research process: section 3.2 explains the methods employed for severity classification, and section 3.3 covers the modelling choices for treatment effect investigations. Third, chapter 4 introduces the data in this study. This part presents the dataset, model variables, sample selection process, data pre-processing and finally, a descriptive analysis of the dataset. The large emphasis on the extraction and pre-treatment of the data is justified as it composed a large part of this work. Lastly, chapters 5 and 6 discuss the results of this study more thoroughly. In addition to a summary and discussion of the study results, chapter 6 is also reserved for discussing the limitations, ethical considerations, and possible future research avenues.

2 Literature Review

This chapter reviews recent literature on OSA and its treatment methods. First, section 2.1 concerns an overview of OSA. Subsequently, section 2.2 examines the most common OSA treatment methods. Then, section 2.3 presents standard concepts associated with longitudinal data and section 2.4 covers hypothesis development.

2.1 Obstructive Sleep Apnoea

OSA is a common sleep-related breathing disorder characterised by the cessation of breath resulting from recurrent upper airway collapses during sleep (e.g., Peppard et al., 2013; Rossi et al., 2021; Punjabi, 2008; Bikov et al., 2020). Obstructive events generate progressive asphyxia, often caused by the softening of the muscles in the back of the throat (Cowie, 2017; Spicuzza et al., 2015). The apnoea-hypopnoea index (AHI) is a parameter that measures the number of obstructive events per hour of sleep (Duodecim, 2022). Consequently, an obstructive event can be either “apnoea” when referring to a complete or near-complete cessation of breath for 10 seconds or more, or “hypopnoea” when referring to a 30% reduction in breathing for 10 seconds or more (Duodecim, 2022). Interrupted ventilation causes a below-normal level of oxygen in the blood, also known as hypoxemia (e.g., Dewan et al., 2015; Farré et al., 2018; Cowie, 2017). Chronic exposure to intermittent hypoxemia generates hypoxia, a state where body tissues become deprived of adequate oxygen supply (Cowie, 2017). Symptoms of OSA are categorised into diurnal and nocturnal symptoms. During the daytime, patients experience tiredness, cognitive impairment, decreased libido, sore throat, mood swings, morning headache and a tendency towards depression. (e.g., Jacobsen et al., 2013; Spicuzza et al., 2015). Nocturnal symptoms include irregular and loud snoring, excessive perspiration, apnoeic sleep episodes, arousal, insomnia, and nightmares (Cowie, 2017; Duarte et al., 2022).

2.1.1 Disease Progression and Morbidities

According to a large body of research, OSA has the characteristics of a chronic condition that tends to progress slowly over the years (Heatley et al., 2013; Sahlman et al., 2007; White & Younes, 2012). Marin-Oto et al. (2019) presented a framework with four stages to describe OSA disease progression: First, the ‘stage of susceptibility’ is characterised by initial symptoms such as snoring. Second, the ‘pre-symptomatic stage’ is characterised by nocturnal symptoms; however, patients at this stage are less likely to report diurnal

symptoms. Thirdly, the condition evolves to a 'stage of clinical illness', which is characterised by the development of morbidities. Finally, untreated patients might experience disability or death due to the unbalance of the cardiovascular system. Figure 3 outlines this process:

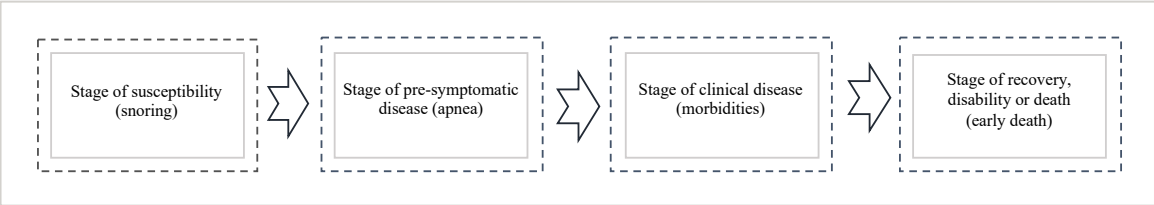


Figure 3. OSA progression chart. (Adopted from Marin-Oto et al., 2019)

Studies have shown that OSA severity progression depends mostly on weight gain and, to a less degree, on time (Newman et al., 2005; Berger et al., 2009; Schwartz et al., 2008). Furthermore, evidence suggests a bidirectional relationship between OSA and other coexisting health conditions (Gleeson et al., 2022; Dewan, 2015; Bonsignore et al., 2019; Rossi et al., 2021). A critical factor influencing the development of morbidities is oxygen deprivation caused by interrupted ventilation (Cowie, 2017). Oxygen deprivation generates low levels of oxygen in the blood and produces a state called hypoxemia, which can consequently trigger chronic hypoxia (e.g., Dewan et al., 2015; Farré et al., 2018; Cowie, 2017). Over time, sleep deprivation, recurrent apnoeic episodes and hypoxia impact the proper functioning of the metabolism, affecting the cardiovascular system and altering the metabolic balance (Spicuzza et al., 2015; Drager et al., 2010; Kent et al., 2011).

2.1.2 Diagnosis of the Condition

In Finland, OSA diagnosis is based on three factors: (1) physical examination, (2) anamnesis, and (3) overnight polysomnogram (Duodecim, 2022). First, physicians conduct extensive physical examinations to evaluate the respiratory, cardiovascular, and nervous system (Aro et al., 2019). Second, patients who report signs indicating OSA presence are referred for anamnesis (Aro et al., 2019). In anamnesis, health experts evaluate patient history, occupational factors, allergies, lifestyle choices and coexisting health conditions. In addition, sleepiness symptoms are assessed with the Epworth Sleepiness Scale (ESS) survey, which measures a patient's general daytime sleepiness (Aro et al., 2019). Thirdly, patients are directed to a sleep study at home or a hospital (Aro et al., 2019). The overnight polysomnogram conducted in a sleep laboratory is the standard diagnostic test (e.g., Punjabi et al., 2008; Spicuzza et al., 2015; Kapur et al., 2017). However, the lack of

standardisation in sleep studies challenges the comparativeness of sleep study results across laboratories (Hirshkowitz, 2016).

2.1.3 Severity Classification

The *Finnish Current Care Guidelines* (2022) follow international standards and hence consider the apnoea-hypopnoea index (AHI), oxygen saturation percentage (SaO2%) and daytime sleepiness as the three critical indicators of OSA severity. The severity levels with their respective measurement metrics and thresholds are explained in Table 1.

Table 1: The Finnish Severity Classification Guidelines

	Measurement Metrics		
Severity level	AHI	SaO2%	Daytime sleepiness
Mild	5 – 15	On average ≥ 90 **and** minimum ≥ 85	Minor daytime sleepiness in social and occupational situations.
Moderate	16 – 30	On average < 90 **and** minimum ≥ 70	Daytime sleepiness in situations that require moderate concentration.
Severe	> 30	On average < 90 **and** minimum < 70	Daytime sleepiness in situations that require deep concentration.

Adapted from the *Finnish Current Care Guidelines* (2022)

AHI is the most widely recognised OSA severity indicator in scientific societies (Nishibayashi, 2008; Won, 2020). A significant number of treatment analyses consider AHI as a metric to determine the successfulness of a treatment intervention (e.g., Khan et al., 2009; Cielo et al., 2019). Nevertheless, the connection between AHI and OSA severity remains unclear (Linz, 2019; Prasad, 2016; Ahmadi, 2009; Hudgel, 2016). Despite standardisation attempts, uncertainty around the hypopnoea definitions persists (Berry et al., 2012; Ruehland et al., 2009; Hudgel et al., 2016). Some sources measure the nocturnal decrease in airflow, while others track the decrease in oxyhaemoglobin saturation (Hudgel, 2016; Ruehland et al., 2009). Another OSA severity indicator is the percentage of oxygen-saturated haemoglobin molecules in arterial blood (SaO2%) (Hafen et al., 2022). A healthy level of SaO2% is approximately 95%; levels below 90% require external oxygen supplementation, whereas levels below 70% are life-threatening. (Hafen et al., 2022). In

addition to factors such as AHI and SaO2%, OSA severity assessment considers a patient's subjective perception of daytime sleepiness. Although occasionally, patients with higher AHI tend to rank lower in sleep quality than individuals with lower AHI (McNicholas, 2017; Macey et al., 2010).

Nevertheless, more variables than AHI or SaO2% should be considered in measuring OSA severity (Hudgel, 2016). The diagnostic and coding manual of the American Academy of Sleep Medicine (2017) points out that *"...a single numerical cut point (such as apnoea index) is often not an appropriate division between levels of severity and clinical judgement of several indexes of severity is considered superior*". The ambiguity in OSA severity raises difficulties in generalising inferences among clinical studies. In addition to this, the polysomnogram (PSG) scoring criteria has experienced changes over time, but the severity classification standards have not been calibrated to adjust to these changes (Hudgel, 2016).

2.1.4 Risk Factors

Findings from extensive population studies have contributed to a better understanding of the OSA risk factors. Table 2 summarises structural and non-structural OSA risk factors.

Table 2: OSA Risk Factors

Structural Risk Factors	Non-structural Risk Factors
Anatomic variation	Obesity
Facial deformations	Male sex
Shorter jaw length	Age
Abnormalities in skull shape	Postmenopausal state
Abnormal growth of tonsils	Smoking
Down syndrome	Family history
Pierre Robin, Marfan, and Prader-Willi syndromes	Habitual snoring
High, arched palate (particularly in women)	Supine sleep position
Inferior displacement of the hyoid	Substance abuse

Adapted from Buchanan et al. (2016) 'Cone-beam CT analysis of patients with obstructive sleep apnea compared to normal controls.'

A substantial contributor to the increasing prevalence of OSA is the rising obesity among the population (Newman et al., 2005; Berger et al., 2009; Schwartz et al., 2008; Bikov, 2020). According to Tuomilehto et al. (2013), approximately 70% of OSA patients suffer from obesity. Furthermore, the study conducted by Ong et al. (2013) shows a reciprocal

relationship between weight gain and OSA, suggesting that the metabolic unbalance caused by OSA contributes to further weight gain or the inability to lose weight. Other non-structural risk factors include male sex, age, menopause, family history, and behavioural factors such as alcohol consumption and smoking (Duodecim, 2022). In contrast, structural risk factors entail risks associated with abnormal skeletal structures and excessive soft tissue in the upper airway (Buchanan et al., 2016). For example, a narrow upper airway is a common abnormality in OSA diagnosed individuals (Junior, 2010). Furthermore, small mandibles, large tongue area, inferior hyoid bones and narrow posterior airway spaces are all more common in nonobese OSA patients (Yu et al., 2003).

2.2 Treatment Methods

This section covers essential information regarding the relevant treatment interventions in this study.

2.2.1 Treatment Methods and Indications

This study considers four OSA treatment interventions: CPAP, OA, upper airway surgery, and bariatric surgery. Table 3 presents the treatment indications for each treatment. Treatment indication refers to the reasoning behind any treatment prescription.

Table 3: Treatment Methods and Treatment Indications

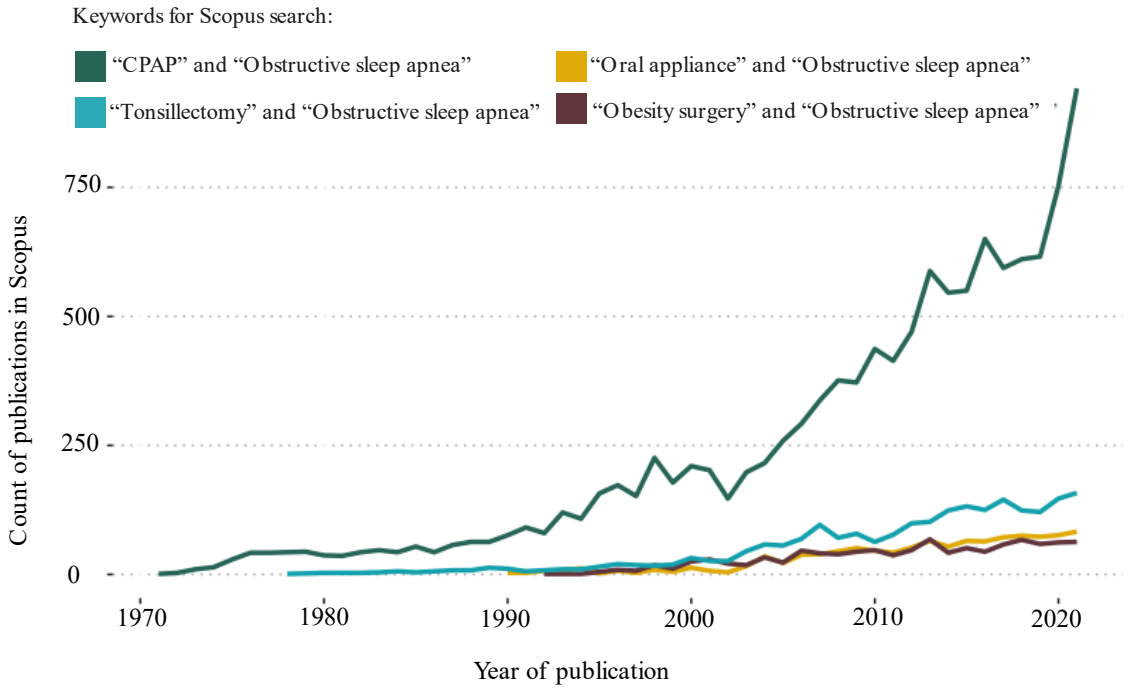
Treatment Methods	Treatment Indications
CPAP	Clinical symptoms include daytime sleepiness, cognitive impairment, sleep problems or cardiovascular disease (*Finnish Current Care Guidelines,* 2022).
Upper airway surgery	Firstly, an obstructive abnormality of the upper airway (e.g., large tonsils). Secondly, contraindications for non-surgical therapies. Thirdly, to improve other OSA therapies, such as OA (*Finnish Current Care Guidelines,* 2022).
Bariatric surgery	A body-mass index (BMI) > 35 kg/m2, aged 18-60 and showing a clear motivation for a lifestyle change (*Finnish Current Care Guidelines,* 2022).
OA	Intolerance to CPAP therapy or patient preference for alternative therapy (American Academy of Sleep Medicine, 2017).

OSA treatment practice focuses on symptom improvement and preventing long-term health risks. CPAP is the most common OSA treatment and often the first intervention offered to patients seeking symptomatic relief (Bachour & Bäck, 2015). Although a crucial factor in OSA treatment is the interaction of the upper airway anatomy, each patient's

specific pathology is not always assessed as part of the diagnostic process (Duodecim, 2022). For this reason, treatment selection is often a trial-and-error process that starts with the CPAP method (Kapur et al., 2017; Pépin et al., 2022).

2.2.2 Previous Research on OSA Treatment Effect

Previous research on individual OSA treatment interventions has been proliferating, but comparative research between treatment interventions is scarce (Venema et al., 2020). Figure 4 displays the growing number of publications on individual OSA treatment methods.



Data from Scopus database (04/2022)

Figure 4. Scopus Publication Volume on OSA Treatment Methods (1970-2021)

The prevalence of CPAP as a standard OSA treatment intervention explains the growth of publications containing the keyword 'CPAP'. Despite a high number of yearly publications, a significant limitation in OSA treatment research is the lack of long-term follow-up research (Moeller et al., 2014; Martínez-García et al., 2012; Moxness et al., 2014; Marin et al., 2005; Venema et al., 2020). Long-term research has been challenging to implement before 2010 due to the lack of EHR data (DesRoches et al., 2013). Furthermore, no long-term observational analyses containing the four most common treatment interventions were found during this study. The following subsections, 2.2.3, 2.2.4, 2.2.5 and 2.2.6, explore each of the four treatment interventions in detail.

2.2.3 Continuous Positive Airway Pressure (CPAP)

CPAP is recognised as the standard treatment method for OSA patients (e.g., Bachour & Bäck, 2015; Antic et al., 2011; Caples et al., 2005). The CPAP machine creates pressure that supports the opening of the collapsing upper airway during sleep (Issa et al., 1984; Padma et al., 2007; Li et al., 2020). According to a substantial number of studies, CPAP effectively controls the symptoms of OSA and improves daily functioning, cognitive processing, and quality of life (e.g., Weaver, 2019; Vakulin et al., 2017; Ryan et al., 2005; Nural et al., 2013).

Despite previous evidence supporting the effectiveness of CPAP, not all studies corroborate the effectiveness of CPAP in treating OSA (e.g., Shapiro et al., 2010). Some specific factors, such as nose, throat, or glottis lesions, can contribute to the failure of CPAP treatment (Li et al., 2020). Furthermore, treatment adherence is crucial for maintaining CPAP efficacy, and failure in treatment commitment can lead to the re-emergence of symptoms (Weaver et al., 2007). Nonetheless, failing to consider CPAP adherence often generates bias in research outcomes (Shapiro et al., 2010).

2.2.4 Oral Appliance (OA)

Oral appliances (OA) are custom-fitted dental devices that maintain a patient's airway open during sleep (Sutherland et al., 2014). The device repositions the lower jaw slightly forward and enlarges the upper airway by simultaneously preventing the tongue from blocking the airway during sleep (e.g., Sutherland et al., 2014). A large body of research demonstrates the efficacy of OA in reducing obstructive breathing events (Okuno, 2016; Zhu et al., 2015; Ilea et al., 2021). As a result, OA is considered an alternative to CPAP, and numerous patients prefer OA due to its user-friendliness (Gotsopoulos et al., 2002). Although clinical trials confirm that CPAP is more efficient in reducing OSA symptoms, this superiority does not necessarily translate into more effective health outcomes in clinical practice (Sutherland et al., 2014; Balk et al., 2011). In addition, previous comparative research has reported higher OA adherence compared to CPAP. Thus, OA's inferiority in treatment efficacy could be compensated by greater treatment adherence (Sutherland et al., 2015; Li et al., 2013). Nonetheless, there might be adverse effects associated with OA compared to CPAP. For example, short-term adverse effects include painful teeth, increased salivation, discomfort, and long-term adverse effects comprehend dentoskeletal alterations such as teeth movements (Minagi et al., 2018; Baldini et al., 2021).

2.2.5 Upper Airway Surgery

Upper airway surgery attempts to surgically modify dysfunctional anatomical areas connected to the upper airway area (e.g., Mackay et al., 2020). The procedure consists of enlarging a patient's upper airway by reducing the amount of soft tissue or widening the airways with bony reconstructions (e.g., Kemppainen et al., 2019). Consequently, a stiffer soft palate is less likely to touch the back wall of the throat as the muscles relax during sleep. Upper airway surgery has been proven to normalise symptoms, reduce AHI values, and improve SaO2% levels (Kemppainen et al., 2019). A standard surgical procedure for OSA is uvulopalatopharyngoplasty (UPPP) (Won et al., 2008). UPPP is often performed in conjunction with tonsillectomy (Won et al., 2008). Furthermore, research has shown a surprising success rate in tonsillectomy as an OSA treatment (e.g., Khan et al., 2009; Lee et al., 2012). Further developments in surgical upper airway interventions include multilevel surgery that combines various surgical methods to attain higher treatment effectiveness (Mackay et al., 2020). Nevertheless, multilevel surgery is beyond the scope of this study.

2.2.6 Bariatric Surgery

Weight loss reduces AHI in obese patients; hence obesity surgery has been proven an efficacious option for treating OSA (Sarkhosh et al., 2013; Cowie, 2017; Fritscher et al., 2007; Greenburg, 2009; Rao, 2009; Ashrafian, 2015). As of 2022, the two main types of bariatric surgeries performed in Finland are sleeve gastrectomy and the Roux-en-Y gastric bypass procedure (Duodecim, 2022). However, data in this study include a timeframe where the laparoscopic adjustable gastric banding procedure was still conducted in Finland, and it is thus considered in this study.

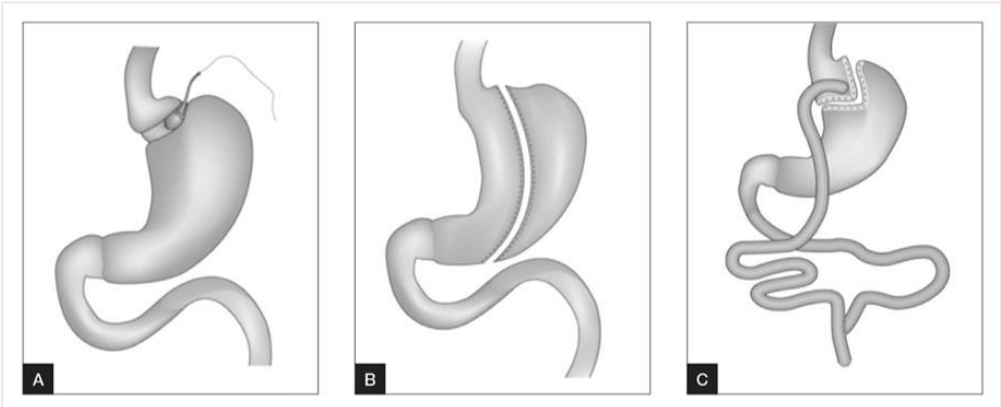


Figure 5. The Three Bariatric Surgery Types in Finland: (A) Laparoscopic Adjustable Gastric Banding, (B) Sleeve Gastrectomy and (C) Roux-en-Y Gastric Bypass Procedure. (Dixon et al., 2011/ permission granted)

Firstly, laparoscopic adjustable gastric banding (A) is a minimally invasive weight loss surgery. The procedure consists of placing a restrictive silicone ring around the upper part of the stomach, thus decreasing food intake by only filling the top part of the stomach (e.g., Himpens et al., 2011). Secondly, sleeve gastrectomy (B) consists of removing a significant part of the stomach to restrict food intake (Sarkhosh, 2013). Thirdly, Roux-en-Y gastric bypass (C) alters the gastrointestinal physiology by restricting food intake and enabling faster absorption of nutrients to the small intestine (Abdeen et al., 2016). The following section, 2.3, explains essential concepts regarding EHR treatment effect research.

2.3 Treatment Effect Research in Longitudinal Study Settings

This section explains relevant concepts to EHR studies. First, subsections 2.3.1 and 2.3.2 explain treatment effect research and longitudinal studies. Then, subsection 2.3.3 explores the basics of multistate models in life history analysis and subsection 2.3.4 presents essential concepts in survival analysis. Finally, subsection 2.3.5 dives into some critical challenges in longitudinal studies.

2.3.1 The Practice of Treatment Effect Research

Treatment effect research aims to assess the effectiveness of different treatments or clinical interventions. Moreover, the treatment effect measures whether there are any benefits in prescribing a specific treatment over others (Yao et al., 2018). Efficacy and effectiveness are often distinguished in the following way: treatment efficacy refers to the success of an intervention under ideal circumstances, whereas effectiveness refers to treatment performance in the real world, where conditions cannot be controlled (Sutherland et al., 2015). EHR data proceed from actual medical visits; hence longitudinal studies often represent the latter. According to Sutherland et al. (2015), a significant decrease in AHI indicates an effective OSA treatment. Furthermore, a large body of research (Nishibayashi et al., 2008; Johansson et al., 2009; Boyd et al., 2013; Matiello et al., 2010) has based the evaluation of OSA treatment effect solely on AHI values. However, as discussed in subsection 2.1.2, accounting only for AHI to determine disease severity is controversial.

2.3.2 Longitudinal Studies involving EHR

Longitudinal studies consist of repeatedly measured variables (e.g., blood pressure) within the same clusters (e.g., patients) (Twisk, 2013, p.1). Longitudinal research is observational, as subjects are not explicitly involved in the study process (Gail et al., 2019; Twisk, 2013,

p.7; Cook & Lawless, 2007, p.16). EHRs often serve as a data source for longitudinal studies. EHRs are records of a patient's medical history tracked and managed by healthcare professionals. For every individual patient, there is typically a longitudinal record of medications, medical visits and biomarkers that proceed from laboratory data. The primary benefits of EHRs are improved patient care and reduced errors, as EHRs provide easy access to complete patient histories. The rapid adoption of EHR systems has resulted in the increased availability of data generated as a byproduct of routine healthcare visits (Lin et al., 2022). EHR data convey benefits to biomedical research, as access to large patient datasets enables a wide variety of longitudinal research designs (Huang et al., 2020). Nonetheless, medical visits are often irregular, generating a discontinuity in measurements as well as substantial information loss (Huang et al., 2020).

2.3.3 Multistate Models for Studying the Disease Process

Multistate modelling is an established statistical technique commonly applied for describing longitudinal data (e.g., Andersen & Keiding, 2002). These models are utilised to model the relationships between different states (Cook & Lawless, 2007, p.14). A multistate process consists of a finite number of states, and any transition between states may be considered (Andersen & Keiding, 2002). The procedure is a stochastic process in which a study subject can occupy one state out of discrete states at different time points (Allignol, 2011). In probability theory, a stochastic process reflects a set of variables that represent the evolution of a process (Allignol, 2011). This way, multistate models can be applied to model the probability of an individual transitioning from one disease state to another (e.g., Ching, 2013, p.3; Allignol, 2011). The complexity of a multistate model depends on the number and progressiveness of the states. Multistate models typically satisfy the Markov property, which assumes that every event depends on a previous event (e.g., Zhang et al., 2010). A multistate model is an extension of classical survival analysis; thus, the following subsection, 2.3.4, explains the most basic concepts regarding survival analysis.

2.3.4 Key Concepts in Survival Analysis

Survival analysis is a method for statistical analysis where the outcome of interest is the time T^* until an event occurs (e.g., Miller et al., 2011, p.8). An event can be death, disease incidence or relapse, and time is calculated from the beginning of the follow-up of an individual (Kleinbaum & Klein, 2012, p.4). Survival functions are applied in several fields,

but they are ubiquitous in biomedical studies to analyse disease occurrence or life expectancy among treated and untreated patients (e.g., Liu, 2012; Fizazi et al., 2012; Mueller, 2018). In this study, survival regression is applied along with multistate models to investigate failure rates over time. Standard statistical regression is inappropriate because survival times are often incompletely observed (Kleinbaum & Klein, 2012, p.23). Furthermore, patients often enter and leave clinical studies at different time points. Therefore, *censoring* and *truncation* account for the distinct entry and exit times in survival models (Koul et al., 1995). Censored individuals are study subjects that have not experienced the outcome of interest within the observation period (e.g., Kleinbaum & Klein, 2012, p.286). This study assumes right-censoring. Right-censoring happens when the survival time exceeds the censoring time or when a study subject abandons the study before the event of interest occurs (Kleinbaum & Klein, 2012, p.286). Furthermore, survival time could exceed censoring time if an individual remains alive after the study has terminated or the subject is lost to follow-up (Kleinbaum & Klein, 2012, p.17). In epidemiologic studies, patients are followed from the time of study entry and not from time 0 (in the relevant time scale, such as age) (Leung et al., 1997). Thus, in addition to right-censoring, longitudinal studies are subject to left-truncation (Aalen et al., 2008, p.4). Left-truncation occurs when records hold data from subjects who have already survived until the study entry (Aalen et al., 2008, p.4).

2.3.5 Challenges in EHR Longitudinal Studies

The complexities and limitations of EHR data make medical findings prone to biases (Huang et al., 2020). Some challenges relevant to this study include the following:

- *Attrition:* Longitudinal studies experience attrition as individuals withdraw from the study or are lost to follow-up (Casey et al., 2016). Attrition can result from death or participants moving away.
- *Correlated responses:* Statistical analysis of longitudinal EHR data often requires methods that consider the correlations in response measurements (Hedeker, 2003).
- *Data missingness:* The large amount of EHRs pose challenges in ensuring that the data is collected consistently and accurately (Huang et al., 2020). The data collected from EHR databases reflect not only the health of the patients but also the interactions between patients and healthcare professionals. Measurements are irregular since individuals are observed at discrete time points during routine medical visits. Therefore, understanding assumptions related to missing data is vital

for finding a suitable missing value imputation method (Twisk, 2013, p.215). Data can be classified as missing completely at random (MCAR), missing at random (MAR), or missing not at random (MNAR) (e.g., Sainani et al., 2015). Table 4 explains the definitions behind the three distinct types of missingness.

Table 4: Types of Missingness

Type of missingness	Description
Missing Completely at Random (MCAR)	Missingness is not dependent on any observed or unobserved variables. There are no systematic differences between missing and observed data.
Missing at Random (MAR)	Missingness may depend on observed variables but does not depend on unobserved variables. Therefore, there is a systematic relationship between the missing values and observed data.
Missing Not at Random (MNAR)	Missingness is dependent on unobserved variables. Thus, data missingness is systematically related to unobserved data (factors not measured by the researcher).

Adapted from *Sainani et al. (2015) 'Dealing with missing data'*

Other essential factors in handling EHR data for research purposes include ensuring the privacy of the study subjects and maintaining a functional data management architecture for the project (Huang et al., 2020). The following section, 2.4, explains the hypothesis development procedure.

2.4 Hypothesis Development for OSA Treatment Effect Analysis

The purpose of this study is to compare the four major OSA treatment interventions with one another. To accomplish that, we must first validate the OSA severity levels as a prerequisite for comparing treatment outcomes. OSA severity levels are relevant as they help define treatment effectiveness. Therefore, this study considers the following research questions:

RQ1: Are there statistical differences between OSA severity levels?
RQ2: Do the different OSA treatment interventions lead to different outcomes?

The literature review accentuates the problem statement of this study. Previous research on OSA treatment interventions is limited to short-term follow-ups of one or two treatments,

but long-term comparative research is scarce (Moeller et al., 2021; Martínez-García et al., 2012; Moxness et al., 2014; Marin et al., 2005; Venema et al., 2020). Furthermore, outcomes from distinct research publications are hard to compare. For example, the definition of OSA severity has constantly evolved, and there is inconsistent evidence on the factors that should be considered when assessing OSA severity (e.g., Edwards, 2014; Bakker, 2014; Korkalainen, 2019). Studies have often employed AHI as an OSA severity indicator (e.g., Khan et al., 2009; Cielo et al., 2019). Notwithstanding, previous evidence indicates that the OSA severity indicator should consider more variables than AHI (Linz, 2019; Hudgel, 2016; Kapur et al., 2017). Therefore, as a requirement for assessing patient relapse, this study investigates whether there is a statistical relationship between OSA severity (when defined solely with AHI) and other patient biomarker data. The first hypothesis is developed to validate and support the harmonisation of the severity variable.

Research hypothesis 1, (H_1): There are statistical associations between OSA severity and biomarker variables. Specifically, that OSA severity increases in ODI and decreases in SaO2%.

Previous literature supports a significant relationship between SaO2% and OSA severity, and patients with increased OSA severity tend to have lower oxygen saturation levels (Kainulainen et al., 2019; Myllymaa et al., 2015; Dewan et al., 2015; Farré et al., 2018; Hudgel, 2016). Similarly, we argue that the results of this study show a significant relationship between increased OSA severity and decreased oxygen saturation levels. In addition to oxygen saturation, the *Finnish Current Care Guidelines 2022* also consider ESS a relevant OSA severity indicator. Nevertheless, previous studies have reported a poor association between AHI and ESS (McNicholas, 2017; Macey et al., 2010). Therefore, this study does not expect any evident relationship between AHI-defined OSA severity and ESS. After validating the severity indicator, the second research hypothesis tests differences between treatment outcomes.

Research hypothesis 2, (H_2): There are statistical differences between the four OSA treatment arms. CPAP is expected to have the most substantial effect in decreasing the risk of relapse and relapse recurrence in OSA.

This research hypothesis tests whether the data provide statistical evidence on differences between treatment outcomes. Each of the four treatment arms represents a subgroup of participants that received a specific OSA treatment intervention. Previous studies support the effectiveness of CPAP over other OSA treatment interventions (Vakulin et al., 2017; Bäck & Bachour, 2015; Sutherland et al., 2015; Ryan, 2005). For example, studies argue that CPAP is more effective in reducing OSA symptoms than OA (Sutherland et al., 2014; Balk et al., 2011). Similarly, we expect CPAP to show superiority over other OSA interventions. However, although CPAP is the standard OSA treatment method, poor treatment adherence is common (Shapiro et al., 2010; Weaver et al., 2007). Moreover, it has been shown that CPAP does not enable a permanent cure, and withdrawal from treatment often leads to the re-emergence of symptoms (e.g., Weaver et al., 2007; Rasheid et al., 2013). Therefore, CPAP might still show relatively high estimates for relapse recurrence.

Furthermore, there is conflicting research on the relative effectiveness of bariatric surgery as an OSA treatment intervention. While some studies imply that bariatric surgery could prevent the recurrence of OSA and enable a permanent cure (Verse, 2005; Rasheid et al., 2013), others state that Bariatric surgery does not solely suffice for treating OSA (e.g., Peromaa-Haaviso et al., 2016). Additionally, it has been shown that the effectiveness of upper airway surgery diminishes over time (Lin et al., 2008). Thus, despite the disadvantages concerning CPAP treatment adherence, we expect that CPAP might generally be more effective in treating OSA than other interventions.

3 Research Methodology

This chapter covers the research methodology. Section 3.1 introduces the research process, section 3.2 presents the methods for the severity classification, and section 3.3 explains the methods for the treatment analysis.

3.1 Defining the Research Process

The purpose of this section is to structure the research procedure by briefly describing the research objectives and outlining the necessary research phases. This study aims to analyse and compare the treatment effect of four OSA treatment interventions. A key indicator to measure treatment response is *relapse*, defined as the worsening or reappearance of a disease after improvement or recovery. Consequently, OSA severity is the variable of primary interest for investigating relapse. Therefore, we propose a two-phased study design, where we first statistically validate the severity indicator and then model the relapse rates with a Markov procedure. After this, we compare the relapse behaviour between the four treatment arms. Figure 6 outlines the two phases of this study:

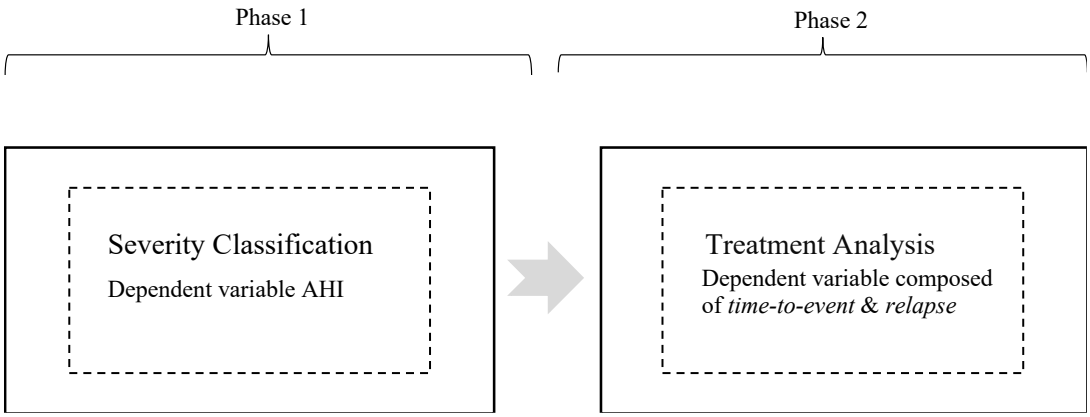


Figure 6. Illustration of Study Process

The *first phase* of this study consists of severity classification. As noted in subsection 2.1.3, OSA severity definitions are subject to controversies in clinical studies. Thus, it is necessary to quantify disease severity across treatment arms. Therefore, the *first phase* of the study considers an ordinal logistic regression analysis with an AHI severity indicator as a dependent variable. *The second phase* of the study consists of the treatment analysis and considers the 'relapse' variable built from the 'severity' variable as an input for modelling. This phase is based on a Markov state-transition model that investigates relapse among

treatment arms. Now, we have provided a summary of the research procedure. Section 3.2 explains the methodology for the first phase of the study.

3.2 Severity Classification

The objective of severity classification is to estimate the influence of physiological factors on OSA severity and thus validate the severity division applied later in this study. The dependent variable 'severity' is an ordinal response variable ranging from 1 to 4, with a natural ordering in levels. Each of the timestamped EHRs is classified as either (1) 'no OSA', (2) 'mild OSA', (3) 'moderate OSA' or (4) 'severe OSA'. The numeric coding of an ordinal variable is a naming convention, and these names should be considered labels rather than values (McCullagh, 1980). The division into severity states is based on the AHI thresholds stated in the *Finnish Current Care Guidelines* (2022). Figure 7 shows the AHI thresholds that dictate the rules for labelling the patient records in this study.

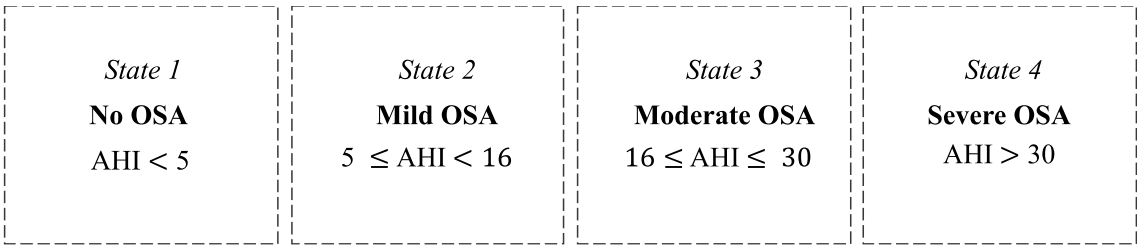


Figure 7. Four Severity Categories.

The EHR data of this study are in repeated measures format. Repeated measures data violate the assumption of independence by nature, and therefore standard ordinal logistic regressions are outside the scope of this study (Bakdash, 2017). Two model types are common for modelling individual responses that are correlated with one another: Generalised Linear Mixed Models (GLMM) and Generalised Estimating Equations (GEE) (Twisk, 2013, p.235). Theoretically, both approaches are equally appropriate and highly suitable for longitudinal data (Twisk, 2013, p.235). However, GLMM is known to be slightly more flexible than GEE (Twisk, 2013, p.81). This study employs a GLMM procedure called Cumulative Link Mixed Model (CLMM) to model clustered ordinal outcome data (Christensen, 2019). The CLMM is an extension of Cumulative Link Models (CLM) (Christensen, 2020). The following subsection, 3.2.1, briefly explains the theory behind CLM.

3.2.1 Ordinal Response Models

Ordinal regression models are often applied to analysing ordinal outcomes (McCullagh, 1980). Cumulative Link Models (CLMs) are ordinal regression models where continuous outcomes are categorised in ordinal classes according to thresholds (McCullagh, 1980). CLMs are based on the cumulative (or accumulated) response probabilities (Christensen et al., 2019). The cumulative response probabilities indicate the likelihood of a randomly selected observation falling in a specific class or below (Agresti, 2010). For example, in the OSA model, $P(Y \leq 2)$ would refer to the probability of being in state 'No OSA' or 'Mild OSA' versus being in state 'Moderate OSA' or above. The general form of CLM is noted as follows:

$$G^{-1}[P(Y \leq j)] = \alpha_j - X\beta \tag{1}$$

Where:
Y is the ordinal response variable,
j is the level of an ordered category with J levels. (e.g., j=1 equals 'no OSA'),
X is the model matrix,
β *represents the intercept (*β_0*) and the coefficients for each regressor (*$\beta_1, \beta_2, \beta_3 \ldots \beta_m$*),*
α_j *represents the threshold for level j, j=1,…J for an ordinal variable with J levels,*
G^{-1} *is the link function (logit for this study)*

The link function relates the expected value of the response to the linear model predictors (Greenwell et al., 2018; Christensen et al., 2019). Generally, *logit* is the most common link function due to its simple interpretability and computational convenience (Agresti, 2010). This study applies the logit link function for modelling. The logit link represents the inverse cumulative density function of a logistic probability distribution (Agresti, 2010). Furthermore, CLMs are known to have two main assumptions. One is the proportional odds (PO) assumption, and the other is the independence assumption (Christensen, 2015). following subsection 3.2.2 explains how to extend the CLM to deal with repeated-measures data, while subsection 3.2.3 covers the PO assumption.

3.2.2 Random Effects Account for Patient Clusters

Logistic regression models assume independence of observations (Hedeker, 2003). Nevertheless, repeated-measures longitudinal data violate this assumption by nature. In this study, multiple patient measurements originate from a single patient; therefore, each

patient forms a nested unit of multiple measurements. Thus, to consider the correlated observations, we include a random effect as a grouping structure to the CLM (e.g., Hedeker et al., 2009). Furthermore, the CLMM assumes normally distributed random effects (Christensen, 2018). Moreover, this study applies an extension of the CLM that includes random effects in the location part of the predictor. Thus, the cumulative link mixed model formula with random effects is written as follows:

$$G^{-1}[P(Y \leq j)] = \alpha_j(Z_{t[i]}u_t - X_i\beta) \quad (2)$$

Where:

u_t is the vector of coefficients for the group-level predictor,

$Z_{t[i]}$ represents the group-level predictors for observation *i* in cluster *t*

The random effect considers the intra-cluster correlations between observations (Hedeker, 2003). However, model estimates may be unstable for small observation clusters (Liang, 1993). Random effects deploy partial pooling, while fixed effects do not (Bartels, 2008). Partial pooling happens when data are grouped, and the effect estimates are based partially on the more abundant data from other groups (Bartels, 2008). Partial pooling then masks group-level variation between samples (Bartels, 2008).

3.2.3 The Proportional Odds Assumption

An ordinal regression model assumes a common slope for the effect of any of the explanatory variables in the model (McDonald et al., 2010). The PO assumption considers that the differences regarding the logit of the cumulative probability for $Y \leq j$ are constant for the values of X.

$$G^{-1}[P(Y \leq j \mid X)] - G^{-1}[P(Y \leq i \mid X)] = \alpha_j - \alpha_i \quad (3)$$

The PO assumption is satisfied when no predictor variable disproportionally affects a specific level of the response variable (Brant, 1990). A violation of the assumption indicates that the effects of a predictor variable differ across cutpoint equations in the model (Argresti, 2010). Therefore, the modelling approach fails because it cannot reduce the model's coefficients to a single set across all ordinal response levels (Christensen, 2019). The PO assumption can be tested via a likelihood ratio test (LRT) (Murphy et al., 1997). The test hypothesis is that the model fit does not improve by relaxing the PO

assumption. Therefore, significant *p*-values from this test indicate model failure (Christensen, 2015). Furthermore, failing the PO assumption results in unreliable model parameters (Christensen, 2019). In addition to the PO assumption, CLMMs make assumptions about an underlying latent distribution, but they are robust to violations of these assumptions.

3.2.4 CLMM Model Diagnostics

Model diagnostics refer to a set of procedures applied to assess the validity of the results of a regression analysis (Greenwell et al., 2018). Methods for model diagnostics include graphical methods, quantitative approaches, and hypothesis tests. Nevertheless, there is limited availability of diagnostic tools for ordinal regression models (Liu et al., 2018). For example, traditional goodness-of-fit metrics are generally unavailable for GLMM ordinal models (Lorenzo-Arribas, 2019). Furthermore, the nature of ordinal outcomes poses challenges in defining residual statistics that are valid and simple to interpret (Liu et al., 2018). As advised in the work of O'Connell and Liu (2011), this thesis assesses model residuals by converting the ordinal response variable back to a continuous variable and applying a linear mixed-effects model residual analysis. In addition to analysing model residuals, this study applies an LRT to test whether the final model explains the outcome better than a reduced model. The LRT is formulated as follows:

$$LRT = -2ln\left(\frac{L(m_1)}{L(m_2)}\right) \tag{4}$$

Where m_1 represents the reduced model and m_2 the full model.

The LRT assesses how well a model explains an outcome compared to a model with fewer predictors (Christensen, 2018). This study builds the reduced model based on a few non-significant independent variables from the original model. Thus, the LRT examines whether one model fits the data significantly better than a reduced model.

3.3 Treatment Analysis

This section presents the methodology for the treatment analysis. The section is composed in the following manner: subsections 3.3.1 and 3.3.2 discuss the objectives and problem characterisation of treatment analysis. The following subsections, 3.3.3 and 3.3.4, explain the approach for modelling patient relapse. Finally, subsection 3.3.5 discusses the OSA

multistate model; subsections 3.3.6 and 3.3.7 cover the estimators, and subsection 3.3.8 explains hypothesis testing.

3.3.1 Objectives of Treatment Analysis

The objective of treatment analysis is to compare the clinical effectiveness (or *ineffectiveness*) of four OSA treatment alternatives. The treatment interventions considered in this study are CPAP, OA, upper airway surgery and bariatric surgery. Treatment effectiveness is measured by treatment failure, which is characterised by an occurrence of relapse after treatment prescription. Consequently, the analysis aims to understand 1) the risk of relapse and 2) the progressiveness of the disease. Disease progressiveness is measured by the cumulative recurrence of relapse. After briefly defining the objectives of the study, the following subsection, 3.3.2, defines the problem characterisation more precisely.

3.3.2 Problem Characterisation

Firstly, each study subject is randomly allocated to a treatment subsample according to database treatment identifications (see: Table 8). Each treatment subsample represents a treatment arm. A patient's observation period begins after treatment prescription. Patients can start at any state, but they transition from one severity state into another in a predetermined order (Figure 8). A transition into a healthier state indicates effective OSA treatment. Consequently, a transition into a state of increased severity indicates ineffective treatment or *relapse*. Figure 8 illustrates the severity states, the transition order, and all three transition alternatives present in this study:

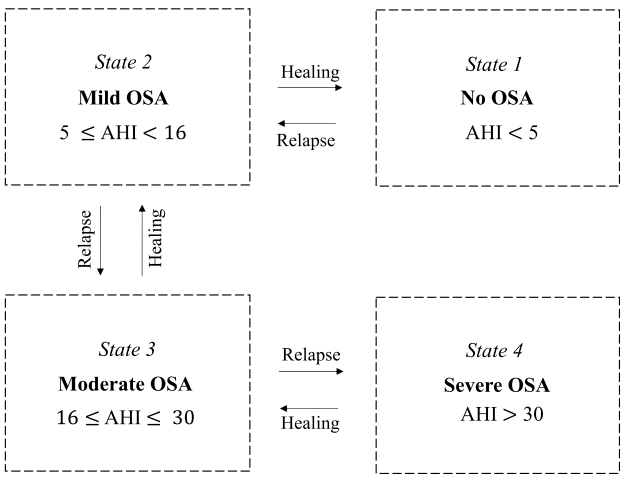


Figure 8. OSA Progression Diagram.

A transition from ‘no OSA’ to ‘mild OSA’ is considered a relapse instead of falling ill because a prerequisite for entering the study sample is an OSA diagnosis (AHI>5). Therefore, all study subjects in the state of ‘no OSA’ have previously been in the state of ‘mild OSA’. Furthermore, transitioning to a state of increased severity starts to count as *relapse* from the *second time* the subject enters the severity state.

The focus is to study ineffective treatment, measured by 1) the probability of relapse and 2) cumulative relapse recurrence. A distinct patient could undergo multiple transition types during observation, but we only include relapses between consecutive states. Therefore, this study only accounts for three relapse types:

- Relapse from ‘no OSA’ to ‘mild OSA’
- Relapse from ‘mild OSA’ to ‘moderate OSA’
- Relapse from ‘moderate OSA’ to ‘severe OSA’

Furthermore, cumulative relapse recurrence measures the repetitiveness of each relapse type. Thus, cumulative relapse recurrence captures the number of times each relapse type is estimated in each treatment category. Figure 9 illustrates cumulative relapse recurrence.

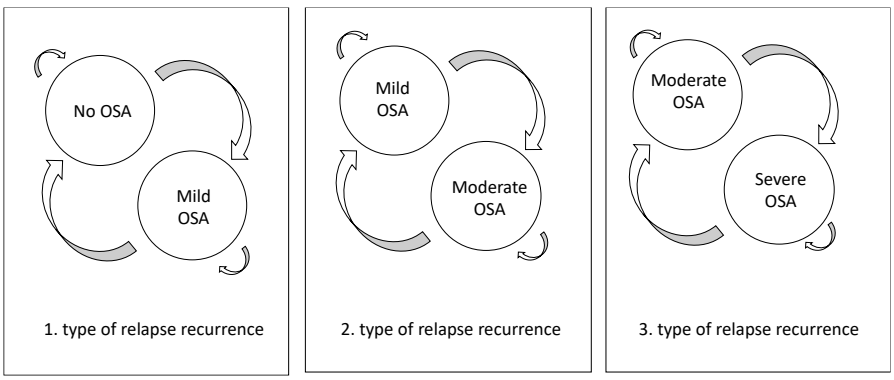


Figure 9. Illustration of Cumulative Relapse Recurrence by Type of Relapse

This subsection explained the problem characterisation of this study. The subsequent subsection, 3.3.3, covers the modelling strategy for relapsing events.

3.3.3 Multistate Analysis and Back Transitions

Multistate models can consider back transitions (relapses) to previously occupied states (Andersen & Keiding, 2002). If a process is assumed to be present in a particular state E_1 , the return of the process to E_1 may be classified as a recurrent event. Figure 10 illustrates

a simplified Markov Chain transition diagram with back transitions for an OSA multistate model.

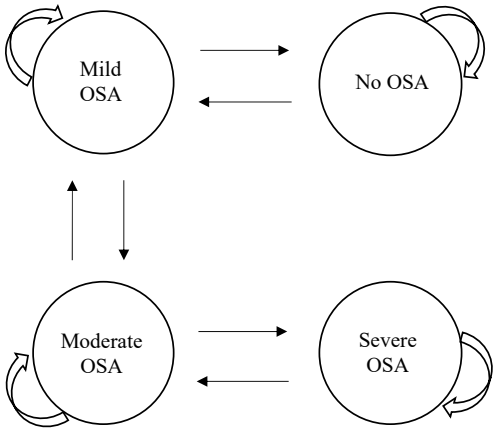


Figure 10. Markov Chain Diagram to Model the Relapsing Events in OSA

The graphical representation of a multistate model consists of nodes and arrows. The nodes denote the states, and the arrows denote the transitions and describe the disease pathway. Every node between two states represents a dichotomous event that describes the transition from one state to another. In a multistate setting, individuals can start from different initial states and move forward through a sequence of states. Transition probabilities or transition intensities describe the transitions from one state to another. (Hougaard, 1999)

3.3.4 Markov Chain Models

Multistate models often rely on the *Markov property*, which refers to the *memoryless property* of a stochastic process (e.g., Aalen et al., 2008, p.463). The memoryless property indicates that the next state of a Markov model depends on the most recent point in the trajectory, and not on the history of the process (Aalen et al., 2008, p.463). A random variable τ has the memoryless property if for all t, $h \geq 0$ it satisfies the following condition:

$$P(\tau > t + h | \tau > t) = P(\tau > h) \quad (5)$$

Where the probability of surviving for time h given, we have survived to time t is the same as the (unconditional) probability of surviving for time h.

Two major types of multistate models are present in literature: *discrete-time Markov chain processes* and *continuous-time Markov chain processes* (Wan et al., 2016; Geweke et al., 1986; Aziz et al., 1996). Both techniques are related and can be applied to model state transitions (Aalen et al., 2008, pp.463-465). However, the assumptions behind these techniques differ and might generate different results (Wan et al., 2016). Discrete-time multistate models assume evenly spaced data (Wan et al., 2016). Nevertheless, EHR observations are often inconsistent since longitudinal data are subject to irregular observations (Aalen et al., 2008, p.17). Therefore, this study employs a continuous-time multistate model for unevenly spaced measurements. Furthermore, Markov models can be either *time-homogeneous* or *time-inhomogeneous* (Aalen et al., 2008, p.465). Time-homogeneity indicates that the transition probabilities are independent of time (Aalen et al., 2008, p.463). In non-homogeneous chains, transition probabilities often vary across time (Aalen & Johansen, 1978).

In this study, we propose a time-inhomogeneous continuous-time Markov model to account for the changes in transition probabilities as subjects age. Therefore, we denote an inhomogeneous continuous-time Markov process X has transition probabilities that depend on the start and end time of the transition period in the following manner:

$$P_{xy}(s,t) = P(X_t = y|\, X_s = x) = P(\, X_t = y \mid X_s = x, Past\,), s \leq t. \qquad (6)$$

The underlying Markov process is time-inhomogeneous, as the transition probabilities depend on the actual time interval $[s,t]$. However, the Markov chain still has the Markov property that conditions on the past state y, given the present is state x. (Beyersmann et al., 2011, p.30)

3.3.5 Obstructive Sleep Apnoea Model

Formally, a Markov chain is specified by the following components:

(1) a set of N states,

(2) a transition probability matrix

(3) initial probability distribution over states.

In this study, with K states of disease severity, the underlying process is defined as $X(t)\ \varepsilon\ \{1,2,\ldots,\mathrm{K}\}, t \geq 0$. Where $X(t)$ denotes the occupied severity state at time *t.* If a patient is observed at times $T = (t_0, t_1, \ldots, t_m)$, then we can define $X = (X_0, X_1, \ldots, X_m)$ and the respective occupied severity states are denoted as follows: $Xl = X(tl), l = 1,2,\ldots,m.$ Consider the following transition diagram for a four-state Markov process:

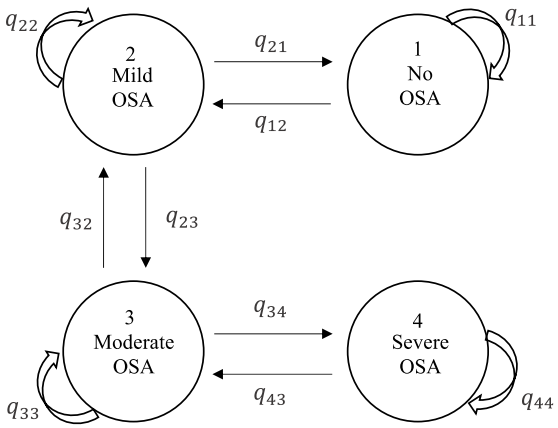


Figure 11. Transition Diagram of the Relapsing-remitting Process for OSA

The Markov chain behaves according to the values of a transition matrix. These values specify the probabilities of the transitions between different states (Aalen & Johansen, 1978). The following matrix may then represent the Markov chain from Figure 11:

$$Q = \begin{bmatrix} q_{11} & q_{12} & 0 & 0 \\ q_{21} & q_{22} & q_{23} & 0 \\ 0 & q_{32} & q_{33} & q_{34} \\ 0 & 0 & q_{43} & q_{44} \end{bmatrix} \tag{7}$$

Where $q_{11}, q_{12}\ q_{21}, q_{22}, q_{23}, q_{32}, q_{33}, q_{34}, q_{43}, q_{44}$ are all transition rates in a *Q-matrix*

The transition intensity matrix describes the Markov chain and the directions of the transitions between health states. The elements of one row represent the probabilities of transitioning from a single state to any of the other states or staying in the same state. The following subsection, 3.3.6, explains the Aalen-Johansen estimator applied for modelling the transition probabilities.

3.3.6 Aalen-Johansen Estimator to Compute the Risk of Relapse

This study employs the Aalen-Johansen estimator to model the empirical transition matrix and compute the transition probabilities (Aalen & Johansen, 1978). The Aalen-Johansen estimator is suitable for *continuous-time inhomogeneous* Markov multistate models for

left-truncated and *right-censored data* (Aalen & Johansen, 1978). The model is derived from a Nelson-Aalen estimator of integrated transition hazards (Aalen & Johansen, 1978). The Aalen-Johansen estimator fits on the residuals of each transition time and builds on the assumption that the data are Markovian (Meira-Machado, 2006). Nevertheless, the Markov property might not be fulfilled for real data, and thus it needs to be demonstrated with statistical testing (Jackson, 2011). The Aalen-Johansen estimator employs a plug-in estimate to estimate transition probabilities. Therefore, the transition matrix $P(s,t)$ is estimated with the Aalen-Johansen estimator $\hat{P}(s,t)$:

$$\hat{P}(s,t) = \prod_{s<u\leq t} (\mathrm{I} + \Delta\hat{A}(\mathrm{u})) \tag{8}$$

Where:

I is the identity matrix,

$\hat{A}$ is the matrix from the Nelson-Aalen estimator

The Aalen-Johansen estimator, also known as the empirical transition matrix, is a finite matrix product over all event times u in (s, t] and matrices $I + \Delta\hat{A}(\mathrm{u})$ (Beyersmann et al., 2011, p.33). In clinical studies, individuals might leave the study before the study terminates. Therefore, we cannot observe the actual time a study subject leaves, but a 'censoring time' instead (Kleinbaum & Klein, 2008, p.5). In this study, patients are assumed to be right-censored and followed until their last relapse. Nevertheless, for the analysis to be valid, censoring times must be independent of the times at which an event occurs (Aalen et al., 2008, p.58). Independent censoring essentially means that within any sample of interest, censored patients should represent all the individuals that remained at risk in the sample (Aalen et al., 2008, p.58). Furthermore, both the Aalen-Johansen estimator and the Nelson-Aalen estimator assume independent censoring (Aalen et al., 2008, p. 60). this study tests independent censoring via visual assessment. As noted in subsection 2.3.4, left-truncation, also known as late entrance, happens when individuals enter the study after time origin *t=0* (e.g., Borgan, 2014). The following subsection, 3.3.7, covers the theory behind the Nelson-Aalen Estimator for survival functions.

3.3.7 Nelson-Aalen Estimator to Estimate Cumulative Relapse Recurrence

The Nelson-Aalen estimator is a non-parametric survival regression model employed to analyse recurrent events (Aalen et al., 2008, p.71). The Nelson-Aalen estimator can consider any Markov process with a finite number of states to model a subject's life history (Borgan, 1997). The method measures the risk of failure over time (Klein, 1991). The risk of failure is the probability of a subject experiencing an event of interest, such as relapse (Klein, 1991). Furthermore, recurring events of the same type might be considered for each study subject (Borgan, 1997). Thus, the estimator estimates the cumulative number of expected events within a certain period (Borgan, 1997). Non-parametric methods are distribution-free, meaning they do not have assumptions about underlying distributions of the survival times (Györfi et al., 2002). The Nelson-Aalen estimator $\hat{A}_{NA}(t)$ applies a hazard function to interpret the cumulative risk of failure (Borgan, 1997). The estimator is formulated as follows:

$$\hat{A}_{NA}(t) = \sum_{T_{(j)} \leq t} \frac{d_j}{\mathrm{n}_j}, \tag{9}$$

Where

d_j is the number of observed relapses at time *j*

n_j is the number of individuals at risk just prior to time *j*

The Nelson-Aalen estimator is a right-continuous step function with increments d_j/n_j at failure times (Aalen et al., 2008, p.72). The numerator represents the increments in a transition-specific counting process (Beyersmann et al., 2011, p.22). The denominator shows all patients who entered the specific state *x* before time *j* and are still present in state *x* again or censored (Beyersmann et al., 2011, p.22). The variance of the Nelson-Aalen estimator is formulated as follows:

$$\widehat{\sigma^2}(t) = \sum_{T_{(j)} \leq t} \frac{(n_j - d_j) d_j}{(n_j - 1) {n_j}^2}. \tag{10}$$

The numerator represents the number of observed transitions at time *j,* and the denominator includes the number of patients at risk in state *x* prior to time *j* (Aalen et al., 2008, p.72). A confidence interval (CI) yields the probability that an interval produced by

the model includes the actual value of the estimator (e.g., Beyersmann et al., 2011, p.34). Variance can be applied to construct an approximately 95% CI in the following manner:

$$\hat{A}_{NA}(t) \pm \hat{\sigma}(t) \cdot 1.96. \quad (11)$$

Where 1.96 represents the 0.975 quantiles of the standard normal distribution.

3.3.8 Hypothesis Testing: Differences Between Treatment Arms

Hypothesis testing is a statistical inference method that tests whether the results of an experiment are statistically significant (e.g., Braumoeller, 2004). Hypothesis testing consists of two statistical hypotheses: *(1)* the null hypothesis and *(2)* the alternative hypothesis. The null hypothesis, H_0 is the one being tested, while the alternative hypothesis, H_A is the suspected outcome. Hypothesis testing is subject to two types of errors. We can either reject H_0 when H_0 is true, leading to a false positive (type-I error) or reject H_1 when H_1 is true and generate a false negative (type-II error) (Braumoeller, 2004).

The log-rank test serves time-to-event studies to test the null hypothesis of no difference between study subgroups (e.g., Yang & Prentice, 2010). More precisely, the log-rank test assesses whether the time until relapse differs significantly from one treatment arm to another (Yang & Prentice, 2010). Consequently, it compares the distribution curves of the sample subgroups, as similarity in distribution curves implies similarity in event rates (Lakatos, 1988). Furthermore, the log-rank test can handle well right-censored data, which makes it a suitable approach for this study (Zhao, 2004). This study examines the differences between *four* treatment subsamples and applies the log-rank test to compare the treatment arms with one another. For analysing the distribution curves of four subsamples, it is necessary to apply a generalisation of the log-rank test to account for more than two groups (Bland & Altman, 2004). Therefore, we propose the following hypotheses for the log-rank test:

H_0: There are no statistical differences between the four treatment arms.
H_A: There are statistical differences between the four treatment arms.

If the p-value of the test is less than the predetermined significance level of 5%, then we can reject the null hypothesis. Consequently, there is sufficient evidence to conclude that

the groups differ in time-to-event (until relapse). The log-rank statistic is approximately distributed as a chi-square statistic and formulated as follows:

$$\sum \frac{(\sum O_{jt} - \sum E_{jt})^2}{\sum E_{jt}}. \quad (12)$$

Where

$\sum O_{jt}$ is the sum of the observed number of events (O) in the j[th] treatment arm over time, t (e.g., j = 1,2)

$\sum E_{jt}$ represent the sum of the expected number of events (E) in the j[th] treatment arm over time, *t.*

Furthermore, the log-rank test assumes similar assumptions as the Kaplan Meier survival curve (Bland & Altman, 2004). These assumptions include that the censoring patterns must be similar for all treatment arms, *and* the proportional hazards (PH) assumption must be satisfied (Bland & Altman, 2004). However, this study does not assess the PH assumption, as the log-rank test is still statistically valid under non-PH (Lin et al., 2020). Although the log-rank test might suffer from substantial power loss due to the unmet assumption (Lin et al., 2020).

4 Data

This section describes the data in this study, along with preparatory measures. Section 4.1 entails an overview of the database, and section 4.2 explains the data architecture. Sections 4.3 and 4.4 present essential model variables and the treatment identification criteria. Section 4.5 describes the sample selection procedure, and section 4.6 outlines the necessary pre-processing steps. Finally, section 4.7 presents the exploratory analysis.

4.1 Description of Data

The health data acquired for this study was a fully anonymised data set obtained from Auria Biobank (2019). Auria Biobank operates together with Turku University Hospital to collect biological samples from patient visits for research initiatives. The data were composed of patient information collected during individual medical care visits. Overall, the database contained the pseudonymised EHRs of approximately 24 700 unique patients tracked between 2003 and 2019. The patient cohort consisted of a heterogenous population, and all the patients had been diagnosed with sleep apnoea (ICD code G47.3, including both OSA and CSA patients). Table 5 describes the PostgreSQL (www.postgresql.org) database tables utilised for this thesis:

Table 5: PostgreSQL Tables Utilised for this Thesis

Table name	Table description	Distinct patients	Distinct records
Customers	Records with descriptive patient information such as gender, birthdate, and death date.	24 669	24 749
Features	Records with timestamped biomarker values such as blood pressure, AHI, computed BMI and SaO2%, among others.	21 786	1 651 360
Visits	Records about visits to specialists, including time stamps, diagnosis codes, visit specifications and descriptions.	24 375	2 285 851
Procedures	Records from patient procedures such as surgeries, X-ray screenings and other physical examinations.	24 475	957 248
Resmed	Records collected from PAP machines. The table contains session times, device types, dates, and clinical metrics such as mask leakage and AHI values.	7 649	617 844

4.2 Data Architecture and Tools

The raw patient data were stored in tables inside a PostgreSQL database. The data were accessed and explored with the help of SQL (Structured Query Language) queries. Data analysis tasks, such as severity classification and treatment pre-processing, required a connection between the PostgreSQL database and the programming applications Python (www.python.org) and R (www.R-studio.org). Figure 12 illustrates this process:

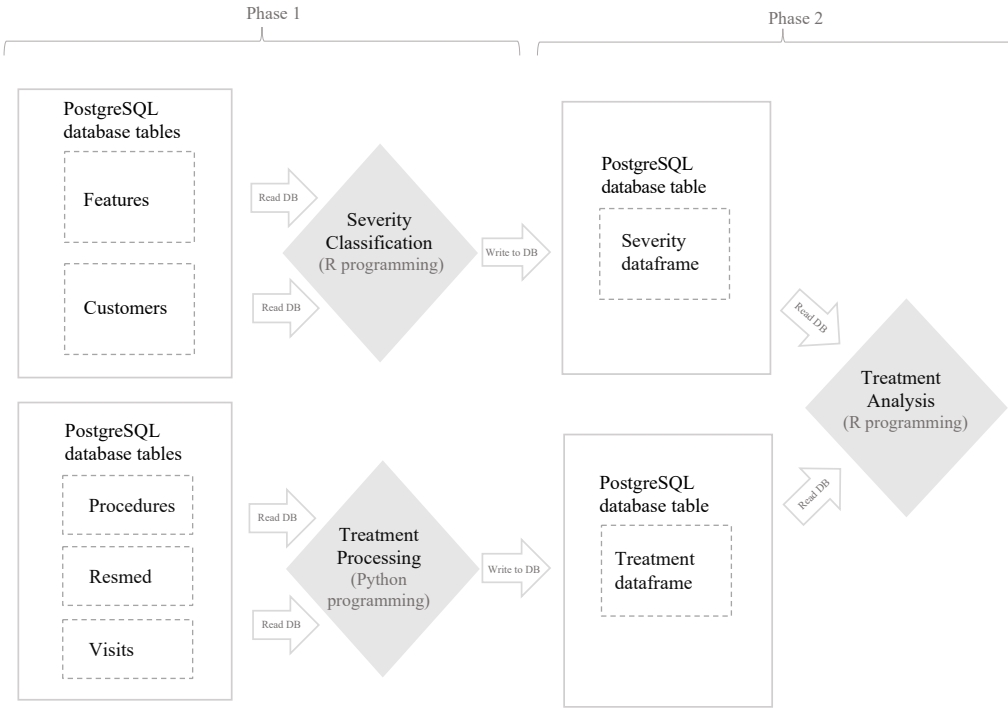


Figure 12. Data Architecture

After acquiring all the necessary information for modelling, the newly processed tables were inserted back into the PostgreSQL database as tables under a new section created for this project. Finally, treatment analysis was completed in the R programming application, so a connection between the new tables and R was established to run the final models.

4.3 Model Variables

This section explains the variables considered in this study. First, the section introduces the dependent variable (AHI) and continues to explain the explanatory biomarker variables. This study defined the AHI-derived severity variable according to the prevalent severity classification thresholds in the *Finnish Current Care Guidelines (2022).* AHI was considered an ordinal variable with four severity levels, and the exact severity thresholds are described in Table 6.

Table 6: AHI Thresholds for the Severity Variable

Severity level	Dependent variable AHI
No OSA	$AHI < 5$
Mild OSA	$5 \leq AHI < 16$
Moderate OSA	$16 \leq AHI \leq 30$
Severe OSA	$AHI > 30$

Previous studies have applied different derivations to create an OSA severity variable, both including and excluding SaO2% values (Kainulainen et al., 2019; Myllymaa et al., 2015). Nonetheless, the *severity classification* phase of this study utilised only AHI as a response variable and biomarker values such as SaO2% and BMI as explanatory variables. Table 7 presents all the biomarker variables and their descriptions, data types and inclusion ranges.

Table 7: Variable Descriptions and Inclusion Ranges

Variable	Variable description	Type	Inclusion range
AHI	The Apnoea Hypopnoea Index (AHI) refers to the number of apnoeas or hypopnoeas per hour of sleep (Duodecim, 2022). The hourly AHI values are averaged per night.	Numerical	0–120
ODI	The Oxygen Desaturation Index (ODI) measures the number of desaturation events per hour. A desaturation event is described as a drop of either 4% (ODI4) or 10% (ODI10) below baseline levels. (Iber et al., 2007)	Numerical	0–130
Age	Age at diagnosis.	Numerical	18–101
BMI	The Body Mass Index (BMI) measures the weight of an individual in kilograms divided by the square of height in meters.	Numerical	10–80
Diastolic BP	Diastolic Blood Pressure is the pressure in the arteries when the heart rests between beats.	Numerical	30–190
Systolic BP	Systolic Blood Pressure measures the pressure the blood exerts against the artery walls as the heart beats.	Numerical	20–250
Glu	The Blood Glucose Level (Glu) measures the glucose concentration in the blood.	Numerical	0–25
SaO2	Arterial Oxygen saturation (SaO2) refers to the percentage of oxygen in the blood.	Numerical	60–100
Min SaO2	The minimum SaO2 value per night.	Numerical	60–100
Deps	Patient values for Depressiveness questionnaire.	Numerical	0–30
ESS	Patient values for ESS questionnaire.	Numerical	0–25

Inclusion ranges were discussed with healthcare professionals and adjusted to contain realistic values. Additionally, the Isolation Forest algorithm was applied as a more advanced outlier detection method.

4.4 Treatment Identifications

Treatment identifications were conducted manually from the database tables. All patients in the data were diagnosed with OSA; therefore, any treatment suitable for OSA was considered as such. Table 8 conveys an overview of the treatment variables in this study.

Table 8: Treatment Variables and Identification Criteria

Treatment	Identification criteria	Identified patients
CPAP	Patient records with information on CPAP devices. A total of 10 different device types were identified.	13 882, (11 874*)
Upper airway surgery	Patient records with visit type specifications such as "uvulopalatopharyngoplasty", "tonsillectomy", "adenoidectomy", and other upper airway surgery specifications.	1 921, (1 674*)
OA	At least one visit to an oral specialist with the following visit type specification: "oral appliance", "mouthguard", or "dental appliance."	1 002, (1 002*)
Bariatric Surgery	Patient records with specialist visits to areas of bariatric surgery and a recorded BMI ≥ 35 (The BMI threshold is aligned with the *Finnish Current Care Guidelines, 2022*). The three identified bariatric surgery types were 1) laparoscopic adjustable gastric banding, 2) sleeve gastrectomy, and 3) Roux-en-Y gastric bypass procedure.	704, (629*)

*Count of patients assigned to each final treatment arm after handling patients with multilevel treatment plans.

In addition to the four official OSA treatment methods, other treatment methods were present in the data. Nevertheless, this study only considers the four most official OSA treatment methods. Furthermore, some unique patients were assigned to two or more treatment methods during the follow-up. Unique patients treated with multiple treatments posed a challenge in the research design. For example, if we evaluate the efficiency of CPAP, but a portion of the patients treated with CPAP has been treated with upper airway surgery afterwards, connecting a patient's relapse solely to upper airway surgery or CPAP is ambiguous. There are two established approaches for reducing the bias that treatment

reallocation causes in clinical trials (Gupta et al., 2011). (1) An intention-to-treat approach consists of analysing patients according to the treatment prescription that they have been assigned at the beginning of the treatment plan. Although an intention-to-treat approach is often considered conservative, the approach was unfeasible for this study due to a lack of data on treatment intention. (2) A per-protocol approach considers only the received treatment instead of the randomised treatment. This study followed the per-protocol approach, considered only the first treatment received and excluded the patients at the point of switching treatments. Figure 13 explains the per-protocol approach relevant to this thesis.

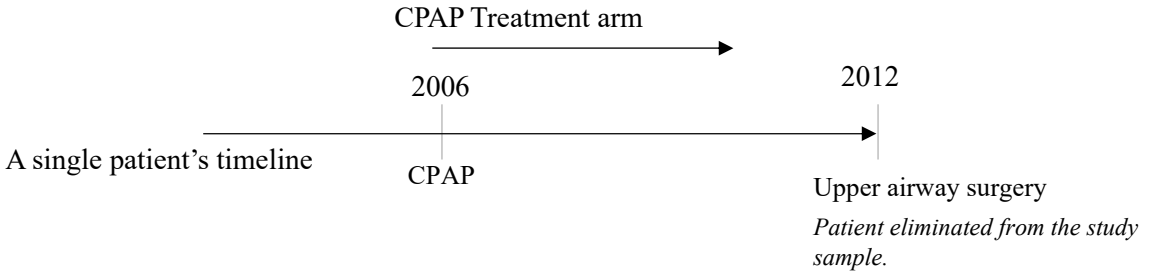


Figure 13. Example of the Per-protocol Approach in a Patient's Timeline

The per-protocol approach was applied to all patients treated with multilevel treatment plans. Thus, only the first treatment method was included per patient. The following section, 4.5, explains the reasoning behind the sample selection process.

4.5 The Sample Selection Process

This section describes the sample selection process of this study. Firstly, patients under 18 years old were excluded from the sample population. Age was calculated as age at diagnosis. Furthermore, patients with signs of central sleep apnoea (CSA) were removed from the study sample. CSA patients were identified based on specialist recordings and the treatment machine type. The CSA treatment machine type is the adaptive servo-ventilation (ASV), a non-invasive ventilatory treatment option (Resmed.com, 2022). Additionally, patients with no treatment data were automatically excluded from the study scope. Furthermore, only patients with two state transitions or more were included. Hence, some patients with only one available record were automatically considered outside of the study scope. Table 9 presents the sample selection process step by step:

Table 9: The Sample Selection Process

Exclusion criteria	Eliminated	Patient sample after selection
(All patients at the beginning of the study)	-	(24 669)
Remove individuals aged under 18	2 979	21 690
Remove CSA suspects	382	21 308
Remove patients with not enough observations	5 487	15 821
Remove individuals with no treatment data	642	15 179
Total Sample		**15 179**

After removing all the patients considered beyond the scope of this study, the final sample was composed of 15 179 study subjects. The following section, 4.6, covers all data pre-processing steps relevant to this study.

4.6 Data Pre-processing

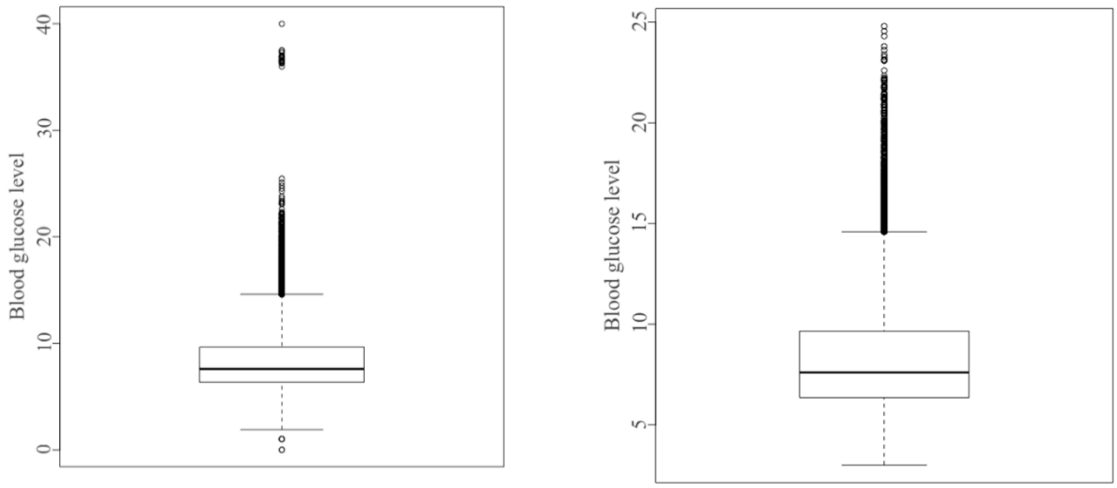
This part explains the data pre-processing steps relevant to this thesis. Subsection 4.6.1 covers time-series irregularities, and subsection 4.6.2 concerns the outlier detection methods. Subsection 4.6.3 explains the strategy for missing value imputation and subsection 4.6.4 presents the multicollinearity tests.

4.6.1 Irregular Time-series Data

Medical care visits are often irregular, which generates irregularity in the collected biomarkers. For patients with more healthcare visits, variables were followed more frequently than for patients not appointed to visits as repeatedly. Furthermore, a high amount of healthcare visits per patient could implicate a higher degree of severity, which could cause data to be biased towards the more severe cases of OSA. *To limit the amount of physiological data, such as blood pressure measurements, each patient was left with only one recorded observation for each biomarker variable per month*. The first approach included only the last monthly observation to reflect the most recent monthly values. However, further data analysis showed high variability between values collected at the beginning of the month compared to those collected at the end of the month. To stabilise this variation, monthly averages were utilised for biomarker observations.

4.6.2 Outlier Detection

Outlier detection is essential for flagging implausible EHRs that can influence the reliability of the study results. Hence, one of the goals of data pre-processing was to modify data accordingly. This study handled anomaly detection with the Isolation Forest algorithm together with predefined inclusion ranges from medical practitioners (presented in Table 7). Figure 14 exemplifies the outlier detection procedure.



a) Variable *glu* before outlier detection b) Variable *glu a*fter outlier detection

Figure 14. An Example of the Outlier Detection Procedure. (A boxplot with lower 25th and upper 75th percentiles as box boundaries, the line inside the box is the median, lower and upper error lines are 10th and 90th percentiles, and filled circles indicate data outside the 10th and 90th percentiles.)

4.6.3 Data Missingness

Missing data refers to any observation that would be intentional to have but is not recorded for any reason. The observations in the dataset were nested within participants, so it was necessary to utilise a missing data imputation method suitable for such a multilevel structure (Twisk, 2013, p.215). After investigating the type of missingness, we considered the data to be consistent with MAR (see definition in Table 4). Furthermore, under the conditions of MAR and MCAR, the *multiple imputation* technique should result in unbiased estimates (Sainani, 2015). The multiple imputation technique calculates several imputation values for each missing value, and the final set of imputed values is then applied to form a completed dataset (Twisk, 2013, p.223). We applied the 'MICE' (Multiple Imputation by Chained Equations) package that had several options to deal with correlated longitudinal data. In this study, the missing values were imputed with a two-

level predictive mean matching (PMM) technique that handles well non-normally distributed variables (White et al., 2011). Furthermore, the MICE model accounted for the multilevel nature of the data. The *patient ID* variable represented the class variable for patient clusters. The final imputation method selection was based on model performance: exploring other alternative imputation methods included the implementation of algorithms such as missForest, K-nearest neighbour and the *last observation carried forward.* The MICE imputation method yielded the highest accuracy scores in preliminary logistic regression trials. Thus, we selected the MICE imputation method with the PMM technique as our missing data imputation method.

4.6.4 Multicollinearity

Finally, it was essential to test for multicollinearity to investigate possible data quality issues regarding correlated independent variables. Therefore, multicollinearity was tested by analysing a correlation matrix (Table 10) and the Variance Inflation Factors (VIF). Both techniques were implemented with methods suitable for repeated measures data.

Table 10: Correlations Between Variables

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1) AHI	1											
(2) BP diast	-0.014	1										
(3) BP syst	0.0066	-0.45	1									
(4) DEPS	0.0057	0.0017	0.0090	1								
(5) ESS	0.0017	-0.011	0.017	-0.080	1							
(6) ODI10	0.00073	-0.029	0.0145	-0.017	0.028	1						
(7) ODI4	0.070	0.0018	0.030	-0.054	0.018	0.070	1					
(8) SaO2%	-0.10	-0.032	0.027	0.17	0.023	0.020	0.094	1				
(9)$SaO2\%_{min}$	-0.12	-0.095	-0.010	-0.016	-0.10	0.024	-0.22	-0.14	1			
(10) BMI	0.012	-0.027	-0.034	-0.017	0.27	0.060	0.094	0.026	0.0064	1		
(11) GLU	0.0026	0.028	-0.028	0.017	0.013	0.10	0.0132	0.0021	0.0070	0.010	1	
(12) Age	0.022	0.11	-0.13	0.0072	0.056	-0.12	-0.010	0.042	0.018	0.16	0.011	1

There were no strong correlations between variables. Furthermore, all VIF values were less than 2, indicating that the VIF values in this study do not indicate multicollinearity problems. Generally, a VIF value greater than 5 indicates multicollinearity (Gujarati & Porter, 2009, p. 340; Alin, 2010). Based on these results, the data in this study does not

suffer from multicollinearity problems. The following section, 4.7, presents a descriptive analysis of the data.

4.7 Descriptive Analysis

This section entails the exploratory analysis. First, subsection 4.7.1 presents an overview of the study sample. Then, subsection 4.7.2 explores differences between the severity levels, and subsection 4.7.3 covers the treatment methods in this study. Finally, subsection 4.7.4 analyses the development of relapses.

4.7.1 Overview of the Study Sample

The final sample included observations from 15 179 diagnosed OSA patients. These patients were treated and observed between the years 2003 - 2019. The sample of this study had a twofold OSA prevalence in males (66.35%) compared to females (33.65%). This ratio is consistent with the information on the *Finnish Current Care Guidelines* (2021), which states that OSA is twice as common in males compared to females. Furthermore, the mean age at diagnosis of the study sample was approximately 57. The mean age coincides with previous research, which has demonstrated an OSA occurrence peak in individuals between 55-59 years of age (Huang et al., 2008; May et al., 2018; Costa et al., 2019). Figure 15 depicts the distribution of age between male and female OSA patients.

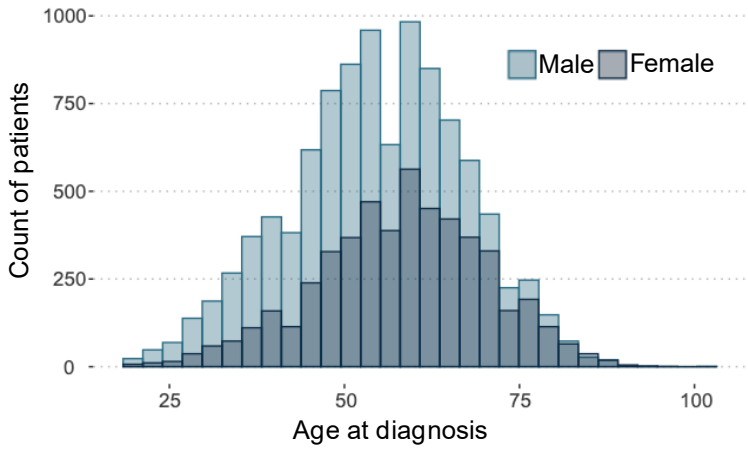


Figure 15. Distribution of Age at Diagnosis by Gender

The distribution of male patients is skewed towards the left, while the distribution of female patients is approximately normal. Left-skewed distribution indicates that a higher volume of male patients was diagnosed younger, while generations over 70 years did not

show significant gender differences in the age of diagnosis. According to the *National Institute of Health* (2022), the menopausal transition starts in women aged 45 to 55. Menopause has been studied as an independent risk factor for developing OSA (Young et al., 2003; Lee et al., 2019). Therefore, menopause could explain the rapid growth of diagnosed female patients aged between 45 and 55.

The mean BMI value in this study was 33.52 kg/m^2, categorised as overweight in the BMI classification scale. Conversely, according to the publication from Abarca-Gómez et al. (2017), the age-standardised BMI among the Finnish population was approximately 26.5 kg/m^2 in 2017. Therefore, the sample of this study coincides with previous evidence that suggests a strong association between obesity and OSA (Newman et al., 2005; Berger et al., 2009; Schwartz et al., 2008). The following subsection, 4.7.2, explores the study sample from the perspective of disease severity.

4.7.2 The Characteristics of The Severity Groups

The severity groups were derived from the AHI thresholds stated in the *Finnish Current Care Guidelines* (2022). Table 11 describes the characteristics of each severity level.

Table 11: Descriptive Statistics of the Sample Population by Severity Levels

	SEVERITY STATES							
	No OSA		Mild OSA		Moderate OSA		Severe OSA	
Variables	Mean	Sd	Mean	Sd	Mean	Sd	Mean	Sd
AHI	3.12	1.25	9.70	2.99	20.9	3.83	35.2	4.35
Glu	8.25	2.76	8.21	2.74	8.24	2.72	8.10	2.69
ESS	6.16	4.35	6.23	4.36	6.50	4.53	7.05	4.64
DEPS	6.39	6.25	6.45	6.29	6.52	6.28	6.67	6.34
ODI10	5.35	11.9	5.60	12.4	6.40	12.99	11.5	18.3
ODI4	11.6	13.6	12.3	14.2	19.7	19.9	42.5	30.2
Age	57.5	12.6	57.8	12.5	57.9	12.8	53.6	13.3
BMI	33.2	7.4	33.1	7.29	33.2	7.34	34.1	8.15
SaO2%	94.5	3.3	94.4	3.29	93.8	3.56	92.3	4.33
SaO2% min	80.4	7.4	79.9	8.59	78.9	9.28	74.4	10.82
Systolic BP	139	20.1	139	20.1	140	20.1	143	20.1
Diastolic BP	78.8	12.6	78.9	12.6	79.9	12.6	81.9	13.5

Data from 84 311 observations from 15 179 study subjects

There were apparent differences between biomarker values at distinct severity states. Studies have often identified direct associations between AHI-calculated OSA severity and SaO2% measurements (Fernandes, 2021; Almazaydeh, 2012). As expected, severe OSA

records show lower SaO2% (and SaO2% min) values than milder patient records. For example, the non-OSA group (mean SaO2: 94.49 %) had similar values to healthy adults, whose SaO2 values are approximately 95% (Hafen et al., 2022). In addition to AHI and SaO2%, ESS is often considered when diagnosing patient severity. The data demonstrated that ESS values increased in proportion to severity categories. Thus, low ESS scores were more common in the 'no OSA' and the 'mild OSA' states than in the higher severity levels. The following subsection, 4.7.3, explains the results from treatment identifications.

4.7.3 Treatment Types

The final sample included 15 179 study subjects that had been prescribed to a total of 17 509 treatment methods. Individuals treated with two or more treatment methods were assigned to a treatment arm according to the *per-protocol* approach. The per-protocol approach only considers the first assigned treatment per patient. Thus, patients with multiple treatments were excluded from the study after switching treatments. Figure 16 summarises the final treatment division relevant to modelling.

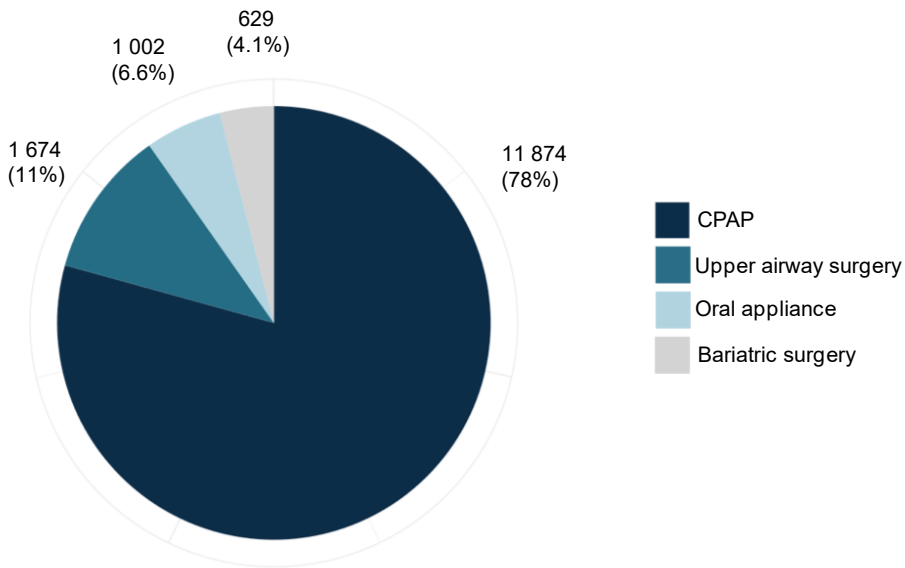


Figure 16. Treatment Distribution Among Patient Sample

CPAP was the most widely prescribed treatment method in the study sample (11 874 prescriptions, 80% of the sample). Evidence-based medical research often supports the efficaciousness of CPAP as the first treatment method (Rotenberg et al., 2016). As CPAP is frequently the first treatment a patient receives, it is often present in different treatment combinations (Bachour & Bäck, 2015). Analysis showed that the prescription of CPAP had increased the most over time, while other treatment methods had only moderate

changes. Furthermore, treatment pre-processing showed that almost 2 000 individual patients had undergone some nasal intervention, such as the elimination of nasal polyps. Approximately 75% of the patients with a nasal intervention had also been prescribed to another treatment, such as CPAP or upper airway surgery. This suggests that various patients in the study sample were assigned to a multilevel treatment plan. Nevertheless, this study does not consider the effects that multilevel treatment could have on relapsing patients.

4.7.4 Patient Relapses

After transforming the data into a time-to-event format, the data set was composed of one row per patient transition. There was a total of 84 311 transitions that can be divided into three transition types: (1) no transition (40 421 observations), (2) transition to a healthier state (22 455 observations) and (3) transition to a less healthy state (21 435 observations). Thus, there was a total of 21 435 relapses in the data sample. Although not all 21 435 observations were accounted for modelling as the model only considers consecutive relapses. This study focuses on treatment effects; therefore, it was of particular interest to understand more details on the relapsing process regarding each treatment arm. Figure 17 depicts the yearly relapses by treatment category.

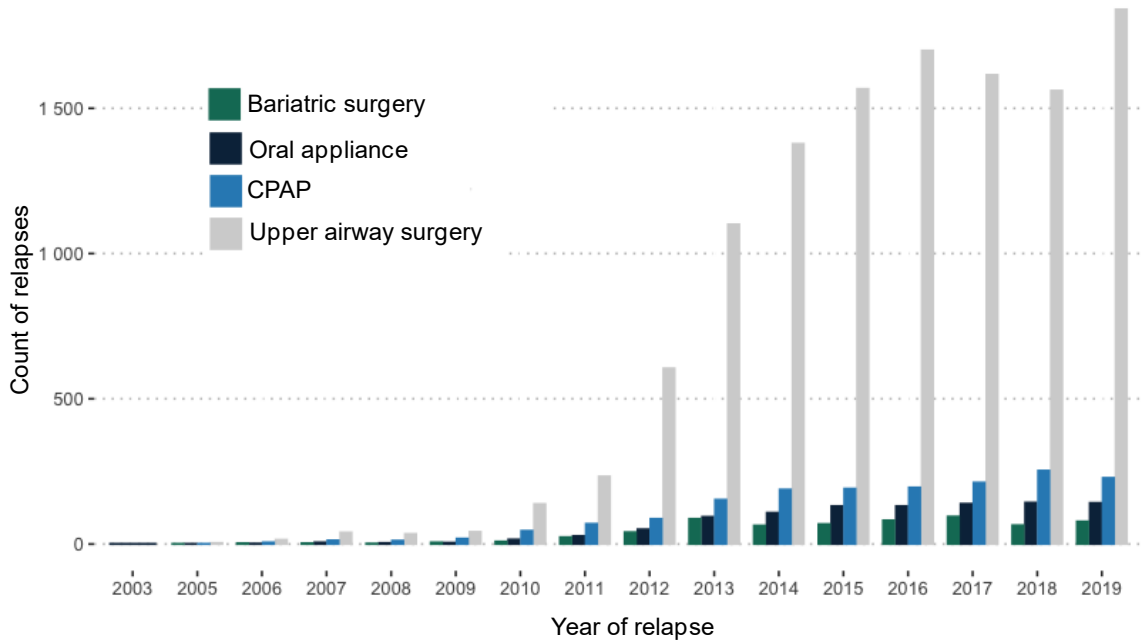


Figure 17. Total Yearly Relapses in Each Treatment Arm

Figure 17 demonstrates a pattern of growth and decay in the relapses of patients treated with CPAP. Other treatment arms experienced moderate yearly changes. Further analysis showed that the prescriptions of CPAP had increased simultaneously. Thus, there could be a connection between treatment effectiveness in patients treated with CPAP and a decreasing rate of relapses. This exploratory section investigated the data set from different perspectives relevant to this study. The subsequent chapter 5 presents the empirical findings and results of the study.

5 Empirical Findings and Results

This chapter presents the results of this study. Section 5.1 covers the results of the severity classification analysis, and section 5.2 explains the results of the treatment analysis.

5.1 Severity Classification

This section presents the results of severity classification. First, the effects of covariates on OSA severity were examined with a CLMM ordinal logistic regression procedure (equation 2). The random effects structure of the model accommodates the repeated measures format of the data. Table 12 presents the results for the CLMM model.

Table 12: Results from CLMM on Associations between Patient Biomarkers and OSA severity

	PARAMETER ESTIMATES			
Covariates	Exp.	Coefficient	95% CI	p-value
BP diastolic	+	0.038	0.033, 0.096	0.33
BP systolic	+	0.027	0.028, 0.10	0.42
SaO2% min	-	- 0.39	- 0.45, - 0.12	< 0.01 **
SaO2%	-	- 0.16	- 0.18, - 0.025	< 0.001 ***
DEPS	+	0.082	- 0.014, 0.10	0.51
ESS	+	0.092	0.050, 0.11	0.094
ODI10	+	0.12	0.11, 0.14	< 0.01 **
ODI4	+	0.40	0.33, 0.46	< 0.001 ***
Glu	+	0.0011	- 0.07, 0.05	0.71
BMI	+	0.025	- 0.029, 0.030	0.52
Age	+	0.032	0.028, 0.045	0.33

Note: Asterisks represent significance levels: *p < 0.05, **p < 0.01, ***p < 0.001. Variables are defined in Table 7.

The results demonstrated that from eleven (11) variables, only four (4) were statistically significant. SaO2%, SaO2% min, ODI4 and ODI10 were significant predictors of OSA severity. The SaO2% and SaO2% min coefficients were negative, indicating that higher oxygen saturation is associated with lower OSA severity. Moreover, one unit increase in SaO2% was associated with a 0.78 decrease in the expected severity value on the odds scale, given that all other variables are held constant. Conversely, ODI predictors show positive coefficients, which indicates that an increase in ODI was associated with an

increase in OSA severity. Based on the results of this study, one unit increase in ODI4 is associated with a 1.49 increase in severity on the odds scale. To summarise, the results demonstrate that a drop in oxygen saturation is associated with increased OSA severity.

5.1.1 Research Hypothesis 1: The Determinants of OSA Severity

This subsection evaluates the plausibility of the research hypothesis 1 (**H1**) developed for severity classification in section 2.4. This hypothesis aims to validate and support the harmonisation of the severity variable.

H1: There are statistical associations between OSA severity and biomarker variables. Specifically, that OSA severity increases in ODI and decreases in SaO2%.

The associations between OSA severity and biomarker variables were analysed with the LRT (equation 4). The LRT evaluated how well a model explains an outcome compared to a model with fewer predictors. Therefore, we built another ordinal logit model with only two covariates (systolic and diastolic blood pressure, as they were statistically non-significant in the original model). Then, we statistically compared it to the original model from Table 12. The LRT resulted in a significant p-value ($p<0.001$), indicating that the original model from Table 12 explains the outcome variable better than the reduced model. Therefore, we assume a significant negative association between OSA severity and SaO2% (including min SaO2%). Additionally, we assume a significant positive association between OSA severity and ODI (including ODI4 & ODI10). However, the model did not show any other significant associations between variables. Thus, we conclude that there are differences in the OSA severity levels of this study, and covariates related to oxygen saturation explain those differences.

5.1.2 Reliability of the Results: Ordinal Regression

The association between oxygen saturation and OSA severity seem reliable as it is consistent with previous research. This study assessed model residuals by converting the ordinal response variable back to a continuous variable and applying a linear mixed-effects model residual analysis. The residuals were normally distributed, although slightly skewed towards the right. However, the distribution was not radically different from a normal distribution. Therefore, the model might not fit perfectly, but it serves the purpose of validating the differences between OSA severity levels. Furthermore, multicollinearity

analysis did not show significant correlations between variables (subsection 4.6.4), and the random effects were normally distributed. The results of the PO assumption test were non-significant, which concludes that the PO assumption is reasonable; thus, an ordinal regression model is valid for this data.

5.2 Treatment Analysis

This section presents the results of the treatment analysis. First, subsection 5.2.1 explains the results regarding relapse probabilities, and subsection 5.2.2 presents relapse recurrence.

5.2.1 Relapse Probabilities Measure the Risk of Relapse

The probabilities of relapse measured treatment inefficiency and were computed with a continuous-time inhomogeneous Markov chain procedure via an Aalen-Johansen estimator (equation 8). Figure 18 presents the relapse probabilities for each treatment arm.

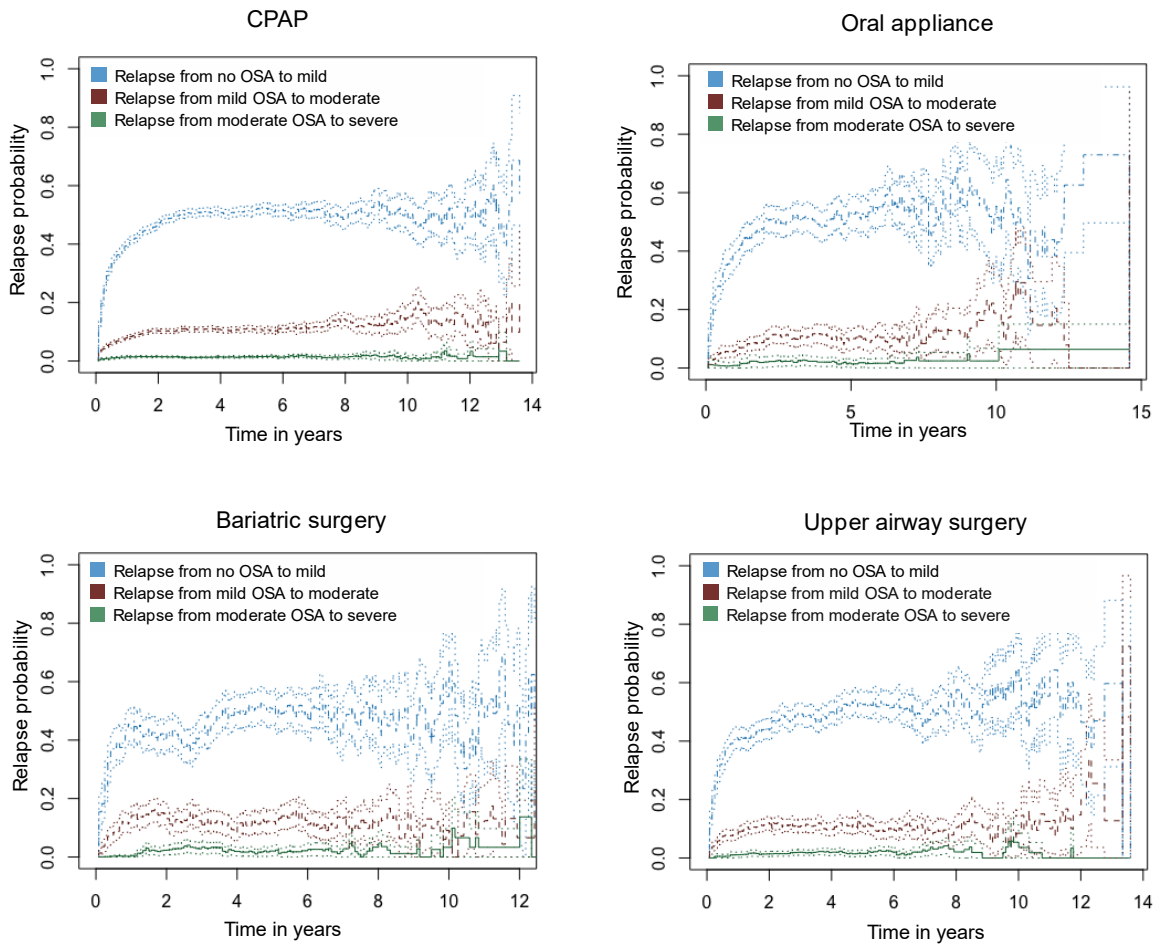


Figure 18. Aalen-Johansen Relapse Probabilities in Each Treatment Arm (95% CI)

As noted in the methodology subsection 3.3.2, the model considered three types of relapses: *(1) relapse from 'no OSA'* to *'mild OSA'*, *(2) relapse from 'mild OSA'* to *'moderate OSA'* and *(3) relapse from 'moderate OSA'* to *'severe OSA'*. The patients were followed between the years 2004 and 2019. The y-axes represent the probabilities of relapse. The x-axes represent the time in years. The following bullet points explain the results of each relapse type.

- **Relapse from 'no OSA' to 'mild OSA'**: This relapse category held the highest probability values among all treatment arms. The group treated with upper airway surgery had the highest *average* probability of relapse (50%). In comparison, the group treated with bariatric surgery had the lowest *average* relapse probability (49%). The probabilities of relapse ranged from 36% to 73%. Moreover, the variance in each treatment arm increased as time passed, and the number of subjects in the risk set decreased. Especially the variances for the patients treated with OA and bariatric surgery were high. Nevertheless, a higher sample size stabilised the variation for CPAP. For patients treated with upper airway surgery and OA, the transition probabilities decreased towards the end of the observation period. A decrease in transition probabilities could indicate that the risk of relapse decreased as time passed. Nonetheless, another explanation for decay could be that the set of patients at risk became smaller as time passed.
- **Relapse from 'mild OSA' to 'moderate OSA'***:* The average relapse probability in this relapse category was 13% for all treatment arms. The probabilities of relapse ranged from 0% to 60%, with increments towards the end of the observation period. The group treated with bariatric surgery had the highest *average* relapse probability (22%), while the group treated with CPAP had the lowest *average* relapse probability (12%) within this relapse category. Patients treated with OA and upper airway surgery had decreasing probabilities of relapse.
- **Relapse from 'moderate OSA' to 'severe OSA'***:* The average relapse probability in this relapse category was 1.7% for all treatment arms. The yearly values ranged between 0% and 6.4%. Patients treated with OA had the highest *average* probability of relapse (3.1%), and those treated with CPAP had the lowest *average* probability of relapse (1.6 %). All groups had a slight increase in this relapse category after year 8.

Table 13 shows the numerical values for relapse probabilities, 95% CI levels and the number of patients at risk in each relapse type.

Table 13: Probabilities of Relapse, 95% CIs and Number of Patients at Risk.

		TREATMENT METHOD											
		CPAP			Upper Airway Surgery			Oral Appliance			Bariatric Surgery		
		Prob	CI95%	at risk	Prob	CI95%	at risk	Prob	CI95%	at risk	Prob	CI95%	at risk
2004	1 → 2	40%	*37%-41%*	2 358	40%	*38%-42%*	385	40%	*37%-45%*	200	44%	*42%-45%*	113
2006	1 → 2	51%	*50%-52%*	1 848	48%	*45%-52%*	280	51%	*46%-55%*	153	39%	*33%-44%*	143
2008	1 → 2	51%	*49%-53%*	1 865	50%	*50%-57%*	267	50%	*44%-55%*	108	50%	*44%-56%*	154
2010	1 → 2	52%	*46%-56%*	1 012	54%	*52%-66%*	148	53%	*43%-63%*	30	48%	*43%-52%*	94
2012	1 → 2	48%	*44%-60%*	788	58%	*51%-66%*	44	52%	*35%-68%*	15	41%	*31%-51%*	27
2014	1 → 2	57%	*30%-85%*	30	53%	*39%-67%*	47	53%	*35%-69%*	10	35%	*18%-51%*	13
2016	1 → 2	55%	*31%-79%*	2	36%	*0%-87%*	0	44%	*29%-67%*	5	0%	*0%-0%*	1
2018	1 → 2	-	-	-	-	-	-	-	-	-	-	-	-
2004	2 → 3	0.22%	*0.10%-0.20%*	3 145	0%	*0%-0%*	470	0.3%	*0%-1.0%*	310	0%	*0%-0%*	143
2006	2 → 3	11%	*9.1%-12%*	2 694	11%	*9%-13%*	390	11%	*8.2%-14%*	248	13%	*9.6%-17%*	124
2008	2 → 3	12%	*10%-13%*	1 900	10%	*8%-13%*	352	8.7%	*5.5%-12%*	145	14%	*9.8%-18%*	137
2010	2 → 3	13%	*10%-17%*	1 578	10%	*8%-14%*	230	12%	*6.2%-18%*	50	13%	*6.8%-21%*	122
2012	2 → 3	13%	*7.9%-19%*	900	11%	*7%-15%*	82	20%	*6.7%-33%*	13	10%	*3%-21%*	28
2014	2 → 3	22%	*6.6%-18%*	45	12%	3%-21%	24	19%	*7.5%-34%*	10	8.7%	*5%-20%*	12
2016	2 → 3	13%	*0%-46%*	20	14%	0%-36%	0	73%	*0.52%-0.96%*	1	2.3%	*0%-30%*	1
2018	2 → 3	-	-	-	-	-	-	-	-	-	-	-	-
2004	3 → 4	0%	*0%-0%*	1 276	0%	*0%-0%*	162	0%	*0%-0%*	81	0%	*0%-0%*	58
2006	3 → 4	12%	*0%-0.12%*	1 000	1.2%	*0.10%-2%*	150	2.2%	*0.45%-4.1%*	56	3.8%	*1.7%-6%*	55
2008	3 → 4	0.2%	*0%-0.53%*	541	1.4%	*1%-2.2%*	91	2%	*0.34%-3.6%*	45	3%	*1.4%-4.6%*	32
2010	3 → 4	1.7%	*0.10%-3.9%*	217	2.2%	*0%-4%*	40	1.6%	*0.41%-3.2%*	36	1.7%	*0%-3.3%*	43
2012	3 → 4	1.7%	*0.10%-4%*	41	2.8%	*0%-5%*	17	1.7%	*0.22%-3.1%*	22	4.7%	*0%-7.4%*	38
2014	3 → 4	1.7%	*0%-4.3%*	18	3.6%	*0%-8.4%*	6	2.4%	*0.13%-0.5%*	10	6.6%	*0%-15%*	11
2016	3 → 4	0%	*0%-0%*	0	0%	*0%-0%*	2	0%	*0%-6.7%*	7	0%	*0%-0%*	1
2018	3 → 4	-	-	-	-	-	-	-	-	-	-	-	-

Prob = Probability of Relapse

CI95% = 95% Confidence Interval

at risk = Patients at risk

'Patients at risk' refers to a subset of the sample at each time point, which can increase or decrease depending on whether patients enter or leave the study. As shown in the table, there is frequently a higher number of patients relapsing from the state 'moderate' compared to the other states. The following subsection, 5.2.2, covers the results from the Nelson-Aalen estimator.

5.2.2 Cumulative Relapse Recurrence and Disease Progressiveness

This subsection explains the results of the Nelson-Aalen estimator that estimates the cumulative relapse recurrence (equation 9). The estimator provides the number of times relapse could be expected during the observation period. Figure 19 presents the Nelson-Aalen estimates for the cumulative intensities of relapse in each treatment arm:

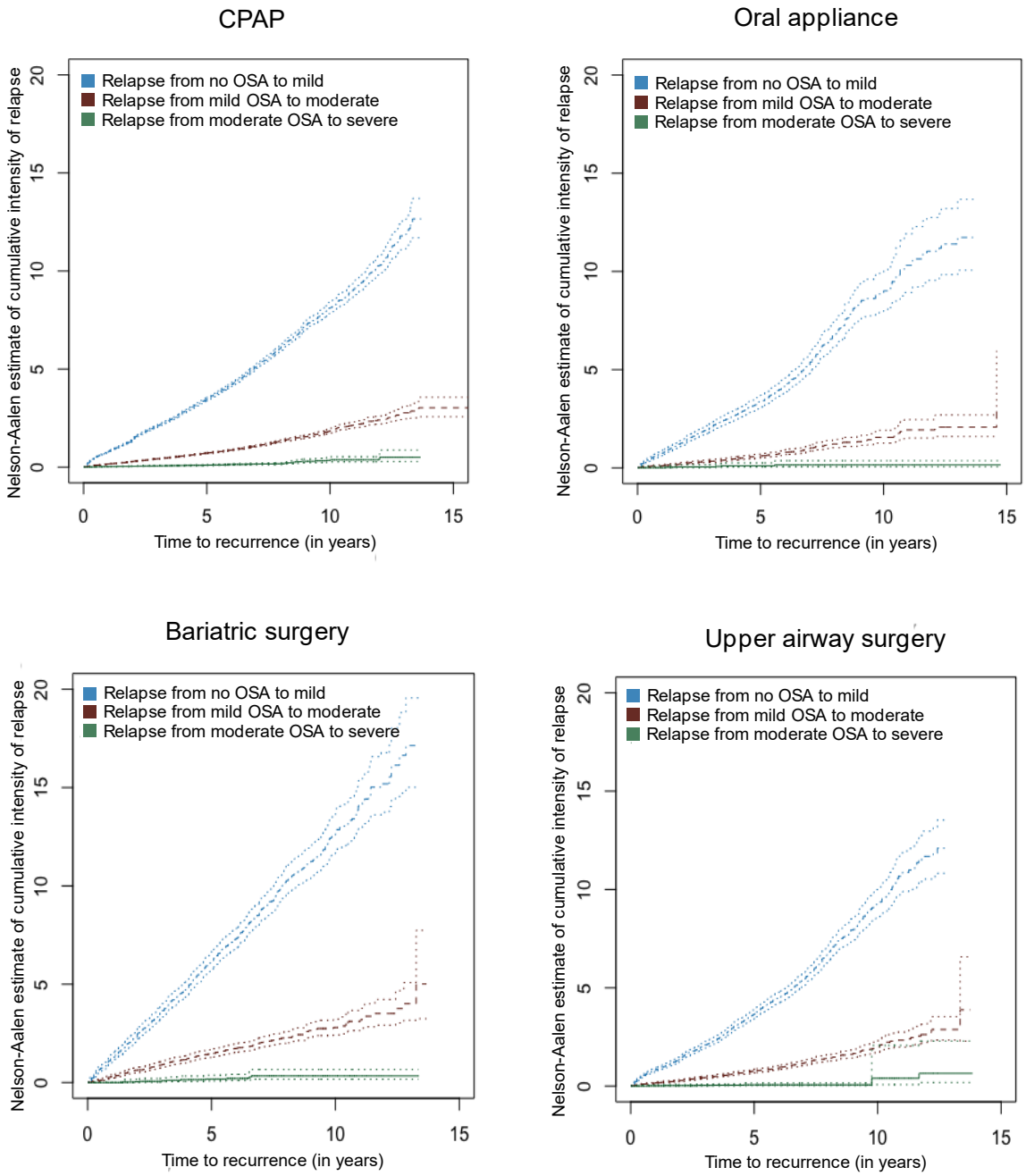


Figure 19. Nelson-Aalen Estimates for Relapse Types and Treatment Arms (with 95% CI)

The y-axes represent the Nelson-Aalen estimates of the cumulative intensities of relapse. The x-axes represent the time to occurrence in years. The Nelson-Aalen estimator curve demonstrates how the hazard rate changes over time. Thus, the 'slope' of the estimated curve indicates the relapse behaviour in different treatment arms. Steeper slopes indicate a higher risk of relapse. Furthermore, the approximate linearity suggests constant risk. The following bullet points compare each treatment method within a relapse type:

- **Cumulative relapse recurrence from 'no OSA' to 'mild OSA':** The patients that relapsed from the state 'no OSA' (AHI<5) had the highest estimates for relapse recurrence in all four treatment arms. High estimates indicate a faster accumulation of risk. The hazard rates were similar in all treatment arms, but slight differences were apparent: the patients treated with CPAP had a steadily increasing failure rate over the observation period. The subsamples treated with OA and upper airway surgery started with high estimate rates at the beginning of the observation period, but the rates decreased as time passed. A decay in hazard rates indicates that the risk of relapse recurrence decreases over the observation period. The patients treated with OA and upper airway surgery had similar estimate curves, but those treated with OA had the lowest number of cumulative relapses (11.7 relapses). The group treated with bariatric surgery had the steepest slope and the highest estimate of cumulative relapses (17 relapses). Thus, bariatric surgery was the treatment arm that accumulated the most risk of relapse during the observation period.
- **Cumulative relapse recurrence from 'mild OSA' to 'moderate OSA':** The results from this relapse type indicate that patients at state 'mild OSA' experienced significantly fewer cumulative relapses than those relapsing from state 'no OSA'. Patients treated with OA and CPAP had the slowest rates of failure; thus, these treatment types accumulated the least risk of relapse recurrence in this state. Patients treated with CPAP had the lowest estimate of cumulative relapse recurrence (three relapses). All treatment arms had increasing failure rates, and the arm treated with bariatric surgery reached the highest estimate of cumulative relapse recurrence (five relapses).
- **Cumulative relapse recurrence from 'moderate OSA' to 'severe OSA':** Overall, the results of this relapse type had the lowest estimates for cumulative relapse recurrence. The estimates did not reach a higher value than one during the observation period. Estimates below one indicate that the model does not estimate disease

recurrence in patients relapsing from ‘moderate’ to ‘severe OSA’. In this category, upper airway surgery had the highest rate of failure (0.65 relapses), while OA had the lowest (0.16 relapses).

Table 14 shows the numerical values for Nelson-Aalen estimates, 95% CI levels and the number of patients at risk.

Table 14: Nelson-Aalen Estimates, 95% CIs and Number of Patients at Risk.

		TREATMENT METHOD											
		CPAP			Upper Airway Surgery			Oral Appliance			Bariatric Surgery		
		NA	CI95%	at risk	NA	CI95%	at risk	NA	CI95%	at risk	NA	CI95%.	at risk
2004	1 → 2	0.79	*0.76-0.84*	2 358	0	*0-1.7*	385	0	*0-0*	200	1.3	*1.1-1.5*	113
2006	1 → 2	1.5	*1.4-1.5*	1 848	1.9	*1.8-2.1*	280	1.6	*1.4-1.8*	153	3.6	*3.3-3.9*	143
2008	1 → 2	2.7	*2.6-2.8*	1 865	3.6	*2.8-4.8*	267	3.3	*3.2-3.6*	108	6.1	*5.7-6.5*	154
2010	1 → 2	4.2	*4.1-4.3*	1 012	4.4	*4.1-4.6*	148	4.6	*6.6-7.2*	56	8.6	*8-9.2*	94
2012	1 → 2	6	*5.9-6.2*	788	7.7	*7.2-8.3*	44	8.7	*6.1-7.2*	30	11	*10-12*	27
2014	1 → 2	10	*9.8-12*	30	10	*9.5-11*	47	9.8	*8.6-11*	10	14	*13-15*	13
2016	1 → 2	13	*12-14*	2	12	*11-14*	0	11	*10-14*	5	17	*15-20*	1
2018	1 → 2	-	-	-	-	-	-	-	-	-	-	-	-
2004	2 → 3	0.16	*0.14-0.27*	3 145	0	*0-0*	470	0	*0-0*	310	0.30	*0.23-0.41*	143
2006	2 → 3	0.42	*0.40-0.44*	2 694	0.41	*0.35-0.48*	390	0.30	*0.21-0.55*	248	0.89	*0.74-1.1*	124
2008	2 → 3	0.71	*0.67-0.72*	1 900	0.88	*0.79-1*	352	0.54	*0.47-0.76*	145	1.5	*1.2-1.7*	137
2010	2 → 3	1.1	*0.92-1.1*	1 578	0.92	*0.67-1*	230	0.76	*0.50-0.81*	50	2	*1.8-2.3*	122
2012	2 → 3	1.5	*1.5-1.6*	900	1.6	*8-15.1*	82	1.2	*1.4-2.2*	30	2.7	*2.3-3*	28
2014	2 → 3	2.1	*1.9-2.3*	45	1.6	2-2.8	24	1.8	*1.4-2.7*	10	3.1	*2.6-3.7*	12
2016	2 → 3	3.1	*2.6-3.6*	20	3.9	2.3-6.6	0	2.1	*1.6-2.6*	2	5.0	*3.2-5.1*	1
2018	2 → 3	-	-	-	-	-	-	-	-	-	-	-	-
2004	3 → 4	0.032	*0.021-0.45*	1 276	0	*0-0*	162	0	*0-0*	81	0	*0-0*	58
2006	3 → 4	0.061	*0.052-0.082*	1 000	0.03	*0.012-0.091*	150	0.015	*0.010-0.019*	56	0.092	*0.031-0.24*	55
2008	3 → 4	0.093	*0.063-0.12*	541	0.06	*0.031-0.15*	91	0.023	*0.012-0.034*	45	0.17	*0.082-0.37*	32
2010	3 → 4	0.16	*0.12-0.20*	217	0.06	*0.034-0.16*	40	00.13	*0.065-0.15*	36	0.33	*0.16-0.66*	43
2012	3 → 4	0.30	*0.21-0.41*	41	0.40	*0.083-2.1*	17	00.14	*0.063-0.20*	22	0.34	*0.17-0.66*	38
2014	3 → 4	0.37	*0.26-0.54*	18	0.43	*0.086-2.1*	6	00.15	*0.071-0.37*	10	0.35	*0.22-0.67*	11
2016	3 → 4	0.50	*0.28-0.87*	0	0.65	*0.18-2.3*	2	00.16	*0.093-0.37*	7	0.35	*0.17-0.66*	1
2018	3 → 4	-	-	-	-	-	-	-	-	-	-	-	-

NA = Nelson-Aalen estimate

CI95% = Confidence interval of 95%

at risk = Patients at risk

This concludes the analysis regarding the OSA treatment effect. The following subsection, 5.2.3, examines the research hypotheses. After that, subsection 5.2.4 discusses the reliability of the results.

5.2.3 Research Hypothesis 2: Differences in Treatment Outcomes

This subsection evaluates the plausibility of research hypothesis 2 **(H2)** developed for treatment analysis in section 2.4. Hypothesis **H2** demonstrates whether the data shows a statistical difference between treatment arms. Thus, the research hypothesis is presented in the following way:

H2: There are statistical differences between the four OSA treatment arms. CPAP is expected to have the most substantial effect in decreasing the risk of relapse and relapse recurrence in OSA.

We applied a log-rank test to compare the four treatment arms. The result of the log-rank test was statistically significant ($p<0.001$), indicating that we can reject the null hypothesis of no statistical difference between the treatment arms. Therefore, there is sufficient evidence to conclude that there are significant differences between the four treatment arms. Furthermore, we expected CPAP to be the most effective treatment for OSA. Based on the results of this study, the relative effectiveness of OSA treatment depends on the severity state; thus, CPAP is the most effective intervention for treating mild-to-moderate OSA. Therefore, we conclude that there are differences between treatment arms, and CPAP is superior to other methods in mild-to-moderate OSA.

5.2.4 Reliability of the Results: Treatment Effect Analysis

This subsection explains the factors behind the reliability of the results from treatment analysis. There were evident similarities between the results of the Nelson-Aalen estimator and the Aalen-Johansen estimator in terms of relapse types. Thus, the results from both models support the main finding regarding the higher risk of relapse (or relapse recurrence) in milder states. Furthermore, CIs surrounding transition probabilities and Nelson-Aalen estimates were acceptable but got wider towards the end of the observation period. Therefore, the unreliability of the results increases as the set of patients at risk becomes smaller. Additionally, the high variance in the probabilities of relapse makes them hard to compare with averages. Furthermore, we assessed the general goodness-of-fit of the models via graphical evaluation. We compared the observed relapse probabilities with the expected transition probabilities within each treatment arm. The results from this analysis showed that while the differences between observed and predicted transition probabilities were significant at the 0.05 level of significance for some periods, the differences were

minor. Furthermore, the discrepancies were more evident at the end of the observation period. Thus, it is safe to assume that the model fitted satisfactorily.

Nonetheless, to rely on the inference from transition intensities, it was required to investigate whether the Markov assumption holds. The following hypotheses were developed for testing the Markov assumption:

H_0: The sequence is first-order Markov

H_1: The sequence is not first-order Markov

The procedure applies a Kolmogorov-Smirnov test for uniformity and a Ljung-Box test for independence. The smallest adjusted p-value represents the overall p-value from the test. If the p-value is less than the significance level, then the sequence is not first-order Markov.

Table 15: The Markov Assumption Test

Markov Assumption Test	
Test	P
Ljung-Box	0.23
Kolmogorov-Smirnov	0.37

Asterisks represent significance levels: *p < 0.05, **p < 0.01, ***p < 0.001

Table 15 shows that the Box-Ljung test had the smallest p-value (0.23) for this test. As the p-value was higher than a significance level of 0.05, the null hypothesis cannot be rejected. Therefore, the sequence is assumed to be first-order Markov. Furthermore, we conclude that the Markov property is satisfied, and any current state of the process was enough to predict the future state of the process. Additionally, it was necessary to investigate whether independent censoring holds in the data. Independent censoring assumes that within any subgroup of interest, the censored subjects should represent the subgroup that remained at risk (Aalen et al., 2008, p.58). Independent censoring was tested via visual assessment by plotting the non-censored and censored values against time and evaluating patterns across both groups. The assumption of independent censoring was especially relevant for the Aalen-Johansen estimator and the Nelson-Aalen estimator (and, to some extent, relevant for the log-rank statistic).

Moreover, the log-rank statistic assumes that the censoring patterns must be similar for all treatment arms (Bland & Altman, 2004). The analysis of censored values did not show any apparent differences between the treatment arms. Thus, it is safe to assume that

the similarity assumption of censored values across treatment arms is satisfied for the log-rank test. However, the log-rank statistic also assumes that the PH assumption is met. This study did not assess the PH assumption, as the log-rank test is still statistically valid under non-PH (Lin et al., 2020). Although the log-rank test might suffer from substantial power loss due to the unmet assumption, the moderate sample sizes might mitigate this weakness (Lin et al., 2020). Nevertheless, the unassessed PH assumption weakens the reliability of the log-rank test.

6 Conclusions & Discussion

The purpose of this chapter is to discuss the findings of this study. This chapter is constructed as follows. First, section 6.1 summarises the key findings of this study, and section 6.4 discusses the limitations of the study. Then, section 6.5 considers future research avenues. Finally, section 6.6 elaborates on the ethical considerations of this study, and section 6.7 presents the final conclusions.

6.1 Summary

This study investigated the treatment effect of four OSA treatment methods: CPAP, upper airway surgery, OA, and bariatric surgery. Treatment (in)effectiveness was measured by 1) the risk of relapse and 2) the progressiveness of the disease. The study was a retrospective cohort study where all participants had been diagnosed with OSA and treated within the Hospital District of Southwest Finland. The data consisted of irregular time-stamped EHRs, and the final sample included 15 179 patients matching the sample selection criteria. The data were collected from an EHR database of patients followed between 2003 and 2019.

A large part of the work involved extracting and cleaning study features from the EHR database. Treatment identifications were processed from raw patient data, which resulted in size variations among the treatment subsamples. The treatment identification procedure showed signs of patients with multilevel treatment plans. However, only the first treatment method was considered for patients treated with two or more treatment methods. Study findings from the pre-processing phase were visualised as an exploratory analysis presented in section 0. The analysis of patient biomarker data showed clear signs of OSA among the patient sample. Biomarker values such as blood pressure and ODI were higher in the study sample than in healthy adults. Additionally, as expected, values from subjects categorised as severe OSA patients showed lower SaO2% values and higher ESS values. Moreover, the exploratory analysis showed that the number of relapses had decreased over time. Especially patients treated with CPAP showed significant decay in relapses over time. Now that we have summarised the study process, the following sections discuss the findings in more detail. The following sections are divided according to the two phases of this study: section 6.2 presents the findings from severity classification, and section 6.3 covers treatment analysis.

6.2 Severity Classification

The objective of severity classification was to estimate the influence of physiological factors on OSA severity levels and thus validate the severity division relevant to calculate patient relapse. The following research question guided in analysing severity classification:

RQ1: Are there statistical differences between OSA severity levels?

The study results demonstrated statistical differences between the four OSA severity levels. To summarise, this study considered the effect of eleven biomarker variables on AHI-derived OSA severity. Four variables related to oxygen saturation (SaO2%, SaO2% min, ODI4 and ODI10) were statistically significant predictors of OSA severity. This finding indicates that patients with higher OSA severity tend to experience lower oxygen saturation levels. The association between low oxygen saturation and high OSA severity is consistent with clinical consensus (Fernandes, 2021; Almazaydeh, 2012; Sahib, 2022). Furthermore, according to theory, the low levels of oxygen in the blood produce a state called hypoxemia (Dewan et al., 2015; Farré et al., 2018; Cowie, 2017). Prolonged exposure to hypoxemia can generate a chronic state of intermittent hypoxia (Fletcher et al., 2001; Tobaldini et al., 2017). Therefore, moderate-to-severe OSA patients with lower SaO2% levels are at a higher risk of chronic intermittent hypoxia. Furthermore, hypoxia relates to other risks, such as autonomic deregulation and increased risk of cardiopulmonary diseases (Fletcher et al., 2001; Tobaldini et al., 2017). The findings of this study emphasise the importance of treating OSA at milder states before the disease evolves in severity.

6.3 Treatment Analysis

The objective of treatment analysis was to compare the clinical effectiveness (or *ineffectiveness*) of four OSA treatment alternatives. This study analysed the outcomes of CPAP, OA, upper airway surgery and bariatric surgery. The following research question was considered to guide the investigation:

RQ2: Do the different OSA treatment interventions lead to different outcomes?

The results of this study indicated statistically significant differences between treatment outcomes. Nevertheless, regardless of significant differences, all relapse types (i.e., relapse from 'no OSA' to 'mild OSA') had similar value ranges in all treatment arms. Therefore, the following subsections explain the findings of this study in the following order: subsection 6.3.1 summarises the common findings among treatment arms. In contrast, subsection 6.3.2 entails a comparative analysis of treatment outcomes.

6.3.1 Summary of OSA Treatment Effect Analysis

This study demonstrated that non-OSA patients (AHI<5) had the highest risk of relapse. Non-OSA patients had approximately a 50% risk of relapse and a high relapse recurrence of nearly 12 relapses in all treatment arms. The high risk of relapse in asymptomatic patients accentuates the difficulty of achieving a permanent cure for OSA. Therefore, to avoid disease recurrence in treated OSA patients, it might be necessary to continue with non-surgical treatment options after a treatment intervention, even if all symptoms appear to have healed. Furthermore, disease severity assessment should always consider the subjective symptoms of a patient, as there is a high likelihood of disease presence after momentary healing.

Igelström et al. (2017) demonstrated that patients with greater severity at baseline often experience more significant AHI reductions. Consequently, the categorisation of AHI might be too insensitive to capture the actual disease progression. Thus, the shorter AHI range in non-OSA patients compared to mild OSA patients partly affects the high likelihood of relapse in non-OSA patients. Furthermore, patients relapsing from 'mild OSA' ($5 \leq AHI \geq 15$) had approximately a 15% risk of relapse and a moderate relapse recurrence (approx. three relapses) in all treatment arms. Moreover, patients with 'moderate OSA' ($15 < AHI \geq 30$) had the lowest probability of transitioning into a severe OSA state ($AHI < 30$). The probabilities of relapse were less than 6% for all treatment arms. The results for cumulative relapse recurrence were aligned with this finding, and estimates for relapse recurrence did not pass one unit of relapse in any treatment arm.

To summarise, the results indicate that patients with milder forms of the condition are more at risk of relapsing than patients with increased disease severity. Similarly, Lin et al. (2008) reported that patients with severe OSA benefit more from treatment than patients with milder OSA. Thus, it could be further hypothesised that all four OSA treatments are more effective in preventing relapse and relapse recurrence in the more severe OSA states. However, there is not enough evidence to support this hypothesis, as there was no control

group to indicate the relapse behaviour of untreated patients. Furthermore, although study results were similar in differing treatment arms, some differences were apparent. The results from this study support previous evidence indicating that different treatments might be more beneficial at differing severity levels (Chang, 2020; Campos-Rodriguez, 2012). The following subsection, 6.3.2, presents the comparisons between treatment outcomes.

6.3.2 Comparisons of Outcomes by Treatment Arms

Table 16 summarises the main findings of the comparative analysis.

Table 16: Summary Table: Relapse Probabilities and Cumulative Relapse Estimates

RESULTS SUMMARY TABLE

Treatment method			Relapse: 'no OSA' to 'mild'		Relapse: 'mild' to 'moderate'		Relapse: 'moderate' to 'severe'	
			Relapse P%	Cumulative	Relapse P%	Cumulative	Relapse P%	Cumulative
		Min	40 %		6 %		1 %	
	CPAP	Max	57 %	12.7	21 %	**3**	3 %	0.501
		Avg	50 %		**12.3** %		**1.4** %	
	Upper	Min	36 %		8 %		0 %	
	Airway	Max	58 %	12.1	36 %	4	5 %	**0.652**
	Surgery	Avg	**50.2** %		14.5 %		1.6 %	
		Min	40 %		6 %		0.88 %	
	OA	Max	73 %	**11.7**	29 %	**3**	6 %	**0.163**
		Avg	50.1 %		16 %		**3.1** %	
		Min	39 %		7 %		0 %	
	Bariatric	Max	68 %	**17.1**	60 %	**5**	3 %	0.352
	Surgery	Avg	**48.7** %		**22.1** %		2 %	

The following analysis elaborates on the study findings regarding each treatment arm:

CPAP: Patients treated with CPAP had relatively low relapse probabilities in all three relapse categories. CPAP was the most effective treatment in the category 'mild OSA' by having an overall probability of relapse of 12.3% and an estimated cumulative relapse recurrence of 3 relapses. In addition, patients treated with CPAP had the lowest average relapse probability when relapsing from 'moderate OSA'. Recent research has

demonstrated that CPAP is more effective for moderate and severe OSA than for milder forms of the condition (McEvoy, 2016; Chowdhuri et al., 2016). The results of this study strengthen this finding by suggesting better suitability of CPAP for more severe OSA. Furthermore, clinical studies have reported CPAP as superior to OA (Sutherland et al., 2014; Balk et al., 2011). Based on the results of this study, the superiority of CPAP is unclear: although CPAP did perform slightly better than OA at the state 'moderate OSA', it was less effective for those at the state 'no OSA'. Nevertheless, the results of this study align with the general recognition of the effectiveness of CPAP therapy. Furthermore, the results of this study support the prescription of CPAP to patients at all severity levels.

Upper Airway Surgery: Upper airway surgery did not perform best in any of the three relapse categories. Verse et al. (2011) argue that patients with mild-to-moderate OSA have an equal response rate for upper airway surgery as for CPAP. The results of this study support equal performance between CPAP and upper airway surgery when relapsing from 'no OSA'. Conversely, upper airway surgery did perform worse than CPAP in 'moderate OSA' states. Furthermore, it has been shown that the effectiveness of upper airway surgery diminishes over time (Sher, 2002). The results of this study align with this argument, as the high estimates for cumulative relapse recurrence could demonstrate long-term ineffectiveness for upper airway surgery. Furthermore, Lin et al. (2008) researched upper airway surgery and discovered that OSA severity could not predict treatment outcomes. Nevertheless, this study does not coincide with the findings of Lin et al. (2008), as treatment failure was less likely to occur in OSA states of increased severity. Therefore, there is a higher relapse probability in milder OSA states. To conclude, this study ranks upper airway surgery as a less effective treatment method in all three relapse types. The results of this study show that upper airway surgery would be most suitable for patients with mild OSA.

***OA*:** Patients treated with OA performed comparatively well within all relapse types. OA is seen as an alternative to CPAP because of its non-invasive nature, but clinical studies often report OA as less effective than CPAP (Sutherland et al., 2015; Li et al., 2013; Marklund et al., 2017; Lam et al., 2007). Conversely, the results of this study do not show clear inferiority of OA compared to CPAP. Furthermore, OA was the most effective in preventing recurrence in all relapse categories. OA's effectiveness regarding disease progressiveness could be explained by better treatment adherence (Sutherland et al., 2015;

Li et al., 2013). Previous studies coincide with the effectiveness of OA, especially for the milder forms of OSA (Ilea et al., 2021; Ferguson et al., 2006). Nevertheless, the results of this study show a comparatively high risk of relapse in patients relapsing from 'moderate OSA'. To summarise, this study supports the effectiveness of OA and would recommend OA for patients with recurrent relapses. The results of this study would support OA as a non-invasive treatment alternative for CPAP.

Bariatric Surgery: This study ranks bariatric surgery as a comparatively less effective treatment method. This result aligns with previous findings. For example, Peromaa-Haavisto et al. (2016) argue that bariatric surgery is insufficient for treating OSA, even though weight loss treatment in OSA patients is justified. The results of this study support this argument, as bariatric surgery shows low effectiveness in treating milder OSA forms. Therefore, patients treated with bariatric surgery could benefit from a multilevel treatment plan that combines bariatric surgery with other treatments. Furthermore, the results of this study show that bariatric surgery is not successful in promoting a permanent cure. Permanent cure is unlikely as patients treated with bariatric surgery had the highest relapse recurrence estimate in non-OSA and mild patients, suggesting poor long-time treatment response. This finding does not coincide with previous findings claiming that obesity surgery could effectively promote a permanent cure (Verse et al., 2005; Rasheid et al., 2013; Boot et al., 2000). Therefore, we conclude that non-surgical interventions and upper airway surgery are more effective than bariatric surgery in preventing recurrent relapse.

6.4 Limitations

The findings of this study are subject to a variety of limitations. The following subsections explain the limitations relevant to this study.

6.4.1 Intermittent Observations and Missing Data

A standard limitation in longitudinal studies is the intermittent nature of biomarker measurements. Longitudinal EHRs are snapshots of a patient's timeline; therefore, the observed disease progression does not always match the underlying health status. For example, a patient could experience more relapses than recorded in the EHR data or a patient could spend most of the time in a relapsed state but become healthier on the day the measurements were taken and then relapse back to a less healthy state. In both cases, measurements would not record the true status of health. Another consequence of

intermittent data collection is the complexity of capturing the true disease progressiveness. The scope of this thesis was limited to only analysing relapses from consecutive states, and therefore relapses that skipped states were excluded. An example of a relapse that does not happen between two consecutive states would be a relapse from mild-to-severe OSA. Here the patient records would not show any visit to the state 'moderate' before entering to state 'severe'. Therefore, a missing record for state 'moderate' would be a consequence of irregular visits that fail to capture the true progressiveness of a disease. Not accounting for all relapse types generates a loss of information; but taking all possible transitions into account would overcomplicate the analysis. Misclassification models account for the issues that arise from the intermittent data. However, for this study, there were no software packages for connecting the Nelson-Aalen estimator to a Hidden Markov Model of misclassified states.

In addition to misclassified states, data quality showed critical data missingness. There were less biomarker data available for the early years of the observation period compared to later years. Missing data reduces the statistical validity of a study and produces biased estimates and invalid conclusions (Huang et al., 2020). Thus, it was essential to find an imputation method that could handle a high level of missingness, non-normally distributed biomarker variables and nested EHRs. Another limitation was that CLMM is not equipped to handle censored data. Censored data is a special type of missing data in time-to-event analyses (Koul et al., 1995). Furthermore, the treatment analysis assumed the data to be right-censored, but the data could also be interpreted as interval-censored. Unfortunately, identifiability problems prevented further evaluating the assumption of right-censoring (see: Beyersmann et al., 2011, p.23). However, there is a high likelihood that the data is right-censored, as right-censoring is the most common type of incompletely observed time-to-event data (Beyersmann et al., 2011, p.23).

6.4.2 Research Design

The lack of a control group proposed a weakness in the study design. Clinical investigations often approach treatment effect research by comparing the probability of an event on individuals exposed to an intervention with those not exposed. This comparison is made via a control group, a group of subjects identical to the treatment sample except that they lack the intervention of interest (Malay et al., 2012). A control group is beneficial as it enables greater comparability of outcomes, and the observed effects in the intervention group are more directly attributed to the treatment alone (Malay et al., 2012). However, the

original data did not distinguish between treated and untreated patients. Additionally, the study results assume treatment adherence, but treatment adherence had not been documented in the data. Treatment adherence is relevant for CPAP and OA.

6.4.3 The Research Sample

This study applied the *per-protocol* approach in handling patients with multiple treatments. The per-protocol approach consists of excluding patients at the point of switching treatments. Although this is a valid method, the approach is often considered less conservative because it increases the probability of making a type I error (Gupta, 2011; Currow, 2012). In addition, the method might overestimate the treatment effect. However, the data acquired for this thesis did not have the treatment intention information required for conducting the research utilising the more conservative *intention-to-treat* approach. Furthermore, another limitation of the per-protocol approach was that patients might have been treated with another OSA intervention before participating in this study.

The data comprised only specialised EHRs; therefore, patients with milder OSA had been allocated to primary healthcare units. This generated an over-representation of patients with a more severe condition. Consequently, the over-representation of severe patients led to higher sample sizes in patients relapsing from moderate OSA but smaller sample sizes in patients relapsing from milder OSA states. Smaller sample sizes in milder patients resulted in larger CIs and uncertainty in the results. Furthermore, the treatment arms differed in sample sizes. Differing sample sizes are common in epidemiologic studies and pose challenges in comparing the results. Unequal sample sizes generate unequal variances and a general loss of power.

Furthermore, the generalisation of the results between OSA and bariatric surgery must be cautious, as the leading medical condition in the ‘bariatric surgery’ treatment arm has likely been obesity rather than OSA. Therefore, although bariatric surgery is an OSA treatment method, the analysis should consider the possible bias generated by assuming that the surgery was performed as OSA treatment.

6.5 Future Research

This study points to several avenues for future research on OSA severity and treatment effect analysis. The adoption of EHR data collection systems has increased rapidly since 2010. Consequently, EHR databases with extended observation periods are becoming more

available for research purposes. As noted in section 6.4, EHR data is subject to intermittent observations that cannot capture the actual underlying process of a disease. Future research could apply Hidden Markov models to model the underlying disease progression behind irregular sampling times. Additionally, the definition of OSA severity requires reassessment in clinical studies. The lack of generalisation in OSA severity makes studies hard to compare. Therefore, more in-depth research on OSA severity assessment would support OSA severity standardisation. Based on the results of this study, future research could consider a combination of AHI and ODI (or SaO2%) to construct a new OSA severity indicator.

Furthermore, future OSA treatment effect research could incorporate the analysis of covariates to gain more information on patient characteristics that influence relapse. OSA is considered a heterogenic disease, meaning that the efficiency of treatment varies depending on the age, sex, and ethnicity of the patient (Geovanini, 2018). Associating the OSA treatment effect with patient characteristics such as obesity would emphasise the role of individual attributes in the relapsing process. For example, distinct patient phenotype clusters could have differing relapse responses depending on the OSA treatment type.

Future OSA treatment effect research could consider investigating the relapse response for multilevel treatment and reviewing different treatment combinations in conjunction. Instead of analysing individual treatment responses, the study could focus on comparing patient groups that have undergone a set of distinct treatments. The data set contained approximately 1 000 records for nasal interventions, and 70% were associated with either CPAP or upper airway surgery. Thus, nasal surgery is considered a supportive treatment method and should be recognised in patient responses to treatments. Furthermore, including a control group of non-treated OSA patients would be beneficial. Although finding a subsample of non-treated patients in a long-term EHR research design can pose practical challenges.

6.6 Ethical Considerations

This study considered important ethical principles to guide the research procedure. The study was conducted under a research permit (No T164/2019) from Turku University Hospital, Finland. The ethical risks of epidemiologic studies are minimised by protecting the confidentiality of patient data. The data set of this study was fully anonymised by masking direct and indirect patient identifiers. The anonymised data offered increased

protection of patient privacy during the research process. Additionally, any access to data was controlled by establishing appropriate levels of security in data transfer activities. Furthermore, this study utilised a legally binding data-sharing agreement. This agreement prohibited downloading or sharing data with third parties. Additionally, any attempt to identify patients was prohibited. Furthermore, no underage patients were included in the study. Finally, an extensive discussion on research limitations was conducted to ensure the understanding of possible biases in clinical decision-making processes.

6.7 Conclusions

The results of this study indicate that non-OSA patients had approximately a 50% relapse risk and the highest relapse recurrence estimates in all treatment arms. The high relapse rate accentuates the difficulty of finding a permanent cure for OSA. Approximately 26% of the Finnish population suffers from mild OSA; consequently, the number of outpatient visits continues to expand rapidly (Bachour & Avellan-Hietanen, 2021; Mattila, 2022). Thus, there might be a connection between the increased relapse potential in milder patients and the growing number of patients seeking medical care. The results of this study demonstrate that treating OSA is a long-term process, and long-term diseases imply high costs covered by public welfare. However, the study demonstrated that relapse recurrence is the lowest in states of increased OSA severity. A lower risk of relapse in severe states could indicate that the disease progression slows down in treated patients as the severity increases.

The study results increased our understanding of the differences between OSA treatment outcomes. Distinct treatments are applicable at differing OSA severity levels. Furthermore, the results of this study support the effectiveness of CPAP as the standard OSA treatment method. CPAP effectively treated OSA at all severity levels, but especially at *mild and moderate* OSA states. Furthermore, patients treated with OA had the lowest risk of recurrent relapses, and thus it could be further argued that OA supports the prevention of OSA disease progression. Thus, the results of this study would support the prescription of OA as a non-surgical treatment alternative for CPAP. Moreover, the study results ranked bariatric surgery as the least effective OSA treatment method, particularly in milder severity levels. One factor contributing to the failure of bariatric surgery is that it does not address the disease directly. Bariatric surgery is performed in the abdominal area, but the detriments for OSA are often the excessive body tissue around the neck area.

In addition to the comparative analysis of the OSA treatment effect, the study demonstrated a statistical difference between OSA severity levels. Furthermore, the study results demonstrated that higher OSA severity is associated with lower oxygen saturation levels. This finding emphasises the importance of early diagnosis and treatment, as prolonged exposure to low oxygen saturation levels increases the risk of chronic intermittent hypoxia (Fletcher et al., 2001; Tobaldini et al., 2017).

7 References

Aalen, O. & Johansen, S. (1978). An empirical transition matrix for non-homogeneous Markov chains based on censored observations. *Scandinavian Journal of Statistics*, 141-150.

Aalen, O., Borgan, O., & Gjessing, H. (2008). Survival and event history analysis: a process point of view. *Springer Science & Business Media*.

Abarca-Gómez, L., Abdeen, Z. A., Hamid, Z. A., Abu-Rmeileh, N. M., Acosta-Cazares, B., Acuin, C. & Cho, Y. (2017). Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128· 9 million children, adolescents, and adults. *The lancet*, *390*(10113), 2627-2642.

Abdeen, G., & Le Roux, C. W. (2016). Mechanism underlying the weight loss and complications of Roux-en-Y gastric bypass. Review. *Obesity surgery*, *26*(2), 410-421.

Agresti, A. (2010). *Analysis of ordinal categorical data* (Vol. 656). John Wiley & Sons.

Ahmadi, N., Shapiro, G. K., Chung, S. A., & Shapiro, C. M. (2009). Clinical diagnosis of sleep apnea based on single night of polysomnography vs. two nights of polysomnography. *Sleep and Breathing*, *13*(3), 221-226.

Alin, A. (2010). Multicollinearity. *Wiley interdisciplinary reviews: computational statistics*, *2*(3), 370-374.

Allignol, A., Schumacher, M., & Beyersmann, J. (2011). Empirical transition matrix of multi-state models: the etm package. *Journal of Statistical Software*, *38*, 1-15.

Almazaydeh, L., Elleithy, K., & Faezipour, M. (2012). Detection of obstructive sleep apnea through ECG signal features. *IEEE International Conference on Electro/Information Technology (pp. 1-6). IEEE.*

Andersen, P. K., & Keiding, N. (2002). Multi-state models for event history analysis. *Statistical methods in medical research*, 11(2), 91-115.

Antic, N. A., Catcheside, P., Buchan, C., Hensley, M., Naughton, M. T., Rowland, S., ... & McEvoy, R. D. (2011). The effect of CPAP in normalizing daytime sleepiness, quality of life, and neurocognitive function in patients with moderate to severe OSA. *Sleep*, *34*(1), 111-119.

Aro. A., Myllylä. M., Anttalainen U., Saaresranta., T. (2019). 'Uniapneaepäily', *Lääketieteellinen aikakausikirja Duodecim*.

Ashrafian, H., Toma, T., Rowland, S. P., Harling, L., Tan, A., Efthimiou, E., ... & Athanasiou, T. (2015). Bariatric surgery or non-surgical weight loss for obstructive sleep apnoea? A systematic review and comparison of meta-analyses. *Obesity surgery*, *25*(7), 1239-1250.

Aziz, A., Sanwal, K., Singhal, V., & Brayton, R. (1996). Verifying continuous time Markov chains. In *International Conference on Computer Aided Verification* (pp. 269-276). Springer, Berlin, Heidelberg.

Bachour L., Bäck, A., (2015). Obstruktiivinen uniapnea aikauisilla. *Lääketieteellinen aikakausikirja Duodecim*. 131(16):1477–9.

Bachour, A & Avellan-Hietanen, H (2021). ' Obstruktiivinen uniapnea aikuisilla', *Suomen lääkärilehti*, Vuosikerta. 76, Nro 14, Sivut 865-870. Accessed on 10/10/2022 in https://www.laakarilehti.fi/pdf/2021/SLL142021-865.pdf

Bakdash, J. Z., & Marusich, L. R. (2017). Repeated measures correlation. *Frontiers in psychology*, *8*, 456.

Bakker, J. P., Edwards, B. A., Gautam, S. P., Montesi, S. B., Durán-Cantolla, J., Barandiarán, F. A., ... & Malhotra, A. (2014). Blood pressure improvement with continuous positive airway pressure is independent of obstructive sleep apnea severity. *Journal of Clinical Sleep Medicine*, *10*(4), 365-369.

Baldini, N., Gagnadoux, F., Trzepizur, W., Meslier, N., Dugas, J., Gerves-Pinquie, C., ... & Kün-Darbois, J. D. (2021). Long-term dentoskeletal side effects of mandibular advancement therapy in patients with obstructive sleep apnea: data from the Pays de la Loire sleep cohort. *Clinical Oral Investigations*, 1-12.

Balk, E. M., Moorthy, D., Obadan, N. O., Patel, K., Ip, S., Chung, M., ... & Lau, J. (2011). *Diagnosis and treatment of obstructive sleep apnea in adults.*

Bartels, B. (2008). Beyond" fixed versus random effects": a framework for improving substantive and statistical analysis of panel, time-series cross-sectional, and multilevel data. *The Society for Political Methodology*, *9*, 1-43.

Berger, G., Berger, R., & Oksenberg, A. (2009). Progression of snoring and obstructive sleep apnoea: the role of increasing weight and time. *European Respiratory Journal*, *33*(2), 338-345.

Bergeron, M., Duggins, A. L., Cohen, A. P., Leader, B. A., & Ishman, S. L. (2020). The impact of persistent pediatric obstructive sleep apnea on the Quality of Life of Patients' families. *International Journal of Pediatric Otorhinolaryngology*, *129*, 109723.

Berry, R. B., Budhiraja, R., Gottlieb, D. J., Gozal, D., Iber, C., Kapur, V. K., ... & Tangredi, M. M. (2012). Rules for scoring respiratory events in sleep: update of the 2007 AASM manual for the scoring of sleep and associated events: deliberations of the sleep apnea definitions task force of the American Academy of Sleep Medicine. *Journal of clinical sleep medicine*, *8*(5), 597-619.

Beyersmann, J., Allignol, A., & Schumacher, M. (2011). *Competing risks and multistate models with R*. Springer Science & Business Media.

Bikov, A., & Dragonieri, S. (2020). Obstructive Sleep Apnea.

Bland, J. M., & Altman, D. G. (2004). The logrank test. *Bmj*, *328*(7447), 1073.

Bonsignore, M. R., Baiamonte, P., Mazzuca, E., Castrogiovanni, A., & Marrone, O. (2019). Obstructive sleep apnea and comorbidities: a dangerous liaison. *Multidisciplinary respiratory medicine*, *14*(1), 1-12.

Boot, H., van Wegen, R., Poublon, R. M., Bogaard, J. M., Schmitz, P. I., & van der Meché, F. G. (2000). Long-term results of uvulopalatopharyngoplasty for obstructive sleep apnea syndrome. *The Laryngoscope*, *110*(3), 469-475.

Borgan, Ø. (1997). Three contributions to the encyclopedia of biostatistics: The Nelson-Aalen, Kaplan-Meier, and Aalen-Johansen. *Preprint series. Statistical Research Report http://urn. nb. no/URN: NBN: no-23420*.

Borgan, Ø. (2014). Nelson–Aalen Estimator. *Wiley StatsRef: Statistics Reference Online*.

Boyd, S. B., Walters, A. S., Song, Y., & Wang, L. (2013). Comparative effectiveness of maxillomandibular advancement and uvulopalatopharyngoplasty for the treatment of moderate to severe obstructive sleep apnea. *Journal of Oral and Maxillofacial Surgery*, *71*(4), 743-751.

Brant, R. (1990). Assessing proportionality in the proportional odds model for ordinal logistic regression. *Biometrics*, 1171-1178.

Braumoeller, B. (2004). "Hypothesis testing and multiplicative interaction terms." *International organization* 58, no. 4: 807-820.

Buchanan, A., Cohen, R., Looney, S., Kalathingal, S., & De Rossi, S. (2016). Cone-beam CT analysis of patients with obstructive sleep apnea compared to normal controls. *Imaging science in dentistry*, *46*(1), 9-16.

Campos-Rodriguez, F., Martinez-Garcia, M. A., de la Cruz-Moron, I., Almeida-Gonzalez, C., Catalan-Serra, P., & Montserrat, J. M. (2012). Cardiovascular mortality in women with obstructive sleep apnea with or without continuous positive airway pressure treatment: a cohort study. *Annals of internal medicine*, *156*(2), 115-122

Caples, S. M., Gami, A. S., & Somers, V. K. (2005). Obstructive sleep apnea. *Annals of internal medicine*, *142*(3), 187-197.

Casey, J. A., Schwartz, B. S., Stewart, W. F., & Adler, N. E. (2016). Using electronic health records for population health research: a review of methods and applications. *Annual review of public health*, *37*, 61-81.

Chang, H. P., Chen, Y. F., & Du, J. K. (2020). Obstructive sleep apnea treatment in adults. *The Kaohsiung journal of medical sciences*, *36*(1), 7-12.

Ching, W. K., & Ng, M. K. (2013). Markov chains. *Models, algorithms and applications*.

Chowdhuri, S., Quan, S. F., Almeida, F., Ayappa, I., Batool-Anwar, S., Budhiraja, R., ... & Slyman, A. (2016). An official American Thoracic Society research statement: impact of mild obstructive sleep apnea in adults. *American journal of respiratory and critical care medicine*, *193*(9), e37-e54.

Christensen, R. H. B. (2015). Analysis of ordinal data with cumulative link models—estimation with the R-package ordinal. *R-package version*, *28*, 406.

Christensen, R. H. B. (2018). Cumulative link models for ordinal regression with the R package ordinal. *Submitted in J. Stat. Software*, *35*.

Christensen, R. H. B. (2019). A Tutorial on fitting Cumulative Link Mixed Models with clmm2 from the ordinal Package. *Tutorial for the R Package ordinal https://cran. r-project. org/web/packages/ordinal/Accessed*, *1*.

Cielo, C. M., & Tapia, I. E. (2019). Diving deeper: rethinking AHI as the primary measure of OSA severity. *Journal of Clinical Sleep Medicine*, *15*(8), 1075-1076.

Cook, R. J., & Lawless, J. F. (2007). *The statistical analysis of recurrent events* (pp. 128-133). New York: Springer.

Costa, J. C., Rebelo-Marques, A., Machado, J. N., Gama, J. M. R., Santos, C., Teixeira, F., & Moita, J. (2019). Validation of NoSAS (Neck, Obesity, Snoring, Age, Sex) score as a

screening tool for obstructive sleep apnea: analysis in a sleep clinic. *Pulmonology*, *25*(5), 263-270.

Cowie, M. R. (2017). Sleep apnea: state of the art. *Trends in cardiovascular medicine*, *27*(4), 280-289.

Currow, D. C., Plummer, J. L., Kutner, J. S., Samsa, G. P., & Abernethy, A. P. (2012). Analyzing phase III studies in hospice/palliative care. A solution that sits between intention-to-treat and per protocol analyses: the palliative-modified ITT analysis. *Journal of pain and symptom management*, *44*(4), 595-603.

DesRoches, C. M., Charles, D., Furukawa, M. F., Joshi, M. S., Kralovec, P., Mostashari, F., ... & Jha, A. K. (2013). Adoption of electronic health records grows rapidly, but fewer than half of US hospitals had at least a basic system in 2012. *Health Affairs*, *32*(8), 1478-1485.

Dewan, N. A., Nieto, F. J., & Somers, V. K. (2015). Intermittent hypoxemia and OSA: implications for comorbidities. *Chest*, *147*(1), 266-274.

Drager, L. F., Jun, J. C., & Polotsky, V. Y. (2010). Metabolic consequences of intermittent hypoxia: relevance to obstructive sleep apnea. *Best practice & research Clinical endocrinology & metabolism*, *24*(5), 843-851.

Duarte, R. L., Magalhães-da-Silveira, F. J., & Gozal, D. (2022). Clinical and polysomnographic predictors of suboptimal auto-adjusting CPAP titration in adult OSA patients: a single-center study. *European Archives of Oto-Rhino-Laryngology*, 1-8.

Edwards, B. A., Eckert, D. J., McSharry, D. G., Sands, S. A., Desai, A., Kehlmann, G., ... & Malhotra, A. (2014). Clinical predictors of the respiratory arousal threshold in patients with obstructive sleep apnea. *American journal of respiratory and critical care medicine*, *190*(11), 1293-1300.

Farré, R., Montserrat, J. M., Gozal, D., Almendros, I., & Navajas, D. (2018). Intermittent hypoxia severity in animal models of sleep apnea. *Frontiers in physiology*, 1556.

Ferguson, K. A., Cartwright, R., Rogers, R., & Schmidt-Nowara, W. (2006). Oral appliances for snoring and obstructive sleep apnea: a review. *Sleep*, *29*(2), 244-262.

Fernandes, E. R., Pires, G. N., Andersen, M. L., Tufik, S., & Rosa, D. S. (2021). Oxygen saturation as a predictor of inflammation in obstructive sleep apnea. *Sleep and Breathing*, 1-8.

Fizazi, K., Scher, H. I., Molina, A., Logothetis, C. J., Chi, K. N., Jones, R. J., ... & COU-AA-301 Investigators. (2012). Abiraterone acetate for treatment of metastatic castration-

resistant prostate cancer: final overall survival analysis of the COU-AA-301 randomised, double-blind, placebo-controlled phase 3 study. *The lancet oncology*, *13*(10), 983-992.

Fletcher, E. C. (2001). Invited review: Physiological consequences of intermittent hypoxia: systemic blood pressure. *Journal of Applied Physiology*, *90*(4), 1600-1605.
Fritscher, L. G., Mottin, C. C., Canani, S., & Chatkin, J. M. (2007). Obesity and obstructive sleep apnea-hypopnea syndrome: the impact of bariatric surgery. *Obesity surgery*, *17*(1), 95-99.

Gail, M. H., Altman, D. G., Cadarette, S. M., Collins, G., Evans, S. J., Sekula, P., ... & Woodward, M. (2019). Design choices for observational studies of the effect of exposure on disease incidence. *BMJ open*, *9*(12), e031031.

Geovanini, G. R., Wang, R., Weng, J., Jenny, N. S., Shea, S., Allison, M., ... & Redline, S. (2018). Association between obstructive sleep apnea and cardiovascular risk factors: variation by age, sex, and race. The Multi-Ethnic Study of Atherosclerosis. *Annals of the American Thoracic Society*, *15*(8), 970-977.

Geweke, J., Marshall, R. C., & Zarkin, G. A. (1986). Mobility indices in continuous time Markov chains. *Econometrica: Journal of the Econometric Society*, 1407-1423

Gleeson, M., & McNicholas, W. T. (2022). Bidirectional relationships of comorbidity with obstructive sleep apnoea. *European Respiratory Review*, *31*(164).

Gotsopoulos, H., Chen, C., Qian, J., & Cistulli, P. A. (2002). Oral appliance therapy improves symptoms in obstructive sleep apnea: a randomized, controlled trial. *American journal of respiratory and critical care medicine*, *166*(5), 743-748.

Greenburg, D. L., Lettieri, C. J., & Eliasson, A. H. (2009). Effects of surgical weight loss on measures of obstructive sleep apnea: a meta-analysis. *The American journal of medicine*, *122*(6), 535-542.

Greenwell, B. M., McCarthy, A. J., Boehmke, B. C., & Liu, D. (2018). Residuals and diagnostics for binary and ordinal regression models: an introduction to the sure package. *R J.*, *10*(1), 381.

Gujarati, D. & Porter, D. (2009). Basic Econometrics. 5th edition. McGraw-Hill Inc., New York, 922 s.

Guo, J., Sun, Y., Xue, L. J., Huang, Z. Y., Wang, Y. S., Zhang, L., ... & Yuan, L. X. (2016). Effect of CPAP therapy on cardiovascular events and mortality in patients with obstructive sleep apnea: a meta-analysis. *Sleep and Breathing*, *20*(3), 965-974.

Gupta, S. K. (2011). Intention-to-treat concept: a review. *Perspectives in clinical research*, *2*(3), 109.

Györfi, L., Kohler, M., Krzyzak, A., & Walk, H. (2002). *A distribution-free theory of nonparametric regression* (Vol. 1). New York: Springer.

Hafen, B. B., & Sharma, S. (2022). Oxygen saturation. 2021 Aug 12. *StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing*. Accessed on 23.07.2022 on www.ncbi.nlm.nih.gov/books/NBK525974/

Heatley, E. M., Harris, M., Battersby, M., McEvoy, R. D., Chai-Coetzer, C. L., & Antic, N. A. (2013). Obstructive sleep apnoea in adults: a common chronic condition in need of a comprehensive chronic condition management approach. *Sleep medicine reviews*, *17*(5), 349-355.

Hedeker, D. (2003). A mixed-effects multinomial logistic regression model. *Statistics in medicine*, *22*(9), 1433-1446.

Hedeker, D., Demirtas, H., & Mermelstein, R. J. (2009). A mixed ordinal location scale model for analysis of Ecological Momentary Assessment (EMA) data. *Statistics and its Interface*, *2*(4), 391.

Himpens, J., Cadière, G. B., Bazi, M., Vouche, M., Cadière, B., & Dapri, G. (2011). Long-term outcomes of laparoscopic adjustable gastric banding. *Archives of surgery*, *146*(7), 802-807.

Hirshkowitz, M. (2016). Polysomnography challenges. *Sleep medicine clinics*, *11*(4), 403-411.

Hougaard, P. (1999). Multi-state models: a review. *Lifetime data analysis*, *5*(3), 239-264.

Huang, B., Qiu, T., Chen, C., Neace, A., Zhang, Y., Yue, X., ... & Brunner, H. (2020). Comparative effectiveness research using electronic health records data: ensure data quality. SAGE Publications Ltd.

Huang QR, Qin Z, Zhang S, et al. (2008). Clinical patterns of obstructive sleep apnea and its comorbid conditions: a data mining approach. J Clin Sleep Med; 4: 543–550

Hudgel, D. W. (2016). Sleep apnea severity classification revisited. *Sleep*, *39*(5), 1165-1166.

Iber C, Ancoli-Israel S, Chesson A, Quan S. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*, 1st edn. Westchester, IL, American Academy of Sleep Medicine, 2007.

Igelström, H., Åsenlöf, P., Emtner, M., & Lindberg, E. (2018). Improvement in obstructive sleep apnea after a tailored behavioural sleep medicine intervention targeting healthy eating and physical activity: a randomised controlled trial. *Sleep and Breathing*, *22*(3), 653-661.

Ilea, A., Timuș, D., Höpken, J., Andrei, V., Băbțan, A. M., Petrescu, N. B., ... & Mesaros, A. (2021). Oral appliance therapy in obstructive sleep apnea and snoring-systematic review and new directions of development. *CRANIO®*, *39*(6), 472-483.

Imbens, G. W., & Kolesar, M. (2016). Robust standard errors in small samples: Some practical advice. *Review of Economics and Statistics*, *98*(4), 701-712.

Issa, F. G., & Sullivan, C. E. (1984). Upper airway closing pressures in obstructive sleep apnea. *Journal of Applied Physiology*, *57*(2), 520-527.

Jackson, C. (2011). Multi-state models for panel data: the msm package for R. Journal of statistical software, 38, 1-28.

Jacobsen, J. H., Shi, L., & Mokhlesi, B. (2013). Factors associated with excessive daytime sleepiness in patients with severe obstructive sleep apnea. *Sleep and Breathing*, *17*(2), 629-635.

Johansson, K., Neovius, M., Lagerros, Y. T., Harlid, R., Rössner, S., Granath, F., & Hemmingsson, E. (2009). Effect of a very low energy diet on moderate and severe obstructive sleep apnoea in obese men: a randomised controlled trial. *Bmj*, *339*.

Junior, C. M. C., & Palombini, L. (2010). Evaluation of the upper airway in obstructive sleep apnoea. *Indian J Med Res*, *131*, 230-235.

Kainulainen, S., Töyräs, J., Oksenberg, A., Korkalainen, H., Sefa, S., Kulkas, A., & Leppänen, T. (2019). Severity of desaturations reflects OSA-related daytime sleepiness better than AHI. *Journal of Clinical Sleep Medicine*, *15*(8), 1135-1142.

Kapur, V. K., Auckley, D. H., Chowdhuri, S., Kuhlmann, D. C., Mehra, R., Ramar, K., & Harrod, C. G. (2017). Clinical practice guideline for diagnostic testing for adult obstructive sleep apnea: an American Academy of Sleep Medicine clinical practice guideline. *Journal of Clinical Sleep Medicine*, *13*(3), 479-504.

Kemppainen S., Ikävalko T., Kellokoski J., Seppä J., Savolainen, J., (2019) Ortognaattinen hoito obstruktiivisen uniapnean hoitovaihtoehtona. Accessed on 23.05.2022 in https://www.duodecimlehti.fi/duo14753

Kent, B. D., Mitchell, P. D., & McNicholas, W. T. (2011). Hypoxemia in patients with COPD: cause, effects, and disease progression. *International journal of chronic obstructive pulmonary disease*, *6*, 199.

Khan, A., Ramar, K., Maddirala, S., Friedman, O., Pallanch, J. F., & Olson, E. J. (2009). Uvulopalatopharyngoplasty in the management of obstructive sleep apnea: the mayo clinic experience. In *Mayo Clinic Proceedings* (Vol. 84, No. 9, pp. 795-800). Elsevier.

Klein, J. P. (1991). Small Sample Moments of Some Estimators of the Variance of the Kaplan-Meier and Nelson-Aalen Estimators. *Scandinavian Journal of Statistics*, *18*(4), 333–340. http://www.jstor.org/stable/4616215

Kleinbaum, D. G., & Klein, M. (2012). *Survival analysis: a self-learning text* (Vol. 3). New York: Springer.

Korkalainen, H., Töyräs, J., Nikkonen, S., & Leppänen, T. (2019). Mortality-risk-based apnea–hypopnea index thresholds for diagnostics of obstructive sleep apnea. *Journal of sleep research*, *28*(6), e12855.

Koul, H. L., & Deshpande, J. V. (1995). Analysis of Censored Data: Proceedings of the Workshop on Analysis of Censored Data, December 28, 1994-January 1, 1995, University of Pune, Pune, India. IMS.

Lakatos, E. (1988). Sample sizes based on the log-rank statistic in complex clinical trials. *Biometrics*, 229-241.

Lam, B., Sam, K., Mok, W. Y., Cheung, M. T., Fong, D. Y., Lam, J. C., ... & Mary, S. M. (2007). Randomised study of three non-surgical treatments in mild to moderate obstructive sleep apnoea. *Thorax*, *62*(4), 354-359.

Lee, J. M., Weinstein, G. S., O'Malley Jr, B. W., & Thaler, E. R. (2012). Transoral robot-assisted lingual tonsillectomy and uvulopalatopharyngoplasty for obstructive sleep apnea. *Annals of Otology, Rhinology & Laryngology*, *121*(10), 635-639.

Lee, J., Han, Y., Cho, H. H., & Kim, M. R. (2019). Sleep disorders and menopause. Journal of Menopausal Medicine, 25(2), 83-87.

Leung, K. M., Elashoff, R. M., & Afifi, A. A. (1997). Censoring issues in survival analysis. *Annual review of public health*, *18*(1), 83-104.

Li, H. Y., Lee, L. A., Tsai, M. S., Chen, N. H., Chuang, L. P., Fang, T. J., ... & Cheng, W. N. (2020). How to manage continuous positive airway pressure (CPAP) failure–hybrid surgery and integrated treatment. *Auris Nasus Larynx*, *47*(3), 335-342.

Li, H. Y., Wang, P. C., Chen, Y. P., Lee, L. A., Fang, T. J., & Lin, H. C. (2011). Critical appraisal and meta-analysis of nasal surgery for obstructive sleep apnea. *American Journal of Rhinology & Allergy*, *25*(1), 45-49.

Li, W., Xiao, L., & Hu, J. (2013). The comparison of CPAP and oral appliances in treatment of patients with OSA: a systematic review and meta-analysis. *Respiratory care*, *58*(7), 1184-1195.

Liang, K. Y., & Zeger, S. L. (1993). Regression analysis for correlated data. *Annual review of public health*, *14*(1), 43-68.

Lin, H. C., Friedman, M., Chang, H. W., & Gurpinar, B. (2008). The efficacy of multilevel surgery of the upper airway in adults with obstructive sleep apnea/hypopnea syndrome. *The Laryngoscope*, *118*(5), 902-908.

Lin, S. C., & Tunalilar, O. (2022). Rapid adoption of electronic health record and health information exchange among assisted living communities, 2010–2018. *Journal of the American Medical Informatics Association*, *29*(5), 953-957.

Lin, R. S., Lin, J., Roychoudhury, S., Anderson, K. M., Hu, T., Huang, B., ... & Cross-Pharma Non-Proportional Hazards Working Group. (2020). Alternative analysis methods for time to event endpoints under nonproportional hazards: a comparative analysis. *Statistics in Biopharmaceutical Research*, *12*(2), 187-198.

Linz, D., Baumert, M., Desteghe, L., Kadhim, K., Vernooy, K., Kalman, J. M., ... & Lau, D. H. (2019). Nightly sleep apnea severity in patients with atrial fibrillation: potential applications of long-term sleep apnea monitoring. *IJC Heart & Vasculature*, *24*, 100424.

Liu, D., & Zhang, H. (2018). Residuals and diagnostics for ordinal regression models: a surrogate approach. *Journal of the American Statistical Association*, *113*(522), 845-854.
Liu, X. (2012). *Survival analysis: models and applications*. John Wiley & Sons.

Lorenzo-Arribas, A. (2019). *Statistical methods for the analysis of ordinal response data* (Doctoral dissertation, University of Southampton).

Macey, P. M., Woo, M. A., Kumar, R., Cross, R. L., & Harper, R. M. (2010). Relationship between obstructive sleep apnea severity and sleep, depression and anxiety symptoms in newly-diagnosed patients. *PloS one*, *5*(4), e10211.

MacKay, S., Carney, A. S., Catcheside, P. G., Chai-Coetzer, C. L., Chia, M., Cistulli, P. A., ... & McEvoy, R. D. (2020). Effect of multilevel upper airway surgery vs medical management on the apnea-hypopnea index and patient-reported daytime sleepiness among patients with moderate or severe obstructive sleep apnea: the SAMS randomized clinical trial. *Jama*, *324*(12), 1168-1179.

Malay, S., & Chung, K. C. (2012). The choice of controls for providing validity and evidence in clinical research. *Plastic and reconstructive surgery*, *130*(4), 959.

Mandereau-Bruno, L., Leger, D., & Delmas, M. C. (2021). Obstructive sleep apnea: A sharp increase in the prevalence of patients treated with nasal CPAP over the last decade in France. *PLoS one*, *16*(1), e0245392.

Marin-Oto, M., Vicente, E. E., & Marin, J. M. (2019). Long term management of obstructive sleep apnea and its comorbidities. *Multidisciplinary respiratory medicine*, *14*(1), 1-9.

Marin, J. M., Carrizo, S. J., Vicente, E., & Agusti, A. G. (2005). Long-term cardiovascular outcomes in men with obstructive sleep apnoea-hypopnoea with or without treatment with continuous positive airway pressure: an observational study. *The Lancet*, *365*(9464), 1046-1053.

Marklund, M. (2017). Update on oral appliance therapy for OSA. *Current Sleep Medicine Reports*, *3*(3), 143-151.

Martínez-García, M. A., Campos-Rodríguez, F., Catalán-Serra, P., Soler-Cataluña, J. J., Almeida-Gonzalez, C., De la Cruz Morón, I., ... & Montserrat, J. M. (2012). Cardiovascular mortality in obstructive sleep apnea in the elderly: role of long-term continuous positive airway pressure treatment: a prospective observational study. *American journal of respiratory and critical care medicine*, *186*(9), 909-916.

Matiello, M., Nadal, M., Tamborero, D., Berruezo, A., Montserrat, J., Embid, C., ... & Mont, L. (2010). Low efficacy of atrial fibrillation ablation in severe obstructive sleep apnoea patients. *Europace*, *12*(8), 1084-1089.

Mattila, T., Hasala, H., Kreivi, H. R., Avellan-Hietanen, H., Bachour, A., Herse, F., ... & Vasankari, T. (2022). Changes in the societal burden caused by sleep apnoea in Finland from 1996 to 2018: a national registry study. *The Lancet Regional Health-Europe*, *16*, 100338.

May, A. M., Gharibeh, T., Wang, L., Hurley, A., Walia, H., Strohl, K. P., & Mehra, R. (2018). CPAP adherence predictors in a randomized trial of moderate-to-severe OSA enriched with women and minorities. *Chest*, *154*(3), 567-578.

McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society: Series B (Methodological)*, *42*(2), 109-127.

McEvoy, R. Doug, Nick A. Antic, Emma Heeley, Yuanming Luo, Qiong Ou, Xilong Zhang, Olga Mediano et al. (2016). "CPAP for prevention of cardiovascular events in obstructive sleep apnea." *New England Journal of Medicine* 375, no. 10 :919-931.

McNicholas, W. T. (2017). Obstructive sleep apnoea of mild severity: should it be treated?. *Current Opinion in Pulmonary Medicine*, *23*(6), 506-511.

Meira-Machado, L., De Una-Alvarez, J., & Cadarso-Suarez, C. (2006). Nonparametric estimation of transition probabilities in a non-Markov illness–death model. *Lifetime Data Analysis*, *12*(3), 325-344.

Miller Jr, R. G. (2011). *Survival analysis*. John Wiley & Sons.

Minagi, H. O., Okuno, K., Nohara, K., & Sakai, T. (2018). Predictors of side effects with long-term oral appliance therapy for obstructive sleep apnea. *Journal of Clinical Sleep Medicine*, *14*(1), 119-125.

Moeller, J. L., Paskay, L. C., & Gelb, M. L. (2014). Myofunctional therapy: a novel treatment of pediatric sleep-disordered breathing. *Sleep Medicine Clinics*, *9*(2), 235-243.

Moxness, M. H. S., & Nordgård, S. (2014). An observational cohort study of the effects of septoplasty with or without inferior turbinate reduction in patients with obstructive sleep apnea. *BMC Ear, Nose and Throat Disorders*, *14*(1), 1-5.

Mueller, C., Perera, G., Hayes, R. D., Shetty, H., & Stewart, R. (2018). Associations of acetylcholinesterase inhibitor treatment with reduced mortality in Alzheimer's disease: a retrospective survival analysis. *Age and ageing*, *47*(1), 88-94.

Murphy, S. A., Rossini, A. J., & van der Vaart, A. W. (1997). Maximum likelihood estimation in the proportional odds model. *Journal of the American Statistical Association*, *92*(439), 968-976.

Myllymaa, S., Myllymaa, K., Kupari, S., Kulkas, A., Leppänen, T., Tiihonen, P., ... & Töyräs, J. (2015). Effect of different oxygen desaturation threshold levels on hypopnea scoring and classification of severity of sleep apnea. *Sleep and Breathing*, *19*(3), 947-954.

Newman AB, Foster G, Givelber R, Nieto FJ, Redline S, Young T (2005) Progression and regression of sleep-disordered breathing with changes in weight: the sleep heart health study. Arch Intern Med 165:2408–2413

Nishibayashi, M., Miyamoto, M., Miyamoto, T., Suzuki, K., & Hirata, K. (2008). Correlation between severity of obstructive sleep apnea and prevalence of silent cerebrovascular lesions. *Journal of Clinical Sleep Medicine*, *4*(3), 242-247.

Nural, S., Günay, E., Halici, B., Celik, S., & Ünlü, M. (2013). Inflammatory processes and effects of continuous positive airway pressure (CPAP) in overlap syndrome. *Inflammation*, *36*(1), 66-74.

O'Connell, A. A., & Liu, X. (2011). Model diagnostics for proportional and partial proportional odds models. *Journal of Modern Applied Statistical Methods*, *10*(1), 15.

Okuno, K., Pliska, B. T., Hamoda, M., Lowe, A. A., & Almeida, F. R. (2016). Prediction of oral appliance treatment outcomes in obstructive sleep apnea: a systematic review. *Sleep medicine reviews*, *30*, 25-33.

Ong, C. W., O'Driscoll, D. M., Truby, H., Naughton, M. T., & Hamilton, G. S. (2013). The reciprocal interaction between obesity and obstructive sleep apnoea. *Sleep medicine reviews*, *17*(2), 123-131.

Padma, A., Ramakrishnan, N., & Narayanan, V. (2007). Management of obstructive sleep apnea: A dental perspective. *Indian Journal of dental research*, *18*(4), 201.

Pépin, J. L., Eastwood, P., & Eckert, D. J. (2022). Novel avenues to approach non-CPAP therapy and implement comprehensive obstructive sleep apnoea care. *European Respiratory Journal*, *59*(6).

Peppard, P. E., Young, T., Barnet, J. H., Palta, M., Hagen, E. W., & Hla, K. M. (2013). Increased prevalence of sleep-disordered breathing in adults. *American journal of epidemiology*, *177*(9), 1006-1014.

Peromaa-Haavisto, P., Tuomilehto, H., Kössi, J., Virtanen, J., Luostarinen, M., Pihlajamäki, J., ... & Victorzon, M. (2017). Obstructive sleep apnea: the effect of bariatric surgery after 12 months. A prospective multicenter trial. *Sleep medicine*, *35*, 85-90.

Prasad, B., Usmani, S., Steffen, A. D., Van Dongen, H. P., Pack, F. M., Strakovsky, I., ... & Weaver, T. E. (2016). Short-term variability in apnea-hypopnea index during extended home portable monitoring. *Journal of Clinical Sleep Medicine*, *12*(6), 855-863.

Punjabi, N. M. (2008). The epidemiology of adult obstructive sleep apnea. *Proceedings of the American Thoracic Society*, *5*(2), 136-143.

Rao, A., Tey, B. H., Ramalingam, G., & Poh, A. G. (2009). Obstructive sleep apnoea (OSA) patterns in bariatric surgical practice and response of OSA to weight loss after laparoscopic adjustable gastric banding (LAGB). *Ann Acad Med Singapore*, *38*(7), 587-7.

Rasheid, S., Banasiak, M., Gallagher, S. F., Lipska, A., Kaba, S., Ventimiglia, D., ... & Murr, M. M. (2013). Gastric bypass is an effective treatment for obstructive sleep apnea in patients with clinically significant obesity. *Obesity surgery*, *13*(1), 58-61.

ResMed, CPAP-laitteet (CPAP machines), (2022). Accessed on 25.08.2022 in https://shop.resmed.com/FI/fi/cpap-laitteet/c/110

Rossi, F. H., & Tsakadze, N. (2021). Updates in Sleep Neurology and Obstructive Sleep Apnea.

Rotenberg, B. W., Vicini, C., Pang, E. B., & Pang, K. P. (2016). Reconsidering first-line treatment for obstructive sleep apnea: a systematic review of the literature. *Journal of Otolaryngology-Head & Neck Surgery*, *45*(1), 1-9.

Ruehland, W. R., Rochford, P. D., O'Donoghue, F. J., Pierce, R. J., Singh, P., & Thornton, A. T. (2009). The new AASM criteria for scoring hypopneas: impact on the apnea hypopnea index. *sleep*, *32*(2), 150-157.

Ryan, S., Taylor, C. T., & McNicholas, W. T. (2005). Selective activation of inflammatory pathways by intermittent hypoxia in obstructive sleep apnea syndrome. *Circulation*, *112*(17), 2660-2667.

Ryan, C. F. (2005). Sleep· 9: An approach to treatment of obstructive sleep apnoea/hypopnoea syndrome including upper airway surgery. *Thorax*, *60*(7), 595-604.

Sahib, A., Roy, B., Kang, D., Aysola, R. S., Wen, E., & Kumar, R. (2022). Relationships between brain tissue damage, oxygen desaturation, and disease severity in obstructive sleep apnea evaluated by diffusion tensor imaging. *Journal of Clinical Sleep Medicine*, jcsm-10192.

Sahib, A., Roy, B., Kang, D., Aysola, R. S., Wen, E., & Kumar, R. (2022). Relationships between brain tissue damage, oxygen desaturation, and disease severity in obstructive sleep apnea evaluated by diffusion tensor imaging. *Journal of Clinical Sleep Medicine*, jcsm-10192.

Sahlman J, Pukkila M, Seppä J, Tuomilehto H. (2007) Evolution of mild obstructive sleep apnoea after different treatments. *Laryngoscope* 2007;**117**:1107–1111.

Sainani, K. L. (2015). Dealing with missing data. *PM&R*, *7*(9), 990-994.

Sarkhosh, K., Switzer, N. J., El-Hadi, M., Birch, D. W., Shi, X., & Karmali, S. (2013). The impact of bariatric surgery on obstructive sleep apnea: a systematic review. *Obesity surgery*, *23*(3), 414-423.

Schwartz A, Patil S, Laffan A, Polotsky V, Schneider H, Smith P (2008) Obesity and obstructive sleep apnea. Pathogenic mechanisms and therapeutic approaches. Proc Am Thorac Soc 5:185–192

Shapiro, G. K., & Shapiro, C. M. (2010). Factors that influence CPAP adherence: an overview. *Sleep and Breathing*, *14*(4), 323-335.

Sher, A. E. (2002). Upper airway surgery for obstructive sleep apnea. *Sleep medicine reviews*, *6*(3), 195-212.

Sleep Apnea (obstructive sleep apnea in adults). The Finnish Current Care Guidelines. Working group set up by the Finnish Medical Society Duodecim and the Finnish Cardiac Society. Helsinki: The Finnish Medical Society Duodecim, 2017 (referred 3/12/2022). Available online at: www.kaypahoito.fi

Spicuzza, L., Caruso, D., & Di Maria, G. (2015). Obstructive sleep apnoea syndrome and its management. *Therapeutic advances in chronic disease*, *6*(5), 273-285.

Sutherland, K., Phillips, C. L., & Cistulli, P. A. (2015). Efficacy versus effectiveness in the treatment of obstructive sleep apnea: CPAP and oral appliances. *J Dent Sleep Med*, *2*(4), 175-181.

Sutherland, K., Vanderveken, O. M., Tsuda, H., Marklund, M., Gagnadoux, F., Kushida, C. A., & Cistulli, P. A. (2014). Oral appliance treatment for obstructive sleep apnea: an update. *Journal of Clinical Sleep Medicine*, *10*(2), 215-227.

Tobaldini, E., Costantino, G., Solbiati, M., Cogliati, C., Kara, T., Nobili, L., & Montano, N. (2017). Sleep, sleep deprivation, autonomic nervous system and cardiovascular diseases. *Neuroscience & Biobehavioral Reviews*, *74*, 321-329.

Tolonen H, Reinikainen J, Zhou Z, Härkänen T, Männistö S, Jousilahti P, Paalanen L, Lundqvist A, Laatikainen T. (2022). Development of non-communicable disease risk factors in Finland – projections up-to 2040. Scandinavian Journal of Public Health

Tuomilehto, H., Seppä, J., & Uusitupa, M. (2013). Obesity and obstructive sleep apnea–clinical significance of weight loss. *Sleep medicine reviews*, *17*(5), 321-329.

Twisk, J. W. (2013). *Applied longitudinal data analysis for epidemiology: a practical guide*. cambridge university press.

Vakulin, A., Yap, C., McEvoy, R. D., & Chai-Coetzer, C. L. (2017). Obstructive sleep apnoea: Therapeutic options and challenges. *Clinical Medicine Insights: Therapeutics*, *9*, 1179559X17711932.

Venema, J. A., Doff, M. H., Joffe-Sokolova, D., Wijkstra, P. J., van der Hoeven, J. H., Stegenga, B., & Hoekema, A. (2020). Long-term obstructive sleep apnea therapy: a 10-year follow-up of mandibular advancement device and continuous positive airway pressure. *Journal of Clinical Sleep Medicine*, *16*(3), 353-359.

Verse, T. (2005). Bariatric surgery for obstructive sleep apnea. *Chest*, *128*(2), 485-487.

Verse, T., & Hörmann, K. (2011). The surgical treatment of sleep-related upper airway obstruction. *Deutsches Ärzteblatt International*, *108*(13), 216.

Vijayan, V. K. (2012). Morbidities associated with obstructive sleep apnea. *Expert review of respiratory medicine*, *6*(5), 557-566.

Wan, L., Lou, W., Abner, E., & Kryscio, R. J. (2016). A comparison of time-homogeneous Markov chain and Markov process multi-state models. *Communications in Statistics: Case Studies, Data Analysis and Applications*, *2*(3-4), 92-100.

Weaver, T. E., & Chasens, E. R. (2007). Continuous positive airway pressure treatment for sleep apnea in older adults. *Sleep medicine reviews*, *11*(2), 99-111.

White, D. P., & Younes, M. K. (2012). Obstructive sleep apnea. *Comprehensive Physiology*, *2*(4), 2541-2594.

White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: issues and guidance for practice. *Statistics in medicine*, *30*(4), 377-399.

Won, C. H. (2020). When will we ditch the AHI? *Journal of Clinical Sleep Medicine*, *16*(7), 1001-1003.

Won, C. H., Li, K. K., & Guilleminault, C. (2008). Surgical treatment of obstructive sleep apnea: upper airway and maxillomandibular surgery. *Proceedings of the American Thoracic Society*, *5*(2), 193-199.

Yang, S., & Prentice, R. (2010). Improved logrank-type tests for survival data using adaptive weights. *Biometrics*, *66*(1), 30-38.

Yao, L., Li, S., Li, Y., Huai, M., Gao, J., & Zhang, A. (2018). Representation learning for treatment effect estimation from observational data. *Advances in Neural Information Processing Systems*, *31*.

Young T, Finn L, Austin D, Peterson A. (2003). Menopausal status, and sleep-disordered breathing in the Wisconsin Sleep Cohort Study. Am J Respir Crit Care Med. 167:1181–1185.

Yu, X., Fujimoto, K., Urushibata, K., Matsuzawa, Y., & Kubo, K. (2003). Cephalometric analysis in obese and nonobese patients with obstructive sleep apnea syndrome. *Chest*, *124*(1), 212-218.

Yuan, X., Fang, J., Wang, L., Yao, L., Li, L., Zhan, X., ... & Wei, Y. (2015). Adequate continuous positive airway pressure therapy reduces mortality in Chinese patients with obstructive sleep apnea. *Sleep and Breathing*, *19*(3), 911-920.

Zhang, Y. F., Zhang, Q. F., & Yu, R. H. (2010). Markov property of Markov chains and its test. In *2010 International Conference on Machine Learning and Cybernetics* (Vol. 4, pp. 1864-1867). IEEE.

Zhang, Z., & Sun, J. (2010). Interval censoring. *Statistical methods in medical research*, *19*(1), 53-70.

Zhao, Q., & Sun, J. (2004). Generalized log-rank test for mixed interval-censored failure time data. *Statistics in medicine*, *23*(10), 1621-1629.

Zhu, Y., Long, H., Jian, F., Lin, J., Zhu, J., Gao, M., & Lai, W. (2015). The effectiveness of oral appliances for obstructive sleep apnea syndrome: A meta-analysis. *Journal of dentistry*, *43*(12), 1394-1402.