UNIVERSITY OF PADUA  FACULTY OF ENGINEERING

Department of Information Engineering

Ph.D. School in Information Engineering
Section of Bioengineering

23rd cycle

A MIXED-EFFECT MULTINOMIAL MARKOV-CHAIN MODEL FOR DESCRIBING SLEEP ARCHITECTURE IN INSOMNIAC PATIENTS

School Head: Prof. Matteo Bertocco, Ph.D.

Supervisor: Prof. Claudio Cobelli, Ph.D.

PhD Student: Roberto Bizzotto

A papà

Table of contents

Preface i
Summary iii
Sommario v
Abbreviations vii
1 Introduction 1
2 Fundamentals of sleep and insomnia 5
2.1 Sleep 6
2.1.1 Sleep stages and architecture 7
2.1.2 Neuroanatomy and regulation of sleep 10
2.2 Sleep disorders 13
2.2.1 Classification 13
2.2.2 Insomnia prevalence and comorbities 14
2.2.3 Insomnia treatment 15
2.3 Conclusions 18
2.4 Figures 19
3 Mixed-effect PK-PD modeling and categorical data modeling 23
3.1 Mixed-effect models 26
3.1.1 Non-linear mixed-effect modeling: theory 27
3.1.2 Non-linear mixed-effect modeling: NONMEM 30
3.1.3 Laplace approximation 31
3.1.4 Model selection 32
3.2 Categorical data modeling 34
3.2.1 Markov-chain models 36
3.2.2 Basic theory on Markov chains 38
3.2.3 Link functions 40
3.3 Conclusions 43
3.4 Tables and figures 44
4 Multinomial logistic functions in a Markov-chain model of sleep architecture 47

Table of contents

4.1 Materials 49
4.2 A new Markov-chain model 50
4.2.1 Methods 50
4.2.2 Results 55
4.2.3 Discussion 57
4.3 The final Markov-chain model 59
4.3.1 Methods 59
4.3.2 Results 60
4.3.3 Discussion 63
4.4 Tables and figures 66
5 Model evaluation 83
5.1 Internal evaluation 85
5.1.1 Simulation-based methods 87
5.1.2 Results 89
5.1.3 Discussion 90
5.2 External evaluation 92
5.2.1 Materials 92
5.2.2 Methods 93
5.2.3 Results 93
5.2.4 Discussion 94
5.3 Tables and figures 96
6 Covariate analysis 107
6.1 Methods 109
6.2 Results 112
6.3 Discussion 113
6.4 Tables and figures 115
7 Conclusions 117
Appendix A 121
Appendix B 125
Appendix C 127
References 135

Preface

The work described in this thesis has been done in collaboration with the Clinical Pharmacology & Modelling and Simulation Group - Drug Discovery in Neuroscience – GlaxoSmithKline (GSK), with the coordination by Stefano Zamuner, Ph.D and Roberto Gomeni, PhD.

Moreover, I spent six months during my Ph.D. in Uppsala University - Department of Pharmaceutical Biosciences, with the coordination by Prof. Mats O. Karlsson.

Great help came also from Prof. Giuseppe De Nicolao, Prof. Andrew C. Hooker and Enrica Mezzalana.

Thanks to all of them.

Summary

The decrease of sleep quality highly compromises the physical and mental well-being of the human body and is a common disorder affecting a large segment of world's population. The quality of nocturnal sleep is determined by its internal structure, i.e. the pattern through different physiological conditions of the body during the night. This structure, called 'sleep architecture', can be expressed by different stages (awake, stage 1 and 2 of light sleep, slow-wave sleep and REM sleep) and the transitions between them during nighttime, and can be modified through drugs for insomnia treatment. The sequence of sleep stages, objectively assessed every 30 seconds through polysomnography (PSG), constitute the so-called 'PSG signal' and can be seen as a finite succession of categorical, nominal (i.e., non-ordered) data.

In the context of modeling of drug effects (pharmacodynamic, PD) and correlations between drug exposure (pharmacokinetics, PK) and drug effect, the analysis of categorical nominal polychotomous data has been explored only recently. The interest on the subject has strongly grown, also because many other pharmacodynamic data provided by clinical studies during drug development share the same characteristics. PK-PD modelling of categorical data requires specific methodologies. When dealing with nominal polychotomous data, the most interesting approach is the use of mixed-effect non-homogeneous Markov-chain models, whose parameters are related to the evolution of the probabilities of transitioning between different states of the chain for increasing values of the independent variable. Despite their relevance, such models present many aspects which have only been partially investigated in the literature so far.

This thesis is dedicated to the introduction of multinomial logistic functions as link functions for describing transition probabilities in the Markov-chains with more than two states. Binary logistic functions have previously been used, instead. A new model for sleep architecture is therefore implemented and evaluated, using PSG placebo data obtained from a clinical study in patients affected by primary insomnia. Parameter estimation is accomplished through maximization of Laplace-approximated likelihood, using NONMEM VI. Model evaluation is performed through standard techniques, like inspection of goodness-of-fit plots, bootstrap and simplified posterior predictive check.

Later on in the thesis, the new multinomial Markov-chain model is further developed by combining the strengths of other existing models, and by adding additional components. The major investigated features are the predictors of the multinomial logits, the model parameterization, the relevance of the vari-

ous stages and transitions, and the description of the inter-individual variability.

The final model is internally evaluated through simplified posterior predictive check and two other diagnostics based on Monte Carlo simulation: visual predictive check, implemented not only on stage frequencies (as done in the literature) but also on transition frequencies along the night; and visual estimation check, introduced here for the first time in the context of PK-PD mixed-effect modeling. This tool aims to evaluating the capability of accurately and precisely estimating model parameters through a graphic description of accuracy and precision on the estimation of transition probabilities time-course. The three diagnostics show an overall good performance of the developed multinomial Markov-chain model in describing and reproducing the data, and of the employed estimation technique in producing robust estimates of the model parameters.

The final model is also externally evaluated, using data from a new clinical study in patients with the same conditions as the original study. The evaluation is mainly performed looking at minimized objective function values and at new simplified posterior predictive checks. The new proposed model is shown to adequately describe also the new data.

In the last part of the thesis, stepwise covariate modeling is adopted for investigating the appropriate structural form of a second stage model in which age, body mass index and gender effects are integrated in the base model. The statistical relevance of these covariate effects is computed on the original insomniac population, together with the entity of the effects themselves. Interesting and novel results are shown, depicting how each of the three covariates affects some of the transition probabilities of the multinomial Markov-chain model, during specific nighttime intervals.

Sommario

La riduzione della qualità del sonno compromette considerevolmente il benessere psico-fisico del corpo umano ed è un disturbo comune ad un'ampia fetta della popolazione mondiale. La qualità del sonno notturno è determinata dalla sua struttura interna, ossia il percorso attraverso diverse condizioni fisiologiche dell'individuo nella notte. Tale struttura, chiamata 'architettura del sonno', può essere espressa attraverso diversi stadi (veglia, stadi 1 e 2 del sonno leggero, sonno profondo e sonno REM) e le transizioni tra di essi durante la notte, e può essere modificata da farmaci preposti al trattamento dell'insonnia. La sequenza di stadi del sonno, oggettivamente determinabili ogni 30 secondi attraverso polisonnografia, costituisce il cosiddetto 'segnale PSG' e può essere vista come una successione finita di dati categoriali nominali (cioè, non ordinati).

Nel contesto della modellistica degli effetti farmacologici (farmacodinamica, PD) e delle correlazioni tra esposizione al farmaco (farmacocinetica, PK) ed effetti farmacologici, l'analisi di dati categoriali nominali policotomi è stata esplorata solo recentemente. L'interesse su di essa è cresciuto fortemente, anche perchè molti altri dati farmacodinamici provenienti da studi clinici fatti per sviluppare nuovi farmaci condividono le stesse caratteristiche. La modellistica PK-PD di dati categoriali richiede metodologie specifiche. Quando i dati sono nominali policotomi, l'approccio più interessante è l'utilizzo di modelli ad effetti misti a catena di Markov non omogenea, i cui parametri sono legati all'evolversi delle probabilità di transizione tra i diversi stati della catena al variare della variabile indipendente. Questi modelli però hanno presentato fin'ora molti aspetti specifici che in letteratura sono stati sviscerati solo parzialmente.

Questa tesi è dedicata all'introduzione delle funzioni logistiche multinomiali come funzioni link in grado di descrivere le probabilità di transizione nelle catene di Markov con più di due stati. Precedentemente invece, erano state utilizzate funzioni logistiche binarie. Nella tesi viene quindi implementato e validato un nuovo modello dell'architettura del sonno, facendo uso di dati PSG ottenuti da uno studio clinico in soggetti con diagnosi di insonnia primaria, ai quali era stato somministrato placebo. La stima parametrica viene effettuata tramite massimizzazione della verosomiglianza approssimata con metodo di Laplace, utilizzando NONMEM VI. La validazione del modello avviene tramite tecniche consolidate, come l'ispezione dei goodness-of-fit plot, il bootstrap ed il simplified posterior predicitve check.

Nel prosieguo della tesi il nuovo modello multinomiale a catena di Markov viene ulteriormente sviluppato fondendo i punti di forza di altri modelli

esistenti nell'ambito considerato, ed aggiungendo nuovi elementi. I predittori delle funzioni logit multinomiali, la parametrizzazione del modello, la rilevanza di transizioni e stadi diversi, la descrizione della variabilità interindividuale sono i principali ambiti di analisi.

Il modello finale viene validato internamente attraverso simplified posterior predictive check e due altri metodi diagnostici basati su simulazione Monte Carlo: il visual predictive check, implementato non solo sulle frequenze degli stadi (come fatto in letteratura), ma anche sulle frequenze delle transizioni nel corso della notte; ed il visual estimation check, introdotto qui per la prima volta nel contesto della modellistica PK-PD ad effetti misti. Questo strumento ha l'obiettivo di validare la capacità di stimare i parametri del modello in modo preciso ed accurato attraverso una descrizione grafica dell'accuratezza e della precisione nella stima delle probabilità di transizione nel corso della notte. I tre strumenti diagnostici mostrano una buona performance nel descrivere e riprodurre i dati, per quanto riguarda il modello multinomiale a catena di Markov sviluppato, e nel produrre stime robuste dei parametri del modello, per quanto concerne il metodo di stima adottato.

Il modello finale viene validato anche esternamente, con dati ottenuti da un nuovo studio clinico in pazienti con condizioni uguali a quelle dei pazienti dello studio originale. La validazione viene effettuata valutando principalmente i valori minimizzati della funzione obiettivo ed i nuovi simplified posterior predictive check. I suoi risultati mostrano che il nuovo modello proposto è in grado di descrivere adeguatamente anche i nuovi dati.

Nell'ultima parte della tesi, il processo di stepwise covariate modeling viene usato allo scopo di scegliere la forma strutturale appropriata di un modello del secondo stadio in cui gli effetti di età, indice di massa corporea e sesso vengono integrati nel modello base. La significatività statistica di tali effetti viene quindi calcolata sulla popolazione originaria di soggetti insonni, insieme all'entità degli effetti stessi. Gli innovativi ed interessanti risultati di questa analisi mostrano come ciascuna covariata influenza alcune probabilità di transizione del modello multinomiale a catena di Markov, durante specifici intervalli della notte.

Abbreviations

AW AWake
BIC Bayesian Information Criterion
BMI Body Mass Index
BSV Between-Subject Variability
CV Coefficient of Variation
dof degrees of freedom
EBE Empirical Bayes Estimate
IIV Inter-Individual Variability
LLR Log-Likelihood Ratio test
LPS Latency to Persistent Sleep
ML Maximum Likelihood
NLME Non-Linear Mixed Effects
OFV Objective Function Value
PK Pharmacokinetics
PD Pharmacodynamics
PSG Polysomnography
REM Rapid Eyes Movement sleep
RUV Residual Unexplained Variability
sPPC simplified Posterior Predictive Check
ST1 STage 1 sleep
ST2 STage 2 sleep
SWS Slow-Wave Sleep
TST Total Sleep Time
VEC Visual Estimation Check
VPC Visual Predictive Check
WASO Wake After Sleep Onset

Chapter 1

Introduction

Sleep is the natural state of bodily rest. Sleep disorders affect a large portion of world-wide population –prevalence is thought to be approximately 10% (Mai & Buysse, 2008)– and their effects are far-reaching: medical, psychiatric, personal and societal spheres are all substantially involved. Among other things, sleep pathologies affect quality of life because of comorbid conditions and impaired interpersonal relationships (Mai & Buysse, 2008).

The appropriate diagnosis and treatment of these disorders still represent great challenges for clinicians and pharmaceutical companies. The latter have made large investments in this research field, trying to develop safer and more effective drugs able to regulate the sleep-wake alternation. However, sleep is not a homogeneous state of unconsciousness, but it is characterized by an internal structure, called 'sleep architecture', defined by different stages and the transitions between them. Such architecture determines the sleep quality and, therefore, the physical and mental well-being and performance of the human being(Roth & Reehrs, 2003; Walsh, 2004; Avidan, 2003). Therefore, in case of sleep disturbances, it is of particular importance that the therapeutic interventions maintain or restore the physiological sleep internal structure (Gimenez, Clos, Romero, Grasa, Morte, & Barbanoj, 2007; Penzel & Kesper, 2006; Stanley, 2005; Burgess, Holmes, & Dawson, 2001).

Sleep stages are typically assessed using polysomnography, which includes electroencephalography (assessment of brain activity), electrooculography (assessment of eye movements) and electromyographic measurements (determination of skeletal muscle activity). Based on the combination of these techniques, different stages can be recognized: wake, light sleep (stage 1 and 2), deep sleep (stage 3 and 4), and Rapid Eyes Movement (REM) sleep (Rechtschaffen & Kales, 1968).

The times spent in the different phases are typically aggregated over nighttime, resulting in pharmacodynamic endpoints with recognized clinical relevance, such as sleep onset (Latency to Persistent Sleep, LPS), sleep maintenance (Wake After Sleep Onset, WASO), and Total Sleep Time (TST) (Roth, Walsh, Krystal, Wessel, & Roehrs, 2005; Erman, Seiden, Zammit, Sainati, & Zhang, 2006). These metrics, however, do not preserve information on the sleep internal structure. Mathematical models are needed, instead, for characterizing sleep architecture and, since the sleep stages during nighttime can be

described as a sequence of correlated discrete random variables, Markov-chain models represent a valuable methodology to analyze these data.

In two recent works (Karlsson, et al., 2000; Kjellsson, Ouellet, Corrigan, & Karlsson, 2008) sleep data were modeled through mixed-effect Markov-chain models, and the transition probabilities were described as binary logistic piecewise linear functions. The choice of binary logistic functions as links between transition probabilities and model parameters brings to a long and elaborate model building process and resorts to a parameterization which is not completely physiological and robust. Instead, multinomial logistic functions are more suited to the nature of the described data and can simplify the model structure.

Therefore, the first aim of this work is the implementation of a mixed-effect Markov-chain model with multinomial logistic functions for describing sleep data obtained in patients suffering from primary insomnia and treated with placebo. More detailed features on the model parameterization, the predictors of the parameters, the description of the inter-individual variability on model parameters, the model structure able to ease the later inclusion of potential covariate and drug effects will also be investigated.

For a mixed-effect model used in the pharmaceutical field, the evaluation of adequacy to data is very important (Karlsson & Savic, 2007). For example, it is relevant to evaluate the reliability of the subsequent clinical trial simulations performed with the developed model itself. Literature is still lacking in assessed methodologies for a complete model evaluation when categorical (nominal) data are involved. The second aim of this thesis is to introduce, implement and discuss consolidated and innovative techniques for evaluating the proposed Markov-chain model, by means of the learning dataset, datasets created via Monte Carlo simulation and a further set of real data, used as validation dataset.

Despite the frequency with which sleep stage data are used in the evaluation of patients with sleep disorders or daytime sleepiness, only few times the heterogeneity of sleep architecture among individuals has been described. Most previous studies have included small samples, representing a limited range of demographic conditions (Redline, Kirchner, Quan, Gottlieb, Kapur, & Newman, 2004). Published studies with large sample size, instead, never investigated the effect of covariates like gender, weight, race, etc., on the internal structure of sleep, but only on aggregated pharmacodynamics endpoints which contains just part of the information available. With the model illustrated here, this analysis becomes instead possible. One further aim of this work is therefore to perform a covariate analysis on the developed multinomial Markov-chain model, using gender, body mass index (BMI) and age as potential covariates. Besides the clinical importance of the results, also the identification of the appropriate structural form of a second stage model for defining the covariate effects will deserve attention, considering the specificity of the modeled data.

The thesis is organized as follows. Chapter 2 and Chapter 3 provide some background to the topics of this work, i.e., the characterization of sleep and the use of PK-PD models for describing categorical data. Chapter 4 describes the Markov-chain model with multinomial functions proposed here to model sleep data. A base model and its development are discussed in sequence, and the results of the application to real data from a clinical study with insomniac subjects are shown. Chapter 5 presents internal and external evaluation of the final model, through partially new diagnostics. Finally, a second-stage model taking into account the effects of age, sex and BMI on the parameters is developed and estimated in Chapter 6.

Chapter 2

Fundamentals of sleep and insomnia

The cyclic repetition of sleep and wakefulness states is essential to the basic functioning of all higher animals, including humans. As understanding of the neurobiology of sleep increases, clinicians no longer view it as a passive state (i.e., as absence of wakefulness). Sleep is an active neurobehavioral state that is maintained through a highly organized interaction of neurons and neural circuits in the central nervous system (CNS).

Moreover, sleep is not a homogeneous state of unconsciousness but it is characterized by an internal structure, called 'sleep architecture' and defined by different stages and the transitions between them. It has been demonstrated that the maintenance of such architecture is fundamental to determine sleep quality and, therefore, the physical and mental well-being. On the other hand, disorders in the natural pattern of sleep may lead to adverse consequences and may seriously affect patients' health, productivity and life quality (Roth & Reehrs, 2003; Walsh, 2004; Avidan, 2003).

The prevalence of sleep disturbances indicates that it is a very common problem affecting both men and women, elderly and young population. The causes of sleep problems are different and often are related to other clinical pathologies. For these reasons, the appropriate diagnosis and the treatment of sleep disorders is becoming more and more relevant. However, the mechanisms regulating sleep and their purposes are not completely clear and the deep understanding of sleep architecture and patterns represents a great challenge to clinicians and pharmaceutical companies, which are leading an intense research in this field.

This chapter aims, first of all, at providing the basics of physiology, neuroanatomy and regulation of sleep. Then, it introduces the physiological reasons for which we sleep, the consequences of sleep disturbances and the available pharmaceutical treatments to manage them.

2.1 Sleep

The first attempt to describe the pattern of human sleep was made in 1930 by Berger, the father of electroencephalography (Berger, 1930). He obtained the first sleep recording and noted that the alpha rhythm disappeared when his subject fell asleep. A second important achievement was made in 1937 when Loomis et al. (Loomis, Harvey, & Hobart, 1937) published the first continuous overnight EEG sleep recording in humans and proposed a scheme, the so called 'sleep staging', to summarize the EEG recording in a reduced dataset. They proposed a classification of the EEG activity recorded during sleep into 5 stages: A, B, C, D and E, on the basis of the predominant EEG rhythm in a fixed time interval.

In 1953 Aserinsky and Kleitmain (Aserinsky & Kleitman, 1953) discovered episodic electro-oculagraphic (EOG) activity occurring during sleep stage B every 90-120 minutes. Initially this activity was supposed to be an artifact due to instrumentation, but subsequent studies demonstrated that these episodes were actually occurring. These events were called 'Rapid Eye Movements' (REMs).

The authors tried also to establish the relation existing between REM stage and dreaming. It was noted that dreaming happened in 20 of the 27 instances after the awakening from REM sleep stage. In 1957 Dement and Kleitman (Dement & Kleitman, 1957) suggested a new classification for sleep stages: sleep stages were divided into four Non-REM (NREM) stages and a REM stage.

The next major milestone on understanding sleep architecture was made in 1959, when Jouvet (Jouvet, Michel, & Courjon, 1959) observed by the electromyography (EMG) technique muscular atonia related to the REM stage. He also introduced the concept that REM stage was a state in which the brain was 'active'.

The staging criteria were standardized in 1968, when Rechtschaffen and Kales developed and published 'A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects' (Rechtschaffen & Kales, 1968), establishing the major rules for classifying sleep stages during a standardized sleep recording. This manual, generally reported as the 'R & K Manual', received a general consensus and became a gold standard in sleep measurements. In the 'R & K Manual', NREM sleep was divided into four stages (stages 1, 2, 3, 4), with slow-wave sleep or deep sleep comprising stages 3 and 4 and, in contrast, light-sleep comprising stages 1 and 2. REM sleep was sometimes reported as stage 5.

In 2004, the American Academy of Sleep Medicine (AASM) proposed several changes in the scoring system indicated by the 'R & K standard', the most

significant being the combination of stages 3 and 4 into a unique stage, called Stage N3. These proposed changes were published in 2007 in 'The AASM Manual for the Scoring of Sleep and Associated Events' (Iber, Ancoli-Israel, Chesson, & Quan, 2007).

2.1.1 Sleep stages and architecture

Sleep physicians define human sleep on the basis of an individual's observed behavior and accompanying physiologic changes in the brain's electrical activity as the brain transitions between wakefulness and sleep. Behaviorally, human sleep is characterized by reclined position, closed eyes, decreased movement, and decreased responsivity to internal and external environment.

Polysomnography

Physiologically, sleep consists of two strikingly different states, rapid-eye-movement (REM) sleep and non-REM (NREM) sleep. NREM sleep can be subdivided further into four stages. Polysomnography is the 'gold standard' technique that simultaneously records the three physiologic measures that define the main stages of sleep and wakefulness. These measures include muscle tone, recorded through electromyogram (EMG); eye movements, recorded through electro-oculogram (EOG); and brain activity, recorded through electroencephalogram (EEG). The clinical polysomnogram, the purpose of which is to detect findings that are characteristic of certain sleep disorders, includes, in addition to these three variables, the following: monitors for airflow at the nose and mouth, respiratory effort strain gauges placed around the chest and abdomen, and noninvasive oxygen saturation monitors that function by introducing a beam of light through the skin. Other parameters include the electrocardiogram and EMG of the anterior tibialis muscles, which are intended to detect periodic leg movements. Finally, a patient's gross body movements are monitored continuously by audiovisual means.

Stage classification

To classify sleep, recommendations have been introduced in 1968 by a committee chaired by Rechtschaffen and Kales (Rechtschaffen & Kales, 1968). They introduced discrete sleep stages based on the observed EEG waves and patterns as well as EOG patterns and mental or submental muscle tone de-

rived by EMG (see Figure 2.1 and Figure 2.2). The EEG, EOG, and EMG electrodes are attached at well standardized positions. Sleep recording is classified into epochs of 30 s duration.

The EEG pattern of drowsy wakefulness consists of low-voltage rhythmic alpha activity (8–13 Hz). In stage 1 of NREM sleep, the low-voltage mixed frequency theta waves (4–8 Hz) replace alpha rhythm of wakefulness. Slow asynchronous eye movements are seen on the EOG in the beginning of stage 1 sleep and disappear in a few minutes. The muscle activity is highest during wakefulness and diminishes as sleep approaches. Individuals with behavioral characteristics of sleep and polysomnographic characteristics of stage 1 sleep may or may not perceive themselves as sleeping. Stage 1 is viewed as a 'shallow' sleep, during which an individual can be easily aroused.

With transition to stage 2, EEG patterns called 'sleep spindles' and 'K complexes' appear on the EEG. Sleep spindles are 12-14 Hz synchronized EEG waveforms with duration of 1.5 seconds. Sleep spindle waves arise as a result of synchronization of groups of thalamic neurons by the GABAergic thalamic spindle pacemaker. The origin of K complexes is unknown. With the onset of stage 2, the arousal threshold increases, and a more intense stimulus is needed to arouse a sleeper.

Stages 3 and 4 of NREM sleep are defined by synchronized high-amplitude (>75 μV) and slow (0.5–2 Hz) delta wave EEG pattern. Stages 3 and 4 collectively are referred as 'deep sleep', 'delta sleep', or 'slow-wave sleep'. By definition, delta waves account for 20% to 50% of EEG activity during stage 3 and greater than 50% of EEG activity during stage 4 of sleep. Slow-wave sleep is associated with a higher arousal threshold than 'lighter' stages of NREM sleep. No eye movements are detected on the EOG during stages 2, 3, and 4 of NREM sleep. The EMG tracks continued muscle tone decline as NREM sleep 'deepens' from stages 1 to 4.

The cortical EEG pattern of REM sleep is characterized by low voltage and fast frequencies (alpha or 8–13 Hz). This EEG pattern is referred as activated or desynchronized and also is seen in the state of relaxed wakefulness (with eyes closed). Activated refers to an active mind (dreams) and the EEG pattern characteristic of wakefulness. Paradoxically, individuals in REM sleep, although activated, are behaviorally less responsive than during the wake state (Siegel, 2005). Desynchronized refers to the random-appearing wave pattern seen on the REM sleep EEG, which is contrasted with the synchronized uniform wave pattern seen on the NREM sleep EEG (Siegel, 2005). To be scored as REM sleep, a polysomnographic tracing must contain an activated EEG pattern and muscle atonia (EMG) and the presence of rapid eye movements (EOG). REM sleep can be subdivided further into two stages: tonic and phasic. The tonic stage is continuous and shows muscle atonia and desynchronized EEG as two main features. Superimposed on the tonic stage of REM are intermittent phasic events, which include bursts of rapid eye movements and irregularities of respiration and heart rate.

Sleep architecture

The sequence of sleep stages across the night is called 'sleep architecture'. Figure 2.3 depicts how such internal structure looks like in a healthy person. Sleep typically begins with a 'shallow' stage 1 of NREM sleep and 'deepens' to NREM sleep stages 2, 3, and 4, which are followed by the first brief episode of REM sleep in approximately 90 minutes. After the first sleep cycle, NREM and REM sleep continue to alternate in a predictable fashion, each NREM-REM cycle lasting approximately 90 to 120 minutes (Sinton & McCarley, 2004). In the course of the night, sleep cycles recur three to seven times.

Stage 1 of NREM sleep, which lasts only a few minutes, serves as a transition from wakefulness to sleep and later during sleep it serves as a transition between REM-NREM sleep cycles. Typically, stage 1 constitutes 2% to 5% of total sleep time. An increase in the amount or percentage of stage 1 sleep may be a sign of sleep disruption.

The brief first period of stage 1 NREM sleep is followed by 'deeper' stage 2, which lasts approximately 10 to 20 minutes. Stage 2 sleep normally constitutes 45% to 55% of the total sleep time.

Stage 2 sleep progresses to stages 3 (lasting a few minutes) and 4 (lasting 40 minutes). Stage 3 constitutes 5% to 8% of the total sleep time, and stage 4 constitutes 10% to 15% of the total sleep time. Stages 3 and 4 of NREM sleep predominate during the first third of the night.

The first REM period is brief and occurs approximately 90 minutes after sleep onset; subsequent REM cycles occur approximately 90 to 120 minutes apart. REM sleep episodes become longer as the night progresses, and the longest REM periods are found in the last third of the night (Carskadon & Dement, 2005). NREM sleep accounts for 75% to 80% and REM sleep accounts for 20% to 25% of the total sleep time (Sinton & McCarley, 2004; Siegel, 2005). These proportions commonly vary with age.

Key summary statistics of sleep

Some important characteristics of sleep can be derived from plysomnographic data. Such characteristics are used in the clinical practice for quantifying both the severity of sleep disorders and hypnotics efficiency. These parameters are called 'aggregated parameters' because they reflect the overall trend of sleep during the night. The following definitions of sleep parameters are listed for reference:

- total recording time (TRT): the duration of time from the start to the end of a recording, usually 8 hours;
- time in bed (TIB): the duration of time from 'light off' to final awakening;

- sleep onset (SO): the first epoch followed by 19 epochs of 'non awake' stages;
- sleep period time (SPT): the duration of time from SO to final awakening;
- latency to persistent sleep (LPS): the duration of time from 'light off' to SO;
- wake after sleep onset (WASO): the total time spent awake during SPT;
- total sleep time (TST): the amount of actual sleep time in a recording;
- sleep efficiency (SE): the percentage ratio of total sleep time to time in bed, i.e. TST/TIB*100;
- time spent in each of the sleep stages (tAW, tST1, tST2, tST3, tST4, tREM);
- number of transition to each stage (nAW, nST1, nST2, nST3, nST4, nREM);
- mean extension of each stage (meanAW, meanST1, meanST2, meanST3, meanST4, meanREM).

2.1.2 Neuroanatomy and regulation of sleep

All human physiological functions are embedded in the circadian day-night cycle. The sleep-wake cycle is closely linked to the circadian cycle and both influence each other. The circadian system is well described and an anatomical circadian clock has been identified: the suprachiasmatic nucleus in the anterior hypothalamus controls the timing of most circadian rhythms in mammals. In contrast, no single neural system identified so far is responsible for the generation of sleep or wakefulness (Jones, Basic mechanisms of sleep-wake states, 2000).

Most neuroanatomical aspects of sleep are obtained by the observation of the effects of lesions in particular regions. Wakefulness is maintained by multiple neural systems that extend from the brainstem reticular formation into the thalamus and through the posterior hypothalamus up to the basal forebrain. Sleep is promoted by neurons in the lower brainstem and upper forebrain that inhibit wake-generating neurons to dampen cortical activation and behavioral arousal (Jones, 2004).

The arousal system involved in wake generation utilizes a number of different neurotransmitters such as glutamate, noradrenaline, acetylcholine, dopamine, glutamic acid, histamine and orexine. Slow-wave sleep occurs through the inhibition of the arousal systems. The key neurotransmitters involved in sleep generation are instead adenosine, GABA, acetylcholine during REM sleep, glycine and some immune modulators (Krueger & Majde, 2003).

Regulation of sleep and wakefulness

According to the so-called 'two-process model', the occurrence, duration, intensity and internal structure of sleep are determined by two oscillatory processes called process 'S' or sleep homeostasis and process 'C' or circadian rhythm. The timing of sleep is determined by process C, which is strongly influenced by the light-dark cycle. Process S is homeostatic and increases during wakefulness until it reaches a circadian upper threshold and sleep is initiated.

Both processes interact closely and can be related to neural activities. The circadian rhythm in particular is determined by the rhythmic activity of the suprachiasmatic nuclei, being primarily responsible for changes in body temperature and endocrine secretions (mostly melatonin). This means that these variables vary closely with the circadian rhythm and vice versa, i.e., changes in body temperature and melatonin influence the circadian rhythm. The latter produces two 'opening windows' for sleep occurrence. The main window 'opens' in the late evening, accompanied by a drop in body temperature. The other one, less pronounced, 'opens' in the early afternoon. Important zeitgeber for the circadian clock are light, as mentioned already, but also ambient temperature changes, noise, nutrition, and social contacts.

The intrinsic circadian rhythm is a little bit longer than 24 h, on average by 15 min. The duration of REM sleep is strongly linked to the circadian sleep process, and the longest REM sleep in bedrest studies is found roughly 1–2 h after the body temperature has reached its minimum (Dijk & Czeisler, 1995).

From the results of circadian research it is evident that the homeostatic sleep regulation and the circadian process interact strongly with each other and influence each other. A clear experience of these interactions can be observed at flights across several time zones during intercontinental flights.

Biologic functions of sleep

Many theories attempt to explain the biologic function of sleep, without a clear winner. One such theory posits that sleep serves a restorative function for the brain and body. Normal sleep is subjectively associated with feeling refreshed on awakening. REM sleep is associated with increased CNS synthesis of proteins and is crucial for the CNS development of infants and young humans and animals. Growth hormone secretion is increased, while cortisol

secretion is decreased during sleep. All these can be used to support the restorative theory of sleep (Chokroverty, 2003).

Another theory of sleep function proposes that sleep has a central role in reinforcement and consolidation of memory. Sleep deprivation experiments have highlighted the important role of REM sleep in memory function (Chokroverty, 2003). Another theory suggests that sleep is important for thermoregulatory function. Experiments have shown that total sleep deprivation results in thermoregulatory abnormalities, NREM sleep maintains thermoregulatory function, and REM sleep is associated with impaired thermoregulatory responses (e.g., shivering and sweating) (Chokroverty, 2003).

Since the middle 1950s, when REM sleep was identified, sleep research has focused on understanding the physiology of dreams. Most dreams (about 80%) occur during REM sleep; the remainder occurs during NREM sleep. REM sleep dreams are more complex, have more emotional valence, can be bizarre, and are easier to recall. NREM sleep dreams are more logical and realistic, but more difficult to recall possibly because awakening from NREM sleep leaves a person feeling more confused and disoriented than awakening from REM sleep. During REM sleep, neuronal signals originating from the brainstem are transmitted to the cerebral hemispheres and stimulate the cortical association areas to produce images that compose dreams (Chokroverty, 2003).

2.2 Sleep disorders

Disruptions in the correct maintenance of sleep architecture and sleep-wake balance may lead to serious consequences for individuals and society in general, compromising both productivity and wellness. Sleep disorders are very common complaints affecting a large segment of world's population. Sleep disorders consequences represent a substantial economic burden and, for this reason, it is a major objective for clinicians properly diagnosing and treating this kind of pathology.

The guidelines for diagnosing sleep disorders are listed in the 'Diagnostic and Statistical Manual of Mental Disorders-Fourth Edition-Text Revision' (DSM-IV-TR) (American Psychiatric Association, 2000). Pharmaceutical companies have invested a lot of resources to develop new hypnotic drugs with a more safe and effective profile for the treatment of insomnia.

2.2.1 Classification

According to the 'Diagnostic and Statistical Manual of Mental Disorders-Fourth Edition-Text Revision' (DSM-IV-TR) (American Psychiatric Association, 2000), sleep disorders can be divided into two main categories.

- The first one, but less frequent, is made of disorders characterized by excessive sleepiness and, as its consequences, extreme daytime fatigue and the natural tendency of the individual to fall asleep at inappropriate times. An example of such disorders is narcolepsy, often confused with insomnia since the subject usually experiences disturbed nocturnal sleep and an abnormal daytime sleep pattern.
- The second category is insomnia, defined as difficulty in initiating or maintaining sleep or non-restorative sleep. Insomnia can be transient (i.e., lasting 2 to 3 days), short-term (i.e., lasting fewer than 3 weeks), or chronic (i.e., lasting at least 1 month) (Sateia, Doghramji, & Hauri, Evaluation of chronic insomnia, 2000). Transient and short-term insomnia are more likely to be precipitated by acute environmental events (e.g., being dismissed from a job, having a death in the family) than is chronic insomnia. Transient and short-term insomnia typically remit upon removal of or adaptation to the precipitating stressor.

The causes of insomnia are various. The major diagnostic classification scheme for sleep disorders (American Psychiatric Association, 2000) categorizes insomnia into primary and secondary forms.

- Primary insomnia is the difficulty in initiating or maintaining sleep in the absence of known comorbidities and other factors to which the sleep disturbance potentially could be attributed.
- Secondary insomnia is related to a psychiatric disorder such as depression, to an organic factor such as a medical condition, or to a substance use or abuse.

For a diagnosis of insomnia, at least one of the following forms of daytime impairment related to the nighttime sleep difficulty needs to be reported by the patient:

- fatigue or malaise;
- attention, concentration, or memory impairment;
- social or vocational dysfunction or poor school performance;
- mood disturbance or irritability;
- daytime sleepiness;
- motivation, energy, or initiative reduction;
- proneness for errors or accidents at work or while driving;
- tension, headaches, or gastrointestinal symptoms in response to sleep loss;
- concerns or worries about sleep.

2.2.2 Insomnia prevalence and comorbities

It is estimated that approximately 20% to 30% of adults worldwide have insomnia but that only approximately half of those having insomnia are diagnosed (Zisapel, 2004). In epidemiologic studies, prevalence estimates for insomnia range from approximately 5% to 35% (Sateia & Nowell, Insomnia, 2004). The variability in results of epidemiologic studies has been attributed to between-study differences in descriptors of insomnia, in age groups studied, and in degrees of severity and chronicity required to meet the definition of insomnia.

The prevalence of insomnia increases with age. In community-based research, more than half of non-institutionalized elderly patients were found to have chronic insomnia (Foley, Monjan, & Brown, 1995). Furthermore, insomnia is more common in women than in men. Women are approximately 1.5 times more likely than men to suffer from insomnia (Walsh, 2004).

Insomnia also appears to be more common among individuals with medical comorbidities (e.g., chronic somatic pain) or psychiatric comorbidities (e.g., depression) than among individuals not having comorbidities (Walsh, 2004). The nature of the relationship between insomnia and these comorbidities can be difficult to determine. Insomnia could contribute to causing the comorbidities or be caused by them (or both), or insomnia might simply be correlated with specific comorbidities in the absence of a causal relationship. Insomnia is a common feature of psychiatric disorders including depression, anxiety, schizophrenia, and eating disorders (Benca, 2005).

Insomnia is associated with adverse personal and economic consequences. Across several studies, insomnia has been shown to impair mood and cognition and to reduce functional ability (Pilcher & Huffcutt, 1996). Excessive daytime sleepiness can impair physical, mental, and psychosocial functioning. Insomnia has also been linked to impairments in physical, psychological, and social dimensions of health-related quality of life, to increased work absenteeism and to more frequent on-the-job accidents (Leger, 2000).

2.2.3 Insomnia treatment

The purpose of any insomnia treatment is the improvement of the patient's quality of life, through the identification and removal of any existing problem that may cause insomnia. This goal is usually reached through both pharmacologic therapy and educational and behavioural aid.

In the past, bromides, barbiturates, paraldehyde and methaqualone have been used as hypnotics, but although they have proved sedating properties, they also are cause of significant toxicity problems. Consequently, their use is no more recommended. Current hypnotic drugs indicated in treating insomnia include traditional benzodiazepines and non-benzodiazepines. The non-benzodiazepines are positive allosteric modulators of the GABA-A receptor. Like the benzodiazepines, they exert their effects by binding to and activating the benzodiazepine site of the receptor complex.

Benzodiazepines bind to a specific site (the benzodiazepine site) on the GABA-A receptor complex to open chloride channels that span neuronal

membranes (Sullivan, Petroski, & Verge, 2004). The entry into neurons of negatively charged chloride hyperpolarizes the cells with a resultant decrease in excitability. As GABA receptors are present at 40% of neuronal synapses, the decrease in excitability is widespread. This mechanism is responsible for the anxiolytic, muscle relaxant, and anticonvulsant properties of benzodiazepines as well as their hypnotic effects. Traditional benzodiazepines have been available since the 1960s. These types of medications vary significantly in their elimination half-lives, ranging from few hours to few days. This is the major concern about their use, since long half-life time values produce daytime undesired effects, like daytime impairment and higher risk of accidents and falls. Among the benzodiazepines approved for the treatment of insomnia, the ones with shorter half-lives, such as estazolam, triazolam, and temazepam, are recommended. Longer-acting benzodiazepines such as nitrazepam and diazepam have residual effects that may persist into the next day and are, in general, not recommended.

The side effects and the risks of tolerance and dependence with benzodiazepines and other older sleep aids motivated a search for sleep aids with improved risk/benefit ratios. This search resulted in the introduction of the first generation of nonbenzodiazepine hypnotics, including zolpidem, zopiclone, eszopiclone, and zaleplon, available since the 1990s. These medications are at least as effective as the benzodiazepines but have better safety and tolerability profiles. The sleep-promoting effects of these nonbenzodiazepine hypnotics, like those of the benzodiazepines, are attributed to interaction of the drugs with portions of the GABA-A receptor complex. Their advantages appear to come from their more selective mode of interaction with the GABA-A receptor combined with a reduced duration of action compared with the benzodiazepines (Foster, Pelleymounter, & Cullen, 2004). The shorter half-life time (1-2 hours), reduces the risk of morning residual effects. Moreover, the nonbenzodiazepine action implies a rapid sleep onset, so that the patient can take them just before going to bed. The short term safety and efficacy of these new hypnotics have been well proved and clinical experience supports safety during long-term intermittent use. Continuous long term use is instead not recommended as tolerance, dependence and addiction can occur.

Pharmaceutical companies are currently looking for new mechanisms and molecules with potential hypnotic effect. Melatonin analogs are an example, even if results of clinical studies of effects of exogenous melatonin on sleep are inconsistent so far (Turek & Gillette, 2004). Tricyclic antidepressants could also contribute to sleep-modulating effects, but their mechanism of action in insomnia is not known and has not been systematically assessed (Baldessarini, 2001). $5HT_2$ receptor antagonists are in development for insomnia as well. The limited results available to date suggest that 5HT2 antagonists may be effective for sleep maintenance but not for sleep induction. Finally, several lines of evidence point to an integral role of orexins (also known as hypocretins) in regulating sleep and wakefulness. Several compounds that

antagonize one or both orexin receptors have been synthesized (Nishino, 2007) or are being investigated for the treatment of insomnia in humans.

2.3 Conclusions

This chapter has highlighted that sleep is a phenomenon with high degree of complexity. Many mechanisms are involved in the neuroanatomy of sleep, as well as in the regulation of the wake-sleep alternation. Available drugs against insomnia are still far from completely restoring the quality of sleep in insomnia patients. With the growing evidence of the importance to maintain a physiological pattern in the sleep architecture, there is still room to find better treatments to insomnia. For example, hypnotic medications currently in use are especially demonstrated to positively influence the Latency to Persistent Sleep (LPS) (Elie, Ruther, & Farr, 1999; Hajak & Bandelow, 1998; Stone, et al., 2002). But the capability of the different drugs of restoring the normal sleep architecture has not been fully investigated. The internal structure can be easily registered obtaining a large amount of information in terms of the sequence of sleep stages recognizable at each 30-second interval. If such information is not aggregated but entirely used, deeper knowledge can be gained on both sleep physiology and hypnotics efficacy. This is the roadmap that pharmaceutical companies are now undertaking in their new investigations.

Hence, robust new methods need to be developed to fully describe data obtained from undergoing clinical studies in insomnia and for informing drug development strategies and decision making. In this context, mixed-effect pharmacokinetic-pharmacodynamic models are promoted by industry, academia and regulatory agencies (Miller, Ewy, Corrigan, Ouellet, Hermann, & Kowalski, 2005; Administration, 1999): they allow the characterization of the observed data, the prediction of drug effects under alternative dosing strategies, and the understanding of the physiological system.

The development of a new mixed-effect model for describing sleep architecture is the primary aim of this thesis, since models previous developed suffer from some important faults. This model will be widely evaluated for assessing how reliable it can be in describing real data or forecasting new scenarios. It will be used for describing data registered on insomniac patients taking placebo, for which age, gender and body mass index values are available.

In summary, the application of this model can help to better understand sleep physiology in primary insomnia, including the effect of insomniac individual's covariates on the internal sleep structure (not only the sleep aggregated parameters, as done before). Finally the developed model can be applied to assess sleep differences after drug and placebo intake, especially with respect to sleep architecture, a fundamental feature to understand the restorative or non restorative properties of a hypnotic.

2.4 Figures

Figure 2.1 Typical EEG, EOG and EMG activities related to the awake, stage 1, and stage 2 sleep.

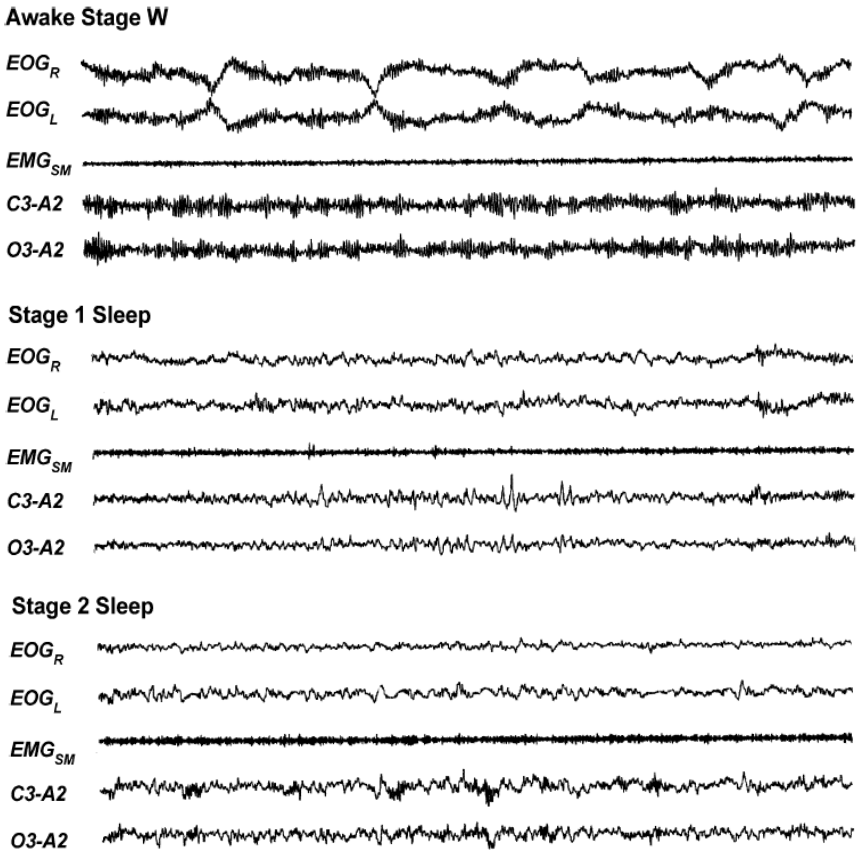


Figure 2.2 Typical EEG, EOG and EMG activities related to the stage 3, stage 4 and REM sleep.

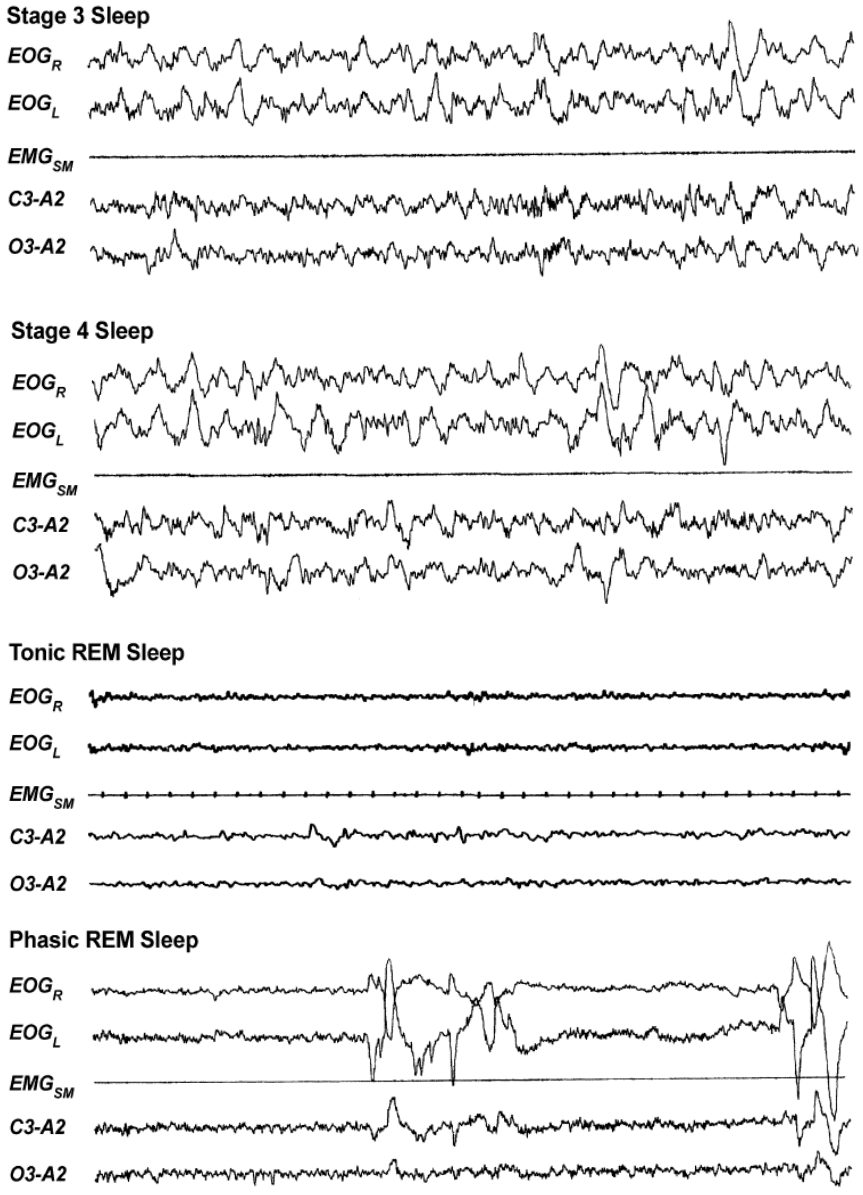
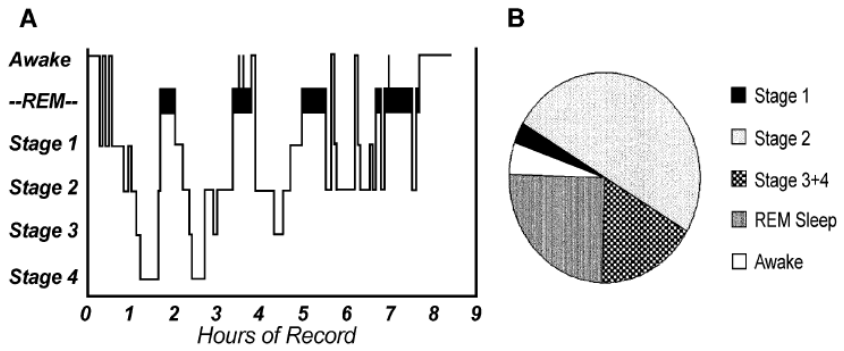


Figure 2.3 Normal sleep pattern. On the left a normal sequence of sleep stages during the night. On the right a diagram showing the total time spent in each stage.



Chapter 3

Mixed-effect PK-PD modeling and categorical data modeling

Data can be considered as continuous or categorical. Continuous data consist of variables with an infinite number of values, while categorical data consist of variables taking values in a finite set and can therefore be placed into mutually exclusive categories. Categorical data can be further divided into nominal or ordinal. The nominal ones are non-ordered and can be classified by 'names', while the ordinal ones have a natural order and can be organized into 'levels', even if the exact distance between the levels is generally unknown. Examples of nominal categorical variables are race or sex, while examples of ordinal categorical variables are age, patient compliance, values from sedation scales, scoring rating scales (e.g., HAMD in depression, Barthel in stroke, WOMAC in osteoporosis, ACR in rheumatoid arthritis) or pain intensity scales (e.g., absent, mild, moderate, severe).

As described in Chapter 1, the main objective of this work is the development, evaluation and application of a model able to describe polisomnography data (or 'sleep data'). These data can be considered as categorical: information derived from PSG recordings are 'categorized' into 6 mutually exclusive categories (AW, ST1, ST2, ST3, ST4, REM), according to the criteria illustrated in Paragraph 2.1.1. Such categories do not have a recognized natural order, therefore they can be considered as nominal categorical data. Moreover, the categories are more than two, so data are polychotomous (vs. dichotomous or binary).

An important aspect of clinical trials is to determine the pharmacokinetics (PK) of a drug, describing the time course of the systemic exposure following a given dose (Peck, et al., 1992). Even more essential is to investigate whether a drug has an effect and how this effect varies with the systemic exposure: that is, the pharmacodynamics (PD) of the drug. To assess PD, categorical data are commonly measured repeatedly within the same patient in clinical trials for describing both disease symptoms (through evaluation of clinical scores) and drug effects (through evaluation of clinical efficacy and adverse event severity).

Traditionally, when analyzing data from a confirmatory clinical study, the observations are analyzed by performing statistical testing, using for example

t-test or Wilcoxon signed-rank test, which compare the measurements before and after drug administration at a certain time point. Another approach is to perform statistical testing on the measurements for the active and the placebo arm in the study. This type of analysis, even though commonly used, answers only a narrow question. Mixed-effect analysis, the method used in PK-PD modeling, is instead more appealing from a learning perspective (Sheiner, Beal, & Dunne, 1997) as it provides the opportunity to explore the time course of the effect using all measurements obtained at different times within the same individual. Moreover, it allows the modeler to establish both population mean and individual responses, which is the key objective of any modeling analysis since the drug under development has to be effective and safe for the whole population.

The first part of this chapter briefly illustrates the 'mixed-effect modeling approach'. Then it recalls the basics on the usual software platform used for implementing this approach, NONMEM (ICON Development Solutions) (Beal & Sheiner, 1998). This tool is capable of estimating non-linear mixed-effect models through the likelihood maximization. It provides three estimation methods for the likelihood estimation, namely 'First Order', 'First Order Conditional Estimation' and 'Laplace'. Only the latter can be applied when data to model are categorical, therefore only Laplace is presented here. Various methods and diagnostics are implemented when developing a model in order to select between different alternatives and evaluate the final model. The main method for model selection (log-likelihood ratio test) is introduced in this chapter.

The specificities of nominal categorical data modeling are discussed in the second part of the chapter. First of all, a specific expression for the likelihood is presented. Such expression requires that the model describes the probability distributions of the potential outcomes instead of the outcome values themselves. That is, the model to develop needs to be a 'probability model'.

When assuming that the probability of a category is function of only external predictors, like drug exposure, covariates, time, etc., we exclude that observations in a specific subject can be dependent on each other. This cannot be the case for sleep stages. Therefore, instead of implementing a probability model, we need to develop a so-called 'transition model'. The specific transition model adopted in this work is the Markov-chain model.

In the context of mixed-effect modeling, it becomes useful to mathematically transform the probability values to be described through the so-called 'link functions', in order to better estimate the probability distributions of the individual values of the parameters. Works published so far in the context of population PK-PD modeling used link functions only as transformations of binomial probabilities, even if polychotomous data were involved.

The concepts of probability and transition models, some fundamentals on Markov chains, a brief review on link functions and the disadvantages of mod-

eling binomial probabilities even when data are polychotomous are introduced in the second part of this chapter.

3.1 Mixed-effect models

Let the gathered data arise from *M* subjects participating in a certain clinical study and let *X* be the data matrix, where each row contains individual data on N different times (i.e., data are 'longitudinal', or 'repeated-measure'):

$$X = \begin{bmatrix} X_1^T \\ X_2^T \\ \vdots \\ X_M^T \end{bmatrix} = \begin{bmatrix} x_{11}\ x_{12} \cdots x_{1N} \\ x_{21}\ x_{22} \cdots x_{2N} \\ \vdots\ \ \vdots\ \ \ldots\ \ \vdots \\ x_{M1} x_{M2} \cdots x_{MN} \end{bmatrix}. \quad [4.1]$$

Suppose such data need to be described through a mathematical model. Classical approaches, e.g. Least Squares (LS), Weighted Least Squares (WLS), Maximum Likelihood (ML) estimate, Bayesian approach, etc., are generally designed for individual fitting:

$$X_i = f_i(z, \varphi_i) + v_i, \quad [4.2]$$

where X_i are the individual data, f_i the corresponding model prediction, z the independent variable, φ_i the individual parameters and v_i the vector of the random error affecting individual data. The random error is generally due to measurement errors and noise and it is supposed to be drawn from a Gaussian distribution with zero mean and covariance matrix equal to Σ:

$$v_i \sim N(0, \Sigma). \quad [4.3]$$

These approaches present some identification problems in case of noisy or sparse individual data: individual estimates in certain cases may be not accurate or impossible to obtain. Furthermore, information at the mean and individual levels is often needed. Consequently, when working in similar contexts, it is suitable to build models and identify the corresponding parameters using the so called 'population approaches', or 'mixed-effect approaches'.

Let the individual parameters of the population under examination belong to a certain distribution (gaussian, lognormal, etc.), characterized by mean θ and covariance Ω:

$$\varphi_i \sim Distr(\theta, \Omega). \quad [4.4]$$

It is then possible to think that individual predictions from the model are obtained using the realizations of such distribution as individual parameters. The population approach allows to investigate, with varying levels of precision, both the mean parameter and the variability in the population, i.e. the first two moments of the parameters distribution, $\Theta = (\theta, \Omega)$.

Mixed-effect approaches can be implemented in various ways: starting from the ones with more limitations and less informative power, the naïve average data approach, the naïve pooled data approach, the standard two-stage

approach, the iterative two-stage approach, and the 'Non-Linear Mixed-Effect (NLME) approach'. The latter is one of the most interesting for population analysis, and it is particularly well suited for biological and medical data, which display heterogeneity of responses to stimuli and treatments. NLME was used in this work and is introduced in the next paragraphs.

3.1.1 Non-linear mixed-effect modeling: theory

The non-linear mixed-effect approach is based on the assumption that the (unknown) process to be described is characterized by a typical behaviour which is common to the whole population and by some sources of variability that make the individual behaviours differ from the typical one. The latter is determined by the so called 'fixed effects', while the identified sources of variability are of two (or more) different types and are called 'random effects'. The first usual source of variability is the intrinsic difference that exists among subjects: one individual is obviously different from another one. This is called 'inter-individual' or 'between-subject' variability (IIV and BSV, respectively). Mostly in the medical field, to understand the variability between subjects is as important as to understand the characteristics of the typical individual. The second usual source of variability is the 'residual error' (also called 'noise or 'intra-individual error'): this is the difference between the predictions of the model for the observations measured in the same individual. It is also called 'intra-individual' or 'within-subject' or 'residual unexplained' variability (RUV). For taking into account all of these assumptions, the NLME approach specifies the model in a hierarchical fashion, integrating an 'individual' model and a 'population' one. In this way, it allows to estimate both the vector of population characteristics, $\Theta = (\theta,\ \Omega)$, and the individual parameters, φ_i.

Individual model

The individual model is aimed to describe individual data specifying the relationship between the dependent variables, independent variables and individual parameters.

Let

z be the independent variable, for example 'time' in a time series;

z_t be the t-th value of the independent variable, t = 1, …, N;

x_{it} be the t-th observation of the i-th individual, t = 1, …, N and i = 1, …, M;

φ_i be the vector of model parameters of subject *i*.

Each individual measure, x_{it}, can be described by the individual model in this way:

$$x_{it} = f_i(z_t, \varphi_i) + v_{it}, \quad \forall t = 1, \dots, N, \tag{4.5}$$

where $f_i(z_t, \varphi_i)$ is the individual model prediction and v_{it} is the residual error.

Using a vector notation:

$$\begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{iN} \end{bmatrix} = X_i = f_i(z, \varphi_i) + v_i = \begin{bmatrix} f_i(z_1, \varphi_i) \\ f_i(z_2, \varphi_i) \\ \vdots \\ f_i(z_N, \varphi_i) \end{bmatrix} + \begin{bmatrix} v_{i1} \\ v_{i2} \\ \vdots \\ v_{iN} \end{bmatrix}. \tag{4.6}$$

The residual errors v_{it} are classically assumed independently normally distributed with

$$E[v_i|\varphi_i] = 0, \tag{4.7}$$

$$Cov(v_i|\varphi_i) = R_i(\varphi_i, \xi), \tag{4.8}$$

where R_i is a diagonal matrix depending on ξ (a constant characteristic across individuals) and possibly on the individual parameters (according to the error model structure). Therefore,

$$v_i \sim N\big(0, R_i(\varphi_i, \xi)\big). \tag{4.9}$$

Population model

A model for φ_i is also needed in order to account for inter-individual variability among the φ_i's. In particular, the population model relates the individual parameters to the covariate vector, the fixed effects and the inter-individual random effects.

Let

a_i be the covariate vector, i.e. the set of individual values for weight, age, etc.,

η_i be the vector of inter-individual random effects associated with the subject *i*,

θ be the vector of fixed effects.

A general population model is given by

$$\varphi_i = d(\theta, \eta_i, a_i), \tag{4.10}$$

where d is a multi-dimensional function, and η_i are supposed to be drawn from a normal distribution having zero mean and Ω covariance matrix, i.e.,

$$\eta_i \sim N(0, \Omega). \tag{4.11}$$

Eventually, the model can be expressed as follows:

$$X_i = f_i(z, d(\theta, \eta_i, a_i)) + v_i,\ \eta_i \sim N(0, \Omega),\ v_i \sim N\big(0, R_i(\varphi_i, \xi)\big). \qquad [4.12]$$

Most of the non-linear mixed-effect modelling methods estimate the parameters using a Maximum Likelihood (ML) approach: the data probability is given by a function of the model parameters and parameter estimates are chosen to maximize this probability.

The overall likelihood is the product of all individual likelihoods L_i and since the likelihood must account for the random effects on the individual level, the individual likelihood is expressed as the integral over all possible values of η_i:

$$L(\theta, \Omega) = \prod_{i=1}^{M} L_i = \prod_{i=1}^{M} \int l(X_i|\theta, \eta_i, a_i) h(\eta_i|\Omega) d\eta_i, \qquad [4.13]$$

where h is a multivariate normal density function with zero mean and covariance matrix Ω.

The population parameters $\Theta = (\theta, \Omega)$ can be estimated by maximizing $L(\theta, \Omega)$ or, similarly, minimizing the following Objective Function Value (OFV):

$$\mathrm{OFV} = -2\log\big(\mathrm{L}(\theta, \Omega)\big). \qquad [4.14]$$

Each time the likelihood is computed, during the maximization process, individual parameters φ_i can be derived. In particular, the inter-individual random effects are estimated with a 'Maximum A Posteriori' (MAP) approach using as prior for the their distribution

$$\eta_i \sim \mathrm{N}\big(0, \hat{\Omega}\big), \qquad [4.15]$$

where $\hat{\Omega}$ is the previously estimated population variability.

In particular, calling $\mathrm{R_i}\big(\varphi_i, \xi\big) = \mathrm{R_i}$, the η_i estimates are obtained through the minimization

$$\hat{\eta}_i = \mathrm{argmin}_{\eta_i}\big[\mathrm{X_i} - \mathrm{f_i}\{\mathrm{d}\big(\hat{\theta}, \eta_i, \mathrm{a_i}\big)\}\big]^{\mathrm{T}} {\mathrm{R_i}}^{-1}\big[\mathrm{X_i} - \mathrm{f_i}\{\mathrm{d}\big(\hat{\theta}, \eta_i, \mathrm{a_i}\big)\}\big] + {\eta_i}^{\mathrm{T}} \hat{\Omega}^{-1} \eta_i, \qquad [4.16]$$

and called 'Empirical Bayes Estimates' (EBE's). Once the inter-individual random effects are computed, the individual parameters can be obtained according to the population model

$$\hat{\varphi}_i = \mathrm{d}\big(\hat{\theta}, \hat{\eta}_i, \mathrm{a_i}\big), \qquad [4.17]$$

and are called 'Post-hoc estimates'.

3.1.2 Non-linear mixed-effect modeling: NONMEM

NONMEM version VI (Icon Development Solutions) (Beal & Sheiner, 1998) is a software platform that allows to perform population analysis through the non-linear mixed-effect approach and whose name stands for 'NON linear Mixed Effect Modeling'.

The basic steps for running NONMEM are:

(a) to organize data input,
(b) to write the control file, which specifies the mixed-effect model,
(c) to run the model and obtain model parameter estimates.

(a) NONMEM needs 'data input' files organized into records with some pre-defined items as follows. The subject number, called 'ID', is the first item in each record. The records appear in subject order and, within a subject, they are organized by time if a time series is being analyzed. Time specification, called 'TIME', is usually the second item in each record. The dependent variable, called 'DV', is the third one. Additional items can be added if the model requires so, for example subject covariates, such as weight, height, gender, etc.

(b) Once data are organized, the model is specified in the 'control file', or 'control stream', which for an estimation problem with categorical data typically contains the following control elements.

$PROB	States a name for the problem being solved.
$DATA	Specifies the name of the data file.
$INPUT	List the names of the data records in the input file, in the exact order of the data file columns.
$PRED	Describes the routine predicting the observations. It is the main part of the control stream and it specifies the mixed-effect model.
$THETA	Lists the initial estimates of the fixed-effect parameters.
$OMEGA	Lists the initial estimates of the variance-covariance matrix for the inter-individual random effects.
$SIGMA	Lists the initial estimates of the variance of the intra-

	individual random effects.
$EST	Provides the parameters that control the estimation process. Typically includes: METHOD (indicating which estimation method has to be applied), MAX (maximum number of iterations), LIKELIHOOD (indicating whether the likelihood is defined by the modeller; this option is necessary in case of categorical data modelling).
$COV	Implies the estimation of the full variance-covariance matrix of the parameter estimates. This step is useful to get standard errors of the estimated parameters.
$TABLE	Produces an output table of the results.

One-line comments can be specified after typing a semicolon. NONMEM needs specific keywords for each feature of a mixed-effect model, in particular:

(i) THETA is a fixed effect parameter, i.e. an element of θ.;
(ii) ETA is an inter-individual random effects, i.e. an occurrence of η_i;
(iii) EPS is an intra-individual random errors, i.e. an occurrence of v_i;
(iv) Y is an observation, x_{it}, or the correspondent user-defined likelihood (if LIKELIHOOD is specified in $EST).

(c) NONMEM estimates model parameters with a ML approach. The likelihood of non-linear mixed-effect models, Equation [4.13], is often difficult to compute exactly because of non-linearity of the model in the random effects. To deal with this problem, three approximation methods have been implemented in NONMEM: the 'First Order', the 'First Order Conditional Estimation' and the 'Laplace' methods, with high, medium and low levels of approximation, respectively. Within NONMEM, the Laplace method is the only one applicable with categorical data. Therefore only this likelihood approximation is introduced in the following paragraph.

3.1.3 Laplace approximation

The Laplace method is based on the Laplacian approximation of the exact marginal (i.e., individual) likelihood specified by the hierarchical non-linear model (Equation [4.13]). Given a complex integral, $\int b(x)dx$, $b(x)$ can be re-

expressed as $e^{logb(x)} = e^{g(x)}$ and $g(x)$ can be approximated by a second-order Taylor expansion around a point x_0 as:

$$g(x) \simeq g(x_0) + (x - x_0)g'(x_0) + \frac{(x - x_0)^2}{2!}g''(x_0). \quad [4.18]$$

The approximated integration is called a first order Laplacian approximation to the true integration and is the following:

$$\int b(x)dx = \int e^{g(x)}dx \approx \int e^{g(x_0)+(x-x_0)g'(x_0)+\frac{(x-x_0)^2}{2!}g''(x_0)}dx. \quad [4.19]$$

When considering η_i as x and therefore $l(X_i|\theta, \eta_i, a_i)h(\eta_i|\Omega)$ as $b(x)$, this approximation allows to compute the approximated likelihood in closed form. The Taylor expansion is computed around the conditional estimates of η_i. Other details on how to derive the closed form and how NONMEM implements the likelihood maximization can be found in (Wang, 2007).

3.1.4 Model selection

When deciding to include a parameter in a model, objective measurements are needed to verify that the inclusion is relevant. Relevance can be defined as physiological plausibility, clinical impact or statistical significance. A combination of all these is usually used in combination with the assessment of the predictive performance of the model. Physiological plausibility as well as clinical impact are depending on the drug and disease and are therefore decided based on subject-matter information (Sheiner & Wakefield, 1999). Criteria for a parameter inclusion based on physiological plausibility and clinical impact need to be defined pre-analysis in order to make these measurements objective.

Statistical significance

The most commonly used test within PK-PD modeling to assess statistical significance is the Log-Likelihood Ratio (LLR) test. The ratio between the likelihood of the model with the new parameter included (the full model) or excluded (the reduced model), given that the models are nested, is assumed to be χ^2-distributed with the number of differing parameters between the models as the degrees of freedom (dof). Models are nested if the full model can collapse into the reduced model. The parameter can thus be included in the model based on a statistical significance level. This level is corresponding to

the type I error, i.e., the risk of including a parameter that does not belong to the model. The usual threshold on the level of significance for parameter inclusion-exclusion is p=0.05.

The objective function value calculated by NONMEM is proportional to the OFV presented in Equation [4.14], and the difference in OFV between two models is then the likelihood ratio. Thus, the OFV can be used to perform statistical testing between nested models. A number of studies have been performed to assess the robustness of the LLR test (Wahlby, Jonsson, & Karlsson, 2001; Wahlby, Bouw, Jonsson, & Karlsson, 2002), but only one has been performed with categorical data (Wahlby, Matolcsi, Karlsson, & Jonsson, 2004). The investigated data in this study was ordinal, originating from and analyzed with the proportional odds model. It was concluded that the LLR test performs well, with type I errors equal to or lower than expected.

3.2 Categorical data modeling

As a starting point, the simplest case of modeling independent binary pooled data is considered. That is, data are observed once per individual (or if repeated measures are taken, they are considered as independent), take two values only (e.g., DV=1 or DV=0, i.e. response or no-response) and are considered as belonging to one unique individual. In this case the easiest way to model the data is to define the likelihood of each DV as its probability p if DV=1 and 1-p if DV=0. The value of π needs to be estimated in order to maximize the overall likelihood, defined as

$$L(\pi) = p^k(1-p)^{m-k}, \qquad [4.20]$$

where k is the occurrence of DV=1 and m is the number of observations. In this case the NONMEM control stream would look like this:

```
$PROB Probability model for independent binary pooled data
$DATA data
$INPUT ID TIME DV
$PRED
  PROB = THETA(1)
  IF(DV.EQ.1) Y = PROB
  IF(DV.EQ.0) Y = 1-PROB
$THETA (0,.5,1) ; fixed effect constrained between 0 and 1
$EST LIKE
```


The control element for the model estimation ($EST) states that the entity to model (Y) is in this case a user-defined likelihood, not an observation like it usually happens in PK-PD modeling. As seen with Equation [4.20], the likelihood can be expressed in an exact closed form, therefore no approximation methods are adopted in this case. Moreover, since the probability of a certain outcome, rather than the values of the outcome itself, is modeled, no residual error is defined in the model.

If we want to consider that data come from different individuals with different characteristics, the naïve pooled approach needs to be replaced with a population one, and the best one, as already seen, is the non-linear mixed-effect approach. In this case the individual probability p_i deviates from the typical individual probability but still needs to be constrained between 0 and 1. This characterization can be obtained considering a new individual variable with values in (-∞, +∞), called g_i and defined as the sum of the typical value g

and the random deviation η_i from g taken from a normal distribution with zero mean:

$$g_i = g + \eta_i, \qquad \eta_i \sim N(0, \Omega). \qquad [4.21]$$

The functions able to transform values in the non-bounded scale (-∞, +∞) back to values in the probability scale (0, 1) are called 'link functions' and are introduced in Paragraph 3.2.2. The most used ones are the 'logistic' functions. The models implementing this approach are called 'logistic models', regardless of the type of link function used (logistic or not).

The NONMEM code becomes similar to the following:

```
$PROB Mixed-effect probability model for independent binary data
$DATA data
$INPUT ID TIME DV
$PRED
  G = THETA(1)+ETA(1)
  PROB = ... ; the link function of G is inserted here
  IF(DV.EQ.1) Y = PROB
  IF(DV.EQ.0) Y = 1-PROB
$THETA .1 ; non-constrained fixed effect
$OMEGA .5
$EST LIKE METHOD=1 LAPLACE
```


The ETA's are therefore included in the model within a link function, which is always highly non-linear, hence the likelihood becomes intractable again (see Equation [4.13]). Among the estimation methods available in NONMEM VI, Laplace is the one introducing a lower level of approximation, since it uses a second order Taylor expansion instead of a first order one for linearizing the model about the ETA's. Anyway, NONMEM does not allow First Order and First Order Conditional Estimation when the option LIKE is used, thus the choice of the Laplace approximation is compulsory. This choice is expressed by adding 'METHOD=1 LAPLACE' in the $EST record. Other methods not implemented in NONMEM, such as the Gaussian Quadrature, have been shown to perform somewhat better in some situations, but often with stability issues (Jonsson, Kjellsson, & Karlsson, 2004; Plan, Maloney, Traconiz, & Karlsson, 2008).

Let me consider now that more outcomes x_{it} can be registered for a unique subject i at different values for the independent variable z_t (time for example), i.e. data are longitudinal. In this general case the mixed-effect model for independent binary data becomes

$$g_{it} = f'_i(z_t, d(\theta, \eta_i, a_i)), \qquad \eta_i \sim N(0, \Omega), \qquad [4.22]$$

where the covariate effects are also included and f' is the individual model prediction in the scale (-∞, +∞). But the probability p_{it} of a certain outcome for the *i*-th subject at time z_t is given by a link function of g_{it}, hence the final probability model can be written as the following:

$$p_{it} = f_i(z_t, d(\theta, \eta_i, a_i)), \quad \eta_i \sim N(0, \Omega). \qquad [4.23]$$

This equation is similar to Equation [4.12], and can be used as a user-defined model for each observation likelihood when DV=1 (1-p_{it} is used when DV=0). Consequently, the likelihood $l(X_i|\theta, \eta_i, a_i)$ of Equation [4.13] can be expressed by the product of the modeled probabilities of the outcomes.

Similar considerations can be derived for polychotomous longitudinal data. Therefore, the two major differences between models for continuous and categorical data when the non-linear population approach is applied can be summarized as follows:

- the likelihood or probability of a certain observation is used for the parameters estimation (instead of the observation itself);
- the residual error is not considered.

Categorical data can be modeled in many different ways, according to their specificities: data can be dichotomous or polycothomous, nominal or ordinal, and correlation (dependence) can be considered or not when data are repeatedly measured in a subject. Count and time-to-event data are other specific kinds of categorical data: examples of count data are the numbers of acid refluxes, emetic episodes or epilepsy seizures per time interval; examples of time-to-event data are time until cardiovascular death, until AIDS for HIV patients, or until a side-effect like nausea.

Models that have been recently widely used in population PK-PD modeling of categorical data (mostly for the PD part of it) are specified in Table 3.1 for the respective data type involved. The focus of this work is on models on polychotomous longitudinal nominal data considering the correlation between the repeated measures.

3.2.1 Markov-chain models

Markov-chain mixed-effect models have been used for modeling of drug compliance (Girard, Blaschke, Kastrissios, & Sheiner, 2002), spontaneously reported side effects (Zingmark, Kagedal, & Karlsson, 2005), and sleep stages (Karlsson, et al., 2000). Compared to other models for categorical data, these

ones can be called 'transition models', since they describe the probabilities of transitioning between categories instead of the probabilities of the categories themselves. To model transitions between states means taking into consideration that the states may be correlated.

In fact, if observations in a subject show serial correlation, considering them as independent may bring to both modeling and simulation misspecifications. In the modeling process the information content may be overestimated, the time-course of dependence (correlation) would be ignored and inter-individual variability may be overestimated. In the simulation process realistic time courses of the PD variable would be hardly produced.

Sleep architecture provides us with a good example of how correlation may be fundamental in describing the data. Even if the only awake and asleep states are considered, the reader can easily understand that during nighttime their correlation is high: after moving from wake to sleep, many observations of sleep can be registered before observing wake again, and vice versa. If correlation is not modeled, i.e. the probabilities of the two states are modeled instead of the transitions between them, this feature will not be caught.

The NONMEM control stream for a simple model of sleep/wake data could look like the following (not considering inter-individual variability for simplicity):

```
$PROB Transition probabilities between sleep and wake
$DATA data
$INPUT DV PDV
;PDV = Value of immediately preceding observation
$PRED
  P10 = THETA(1)
  P01 = THETA(2)
  IF(PDV.EQ.0.AND.DV.EQ.1) Y=P10
  IF(PDV.EQ.0.AND.DV.EQ.0) Y=1-P10
  IF(PDV.EQ.1.AND.DV.EQ.0) Y=P01
  IF(PDV.EQ.1.AND.DV.EQ.1) Y=1-P01
$THETA  (0,.1,1)    ; PROB AWAKE GIVEN ASLEEP
$THETA  (0,.1,1)    ; PROB ASLEEP GIVEN AWAKE
$ESTIM LIKE
```


In the model above, correlation is considered only between states occurring at two consecutive time points. In other words, a first-order Markov-

chain model is implemented, as next paragraph will define more formally. However, the probability of being in a specific state at time t can generally depend on a longer history of the chain. Moreover, each transition probability can be described as time-independent ('homogeneous') or time-dependent ('non-homogeneous'). In the latter case a functional form of the time dependence needs to be chosen.

3.2.2 Basic theory on Markov chains

Let a random process be a finite sequence $X = \{X_t\}_{t\geq 1} = \{X_1, X_2, \dots\}$ of random variables taking values in a discrete set $\mathcal{S}$. The elements of $\mathcal{S}$ are called 'states' of the system and thus $\mathcal{S}$ the 'state space'.

The index t of X_t is usually thought as a 'time index', even though it is not necessarily related to the concept of time, but rather expressing the ordered evolution of the process. Consequently, X_t represents the state of the process at 'time' z_t, where z is the 'time' vector.

The process $X = \{X_1, X_2, \dots\}$ is called 'Markov chain' if it satisfies the 'Markov property':

$$P(X_{t+1} = x_{t+1} \mid X_t = x_t,\ X_{t-1} = x_{t-1}, \dots,\ X_1 = x_1) = P(X_{t+1} = x_{t+1} \mid X_t = x_t), \quad [4.24]$$

for every sequence $x_1, \dots, x_t, x_{t+1}$ of elements of $\mathcal{S}$ and for every $t \geq 1$.

This property states that the probability of an event one step into the future conditioned on the entire past up to the present time t is equal to the conditional probability of the future event given just the present one. In particular, a sequence of such random variables is said to be a 'first-order Markov-chain process' because each outcome depends exclusively on the previous state.

The process X is an 'Nth-order Markov-chain process' if the dependency between the random variables constituting the process involves N successive steps in the sequence, i.e., if the probability of the future outcome is conditioned on the N previous states:

$$P(X_{t+1} = x_{t+1} \mid X_t = x_t, \dots,\ X_1 = x_1) = P(X_{t+1} = x_{t+1} \mid X_t = x_t, \dots,\ X_{t-N} = x_{t-N}). \quad [4.25]$$

Given a first-order Markov-chain process $X = \{X_1, X_2, \dots\}$ with k and m in its state space $\mathcal{S}$, the conditional probability

$$p_{km}(t) = P(X_{t+1} = m \mid X_t = k) \geq 0 \qquad [4.26]$$

is called the 'transition probability' from k to m at time z_t. If the transition probabilities do not depend on time, i.e. $p_{km}(t) = p_{km}(t+h) = p_{km}$, $\forall t \in \mathbb{N}^+, \forall h \in \mathbb{N}^+$, the Markov chain is said to be 'time-homogeneous'.

Assuming a finite state space $\mathcal{S}$, that is $\mathcal{S} = \{0, 1, \dots, S\}$, it is useful to collect the transition probabilities from/to the states of $\mathcal{S}$ in a matrix:

$$P(t) = \begin{bmatrix} p_{00}(t) & p_{01}(t) & \cdots & p_{0S}(t) \\ p_{10}(t) & p_{11}(t) & \cdots & p_{1S}(t) \\ \vdots & \vdots & \vdots & \vdots \\ p_{S0}(t) & p_{S1}(t) & \cdots & p_{SS}(t) \end{bmatrix}, \qquad [4.27]$$

which is called 'transition probability matrix'. Each row of $P(t)$ represents all the transition probabilities from a single state of $\mathcal{S}$: therefore, the probabilities in each row must sum up to 1:

$$\sum_{m=0}^{S} p_{km}(t) = 1\,, \qquad \forall\, i \in \mathcal{S}. \qquad [4.28]$$

Assuming to observe M independent realizations of the process of the same length N:

$$\begin{gathered} X^{(1)} = \{x_{11}, x_{12}, x_{13}, \dots, x_{1N}\} \\ X^{(2)} = \{x_{21}, x_{22}, x_{23}, \dots, x_{2N}\} \\ \dots\dots \\ X^{(M)} = \{x_{M1}, x_{M2}, x_{M3}, \dots, x_{MN}\}, \end{gathered} \qquad [4.29]$$

it is possible to define some statistics depending on the data:

- N_{km} ≜ number of transitions from state k to state m in all the realizations,
- N_k ≜ number of transition from state k,
- TR ≜ total number of transitions in the data,
- SO_k ≜ number of occurrences of state k in the data.

Once these statistics are available, the frequency of occurrence of each stage is computed as

$$\hat{\rho}_k = \frac{SO_k}{\sum_{k=0}^{S} SO_k}, \qquad \forall\, k \in \mathcal{S}, \qquad [4.30]$$

and the 'transition frequencies' between stages are calculated as

$$\hat{f}_{km} = \frac{N_{km}}{N_k}, \qquad \forall\, k, m \in \mathcal{S}. \qquad [4.31]$$

Every single realization of the process can be considered as a path in time through the state space, and its probability

$$P((X_1, \dots, X_t) = (x_1, \dots, x_t)) \qquad [4.32]$$

is the joint probability of $(X_1, \dots, X_t)$.

For a Markov chain X and for any path $\{x_1, x_2, x_3, \dots, x_t\}$, the conditional probability of the path conditioned on the first value is the product of the transition probabilities between successive states of the path:

$$\begin{aligned} P\big((X_2, \dots, X_t) = (x_2, \dots, x_t) \big| X_1 = x_1\big) \\ = p_{x_1 x_2} p_{x_2 x_3} \cdots p_{x_{t-1} x_t}, \end{aligned} \qquad [4.33]$$

and consequently the probability of the path is:

$$\begin{aligned} P\big((X_1, \dots, X_t) = (x_1, \dots, x_t)\big) \\ = P(X_1 = x_1)\, p_{x_1 x_2} p_{x_2 x_3} \cdots p_{x_{t-1} x_t}. \end{aligned} \qquad [4.34]$$

3.2.3 Link functions

Different link functions (Agresti, 2002) are used to ensure that the probability of an event ranges from zero to one, while the estimated parameter ranges from –infinity to infinity. The 'logit' or 'log of odds' transformation is by far the most used transformation of categorical data, possibly for its mathematical tractability. The cumulative distribution of the observations is assumed to be logistically distributed and it is defined as

$$g(p) = \text{logit}(p) = \ln \frac{p}{1-p}. \qquad [4.35]$$

The link function which transforms back the logit values into probability values is the so-called 'logistic' function:

$$p(g) = \frac{\exp(g)}{1+\exp(g)}. \qquad [4.36]$$

Other transformations are also possible, for example the 'probit' one. This transformation assumes that the cumulative distribution of the observations is the inverse of the cumulative normal distribution:

$$g(p) = \text{probit}(p) = \Phi^{-1}(p), \qquad p(g) = \Phi(g). \qquad [4.37]$$

The 'log-log' transformation assumes instead that

$$g(p) = \text{log-log}(p) = \ln(-\ln(p)),$$
$$p(g) = \exp\big(-\exp(g)\big). \qquad [4.38]$$

Another quite common transformation is the 'complementary log-log', which assumes that the cumulative distribution is the Gumbel distribution:

$$g(p) = \text{complementary log-log}(p) = \ln\ln(1-p),$$
$$p(g) = 1 - \exp\big(-\exp(g)\big). \qquad [4.39]$$

The different transformations give rise to slightly different cumulative probability distributions, schematically depicted in 3). Notable is that both the logit and the probit transformations are symmetric while the complementary log-log is not. The largest differences are also seen at the tails of the distributions.

The major drawback of using link functions in a model is that once we model a logit, for example, the parameters used to relate the chosen predictors of the logit to the logit itself loose their direct physiological meaning. If, for instance, a logit is modeled as function of dose (the predictor) through a linear model

$$g = Baseline + Slope \cdot Dose, \qquad [4.40]$$

than the meaning of the parameter *Slope* with respect to the probability outcome is lost: the diagram of the relation between dose and g is a straight line, while the diagram of the relation between dose and probability is not.

Another drawback comes from the definition of the uncertainties on the parameter estimates through standard errors (se) and relative standard errors (rse). The same standard error (se $= 0.2$) on two estimates of a logit parameter, $\hat{g} = 0$ and $\hat{g} = 2$, for example, reflects a certain uncertainty on the estimated probability value $\hat{p} = 0.50$ (i.e. $p \in [0.45,0.55]$) and lower uncertainty on the estimated probability value $\hat{p} = 0.88$ (i.e. $p \in [0.86,0.90]$), respectively. Relative standard errors, instead, cannot be used at all: if a logit estimate is close to zero, this measure explodes.

Each of the link functions introduced above is defined on probabilities of binary events. No other link functions have been implemented in the literature within mixed-effect modeling. Nevertheless, two published works handled polychotomous nominal data with population models (Karlsson, et al., 2000; Kjellsson, Ouellet, Corrigan, & Karlsson, 2008), in particular sleep data with more than two categories. The solution they provided was to consider each transition as a binary event, i.e. an event with only two possible outcomes:

- moving to one specific sleep stage S;
- moving to a sleep stage other than S.

In Karlsson's work (Karlsson, et al., 2000), six different stages were considered, hence a total of 30 transitions from one stage to another were possible. Not every transition was equally likely though, and certain transitions hardly

ever occurred in practice. Therefore, 17 transitions were modeled at the end. In Kjellsson's work (Kjellsson, Ouellet, Corrigan, & Karlsson, 2008) 16 transitions were modeled instead. As no parameters were shared between the different transitions, all transitions were modeled separately. It is evident that working with 17 or 16 models causes a very extensive work. Moreover, the reader needs to consider that the frequent sampling of sleep contributes to the complexity of the model building as most software for mixed-effect modeling has an upper limit for the number of observations allowed per individual. In addition, the run-times of the models increase rapidly with increasing number of observations.

With polychotomous data (in a mixed-effect transition model), the sum of all probabilities of transitions starting from a certain state needs to be equal to 1, both for the individuals in the population and for the typical individual. In the binary-logit approach this constrain is enforced from estimating one transition probability as 1 minus the sum of all the other previously estimated transition probabilities from the same state (Karlsson, et al., 2000). But this strategy does not provide precise information on the inter-individual distribution for the parameters involved in the description of the lastly modeled transition.

3.3 Conclusions

This chapter provided an introduction to the estimation approach and software platform which will be adopted for modeling sleep data. It also presented the specificities of models dealing with categorical data. When model estimation is based on the maximization of categorical data likelihood, the main conclusions are the following:

- the likelihood or probability of the observations need to be used for parameter estimation, instead of the observations themselves;
- the residual error cannot be considered.

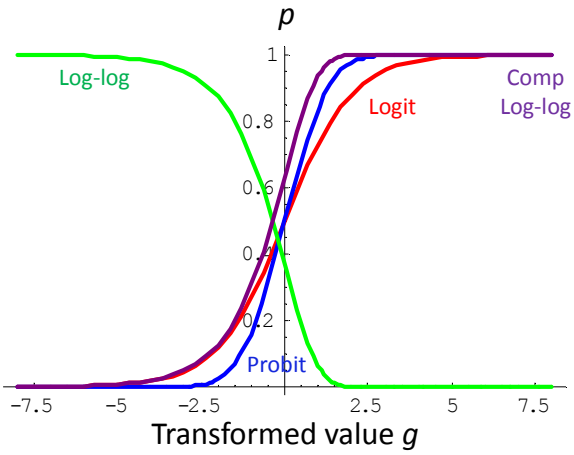
Sleep data are categorical polychotomous longitudinal data and have already been modeled twice within the non-linear mixed-effect approach (Karlsson, et al., 2000; Kjellsson, Ouellet, Corrigan, & Karlsson, 2008). In both cases the link functions adopted were logit functions and the events to model were dichotomized. This strategy presents some drawbacks, which this thesis aims to remove. Next chapter will propose new link functions, the so-called 'multinomial logistic functions', as the appropriate tool to parameterize sleep data models.

3.4 Tables and figures

Table 3.1 Different relevant types of categorical data, with examples of the respective models used in the literature for their description in the population PK-PD modeling context.

Correlation considered	Data	Model
no	binary data	logistic models (Zeger & Liang, 1992)
	ordinal data	proportional (Sheiner L. B., 1994) or differential odds models (Kjellsson, Zingmark, Jonsson, & Karlsson, 2008), models for continuous data, models for binary data (Armstrong & Sloan, 1989; Stromberg, 1996; Sankey & Weissfeld, 1998)
	count data	Poisson models (Snoeck & Stockis, 2007)
yes/no	[repeated] time - to - [categorical] event data, repeated categorical events per time interval	time-to-event models (Cox, Veyrat-Follet, Beal, Fuseau, Kenkare, & Sheiner, 1999; Spruance, Reid, Grace, & Samore, 2004; Plan, Karlsson, & Karlsson, 2010)
yes	binary data	logistic models with Markov features (Gallop, Ten Have, & Crits-Christoph, 2006)
	ordinal data	proportional odds models with Markov features (Zingmark, Kagedal, & Karlsson, 2005; Ito, Hutmacher, Liu, Qiu, Frame, & Miller, 2008; Henin, et al., 2009)
	count data	Poisson models with Markov features (Traconiz, Plan, Miller, & Karlsson, 2009)
	nominal data	compartmental chains (Bergstrand, Soderlind, Weitschies, & Karlsson, 2009) Markov-chain models (Karlsson, et al., 2000)

Figure 3.1 The cumulative probability distributions obtained with different link functions used for transforming categorical data: the logit, the probit, the log-log and the complementary log-log.



Chapter 4

Multinomial logistic functions in a Markov-chain model of sleep architecture

Some attempts to apply a Markov-chain model for describing sleep data are reported in the literature: in 2002 Gregory and Cabeza (Gregory & Cabeza, 2002) used this approach for describing the internal architecture of sleep in rats. However, they modeled sleep as a two-state process, considering only REM and non-REM sleep. Kemp et al. (Kemp & Kamphuisen, 1986) modeled rates of transition among the various sleep stages, but either assumed constant rates throughout the night or estimated their dynamics by smoothing observed transition frequencies by hand. More recently, Karlsson et al. (Karlsson, et al., 2000) and Kjellsson et al. (Kjellsson, Ouellet, Corrigan, & Karlsson, 2008) proposed a Markov-chain model using a mixed-effect approach and modeling the transition probabilities among different stages as binary logistic functions.

Building models for sleep data using data from a clinical study is more demanding than most categorical data analyses. The most recent and interesting model choice for sleep data is the Markov-chain model, as measuring sleep each 30-second generates data where the observations are dependent on previous observations. The quite large number of categories, the many observations and the potential numerous study arms in a clinical study all adds to the complexity of the model building, apart from the complexity the data itself adds.

This chapter investigates the development of a mixed-effect Markov-chain model in which binary logistic functions are replaced by multinomial logistic functions, in order to strongly reduce the model building process. By the way, the new functions are also more suited to the polychotomous nature of the data to describe, as already introduced in Paragraph 3.2.2. Paragraph 4.1 describes the clinical study providing the learning dataset of this analysis, and the characteristics of the insomniac subjects involved. Paragraph 4.2 describes a base first-order Markov-chain model in which the new link functions are applied. Its performance in describing the data is also analyzed. This model and its evaluation have been recently published by the author of this thesis and

coauthors (Bizzotto, Zamuner, De Nicolao, Karlsson, & Gomeni, 2010). The model is then further developed removing the first-order assumption on the Markov structure and investigating several other features, some of which were already discussed by the authors of previously published sleep models. The evaluation of the final model and its use for covariate analysis are left to Chapter 5 and Chapter 6.

4.1 Materials

Data were obtained from a polysomnograghic (PSG) multi-centre, randomized, double-blind, placebo-controlled, parallel group study designed to investigate a new candidate drug (Figure 4.1). Male and female subjects diagnosed with primary insomnia were chosen as feasible candidates for the study. The eligibility of the subjects was determined on the basis of specific PSG variables (e.g., LPS, WASO) obtained after a screening period consisting of a first clinical screening visit followed by a 2-night PSG recording in a sleep laboratory. After a week of daily placebo administration, subjects were randomized in the study, each arm assuming placebo or two different doses of the drug for 28 days before bedtime. PSG was recorded in three occasions in two consecutive days (1-2, 13-14 and 27-28) for each arm of the study.

Subjects taking part in the selection for this study had a diagnosis of primary insomnia and insomnia symptoms for at least three months, according to the 'Diagnostic and Statistical Manual of Mental Disorders - Fourth Edition - Text Revision (DSM - IV - TR)' (American Psychiatric Association, 2000), criteria 307.42. For being included in the study, the mean of PSG variables obtained after the two screening nights had to fall within the following ranges:

- mean TST: between 240 and 390 minutes;
- mean LPS: at least 30 minutes and not less than 20 minutes on either night;
- mean WASO: 60 minutes or more and neither night less than 45 minutes.

The model described in this chapter is developed based only on the first night of the double-blind treatment nights, from M=116 patients treated with placebo. Since few epochs of stages 3 and 4 were reported, they were merged in a single stage called ‘slow-wave sleep’ stage. Hence, the stages considered are the awake stage (AW), stage 1 sleep (ST1), stage 2 sleep (ST2), slow-wave sleep (SWS) and REM sleep (REM), and the state space is $\mathcal{S}$ = {AW, ST1, ST2, SWS, REM}.

The sequence of sleep stages can be described as in Equation [4.29], where N = 960 is the number of samples for each subject and M = 116 is the number of subjects. In fact, PSG signals are recorded for 8 hours along the night and they are translated into sleep stages at each 30-second time interval, called ‘epoch’.

Age, gender and body mass index (BMI) were available for each patient. Demographic statistics are reported in Table 4.1.

4.2 A new Markov-chain model

This section is aimed at developing a new mixed-effect Markov-chain model for the data described above. The innovative element in this model is the introduction of multinomial logistic functions in place of the binary ones adopted in previous analysis of sleep data (Karlsson, et al., 2000; Kjellsson, Ouellet, Corrigan, & Karlsson, 2008).

4.2.1 Methods

Multinomial logistic functions

Let x_{it} represent the state (i.e. the sleep stage) of the *i*-th patient at epoch *t* and let each single realization of the process (i.e. each patient's sequence) obey to a first-order Markov-chain.

Let then the binary variable y_{ikmt} represent the transition of the *i*-th individual from state *k* at epoch (*t*-1) to state *m* at epoch *t*, that is:

$$y_{ikmt} = \begin{cases} 1 \text{ if } x_{i(t-1)} = k \text{ and } x_{it} = m, \text{ with } k, m \in \mathcal{S} \\ 0 \text{ otherwise} \end{cases} \quad [5.1]$$

Then, for given values of $k \in \mathcal{S}$ (i.e. the starting state of a transition) and for a given time $t \in \{1, 2, \dots, N\}$, the vector

$$\bar{y}_{ikt} = [y_{ikAWt}, y_{ikST1t}, \dots, y_{ikREMt}]^T \quad [5.2]$$

is a multinomial random variable representing all the possible transitions from state *k* at time *t*. This multinomial random variable is characterized by its probability vector:

$$\bar{p}_{ik}(t) = [p_{ikAW}(t), p_{ikST1}(t), \dots, p_{ikREM}(t)]^T, \quad [5.3]$$

where $p_{ikm}(t) = P(x_{it} = m | x_{i(t-1)} = k) \geq 0$ is the probability of moving from *k* to *m* at time *t*. Since $\bar{p}_{ik}(t)$ is the *k*-th row of the transition probability matrix characterizing the process at time *t*, than

$$\sum_{m \in \mathcal{S}} p_{ikm}(t) = 1. \quad [5.4]$$

The model for the transition from the state k of the chain is therefore

$$(\bar{y}_{ikt}|\bar{p}_{ik}(t)) \sim Multinomial(\bar{p}_{ik}(t)). \quad [5.5]$$

In our context, the transition probabilities represent the model parameters to be estimated according to the sleep data. To avoid estimates constrained between 0 and 1, such parameters are described through link functions. The natural way to identify parameters of a multinomial distribution is to transform them into multinomial logit functions, which can be defined as the following for all $i \in \{1, ..., M\}$, k, m and $r \in \mathcal{S}$, and $t \in \{1, ..., N\}$:

$$g_{ikm\_r}(t) = \log \frac{p_{ikm}(t)}{p_{ikr}(t)}. \quad [5.6]$$

For each triple (i, k, t), a certain value of r - called 'reference state' - is taken from $\mathcal{S}$, and all the values in $\mathcal{S}$ are taken for m. In such a way, 5 different logit functions are defined, one of which results to be equal to zero. Using the different logits and recalling that the sum of the probabilities conditional on $x_{i,t-1} = k$ is equal to one, the transition probabilities $p_{ikm}(t)$ are expressed as 'generalized logistic functions' for the triple (i, k, t) and all the values of m in $\mathcal{S}$:

$$p_{ikm}(t) = \frac{\exp\left(g_{ikm\_r}(t)\right)}{\sum_{m\in\mathcal{S}} \exp\left(g_{ikm\_r}(t)\right)}. \quad [5.7]$$

Hence, instead of the probability vectors $\bar{p}_{ik}(t)$ defined in Equation [5.3], the parameters of the model are the corresponding logit vectors:

$$\bar{g}_{ik\_r}(t) = \left[g_{ikAW\_r}(t), g_{ikST1\_r}(t), \dots, g_{ikREM\_r}(t)\right]^T, \quad k \in \mathcal{S}, \mathrm{i} \in \{1, \dots, M\}, t \in \{1, \dots, N\}, \quad [5.8]$$

which fully characterize the model. Note that if no correlation is assumed between logits with different values for k, i.e. for different stages of departure, the model can be divided into five different smaller models. Each sub-model, referred to as 'sub-model k', describes the transitions from a specific sleep stage and its parameters can be identified separately from the others.

Each sub-model is estimated using a non-linear mixed-effect approach for taking the variability of the population into consideration. That is, each individual logit $g_{ikm\_r}(t)$ is thought to be normally distributed around its typical value:

$$g_{ikm\_r}(t) \sim N\left(g_{km\_r}(t), \omega^2{}_{km\_r}(t)\right), \quad [5.9]$$

where $g_{km}(t)$ is the typical value for the logit and $\omega^2{}_{km}(t)$ is the variance of the inter-individual distribution. Considering the vector-matrix notation

$$\bar{g}_{ik\_r}(t) \sim N\left(\bar{g}_{k\_r}(t), \Omega_{k\_r}(t)\right), \quad [5.10]$$

where $\bar{g}_{ik\_r}(t)$ is the vector of Equation [5.8] and is assumed normally distributed around the vector of population values for the logit functions $\bar{g}_{k\_r}(t)$ with covariance matrix $\Omega_{k\_r}(t)$.

Nighttime as a predictor of logits

The model parameters are assumed time-varying considering the Markov-chain model as non-homogeneous. Temporal dependence are implemented assuming that parameters are piecewise linear functions of nighttime as suggested by (Karlsson, et al., 2000). In order to choose the nighttime values of the knots (called 'break-points') between the different line segments, Karlsson et al. proposed two different criteria:

- equally spaced break-points;
- equi-informative break-points such that all intervals contain the same amount of data.

The data available for this project were initially modeled with a Markov-chain model parameterized as in Karlsson et al. (Bizzotto, Zamuner, Nucci, Cobelli, & Gomeni, 2008), i.e. using binary logistic functions instead of the multinomial ones. In that context, it was found that the selection of equi-informative break-points guaranteed better convergence properties as well as a more accurate description of night dynamics. Accordingly, six equi-informative nighttime values for the break-points, collected in the vector

$$BP = [BPA, BPB, BPC, BPD, BPE, BPF]^T, \quad [5.11]$$

are fixed in the model.

The typical individual logits at the break-points can be expressed with the vector:

$$\bar{g}_{ikm\_r} = \left[\_, \_, \_, \_, \_rE, g_{km\_rF}\right]^T, \quad [5.12]$$

where $g_{km\_rA}, g_{km\_rB}, g_{km\_rC}, g_{km\_rD}, g_{km\_rE}$ and g_{km_rF} are the typical individual values of the logit *km* at times *BP*. The individual logits at the break-points can be expressed with the vector:

$$\begin{aligned} &\bar{g}_{ikm\_r} \\ &= \left[g_{ikm\_rA}, g_{ikm\_rB}, g_{ikm\_rC}, g_{ikm\_rD}, g_{ikm\_rE}, g_{ikm\_rF}\right]^T \\ &= \left[\left(g_{kmr\_A} + \hat{g}_{ikm\_r}\right), \left(g_{kmr\_B} + \hat{g}_{ikm\_r}\right), \left(g_{km\_rC}\right.\right. \\ &\left.\left. + \hat{g}_{ikm\_r}\right), \left(g_{km\_rD} + \hat{g}_{ikm\_r}\right), \left(g_{km\_rE} + \hat{g}_{ikm\_r}\right), \left(g_{km\_rF}\right.\right. \\ &\left.\left. + \hat{g}_{ikm\_r}\right)\right]^T, \end{aligned} \quad [5.13]$$

where $\hat{g}_{ikm\_r}$ is the individual deviation from the logit *km* and is an Empirical Bayes Estimate (EBE) (Equation [4.16]). It is implicit in this equation that the individual deviations of logit functions from the typical values are constrained to be constant at the different break-points, meaning that

$$\Omega_{k\_r}(t) = \Omega_{k\_r} \quad \forall t \in BP. \qquad [5.14]$$

Moreover, a diagonal Ω_{k_r} is assumed. These hypothesis were formulated in order to avoid over-parameterization and the consequent over-fitting of experimental data. In the end, the parameters to be identified are $\bar{g}_{km\_r}$ and $\Omega_{k\_r}, \forall\, k, m \in \mathcal{S}$.

Other model features

Each individual is assumed to be awake in the first epoch of the night. No correlations are assumed between parameters with different values for k (i.e. regarding transitions from different sleep stages), so that each sub-model k, $k \in \mathcal{S}$, can be identified separately from the others.

The reference state r in the definitions of the logit functions (Equation [5.6]) is different from one sub-model to the other. Considering a generic sub-model k, we choose as reference state the sleep stage r in $\mathcal{S}$ such that y_{ikrt} equals 1 less frequently across t and i. In such a way, we avoid identifying logarithmic functions with nearly zero arguments, whose sensitivity is high, thereby obtaining more robust estimates of fixed and random components of the logits.

Each sub-model is identified using a maximum likelihood approach. Considering sub-model k, the likelihood contribution from individual i is given by

$$L_{ik\_r} = \int \left[\prod_{t:x_{i(t-1)}=k} p_{ikm(t)} \left(\bar{g}_{ik\_r}(t) \right) \right] h(\eta_i | \Omega_{k\_r}) d\eta_i. \qquad [5.15]$$

where h is a multivariate normal density with zero mean and covariance matrix Ω_{k_r}.

The overall likelihood function L_{k_r} for sub-model k is then the product of the contributions from all the individuals. The likelihood L_{k_r} is maximized using the Laplacian method in NONMEM version VI (ICON Development Solutions) (Beal & Sheiner, 1998), with the centering option.

Model evaluation

During model development, each sub-model is evaluated through inspection of the objective function value (OFV), which provides a statistic indicator of the relevance of inclusion of new parameters in a model (see Paragraph 3.1.4). Precision of parameter estimates is checked looking at the variance-covariance matrix of the parameter estimates.

Goodness-of-fit inspection is then performed by comparing the observed transition frequencies $\hat{f}_{km}$, $\forall\, k, m \in \mathcal{SS}$ (Equation [4.31]) when all data are merged and grouped by time with the average of individual transition proba-

bility estimates merged and grouped in the same manner. Since observed frequencies of transitions $\hat{f}_{km}$ from stage k to m at epoch t is computed just on subjects which are in stage k at time t-1, the transition frequencies $\hat{f}_{kmj}$ are evaluated over time periods, indexed by j, each covering one tenth of night time; the average of individual transition probability estimates is computed on the same time intervals and the same subjects. The choice of 10 intervals of equal width (48 minutes) is based on two key features: to adequately follow sleep physiological variations across the night and to include a significant amount of data in each interval for computing confident frequencies.

In order to provide precision on estimated transition frequencies, the 90% confidence intervals are computed as

$$\hat{f}_{kmj} \pm 1.645 \sqrt{\frac{\hat{f}_{kmj}\left(1-\hat{f}_{kmj}\right)}{SO_{kj}}}, \qquad [5.16]$$

where SO_{kj} is the total number of observations of stage k in the time period j across all subjects (Bland, 2000). However, this evaluation ensures only the agreement between naïve pooled means of predictions and observations. It cannot be used, instead, for evaluating the goodness of population parameter estimates.

A further evaluation of sub-model performance after its estimation is achieved comparing the distributions of individual posterior estimates with the estimated between-subject variability. To this end, p-values for the null hypothesis that the $\hat{g}_{ikm\_r}$ values are distributed around the components of $\bar{g}_{km\_r}$ are computed $\forall\, k, m \in \mathcal{S}$ and checked. Moreover, the η-shrinkage S_{km_r} (Karlsson & Savic, 2007) is computed $\forall\, k, m \in \mathcal{S}$ as follows:

$$S_{km\_r} = 1 - \frac{SD\left(\hat{g}_{ikm\_r}\right)}{\omega^{m}_{k\_r}}, \qquad [5.17]$$

where SD is a standard deviation and $\omega^{m}_{k\_r}$ are the square roots of the diagonal elements of Ω_{k_r}.

Once all the sub-models are assessed, identified, and evaluated, they are merged together in a unique model. This is used to produce, through Monte-Carlo simulation, 100 new datasets with the same number of individuals as the original one.

The model performance in the prediction of the sleep aggregated parameters (WASO, LPS, etc.) is then tested through the implementation of the 'simplified Posterior Predictive Check' (sPPC), as suggested by (Karlsson, et al., 2000). Since sleep aggregated parameters (or at least some of them) are usually considered as the 'efficacy endpoints' of clinical studies aimed to test efficacy of hypnotic drugs, this term will also be used as synonymous of 'aggregated parameters'. The individual values of each of these parameters, derived from the observed data, are compared to the corresponding values computed from

the simulated data. In particular, for any given parameter, the median of the individual values is computed in each dataset (observed or simulated) and the relative deviations of medians are calculated as follows:

$$RelativeDeviation = \frac{MedianEndpoint_{simulated} - MedianEndpoint_{observed}}{MedianEndpoint_{observed}}. \quad [5.18]$$

For each endpoint, the distribution of relative deviations is computed and plotted in box-whisker plots. Simulations and goodness-of-fit plots are produced using the R package (R 2.10.1 from the R Development Core, 2009). The considered parameters of interest are the following:

- Latency to Persistent Sleep (LPS),
- Wake After Sleep Onset (WASO),
- Total Sleep Time (TST),
- time spent in each stage (tAW, tST1, tST2, tSWS, tREM),
- time spent in non-REM sleep (tNREM),
- sleep efficiency in 0-2 hours of bed time (SE1),
- 2-4 hours of bed time (SE2),
- 4-6 hours of bed time (SE3),
- 6-8 hours of bed time (SE4),
- mean extension of each sleep stage (meanAW, meanST1, meanST2, meanSWS, meanREM),
- number of transitions to each stage (nAW, nST1, nST2, nSWS, nREM).

4.2.2 Results

The 5 specified sub-models are identified according to the assumptions described above. For 4 of them, NONMEM is unable to compute the variance-covariance matrix of the parameter estimates due to numerical problems in the estimate of the Hessian matrix. Therefore bootstrap analysis is performed for each sub-model, producing the results shown in Table 4.1, where population value estimates and their 90% confidence intervals are reported. The same information is used to build plots in Figure 4.2, showing the 5 time-profiles of the transition probabilities from a specific sleep stage. These plots

highlight the property enjoyed by Markov-chain models built on multinomial logistic functions: by construction, probabilities sum up to one. Moreover, the reliability of likelihood maximization is confirmed by non-zero gradient values for the likelihood surface in the maxima and the narrow confidence intervals. The narrowest intervals are obtained from the model describing transitions from stage 2 (ST2) and REM sleep, while the largest ones are obtained when considering transitions from slow-wave-sleep stage (SWS), for which less data were available.

In most cases there is no need for zero diagonal elements of Ω_{k_r}, as shown in Table 4.3, where values are reported together with their confidence intervals. Such values are used to simulate the distribution of individual probability temporal profiles, which are shown in Figure 4.3 for some transitions chosen as examples. In the same plots, other information is shown with the aim of performing the goodness-of-fit inspection previously described in the 'Model evaluation' section above. A very good match is obtained between mean observed frequencies and mean post-hoc estimates, which lay inside the frequency confidence interval with very few exceptions. The p-values for the null hypothesis that the means of the $\hat{g}_{ikm\_r}$ values, computed $\forall\, k, m \in \mathcal{S}$, are null are close to 1. The estimated η-shrinkage parameters are reported in Table 4.4: their magnitude is high, i.e. >25%, in 4 cases out of 16. The transitions mostly affected by shrinkage are the ones leaving SWS: shrinkage results greater than 40% for the logit value defined on the ratio between the probability of staying in SWS and the probability to move to REM sleep.

Simplified Posterior Predictive Check indicates a good agreement between simulated and observed efficacy endpoints in most cases (Figure 4.4). Only 3 out of 23 median aggregated parameters computed from the real study falls outside the ranges of the median values for the simulated studies: each of them, i.e. median time spent in SWS (tSWS), median number of awakenings (nAW) and median number of transitions to SWS (nSWS), is under-predicted. Figure 4.5 splits night time into 10 consecutive intervals: in each of them the median relative frequency of the various sleep stages is represented for both the simulated datasets and the observed one. Such representation gives some hints on how the deviations observed for the efficacy endpoints result from a precise temporal dynamics. In particular, in the first 48 minutes it shows an under-prediction of wake time covered by SWS, while thereafter a constant slight over-prediction of wake time is covered by a little under-prediction of SWS, sooner, and ST2, later.

4.2.3 Discussion

This is the first time multinomial logistic functions have been employed in the implementation of mixed-effect Markov-chain models for analyzing nominal polychotomous pharmacodynamic data.

Two main assumptions are adopted. The first one is the definition of a unique time-independent inter-individual source of variability in the specification of each logit function. In this way, for what concerns between-subject variability identification, time frames where more raw data are observed carry their informative content to frames featuring fewer data. Moreover, for a specific subject, time frames where most observations are reported play a major role in the definition of the post-hoc estimate of the corresponding logit time-profile, which is constrained to be parallel to the typical profile.

The second assumption is the description of the sequence of sleep stages as a first-order Markov chain. The validity of this assumption is assessed via a simulation-based approach: in fact, sPPC generally shows good adherence between simulations and observations. However, the observed discrepancies between predictions and the real dataset suggest that the model may be further improved. Focusing, for example, on the awake stage, the short predicted median time spent in this stage is likely due to the lack of at least one long period of such stage during the night. Also, very short periods of continuous wake (even one epoch only) are less easily predicted, causing an under prediction of the number of awakenings. These aspects can be highlighted also for slow-wave sleep and, even to a certain extent, for stage 2 and stage 1 sleep. In this regard, previous works on Markov-chain models for sleep architecture (Karlsson, et al., 2000; Kjellsson, Ouellet, Corrigan, & Karlsson, 2008) found a significant stage time influence on most transition descriptions, where stage time was defined as the time elapsed since last change in sleep stage. Keeping into account stage times could reduce the deviations described above, in particular if transition probabilities are longer for shorter stage times and shorter for longer ones. Although alternative models may deserve consideration, the first-order Markov chain considered here allows us to perform a first evaluation of multinomial logistic functions in the mixed-effect Markov approach. Also, it produces realistic transition probability descriptions which are immediately interpretable by clinicians and sleep specialists at either estimation or prediction level.

The analysis of the identification steps for the different sub-models reveals that the proposed approach is generally robust provided that an adequate sample size is guaranteed. In this respect, the sub-model SWS, the least supported by data (especially in the second part of the night), exhibits the worst parameter uncertainty and shrinkage; moreover, large between-subject variability is estimated, together with a strong deviation between the typical prob-

ability profile and the observed transition frequency. It is not easy to understand whether these outcomes are entirely due to physiological reasons, but I conjecture that the proposed constrains on variability of logit functions is still insufficient to avoid over-fitting. In such a case, the reduction of the number of break points and sources of variability may lead to some improvement. Improvements could also be obtained by including more PSG night records in the dataset, which is possibly one of the first steps for a further development of the proposed modeling approach.

The proposed implementation of multinomial logistic functions reduces the number of sub-models to be identified: from 20 sub-models using the binary-logit approach (Karlsson, et al., 2000) to 5 sub-models in the new approach (one for each sleep stage). The total number of parameters to be estimated with the two approaches is the same for both typical and random effects. Despite longer run-times, the reduced number of sub-models strongly simplifies model building. Moreover, multinomial logistic functions ensure that the sum of all probabilities of transitions starting from a certain stage is equal to one, both at the individual and the typical subject level. In the binary-logit approach such a constraint was enforced from estimating one transition probability as 1 minus the sum of all the previously estimated transition probabilities from the same sleep stage (Karlsson, et al., 2000): however, it is not generally true that this can be always done and such a methodology does not provide precise information on the shape of the individual profile distribution for the lastly estimated transition probability.

4.3 The final Markov-chain model

Different models have been proposed to characterize the time-course of sleep stages in groups of individuals, in particular those by Karlsson et al. (Karlsson, et al., 2000), Kjellsson et al. (Kjellsson, Ouellet, Corrigan, & Karlsson, 2008) and Bizzotto et al. (Bizzotto, Zamuner, De Nicolao, Karlsson, & Gomeni, 2010). The latter is the one described in Paragraph 4.2. All of them used Markov chains for describing the time-course of transition probabilities between sleep stages in insomnia patients. Karlsson et al. introduced Markov-chain models as tools for describing sleep data, included the so-called 'stage time effects' and parameterized the models through binary logistic functions. The model presented above as 'base model' introduced multinomial logistic functions instead of the binary ones, without including stage time effect. Kjellsson et al. described initial sleeplessness as a new sleep stage, and estimated the knots of the piece-wise linear binary logits instead of fixing them. These are, briefly, the strengths of the existing models for the time-course of sleep stages, which will be clarified in Paragraph 4.3.1.

The objective of the following part of the chapter is therefore to build on and combine these features and add possible additional components, in order to improve the base multinomial Markov-chain model previously proposed. The improvement will be directed to facilitate the assessment of covariate and drug effects and to reduce potential model biases.

4.3.1 Methods

Each step in model development is tested through sPPC (described above in Paragraph 4.2.1, section 'Model evaluation') and through parsimony criteria, i.e., log-likelihood ratio (LLR) test, with hierarchical structures, and Bayesian information criterion (BIC), with non hierarchical structures. Consistency between sub-models is always preferred when these criteria are suggesting slightly different developments in the different sub-models.

Model reduction is attempted by decreasing the number of knots (breakpoints) in the piecewise logit functions and zeroing some transition probabilities. Removal of model biases is instead pursued acting on different model features, first of all the value of the reference state r to be used for each triple (i,

k, t) in Equation [5.6]. Then, significance of values different from zero is tested for each variance-covariance element in the full $\Omega_{k\_r}(t)$. Since internal evaluation showed some misspecifications related to SWS (see Paragraph 4.3.2), and SWS epochs often follow or precede ST2 epochs, a new sub-model is introduced by merging sub-models ST2 and SWS: in this new sub-model correlation terms are tested between individual values of logits defined on ST2 and SWS leaving stages. Finally, in order to convey a more physiological characterization of sleep architecture, two model features implemented by Karlsson et al. (Karlsson, et al., 2000) and Kjellsson et al. (Kjellsson, Ouellet, Corrigan, & Karlsson, 2008) in their Markov-chain models with binary logit functions are introduced in this model (with multinomial logit functions) and tested on our data. The main purpose of such features is to relax the first-order assumption made on the Markov chain. The first feature is letting the logits depend also on other variables, in addition to nighttime: both time elapsed since the last change in sleep stage ('stage time') and time elapsed in a sleep stage since the nighttime beginning (the latter never tested in the literature) are attempted. The second feature is the differentiation of the model behavior between initial sleeplessness and rest of nighttime.

The identification of the sub-models is performed again using NONMEM VI (ICON Development Solutions) (Beal & Sheiner, 1998).

4.3.2 Results

Reference stage

The exploration of the value of the reference state r to be used in $g_{ikm\_r}(t)$ brings to the choice of the same value used for k. Accordingly, r disappears from the logit notation, and Equation [5.6] can be rewritten as

$$g_{ikm}(t) = log\frac{p_{ikm}(t)}{p_{ikk}(t)}, \qquad [5.19]$$

and Equation [5.8] becomes

$$\bar{g}_{ik}(t) \sim N\left(\bar{g}_k(t), \Omega_k(t)\right). \qquad [5.20]$$

Transition probabilities fixed to zero

Transitions for which probability can be fixed to zero are chosen according to their observed frequency $\hat{f}_{km}$(Equation [4.31]). The chosen frequency

threshold is 0.1%. Consequently, the number of logits in each sub-model is reduced as reported in Table 4.5.

Nighttime break-points

The number of break-points in the piecewise linear logit functions of nighttime is reduced from 6 to 3, *BPA*, *BPB* and *BPC*: *BPA* and *BPC* are placed at the nighttime beginning (epoch 2) and end (epoch 960), respectively, *BPB* is estimated in each sub-model as a new parameter (with no inter-individual variability), as suggested in (Kjellsson, Ouellet, Corrigan, & Karlsson, 2008) for binary logit functions. Consequently, the individual logits at the break-points can be expressed with the vector:

$$\bar{g}_{ikm} = [g_{ikmA}, g_{ikmB}, g_{ikmC}]^T = [(g_{kmA} + \hat{g}_{ikm}), (g_{kmB} + \hat{g}_{ikm}), (g_{kmC} + \hat{g}_{ikm})]^T, \quad [5.21]$$

where $g_{kmA}, g_{kmB}, g_{kmC}$ are the typical individual values of the logit km at times BP = [BPA, BPB, BPC] and $\hat{g}_{ikm}$ is the individual deviation from this logit.

Once nighttime break-points are introduced, the matrices $\Omega_k(t)$, $k \in \mathcal{S}$, in Equation [5.20] are replaced by Ω_k, with elements

$$\omega^2{}_{kmn} = cov(\hat{g}_{ikm}, \hat{g}_{ikn}), \qquad m, n \in \mathcal{S}_k. \quad [5.22]$$

Inter-individual variability

The search for triples (k, m, n) bringing to values of $\omega^2{}_{kmn}$ statistically different from zero in the full Ω_k brings to the use of the following variance-covariance matrices:

$$\Omega_{AW} = \begin{bmatrix} \omega^2{}_{AWST1ST1} & \omega^2{}_{AWST1ST2} & 0 \\ \omega^2{}_{AWST2ST1} & \omega^2{}_{AWST2ST2} & 0 \\ 0 & 0 & \omega^2{}_{AWREMREM} \end{bmatrix},$$

$$\Omega_{ST1} = \begin{bmatrix} \omega^2{}_{ST1AWAW} & \omega^2{}_{ST1AWST2} & \omega^2{}_{ST1AWREM} \\ \omega^2{}_{ST1ST2AW} & \omega^2{}_{ST1ST2ST2} & \omega^2{}_{ST1ST2AW} \\ \omega^2{}_{ST1REMAW} & \omega^2{}_{ST1REMST2} & \omega^2{}_{ST1REMREM} \end{bmatrix},$$

$$\Omega_{ST2} = \begin{bmatrix} \omega^2{}_{ST2AWAW} & 0 & 0 & 0 \\ 0 & \omega^2{}_{ST2ST1ST1} & 0 & 0 \\ 0 & 0 & \omega^2{}_{ST2SWSSWS} & 0 \\ 0 & 0 & 0 & \omega^2{}_{ST2REMREM} \end{bmatrix}, \quad [5.23]$$

$$\Omega_{SWS} = \begin{bmatrix} \omega^2{}_{SWSAWAW} & 0 & 0 \\ 0 & \omega^2{}_{SWSST1ST1} & 0 \\ 0 & 0 & \omega^2{}_{SWSST2ST2} \end{bmatrix},$$

$$\Omega_{REM} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \omega^2{}_{REMST1ST1} & 0 \\ 0 & 0 & \omega^2{}_{REMST2ST2} \end{bmatrix}.$$

No significant improvements are achieved by introducing correlation terms between individual logits in sub-model ST2-SWS (unification of sub-models ST2 and SWS): therefore, these parameters are not included in the final model.

Stage time effect

Stage time t_s is assumed to modify each logit at its three nighttime break-points according to an additive piecewise linear model. Three break-points for each sub-model k are again chosen for the stage time effect, $s_{km}(t_s)$: *BPsa* at t_s = 1 epoch (the minimum stage time that can be observed), *BPsc* at the maximum observed time elapsed since the last change in state k, and *BPsb* considered as a model parameter constrained in the interval (*BPsa*, *BPsc*). Therefore, the vector of individual logits at the nighttime break-points (Equation [5.21]) becomes function of t_s:

$$\begin{aligned} \bar{G}_{ikm}(t_s) &= [G_{ikmA}(t_s), G_{ikmB}(t_s), G_{ikmC}(t_s)\,] \\ &= [(g_{kmA} + \hat{g}_{ikm} + s_{km}(t_s)), (g_{kmB} + \hat{g}_{ikm} + \\ & s_{km}(t_s)), (\, g_{kmC} + \hat{g}_{ikm} + s_{km}(t_s))]. \end{aligned} \qquad [5.24]$$

The introduction of time elapsed in a sleep stage since nighttime beginning does not produce significant improvements: therefore, this predictor is not included in the final model.

Initial sleeplessness

In sub-model AW, the 8-hour nighttime is divided into 2 parts: the first part ranges from the beginning of nighttime to t = IS, where IS is the first epoch of the night in which a non-awake state is observed in a specific subject, and the second one covers the remaining part of the night. In the second time interval the logits are modeled as previously described changing only the position of the first nighttime break-point: *BPA* = IS. During initial sleeplessness new logits are modeled, again as piecewise linear functions, but without inter-individual variability or stage time effects. In particular, three additional break-points are defined: *BP1* at nighttime beginning (epoch 2), *BP3* at the maximum IS observed in the data for the specific sub-model, and the central *BP2* considered as a model parameter. The feasibility of using IS as the first epoch of persistent sleep is also tested, but it is not supported by the data.

A final model file, the one for estimating sub-model AW, is shown in Appendix A as an example, and some lines of the data file used for that control stream are shown in Appendix B. Condition numbers for the final sub-models are in the range 6.9 – 25.8 (they were not available for the base sub-models

(Paragraph 4.2.1) since the R matrix, i.e. the inverse Hessian, cannot be computed in those cases).

The estimated parameter values are shown in Table 4.6. Eight parameters, involved in the computation of logits defined on ratios close to zero in sub-model AW, are fixed to -10, a value which is close enough to zero in terms of probability ratios. When these parameters were not considered fixed, they had high CVs and likelihood minimization by NONMEM was rarely successful. However, they cannot be discarded because they are involved in limited time intervals (nighttime or stage time), and the piecewise linear functions of time needs to be defined in their entire domain.

Transition probability profiles are computed from the estimated parameters and are shown in Figure 4.6. Estimated stage time effects are shown in Figure 4.7.

In each sub-model, an important reduction in OFV is achieved, with respect to the base model (see Table 4.7). Most of OFV reduction is due to the introduction of stage time effect: in cases where covariance elements and initial sleeplessness differentiation still had to be introduced, the implementation of stage time effect produced in the sub-models AW, ST2, SWS, ST1 and REM reductions of 5110, 1275, 874, 88 and 61 points, respectively. The decrease of OFVs amounts to some tens when introducing the final reference state for the logits (Equation [5.19]), covariance elements for inter-individual variability (Equation [5.23]) and initial sleeplessness differentiation for the logits in sub-model AW. Similar outcomes are obtained using BIC. It is not possible to evaluate the effect of fixing one transition probability to zero (in each sub-model, ST2 excluded) with either OFVs or BIC, since the few observations for which transitions assumed impossible actually happened had to be removed.

4.3.3 Discussion

The work just presented (Paragraph 4.3) is dedicated at combining the best features of similar models available in the literature for the development of a Markov-chain model able to describe transitions between sleep stages through multinomial logistic functions. One of these models is the one previously described (Paragraph 4.2), published in 2010 (Bizzotto, Zamuner, De Nicolao, Karlsson, & Gomeni, 2010) and used here as a base for model development. The aim of the new analysis is to improve the model predictive performance while preserving model simplicity, according to the principle of par-

simony. The latter is instrumental to developing second stage models accounting for covariate and drug effects.

The major change adopted during the model development process, with respect to the base model, is the introduction of stage time (time elapsed since the last change in sleep stage) as a predictor of logit values, in addition to nighttime. Its multiplicative effects on ratios of transition probabilities are found to change greatly during stage time, so that the sensitivity of the logits to stage time is comparable or even higher than the sensitivity to nighttime. For this reason, the degrees of freedom in the parameterization of the relationship between logit functions and nighttime can be set as equal to the degrees of freedom in the relationship with stage time (break points number in piecewise linear functions of nighttime lowers from 6 to 3). And, for the same reason, it would be interesting to test if the inclusion of inter-individual variability (IIV) and covariates effects can significantly modify the individual profiles of stage time effects, besides the individual profiles of nighttime effects.

The choice to select the sleep stage recorded at epoch t-1 as the reference stage in the definition of logits at epoch t allows an easy interpretation of plots of stage time effects vs. stage time: increasing values in the profiles indicate higher probability to exit from the current sleep stage, and vice versa. These profiles result to be approximately L-shaped in many cases, meaning that transitions to new states happen with higher probability in the first minutes than later on. However, high early transition probabilities may be partially related to sleep scoring difficulties when sleep stages are changing (Karlsson, et al., 2000). As exceptions to the 'L-shape rule', there are transitions which become likely again (U-shape) when stage time reaches high values: this is the case for transitions from ST1 and ST2 to REM, from SWS to AW and from REM to ST2.

Since it is shown that median ST1 time (1 epoch only) anticipates the decline in ST1 time effect, and that ST1 is the stage with higher probability of transitioning to other stages during nighttime, it can be claimed that stage 1 sleep is a state of 'fast transition' towards other more stable sleep states (Carskadon & Dement, 1989). Another way to explain this is thinking that the separation of the physiological stages of sleep in 5 states (AW, ST1, ST2, SWS and REM) is the discretization of a continuum done with some degree of grossness. SWS, for example, is already explicitly used as a state in which characteristics of stage 3 sleep and stage 4 sleep are aggregated together (Rechtschaffen & Kales, 1968). Similarly, each of the five labels used for the recorded sleep stages likely aggregates an interval of different characteristics changing on a continuous domain. The recorded sleep stages can be thought as refractory aggregated states, or as observable discrete states on the top of a layer of hidden continuous (or at least more refined) states. A more refined discretization of sleep states (and of nighttime) would probably make unnecessary the use of stage time effects (i.e., semi-Markov models). According to this hypothesis, the degree of aggregation of underling more refined states seems lower in ST1 than in other states. The same can be thought about REM

sleep, since in this case stage time effect was estimated to be quite flat over stage time. In fact, the removal of the first order assumption on our Markov-chain model was less important for ST1 and REM, as confirmed by the small values of OFV drop in the two sub-models.

Inclusion of stage time effect and of the other features described in the results section produces an improvement of model predictive performance, which is shown in the chapter dedicated to model evaluation (see Paragraph 5.1.3). Among the others, the introduction of specific parameters for initial sleeplessness results to be important. Incidentally, initial sleeplessness can be thought as a sixth sleep stage (Kjellsson, Ouellet, Corrigan, & Karlsson, 2008), which cannot be observed after the first epoch of sleep occurs (the AW state, instead, can be seen thereafter).

As part of model development, significant correlations between logits from a specific sleep stage are also investigated: the results highlight that diagonal variance-covariance matrices are not optimal in sub-models where the probability of staying in a state is not clearly dominating on the probabilities of transitioning to other states. Moreover, correlations between individual logits of sub-models ST2 and SWS are tested, since aggregated sleep parameters related to SWS are predicted with some bias (see sPPC, Figure 5.1). No improvements are obtained when including such correlations, but the bias can be justified as well, as specified later in Paragraph 5.1.3.

4.4 Tables and figures

Table 4.1 Statistics on age, BMI and gender.

Age[a]	BMI[a]	# M	# F
44 (18; 64)	26.9 (17.0; 33.8)	38	78

[a] Values are reported as median (min, max).

Table 4.2 Estimates of typical probabilities with 90% confidence intervals from bootstrap (base model).

Leaving stage k	Time t (min)	Typical probability values[a]				
		$p_{kAW}(t)$	$p_{kST1}(t)$	$p_{kST2}(t)$	$p_{kSWS}(t)$	$p_{kREM}(t)$
AW	1	0.98 (0.96; 0.99)	0.02 (0.01; 0.04)	0.00 (0.00; 0.00)	0.00 (0.00; 0.00)	0.00 (0.00; 0.00)
	33	0.90 (0.87; 0.92)	0.08 (0.06; 0.11)	0.02 (0.01; 0.03)	0.00 (0.00; 0.00)	0.00 (0.00; 0.00)
	109.5	0.72 (0.65; 0.67)	0.21 (0.17; 0.26)	0.05 (0.03; 0.08)	0.00 (0.00; 0.01)	0.01 (0.01; 0.02)
	246.5	0.73 (0.65; 0.79)	0.19 (0.14; 0.23)	0.06 (0.04; 0.08)	0.00 (0.00; 0.00)	0.02 (0.02; 0.04)
	386.5	0.73 (0.68; 0.78)	0.21 (0.17; 0.25)	0.03 (0.02; 0.05)	0.00 (0.00; 0.00)	0.03 (0.02; 0.05)
	480	0.85 (0.79; 0.89)	0.13 (0.10; 0.18)	0.01 (0.01; 0.02)	0.00 (0.00; 0.00)	0.01 (0.01; 0.02)
ST1	1	0.26 (0.21; 0.31)	0.42 (0.36; 0.46)	0.32 (0.27; 0.38)	0.00 (0.00; 0.01)	0.00 (0.00; 0.01)
	119	0.12 (0.10; 0.14)	0.32 (0.29; 0.36)	0.47 (0.44; 0.51)	0.00 (0.00; 0.00)	0.09 (0.07; 0.11)
	221	0.11 (0.09; 0.14)	0.35 (0.32; 0.38)	0.43 (0.38; 0.47)	0.00 (0.00; 0.00)	0.11 (0.09; 0.14)
	321.5	0.10 (0.08; 0.13)	0.41 (0.38; 0.45)	0.37 (0.31; 0.40)	0.00 (0.00; 0.00)	0.12 (0.10; 0.14)
	406	0.12 (0.10; 0.14)	0.40 (0.36; 0.43)	0.36 (0.34; 0.41)	0.00 (0.00; 0.00)	0.12 (0.10; 0.15)
	480	0.16 (0.14; 0.20)	0.48 (0.43; 0.53)	0.27 (0.22; 0.31)	0.00 (0.00; 0.00)	0.09 (0.05; 0.11)
ST2	1	0.04 (0.04; 0.05)	0.03 (0.02; 0.04)	0.85 (0.83; 0.87)	0.07 (0.06; 0.09)	0.00 (0.00; 0.00)
	120.5	0.04 (0.04; 0.05)	0.03 (0.03; 0.04)	0.85 (0.84; 0.87)	0.06 (0.05; 0.07)	0.02 (0.01; 0.02)
	212	0.04 (0.03; 0.04)	0.03 (0.03; 0.04)	0.89 (0.88; 0.90)	0.03 (0.02; 0.04)	0.01 (0.01; 0.01)
	291.5	0.03 (0.03; 0.04)	0.02 (0.02; 0.03)	0.91 (0.89; 0.92)	0.02 (0.02; 0.03)	0.01 (0.01; 0.01)
	380	0.04 (0.03; 0.05)	0.03 (0.03; 0.04)	0.90 (0.89; 0.91)	0.01 (0.01; 0.01)	0.01 (0.01; 0.02)
	480	0.05 (0.05; 0.06)	0.04 (0.03; 0.05)	0.88 (0.87; 0.90)	0.00 (0.00; 0.01)	0.02 (0.02; 0.03)
SWS	1	0.01 (0.00; 0.03)	0.00 (0.00; 0.02)	0.29 (0.11; 0.48)	0.69 (0.27; 0.77)	0.00 (0.00; 0.41)
	74	0.02 (0.01; 0.03)	0.01 (0.00; 0.01)	0.15 (0.11; 0.20)	0.82 (0.77; 0.87)	0.00 (0.00; 0.00)
	125	0.02 (0.02; 0.03)	0.00 (0.00; 0.01)	0.19 (0.13; 0.28)	0.78 (0.70; 0.84)	0.00 (0.00; 0.00)
	187.5	0.02 (0.01; 0.03)	0.00 (0.00; 0.01)	0.16 (0.11; 0.24)	0.82 (0.69; 0.86)	0.00 (0.00; 0.02)
	274.5	0.01 (0.01; 0.02)	0.00 (0.00; 0.01)	0.33 (0.26; 0.46)	0.66 (0.45; 0.72)	0.00 (0.00; 0.00)
	480	0.04 (0.03; 0.07)	0.00 (0.00; 0.03)	0.38 (0.22; 0.51)	0.57 (0.39; 0.71)	0.00 (0.00; 0.01)

REM	1	0.04 (0.03; 0.06)	0.04 (0.02; 0.08)	0.05 (0.02; 0.13)	0.00 (0.00; 0.00)	0.86 (0.79; 0.90)
	180	0.04 (0.04; 0.06)	0.04 (0.03; 0.05)	0.01 (0.01; 0.02)	0.00 (0.00; 0.00)	0.90 (0.88; 0.91)
	276	0.04 (0.03; 0.05)	0.04 (0.03; 0.05)	0.01 (0.01; 0.02)	0.00 (0.00; 0.00)	0.90 (0.89; 0.92)
	346	0.05 (0.04; 0.06)	0.04 (0.03; 0.05)	0.01 (0.00; 0.01)	0.00 (0.00; 0.00)	0.90 (0.90; 0.92)
	414.5	0.05 (0.04; 0.07)	0.04 (0.03; 0.05)	0.01 (0.01; 0.02)	0.00 (0.00; 0.00)	0.89 (0.88; 0.91)
	480	0.05 (0.04; 0.08)	0.04 (0.02; 0.06)	0.01 (0.01; 0.02)	0.00 (0.00; 0.00)	0.90 (0.86; 0.92)

[a] Values have been rounded to the second decimal place.

Table 4.3 Estimates of inter-individual variability with 90% confidence intervals from bootstrap (base model). Reference states for the definition of logit functions are also reported.

Leaving stage k	Reference state r	Inter-individual variability of logit values[a]				
		$(\omega^{AW}{}_{k,r})^2$	$(\omega^{ST1}{}_{k,r})^2$	$(\omega^{ST2}{}_{k,r})^2$	$(\omega^{SWS}{}_{k,r})^2$	$(\omega^{REM}{}_{k,r})^2$
AW	SWS	1.04 (0.69; 1.31)	0.53 (0.34; 0.74)	0.93 (0.66; 1.30)		0 FIX
ST1	SWS	0.21 (0.09; 0.35)	0.21 (0.11; 0.28)	0.18 (0.11; 0.28)		0.41 (0.23; 0.56)
ST2	REM	0.18 (0.11; 0.24)	0.57 (0.38; 0.79)	0.12 (0.07; 0.16)	0.76 (0.55; 1.00)	
SWS	REM	0 FIX	0 FIX	3.78 (2.68; 4.78)	0.21 (0.10; 0.49)	
REM	SWS	0.66 (0.41; 0.99)	0.83 (0.60; 1.14)	0 FIX		0.24 (0.10; 0.36)

[a] No values are defined in correspondence of logits equal to zero by definition (see Equation [5.6]).

Table 4.4 Shrinkage values for the inter-individual variability parameters of the five estimated sub-models (base model).

Leaving stage *k*	Shrinkage values (%)[a]				
	S_{kAW_r}	S_{kST1_r}	S_{kST2_r}	S_{kSWS_r}	S_{kREM_r}
AW	10	20	20		n. d.
ST1	30	24	27		22
ST2	24	12	25	11	
SWS	n. d.	n. d.	9	47	
REM	21	20	n. d.		32

[a] No values are defined in correspondence of logits equal to zero by definition (see Equation [5.6]).
'n. d.' values are reported when inter-individual variability is fixed to zero.

Table 4.5 Logits for the different sub-models (final model).

AW	ST1	ST2	SWS	REM
0	1: $g_{iST1AW}(t)$	1: $g_{iST2AW}(t)$	1: $g_{iSWSAW}(t)$	1: $g_{iREMAW}(t)$
1: $g_{iAWST1}(t)$	0	2: $g_{iST2ST1}(t)$	2: $g_{iSWSST1}(t)$	2: $g_{iREMST1}(t)$
2: $g_{iAWST2}(t)$	2: $g_{iST1ST2}(t)$	0	3: $g_{iSWSST2}(t)$	3: $g_{iREMST2}(t)$
-	-	3: $g_{iST2SWS}(t)$	0	-
3: $g_{iAWREM}(t)$	3: $g_{iST1REM}(t)$	4: $g_{iST2REM}(t)$	-	0

Table 4.6 Model parameter values (final model).

Sub-model	Parame-ters	Parameter labels and values					
AW	logits at nighttime break-points	TVG1A	TVG2A	TVG3A	TVG1B	TVG2B	TVG3B
		-0.251	-2.66	-10 FIX	-0.209	-1.01	-2.86
		TVG1C	TVG2C	TVG3C			
		-0.203	-2.08	-2.1			
	logits at initial sleepless-ness break-points	TVG11	TVG21	TVG31	TVG12	TVG22	TVG32
		-5.76	-10 FIX	-10 FIX	-3.93	-7.63	-10 FIX
		TVG13	TVG23	TVG33			
		-4.86	-10 FIX	-10 FIX			
	stage time effects at ST[a] break-points	STE1b	STE2b	STE3b	STE1c	STE2c	STE3c
		-2.49	-3.59	-5.67	-6.63	-10 FIX	-10 FIX
	break-points	[b]BPA	BPB	[b]BPC	[b]BPsa	BPsb	[b]BPsc
		[c]IS	[d]0.0679	960 FIX	1 FIX	7.32	265 FIX
		[b]BP1	BP2	[b]BP3			
		2 FIX	13.00	371 FIX			
	variance-covariance for IIV	G1i	OMEGA(2,1)	G2i	G3i		
		0.134	-0.142	0.831	1.07		
ST1	logits at nighttime break-points	TVG1A	TVG2A	TVG3A	TVG1B	TVG2B	TVG3B
		-0.321	-0.0716	-4.15	-1.07	0.492	-1.01
		TVG1C	TVG2C	TVG3C			
		-1.05	-0.251	-1.24			
	stage time effects at ST[a] break-points	STE1b	STE2b	STE3b	STE1c	STE2c	STE3c
		-0.447	-0.52	-0.463	-0.634	-0.246	-3.31
	break-points	[b]BPA	BPB	[b]BPC	[b]BPsa	BPsb	[b]BPsc
		2 FIX	268	960 FIX	1 FIX	3.24	20 FIX
	variance-covariance for IIV	G1i	OMEGA(2,1)	G2i	OMEGA(3,1)	OMEGA(3,2)	G3i
		0.318	0.18	0.389	-0.0637	0.138	0.398
ST2	logits at nighttime break-points	TVG1A	TVG2A	TVG3A	TVG4A	TVG1B	TVG2B
		-2.46	-2.52	-1.71	-4.07	-2.57	-2.6
		TVG3B	TVG4B	TVG1C	TVG2C	TVG3C	TVG4C
		-3.33	-3.34	-2.13	-2.33	-4.59	-3.16
	stage time effects at ST[a] break-points	STE1b	STE2b	STE3b	STE4b		
		-0.905	-0.894	-1.19	-0.993		
		STE1c	STE2c	STE3c	STE4c		
		-1.79	-6.44	0.927	-6.75		
	break-	[b]BPA	BPB	[b]BPC	[b]BPsa	BPsb	[b]BPsc

	points	2 FIX	676	960 FIX	1 FIX	5.62	143 FIX
	variance-covariance for IIV	G1i	G2i	G3i	G4i		
		0.152	0.456	0.786	0.239		
SWS	logits at nighttime break-points	TVG1A	TVG2A	TVG3A	TVG1B	TVG2B	TVG3B
		-3.22	-4.83	-0.526	-3.31	-5.13	0.142
		TVG1C	TVG2C	TVG3C			
		-1.65	-4.55	0.389			
	stage time effects at ST[a] break-points	STE1b	STE2b	STE3b	STE1c	STE2c	STE3c
		-0.999	-0.939	-2.5	0.761	-3.53	-3.73
	break-points	[b]BPA	BPB	[b]BPC	[b]BPsa	BPsb	[b]BPsc
		2 FIX	715	960 FIX	1 FIX	5.9	103 FIX
	variance-covariance for IIV	G1i	G2i	G3i			
		0 FIX	1.4	1.35			
REM	logits at nighttime break-points	TVG1A	TVG2A	TVG3A	TVG1B	TVG2B	TVG3B
		-3.12	-2.87	-3.11	-3.25	-3.02	-4.6
		TVG1C	TVG2C	TVG3C			
		-2.96	-2.98	-4.56			
	stage time effects at ST[a] break-points	STE1b	STE2b	STE3b	STE1c	STE2c	STE3c
		0.351	-0.238	-1.22	0.566	-1.22	1.8
	break-points	[b]BPA	BPB	[b]BPC	[b]BPsa	BPsb	[b]BPsc
		2 FIX	640	960 FIX	1 FIX	13.6	100 FIX
	variance-covariance for IIV	G1i	G2i	G3i			
		0.584	0.849	1.1			

[a] ST stands for stage time.
[b] This parameter can be directly used as a constant in the abbreviated code of the $PRED routine (as shown in the NONMEM model file in Appendix A).
[c] IS is the individual initial sleeplessness length.
[d] Here BPB = IS + (960-IS) * 0.0679.

Table 4.7 OFVs for the 5 sub-models (base and final models).

Sub-model	Base model	Final model
AW	26983	21662
ST1	24264	24086
ST2	49984	48733
SWS	9341	8380
REM	14798	14687

Figure 4.1 Clinical study protocol.

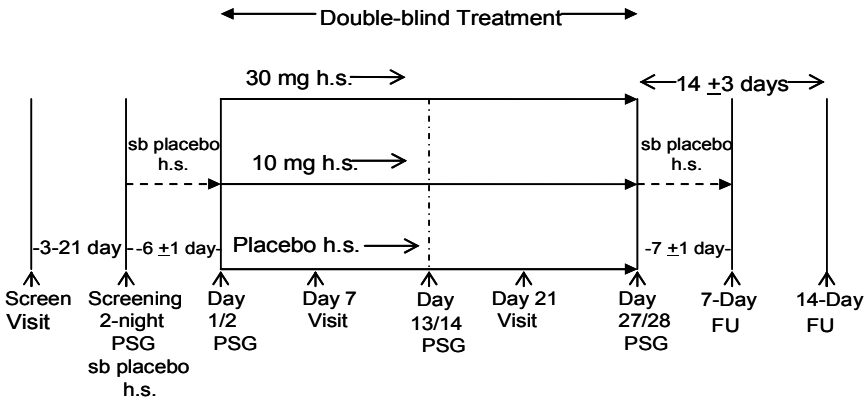


Figure 4.2 Estimated time course of typical transition probabilities with 90% confidence intervals from bootstrap (base model); each panel shows the transition probabilities from a specific sleep stage.

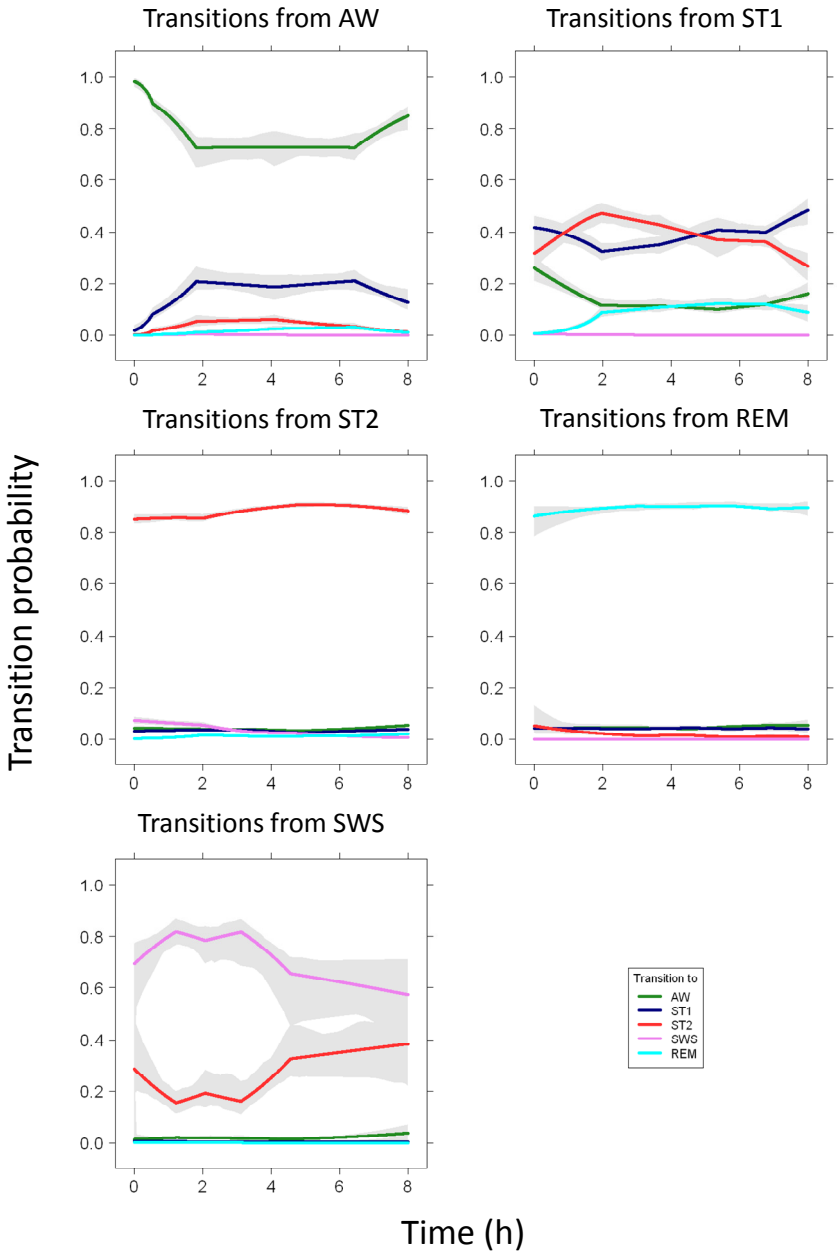


Figure 4.3 Observed time profiles of transition frequencies with 90% confidence intervals; distribution of post-hoc estimates (as mean and 5th-95th percentile); estimated profile of typical transition probability and 5th-95th percentile of the distribution of individual profiles (base model).

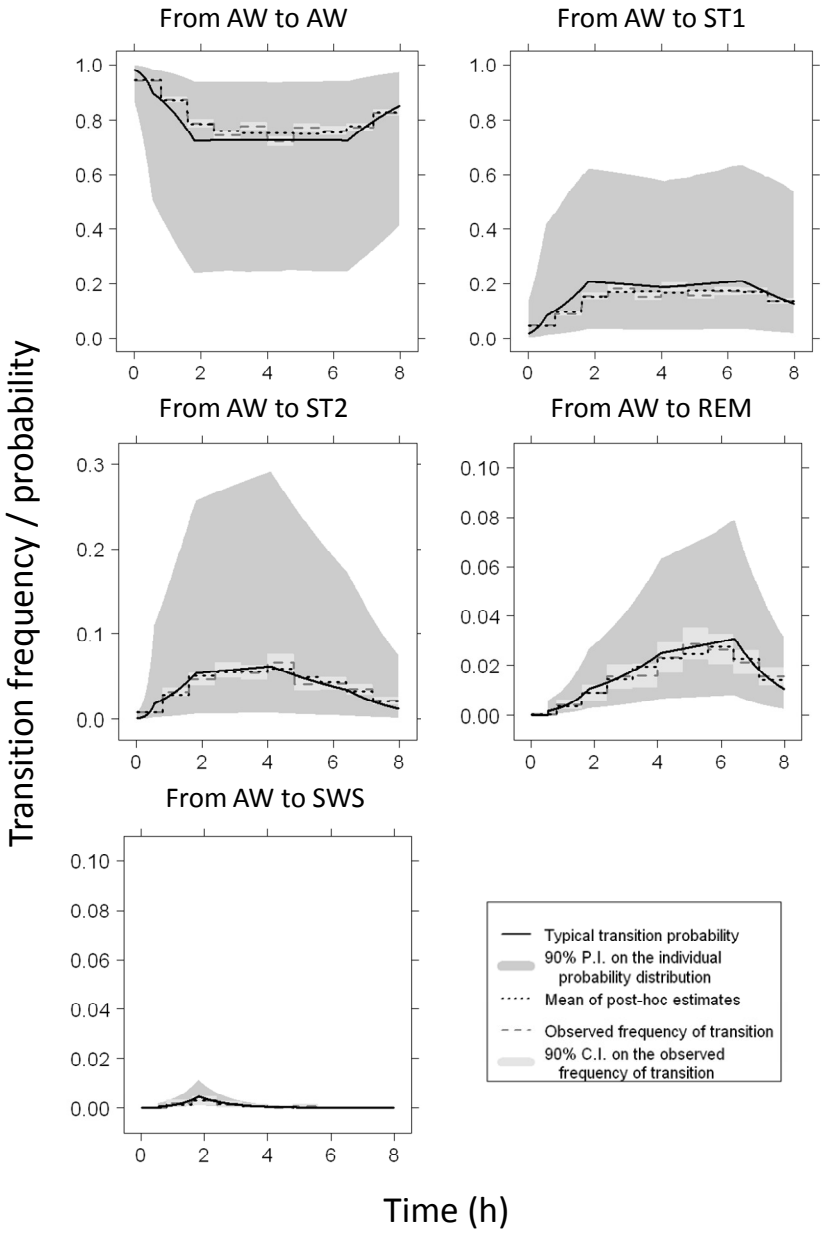


Figure 4.4 Results from posterior predictive check (base model): relative deviations of median efficacy endpoints in 100 simulated clinical studies from parameter medians in the real study.

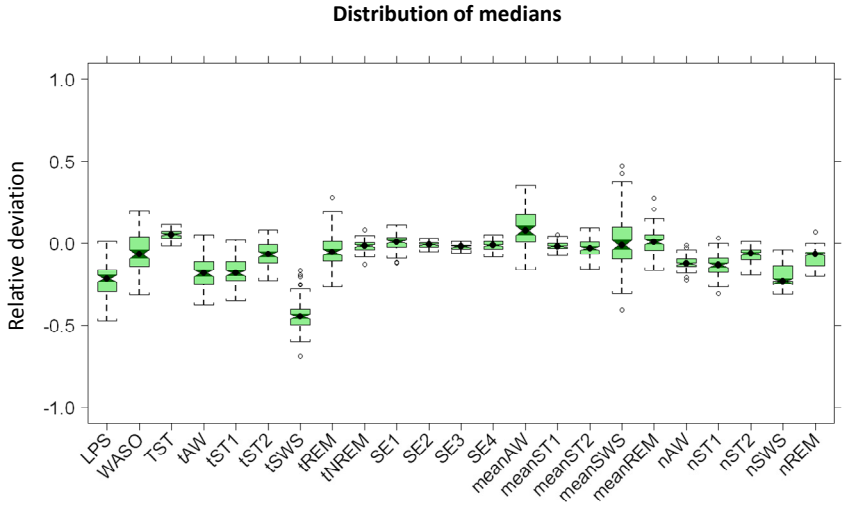


Figure 4.5 Relative frequency of each sleep stage during ten consecutive night periods in the simulated studies and in the observed one: median values are computed for each of 100 simulated studies and their medians are compared to the median value of the real population (base model).

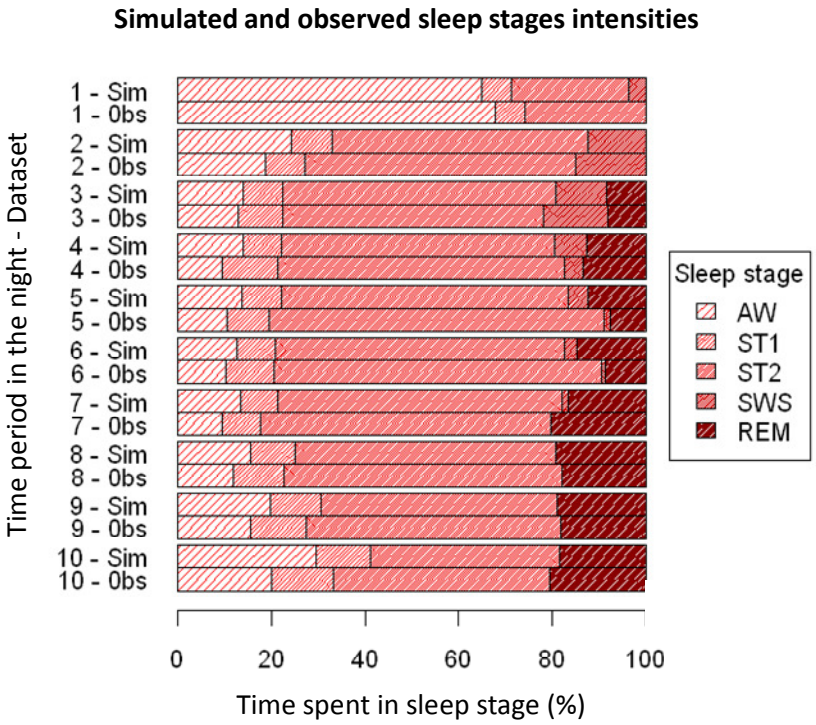


Figure 4.6 Probability profiles for all the transitions between sleep stages (final model). Their computation is done for the median stage times over the nighttime and the whole patient population.

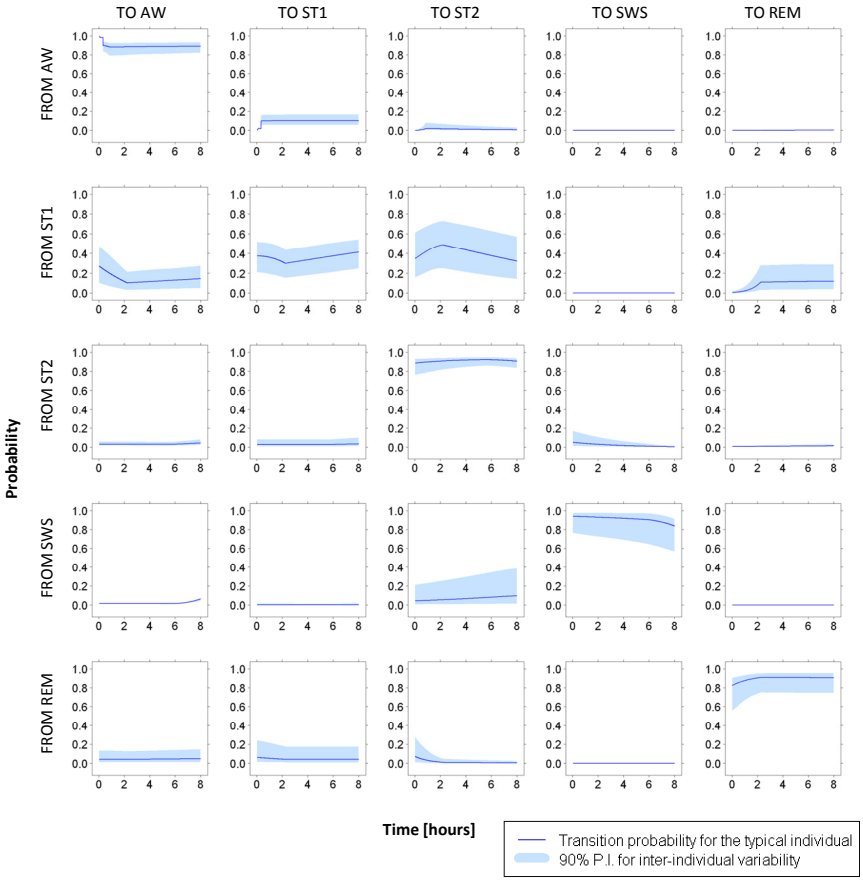
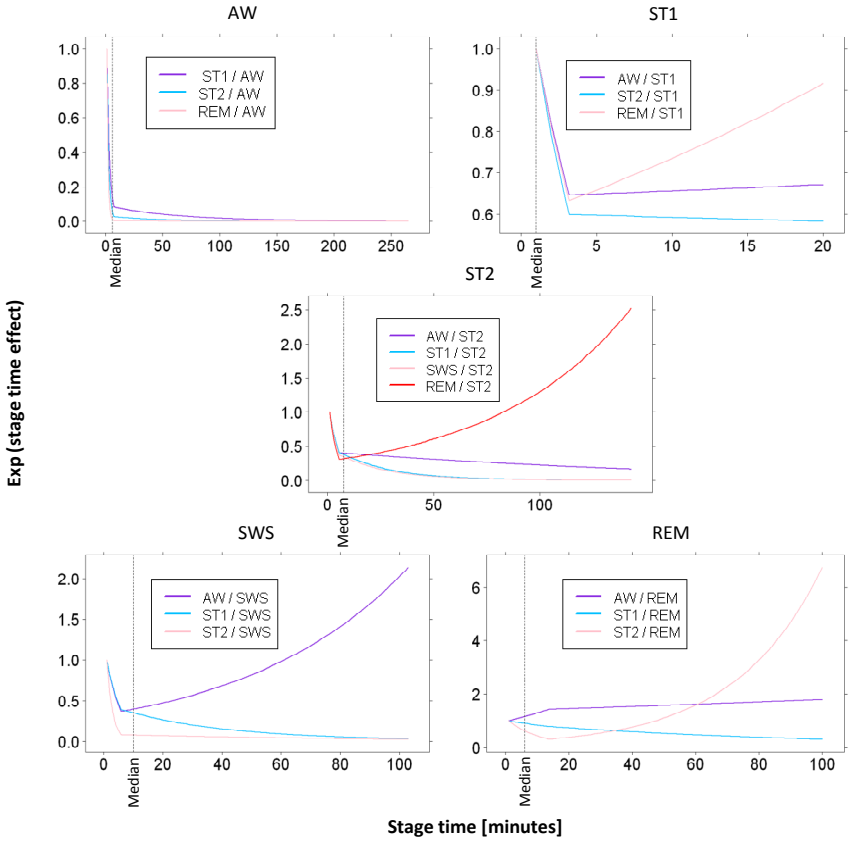


Figure 4.7 Stage time effects estimated in the different sub-models (final model). Exp(stage time effect) is used in order to visualize multiplicative effects on probabilities ratios, instead of additive effects on logits (less intuitive). Median stage times over the nighttime and the whole patient population are also reported in each plot.



Chapter 5

Model evaluation

Mixed-effect models are increasingly used in the pharmaceutical field to analyze and interpret clinical data in a population framework and to perform clinical trial simulations (CTS). In fact the application of CTS during drug development can help to achieve greater efficiency and dosage optimization. Specifically, CTS facilitates the decision making in drug development by assessing uncertainty of predicted trial performance and outcomes in planning of prospective trials (Kimko & Peck, 2011).

One of the major scopes in the development of the Markov-chain model presented in Chapter 4 was to provide a useful tool for clinical trial simulations. This goal is reached when the model capability of describing the data and, therefore, the underlying system is proved. This capability, referred to as 'model adequacy', needs to be thoroughly investigated during both model development and final model selection, in order to avoid the risk that adequate models are rejected and inadequate models are accepted (Karlsson & Savic, 2007).

The first objective of this chapter is therefore to investigate the adequacy of the final Markov-chain model with multinomial logistic functions. The analysis aims to check if the proposed model adequately interprets the underlying process, provides precise predictions for the parameters of interest and avoids biased estimates.

Model evaluation is already described above for the base Markov-chain model with multinomial logistic functions. Paragraph 4.2, especially the section 'Model evaluation' of Paragraph 4.2.1, introduces some evaluation methods for this specific kind of model, i.e.:

- inspection of objective function value (OFV);
- estimation of variance-covariance matrix of the parameter estimates or, equivalently, bootstrap performance;
- comparison between observations and individual model predictions;
- comparison between distribution of individual model predictions and estimated between-subject variability;
- simplified posterior predictive check (sPPC).

A basic form of model evaluation is described above also for the final Markov-chain model of this thesis: in Paragraph 4.3.1 it is stated that model development was guided by sPPC and parsimony criteria (LLR test and BIC), at each step of each sub-model improvement. This chapter aims to provide a more thorough model evaluation for the final Markov-chain model, using partly new diagnostic methods. In fact, model diagnostics have never been systematically investigated and discussed when Markov-chains and, more generically, categorical nominal data are involved. The model evaluation will be performed here using two approaches, referred to as internal and external evaluation. The latter differs from the former because it also uses a new set of real observations, not used for model building or for parameter estimation, and called 'validation dataset'. On the contrary, the data used for model development are called 'learning dataset'.

Internal evaluation is introduced and discussed in Paragraph 5.1. It is presented for the final model developed within this work. Therefore, considerations on the OFV evaluation are avoided in this chapter and left to Paragraph 4.3, since they are useful for guiding model development and not for assessing the final model adequacy. External evaluation is introduced and discussed in Paragraph 5.2, again for the final model.

5.1 Internal evaluation

Statistical theory indicates several diagnostic methods to evaluate model adequacy, most of them involving graphics. The work by Karlsson & Savic (Karlsson & Savic, 2007) presents a detailed overview of the different model diagnostics to be used in the context of mixed-effect modeling, highlighting for each of them pros and cons. The authors divide such methods into five categories:

(a) typical individual prediction-based methods,

(b) individual parameter estimates-based methods,

(c) residual-based diagnostics,

(d) numerical diagnostics,

(e) simulation-based diagnostics.

The (a) and (b) methods are essentially based on graphics: their aim is to visually evaluate whether there is agreement between the dependent variable (DV), i.e. the observations to describe, and population or individual model predictions, respectively. The first method can be misleading when applied to non-linear mixed-effect models and often indicates misspecified models even when they are adequate. The second one suffers from the opposite drawback: in case of sparse information in the individual data, it becomes difficult to recognize misspecified models since excellent predictions are usually produced ('perfect fit' phenomenon). The residual based diagnostics (c), consisting in checking population or individual residuals, are strictly connected to the latter methods and substantially have the same flaws.

Methods (a), (b) and (c), often referred to as 'goodness-of-fit inspection methods', are easily implementable but hardly applicable in the categorical data context, in which only probability of observations is estimated and, consequently, residuals are not considered. In fact, only the (b) method was applied in this work, for evaluating the sub-models of the base Markov-chain model (see section 'Model evaluation' of Paragraph 4.2.1 and Figure 4.3). But, as already mentioned, this method is rather poor and not powerful.

Several types of numerical diagnostics (d) can be implemented for checking models adequacy in case of categorical data. These methods are of essential importance for model comparison (e.g., LLR test) and for the evaluation of model robustness and the detection of possible over-fit (bootstrap and variance-covariance matrix of the parameter estimates). As mentioned in this chapter introduction, LLR test and BIC are not useful for evaluating the final model during the model development process. Bootstrap and variance-covariance matrix of the parameter estimates were not adopted for the final

Markov-chain model of this thesis, since other diagnostics, described below, were considered more appealing and appropriate.

Finally, the simulation-based methods (e) are considered the most interesting tools when dealing with categorical data. These methods appear very attractive because they preserve the same pros of the classical goodness-of-fit inspection methods without being affected by their cons. Such diagnostics consist in comparing a desired statistic derived from raw data with a reference distribution obtained trough stochastic simulation (Monte Carlo simulation). They are useful for showing potential model misspecifications, since simulations mimic real data behavior only when a model is adequate.

In the context of sleep data modeling, applying stochastic simulation-based methods becomes useful to check the model capabilities in predicting the sleep aggregated parameters (WASO, LPS, etc.) used in the clinical practice for quantifying the severity of sleep disorders, but mainly to evaluate the model predictability of sleep physiological patterns whose maintenance has been demonstrated to be relevant. The two inspections are performed here through the implementation of simplified Posterior Predictive Check (sPPC) on the aggregated parameters (Karlsson, et al., 2000) and 'Visual Predictive Check' (VPC) on two specific statistics derived from the data: the transition frequencies (never used before, to my knowledge) and the stage frequencies (Bergstrand, Hooker, & Karlsson, 2009).

However, bias and imprecision in the predictions may depend not only on possible model misspecifications but also on non robust estimation methods. In case of the Markov-chain model with multinomial logistic functions, which is highly non-linear and dealing with categorical data, this issue may be very relevant.

For this reason, a recent approach(Duval & Karlsson, 2002; Jonsson, Kjellsson, & Karlsson, 2004) for checking both the model adequacy and the estimation method performance is implemented here: the 'Stochastic Simulations followed by Estimation' (SSE) based diagnostics. SSE is used here for the visual comparison between the transition probabilities estimated from raw data and the confidence intervals on the correspondent transition probabilities estimated during SSE procedure. To my knowledge, this kind of visual comparison was never implemented before. I propose to call it 'Visual Estimation Check' (VEC), in order to emphasize the similarity with 'VPC' and highlight the focus on the evaluation of the estimation procedure.

5.1.1 Simulation-based methods

Once each different sub-model (awake, stage1, stage2, sws and rem) is identified using the maximum likelihood (ML) approach implemented in NONMEM VI, a set of population parameters, including both typical parameters (θ) and inter-individual variance-covariance (Ω), is available for the simulation step, to be repeated *n* times (*n* is chosen equal to 100).

The aim of this procedure is to generate from the developed model and its estimated parameters new datasets, each made of the PSG outcomes for *M* = 116 (as in the observed dataset) new patients. A model that includes all the 5 sub-models is therefore needed in order to produce a new sequence of sleep stages for each potential patient. This model is written for NONMEM VI and its code is presented in Appendix C. The code can be summarized through the following procedure.

1) The AW state is assumed at the beginning of the night for subject *i*: x_{it}= AW, with *t* = 1 epoch, and starting state *k* = AW.
2) *t* is increased by 1 and a sub-model, *sm*, is chosen according to the *k* value (*sm* = *awake* if *k* = AW, *stage1* if *k* = ST1, *stage2* if *k* = ST2, *sws* if *k* = SWS, *rem* if *k* = REM).
3) The transition probabilities at time *t* are simulated through the sub-model *sm*: the ML estimates obtained from the estimation step on sub-model *sm* are used to sample and re-construct the individual logits at time *t* and thereafter the corresponding individual transition probabilities (*pAW*, *pST1*, *pST2*, *pSWS* and *pREM*).
4) The 5 transition probabilities are placed side by side in a probability scale ranging from 0 to 1.
5) A random variable, called R, is drawn from a uniform distribution in [0, 1] and used to decide which of the 5 possible transitions occurs. If, for example, the transition to ST1 is chosen, than $y_{ikmt} = 1$, with *m* = ST1, and consequently x_{it} = ST1.
6) Variables providing stage times for each sleep stage and initial sleeplessness length (if a non-awake state has already occurred) are updated according to the value of x_{it}.
7) *k* assumes the value of x_{it}.
8) Steps 2 to 7 are repeated until *t* = 960 epochs.
9) Steps 1 to 8 are repeated for *i* in [1, …, *M*].

Dataset for simulation

A brief explanation of the datasets used as input files for the simulation model is provided here.

Each dataset (real and simulated) is initially formed of four columns (Table 5.1) indicating the subject identification number (ID), the PSG recording visit (Visit), the night time expressed in epochs (Time) and the corresponding sleep stage (STAGE), taking values in {0, 1, 2, 3, 5} (0 for AW, 1 for ST1, 2 for ST2, 3 for SWS, 5 for REM). Each column length is therefore M x N = 116 x 960.

During simulation, five more columns called MDV0, MDV1, MDV2, MDV3 and MDV5 ('MDV' stands for 'missing data value'), indicating whether the previous stage is 0, 1, 2, 3, 5 or not, are added to the dataset. For example, MDV0 is equal to 0 if the previous state is AW, and 1 otherwise. Furthermore, five binary columns called AW, ST1, ST2, ST3 and REM are added for indicating if the current state is 0, 1, 2, 3 or 5, respectively. For example, AW is equal to 1 if the current state is AW, and 0 otherwise.

These columns are useful during the estimation step for extracting relevant data for each single sub-model, and during the evaluation process for easily computing the statistics of interest. Furthermore, a column counting for the individual stage time (the number of epochs since the last transition) is added and called STT. Finally, due to the introduction of the Initial Sleeplessness feature for the *awake* sub-model, two more columns are needed: a binary column called SL ('sleep'), which is 1 from the epoch of first non-awake state up to the last epoch for each subject, and one called IS ('Initial Sleeplessness'), that reports the number of awake epochs before the first episode of non-awake (IS includes also the first non-awake epoch).

sPPC

sPPC is performed to assess the model capability in describing and simulating aggregated characteristics of PSG data in the population. This technique is used during the whole model development process and is described in the section 'Model assessment' of Paragraph 4.2.1.

VPC

The final model capability in describing the physiological evolution of the sleep stages and transitions along nighttime is tested through visual predictive check (VPC) (Holford, 2005). Two statistics computed on observed data together with the corresponding confidence interval derived from simulations (the ones used for sPPC) are plotted against nighttime t. These statistics are

- the frequencies of occurrence of each sleep stage, as proposed in (Bergstrand, Hooker, & Karlsson, 2009) for categorical data (Equation [4.30]);
- the transition frequencies between stages, never used before to my knowledge (Equation [4.31]).

Each statistic was computed for each of ten equal intervals in the night-time (48 minutes each).

VEC

Visual estimation check (VEC) is a novel approach to assess robustness and precision of parameter estimates in a mixed-effect model in terms of transition probabilities time-course, and it is introduced in this work. It relies on the combination of stochastic simulation & re-estimation (SSE)(Duval & Karlsson, 2002; Jonsson, Kjellsson, & Karlsson, 2004) and computation. Specifically, all of the 100 simulated datasets are re-identified, using the developed Markov sub-models. From each of the original and the simulated datasets, the estimated parameter values are used for computing the temporal profiles of typical transition probabilities, and for drawing and computing the temporal profiles of individual transition probabilities, from which 5th and 95th percentiles were derived. Consequently, an observed and 100 simulated profiles are obtained to evaluate three statistics: the typical transition probabilities and the 5th and 95th percentiles on inter-individual variability. At the end, the 95% confidence intervals for each statistic-profile derived from simulation are computed and visually compared with each observed statistic-profile.

5.1.2 Results

sPPC

sPPC outcome for the final model is presented in Figure 5.1 and indicates a good agreement between simulated and observed efficacy endpoints in most cases. Only 1 out of 23 median aggregated parameters computed from the real study falls outside the range of median values computed from the simulated studies. This parameter is the time spent in SWS (tSWS), which results under-predicted. Other sPPC plots are produced considering statistics different from the median (not reported here) and they corroborate the overall goodness of the model predictive performance, in terms of both typical outcomes and variability extent in the population.

VPC

Figure 5.2 shows the results of VPC implementation on transition frequencies, $\hat{f}_{km}$, and stage frequencies, $\hat{\rho}_k$. The plots show a general very good agreement between observed and simulated statistics. A slight bias can be detected on transitions from ST2 to REM, from REM to ST2 and from REM to REM, only in the very first period after light off. Simulation-based confidence intervals are generally narrow. The largest ones are observed for transitions from AW to AW and to ST1, from SWS to ST2 and to SWS (especially in the last hours of the night), and from REM to all sleep stages, only in the first hour

VEC

Figure 5.3 illustrates the results from VEC performed on the time-course of transition probabilities. In general, a very good agreement between profiles estimated from raw and simulated data is shown in these plots, with the exception of the transitions from REM to ST2 and from REM to REM at the beginning of the night. Probability confidence intervals on transitions from AW, SWS and REM are larger compared to the other ones and they vary according to the amount of available information (depicted in the last row of Figure 5.2).

5.1.3 Discussion

The Markov-chain model is internally evaluated through three complementary visual diagnostics on categorical data.

sPPC assesses the model capability in predicting overall sleep parameters (usually considered as efficacy endpoints in clinical studies) close to the observed ones.

Visual predictive check (VPC) focuses on the accuracy of sleep description along the independent variable (nighttime is tested here, stage time was not considered): since data are categorical, stage frequencies and transition frequencies, and the uncertainty on their prediction, are considered.

Visual estimation check (VEC) is introduced in this work as a new tool able to validate the capability of estimating accurate and precise model parameters through a graphic description of accuracy and precision on transition probabilities time-courses. The name 'visual estimation check' is chosen because the effect (during nighttime) of possible weaknesses in the estimation method can be visually checked, even if not easily distinguished from the effects of poten-

tial model misspecifications; however, the simultaneous use of VPC and VEC is recommended to overcome this kind of issue.

sPPC, VPC and VEC show that the employed model slightly suffers in a couple of scenarios:

- statistics with high variability despite similar sleep patterns are predicted with some bias (see, for example, the aggregated sleep parameters related to SWS);
- small amount of observations for a specific sleep stage determines small bias (if in the Markov-chain departure) or inflated uncertainty in the VPC and VEC outcomes.

Since outcomes from VPC and VEC are mostly similar, it can be claimed that slight bias and uncertainty in VEC are mostly due to the imperfection of model structure rather than to the estimation of its parameter values. Despite the slight bias just mentioned, the three diagnostic tools show an overall good performance of the developed Markov-chain model in describing the data, and of the employed estimation technique. The maximum likelihood estimator, with Laplacian approximation as implemented in NONMEM VI, is shown in the literature to suffer in case of high η-shrinkage (Kjellsson, 2008). In this work, despite 3 out of 15 values of η-shrinkage being greater than 25%, it reveals instead to be robust.

Comparing the base model and the final model in terms of sPPC (Figure 4.4 vs. Figure 5.1), a general refinement of the predictive performance on overall sleep parameters is depicted. The most significant improvements are obtained on latency to persistent sleep (LPS), number of transitions to AW (nAW) and to ST1 (nST1), and time spent in ST1 (tST1). The introduction of specific parameters for initial sleeplessness is particularly important for the improvement on LPS, while other improvements in the model predictive performance are likely connected mostly with inclusion of stage time effect.

5.2 External evaluation

External model evaluation is useful for understanding if a model built and evaluated on a certain learning dataset can be adopted for modeling other datasets with similar features. If so, the modeler can rely on the fact that the model is suitable for the data used to build the model itself, but also for the physiological process which produces those data and other potential data in similar conditions.

Moreover, this methodology allows to assess if the model misspecifications, or bias or imprecision in the parameter estimates, reported after estimation on the learning dataset, are likely due to a specific dataset or to a systematic issue intrinsic in the model.

Finally, external model evaluation can be applied for checking how the parameter estimates gained from the learning dataset can be different among different datasets.

The literature is quite poor with respect to external evaluation of mixed-effect models, especially when data are categorical. Therefore, the following analysis is considered innovative and experimental. More work will be useful to define a best practice in the external model evaluation procedure.

5.2.1 Materials

Data for external evaluation are obtained again from the placebo arm of a polysomnographic (PSG) multi-centre, randomized, double-blind, placebo-controlled, parallel group study designed to investigate a new candidate drug. The study and the candidate drug are different from those involved in model development and internal evaluation. The new study is called here ‘study B’, while the other one is called ‘study A’.

Study B followed a design similar to the one reported above (Paragraph 4.1) for study A. The only difference was in the inclusion criteria for the PSG parameters, described as follows. The mean sleep parameter values obtained in the two screening PSGs (with single-blinded placebo administration) had to fall within the following ranges:

- mean TST: between 240 and 390 minutes in study A and between 240 and 420 minutes in study B;
- mean LPS: at least 30 minutes and not less than 20 minutes on either night (study A), vs. at least 20 minutes, and not less than 15 minutes on either night (study B);
- mean WASO: 60 minutes or more and neither night less than 45 minutes (both studies).

As done for model development, only data from the first night of the double-blind treatment nights are used. The number of subjects involved, treated again with placebo, is now $M = 81$. Age, gender and body mass index (BMI) are available also for each patient from study B. Demographic statistics are reported in Table 5.1.

5.2.2 Methods

The external evaluation of the final model is performed applying the model to data from study B and looking at objective function values (OFVs), distributions of empirical Bayes estimates (EBEs), parameter values and sPPC. OFVs and EBEs distributions are computed for dataset B using each of the 5 sub-models in two different scenarios: using parameter values estimated from study A and using parameter values estimated from study B. sPPC is performed comparing aggregated parameters computed on study B with corresponding aggregated parameters computed on 100 datasets simulated from parameter values estimated on study B.

5.2.3 Results

The 5 final sub-models are successfully identified using dataset B. The last two columns of Table 5.4 show the estimated OFVs using parameter values estimated from study A and using parameter values estimated from study B. Distributions of EBEs in the two scenarios are not shown here, since η-shrinkage (Equation [5.17]) is high in most occasions (greater than 25%).

Final parameter estimates from study B are shown in Table 5.5. They are used to compute typical probability profiles along nighttime, at stage time = 1 epoch and at median stage time. These profiles are not shown here, as only few small differences can be detected in comparison with previously computed profiles. When using dataset B, variance estimates for inter-individual variability are instead strongly reduced: averages of variances on the logits are reduced by 75%, 39%, 34%, 24%, 18% in sub-models SWS, ST1, REM, AW, and ST2, respectively.

sPPC on median aggregated parameters from the new study is visualized in Figure 5.1. The performance looks very similar to the one shown in Figure 5.1 on data from study A. WASO, tAW and tSWS are even slightly better predicted. The simultaneous comparison with median efficacy endpoints computed from dataset A (Figure 5.1, red dots) highlights a reduced predictive performance for the aggregated parameters which are highly variable in the two studies (see tST1, tSWS, meanAW, meanSWS and nREM).

5.2.4 Discussion

The final Markov-chain model is evaluated on a dataset (from study B) which was not used in the model development procedure. The validation dataset includes less subjects (81 vs 116), whose characteristics are similar to those of the original dataset (from study A). Ten subjects from study B would have been excluded from study A according to its inclusion criteria: these subjects would not have been severe enough, since their TST and LPS values were roughly 5 minutes above and 5 minutes below, respectively, compared with the inclusion criteria of study A.

Nevertheless, the Results paragraph (5.2.3) shows that the proposed model can adequately describe also the new data: the parameters estimated in the two datasets are similar, and the OFVs differ of maximum 300 when using study B with parameter estimates from likelihood maximization on study A or on study B. The new final parameter estimates for the typical individual are used to compute typical probability profiles along nighttime (plots not shown). Few typical probabilities of staying in the different sleep stages are just slightly different compared to the corresponding probabilities estimated from the original dataset:

- transitioning from AW to ST1 is slightly more likely, at about 1-2 hours from light off and short AW stage time;
- moving from ST1 to ST2 is less likely during all nighttime;

- transitioning from ST2 to AW is more likely, in the last hour before light on;
- moving from SWS to ST2 is more likely during all night, at low SWS stage time.

The last difference likely impacts on lower time spent in SWS (tSWS) and mean extension of SWS (meanSWS), detected through sPPC.

5.3 Tables and figures

Table 5.1 Dataset for simulation.

ID	Visit	Time	STAGE
1	3	1	0
1	3	2	0
1	3	3	1
⋮	⋮	⋮	⋮
1	3	960	3
2	3	1	0
⋮	⋮	⋮	⋮

Table 5.2 Dataset after simulation.

ID	Time	STAGE	MDV0	MDV1	...	MDV5	AW	ST1	...	REM	STT	SL	IS
1	1	0	1	1		1	1	0		0	0	0	4
1	2	0	0	1	...	1	1	0	...	0	1	0	4
1	3	1	0	1		1	0	1		0	2	1	4
1	4	0	1	0		1	1	0		0	1	1	4
⋮								⋮					⋮
2	1	0	1	1		1	1	0		0	0	0	30

Table 5.3 Statistics on age, BMI and gender (study A vs. study B).

Study	Age[a]	BMI[a]	# M	# F
A	44 (18; 64)	26.9 (17.0; 33.8)	38	78
B	45 (19; 65))	24.8 (18.7; 34.0)	33	48

[a] Values are reported as median (min, max).

Table 5.4 OFVs for the 5 sub-models, using data from study B.

Sub-model	Using parameter values from estimation on study A	After parameter estimation on study B
AW	11627	11434
ST1	19752	19511
ST2	33772	33441
SWS	5299	5214
REM	8906	8811

Table 5.5 Model parameter values (study B).

Sub-model	Parameters	Parameter labels and values					
AW	logits at nighttime break-points	TVG1A	TVG2A	TVG3A	TVG1B	TVG2B	TVG3B
		-0.304	-4.16	-10 FIX	0.0194	-2.26	-3.01
		TVG1C	TVG2C	TVG3C			
		-0.23	-3.71	-1.95			
	logits at initial sleep-lessness break-points	TVG11	TVG21	TVG31	TVG12	TVG22	TVG32
		-5.62	-10 FIX	-10 FIX	-3.97	-354	-10 FIX
		TVG13	TVG23	TVG33			
		-4.53	-10 FIX	-10 FIX			
	stage time effects at ST[a] break-points	STE1b	STE2b	STE3b	STE1c	STE2c	STE3c
		-2.68	-4.15	-8.35	-6.62	-10 FIX	-10 FIX
	break-points	[a]BPA	BPB	[a]BPC	[a]BPsa	BPsb	[a]BPsc
		[a]IS	[b]0.164	960 FIX	1 FIX	7.47	265 FIX
		[a]BP1	BP2	[a]BP3			
		2 FIX	36.76	371 FIX			
	variance-covariance for IIV	G1i	OMEGA(2,1)	G2i	G3i		
		0.0457	0.0266	1.1	0.657		
ST1	logits at nighttime break-points	TVG1A	TVG2A	TVG3A	TVG1B	TVG2B	TVG3B
		-1.01	-1.02	-5.77	-1.68	0.0468	-1.94
		TVG1C	TVG2C	TVG3C			
		-1.16	-0.484	-1.69			
	stage time effects at ST[a] break-points	STE1b	STE2b	STE3b	STE1c	STE2c	STE3c
		-0.342	-0.18	-0.513	-1.48	-1.28	-0.146
	break-points	[a]BPA	BPB	[a]BPC	[a]BPsa	BPsb	[a]BPsc
		2 FIX	240	960 FIX	1 FIX	3	20 FIX
	variance-covariance for IIV	G1i	OMEGA(2,1)	G2i	OMEGA(3,1)	OMEGA(3,2)	G3i
		0.118	0.0338	0.146	-0.0612	0.0724	0.426
ST2	logits at nighttime break-points	TVG1A	TVG2A	TVG3A	TVG4A	TVG1B	TVG2B
		-2.81	-1.96	-2.03	-4.24	-2.67	-1.94
		TVG3B	TVG4B	TVG1C	TVG2C	TVG3C	TVG4C
		-3.47	-3.32	-0.735	-1.51	-2.7	-2.32
	stage time effects at ST[a] break-points	STE1b	STE2b	STE3b	STE4b		
		-1.12	-1.24	-1.44	-0.94		
		STE1c	STE2c	STE3c	STE4c		
		-2.86	-4.21	1.13	-3.95		
	break-points	[a]BPA	BPB	[a]BPC	[a]BPsa	BPsb	[a]BPsc
		2 FIX	889	960 FIX	1 FIX	6.35	143 FIX

	variance-covariance for IIV	G1i	G2i	G3i	G4i		
		0.109	0.18	1.22	0.144		
SWS	logits at nighttime break-points	TVG1A	TVG2A	TVG3A	TVG1B	TVG2B	TVG3B
		-3.97	-5.28	-0.0835	-2.19	-4.31	0.92
		TVG1C	TVG2C	TVG3C			
		-5.48	-19.7	0.894			
	stage time effects at ST[a] break-points	STE1b	STE2b	STE3b	STE1c	STE2c	STE3c
		-1.14	-0.639	-2.57	-0.657	-0.24	-4.21
	break-points	[a]BPA	BPB	[a]BPC	[a]BPsa	BPsb	[a]BPsc
		2 FIX	1370	960 FIX	1 FIX	518	103 FIX
	variance-covariance for IIV	G1i	G2i	G3i			
		0 FIX	0.256	0.423			
REM	logits at nighttime break-points	TVG1A	TVG2A	TVG3A	TVG1B	TVG2B	TVG3B
		-3.87	-2.68	-2.24	-3.84	-3.12	-4.3
		TVG1C	TVG2C	TVG3C			
		-2.89	-3.21	-4.06			
	stage time effects at ST[a] break-points	STE1b	STE2b	STE3b	STE1c	STE2c	STE3c
		0.29	-0.244	-2.06	1.12	-0.0565	-0.381
	break-points	[a]BPA	BPB	[a]BPC	[a]BPsa	BPsb	[a]BPsc
		2 FIX	657	960 FIX	1 FIX	8.99	100 FIX
	variance-covariance for IIV	G1i	G2i	G3i			
		0.391	0.466	0.807			

[a] ST stands for stage time.
[b] This parameter can be directly used as a constant in the abbreviated code of the $PRED routine (as shown in the NONMEM model file in Appendix A).
[c] IS is the individual initial sleeplessness length.
[d] Here BPB = IS + (960-IS) * 0.164.

Figure 5.1 Results from posterior predictive check (final model): relative deviations of median efficacy endpoints in 100 simulated clinical studies from parameter medians in the real study.

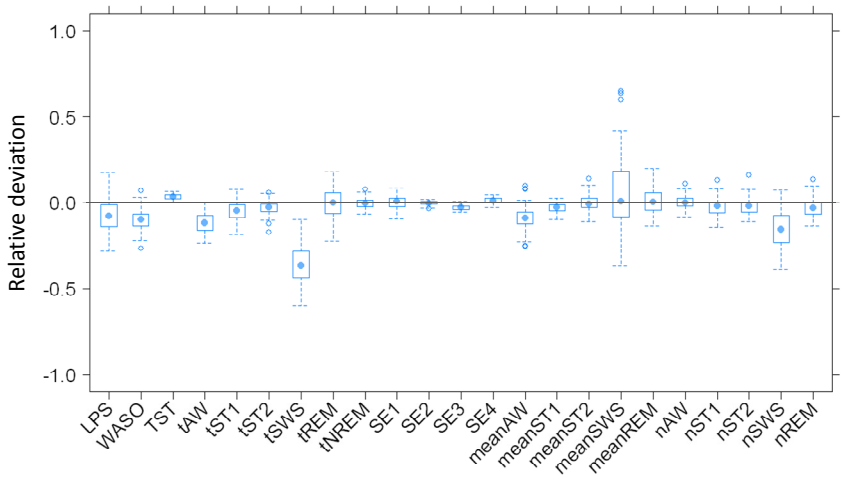


Figure 5.2 Results from visual predictive check (final model) on frequency of transitions (first 5 rows) and stage frequencies (last row). Note that range of y-axis values is larger in plots at positions (4, 3), (4, 4), (6, 1) and (6, 3).

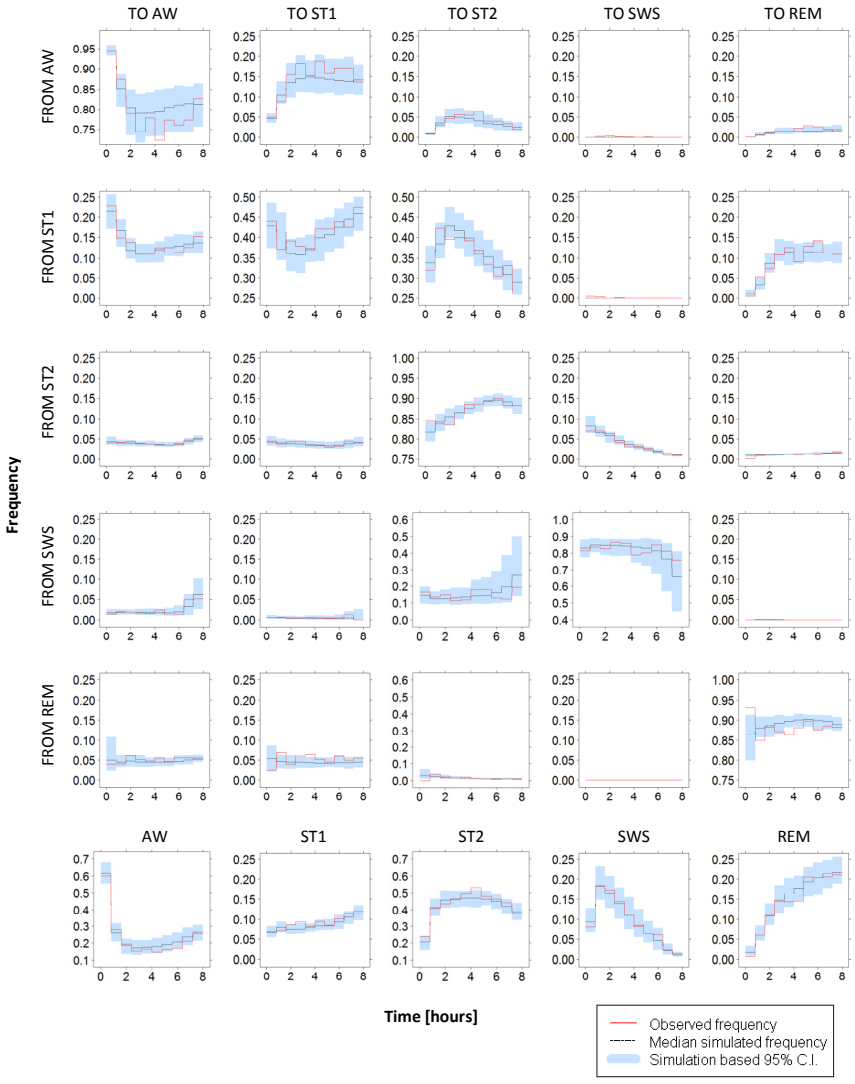


Figure 5.3 Results from visual estimation check on transition probabilities (final model). Note that two different scales are used for y-axis in the different plots. Note also that mixed-color (violet) areas come from the superimposition of pink and blue confidence intervals (see the typical and 5th and/or 95th confidence intervals for sub-models AW, SWS and REM).

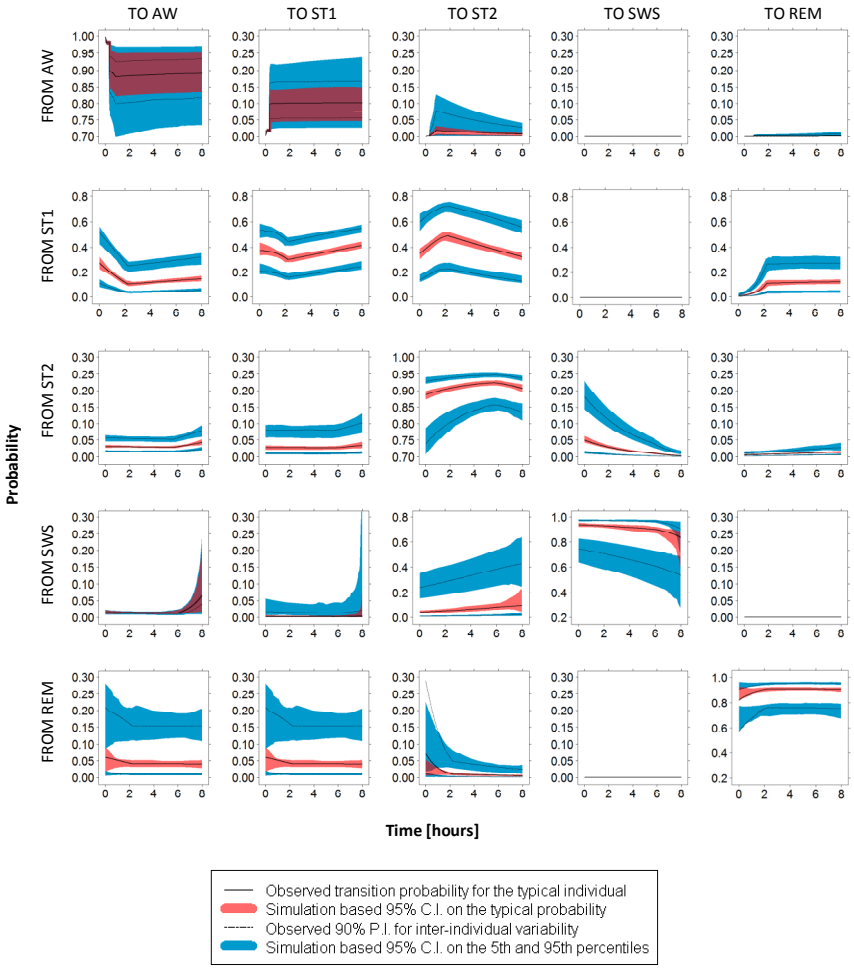
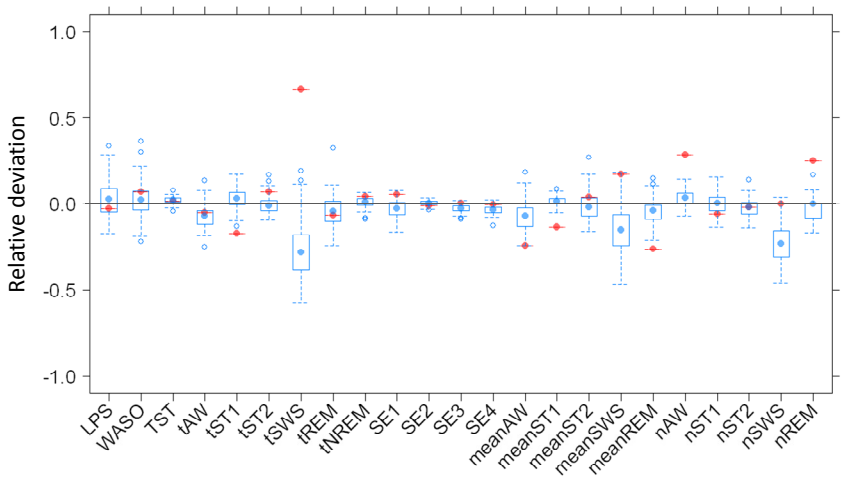


Figure 5.4 Results from sPPC on dataset B: median aggregated parameters computed on dataset B are compared with corresponding median aggregated parameters computed on 100 datasets simulated from parameter values estimated on dataset B. Comparison is shown in terms of relative deviation. Represented parameters are described in the section 'Model evaluation' of Paragraph 4.2.1. Red dots are depicted for visualizing the relative deviation of median values computed from study A, from median values computed from study B.



Chapter 6

Covariate analysis

It is often useful to explain the variability in a parameter using a covariate model that describes the relations between covariates and parameters. A covariate model can be used for identification of patient subpopulations. In fact, in addition to reduce the unexplained variability, a major advantage of the development of a covariate model is to limit potential risks of sub-therapeutic or undesirable effects by individualizing the treatment and the initial dosage regimen in the different subpopulations. Furthermore, such a model is useful for identifying the need for and aiding the design of new studies in the drug development process. If the identified covariate relationships are in line with the literature or prior expectations, the covariate model supports the structure of the other parts of the model. Thus, the development of a covariate model may also be viewed as a component of the model evaluation (Ribbing, 2007).

In an exploratory analysis, there are often a large number of covariates available which would be interesting to test on one or more population model parameters (see Equation [4.10] for the definition of population model). A subset of the potential covariate relations is often selected for the final model. The actual selection of which relations to include can be made after investigating the results of including each of them in the model. These procedures are collectively called 'selection within NONMEM'. Alternatively, the outcomes of the different relations can be investigated outside NONMEM, after which one or several selected models are investigated within NONMEM. The latter alternative is called 'selection outside NONMEM'.

The main advantage of selecting the covariate model outside NONMEM is that an investigation within NONMEM is computer intensive, resulting in long computer run-times and/or high demands on a computer grid. Therefore, a seemingly appealing approach is to perform graphical inspections of the relations between covariates and the EBE's of individual parameters to find the relevant relations. However, if data are sparse, this may lead to shrinkage of the EBE towards the typical parameter value so that a clinically relevant relation may become distorted in its shape or appear as unimportant or falsely important (Wade, Edholm, & Salmonson, 2005; Savic, Wilkins, & Karlsson, 2006).

A statistical evaluation of a relation can be used to pick up even weak trends that may be invisible in a graphical inspection. This is commonly used

for identifying the covariate model using generalized-additive modeling (GAM) (Mandema, Verotta, & Sheiner, 1992). Also in this case, the issue of shrinkage is problematic unless the data contain rich information on the investigated individual parameters. A modest correlation between the population model parameters can result in a false correlation between the individual parameters. Thus, a relation between a covariate and one parameter may induce a false relation between the covariate and several other parameters.

A completely different approach performing the covariate selection outside NONMEM is to use Wald's approximation to the likelihood ratio test (WAM) (Kowalski & Hutmacher, 2001). This method requires the point estimates and the covariance matrix of the estimates from a full model fit including all covariate relations of interest.

In the development of the multinomial Markov-chain model proposed with this thesis, estimating a full model with all the covariate relations of interest was not possible. Moreover, the shrinkage computed for the different sub-models was not generally low and correlation between the population model parameters was high. For these reasons, covariates selection outside NONMEM was not performed.

6.1 Methods

Covariates are often selected within NONMEM in a stepwise manner, e.g. using the procedure stepwise covariate modeling (SCM) (Jonsson & Karlsson, 1998). Stepwise-selection procedures currently dominate model selection in non-linear mixed-effect modeling, not only for the elaboration of the second-stage model, i.e. the model with covariate effects included. This approach can be implemented according to three different algorithms.

In 'forward selection', the model complexity (model size) is increased from a 'base model' (the model with no covariate effects included). The features of interest that can be fitted are evaluated by including them one at a time into the model. The feature that performs best according to the p-value is included into the model if it is statistically significant. Subsequently, all other model features are re-evaluated in the new model and a second feature is included if it is significant. The procedure stops at the full-forward model when no further model features are statistically significant. An alternative approach is 'backward elimination'. The starting point is a model that initially includes all features of interest, called the 'full model'. Elimination is performed until only statistically significant features remain. When this approach is possible, the end result is often as good as an all-subset selection, which investigates all combinations of the different model features of interest. This is not always the case with forward selection (Sauerbrei, 1999). However, the presence of features that are mutually exclusive or inestimable in combination due to correlations between the estimates often prevents the use of backward elimination techniques. Further, many ideas that are generated during an exploratory analysis cannot be included in the initial model. Because of this, backward elimination can only be used for part of the model-building procedure. Another alternative is to use forward selection with a less strict (i.e. a higher) p-value. The full forward model obtained in this manner is then refined by backward elimination using the same or a stricter p-value. This procedure is called 'forward inclusion and backward elimination' and is commonly applied in the selection of a covariate model (Jonsson & Karlsson, 1998).

The stepwise procedures have been extensively investigated in traditional statistics and a few of the associated problems are outlined below.

SCM often uses a p-value as an indicator of when to halt inclusion or deletion of further covariate coefficients, i.e. as a stopping rule. In general, several covariates are investigated, possibly on a number of structural model parameters and in several different functional forms. Therefore, the overall type-I error rate (i.e. the probability of including one or more false covariate coefficients into the model) is much higher than indicated by the required p-value. This type of problem is related to multiple comparisons (O'Neill & Wetherill,

1971). To correct for this, a stricter p-value is often used in the selection, although this can in turn result in omitting relations that are actually important. Furthermore, correction of the p-value is only approximate or even arbitrary. Because of correlations, finding the value that corresponds to the overall type-I error rate is very computer intensive. Thus, although the p-value is used as a criterion for selection based on the ideas of hypothesis testing, the actual strength by which the null hypothesis has been rejected is unknown in case of multiple comparisons.

The coefficients selected by SCM are exaggerated because of selection bias. A relation that seems important is often statistically significant whereas one which by random chance seems less important is left out. In this manner, the selected relations are on average more important than they would have been if the full model had been estimated without selection. A systematic difference is called 'bias'. In this context it is called 'selection bias' since it is caused by two elements in the selection procedure: the requirement of statistical significance and the competition between correlated covariates (Miller A. J., 1984).

Associated with selection bias, SCM is categorical when selecting covariates. A relation is either included or completely excluded from the model. This categorical selection leads to highly variable estimates and reduces the predictive performance of the model.

Compared to all-subset selection, the stepwise approach may not come across the optimal combination of predictors in its search path. This can especially occur when forward selection is used on a set of predictors that perform well together but poorly alone (Berk, 1978). This problem can be overcome by starting a backward elimination of the full covariate model (Sauerbrei, 1999). However, with many covariates to investigate, or many different functional relations or parameters on which effects are possible, this approach becomes impossible or at least very time consuming.

In case of multinomial Markov-chain modeling, five different sub-models need to be analyzed for covariates inclusion. In each of them there are several population parameters which are potentially affected by covariates. Since this type of analysis was never performed before, at least on transition probabilities during nighttime, there are no relevant priors on covariate effects on these parameters. As already pointed out in this chapter introduction, a potential explanatory analysis or, more generally, an investigation done outside NONMEM may be unreliable. In our specific case, an exploratory covariate analysis was considered time consuming and was not performed. Therefore, although the stepwise procedures suffers from some drawbacks, as described above, they are adopted here as the statistic tool for estimation of parameter-covariate relations. The large number of relations to test implies the choice of forward inclusion and backward elimination as the algorithm to implement. The discriminating p-values for covariate effect inclusion (forward search) and exclusion (backward search) are chosen to be 0.05 and 0.01, respectively. Each time a relation is tested on a specific sub-model, the covariate effect is

added to a specific logit at a specific nighttime break-point, as done for stage time effect. By analogy with the characterization of the stage time effect itself, the piecewise linear relation between covariate and parameter is investigated, using the covariate median as independent variable value at which the slope of the effect can change. However, also the linear additive effect is tested as a simpler model with a smaller OFV drop to be considered significant (using p-value=0.05, one added parameter requires a minimum increase in OFV equal to 3.84, while two added parameters require 5.99). Different effects at different break-points are allowed.

The stepwise procedure is implemented using PsN© (Lindbom, Pihlgren, & Jonsson, 2005). This software can aid the use of NONMEM in many different ways, in this particular case by automating the whole process of forward inclusion and backward elimination. To do this, a configuration file needs to be defined assessing the specifics of the process, i.e. the parameters and covariates, their types, the combinations to be tested and the modifications to be made to the NONMEM model file written for the base model. Moreover, the latter needs to be organized according to some specific rules.

The available potential covariates are age, gender and body mass index (BMI). Their distribution is summarized in Table 4.1. As covered in the Discussion paragraph, all of them are shown in the literature to potential influence sleep, therefore they all are investigated here for statistical significance on the model parameters.

6.2 Results

Stepwise covariate modeling brings to OFV reduction in all sub-models, SWS excluded, as indicated in Table 6.1. All of the three analyzed covariates (age, gender and BMI) are included, linearly affecting various model parameters in different night sections. A visual representation of such effects is provided by Figure 6.1, where typical individual probability profiles are shown for different covariates values. Reduction in inter-individual variability is generally not achieved. The application of sPPC to the obtained full model does not show any relevant improvement in the model performance.

6.3 Discussion

Age, gender and BMI are found to be statistically significant predictors of transition probabilities profiles during nighttime in the considered population of insomniac subjects under placebo. However, the predictive performance of the model and the explanation of inter-individual variability are not improved by their inclusion. Each covariate significantly influences specific transitions, in specific nighttime intervals: therefore, the multiple covariate effects could be diluted if evaluated on aggregated sleep parameters. The choice of the covariate-parameter relations to include or exclude in the model is based on p-values which need to be interpreted with caution, since multiple comparisons are involved (Ribbing & Jonsson, Power, selection bias and predictive performance of the population pharmacokinetic covariate model, 2004).

To my knowledge, this is the first analysis where age, BMI and gender are considered potential covariates with respect to transitions between sleep stages. Moreover, transitions are considered here in terms of transition probabilities rather than transition frequencies. In addition, the nature of this model allows understanding in which part of nighttime the effect is significant.

If age increases, probability is found

- to increase for transitions to AW, and to decrease for transitions to ST2 and SWS, in the first hours;
- to increase for the transition from REM to AW, during intermediate hours;
- to increase for transitions to AW, in the last hours.

These effects are consistent with previous findings from the literature (Van Cauter, Leproult, & Plat, 2000; Redline, Kirchner, Quan, Gottlieb, Kapur, & Newman, 2004; Vitiello, 2006; Sahlin, Franklin, Stenlund, & Lindberg, 2009), described in terms of sleep stage percentages, arousal index, sleep efficiency, total sleep time and time spent awake after falling asleep.

Gender is found to affect transition probabilities only in the last part of the night: transitions from AW to ST2 and from REM to AW appear more likely in women. In the literature, it has been reported that females have higher sleep efficiency, higher SWS percentage and lower light sleep (ST1 & ST2) percentage (Redline, Kirchner, Quan, Gottlieb, Kapur, & Newman, 2004), compared with males. Therefore, in this case linking our findings with the ones reported in literature becomes challenging.

Finally, high BMI values translate into a reduction of the following transitions probabilities:

- from ST1 to ST2, ST1 to REM and ST2 to SWS, in the intermediate night hours,
- and from ST2 to REM, in the last hours.

These effects are compatible with lower SWS percentage (Rao, Blackwell, Redline, Stefanick, Ancoli-Israel, & Stone, 2009; Kohatsu, et al., 2006), arousal index (Redline, Kirchner, Quan, Gottlieb, Kapur, & Newman, 2004) and sleep duration(Van Cauter & Knutson, 2008; Kohatsu, et al., 2006) reported in the literature.

It is important to notice that the covariate analysis is performed on insomniac subjects treated with placebo. Although the considered covariates do not appear to be relevant in terms of model predictive performance when including or excluding their effects, their relevance cannot be excluded in a patient population with a wider range of severity. In fact, it is likely that the effect of age, gender and BMI on sleep architecture is highly masked by the insomnia severity. Further applications of this model in different patient populations or healthy subjects are recommended to better characterize and possibly differentiate physiological and patho-physiological sleep architecture.

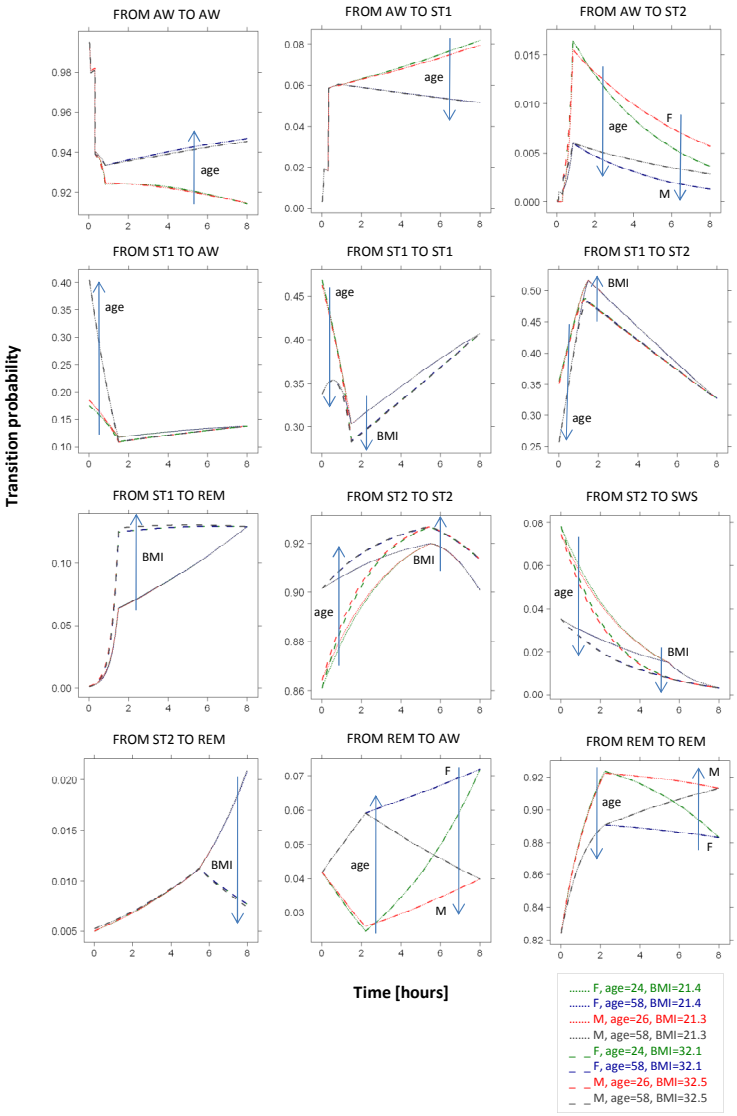
6.4 Tables and figures

Table 6.1 OFVs for the 5 sub-models before and after inclusion of parameter-covariate relationships (study A).

Sub-model	Before inclusion	After inclusion[a]
AW	21662	21605 (7)
ST1	24086	24070 (2)
ST2	48733	48705 (3)
SWS	8380	-
REM	14687	14668 (2)

[a] The number of included covariate effects is indicated between parenthesis.

Figure 6.1 Covariate effects on the typical individual profiles of some transition probabilities. The computation of probability values is done for covariate values chosen as follows: in both the males and females populations of study A, the 5th and 95th percentiles for age and BMI values are computed and used in each of their 4 combinations. Stage times and length of initial sleeplessness are chosen as the median values in the whole population. Effects are shown only on the transitions for which maximum changes in the probability values using the 4 combinations are greater than 0.01.



Chapter 7

Conclusions

Drug development in the insomnia field calls for a better understanding of sleep physiology in order to improve and differentiate novel medicines for the treatment of sleep disorders. On this basis, a proper evaluation of polysomnographic (PSG) data collected in clinical trials conducted to explore clinical efficacy of novel hypnotic compounds should include the assessment of sleep architecture and its drug-induced changes.

PSG data are nominal (non-ordered) correlated data with five different categories. One of the most interesting approaches for modelling this kind of data is the implementation of mixed-effect Markov-chain non-homogeneous models. The latter were previously parameterized using binary logistic functions, which allow a mixed-effect description of transition probabilities between sleep stages.

This work introduces instead the use of multinomial logistic functions as an integrated modelling framework to identify physiologically meaningful parameters. The first aim of this thesis was therefore the implementation of a mixed-effect Markov-chain non-homogeneous model with multinomial logistic functions, and its evaluation based on a few consolidated techniques generally applied for this purpose. The second aim was the improvement of the proposed model through the inclusion of new features or elements already discussed in the literature. The third objective was the assessment and application of a more thorough model evaluation on the final model, by means of the learning dataset, datasets created via Monte Carlo simulation and a further set of real data, used as validation dataset. Finally the thesis aimed at the analysis of the covariates affecting the model parameters and of the functional relations describing these effects.

The learning dataset was obtained from one-night polysomnography measurements on insomniac patients treated with placebo. The inclusion of multinomial logistic functions simplified the model building process and brought to a model structure more suited to this type of data. The evaluation of the base proposed model produced an overall good judgement of the model adequacy, but some observed discrepancies between model predictions and the real dataset suggested that the model could be further improved. The following model development brought to some modifications, mostly in terms of

parameterization of both the population and the individual model, and of inclusion of new predictors for the multinomial logits.

When evaluating the final model, consolidated and innovative diagnostics were discussed and implemented. In particular, visual estimation check (VEC) and visual prediction check (VPC) on transition frequencies were introduced here for the first time in the context of mixed-effect PK-PD modeling. Evaluation through comparison between learning dataset and simulated data (internal evaluation) shown that the maximum likelihood estimator, with Laplacian approximation as implemented in NONMEM VI, is robust enough for this analysis and that the overall performance of the model in both describing and simulating the data is very good. Also the evaluation performed using a new validation dataset (external evaluation) produced good results, suggesting that the model may be adopted on other datasets with similar features and that the parameter estimates gained from the learning dataset may be used as good initial estimates or priors when dealing with new sleep data from similar studies. It was also highlighted that reduced predictive performance may arise mostly for sleep aggregated parameters with high variability in different studies.

Finally, covariate analysis produced interesting novel results on how age, gender and body mass index affect the transition probabilities between sleep stages. The effects could be estimated building a second stage model where covariate-parameter relations were included only when statistically significant, on the basis of a forward inclusion and backward elimination stepwise procedure. Each covariate was found to influence some specific transition probabilities at specific nighttime intervals. These findings are generally in agreement with literature consideration on sleep covariate. However, from this analysis, it appears that the true effect of the considered covariates on the sleep architecture may be highly masked by the insomnia severity of the studied population.

In summary, the model developed and evaluated in this thesis is the first mixed-effect Markov-chain model of nominal polychotomous pharmacodynamic data using multinomial logistic functions. The proposed approach can be considered a robust modeling framework for describing and predicting sleep architecture. Herein, it was applied to data from a population of insomniac patients treated with placebo, but it can be easily extended to account also for drug effect on the transition probabilities and to compare sleep architecture in healthy volunteers and patients. In fact, this model may help to identify the key sleep patterns differentiating the mechanism of action of different hypnotic compounds; in addition it may help to thoroughly diagnose the presence, the absence and the grade of severity of various sleep disorders.

Finally, the proposed framework can be applied to other research areas where non-ordered polychotomous data are used to characterize phase transitions (e.g., in the gastrointestinal tract) or clinical scores (e.g., patient compliance, adverse event severity, disease progression, clinical efficacy). These

models can be clearly applied to a variety of pharmacokinetic-pharmacodynamic applications, such as quantification of drug responses in the study population, identification of covariates, predictions into untried regimens, and simulations of hypothetical trials during a drug development program.

Appendix A

NONMEM model file for sub-model AW:

```
$PROB trAW


$INPUT ID TIME DV=STAG MDV0 MDV STT SL IS
  ; TIME = epoch number (between 1 and 960)
  ; MDV0 = 0 if previous stage==AW, 1 otherwise
  ; STT = stage time
  ; SL = 0 if first sleep stage has not occurred yet, 1 otherwise
  ; has not occurred yet, 1 otherwise
  ; IS = first epoch with SL==1


$DATA data.csv IGNORE=@
  IGNORE=(MDV0.EQ.1)
  IGNORE=(STAG.GT.5)
  IGNORE=(STAG.EQ.3)
```



```
$PRED         ;LOGIT: G1, G2, G3
              ;BP: BPA, BPB, BPC: nighttime break-points
              ;BPs: BPsa, BPsb, BPsc: stage time break-points
              ;BPi: BP1, BP2, BP3: initial sleeplessness break-points
              ;STE: stage time effect

  BPA=IS
  BPC=960
  BPB=(BPC-BPA)*THETA(16)+BPA

  BPsa=1
  BPsb=THETA(17)
  BPsc=265

  BP1=2
  BP3=371             ;max(IS)-1
  BP2=(BP3-BP1)*THETA(18)+BP1


  STE1b=THETA(10)
  STE2b=THETA(11)
  STE3b=THETA(12)
  STE1c=THETA(13)
  STE2c=THETA(14)
  STE3c=THETA(15)

  TVG1A=THETA(1)
  TVG2A=THETA(2)
  TVG3A=THETA(3)
  TVG1B=THETA(4)
  TVG2B=THETA(5)
```

```
  TVG3B=THETA(6)
  TVG1C=THETA(7)
  TVG2C=THETA(8)
  TVG3C=THETA(9)

  TVG11=THETA(19)
  TVG21=THETA(20)
  TVG31=THETA(21)
  TVG12=THETA(22)
  TVG22=THETA(23)
  TVG32=THETA(24)
  TVG13=THETA(25)
  TVG23=THETA(26)
  TVG33=THETA(27)

  IF (STT.LE.BPsb.AND.SL.EQ.1) THEN
    G1A=(TVG1A+ETA(1))*(BPsb-STT)/(BPsb-BPsa)+(TVG1A+ETA(1)+STE1b)*(STT-
BPsa)/(BPsb-BPsa)
    G2A=(TVG2A+ETA(2))*(BPsb-STT)/(BPsb-BPsa)+(TVG2A+ETA(2)+STE2b)*(STT-
BPsa)/(BPsb-BPsa)
    G3A=(TVG3A+ETA(3))*(BPsb-STT)/(BPsb-BPsa)+(TVG3A+ETA(3)+STE3b)*(STT-
BPsa)/(BPsb-BPsa)

    G1B=(TVG1B+ETA(1))*(BPsb-STT)/(BPsb-BPsa)+(TVG1B+ETA(1)+STE1b)*(STT-
BPsa)/(BPsb-BPsa)
    G2B=(TVG2B+ETA(2))*(BPsb-STT)/(BPsb-BPsa)+(TVG2B+ETA(2)+STE2b)*(STT-
BPsa)/(BPsb-BPsa)
    G3B=(TVG3B+ETA(3))*(BPsb-STT)/(BPsb-BPsa)+(TVG3B+ETA(3)+STE3b)*(STT-
BPsa)/(BPsb-BPsa)

    G1C=(TVG1C+ETA(1))*(BPsb-STT)/(BPsb-BPsa)+(TVG1C+ETA(1)+STE1b)*(STT-
BPsa)/(BPsb-BPsa)
    G2C=(TVG2C+ETA(2))*(BPsb-STT)/(BPsb-BPsa)+(TVG2C+ETA(2)+STE2b)*(STT-
BPsa)/(BPsb-BPsa)
    G3C=(TVG3C+ETA(3))*(BPsb-STT)/(BPsb-BPsa)+(TVG3C+ETA(3)+STE3b)*(STT-
BPsa)/(BPsb-BPsa)
  ENDIF
  IF (STT.GT.BPsb.AND.SL.EQ.1) THEN
    G1A=(TVG1A+ETA(1)+STE1b)*(BPsc-STT)/(BPsc-
BPsb)+(TVG1A+ETA(1)+STE1c)*(STT-BPsb)/(BPsc-BPsb)
    G2A=(TVG2A+ETA(2)+STE2b)*(BPsc-STT)/(BPsc-
BPsb)+(TVG2A+ETA(2)+STE2c)*(STT-BPsb)/(BPsc-BPsb)
    G3A=(TVG3A+ETA(3)+STE3b)*(BPsc-STT)/(BPsc-
BPsb)+(TVG3A+ETA(3)+STE3c)*(STT-BPsb)/(BPsc-BPsb)

    G1B=(TVG1B+ETA(1)+STE1b)*(BPsc-STT)/(BPsc-
BPsb)+(TVG1B+ETA(1)+STE1c)*(STT-BPsb)/(BPsc-BPsb)
    G2B=(TVG2B+ETA(2)+STE2b)*(BPsc-STT)/(BPsc-
BPsb)+(TVG2B+ETA(2)+STE2c)*(STT-BPsb)/(BPsc-BPsb)
    G3B=(TVG3B+ETA(3)+STE3b)*(BPsc-STT)/(BPsc-
BPsb)+(TVG3B+ETA(3)+STE3c)*(STT-BPsb)/(BPsc-BPsb)

    G1C=(TVG1C+ETA(1)+STE1b)*(BPsc-STT)/(BPsc-
BPsb)+(TVG1C+ETA(1)+STE1c)*(STT-BPsb)/(BPsc-BPsb)
    G2C=(TVG2C+ETA(2)+STE2b)*(BPsc-STT)/(BPsc-
BPsb)+(TVG2C+ETA(2)+STE2c)*(STT-BPsb)/(BPsc-BPsb)
    G3C=(TVG3C+ETA(3)+STE3b)*(BPsc-STT)/(BPsc-
BPsb)+(TVG3C+ETA(3)+STE3c)*(STT-BPsb)/(BPsc-BPsb)
  ENDIF

  IF (TIME.GE.BPA.AND.TIME.LE.BPB.AND.SL.EQ.1) THEN
    G1=G1A*(BPB-TIME)/(BPB-BPA)+G1B*(TIME-BPA)/(BPB-BPA)
    G2=G2A*(BPB-TIME)/(BPB-BPA)+G2B*(TIME-BPA)/(BPB-BPA)
    G3=G3A*(BPB-TIME)/(BPB-BPA)+G3B*(TIME-BPA)/(BPB-BPA)
  ENDIF
  IF (TIME.GT.BPB.AND.TIME.LE.BPC.AND.SL.EQ.1) THEN
```

```
    G1=G1B*(BPC-TIME)/(BPC-BPB)+G1C*(TIME-BPB)/(BPC-BPB)
    G2=G2B*(BPC-TIME)/(BPC-BPB)+G2C*(TIME-BPB)/(BPC-BPB)
    G3=G3B*(BPC-TIME)/(BPC-BPB)+G3C*(TIME-BPB)/(BPC-BPB)
  ENDIF


  IF (TIME.GE.BP1.AND.TIME.LE.BP2.AND.SL.EQ.0) THEN
    G1=(TVG11+ETA(4))*(BP2-TIME)/(BP2-BP1)+(TVG12+ETA(4))*(TIME-BP1)/(BP2-
BP1)
    G2=(TVG21+ETA(5))*(BP2-TIME)/(BP2-BP1)+(TVG22+ETA(5))*(TIME-BP1)/(BP2-
BP1)
    G3=(TVG31+ETA(6))*(BP2-TIME)/(BP2-BP1)+(TVG32+ETA(6))*(TIME-BP1)/(BP2-
BP1)
  ENDIF
  IF (TIME.GT.BP2.AND.TIME.LE.BP3.AND.SL.EQ.0) THEN
    G1=(TVG12+ETA(4))*(BP3-TIME)/(BP3-BP2)+(TVG13+ETA(4))*(TIME-BP2)/(BP3-
BP2)
    G2=(TVG22+ETA(5))*(BP3-TIME)/(BP3-BP2)+(TVG23+ETA(5))*(TIME-BP2)/(BP3-
BP2)
    G3=(TVG32+ETA(6))*(BP3-TIME)/(BP3-BP2)+(TVG33+ETA(6))*(TIME-BP2)/(BP3-
BP2)
  ENDIF

  PAWp=1/(1+EXP(G1p)+EXP(G2p)+EXP(G3p))
  P1p=EXP(G1p)/(1+EXP(G1p)+EXP(G2p)+EXP(G3p))
  P2p=EXP(G2p)/(1+EXP(G1p)+EXP(G2p)+EXP(G3p))
  P3p=0
  PRp=EXP(G3p)/(1+EXP(G1p)+EXP(G2p)+EXP(G3p))

  Y=0
  IF (STAG.EQ.0) Y=PAW
  IF (STAG.EQ.1) Y=P1
  IF (STAG.EQ.2) Y=P2
  IF (STAG.EQ.5) Y=PR
```



```
$THETA

  -1; TVG1A
  -5; TVG2A
  -10 FIX; TVG3A

  -1; TVG1B
  -1; TVG2B
  -4; TVG3B

  -.1; TVG1C
  -2; TVG2C
  -2; TVG3C

  -2; STE1b
  -3; STE2b
  -5; STE3b

  -5; STE1c
  -10 FIX; STE2c
  -10 FIX; STE3c

  (0,.1,1); REL BPB

  (1,10,265); BPsb
  (0,.4,1); REL BP2

  -5; TVG11
  -10 FIX; TVG21
  -10 FIX; TVG31
```

```
  -5; TVG12
  -5; TVG22
  -10 FIX; TVG32

  -1; TVG13
  -10 FIX; TVG23
  -10 FIX; TVG33


$OMEGA BLOCK(2)

  .1 ;G1i
  -.02 .5 ;G2i

$OMEGA

  .5 ;G3i


$ESTIMATION METHOD=COND LAPLACE LIKE MSFO=msf1


$COVARIANCE MATRIX=R PRINT=E
```

Appendix B

Some lines from the dataset used with sub-model AW:

ID	TIME	STAG	MDV0	MDV	STT	SL	IS
142	52	0	0	0	51	0	55
142	53	0	0	0	52	0	55
142	54	1	0	0	53	0	55
142	55	1	1	0	1	1	55
142	56	1	1	0	2	1	55
…							
126	343	2	1	0	10	1	71
126	344	2	1	0	11	1	71
126	345	0	1	0	12	1	71
126	346	1	0	0	1	1	71
126	347	0	1	0	1	1	71
126	348	1	0	0	1	1	71
126	349	1	1	0	1	1	71

Appendix C

NONMEM control file for the simulation of new datasets:

```
$PROB simulator

$INPUT ID DV=VIS TIME STAGE
  ; TIME = epoch number (between 1 and 960)

$DATA data1.csv IGNORE=@

$ABBREVIATED DERIV2=NO

$PRED         ;LOGIT: G1, G2, G3
              ;BP: BPA, BPB, BPC: nighttime break-points
              ;BPs: BPsa, BPsb, BPsc: stage time break-points
              ;BPi: BP1, BP2, BP3: initial sleeplessness break-points
              ;STE: stage time effect
              ;STT: stage time

  REP=IREP

  AA=THETA(1)+ETA(1)+ETA(2)+ETA(3)+ETA(4)+ETA(5)+ETA(6)+ETA(7)+ETA(8)
  AA=AA+ETA(9)+ETA(10)+ETA(11)+ETA(12)+ETA(13)+ETA(14)+ETA(15)+ETA(16)

  MDV0=1
  MDV1=1
  MDV2=1
  MDV3=1
  MDV5=1

  IF (TIME.EQ.1) THEN
    PST=-1                    ;previous stage
    PSTT=-10                  ;previous stage time
    PSL=0                     ;previous sleep
    SL=0                      ;sleep
    IS=0                      ;first epoch with SL==1
  ENDIF

  SL=0
  IF (PSL.EQ.0.AND.PST.GT.0) SL=1
  IF (PSL.EQ.1) SL=1
  IF (PSL.EQ.0.AND.SL.EQ.1) IS=TIME

  STT=PSTT+1

  IF (PST.EQ.0) MDV0=0
  IF (PST.EQ.1) MDV1=0
  IF (PST.EQ.2) MDV2=0
  IF (PST.EQ.3) MDV3=0
  IF (PST.EQ.5) MDV5=0

  ; break-points
  BPA=2
```

```
BPB=960
BPC=960
BPsa=1
BP1=1
BP2=2
BP3=1

; stage time effect
STE4b=1
STE4c=1

; for stage2 sub-model
TVG4A=1
TVG4B=1
TVG4C=1
DEV4=1

; for awake sub-model
TVG11=1
TVG21=1
TVG31=1
TVG12=1
TVG22=1
TVG32=1
TVG13=1
TVG23=1
TVG33=1

;-------------------------------------------------------------
;  initialization using the previously estimated THETAs
;-------------------------------------------------------------

; awake sub-model
IF (MDV0.EQ.0) THEN
  BPA=IS                        ;first epoch with SL==1
  BPB=(BPC-BPA)*0.0679+BPA
  BPsc=265                      ;max stage time
  BPsb=(BPsc-BPsa)*0.024+BPsa


  BP1=2
  BP3=371                       ;MAX initial sleeplessness
  BP2=(BP3-BP1)*0.0298+BP1

  STE1b=-2.49
  STE2b=-3.59
  STE3b=-5.67

  STE1c=-6.63
  STE2c=-10
  STE3c=-10

  TVG1A=-0.251
  TVG2A=-2.66
  TVG3A=-10
  TVG1B=-0.209
  TVG2B=-1.01
  TVG3B=-2.86
  TVG1C=-0.203
  TVG2C=-2.08
  TVG3C=-2.1

  TVG11=-5.76
  TVG21=-10
  TVG31=-10
  TVG12=-3.93
```

```
  TVG22=-7.63
  TVG32=-10
  TVG13=-4.86
  TVG23=-10
  TVG33=-10

  DEV1=ETA(1)
  DEV2=ETA(2)
  DEV3=ETA(3)
ENDIF

; stage1 sub-model
IF (MDV1.EQ.0) THEN
  BPB=268
  BPsb=3.24
  BPsc=20

  STE1b=-0.447
  STE2b=-0.52
  STE3b=-0.463

  STE1c=-0.634
  STE2c=-0.246
  STE3c=-3.31

  TVG1A=-0.321
  TVG2A=-0.0716
  TVG3A=-4.15
  TVG1B=-1.07
  TVG2B=0.492
  TVG3B=-1.01
  TVG1C=-1.05
  TVG2C=-0.251
  TVG3C=-1.24

  DEV1=ETA(4)
  DEV2=ETA(5)
  DEV3=ETA(6)
ENDIF

; stage2 sub-model
IF (MDV2.EQ.0) THEN
  BPB=676
  BPsb=5.62
  BPsc=143

  STE1b=-0.905
  STE2b=-0.894
  STE3b=-1.19
  STE4b=-0.993

  STE1c=-1.79
  STE2c=-6.44
  STE3c=0.927
  STE4c=-6.75

  TVG1A=-2.46
  TVG2A=-2.52
  TVG3A=-1.71
  TVG4A=-4.07
  TVG1B=-2.57
  TVG2B=-2.6
  TVG3B=-3.33
  TVG4B=-3.34
  TVG1C=-2.13
  TVG2C=-2.33
```

```
  TVG3C=-4.59
  TVG4C=-3.16

  DEV1=ETA(7)
  DEV2=ETA(8)
  DEV3=ETA(9)
  DEV4=ETA(10)
ENDIF

; sws sub-model
IF (MDV3.EQ.0) THEN
  BPB=715
  BPsb=5.9
  BPsc=103

  STE1b=-0.999
  STE2b=-0.939
  STE3b=-2.5

  STE1c=0.761
  STE2c=-3.53
  STE3c=-3.73

  TVG1A=-3.22
  TVG2A=-4.83
  TVG3A=-0.526
  TVG1B=-3.31
  TVG2B=-5.13
  TVG3B=0.142
  TVG1C=-1.65
  TVG2C=-4.55
  TVG3C=0.389

  DEV1=ETA(11)
  DEV2=ETA(12)
  DEV3=ETA(13)
ENDIF

; rem sub-model
IF (MDV5.EQ.0) THEN
  BPB=640
  BPsb=13.6
  BPsc=100

  STE1b=0.351
  STE2b=-0.238
  STE3b=-1.22

  STE1c=0.566
  STE2c=-1.22
  STE3c=1.8

  TVG1A=-3.12
  TVG2A=-2.87
  TVG3A=-3.11
  TVG1B=-3.25
  TVG2B=-3.02
  TVG3B=-4.6
  TVG1C=-2.96
  TVG2C=-2.98
  TVG3C=-4.56

  DEV1=ETA(14)
  DEV2=ETA(15)
  DEV3=ETA(16)
ENDIF
```

```
  G1A=0
  G2A=0
  G3A=0
  G4A=0

  ;--------------------------------
  ; stage time effect interpolation
  ;--------------------------------

  IF (STT.GE.0.AND.STT.LE.BPsb.AND.SL.EQ.1) THEN
    G1A=(TVG1A+DEV1)*(BPsb-STT)/(BPsb-BPsa)+(TVG1A+DEV1+STE1b)*(STT-
BPsa)/(BPsb-BPsa)
    G2A=(TVG2A+DEV2)*(BPsb-STT)/(BPsb-BPsa)+(TVG2A+DEV2+STE2b)*(STT-
BPsa)/(BPsb-BPsa)
    G3A=(TVG3A+DEV3)*(BPsb-STT)/(BPsb-BPsa)+(TVG3A+DEV3+STE3b)*(STT-
BPsa)/(BPsb-BPsa)
    G4A=(TVG4A+DEV4)*(BPsb-STT)/(BPsb-BPsa)+(TVG4A+DEV4+STE4b)*(STT-
BPsa)/(BPsb-BPsa)

    G1B=(TVG1B+DEV1)*(BPsb-STT)/(BPsb-BPsa)+(TVG1B+DEV1+STE1b)*(STT-
BPsa)/(BPsb-BPsa)
    G2B=(TVG2B+DEV2)*(BPsb-STT)/(BPsb-BPsa)+(TVG2B+DEV2+STE2b)*(STT-
BPsa)/(BPsb-BPsa)
    G3B=(TVG3B+DEV3)*(BPsb-STT)/(BPsb-BPsa)+(TVG3B+DEV3+STE3b)*(STT-
BPsa)/(BPsb-BPsa)
    G4B=(TVG4B+DEV4)*(BPsb-STT)/(BPsb-BPsa)+(TVG4B+DEV4+STE4b)*(STT-
BPsa)/(BPsb-BPsa)

    G1C=(TVG1C+DEV1)*(BPsb-STT)/(BPsb-BPsa)+(TVG1C+DEV1+STE1b)*(STT-
BPsa)/(BPsb-BPsa)
    G2C=(TVG2C+DEV2)*(BPsb-STT)/(BPsb-BPsa)+(TVG2C+DEV2+STE2b)*(STT-
BPsa)/(BPsb-BPsa)
    G3C=(TVG3C+DEV3)*(BPsb-STT)/(BPsb-BPsa)+(TVG3C+DEV3+STE3b)*(STT-
BPsa)/(BPsb-BPsa)
    G4C=(TVG4C+DEV4)*(BPsb-STT)/(BPsb-BPsa)+(TVG4C+DEV4+STE4b)*(STT-
BPsa)/(BPsb-BPsa)
  ENDIF

  IF (STT.GT.BPsb.AND.SL.EQ.1) THEN
    G1A=(TVG1A+DEV1+STE1b)*(BPsc-STT)/(BPsc-BPsb)+(TVG1A+DEV1+STE1c)*(STT-
BPsb)/(BPsc-BPsb)
    G2A=(TVG2A+DEV2+STE2b)*(BPsc-STT)/(BPsc-BPsb)+(TVG2A+DEV2+STE2c)*(STT-
BPsb)/(BPsc-BPsb)
    G3A=(TVG3A+DEV3+STE3b)*(BPsc-STT)/(BPsc-BPsb)+(TVG3A+DEV3+STE3c)*(STT-
BPsb)/(BPsc-BPsb)
    G4A=(TVG4A+DEV4+STE4b)*(BPsc-STT)/(BPsc-BPsb)+(TVG4A+DEV4+STE4c)*(STT-
BPsb)/(BPsc-BPsb)

    G1B=(TVG1B+DEV1+STE1b)*(BPsc-STT)/(BPsc-BPsb)+(TVG1B+DEV1+STE1c)*(STT-
BPsb)/(BPsc-BPsb)
    G2B=(TVG2B+DEV2+STE2b)*(BPsc-STT)/(BPsc-BPsb)+(TVG2B+DEV2+STE2c)*(STT-
BPsb)/(BPsc-BPsb)
    G3B=(TVG3B+DEV3+STE3b)*(BPsc-STT)/(BPsc-BPsb)+(TVG3B+DEV3+STE3c)*(STT-
BPsb)/(BPsc-BPsb)
    G4B=(TVG4B+DEV4+STE4b)*(BPsc-STT)/(BPsc-BPsb)+(TVG4B+DEV4+STE4c)*(STT-
BPsb)/(BPsc-BPsb)

    G1C=(TVG1C+DEV1+STE1b)*(BPsc-STT)/(BPsc-BPsb)+(TVG1C+DEV1+STE1c)*(STT-
BPsb)/(BPsc-BPsb)
    G2C=(TVG2C+DEV2+STE2b)*(BPsc-STT)/(BPsc-BPsb)+(TVG2C+DEV2+STE2c)*(STT-
BPsb)/(BPsc-BPsb)
    G3C=(TVG3C+DEV3+STE3b)*(BPsc-STT)/(BPsc-BPsb)+(TVG3C+DEV3+STE3c)*(STT-
BPsb)/(BPsc-BPsb)
    G4C=(TVG4C+DEV4+STE4b)*(BPsc-STT)/(BPsc-BPsb)+(TVG4C+DEV4+STE4c)*(STT-
BPsb)/(BPsc-BPsb)
```

```
  ENDIF

;--------------------------------
; logits interpolation
;--------------------------------

  G1=G1A
  G2=G2A
  G3=G3A
  G4=G4A

  IF (TIME.GE.BPA.AND.TIME.LE.BPB.AND.SL.EQ.1) THEN
    G1=G1A*(BPB-TIME)/(BPB-BPA)+G1B*(TIME-BPA)/(BPB-BPA)
    G2=G2A*(BPB-TIME)/(BPB-BPA)+G2B*(TIME-BPA)/(BPB-BPA)
    G3=G3A*(BPB-TIME)/(BPB-BPA)+G3B*(TIME-BPA)/(BPB-BPA)
    G4=G4A*(BPB-TIME)/(BPB-BPA)+G4B*(TIME-BPA)/(BPB-BPA)
  ENDIF
  IF (TIME.GT.BPB.AND.TIME.LE.BPC.AND.SL.EQ.1) THEN
    G1=G1B*(BPC-TIME)/(BPC-BPB)+G1C*(TIME-BPB)/(BPC-BPB)
    G2=G2B*(BPC-TIME)/(BPC-BPB)+G2C*(TIME-BPB)/(BPC-BPB)
    G3=G3B*(BPC-TIME)/(BPC-BPB)+G3C*(TIME-BPB)/(BPC-BPB)
    G4=G4B*(BPC-TIME)/(BPC-BPB)+G4C*(TIME-BPB)/(BPC-BPB)
  ENDIF

  IF (TIME.GE.BP1.AND.TIME.LE.BP2.AND.SL.EQ.0) THEN
    G1=(TVG11+DEV1)*(BP2-TIME)/(BP2-BP1)+(TVG12+DEV1)*(TIME-BP1)/(BP2-BP1)
    G2=(TVG21+DEV2)*(BP2-TIME)/(BP2-BP1)+(TVG22+DEV2)*(TIME-BP1)/(BP2-BP1)
    G3=(TVG31+DEV3)*(BP2-TIME)/(BP2-BP1)+(TVG32+DEV3)*(TIME-BP1)/(BP2-BP1)
  ENDIF
  IF (TIME.GT.BP2.AND.SL.EQ.0) THEN
    G1=(TVG12+DEV1)*(BP3-TIME)/(BP3-BP2)+(TVG13+DEV1)*(TIME-BP2)/(BP3-BP2)
    G2=(TVG22+DEV2)*(BP3-TIME)/(BP3-BP2)+(TVG23+DEV2)*(TIME-BP2)/(BP3-BP2)
    G3=(TVG32+DEV3)*(BP3-TIME)/(BP3-BP2)+(TVG33+DEV3)*(TIME-BP2)/(BP3-BP2)
  ENDIF

  ;---------------------------------------
  ; transition probabilities as anti-logits
  ;---------------------------------------

  PAW=2
  P1=0
  P2=0
  P3=0

  ; awake
  IF (MDV0.EQ.0) THEN
    PAW=1/(1+EXP(G1)+EXP(G2)+EXP(G3))
    P1=EXP(G1)/(1+EXP(G1)+EXP(G2)+EXP(G3))
    P2=EXP(G2)/(1+EXP(G1)+EXP(G2)+EXP(G3))
    P3=0
    PR=EXP(G3)/(1+EXP(G1)+EXP(G2)+EXP(G3))
  ENDIF

  ; stage1
  IF (MDV1.EQ.0) THEN
    PAW=EXP(G1)/(1+EXP(G1)+EXP(G2)+EXP(G3))
    P1=1/(1+EXP(G1)+EXP(G2)+EXP(G3))
    P2=EXP(G2)/(1+EXP(G1)+EXP(G2)+EXP(G3))
    P3=0
    PR=EXP(G3)/(1+EXP(G1)+EXP(G2)+EXP(G3))
  ENDIF

  ; stage2
  IF (MDV2.EQ.0) THEN
    PAW=EXP(G1)/(1+EXP(G1)+EXP(G2)+EXP(G3)+EXP(G4))
    P1=EXP(G2)/(1+EXP(G1)+EXP(G2)+EXP(G3)+EXP(G4))
```

```
  P2=1/(1+EXP(G1)+EXP(G2)+EXP(G3)+EXP(G4))
  P3=EXP(G3)/(1+EXP(G1)+EXP(G2)+EXP(G3)+EXP(G4))
  PR=EXP(G4)/(1+EXP(G1)+EXP(G2)+EXP(G3)+EXP(G4))
ENDIF

; sws
IF (MDV3.EQ.0) THEN
  PAW=EXP(G1)/(1+EXP(G1)+EXP(G2)+EXP(G3))
  P1=EXP(G2)/(1+EXP(G1)+EXP(G2)+EXP(G3))
  P2=EXP(G3)/(1+EXP(G1)+EXP(G2)+EXP(G3))
  P3=1/(1+EXP(G1)+EXP(G2)+EXP(G3))
  PR=0
ENDIF

; rem
IF (MDV5.EQ.0) THEN
  PAW=EXP(G1)/(1+EXP(G1)+EXP(G2)+EXP(G3))
  P1=EXP(G2)/(1+EXP(G1)+EXP(G2)+EXP(G3))
  P2=EXP(G3)/(1+EXP(G1)+EXP(G2)+EXP(G3))
  P3=0
  PR=1/(1+EXP(G1)+EXP(G2)+EXP(G3))
ENDIF

; probability scale
DEV1=PAW+P1
DEV2=DEV1+P2
DEV3=DEV2+P3

; Initialization of stage ST - first epoch is awake
ST=0

AW=0
ST1=0
ST2=0
ST3=0
REM=0

; random call for the simulation step
IF (ICALL.EQ.4) THEN
  CALL RANDOM (2,R)

  ; if R<PAW transition to AW is taken into account
  ; and the following stage is AW(0)
  IF (R.LT.PAW) THEN
    ST=0
    AW=1

  ; if PAW<=R<PAW+P1 transition to ST1 is taken into account
  ; and the following stage is ST1(1)
  ELSEIF (R.LT.DEV1.AND.R.GE.PAW) THEN
    ST=1
    ST1=1

  ; if PAW+P1<=R<PAW+P1+P2 transition to ST2 is taken into account
  ; and the following stage is ST2(2)
  ELSEIF (R.LT.DEV2.AND.R.GE.DEV1) THEN
    ST=2
    ST2=1

  ; if PAW+P1+P2<=R<PAW+P1+P2+P3 transition to SWS is
  ; taken into account and the following stage is SWS(3)
  ELSEIF (R.LT.DEV3.AND.R.GE.DEV2) THEN
    ST=3
    ST3=1

  ; if R>=PAW+P1+P2+P3 transition to REM is
```

```
    ; taken into account and the following stage is REM(5)
    ELSE
      ST=5
      REM=1
    ENDIF
  ENDIF

  CH=1                              ;change
  IF (ST.EQ.PST) CH=0               ;no change
  PSTT=STT*(CH-1)**2

  IF (TIME.EQ.1) STT=0
  PSL=SL
  PST=ST

$THETA 1

;AW
$OMEGA BLOCK(2) .134 -0.142 .831
$OMEGA 1.07
;ST1
$OMEGA BLOCK(3) .318 .18 .389 -0.0637 .138 .398
;ST2
$OMEGA .152 .456 .786 .239
;ST3
$OMEGA (0 FIX) 1.4 1.35
;REM
$OMEGA .584 .849 1.1

$SIM (123456) (123 UNIFORM) ONLY SUBPROBS=100 NOPREDICTION

$TAB ID TIME ST MDV0 MDV1 MDV2 MDV3 MDV5 MDV
     AW ST1 ST2 ST3 REM STT SL IS REP
     ONEHEADER NOPRINT NOAPPEND FILE=simul
```

References

Administration, F. a. (1999). *Guidance for industry: population pharmacokinetics.* Rockville, Maryland, USA: Department of Health and Human Services.

Agresti, A. (2002). *Categorical Data Analysis.* Haboken, NJ, USA: Wiley.

American Psychiatric Association. (2000). *Diagnostic and Statistical Manula of Mental Disorders (DSM-IV-TR).* Washington DC, USA.

Armstrong, B. G., & Sloan, M. (1989). Ordinal regression models for epidemiologic data. *Am J Epidemiol , 129*, 191-204.

Aserinsky, E., & Kleitman, N. (1953). Regularly occurring periods of eye motility, and concomitant Phenomena. *Science , 118*, 273-274.

Avidan, A. Y. (2003). Insomnia in the geriatric patient. *Clin Cornerstone , 5*, 51-60.

Baldessarini, R. J. (2001). Drugs and the treatment of psychiatric disorders. In J. G. Hardman, & L. E. Limbird, *The Pharmacological Basis of Therape, 10th ed* (pp. 447-484). Goodman and Gilman.

Beal, S. L., & Sheiner, L. B. (1998). *NONMEM User's Guides.* San Francisco, California, USA: NONMEM Project Group.

Benca, R. M. (2005). Diagnosis and treatment of chronic insomnia: a review. *Psychiatr Serv , 56*, 332-343.

Berger, H. (1930). Ueber das elektroenkephalogramm des menschen. *J Psychol Neurol , 40*, 160-179.

Bergstrand, M., Hooker, A. C., & Karlsson, M. O. (2009). Visual Predictive Checks for Censored and Categorical data. *PAGE 17.* St. Petersburg, Russia: Abstr 1604 [www.page-meeting.org/?abstract=1604].

Bergstrand, M., Soderlind, E., Weitschies, W., & Karlsson, M. O. (2009). Mechanistic Modeling of a Magnetic Marker Monitoring Study Linking Gastrointestinal Tablet Transit, In Vivo Drug Release, and Pharmacokinetics. *Clin Pharmacol Ther , 86*, 77-83.

Berk, K. N. (1978). Comparing Subset Regression Procedures. *Technometrics , 20* (1), 1-6.

Bizzotto, R., Zamuner, S., De Nicolao, G., Karlsson, M. O., & Gomeni, R. (2010). Multinomial logistic estimation of Markov-chain models for modeling sleep architecture in primary insomnia patients. *J Pharmacokinet Pharmacodyn , 37*, 137-155.

Bizzotto, R., Zamuner, S., Nucci, G., Cobelli, C., & Gomeni, R. (2008). Mixed effects Markov models for modelling sleep in insomniac patients treated with placebo in a 28 days trial: emphasis on the break points selection. *PAGE 17.* Marseille, France: Abstr 1383 [www.page-meeting.org/?abstract=1383].

Bland, M. (2000). *An introduction to medical statistics, 3rd edition.* Oxford, UK: Oxford University Press.

Burgess, H. J., Holmes, A. L., & Dawson, D. (2001). The relationship between slow-wave activity, body temperature, and cardiac activity during nighttime sleep. *Sleep , 24* (3), 343-349.

Carskadon, M. A., & Dement, W. C. (1989). Normal human sleep: an overview. In M. H. Kryger, T. Roth, & W. C. Dement, *Principles and practice of sleep medicine* (pp. 3-13). W. B. Saunders, Philadelphia, USA.

Carskadon, M. A., & Dement, W. C. (2005). Normal human sleep: an overview. In M. H. Kryger, T. Roth, & W. C. Dement, *Principles and practice of sleep medicine* (pp. 13-23). Philadelphia, Pennsylvania, USA: Saunders.

Chokroverty, S. (2003). In S. Chokroverty, A. H. Wayne, & A. S. Walter, *Sleep and movement disorders* (pp. 23-43). Philadelphia, Pennsylvania: Butterworth Heinemann.

Cox, E. H., Veyrat-Follet, C., Beal, S. L., Fuseau, E., Kenkare, S., & Sheiner, L. B. (1999). A Population Pharmacokinetic-Pharmacodynamic Analysis of Repeated Measures Time-to-Event Pharmacodynamic Responses: The Antiemetic Effect of Ondansetron. *J Pharmacokin Biopharm , 27* (6), 625-644.

Dement, W., & Kleitman, N. (1957). Cyclic variations in EEG during sleep and their relation to eye movement, body motility and dreaming. *Electroencephalogr Clin Neurophysiol , 9*, 673-690.

Dijk, D. J., & Czeisler, C. A. (1995). Contribution of the circadian pacemaker and the sleep homeostat to sleep propensity, sleep structure, electroencephalographic slow waves, and sleep spindle activity in humans. *J Neurosci , 15*, 3526-3538.

Duval, V., & Karlsson, M. O. (2002). Impact of Omission or Replacement of Data below the Limit of Quantification on Parameter Estimates in a Two-Compartment Model. *Pharm Res , 19* (12), 1835-1840.

Elie, R., Ruther, E., & Farr, I. (1999). Sleep latency is shortened during 4 weeks of treatment with zaleplon, a novel nonbenzodiazepine hypnotic. *J Clin Psychiatry , 60*, 536-544.

Erman, M., Seiden, D., Zammit, G., Sainati, S., & Zhang, J. (2006). An efficacy, safety, and dose–response study of Ramelteon in patients with chronic primary insomnia. *Sleep Med , 7*, 17-24.

Foley, D. J., Monjan, A. A., & Brown, S. L. (1995). Sleep complaints among elderly persons: an epidemiologic study of three communities. *Sleep , 18*, 425-432.

Foster, A. C., Pelleymounter, M. A., & Cullen, M. J. (2004). In vivo pharmacological characterization of indiplon, a novel pyrazolopyrimidine sedative-hypnotic. *J Pharmacol Exp Ther , 311*, 547-559.

Gallop, J. R., Ten Have, T. R., & Crits-Christoph, P. (2006). A mixed effects Markov model for repeated binary outcomes with non-ignorable dropout. *Stat Med , 25* (14), 2398–2426.

Gimenez, S., Clos, S., Romero, S., Grasa, E., Morte, E., & Barbanoj, M. J. (2007). Effects of olanzapine, risperidone and haloperidol on sleep after a single oral morning dose in healthy volunteers. *Psycopharmacology , 190*, 507-516.

Girard, P., Blaschke, T. F., Kastrissios, H., & Sheiner, L. B. (2002). A Markov mixed effect regression model for drug compliance. *Stat Med , 17*, 2313-2333.

Gregory, G. G., & Cabeza, R. (2002). A two-state stochastic model of REM sleep architecture in the rat. *J Neurophysiol , 88*, 2589-2597.

Hajak, G., & Bandelow, B. (1998). Safety and tolerance of zolpidem in the treatment of disturbed sleep: A post-marketing surveillance of 16,944 cases. *Int Clin Psycopharmacol , 13*, 157-167.

Henin, E., You, B., VanCutsem, E., Hoff, P. M., Cassidy, J., Twelves, C., et al. (2009). A Dynamic Model of Hand-and-Foot Syndrome in Patients Receiving Capecitabine. *Clin Phramacol Ther , 85* (4), 418–425.

Holford, N. (2005). VPC, the visual predictive check – superiority to standard diagnostic (Rorschach) plots. *PAGE 14.* Pamplona, Spain: Abstr 738 [www.page-meeting.org/?abstract=738].

Iber, C., Ancoli-Israel, S., Chesson, A., & Quan, S. F. (2007). *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specification.* Westchester, Illinois, USA: The American Academy of Sleep Medicine.

Ito, K., Hutmacher, M. M., Liu, J., Qiu, R., Frame, B., & Miller, R. (2008). Exposure-response Analysis for Spontaneously Reported Dizziness in Pregabalin-treated Patient With Generalized Anxiety Disorder. *Clin Pharmacol Ther , 84* (1), 127–135.

Jones, B. E. (2004). Activity, modulation and role of basal forebrain cholinergic neurons innervating the cerebral cortex. *Prog Brain Res , 145*, 157-169.

Jones, B. E. (2000). Basic mechanisms of sleep-wake states. In M. M. Kryger, T. Roth, W. C. Dement, & (eds), *Principles and practice of sleep medicine, ed 3* (pp. 134-154). Philadelphia, Pennsylvania, USA: Saunders.

Jonsson, E. N., & Karlsson, M. O. (1998). Automated covariate model building within NONMEM. *Pharm Res , 15* (9), 1463-1468.

Jonsson, S., Kjellsson, M. C., & Karlsson, M. O. (2004). Estimating Bias in Population Parameters for Some Models for Repeated Measures Ordinal Data Using NONMEM and NLMIXED. *J Pharmacokinet Pharmacodyn , 31* (4), 299-320.

Jouvet, M., Michel, F., & Courjon, J. (1959). Sur un stade d'activite electrique cerebrale rapide au cours du sommeil physiologique. *C R Soc Biol (Paris) , 153*, 1024-1028.

Karlsson, M. O., & Savic, R. M. (2007). Diagnosing model diagnostics. *Clin Pharmacol Ther , 82* (1), 17-20.

Karlsson, M. O., Schoemaker, R. C., Kemp, B., Cohen, A. F., van Gerven, J. M., Tuk, B., et al. (2000). A pharmacodynamic Markov mixed-effects model for the effect of temazepam on sleep. *Clin Pharmacol Ther , 68*, 175-188.

Kemp, B., & Kamphuisen, A. C. (1986). Simulation of human hypnograms using a Markov chain model. *Sleep , 9*, 405-414.

Kimko, H. H., & Peck, C. C. (2011). *Clinical Trial Simulations: Applications and Trends* (Vol. 1).

Kjellsson, M. C. (2008). *Methodological Studies on Models and Methods for Mixed-Effects Categorical Data Analysis.* Uppsala, Sweden: Uppsala Universitet.

Kjellsson, M. C., Ouellet, D., Corrigan, B., & Karlsson, M. O. (2008). Modeling Sleep Data for a New Drug in Development using Markov Mixed-Effects

Models. In M. C. Kjellsson, *Methodological Studies on Models and Methods for Mixed-Effects Categorical Data Analysis.* Uppsala, Sweden: Uppsala Universitet.

Kjellsson, M. C., Zingmark, P. H., Jonsson, E. N., & Karlsson, M. O. (2008). Comparison of proportional and differential odds models for mixed-effects analysis of categorical data. *J Pharmacokinet Pharmacodyn , 35* (5), 483-501.

Kohatsu, N. D., Tsai, R., Young, T., Van Gilder, R., Burmeister, L. F., Stromquist, A. M., et al. (2006). Sleep Duration and Body Mass Index in a Rural Population. *Arch Intern Med , 166*, 1701-1705.

Kowalski, K. G., & Hutmacher, M. M. (2001). Effi cient screening of covariates in population models using Wald's approximation to the likelihood ratio test. *J Pharmacokinet Pharmacodyn , 28* (3), 253-275.

Krueger, J. M., & Majde, J. A. (2003). Humoral links between sleep and the immune system: research issues. *Ann N Y Acas Sci , 992*, 9-20.

Leger, D. (2000). Public health and insomnia: Economic impact. *Sleep , 23(Suppl 3)*, S69-S76.

Lindbom, L., Pihlgren, P., & Jonsson, E. N. (2005). PsN-Toolkit--a collection of computer intensive statistical methods for non-linear mixed effect modeling using NON MEM. *Comput Methods Programs Biomed , 79* (3), 241-257.

Loomis, A. L., Harvey, N., & Hobart, G. A. (1937). Cerebral states during sleep, as studied by human brain potentials. *J Exp Psychol , 21*, 127-144.

Mai, E., & Buysse, D. (2008). Insomnia: Prevalence, Impact, Pathogenesis, Differential Diagnosis, and Evaluation. *Sleep Med Clin , 3* (2), 167-174.

Mandema, J. W., Verotta, D., & Sheiner, L. B. (1992). Building population pharmacokinetic-pharmacodynamic models. I. Models for covariate eff ects. *J Pharmacokinet Biopharm , 20* (5), 511-528.

Miller, A. J. (1984). Selection of Subsets of Regression Variables. *Journal of teh Royal Statistical Society, Series A - General , 147-3*, 389-425.

Miller, R., Ewy, W., Corrigan, B. W., Ouellet, D., Hermann, D., & Kowalski, K. G. (2005). How modeling and simulation have enhanced decision making in new drug development. *J Pharmacokinet Pharmacodyn , 32*, 185-197.

Nishino, S. (2007). The hypocretin/orexin receptor: therapeutic prospective in sleep disorders. *Expert Opin Invest Drugs , 16*, 1785-1797.

O'Neill, R., & Wetherill, G. B. (1971). Th e Present State of Multiple Comparison Methods. *Journal of the Royal Statistical Society, Series B - Methodological , 33* (2), 218-250.

Peck, C. C., Barr, W. H., Benet, L. Z., Collins, J., Desjardins, R. E., Furst, D. E., et al. (1992). Opportunities for integration of pharmacokinetics, pharmacodynamics, and toxicokinetics in rational drug development. *Pharm Res , 9*, 826-833.

Penzel, T., & Kesper, K. (2006). Physiology of sleep and dreaming. *Sleep Apnea , 35*, 13-20.

Pilcher, J. J., & Huffcutt, A. I. (1996). Effects of sleep deprivation on performance: A meta-analysis. *Sleep , 19*, 318-326.

Plan, E. L., Karlsson, K. E., & Karlsson, M. O. (2010). Approaches to Simultaneous Analysis of Frequency and Severity of Symptoms. *Clin Pharmacol Ther , 88*, 255-259.

Plan, E., Maloney, A., Traconiz, I. F., & Karlsson, M. O. (2008). Maximum Likelihood Estimation Methods: Performances in Count Response Models Population Parameters. *PAGE 17.* Marseille, France: Abstr 1372 [www.page-meeting.org/?abstract=1372].

Rao, M. N., Blackwell, T., Redline, S., Stefanick, M. L., Ancoli-Israel, S., & Stone, K. L. (2009). Association Between Sleep Architecture and Measures of Body Composition. *Sleep , 32* (4), 483-490.

Rechtschaffen, A., & Kales, A. (1968). *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects.* Washington DC, USA: Public Health SErvice, US Government Printing Office.

Redline, S., Kirchner, H. L., Quan, S. F., Gottlieb, D. J., Kapur, V., & Newman, A. (2004). The Effect of Age, Sex, Ethnicity, and Sleep-Disordered Breathing on Sleep Architecture. *Arch Intern Med , 164*, 406-418.

Ribbing, J. (2007). *Covariate Model Building in Nonlinear Mixed Effects Models.* Uppsala, Sweden: Uppsala Universitet.

Ribbing, J., & Jonsson, E. N. (2004). Power, selection bias and predictive performance of the population pharmacokinetic covariate model. *J Pharmacokinet Pharmacodyn , 31* (2), 109-134.

Roth, T., & Reehrs, T. (2003). Insomnia: epidemiology, characteristics, and consequences. *Clin Cornerstone , 5*, 5-15.

Roth, T., Walsh, J. K., Krystal, A., Wessel, T., & Roehrs, T. H. (2005). An evaluation of the efficacy and safety of eszopiclone over 12 months in patients with chronic primary insomnia. *Sleep Med , 6*, 487-495.

Sahlin, C., Franklin, K. A., Stenlund, H., & Lindberg, E. (2009). Sleep in women: Normal values for sleep stages and position and the effect of age, obesity, sleep apnea, smoking, alcohol and hypertension. *Sleep Med , 10*, 1025-1030.

Sankey, S. S., & Weissfeld, L. A. (1998). A study of the effect of dichotomizing ordinal data upon modeling. *Communication in Statistics - Simulation and Computation , 27*, 871-887.

Sateia, M. J., & Nowell, P. D. (2004). Insomnia. *Lancet , 364*, 1959-1973.

Sateia, M. J., Doghramji, K., & Hauri, P. J. (2000). Evaluation of chronic insomnia. *Sleep* , 243-261.

Sauerbrei, W. (1999). Th e use of resampling methods to simplify regression models in medical statistics. *Journal of the Royal Statistical Society, Series C - Applied Statistics , 48*, 313-329.

Savic, R., Wilkins, J., & Karlsson, M. O. (2006). (Un)informativeness of Empirical Bayes Estimate-Based Diagnostics test. *The AAPS Journal , 8* (S2).

Sheiner, L. B. (1994). A new approach to the analysis of analgesic drug trials, illustrated with bromfenac data. *Clin Pharmacol Ther , 56*, 309-322.

Sheiner, L. B., Beal, S. L., & Dunne, A. (1997). Analysis of nonrandomly censored ordered categorical longitudinal data from analgesic trials. *J Am Stat Ass , 90*, 1235-1255.

Sheiner, L., & Wakefield, J. (1999). Population modelling in drug development. *Stat Methods Med Res , 8*, 183-193.

Siegel, J. M. (2005). REM sleep. In M. H. Kryger, T. Roth, & W. C. Dement, *Principles and practice of sleep medicine* (pp. 120-135). Philadelphia, Pennsylvania, USA: Saunders.

Sinton, C. M., & McCarley, R. W. (2004). Neurophysiological mechanisms of sleep and wakefulness: a question of balance. *Semin Neurol , 24*, 211-223.

Snoeck, E., & Stockis, A. (2007). Dose–response population analysis of levetiracetam add-on treatment in refractory epileptic patients with partial onset seizures. *Epilepsy Res , 73* (3), 284-291.

Spruance, S. L., Reid, J. E., Grace, M., & Samore, M. (2004). Hazard Ratio in Clinical Trials. *Antimicrob Agents Chemother , 48*, 2787-2792.

Stanley, N. (2005). The physiology of sleep and the impact of aging. *Eur Urol Suppl , 3* (6), 17-23.

Stone, B. M., Turner, C., Mills, S. L., Paty, I., Patat, A., Darwish, M., et al. (2002). Noise-induced sleep maintenance insomnia: hypnotic and residual effects of zaleplon. *Br J Clin Pharmacol , 53* (2), 196-202.

Stromberg, U. (1996). Collapsing ordered outcome categories: a note of concern. *Am J Epidemiol , 144*, 421-424.

Sullivan, S. K., Petroski, R. E., & Verge, G. (2004). Characterization of the interaction of indiplon, a novel pyrazolopyrimidine sedative-hypnotic, with the GABAA receptor. *J Pharmacol Exp Ther , 311*, 537-546.

Traconiz, I. F., Plan, E. L., Miller, R., & Karlsson, M. O. (2009). Modelling overdispersion and Markovian features in count data. *J Pharmacokinet Pharmacodyn , 36* (5), 461-477.

Turek, F. W., & Gillette, M. U. (2004). Melatonin, sleep, and circadian rhythms: rationale for development of specific melatonin agonists. *Sleep Med* , 523-532.

Van Cauter, E., & Knutson, K. L. (2008). Sleep and the epidemic of obesity in children and adults. *Eur J Endocrinol , 159*, S59-S66.

Van Cauter, E., Leproult, R., & Plat, L. (2000). Age-Related Changes in Slow Wave Sleep and REM Sleep and Relationship With Growth Hormone and Cortisol Levels in Healthy Men. *J A M A , 284* (7), 861-868.

Van Cauter, E., Leproult, R., & Plat, L. (861-868). Age-Related Changes in Slow Wave Sleep and REM Sleep and Relationship With Growth Hormone and Cortisol Levels in Healthy Men. *J A M A , 284* (7), 2000.

Vitiello, M. V. (2006). Sleep in Normal Aging. *Sleep Med Clin , 1*, 171-176.

Wade, J. R., Edholm, M., & Salmonson, T. (2005). A guide for reporting the results of population pharmacokinetic analyses: A Swedish perspective. *AAPS Journal , 7* (2).

Wahlby, U., Bouw, M. R., Jonsson, E. N., & Karlsson, M. O. (2002). Assessment of type I error rates when modeling ordered categorical data in NONMEM. *J Pharmacokinet Pharmacodyn , 29*, 251-269.

Wahlby, U., Jonsson, E. N., & Karlsson, M. O. (2001). Assessment of actual significance level for covariate effects in NONMEM. *J Pharmacokinet Pharmacodyn , 28*, 231-252.

Wahlby, U., Matolcsi, K., Karlsson, M. O., & Jonsson, E. N. (2004). Evaluation of type I error rates when modeling ordered categorical data in NONMEM. *J Pharmacokinet Pharmacodyn , 31*, 61-74.

Walsh, J. K. (2004). Clinical and socioeconomic correlates of insomnia. *J Clin Psychiatry , 65(Suppl 8)*, 13-19.

Wang, Y. (2007). Derivation of various NONMEM estimation methods. *J Pharmacokinet Pharmacodyn , 34*, 575-593.

Zeger, S. L., & Liang, K. Y. (1992). An overview of methods for the analysis of longitudinal data. *Stat Med , 11* (14), 1825–1839.

Zingmark, P. H., Kagedal, M., & Karlsson, M. O. (2005). Modelling a Spontaneously Reported Side Effect by Use of a Markov Mixed-Effects Model. *J Pharmacokinet Pharmacodyn , 32* (2), 261-281.

Zisapel, N. (2004). Reviews. Insomnia: principles and management. *BMJ* , 328-355.