UNIVERSITY OF SOUTHAMPTON

FACULTY OF BUSINESS AND LAW

School of Management

Enabling health, independence and wellbeing for patients with bipolar disorder through Personalised Ambient Monitoring

By

Syed Golam Mohiuddin

Thesis for the degree of Doctor of Philosophy

January 2011

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF BUSINESS AND LAW
SCHOOL OF MANAGEMENT

Doctor of Philosophy

ENABLING HEALTH, INDEPENDENCE AND WELLBEING FOR PATIENTS WITH BIPOLAR DISORDER THROUGH PERSONALISED AMBIENT MONITORING

By Syed Golam Mohiuddin

This thesis describes the role of mathematical modelling in the evaluation of an innovative automated system of wearable and environmental sensors to monitor the activity patterns of patients with Bipolar Disorder (BD). BD is a chronic and recurrent mental disorder associated with severe episodes of mania and depression, interspersed with periods of remission. Early detection of transitions between the normal, manic and depressed stages is crucial for effective self-management and treatment. *Personalised Ambient Monitoring* (PAM) is an EPSRC-funded multidisciplinary project involving biomedical engineers, computer scientists and operational researchers. The broad aim of PAM is to build and test a network of sensors (chosen by the patient) to collect and analyse daily activity data in order to identify an 'activity signature' for that individual in various health states. The hypothesis is that small but potentially significant changes in this activity pattern can then be automatically detected and the patient alerted, enabling him/her to take appropriate action.

The research presented in this thesis involves the development and use of a Monte Carlo simulation model to evaluate the potential of PAM without the need for a costly and time-consuming clinical trial. A unique and novel disease state transition model for bipolar disorder is developed, using data from the clinical literature. This model is then used stochastically to test many different scenarios, for example the removal or technical failure of a sensor, or the limited availability of various types of data, for various simulated patient types and a wide range of assumptions and conditions. The feasibility of obtaining sufficient information to derive clinically useful information from a limited set of sensors is analysed statistically. The minimum best set of sensors suitable to detect both aspects of the disorder is identified, and the performance of the PAM system evaluated for a range of personalised choices of sensors.

Contents

List of Figures

List of Tables

Declaration of Authorship

I, Syed Mohiuddin, declare that the thesis entitled "enabling health, independence and wellbeing for patients with bipolar disorder through Personalised Ambient Monitoring (PAM)" and the work presented in it are my own. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University
- where any part of this thesis has been previously submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated
- where I have consulted the published work of others, this is always clearly attributed
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work
- I have acknowledged all main sources of help
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself
- Part of this work has been presented at:
    - Mohiuddin S and Brailsford S. Enabling health, independence and wellbeing for patients with bipolar disorder through Personalised Ambient Monitoring (PAM). 36th ORAHS conference, University of Genoa (18-23 Jul 10)
    - James C, Crowe J, Magill E, Brailsford S, Amor J, Prociow P, Blum J and Mohiuddin S. Personalised Ambient Monitoring of the Mentally Ill. 4th European Congress for Medical & Biomedical Engineering (22-27 Nov 08).

Signed: ……..

Date: …….

Acknowledgements

I have the great pleasure to express my deepest sense of gratitude to my supervisor, Professor Sally Brailsford, for her affectionate guidance and support throughout the course of this work. Thank you for all your valuable comments and constant inspirations from the beginning of my research throughout the writing process of this thesis. Thank you for strengthening my research skills through the brainstorming meetings and sharing your academic experiences. You are someone who I will always remember with respect.

I would also like to thank to all members of the PAM (Personalised Ambient Monitoring) team for the time they spent in offering me with important information to make this thesis more complete. The PAM teamwork broadened my inter-disciplinary formation. I greatly appreciate the time we spent together discussing how to make the PAM project successful.

It would be a crime not to thank my beloved wife, Shama, for her relentless support, inspiration and motivation. Ilham and Arham, our two precious sons, thank to both of you for giving us the pleasant touch to everyday of our lives. Finally, from the bottom of my heart, I thank my parents, family and friends for their support from the distance, your prayers kept me going.

With the oversight of my main supervisor, editorial advice has been sought.
No changes of intellectual content were made as a result of this advice.

1. Introduction

This chapter presents a general introduction to the thesis and describes the motivation for this study, the importance of the topic, the methodology used and the research contributions. The research described in this thesis concerns the role and contribution of Operational Research (OR) modelling in the Personalised Ambient Monitoring (PAM) project.

1.1 The PAM project

PAM is a multidisciplinary project involving the School of Management and the Institute for Sound and Vibration Research at the University of Southampton, the School of Electrical and Electronic Engineering at the University of Nottingham and the Department of Computing Science and Mathematics at the University of Stirling. The project was funded by the Engineering and Physical Sciences Research Council. The aims of this project were to develop a system of unobtrusive sensors that monitor the behaviour patterns of mental health patients, and hopefully detect changes in these behaviour patterns that may signal the early onset of an acute episode of illness. By then issuing an alert to the patient, such an episode could potentially be averted. The fundamental research questions underpinning the whole PAM project were:

- Is it possible to obtain, in an automatic, ambient and unobtrusive manner, 'activity signatures' from mental health patients that provide information about the trajectory of their health status?
- If this is so, can this information be used to assist their healthcare?

1.2 Mental illness

Worldwide, mental illness is a major social and economic problem. One in four of the UK population suffer from a mental health problem at some stage in their lifetime

(Singleton et al, 2001), while one in ten are likely to suffer for a disabling mental health problem at some point in their lives (James et al, 2008). 91 million working days are lost every year in the UK because of mental health problems (Gray, 1999). The cost of mental ill health to the UK in 2000 was estimated by the Mental Health Foundation as £32 billion, of which £12 billion is lost employment and productivity (BOHRF, 2005). Mentally ill people can find it difficult to form personal relationships, and thus often fail to progress at work and can experience breakdowns in personal relationships (Coryell et al, 1993; Calabrese et al, 2003).

1.3 Bipolar disorder

Bipolar disorder (BD) was selected for our study. BD is a severe form of mental illness linked with two types of serious episode, mania and depression, both of which drastically affect quality of life (Vojta et al, 2001; Michalak et al, 2007). Although mania is the defining trait of BD, depression is the major source of severe distress for patients themselves, and patients usually spend much more time in the depressed state than in the manic one (Judd et al, 2002; 2003). The World Health Organization (WHO) estimated depression and depression-allied illnesses to be the greatest cause of ailing health by the year 2020 (WHO, 2001). The prevalence of BD is increasing, and the age of onset is decreasing (Dienes et al, 2006). The prognosis for this disorder remains bleak, with repeated severe episodes interspersed with mild but significant symptomatic periods (Solomon et al, 1995).

Bipolar disorder can lead to significant psychological, functional, occupational and cognitive impairments, including higher rates of unemployment (Tse and Walsh, 2001), lower productivity and annual income (Goetzel et al, 2003), higher work absenteeism (Simon, 2003; Goetzel et al, 2003), episodic antisocial behaviour (APA, 2000), reduction in life expectancy (Goodwin and Jamison, 1990), suicidality (Judd and Akiskal, 2003), and changes in attention, planning and working memory (Ferrier and Thompson, 2002; Ågren and Backlund, 2007). Furthermore, despite the existence of pharmacological treatment, about 30-60% of BD patients fail to regain full social

and occupational functionality between episodes (Calabrese et al, 2003; MacQueen and Young, 2001).

In 1990 the WHO ranked bipolar disorder as 22nd among all diseases in terms of worldwide burden of illness. The disorder was also ranked as the 6th leading disabling illness worldwide (WHO, 2001). BD is associated with high costs for social and health care. Higher dependence on public assistance (Judd and Akiskal, 2003) and increased healthcare use and costs (Judd and Akiskal, 2003; Simon, 2003) have been shown to be closely associated with this disorder. In 2002, the annual cost of managing bipolar disorder in the UK NHS was estimated to be £199 million, of which £70 million was spent on hospital admissions (Gupta and Guest, 2002). However, despite its severely disabling nature, BD can be managed effectively through self-monitoring (AstraZeneca, 2010). Many bipolar patients are reportedly keen to monitor their condition regularly in order to minimise the severity of their episodes.

1.4 Relapses in bipolar disorder

Bipolar disorder is a chronic and recurrent illness associated with great morbidity and mortality (Müller-Oerlinghausen et al, 2002). Sadly, the mortality rate in BD is two to three times higher in comparison with the general population (Müller-Oerlinghausen et al, 2002; Belmaker, 2004). The risk of relapse for a bipolar patient increases over time, and can differ from a few weeks to many months. The need to prevent bipolar relapses was shown in a prospective naturalistic study by Tohen et al (1990), where the risk of relapse was found to be 90% within five years. In the survival analysis by Gitlin et al (1995), the risk of relapse was shown as 73% within five years, and two-thirds of those who relapsed suffered multiple relapses.

Bipolar patients require lifetime maintenance therapy (Müller-Oerlinghausen, 2002). It is easier to treat milder symptoms in the early stage of a relapse than more severe symptoms later in the relapse (Morriss et al, 2004). The importance of analysing the early warning signs (i.e. *prodromes*) of relapse is therefore clear: if BD

patients can identify prodromes early enough, actions can be taken to avert the progress of a full-blown episode. Equally, prodromes of relapse are useful indicators to patients themselves, family members or clinicians, given that extra support may be required to stop prodromes progressing into a full-blown episode. Each episode usually begins with a similar pattern of symptoms that is distinctive for each individual; as such, it is often possible to detect unexpected mood changes leading to an imminent episode.

Common prodromes of mania include decreased need for sleep, increased activity, elevated mood and racing thoughts and speech, while prodromes of depression include interrupted sleep, decreased activity, empty mood and loss of interest (Lam et al, 2001). Bipolar patients are known to be able to detect prodromes about two to four weeks before a full manic or depressive relapse (Lam and Wong, 1997; Altman et al, 1992). They are also known to be able to report prodromes reliably (Molnar et al, 1988; Lam et al, 2001). Generally, BD patients are well-informed about their illness and keen to manage their condition themselves. This self-consciousness can be used as a significant factor in managing the condition, since early detection of prodromes is of explicit importance. It is, thus, possible for BD patients to learn how to identify specific personal "stressors" or triggers associated with manic and depressive relapses, and to develop strategies for minimising the risk of these triggers progressing into full-blown episodes.

1.5 Human suffering associated with BD

The condition itself is highly distressing and disruptive; both for the patient and their family, and the medication required for BD treatment is powerful and has many unpleasant side-effects. However, this is not all that BD patients have to suffer. There may be delays in diagnosis and timely treatment, problems with the delivery of care, a complex system to be navigated when seeking further information and communicating with caregivers, and of course there is a stigma related with all mental disorders. About 25% of bipolar patients were found to have never sought help from health services in a community survey (ten Have et al, 2002). Health

Minister Rosie Winterton declared in 2006 that the UK Government "will help people with long-term conditions such as cancer, or mental health problems, to stay independent and take control of their illness" by prescribing information alongside medicines, to allow people taking control of their own illness (James et al., 2008).

1.6 Pharmacological and psychological treatments

Several drugs such as lithium, olanzapine, valproate, lamotrigine and imipramine are available for the treatment of bipolar disorder, but (in addition to the unpleasant side-effects of these drugs) patients commonly experience multiple relapses and frequent oscillations in symptom severity, despite ongoing maintenance therapy (Tohen et al, 2005). Pharmacological treatments cannot control issues such as medication adherence, early detection of prodromes, alertness of the disorder and improving coping skills. Since drug treatment is only partially successful, psychosocial interventions are often combined with maintenance pharmacotherapy in order to target all aspects of the disorder and thus improve overall treatment outcome. Surveys have shown that many bipolar patients are very keen to use psychosocial therapy and self-management approaches in addition to pharmacological treatment (Lish et al, 1994; Hill et al, 1996).

Lithium has been commonly used, both in the acute phases of the illness and as maintenance therapy, for more than 30 years. However, about 20-40% of patients did not respond to lithium prophylaxis compared with the normal control subjects in Lam et al's (2000) study. Moreover, around 75% of patients on lithium report unwanted side-effects such as irritability and mental distress (Fava et al, 1984, 1987; Johnson and Leahy, 2004). Despite the increased use of lithium, rising admission rates to British hospitals for mania were reported by Dickson and Kendell (1986) and Symonds and Williams (1981). Antidepressants may actually trigger an aggravated episode of bipolar illness (Ågren and Backlund, 2007). The possibility of inducing mania has been reported for almost all antidepressants (Kukopulos et al, 1980; Wehr and Goodwin, 1987; Kupfer et al, 1988; Himmelhoch et al, 1991; Peet, 1994; Altshuler et al, 1995; Joffe et al, 2002), although the reported rates vary.

As part of psychosocial therapy, bipolar patients are encouraged to detect and manage prodromes of an imminent episode. Patients who are able to detect and take action to cope with their prodromes early enough are better able to prevent the prodromes developing into a full-blown episode (Russell and Browne, 2005). Early detection and treatment also help prevent suicidal behaviour (APA, 2002). Self-reporting of daily sleep and mood fluctuations is an established clinical tool for the clinician to comprehensively assess frequency and pattern of bipolar disorder (Bauer et al, 1991; Leverich and Post, 1996).

The American Psychiatric Association has recommended psychosocial intervention as a "cornerstone" of treatment for bipolar disorder (APA, 2002). Teaching patients how to recognise prodromes and to act upon them is a significant element of psychosocial therapy (Lam et al, 2001). In particular, teaching patients to identify the early symptoms of the onset of mania can provide a significant positive impact in patients' lives (Perry et al, 1999). Lam and Wong (1997) state that teaching patients to monitor their moods can contribute significantly to their level of social functioning.

1.7 Self-monitoring in BD

Self-management therapy to identify prodromes is very common today, but in clinical practice most implementations are still paper-based (Baldassano, 2005). In healthcare and clinical research, computer-based data collection has been shown to be more accurate than paper-based data collection by minimising errors in data entry (Whybrow et al, 2003; Lane et al, 2006). Electronic methods for data collection have been reported to be better than paper-based data collection for many reasons, including increased motivation, privacy and secrecy, and reduced time, effort and cost required (Whybrow et al, 2003, Bauer et al (2004).

Although both paper-based and computer-based self-monitoring processes have been shown to benefit patients and their caregivers, this is only to a limited extent. Over time, the usefulness of both systems can be reduced owing to various

"behavioural" factors; e.g. patients may simply forget to write the diary, or may be too depressed (or manic) to do so; alternatively, they may become complacent about the whole process and feel it is unnecessary, while others may feel embarrassed by the effort required to do the self-monitoring (Morriss, 2004). Patients have also been known to fabricate diary entries immediately before a hospital visit, and obviously under such circumstances their recall of events may be incorrect and biased (Kobak et al, 2001).

Of course, the accuracy of both paper-based and computer-based self-monitoring mood and activity diaries is questionable, due to their reliance on patients' commitment, openness and honesty when describing their thoughts and feelings. For instance, especially during the early phases of mania, many patients deny anything is wrong and avoid seeking treatment, even though other people can easily see something unusual is beginning to happen. The longer treatment is delayed in a manic episode, the worse the eventual outcome. Reliance on a paper-based or computer-based tool can lead to delay in initiating the necessary treatment for such an episode. An automated monitoring system would avoid this problem and would facilitate the delivery of timely bipolar healthcare treatment. Hence, the PAM project has involved the development of an unobtrusive network of small wearable and environmental sensors, mobile phones and computers. The PAM system supports the automated self-monitoring of a range of personal activity information to help detect any imminent bipolar transitions. Chapter 3 provides a detailed discussion of the PAM project and how the system works.

1.8 Research questions in this thesis

We differentiate between the research questions addressed by the PAM project as a whole, and those addressed by this thesis. The main aim of the Operational Research modelling in the PAM project was to address the following issue: *can OR modelling approaches help in the design of a system to monitor 'activity signatures' for BD patients*? This broad objective was then subdivided into the following research questions, which are addressed in this thesis:

- Can a “natural history” model be developed for BD, using clinical data from the medical literature?
- What type of modelling approach is best suited for such a model?
- Is adequate clinical data available, and if not, can expert opinion be used instead?
- Can simulation be combined with this natural history model to capture the inherent variability and uncertainty in the real-world system?
- Can this model then be used to compare and evaluate different sets of indicator responses, from different configurations of sensors, in order to describe an individual’s ‘activity signature’?
- Can the model assist the design of PAM by minimising the number of sensors required, for a given individual?

1.8.1 Rationale of research questions

PAM is clearly a highly complex healthcare monitoring system, in which many decisions have to be made concerning the choice and configuration of sensors. There are also issues concerning algorithm choice and the technical reliability of the PAM system itself. In addition to the acceptability issues already mentioned, there are various constraints on PAM. These include energy limitations and the need to prolong the network lifetime, the characteristics and lifestyles of the selected patient group and obviously, economic considerations. Therefore, the number of active sensors should be kept to a minimum. However, if there are too few sensors, the information they provide may be of insufficient value to be useful. In some cases, if a patient is unwilling to accept a particular type of sensor, then PAM may not be effective for that person. OR can help to resolve this conflict through the use of models designed to evaluate different sensor configurations for different patients. OR modelling, which is typified by the use of mathematical and computer based models to practical

problems, has played a vital role in helping PAM to test which configurations of PAM sensors could potentially be useful in practice.

"Natural history" of a recurrent and chronic disease like bipolar disorder is significant for clinical implications. To improve the understanding for clinical practices, it was vitally important to acquire the knowledge of a description of the continuous progression of bipolar disorder in patients from earliest pathological change until recovery over time. Knowledge of sequence of developments of the disorder was fundamental to check if the PAM technology is good enough for effective relapse detection. We had little empirical evidence and no clear data. Without the use of OR simulation, it would have been hugely costly and impracticable to put the human efforts to do this in reality. The costs and benefits of applying simulation modelling have been shown to be far less than trying out the reality (Gordon, 2001; Componation et al, 2003).

The first step was to develop a mathematical model of the "natural history" of bipolar disorder in which some kinds of evidence or observations can be used to calculate the probability that a hypothesis may be true. However, modelling any chronic disease process presents various challenges. The model must be stochastic, since the occurrence of a relapse in a chronic disease like BD is a random event. Moreover, the length of an acute episode varies from patient to patient, due to the inherent variability of the disease process. The progression of patients through the natural history of BD is no different to this. Bipolar disorder is a serious problem and that real-world clinical decisions are to be made in the face of uncertainty, so any models which can reduce risk and uncertainty are a good device. It was imperative to rely on a stochastic process in order to reach conclusions about future events.

Integrating patient specific data within an OR model of a healthcare process can be beneficial to capture resource utilisation. With this in mind, we required to know how to develop a framework in which we can combine the BD disease state transition model with simulation of the longitudinal outcomes of interventions. This framework was then used to test a whole range of different PAM sensor

configurations for different (simulated) patients. The model parameters were derived from the clinical literature. Model parameters were then varied stochastically for each individual, to capture patient-to-patient individuality in terms of dwelling times and pathways. The effects of the possible problems of missing or unreliable data within the system were also explored in the model.

The PAM project as a whole was essentially a feasibility study. There was not enough time to undertake a full clinical trial, although a technical trial of the equipment was performed by the PAM team themselves and a small-scale trial on one patient was carried out (see Section 3.8). Therefore, a significant contribution of this simulation model was to perform an "artificial" clinical trial on simulated patients.

1.9 Modelling in the PAM project

Bipolar episodes can become severely disabling and repeated if effective treatment is missing. To improve treatment outcomes, close analysis of episodes and functioning over time is vital. Longitudinal profiles of bipolar patients are diverse and complex, and these pose a challenge to find a model that best explains the longitudinal patterns of bipolar affected people. The longitudinal course and outcomes of bipolar disorder have largely been studied using descriptive statistics, logistic regression and survival analysis (Marneros and Brieger, 2002). However, it is important to build models for better understanding and analysis of bipolar disorder since a modelling approach may help clarify the dynamic process inherent with the disorder.

How to build a model that understands and analyses the dynamic behaviour of bipolar patients? It is well-known that models of dynamic behaviour may often be best described by a cycle of events that change with respect to time, where the events are the transitions among the states. This is typically called a state-based modelling approach in which the set of valid states of a dynamic process and probabilistic transitions among the states are defined. For bipolar disorder, we clearly needed to use a dynamic approach since the disorder evolves over time. However, the challenge was to define the clinical states required for an OR simulation model. Another

challenge was to construct a model of how the PAM system works and then experiment different policies or decisions using the model to observe what works.

Whether we use a Markov or a discrete-event simulation model, it is impossible for a model to predict any outcome with a hundred percent accuracy. However, it is certainly possible to acquire a high enough prediction through an appropriate method. Pidd (2010) emphasises that "models are approximations, built with some intended use(s) in mind and that they are the product of human thought and ingenuity". What is the use of a model if the outputs of the model are approximations? "Essentially, all models are wrong, but some are useful" (Box and Draper, 1987).

The dynamic "natural history" of bipolar disorder is unpredictable, not just for an individual person, but also from one person to another, so no two people will experience similar patterns of disorder. This makes bipolar disorder modelling stochastic and so difficult. On the one hand, bipolar disorder is known to be complex in nature, and on the other hand, keeping a model simple is argued to be at the centre of good modelling practice. Due to the diversity and distinctive attitudes of the participants involved, selecting the most appropriate method to describe the bipolar disease progression was not simple. However, a Markov model, based on the natural history of bipolar disorder, was developed in Excel. This model was then combined with Monte Carlo simulation to provide information on the clinical value of the information that PAM could deliver for a range of different assumptions, including sensor failure for technical reasons, deliberate removal of a sensor, flexible monitoring, or the limited availability of various types of data.

The literature (Sonnenberg et al, 1994; Barton et al 2004) suggests that Markov modelling and Monte Carlo simulation might be a fruitful line to pursue since interaction between patients is irrelevant, a long time period needs to be modelled and events in the model are recurrent. Furthermore, this approach helps construct a less complex model that would be more advantageous in term of

flexibility, data requirement, speed and interpreting the outcomes. Mental illnesses are usually unpredictable in nature, and Monte Carlo simulation is a familiar approach to deal with uncertainties. The use of Monte Carlo simulation enabled us to explore the uncertainties around the actual performance and effectiveness of the PAM system in practice, but with artificial, simulated patients. The model issues alerts to these simulated patients by comparing their current behaviour with their "normal" behaviour, using a set of decision rules and threshold levels. The aim of the model was to evaluate this automated routine decision making without human intervention.

We estimated the model parameters at the individual level in our Monte Carlo trial. It was fast to run, and was also advantageous to carry out "what-if" experimentation. The Monte Carlo experimentation revealed that the Markov model is able to capture the true dynamic behaviour and detect a patient's true health state over time. The computational demand of our multi-state model was met using the Microsoft Excel (VBA coding) and "@Risk" software.

1.9.1 Challenges for modelling bipolar disorder

To construct the OR simulation model, we had to deal with the following challenges:

- The bipolar illness manifests itself in various forms
- The difficulty of developing the "natural history" model due to the illness's characteristics, e.g. a patient can have a mixed (mania with depression) state
- We did not come across any OR modelling approach of longitudinal monitoring of bipolar disorder in the literature, although OR modelling has been applied to unipolar major depression, e.g. Patten and Lee (2005) used a Markov model to describe the longitudinal course of major depression
- No universally accepted staging models for bipolar disorder were found in the medical literature, compared with physical diseases such as cancer

- Lack of easily measurable criteria: diagnosis is a very individual process, based on self-reports by patients on their moods or behaviour, and the observation and judgment of the psychiatrist.

1.10 Scientific contribution

As in Section 1.8, it is vital to differentiate between the scientific contribution of the PAM project as a whole, and the contribution of this thesis. Therefore, in the following, "*PAM*" is used to denote the project as a whole and "*thesis*" denotes the specific contribution of the modelling research in this thesis:

- The application of multidisciplinary approaches to develop an innovative, acceptable and (hopefully) effective approach to managing BD (*PAM*)
- The use of sensor technology, computer networks and biomedical engineering tools to automatically detect and analyse human activity patterns (*PAM*)
- Developing a model for the natural history of BD, capable of capturing individual variability yet consistent with the clinical literature (*thesis*)
- Using this model to evaluate a technology to provide automatic alerts to BD patients, by designing and testing different prodrome detection algorithms for different simulated patients (*thesis*).

1.11 Research paradigm

PAM is a knowledge generating system in which the activity patterns of bipolar patients are studied. The aim of the research is to acquire scientific knowledge in order to improve patient care. The OR modelling aspect of PAM is mainly concerned with the collection and analysis of secondary data, and the subsequent development and use of a quantitative OR model to conduct experiments to test the hypothetical PAM system (see Figure 1.1). Therefore, the research paradigm of this thesis is a conventional positivistic scientific paradigm.

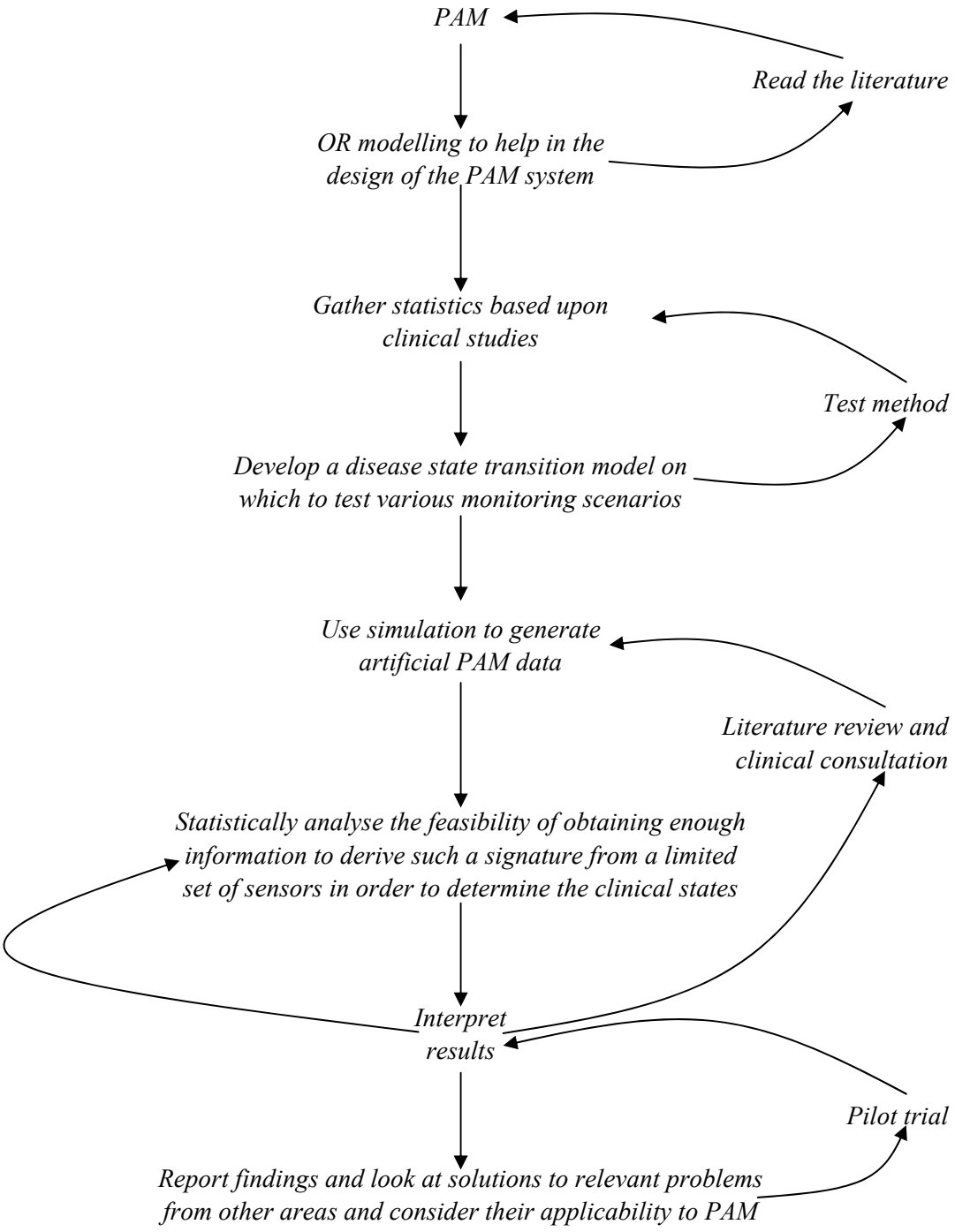


Figure 1.1 The research process in this thesis

1.12 Structure of the thesis

The remainder of this thesis is organised as follows:

- *Chapter 2* describes the epidemiology of bipolar disorder

- *Chapter 3* describes the PAM project in more detail and clarifies the role of OR modelling within the overall project
- *Chapter 4* discusses various diagnostic tests and scoring systems for bipolar disorder. It also compares the sensitivity and specificity of different BD diagnostic scoring systems
- *Chapter 5* briefly reviews applications of modelling and simulation in healthcare. It concludes by summarising the essential aspects of developing OR simulation models
- *Chapter 6* concentrates on mental illness and mental health, and discusses the (limited) use of OR modelling in mental healthcare
- *Chapter 7* explains the structure, data requirements and other components of the simulation model, and explains the assumptions that were made
- *Chapter 8* presents the model results, discusses their accuracy and includes a sensitivity analysis of the key model parameters
- *Chapter 9* contains an overall discussion of the study findings and its limitations. It concludes with a summary of the motivations, goals and findings of this research and provides suggestions for future work.

2. Bipolar Disorder

This chapter provides an overview of bipolar disorder (BD), previously known as manic-depression. As described in Sections 1.3 – 1.5 of Chapter 1, BD is a major problem worldwide both in terms of human suffering and of economic costs.

2.1 Introduction

Bipolar disorder is a mental illness that causes atypical swings in a person's mood, energy and ability to function. Manic behaviour is one extreme of this disorder where the patient feels excessively euphoric and elated, and depression is the other extreme where the patient feels extremely low and even suicidal. The name bipolar disorder officially replaced the older term "manic-depression" in 1980, although Leonhard (1957) first used the term "bipolar" to describe patients who exhibited both mania and depression. Angst and Perris separately showed the distinction between unipolar depression and bipolar disorder in 1966. The symptoms of bipolar disorder are far more acute than the natural "ups and downs" that everyone undergoes in daily life.

It is often difficult for bipolar patients to recognise the degree of impairment they suffer. Some bipolar patients even deny the fact that they are suffering from a mental illness at all, and this flawed awareness of their own illness makes it difficult for them to accept treatment. The deep mood swings of bipolar disorder may last for weeks or even months, and cause great inconvenience in the lives of spouses, family members, friends, colleagues and employers. According to the National Institute of Mental Health (NIMH, 2010), BD patients frequently experience personal, social and job-related problems. Life expectancy has also been shown to be reduced by this disorder (AstraZeneca, 2010).

BD usually starts in the teenage years or early adulthood and carries on throughout life. Some people experience their first symptoms in childhood, and some late in life. BD can start in childhood but may not be diagnosed until the person is

much older. The median age of onset is 25 years (Kessler et al, 2005). The average age at onset ranged from 17.1 to 29.0 years in a general review of six epidemiological surveys (Weissman et al, 1996). Due to its episodic nature, BD is often not identified as a psychological problem, explaining why some patients may suffer unnecessarily for years without proper treatment.

Bipolar disorder is a long-term condition similar to diabetes or heart disease, which must be carefully controlled throughout a person's life. However, BD is treatable, and patients can lead healthy and productive lives. Without treatment, however, the condition normally deteriorates, and can even lead to suicide in addition to other serious symptoms. However, most bipolar patients do not regard their manic episodes (at least in the early stages) as needing treatment, and are reluctant to seek it. BD patients experience very pleasurable feelings during the onset of a manic episode, and understandably they do not want this sense of euphoria to end. Due to this serious judgment problem, patients lose insight into their own condition to the point at which it becomes too late and negative behavioural aspects appear as the manic episode advances. Unfortunately, the National Collaborating Centre for Mental Health (NCCMH, 2006) states that "affective or functional changes occurring prior to the development of bipolar disorder have not been studied systematically".

2.2 The symptoms of mania

The signs and symptoms of mania include (NIMH, 2009; Lam et al, 2001; Morriss, 2004):

- increased energy and activity
- feeling very high (euphoric)
- increased talkativeness – talking very fast and loud
- racing thoughts and ideas, reduced concentration span
- decreased need for sleep
- creative and strong feeling of omnipotence
- increased risk-taking behaviours

- stronger interest in sex
- poor judgment and lack of inhibitions
- spending money more freely (to the point of running up debt)
- provocative, disturbing or aggressive behaviour
- abuse of alcohol and drugs such as cocaine and sleeping medications
- denial that there is any problem.

Good coping strategies with the early signs of mania include deliberately keeping control of one's actions, performing relaxing activities, prioritising and reducing the number of everyday jobs, taking medication as prescribed, delaying impulsive actions, talking to someone to bring reality to one's thoughts, taking time off work, reducing alcohol intake and stopping taking street drugs (Morriss, 2004).

2.3 The symptoms of depression

The signs and symptoms of depression include (NIMH, 2009; Lam et al, 2001; Morriss, 2004):

- persistent sad, apprehensive or empty mood
- fatigue or loss of energy
- loss of interest or pleasure in usual activities such as people, sex and food
- pessimistic feelings that the person is useless and worthless
- less talkative than usual, slow speech
- ideas slow down
- difficulty remembering, concentrating and making decisions
- sleeping too much
- change in appetite
- unwanted weight loss or gain
- recurrent thoughts of death or suicide
- heavy drinking of alcohol.

Good coping strategies for depressive signs include keeping active, meeting people, getting organised, getting the support of family or friends, recognising

impractical thoughts, avoiding negative thoughts, taking medication as prescribed by the doctor, stop worrying and reduce alcohol intake (Morriss, 2004).

2.4 The episodic nature of bipolar disorder

According to the American Psychiatric Association (APA, 2010), a manic episode is characterised by a distinct period of unusually and persistently elevated, expansive or irritable mood, lasting for at least one week. A *hypomanic* episode is characterised by a distinct period consisting of at least four days during which there is a persistently elevated, expansive or irritable mood, clearly dissimilar from a person's usual mood. On the other hand, a major depressive episode is characterised by a period of time where a depressed mood or a loss of interest or pleasure in daily activities lasts unusually and consistently for at least two weeks. A mixed episode is characterised by meeting the diagnostic criteria (except for time period) for both a manic episode and a major depressive episode nearly every day for at least a week.

The differentiable feature of bipolar disorder is the occurrence of at least one manic episode compared with other mood disorders. In addition, BD is a chronic condition because patients who have one manic episode will almost always have further episodes. A person with this condition can experience any number of episodes throughout their life. Without preventive treatment, the NIMH (2009) statistics suggest that a patient may on average undergo four episodes in ten years. The actual length of each episode does not normally vary much over time, but episodes may start suddenly.

Women usually start with a depressive episode, whereas men are more likely to start with a manic episode (AstraZeneca, 2010). Furthermore, women are likely to have more depressive episodes than men. However, for both sexes the length of the remission periods between episodes tends to decrease with time (Goodwin and Jamison, 1990). The pattern of mood cycles differs from one patient to another, but it is usually predictable once identified. BD patients who start with depression can expect more depressive episodes over the course of their illness than those with a

manic onset (Perlis et al, 2005; Turvey et al, 1999). In Perugi et al's (2000) study, about half of BD patients began with depression and half with mania.

2.5 The prevalence of bipolar disorder

People of all races, ethnic groups and socio-economic backgrounds may be affected by this disorder. NIMH has stated that more than two-thirds of bipolar sufferers have at least one close relative either with BD or unipolar major depression. According to Burgess (2006), approximately 2% - 7% of US citizens suffer from BD. Burgess (2006) also estimated that 10 million US people will be affected by this disorder sometimes throughout their lives; and of these, about half will never receive accurate diagnosis or treatment. Also, according to Burgess (2006), there are approximately 723,250 bipolar affected people in the UK, while in each of India and China there are approximately between 12 and 15 million affected people.

The prevalence of BD within European countries has been estimated to be 6% (Pini et al, 2005). A prevalence rate of about 11% was reported by Swiss researchers in a 20-year prospective study of young adults with BD (Angst et al, 2003). Lifetime prevalence rates in Australia vary from 0.45 to 5.5% (Morgan et al, 2005, Goldney et al, 2005). The proportion of the people with BD seems to be increasing, while the age of onset is falling (Dienes et al, 2006).

2.6 Types of bipolar disorder

Bipolar disorder symptoms are characterised by emotional highs and lows whose intensity can be mild, moderate or severe. Mood swings can be frequent or infrequent. Depending upon the symptoms, there are several types of bipolar disorder, as discussed below.

2.6.1 Bipolar type I disorder

People with bipolar type I disorder experience episodes of severe mood swings between mania (at least one manic episode) and depression (with or without a previous episode of depression). This is the most common type of bipolar disorder.

Judd et al (2002) found that patients with bipolar type I spend less time manic than depressed (see Figure 2.1 below). About the same number of men and women are known to be affected by this type (Lloyd et al, 2005). The lifetime prevalence rates of bipolar type I in Europe range from 0.1 to 2.4% (Faravelli et al, 1990; Szadoczky et al, 1998; ten Have et al, 2002; Regeer et al, 2004; Pini et al, 2005). Rates reported in non-European countries range (from 0.3 to 1.5%) less widely than European countries (Weissman et al, 1996). In Australia, a lifetime prevalence rate of 2.5% was reported in a recent study (Goldney et al, 2005).

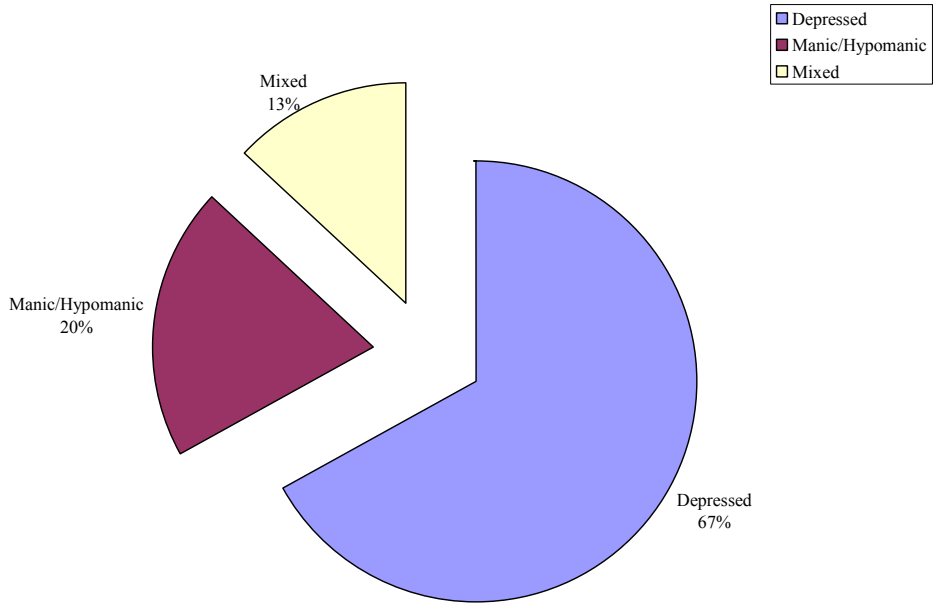


Figure 2.1 Symptoms during bipolar type I disorder. Source: Judd et al (2002)

2.6.2 Bipolar type II disorder

People with bipolar type II disorder experience episodes of severe depression alternating with milder episodes of mania, i.e. hypomania. Bipolar type II disorder became an officially recognised diagnosis by the APA following the introduction of the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV, 1994). This is a milder type of bipolar disorder than bipolar type I. However, bipolar type II can also be debilitating. Judd et al (2003) found that patients with bipolar type II spend considerably more time in the depressive phase of the illness than the manic phase (see Figure 2.2).

Women (29.0%) were more commonly shown to be affected than men (15.3%) (Baldassano et al, 2005). In contrast, Szadoczky et al (1998) found no gender difference in the prevalence of bipolar type II disorder. However, prevalence rates of bipolar type II reported in European studies vary between 0.2 and 2.0% (Faravelli et al, 1990; Szadoczky et al, 1998). Estimates of the lifetime prevalence of bipolar type II disorder reported in the US vary (from 5.5 to 10.9%) more widely than European countries (Angst, 1998; Angst et al, 2003).

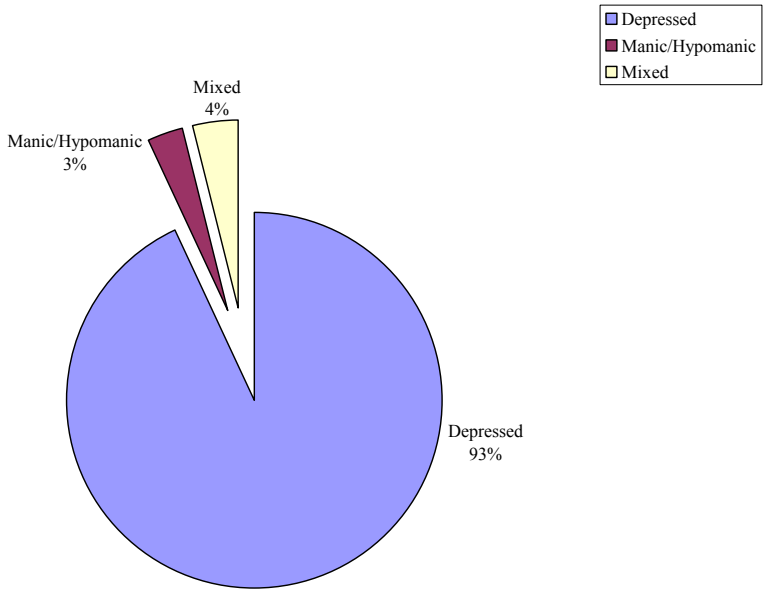


Figure 2.2 Symptoms during bipolar type II disorder. Source: Judd et al (2003)

2.6.3 Cyclothymic disorder

The mood swings during cyclothymia are less severe than those in bipolar types I and II. Symptomatically, this is a mild form of bipolar II disorder consisting of episodes of mood swings between hypomania and a mild form of depression. The prevalence of cyclothymic disorder is roughly the same for men and women. Diagnosing cyclothymia can be difficult since cyclothymic mood swings may appear within the normal range of ups and downs. Cyclothymic disorder episodes tend to last longer than other forms of bipolar disorder, and this may eventually lead to severe mood episodes (Akiskal et al, 1979).

2.6.4 Mixed bipolar disorder

In the context of a mixed state bipolar disorder, symptoms of both mania and depression take place simultaneously. People with any type of bipolar disorder may found themselves in mixed states. Mixed bipolar disorder can often be found in children and adolescents. Around 30 to 40% of bipolar I patients are reported to go through mixed episodes (Swann, 1995; Akiskal et al, 2000), while the figure is around 49% for bipolar II patients (Benazzi, 2000). The risk of mixed bipolar disorder is higher for patients whose immediate family members also suffer from bipolar disorder. Moreover, this is the most unstable state for risk of suicide, because stimulation and agitation of mania are combined with irritable depressive symptoms.

2.6.5 Rapid cycling bipolar disorder

Rapid cycling, which can occur in both bipolar I and II, is the occurrence of four or more mood swings within a year, as defined in DSM-IV. However, this may occur later in the course of illness and increases the risk of severe depression and suicide attempts. Women are more likely than men to experience rapid cycling (Schneck et al, 2004). Akiskal et al (2000) reported a 20% prevalence of rapid cycling in bipolar patients. Papadimitriou et al (2005) state that patients with bipolar type II suffer comparatively more frequently from rapid cycling.

2.6.6 Bipolar disorder Not Otherwise Specified (NOS)

There are occasions when a person evidently appears to be suffering from bipolar disorder, but does not fit the diagnostic criteria for any of the categories discussed above. In a case like this, the person is diagnosed as suffering from bipolar disorder NOS. For example, recurring hypomanic episodes with depressive symptoms in between can be classified under this category.

2.7 Patient descriptions of mood states

The various mood states of bipolar disorder can be thought as a spectrum. Severe mania is at one end, then hypomania, then euthymia (i.e. normal mood), then mild to

moderate depression and then severe depression (see Figure 2.3). There may also be mixed bipolar states, when symptoms of mania and depression occur together. To devise effective treatment plans, it is crucial to be able to identify the various mood states.

Figure 2.3 The range of moods from severe mania to severe depression. Source: NIMH (2010)

The NIMH US website also provided the following descriptions (by bipolar affected people themselves), which offer important insights into the various mood states associated with the disorder:

Mania:

The fast ideas become too fast and there are far too many...overwhelming confusion replaces clarity...you stop keeping up with it--memory goes. Infectious humor ceases to amuse. Your friends become frightened...everything is now against the grain...you are irritable, angry, frightened, uncontrollable, and trapped.

Hypomania:

At first when I'm high, it's tremendous...ideas are fast...like shooting stars you follow until brighter ones appear...all shyness disappears, the right words and gestures are suddenly there...uninteresting people, things, become intensely interesting. Sensuality is pervasive, the desire to seduce and be seduced is irresistible. Your marrow is infused with unbelievable feelings of ease, power, well-being, omnipotence, euphoria...you can do anything...but, somewhere this changes.

Depression:

I doubt completely my ability to do anything well. It seems as though my mind has slowed down and burned out to the point of being virtually useless....[I am]

haunt[ed]...with the total, the desperate hopelessness of it all... Others say, "It's only temporary, it will pass, you will get over it," but of course they haven't any idea of how I feel, although they are certain they do. If I can't feel, move, think, or care, then what on earth is the point?

2.8 Depressive symptoms in bipolar disorder

Although mania (or hypomania) is the defining feature of bipolar disorder, depressive symptoms are more common in the longitudinal course of illness than manic or hypomanic symptoms. According to the outcome of a 12.8-year prospective longitudinal study consisting of 146 patients with bipolar I disorder completing weekly mood ratings (Judd et al, 2002), patients spent about 53% of the time asymptomatic, 32% depressed, 9% manic and 6% in cycling/mixed states (see Figure 2.5). Depression was reported to be over three times more common than mania, and patients were symptomatic about 47% of the time.

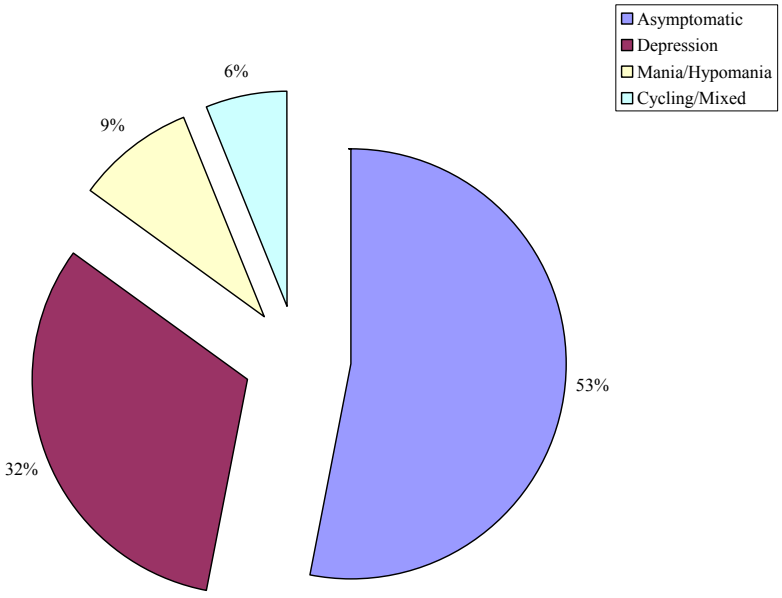


Figure 2.4 The amount of time spent in various states. Source: Judd et al (2002)

In another prospective longitudinal study (13.4-year) consisting of 86 patients with bipolar type II disorder completing weekly mood ratings (Judd et al, 2003), patients spent 46.1% of their time asymptomatic, 50.3% depressed, 1.3% manic and 2.3% cycling/mixed states. Patients were reported to experience mood symptoms

53.9% of the time. Depression was reported to be about thirty-eight times more frequent than mania, but this is probably due to the fact that bipolar II patients do not encounter a severe form of mania. In a yearlong separate prospective study consisting of 258 bipolar patients using the NIMH life chart method (Post et al, 2003), participants spent on average three times more of their time being depressed (33.2%) then being manic (10.8%). The proportion of depressed time did not vary from bipolar I patients to bipolar II, but 76% of the participants were with bipolar I disorder.

Most patients with bipolar disorder are free of symptoms between episodes, but some experience various residual symptoms. Despite treatment, a small number of people sadly encounter continuous symptoms (Hyman and Rudorfer, 2000). However, the natural course of bipolar disorder tends to get worse without treatment. A bipolar patient may undergo more frequent and more severe manic and depressive episodes over time than when the disease first emerged (Goodwin and Jamison, 1990). Effective treatment may help bipolar patients alleviate the regularity and severity of episodes and maintain good quality of life.

2.9 The causes of bipolar disorder

The causes of BD are currently unknown. There appears to be no single cause, but rather a combination of biochemical, genetic and environmental factors which act together to activate the illness. Brain-imaging techniques such as magnetic resonance imaging, positron emission tomography and functional magnetic resonance imaging suggest that the brains of individuals with BD may vary from the brains of normal individuals, although the significance of this is still unclear (Soares and Mann, 1997). The "chemical messenger system" (i.e. the neurotransmitter system) that affects mood may also play a part in causing bipolar disorder, as may hormonal imbalances.

There appears to be a strong genetic influence in bipolar disorder: it is more likely to occur in individuals whose biological relatives also have the condition. Some studies have also shown links between bipolar disorder and schizophrenia, indicating a combined genetic cause (NIMH Genetics Workgroup, 1998). The more

genes a person has in common with a BD sufferer, the more likely that person is to develop BD themselves. However, studies of identical twins suggest that other factors are involved in causing BD (NIMHGW, 1998). If the cause were solely genetic, then if one twin has BD then the other twin would always develop it also, but this is not the case. However, if one twin has BD then the identical twin is more likely to develop it (40 to 80% chance) than a non-identical twin (15 to 20% chance) or any other siblings (5 to 10%) (APA, 2010). Furthermore, a single gene is not the cause of bipolar disorder (Hyman, 1999). It requires several genes operating collectively in combination with other factors to produce the illness.

The environment may be one such factor. "Environmental causes may include problems with self-esteem, significant loss or high stress" (Mayo Clinic, 2010). Another factor is stress, which is often the main psychological cause leading to the onset of BD. Difficult life events such as the death of a loved one or losing one's job are a significant reason of stress. The illness can be activated in women due to childbirth or the menopause. Stressful circumstances may trigger an episode of mania or depression in individuals with BD, but once again, of course stress is not the single fundamental reason.

2.10 Bipolar disorder and suicide

A person with bipolar disorder may commit suicide either by careful planning or as an impulsive act. The lifetime risk for suicide is six times higher for individuals with depression than the general population (Hyman, 2000). Suicide is more common in bipolar depression than in unipolar major depression (Burgess, 2006). Most suicide attempts and most completed suicides occur during the depressive course of the illness (Baldessarini and Tondo, 2003). However, suicidality was reported to be markedly common among patients with mixed mania than patients with pure mania, and non-remission from mixed manic episodes may result multiple suicide attempts (Goldberg et al, 1998).

About 30% of bipolar sufferers attempt suicide at some point during their lives, and of these, about 20% succeed (Burgess, 2006). The standardised mortality ratio for suicide in bipolar disorder is approximately 22.4 for women and 15 for men (Osby et al, 2001). Furthermore, the risk of suicide is greater earlier in the course of the illness. Therefore, early detection of the illness and understanding how best to manage it may significantly reduce the risk of suicidal death. Suicidal feelings and actions can be overcome with proper treatment. Bipolar patients who are more severely affected show noticeably lower suicide rates because of underlying long-term treatment than those who are less severely affected but without proper treatment. Suicidal signs and symptoms include (CR, 2008):

- talking about wanting to die
- feeling hopeless and helpless
- talking about being a burden to others
- misusing alcohol and/or drugs
- writing a suicide note.

2.11 Alcohol, drugs and bipolar disorder

It is very common for bipolar people to abuse alcohol and drugs. Substance abuse has negative impacts on the illness, increases the recovery time, triggers early relapse and makes it difficult to distinguish mania (NIMH, 2010). According to the Mental Illness Fellowship of Australia (MIFA, 2008), BD patients are 11 times more likely to abuse alcohol or drugs than healthy people. Also according to MIFA (2008), substance misuse is the single most common predictor of poor outcomes. Cocaine and amphetamines are key factors in triggering manic episodes, followed by hallucinogens and marijuana. In the case of depression, it is alcohol which is the major precipitating factor. Substance abuse problems emerge due to many factors such as mood symptoms, self-medication of symptoms and risk factors; these may activate the occurrence of both bipolar disorder and substance use disorders (Strakowski and DelBello, 2000).

2.12 Life events and the course of bipolar disorder

The effects of life events are associated with the course of bipolar disorder. For example, a person with bipolar disorder could be at high risk for depression when a negative life event unfolds. To this end, studies such as Kessler et al (1997) and Agid et al (1999) reported an association between early parental death and occurrence of bipolar disorder in later life. Mortensen et al (2003) showed an increased risk for bipolar disorder among children who experienced early suicidal parental death. Ellicot et al (1990) reported a fourfold increased risk of relapse for bipolar patients due to acute negative life events. Furthermore, negative life events increase a time for recovery by threefold (Johnson and Miller, 1997).

In contrast, no independent severe negative life events were shown to predict an increase in manic symptoms in the longitudinal studies carried out by Alloy et al (1999), Johnson et al (2000) and Johnson et al (2004). However, as evidenced by Leibenluft et al (1996), sleep disruption may cause manic symptoms. About 10% of patients with bipolar depression were shown to develop manic symptoms due to sleep disruption (Colombo et al, 1999). People tend to remember major life events for up to a year (Paykel, 1997), while a large percentage of minor events are usually forgotten more quickly (Brown and Harris, 1982). A major life event that is involved in the course of bipolar disorder could be rated in terms of its severity, if it unfolds, by assigning a fixed number of points to assess the patient's condition (Johnson, 2005).

2.13 Incidence of bipolar disorder in UK cities

Lloyd et al (2005) conducted a study to assess the incidence rates of bipolar disorder in three UK cities, including south-east London, Nottingham and Bristol. The overall incidence rates per 100,000 per year in these three cities were found to be 6.2, 3.0 and 1.7 respectively. The study observed no significant gender difference in the incidence of bipolar disorder. The incidence of bipolar disorder was reported significantly lower among White groups than among Black and minority ethnic groups in all three cities. The reported increased incidence rates in Black and minority ethnic groups are in line

with previous similar findings from the UK (Leff et al, 1976; Bebbington et al, 1981; Der and Bebbington, 1987; Van Os et al, 1996).

2.14 Treatment of bipolar disorder

Regardless of the severity of disease, proper treatment can help stabilise mood swings and alleviate the associated symptoms of BD (Sachs et al, 2000). A combination of medication and psychosocial treatment is thought to manage the illness best. Bipolar disorder can affect many areas of life; thus a patient's care plan may involve a psychiatrist, psychologist, family and friends in treating the condition effectively. Usually, a psychiatrist supervises the medication aspects, a psychologist monitors the psychosocial aspects and the others provide encouragement and support in seeking treatment or continuing with the prescribed treatment plan.

It can take some time, and some changes of medication, before a suitable treatment plan is settled. Bipolar patients are usually advised to record their daily activity information even during the normal periods so as to ensure treatment efficacy and delay a future manic or depressive relapse. Bipolar patients those who communicate closely and openly, with their caregivers about their care plans, are likely to be treated most effectively. Medications branded as mood stabilisers are normally prescribed for treatment and prevention purposes. There exist different types of mood stabilisers—a good mood stabiliser can help reduce both manic and depressive episodes.

Lithium, the first mood stabiliser, has been widely used to help manage bipolar disorder. However, Silverstone et al (1998) state that the benefits of lithium are often less in practice than the clinical trial results would suggest, mainly because of poor compliance. Anticonvulsant medications such as carbamazepine or valproate have also been shown to provide mood-stabilising effects. Anticonvulsant medications may also be combined with lithium for maximum effect. However, valproate is not recommended for women of child-bearing age (BPS, 2010). All these drugs are powerful and have unpleasant side-effects. Moreover, Gitlin et al (1995)

state that even aggressive drug treatment can be ineffective in a significant number of cases.

Other medications such as antidepressants are used during depressive episodes in bipolar disorder, but use of antidepressants is controversial since this may trigger manic episodes at times. It often becomes difficult for bipolar patients to cope with medications because of unpleasant side effects. Some patients may also be tempted to discontinue medication during the periods when they feel better. Furthermore, the illness itself may lead a patient to refuse treatment, and the patient may be hospitalised as a result.

In addition to medication, psychosocial interventions such as cognitive behavioural therapy, psycho-education and family therapy are useful in offering help, instruction and direction to bipolar patients and their families. Through cognitive behavioural therapy, bipolar patients learn how to control negative or unjustifiable thought patterns and behaviours. Through psycho-education, bipolar patients learn how to detect early warning signs and thus seek early interventions. Family therapy is a form of psychotherapy that involves the family of a bipolar patient in therapeutic sessions so that the level of distress caused by the patient within the family gets reduced.

2.15 Impact of BD on family caregivers

Family members play a significant role in the care plans of bipolar affected people. Bipolar patients find it easier to manage the illness with support and encouragement from their family. However, family members or other caregivers may often have to deal with the patient's difficult behaviour, such as aggressiveness and wild spending habits during mania or heavy drinking, and excessive withdrawal from others during depression. Both male and female bipolar patients can become hostile and aggressive, especially during mania, and may even commit criminal acts. This is a particular worry for many family caregivers.

Like all serious illnesses, bipolar disorder affects the whole family as well as the patient. A study conducted by Dore and Romans (2001) highlights the notable problems faced by family caregivers in their relationships with patients during periods of illness, with considerable impact on their own employment and financial status, childcare duties and ordinary social relationships. In most cases, family caregivers show considerable tolerance towards the difficulties posed by bipolar patients. Sadly, though, there are cases when family members are unable to cope and withdraw from their caregiving roles.

3. The PAM Project

This chapter starts by explaining the motivation for the PAM (Personalised Ambient Monitoring) project. It then describes what the PAM system involves and how it operates. It also highlights the distinctive contribution of OR modelling to the project.

3.1 Introduction

We have seen in Chapter 2 that bipolar disorder is a serious mental illness which cannot be controlled by drugs alone. Bipolar patients who are trained to use self-help treatments can benefit from greater control over their care and life decisions and can detect early warning signs of serious illness (Morris et al, 2004). To date, self-help interventions have been either manual or computer-based, and are ineffective at identifying the onset of depression (Perry et al, 1999). Although computer-based data entry may be more accurate than manual data entry, both systems still rely on human input and are thus prone to errors. Moreover, manual or computer-based systems do not routinely report physiological and patient context information. Automated data collection through the use of modern technology may overcome these drawbacks.

3.2 Technological interventions in bipolar disorder

The need for innovative approaches that improve the efficiency of homecare is growing, as healthcare services move out of the hospital into the home setting. Examples of such *telecare* projects include the Bath Institute of Medical Engineering (BIME, 2010) project for dementia sufferers and the Mobilising Advanced Technologies for Care at Home (MATCH, 2009) project for ageing people. Other initiatives include Extending QUAlity of Life of older and disabled people (EQUAL, 2004) and Strategic Promotion of Ageing Research Capacity (SPARC, 2007), Smart and Aware Pervasive Healthcare Environment (SAPHE, 2006) and Ubiquitous Computing for Healthcare in the Community (UbiCare, 2003).

In July 1992, the British Government published its future vision and legislation in the policy manuscript *The Health of the Nation* (DoH, 1992), which was subsequently endorsed by the policy manuscript *Our Healthier Nation* (DoH, 1999). Both manuscripts identified the problem of mental illness as one of the key areas for action and targeted mental illness together with coronary heart disease, cancer and strokes for health improvement. Since then further frameworks for action have been developed, most notably the *Health of the Nation Mental Illness Key Area Handbook* (DoH, 1994) and the *National Service Framework for Mental Health* (DoH, 1999).

Research studies such as Malan et al (2004), Van Laerhoven et al (2004), Chakravorty (2006) and Wood et al (2006) propose sensory networks for therapeutic monitoring. Sensory networks can also be used to observe the physiological and environmental information of patients to provide robust monitoring (Blum and Magill, 2008). Sensory data could be combined with self-reported data for the detection of prodromes with the aim of describing the mental health state, thus identifying and minimising both manic and depressive episodes.

However, recording the behaviour patterns of mental health patients using sensory networks to issue automatic alerts in a flexible and accurate manner is a significant research challenge. According to the National Institute of Mental Heath (NIMH) US, bipolar patients are strongly advised to continue self-monitoring even during the normal periods to delay a future relapse and thus to ensure ongoing treatment effectiveness. Besides, bipolar patients may suffer cognitive difficulties not only during a depressive or manic episode, but also during a normal period (Martinez-Arán, 2004). Therefore, without the use of technology, ongoing daily self-monitoring can be unmanageable for many patients.

3.3 The PAM team

PAM (www.pam-research.org) is a collaborative research venture between three UK universities: Southampton, Nottingham and Stirling. The project was funded by the UK Engineering and Physical Sciences Research Council and began at the beginning

of October 2007. The funding for PAM was awarded following a "sandpit" on telecare in the autumn of 2006, entitled "Taking Care to the Patient", where the four academic investigators met for the first time. The PI of PAM was Professor Christopher James, who at the time was based in the Institute of Sound and Vibration Research (ISVR) at Southampton, but is now at the University of Warwick.

The PAM team involves academic researchers from the fields of electronic engineering (Professor John Crowe, Nottingham), biomedical signal processing (Professor Christopher James, ISVR), computer science (Professor Evan Magill, Stirling) and OR (Professor Sally Brailsford, School of Management, Southampton). Each had a PhD student funded on the PAM project. This mix was essential to the overall success of the project, and provided an interesting and distinctive interdisciplinary grouping. Figure 3.1 shows the PAM research team and their research fields. Each investigator had one PhD student on the project. The project's progress was driven by a Steering Group consisting of clinicians, patients and non-academic researchers. Appendix A1 details the people (academics, PhD students and Steering Group) involved in the PAM project. Members of the PAM team explored the potential of technology to enrich healthcare direction, motivation, independence, confidentiality, effectiveness and well-being within a specific area of mental health.

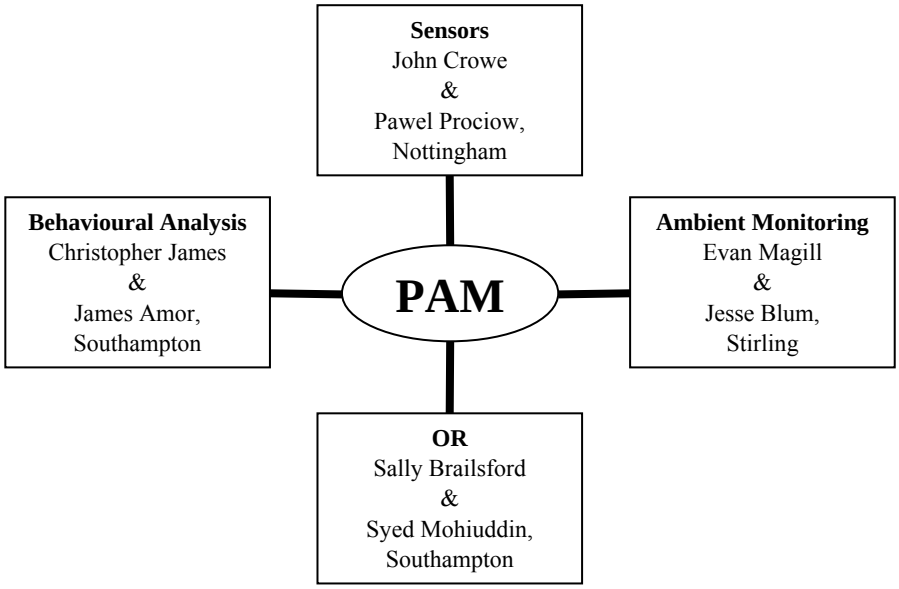


Figure 3.1 The PAM academic team

The first objective of the PAM project was to test whether distinct behavioural patterns (i.e. 'activity signatures') could be identified in healthy individuals. If so, then the second aim was to determine whether these activity signatures could be used to support self-management in bipolar disorder, by providing early warning signs of a change in behaviour which could potentially signal a decline in the patient's health.

3.4 The PAM system

In the PAM project, we were specifically concerned with the management of bipolar disorder. The objectives of managing bipolar disorder usually involve resolving manic and depressive symptoms, preventing ongoing relapse, and achieving well-being (see Figure 3.2):

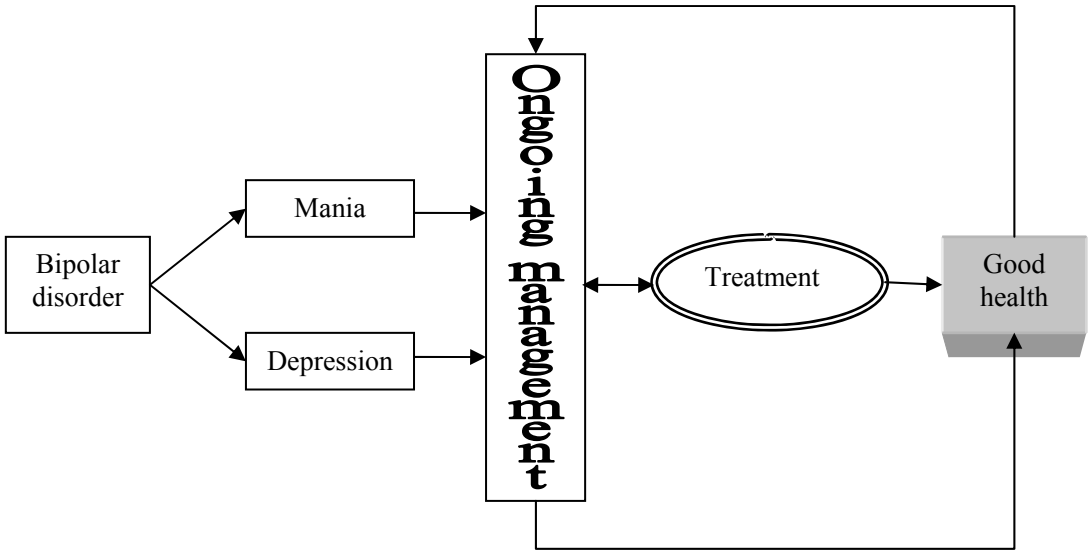


Figure 3.2 Bipolar relapse prevention

The technological aspects of the PAM project involved the development of a system for automated activity monitoring in BD, using a personalised unobtrusive set of small environmental and wearable devices to monitor patients' activity patterns. These devices were built by the Nottingham group. Figure 3.3 below shows how the PAM system collects ambient data through wearable and static environmental sensors from ambulatory and home settings. Examples of these include accelerometers to obtain activity related information and light sensors to detect light levels.

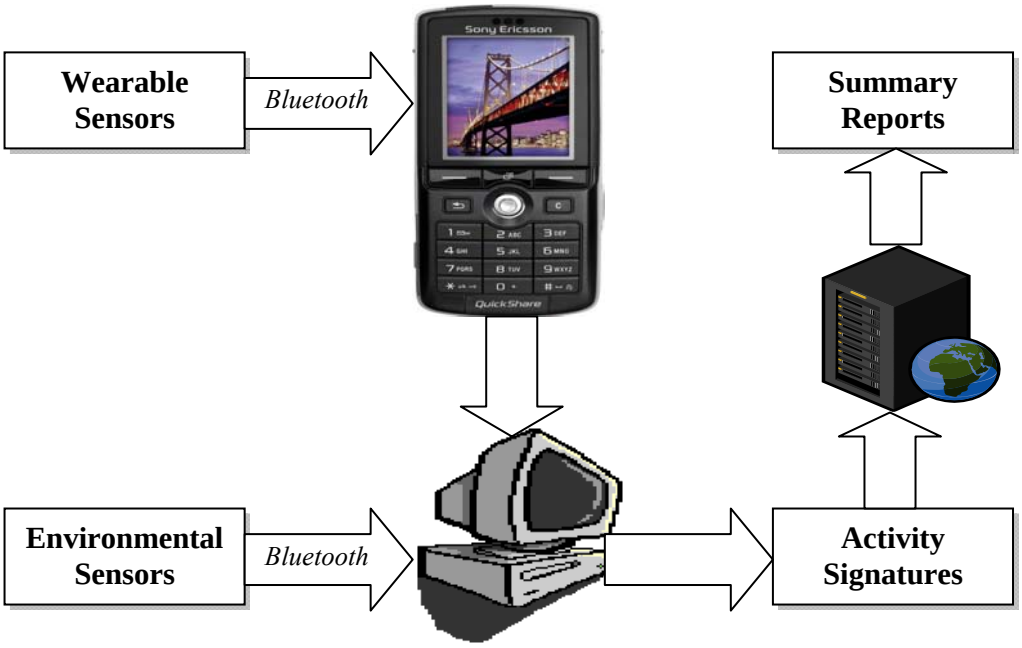


Figure 3.3 Ambient data collection. Source (modified): Prociow and Crowe (2009)

The system includes both wired and wireless communication links between the patient and the monitoring network. Bluetooth transmits the wearable sensor readings to a mobile phone and the environmental sensors readings to a PC. The data is summarised locally using feature extraction algorithms devised by the ISVR group, and then transmitted onwards via the internet to a central server, where algorithms (devised by the Stirling group) are used to detect whether the person's behaviour pattern was within some tolerance of "normal" or whether a significant change may have occurred.

A major concern was the acceptability of the PAM system to patients and their families. Fahrenberg (2006) states that user acceptance issues are at the heart of a successful monitoring system. However, we were reasonably optimistic that BD patients were likely to accept such monitoring, provided they understood its potential benefit. In a study of the ChronoRecord, a computer-based system used to enter daily recordings, Whybrow et al (2003) reported that 83% of the BD patients in their study had no problem accepting the technology. Today, technology is becoming a part of everyday life. Moreover, healthcare-related technologies and modelling have been

recognised as valuable to the UK's future by the Council for Science and Technology (CST, 2010).

The sensor set had to be customisable to the individual needs and preferences of each patient. The P in PAM stands for Personalised and is a key aspect of the system. The PAM system allows patients and their caregivers to select whatever monitoring devices they find acceptable. Patients should be unaware of the sensor devices and they do not need to carry out any special actions. They can simply get on with what they normally do.

3.4.1 Types of sensor

Based on the need to identify changes in key behaviours known to be affected during the course of bipolar disorder, and following discussions with the Steering Group, the full range of potential choices of sensors was chosen. The sensor configuration comprises two main sets: wearable and environmental. The first set includes small devices to be worn or carried by the patients, for example on a mobile phone, while the second includes ambient sensing devices to be set up in their home environment.

Table 3.1 below provides the associations made between specific sensors and bipolar episode prodromes. The sensors are not designed to be completely hidden from sight, but they are also not obtrusive. Tapia et al (2004) suggest that people usually forget about the installation of devices in the home environment within a few days. Rather than adopt a conventional approach of using different sensors each tailored to a particular monitoring task, the PAM system includes a wireless camera within the PAM system to capture images of various areas of interest within a complex environment (e.g. a kitchen) for monitoring different activities within the scene. These images are not identifiable as people or objects but as "blobs" – areas of black and white – simply indicating the presence or absence of movement. Amor and James (2008) provided further details on this matter.

Prodromes	Sensors
Activity level	Accelerometer GPS
Sleep	Light/Noise levels Bed occupancy
Talkativeness	Microphone Phone sensor
Social energy	GPS Bluetooth encounters
Appetite	Fish-eye camera Cupboard/Fridge/Microwave/Cooker doors sensors

Table 3.1 Sensors for specific bipolar episode prodromes

3.4.2 The PAM data processing platform

The PAM system uses a mobile phone to upload its data on a PC via Bluetooth for further processing. All of these aspects were developed and managed by the Stirling team. The ambient sensors also interface with the PC, which is a centre for long-term storage and data fusion. The PAM system then uploads this processed data from the PC to an external server to observe captured behavioural patterns. "Roaming gateway" software transfers operating procedures to the devices and synchronises sensor and network data collection. Middleware fitted on the devices supports multiple tasks such as device registration, task assignment and secure data transfer.

3.4.3 PAM data processing architecture

As shown in Figure 3.4 below, the PAM system uses a three level architecture: the node level, the semantic level and the fusion level, in order to process large volumes of data generated from the sensors. The node level performs feature extraction and converts the raw data into a meaningful form before transmitting to further up the processing chain for more advance manipulation. The semantic level then extracts essential connotation from the features transmitted to it from the node level for data fusion. These two levels were both designed by the ISVR team. The fusion level (developed by Stirling) finally fabricates the semantic features passed on to it in order to impart prognostic competences of the system via particular pattern matching

algorithms to recognise recurring patterns. The data is then fed into the model to monitor variations in a patient's behaviour and produce appropriate alerts.

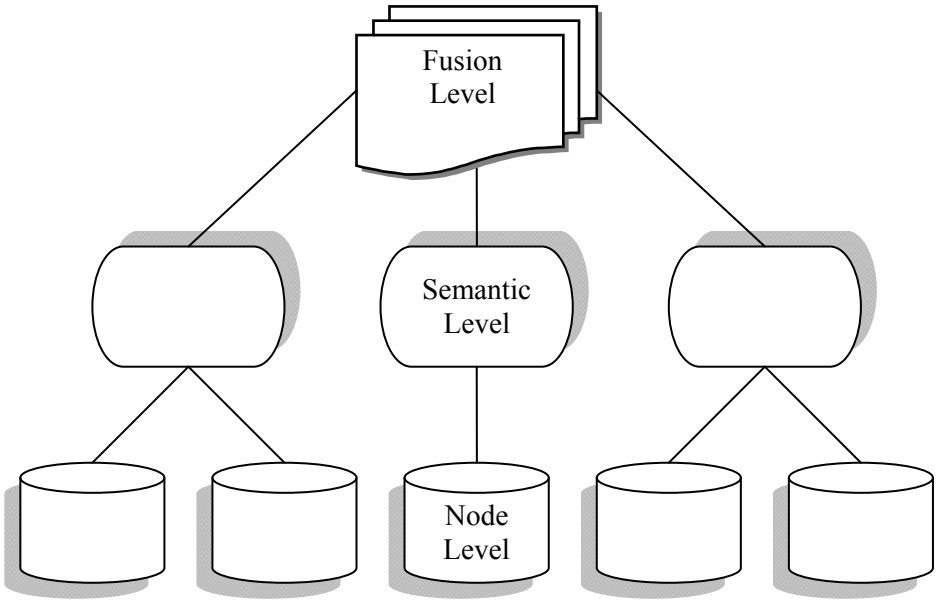


Figure 3.4 Three-level data processing architecture. Source: Amor and James (2008)

3.5 PAM and OR modelling

The PAM team has built an automated unobtrusive system of daily monitoring of bipolar patients'. This was no doubt a challenge, which has been dealt through a multidisciplinary approach. We, the OR researchers, played a key distinguishing role to provide the modelling basis of the work. PAM does not require patients to enter data manually in any format, to reduce the bias of subjective recall from patients. The advantages of ambulatory monitoring, assessment and frequent measures are well documented. With unique features such as the OR modelling basis, and the emphasis on the roaming gateway and dynamic programmability, the PAM system is meant to be seen as an integrated function that provides a unified and coherent care management at home for a better insight into bipolar patients' true behaviours. The system essentially offers personalised monitoring to allow a patient choice of sensors and analysis of behavioural patterns.

The other members of the PAM team provided skills and experience across sensors, medical signal processing, communications and software services, while we dealt with OR simulation modelling. The need of this mix was essential to the success of the project. Within PAM, our role as the operational researchers was to formulate our own model in which to perform the various analyses. Building models for bipolar disorder is very difficult, but we have attempted to build a multi-state transition model that is combined with computer simulation. We used our model to check if daily monitoring can be adjusted and personalised, and may offer as a direct motivator for behavioural change in patients. The model was revised by relevant literature reviews to verify the intervention usefulness.

We assisted the other members of the PAM team to design the PAM system and to direct the analytical conducts within the framework of the system. This role provided us with a chance to augment our involvement to the solution of the problem. The OR simulation model provided insight into the required modalities and sensors. We employed a prognostic modelling analytical concept that used a mathematical equation, computer logic and related tools to estimate the outcomes of specific decision options within the computations of the model rather than actually implementing them, e.g. to produce the preferred outcomes. We opted to use Monte Carlo simulation to provide a more affluent structure to reach at realistic and acceptable decisions. Our model was represented in Microsoft Excel in which we methodically searched for the multitude of decision alternatives to perceive the preferable steps forward.

OR modelling, managed parametrically, allowed us to examine various decision choices within a short time and evaluate their effects to develop an ideal option for our program. In this way, cost was limited and outcomes were faster in a risk-free setting without directly affecting patients before implementation. However, it required us to discover a better mathematical tool to tackle the various issues. We tried to balance all the relevant considerations in developing and disseminating the model that is capable of informing the PAM system. Our model provided good tractability or simplicity for analysis, and yielded useful conclusions within the time

available for decision making. We often communicated with the members of the Pam team, to get maximum benefits of modelling so as to make shared decision making. Figure 3.5 shows the application of the simulation model to the PAM system:

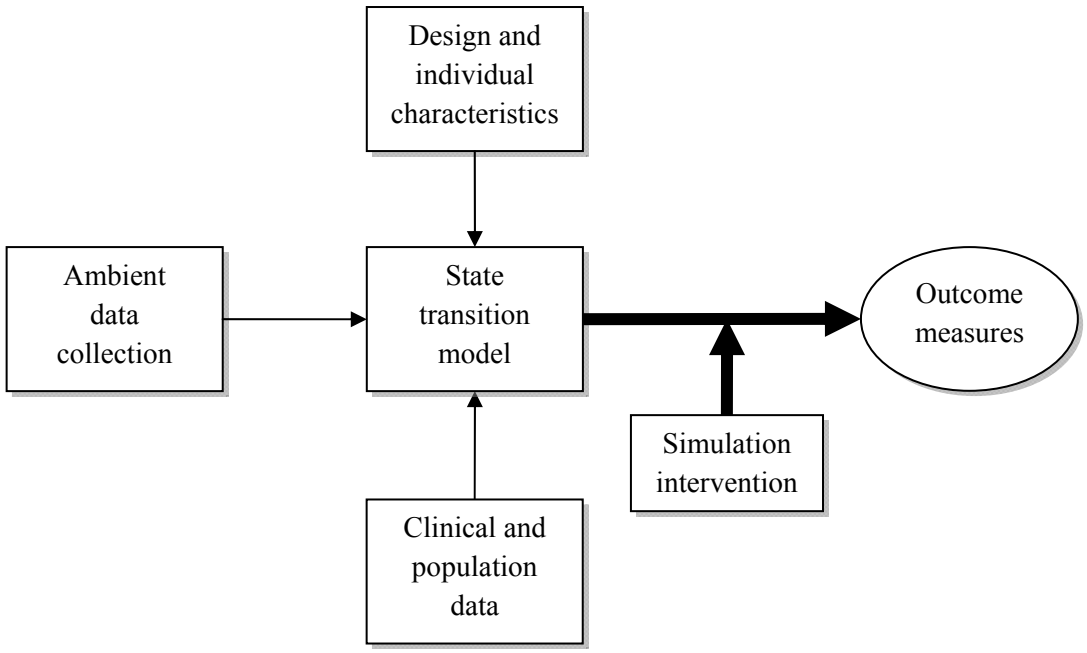


Figure 3.5 The role of simulation to evaluate outcome measures

3.6 Potential benefits through PAM

The PAM team define the potential patient benefits as:

- improved self-management, patient management pathways and patient care
- improved early detection of bipolar prodromes
- better patient confidentiality
- enhanced quality of care and quality of life
- increased productivity of caregivers
- improved applications of drugs and reduced need of hospital admissions
- sustainable support for health services research
- support for potential developments in healthcare technology and management
- improved health, independence and wellbeing status for bipolar patient group
- decreased costs to society in general.

Further to the above benefits, the clinician can quickly be provided with extra useful information through the PAM system. Technological monitoring can surely provide an early warning of any imminent episode; thereby stop patients escalating into a severe state and requiring unnecessary hospital admissions, and help steady the lifestyle of patients. The potential PAM technological benefits are:

- improved evaluation of ambient monitoring in a mental healthcare context
- enhanced the ability to simply plug in and turn on a sensor
- better identification of behavioural signatures from disparate and sparse data
- enhanced intermittent connectivity to sensors
- enhanced the concept of the remote monitoring of bipolar patients
- enhanced the use of OR modelling to guide a technology
- enhanced the use of low-cost sensors at home.

3.7 Acceptability of PAM

One may question about the acceptability of PAM from the perspective of both patient and caregivers. The members of the PAM team have shown through the trials that the system works technically and generates detectable repeated daily activity patterns from different sensors, but will bipolar sufferers actually accept it? Technologies are a part of everyday living at present. Whybrow et al (2003) reported that 4 out of 5 bipolar patients accepted the ChronoRecord, which is a computer-based system to enter daily recordings. Moreover, the concept of the remote monitoring of patients has already been tested, e.g. the Bath Institute of Medical Engineering (BIME) and UbiCare (Ubiquitous Computing for Healthcare in the Community), mentioned earlier in Section 3.2. In any case, the PAM team involved the user community from the start to reflect the patients' ambitions and concerns in the design of the PAM system. The experimental PAM design crucially provides bipolar patients the opportunity to withdraw at any time without compromising the final outcomes.

3.8 Testing the PAM system on real patients

Originally we had planned to carry out a small trial on four patients, but delays in obtaining ethical permission and recruitment difficulties meant that in the end, we were only able to recruit one patient, who participated in the study for about three months. Healthcare ethics is about protecting the human participants in a research study from any kind of physical or psychological harm. Ethics was clearly a major issue for PAM. We obtained NHS ethical approval (see Appendix A7) to run the PAM study.

The participant was given some brief training about the PAM system. She was clearly informed about all aspects of the monitoring system, including how data would be collected, stored and processed. She had to give consent to every aspect of the monitoring system in order to participate in the study, and had the right to withdraw from any part of the monitoring at any point without the need to give a reason. During the study, she was encouraged to report any part of the system that was troublesome. She was also asked to record her own activity in a diary during the study with the purpose of matching the self-recorded data with the automatically logged data obtained from the PAM system, although unfortunately this did not happen.

At the beginning and end of the trial, a member of the PAM team conducted a semi-structured interview with the participant. The main purpose of the interview was to document the participant's opinions regarding the acceptability of individual sensors, the obtrusiveness of sensors and acceptability of the PAM set-up as a whole, any user compliance issues for devices and suggested changes to improve compliance, any issues about data access and privacy, and reasons for not wearing or disabling devices. In addition, she was asked how she felt about the effects of self-monitoring, and was asked to describe her thoughts and feelings about external monitoring. At the end, she was asked about her thoughts and feelings about how life changed via the use of the technology, and for any improvements that could be made to improve compliance.

4. Diagnosis of Bipolar Disorder

The first stage in developing a mathematical model of any disease is to understand its clinical aspects, and to study the medical literature for any pre-existing (and preferably commonly used) staging and diagnostic criteria. This is particularly challenging in a psychiatric condition such as bipolar disorder, because of the qualitative and behavioural nature of the symptoms of this disorder. A few diagnostic tests and scoring systems are available, but most diagnoses are made by a psychiatrist essentially on the basis of a conversation with the patient and a discussion of the person's feelings and behaviour. This chapter presents a review of the literature on this topic.

4.1 Introduction

Chapter 2 discussed the symptoms of BD, which vary from patient to patient. Monitoring symptoms and the frequency and duration of episodes are key for an accurate diagnosis. Without accurate diagnosis, it is difficult to initiate effective intervention. To make matters worse, many bipolar patients avoid treatment, and this usually leads to severe long lasting consequences. Delay in treatment can prevent 30 to 60% of bipolar patients from regaining full functional capabilities (MacQueen and Young, 2001; Calabrese et al, 2003).

4.2 Diagnosis delay and the benefits of screening

The bipolar spectrum, a key area of recent research, includes bipolar I, bipolar II, cyclothymia and bipolar Not Otherwise Specified (NOS), all of which were discussed in Chapter 2. Regrettably, bipolar spectrum disorders frequently go unidentified with common hold-ups of 8 years or more before accurate diagnosis (Lish et al, 1994). This ensues a considerable delay in diagnosis and proper treatment (Hirschfeld et al, 2000), the consequences of which can be severe and may even lead to suicide

(Dunner, 2003). The cost to the NHS of undiagnosed cases of BD may be even higher than the costs of identified cases (Li et al, 2002; McCombs et al, 2007), estimated at £200m in the UK by Gupta and Guest (2002).

However, the early identification of any illness can be improved through screening. There are several screening tools for psychiatric disorders, but only a few specifically for bipolar spectrum disorders. The Diagnostic and Statistical Manual, Fourth Edition, Text Revision (DSM-IV-TR, 2000), by the American Psychiatric Association, is one of the key sources for psychiatric diagnosis. The DSM-IV-TR diagnostic criteria for bipolar disorder are based on the detection of various mood episodes, which must have a distinct onset and end point.

4.3 Structured and semi-structured interviews

The Structured Clinical Interview for DSM-IV (SCID) is the most extensively used clinical evaluation device for the diagnosis of bipolar spectrum disorders (First et al, 1997). A patient's underlying episode type can be determined via a SCID interview (Hilty et al, 1999). It usually takes between one to two hours to accomplish a SCID interview, depending on the patient and who (a clinician or a trained mental health professional) conducts the interview. The SCID interview is split into six separate components covering mood episodes, psychotic symptoms, psychotic disorders, mood disorders, substance use disorders, anxiety, and adjustment and other disorders. These can be conducted one after another to describe a patient's present psychiatric state. The SCID interview has been validated in many diverse populations and its reliability has frequently been shown to be high.

However, the SCID interview may be more reliable in detecting the severe form of BD rather than a milder form (i.e. hypomania). Baldassano (2005) states that the diagnosis of hypomania may require a more delicate approach than the SCID. A semi-structured interview may be more fitting for the diagnosis of bipolar spectrum disorders (Benazzi, 2003), because it permits the clinician to bring up new questions

during the interview and to explore a framework of themes combined with clinical experience and judgment.

4.4 Underdiagnosis or misdiagnosis

Diagnosis and treatment of bipolar disorder are clearly challenging. Burgess (2006) reported that nearly 70% of bipolar patients were found to be misdiagnosed over three times prior to the correct diagnosis. Bipolar II and bipolar NOS types pose a particular challenge, and both types are often misdiagnosed as unipolar major depressive disorder and thus receive improper treatment (Bowden, 2001). One reason for the misdiagnosis of bipolar patients is that they usually spend substantially more time in the depressive phase of the illness than in the manic or hypomanic phase (Judd et al, 2002; Judd et al, 2003).

On the other hand, underdiagnoses can be caused by the patients' lack of insight into mania and failure to involve a caregiver in the diagnostic process, together with the clinicians' failure to agree on the boundaries of the bipolar spectrum (Ghaemi et al, 2002). All bipolar patients go through periods of depression (Goodwin and Jamison, 1990), and especially this primary appearance of bipolar depression may subsequently trigger misdiagnosis as unipolar depression. Depressive patients, who lack understanding of hypomania as pathological, fail to express it to the clinicians spontaneously (Scott, 2002). As such, the clinicians often fail to put forward appropriate direct questions (Angst and Gamma, 2002). One statistic shows that the prevalence of bipolar disorder in patients being treated for unipolar depression in a private psychiatric clinic in northern Italy was 49% (Benazzi, 1997).

Other diagnostic problems include presence of more than one health condition, poor understanding of the classification systems and overlapping symptomatology with a range of type I and type II disorders (Vieta et al, 2000). Anxiety and substance abuse disorders are very common among patients with bipolar disorder. McElroy et al (2001) found anxiety disorders in 40% of bipolar patients and substances abuse in 60% of bipolar patients.

To screen and diagnose bipolar disease in both psychiatric and general practices, clinicians commonly use the operational diagnostic criteria such as DSM-IV and the International Classification of Diseases (ICD-10, 1993). There is some variability between these two major diagnostic classification systems. The DSM-IV criteria for a diagnosis of BD require a single episode of mania or hypomania together with a single depressive episode. In contrast, the ICD-10 criteria require two separate mood episodes including at least one episode of mania for a bipolar diagnosis. Both DSM-IV and ICD-10 may be less consistent and logical than previously thought in the detection of hypomania, which may entail more delicate inquiry than described in SCID and the Composite International Diagnostic Interview (CIDI) (Angst et al, 2005).

It has been reported that about 40% bipolar sufferers are primarily misdiagnosed (Ghaemi et al, 2002), which hinders prompt active intervention. Thus, the true prevalence of bipolar disorder may well be greater than is currently believed (Ghaemi et al, 1999). However, the diagnosis and treatment of bipolar spectrum disorders as well as the subset of depressed patients actually with bipolar disorder may be improved through self-report screening measures (Hirschfeld et al, 2000; Hirschfeld et al, 2005).

4.5 Self-report screening measures

The chance of detecting treatable illness where an intervention may enhance outcome of the illness can be increased through screening. This should be simple to run and acceptable to patients, and able to detect most of the true cases (Wilson and Junger, 1968). The lack of classification systems for the diagnosis of bipolar spectrum and the subsequent rise in prevalence present a clear need for a broader screening approach that may enable a large number of bipolar patients to benefit from effective intervention. It has been reported that bipolar patients' ratings of their own symptoms may differ from their clinicians' opinions (Laje et al, 2002). This may be because the self-report survey is designed for the whole BD spectrum. Another reason for this is

that as mentioned previously, patients have impaired insight into mania (Ghaemi et al, 2002).

Self-report measures are widely used to detect depression, but they have not generally been used for identifying bipolar disorder (Passik et al, 2001). Today, self-report measures have dependably been used to assess manic symptoms (Truman et al, 2002). Furthermore, Benazzi (2002) used evidence derived from a self-report measure as a diagnostic tool for bipolar spectrum disorders. Meyer and Hautzinger (2003) and Angst et al (2005) respectively conducted validation studies of two screening tools, the Hypomanic Personality Scale and the HCL-32, both of which have been developed specifically for identifying hypomania. Hirschfeld et al (2000) developed and later validated the Mood Disorder Questionnaire, which is now one of the most widely used self-assessment screening tools for identifying bipolar spectrum disorders.

Dr Ronald Pies developed the Bipolar Spectrum Diagnostic Scale, which was later validated by Ghaemi et al (2005). There are other self-assessment methods such as the Self-Report Inventory for Mania (Shugar et al, 1992), the Symptom Checklist 90 for a broad range of psychological problems (Hunter et al, 2000) and the Brief Bipolar Disorder Scale (Dennehy et al, 2004), but these are predominantly to assess symptoms and have not been advocated as screening tools (Angst et al, 2005).

4.6 Diagnostic and screening tests in psychiatry

The use of symptom clusters, rather than disease cause or pathology, has been common in the diagnosis of psychiatric disorders for a long time (Warner, 2004). A disease screening test ideally provides information on chronic disease progression at a stage at which an intervention can be useful. However, a screening test is never without difficulty. Screening tests must be safe, cost-effective and acceptable since such tests are carried out within a healthy population. Screening tests are intended to detect the likelihood of having a disease and to indicate the need for further examination for a patient who screens positive. On the other hand, diagnostic tests

should offer more certainty whether an individual has a specific disease. Diagnostic and screening tests in psychiatry are common today. Examples include screening for bipolar disorder (Hirschfeld et al, 2000) and neuropsychological testing to diagnose dementia (De Jager et al, 2003). An overall understanding of the use of tests is vital whether to infer results of clinical trials or for the point of routine clinical practices.

However, the over interpretation of a diagnostic or screening test may lead to incorrect diagnosis; thus the correct evaluation and interpretation of sensitivity, specificity and predictive values are necessary. Sensitivity and specificity indicate how precise a test is, and this information is particularly important to clinicians than patients. However, patients are usually more concerned to know about the likelihood of having a disease (i.e. positive predictive value) or not having a disease (i.e. negative predictive value). These calculations are used in the simulation model to evaluate the effectiveness of the PAM system to detect changes in a person's mental health state.

4.6.1 The sensitivity of a test

The sensitivity of a test measures the proportion of individuals who are correctly identified as having a specific disease. The formula for sensitivity is:

$$Sensitivity = TP / (TP + FN),$$

where TP is the number of true positives (people with the disease who test positive), FN is the number of false negatives (people with the disease who test negative) and $(TP + FN)$ is the total number of patients who have the disease.

4.6.2 The specificity of a test

The specificity of a test measures the proportion of individuals who are accurately identified as not having a specific disease. The formula for specificity is:

$$Specificity = TN / (TN + FP),$$

where *TN* is the number of true negatives (healthy people who test negative), *FP* is the number of false positives (healthy people who test positive) and (*TN* + *FP*) is the total number of healthy people.

4.6.3 The Positive Predictive Value (PPV)

The *PPV* of a test measures the proportion of individuals with positive test results who are correctly diagnosed. The formula for *PPV* is:

$$PPV = TP / (TP + FP),$$

where *TP* is the number of true positives, *FP* is the number of false positives and (*TP* + *FP*) is the total number of patients who test positive.

4.6.4 The Negative Predictive Value (NPV)

The *NPV* of a test measures the proportion of individuals with negative test results who are correctly diagnosed. The formula for *NPV* is:

$$NPV = TN / (TN + FN),$$

where *TN* is the number of true negatives, *FN* is the number of false negatives and (*TN* + *FN*) is the number of patients who test negative.

4.7 The Mood Disorder Questionnaire (MDQ)

The MDQ is a short self-report form, designed to be easy to complete by patients and their caregivers. It is a screening tool to find out the best possible symptom threshold for recognising bipolar disorder. The MDQ form consists of three main questions (MDQ, 2000). Question 1 has 13 brief 'yes' or 'no' items, which are derived from both DSM-IV criteria and clinical experience. In Question 2, patients are asked a further 'yes' or 'no' question to check if any symptoms have occurred within the same time period. Finally, in Question 3, a 4-point scale is used to assess the level of functional impairment.

The MDQ was first validated against the SCID in a sample including 198 patients from five psychiatric outpatient clinics (Hirschfeld et al, 2000). The sensitivity was 0.73 and the specificity was 0.9, confirming the MDQ as a good functional screening tool for bipolar disorder in the community. In a follow up study, Hirschfeld et al (2003) again validated the MDQ against an abbreviated version of SCID within a selected sample consisting of 695 participants from a nationwide epidemiological general US population. The MDQ was shown to have a sensitivity of 0.28 and a specificity of 0.97. According to the authors, the low sensitivity observed in the general population may have been due to the low sensitivity of the abbreviated version of SCID in this population. However, the high specificity of MDQ in this study screened out nearly all the true negatives.

The MDQ has since been validated in various US settings by Miller C et al (2004), Hirschfeld et al (2005), Graves et al (2007) and Kemp et al (2008), and in other countries by Hardoy et al (2005), Weber et al (2005), Konuk et al (2007), Vieta et al (2007) and Twiss et al (2008). Unsurprisingly, the psychometric properties of the MDQ were shown to differ slightly according to the sample population studied. More recently, the MDQ was verified for a UK population. In this study, Twiss et al (2008) measured the sensitivity and specificity of the MDQ for a sample of 127 patients, of whom 54 had BD and 73 had unipolar depression. The overall sensitivity of the MDQ was 0.76 and the specificity was 0.86.

Table 4.1 summarises the performance of the MDQ screening tool in four different studies, all of which are based on using the initial scoring criteria described above:

Study	Population	Sensitivity	Specificity
Twiss et al (2008)	Psychiatric population	0.76	0.86
Miller C et al (2004)	Psychiatric population	0.58	0.67
Hirschfeld et al (2003)	General population	0.28	0.97
Hirschfeld et al (2000)	Psychiatric population	0.73	0.90

Table 4.1 Sensitivity and specificity of the MDQ in four different studies

4.8 The Bipolar Spectrum Diagnostic Scale (BSDS)

The BSDS is another short self-report questionnaire used to screen for bipolar spectrum disorders. This story-based scale is scored using an algorithm requiring endorsement of 19 mood items by one of the following four statements: *i*) "fits me very well", *ii*) "fits me fairly well", *iii*) "fits me to some degree" and *iv*) "does not really describe me at all" (BSDS, 2002a; 2002b). A score between 0 and 25 is then assigned, depending on the strength of the response.

Ghaemi et al (2005) first investigated the BSDS screening tool for entire bipolar spectrum in two psychiatric clinics. A total of 95 patients took part in this study, 68 with BD (44 type I and 24 type II or NOS) and 27 with unipolar major depressive disorder. The performance of the BSDS was validated in its original version against the SCID interviews conducted by psychiatrists. The scale showed a high overall sensitivity of 0.76 (with 0.75 for bipolar type I and 0.79 for bipolar type II) and specificity of 0.85. Miller C et al (2004) had shown that the MDQ was not sensitive enough to detect milder form of BD, but Ghaemi et al (2005) showed that the BSDS is equally sensitive in detecting both severe (type I) and milder (type II) forms of BD. However, the MDQ was more recently shown to perform highly for entire bipolar spectrum disorders by Twiss et al (2008) in distinction; thus making it also helpful in identifying milder form of the illness. Ghaemi et al (2005) suggested using self-report questionnaires such as the BSDS as a complement to the MDQ and a supplement to clinical interviews in screening for bipolar spectrum disorders.

4.9 The CIDI-based bipolar disorder screening scale

The CIDI (Composite International Diagnostic Interview) is another screening tool for detecting bipolar spectrum disorders. This scale starts with asking individuals two "stem questions": a "euphoria" stem question and an "irritability" stem question (CIDI, 2001). If the respondent answers yes to one of these stem questions, s/he then proceeds to the "criterion B screening" question. If the answer is again yes, s/he moves to the 8 or 9 further "criterion B symptom" questions.

Kessler et al (2006) validated this scale among the US household population. The scale was shown to have the ability to identify a high percentage of true cases while minimising the number of false positives. The sensitivity was between 67% and 96%, which is comparable with the MDQ (Twiss et al, 2008) and the BSDS (Ghaemi et al, 2005). However, the two stem questions are not phrased in a simple manner; thus some bipolar patients may respond to these stem questions inappropriately. Hence, a genuine case may be missed simply because of the participant's failure to answer correctly one of the two stem questions to start with, and this is a serious limitation. Furthermore, patients with poor insight may not be able to complete this screening questionnaire without assistance.

4.10 The Hypomanic Personality Scale (HPS)

The HPS is an instrument that could possibly be used to screen for bipolar affective disorders. Eckblad and Chapman (1986) initially developed this scale as a 48-item self-report questionnaire to assess stable hypomanic characteristics. It was shown that participants scoring highly on this scale undergo episodes of mood disorders. This scale was validated in a large sample of adolescents (Klein et al, 1996; Meyer et al, 1998). In a 3-year follow-up study, Meyer et al (1998) found that this scale predicted depressive symptoms but not anxiety. Klein et al (1996) found similar results. Over a 13-year follow-up study, Kwapil et al (2000) reported the scale to be prognostic for bipolar spectrum disorders with a prevalence of 25% for the risk group compared with none of the patients who were part of the control group. Recently, Meyer and Hautzinger (2003) conducted a study to verify this scale in a sample of 212 German university students. Meyer and Hautzinger (2003) replicated the findings of Eckblad and Chapman (1986) that participants with high scores experienced more manic or hypomanic episodes than participants with low scores.

One of the downsides of HPS to its use in general practice settings is the time required to complete a 48-item self-report questionnaire. This scale was developed to target personality traits. Also, the episodic nature of hypomania was not investigated.

The HPS may be more appropriate in the psychiatric setting rather than in general practice.

4.11 The Hypomania Checklist (HCL-32)

The HCL-32 is a self-report questionnaire that includes 32 'yes' or 'no' questions. This tool has mainly been developed to detect bipolar disorder patients who are being misdiagnosed as having major depressive disorder. Recently, Angst et al (2005) tested the performance of the scale in a sample of 426 adult psychiatric patients (266 with bipolar disorder and 160 with major depressive disorder). The factor structure of HCL-32 ("active/elated" hypomania and "risk-taking/irritable" hypomania) was fairly similar in both samples. A score of 14 or more yielded the optimal cut-off for sensitivity (0.80) and specificity (0.51) to differentiate between bipolar disorder and major depressive disorder. The HCL-32 was shown to be sensitive enough for hypomanic symptoms, but did not distinguish between type I and type II BD. Although the sensitivity of the HCL-32 is fairly similar to that of the MDQ, its specificity is much lower. Besides, the HCL-32 is not as quick and easy to administer as the MDQ.

4.12 Use of scales to rate disease severity

The Young Mania Rating Scale (Young et al, 1978), the Altman Self-Rating Mania Scale (Altman et al, 1997), the Internal State Scale (Bauer et al, 2000) and the Bipolar Affective Disorder Dimensional Scale (Craddock et al, 2004) are some of the rating scales that have been used to measure the severity of symptoms of patients with bipolar disorder. All of these scales must be administered by trained clinicians.

Such scales or staging criteria are of great potential value in developing disease state transition models, as they provide a quantitative measurement for representing the disease state. Thus, they are of particular interest to the OR modeller, especially if any particular scale turns out to be universally accepted, such as are common in cancer staging.

4.12.1 The Bipolar Affective Disorder Dimensional Scale (BADDS)

Craddock et al (2004) proposed the BADDS in order to overcome the drawbacks of simple classification and diagnostic scoring systems, which assign BD patients to one of a small number of distinct categories. BADDS is a numerical rating system, building upon the existing categorical classifications, which can be used in addition to traditional diagnostic procedures. BADDS has four dimensions: Mania, Depression, Psychosis and Incongruence. Each dimension uses integers in the range 0 to 100. The dimensional ratings are formulated using all available clinical data, including psychiatric case notes and semi-structured psychiatric interviews. The criteria used to make bipolar spectrum diagnoses follow DSM-IV and ICD-10. The BADDS has been tested for psychiatric assessment in family-genetic studies of bipolar disorder, unipolar depression and puerperal psychosis. Each of 20 patients was rated by two psychiatrists and seven psychologists using the BADDS dimensions to reach a consensus. Inter-rater reliability of this scale was found to be excellent, but sample limitations are clear. BADDS has been used in clinical samples, but it requires further validation study in a bigger sample in non-specialist and community based settings. The idea of dimensional classifications may provide a better structured approach than a conventional categorical approach, but it is clearly labour-intensive in practice and its interpretation in terms of care management may not be easy.

4.13 Long-term self-monitoring of bipolar patients

Some bipolar patients may experience different signs, symptoms, forms and features of the illness over their lifetime (Whybrow et al, 2003). It is difficult to set up medication programmes for such patients unless the early warning signs and the number and length of episodes can be identified (Leverich and Post, 2003). Therefore, long-term monitoring of the illness can be beneficial for patients, caregivers and clinicians to improve effective treatment plans (Leverich and Post, 2003).

To assess the effectiveness of maintenance treatment for bipolar disorder, a patient's course of illness need to be monitored in a systematic and accurate manner.

Driven by this need, self-monitoring tools such as the National Institute of Mental Health Life Chart Method and the Clinical Monitoring Form have been developed to be used by both patients and clinicians to achieve a depiction of the patterns of the illness in a clinically useful layout (Baldassano, 2005).

4.13.1 The Life Chart Method (LCM)

The LCM is the most widely used graphical method among those currently available. It allows a daily assessment of mood status to be documented, based on the level of mood related functional impairment (Denicoff et al, 2000; Leverich and Post, 2003). Hence, an imminent manic or depressive episode can be recognised early on. Life charting provides a retrospective viewpoint on the illness and allows clinicians and patients to recognise features that may cause episodes. An electronic version of the LCM form has recently been developed for use on a handheld device (Scherer, 2002). The LCM was initially developed for use in the research setting by professionals, but patients can also use the forms nowadays (Denicoff et al, 2000). Many interested bipolar patients have been trained to complete both versions of the LCM forms as part of their treatment plan.

4.13.2 The Clinical Monitoring Form (CMF)

The CMF is another method designed to monitor regularly the mood status of bipolar patients. The CMF has been validated against formal rating scales by Sachs et al (2002), and shown to be useful to the clinician to describe a patient's disease progression. The CMF is based on the compilation of factors that may trigger an episode. A computerised version of the form is presently under development.

4.14 Drawbacks of self-monitoring systems

Whether electronic-based or paper-based, a self-monitoring system may suffer from the following problems:

- it may be too time consuming, given the amount of data entry involved

- it suffers from its inbuilt dependence on the accuracy, completeness and honesty of patient-reported data
- use of routine self-report monitoring for clinical purposes may be expensive
- it requires a great deal of energy from patients
- it may not provide a sense of control over a patient's life
- regular monitoring requires motivation which some patients may lack
- intentionally or unintentionally, patients may forget to record the exact details later on (Whybrow et al, 2003)
- existence of other illnesses may prevent patients from recording data on time
- overall data quality may be negatively impacted by data entry errors
- it does not allow patients to document their physiological and contextual details that could be helpful for treatment purposes (Blum and McGill, 2008)
- self-monitoring has been shown to have fallen short of reducing depressive relapses (Perry et al, 1999)
- the National Collaborating Centre for Mental Health state that "most self-report scales are not very specific and are less sensitive to detecting problems with cognition and functional impairment" (NCCMH, 2006).

However, Blum and McGill (2008) state that monitoring via a set of sensors may free patients from the above drawbacks. They also advocated that mood status related data could well be reported via sensor networks as well as self-reporting in order to provide details about the course of a patient mental health, because this may help set up even better therapeutic association between patients and caregivers.

4.15 Performance comparison of bipolar screening tools

Most of the self-appraisal screening tools discussed in this chapter are clinically useful and timesaving. In particular, the BSDS and the MDQ are both sensitive

enough to be used to identify bipolar spectrum disorders in psychiatric practice (Ghaemi et al, 2005; Twiss et al, 2008). Both of these scales can be used as a complement of each other (Ghaemi et al, 2005). Although the MDQ has been validated in psychiatric settings across different countries (Hirschfeld et al 2000; Miller C et al, 2004; Weber et al, 2005; Konuk et al, 2007; Vieta et al, 2007; Twiss et al, 2008), the BSDS still requires further validation. In most cases, evaluation of the sensitivity and specificity of the MDQ met the desired requirement in identifying a high ratio of true positives while minimising the number of true negatives.

An ideal screening instrument for common use in clinical practice should be short and simple to complete. The MDQ tool is slightly shorter than the BSDS (13 as opposed to 19 questions). The MDQ has been successfully validated in a UK psychiatric practice (Twiss et al, 2008), whereas the BSDS has not yet been validated in any UK setting. However, both tools allow fast and simple processing. Table 4.2 shows the performance comparison of these two screening tools:

	MDQ	**BSDS**
Study conducted by	Twiss et al (2008)	Ghaemi et al (2005)
Population	Clinical	Clinical
Sample size	127	95
Validated against	SCID interviews	SCID interviews
Overall sensitivity	0.76	0.76
Overall specificity	0.86	0.85
Sensitivity for bipolar I	0.83	0.75
Sensitivity for bipolar II	0.67	0.79

Table 4.2 Comparative performance of the MDQ and the BSDS

5. OR Modelling in Healthcare

Mathematical Operational Research (OR) modelling approaches have been widely applied within healthcare for many decades and there is a massive literature on this topic. One aspect of this literature deals with the application of modelling techniques to deal with organisational and process issues such as resource allocation and service redesign. Examples of organisational modelling include scheduling outpatient clinics, staffing hospital Emergency Departments and scheduling surgical lists. However, the literature related to this thesis concerns the use of OR approaches to increase medical understanding of diseases, to evaluate treatment interventions and improve clinical decision-making. This chapter focuses on two issues: why healthcare is complex and how this complexity can be overcome.

5.1 Introduction

OR modelling and simulation are well-known scientific methods that have been widely used and proven in manufacturing, logistics, airlines and defence. The use of OR in healthcare began in 1952 (Flagle, 2002) and has been increasing ever since (Eldabi et al, 2008; Naseer et al, 2008; Jahangirian et al 2010, Lagergren, 1998). In the UK, healthcare is regarded as one of the key growth areas for OR (Royston, 2009). As in other industrial sectors, a variety of strategic and operational decisions have to be made within healthcare. The "production process" in healthcare usually consists of a clinical process (i.e. clinicians' decision-making behaviour) and operational processes (i.e. the delivery of a care service, along with the resources required to provide it). The clinical process is the core process in which a patient's health problem is diagnosed and treated through a set of linked decisions and tasks. On the other hand, operational processes involve the resources that support the clinical process. Lessons can be learned in healthcare from the application of

modelling and simulation in other industries. Taylor and Robinson (2006) identify the use of modelling and simulation in healthcare as a research priority.

Few would disagree with the statement that the complexity of any healthcare system is overwhelming. Pressure on healthcare providers is increasing worldwide, for many reasons: quality, safety and performance management, constant organisational restructuring, the need to provide a wide range of services, increased public expectation, demographic changes and the ageing population, government policies, clinical and pharmaceutical advances (and associated costs), and so on.

Healthcare is an enormous business in the UK, where the National Health Service (NHS), the third largest employer in the world, spends more than £70 billion annually that approximates to £8m hourly (DoH, 2006). In the USA, healthcare is regarded as the main domestic industry (Carter, 2002). Healthcare spending has risen rapidly over the last decade and this trend is unlikely to change. Royston (2005) reported over 7% annual growth in NHS spending. The pressures to control costs and to provide various healthcare services and facilities are increasing all the time.

Royston (2005) noted four important performance challenges in healthcare, including *i*) better quality and safety, *ii*) better access, *iii*) better personalisation and more participation and *iv*) better value for money. Healthcare systems require effective and sustainable decision making tools to help clinicians and administrators make better decisions. Decision-making is even more difficult when it involves a group of stakeholders with different perspectives, as is often the case in health. For example, patient benefits and provider costs often conflict when designing services.

However, quantitative models can aid decision-making in health and expert opinions can easily be incorporated. Quantitative modelling enables health service managers to decide in a reasonable and impartial way whether to invest in specific treatment or prevention policies (Cooper et al, 2007). To this end, one may analyse complex design issues and treatment pathways through modelling in a risk-free setting without having to wait years to prove an improvement (Rawlins and

Littlejohns, 2004). It has been argued by Eldabi et al (2007) that modelling and healthcare systems should gain benefits from each other in a "symbiotic" manner.

5.2 Modelling in healthcare

A model represents a simplified form of a real world process. Modelling, as an investigational method, can reduce cost, risk and interruption. Pidd (2010) defines a model as "*an external and explicit representation of part of reality as seen by the people who wish to use that model to understand, to change, to manage and to control that part of reality*". OR modelling is about using common sense, but it may also require using a fair bit of mathematics (Crane, 2007). Figure 5.1 shows a simple idea of OR modelling:

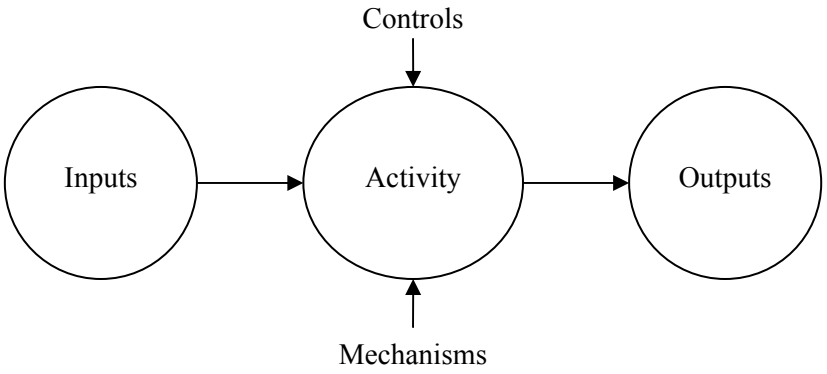


Figure 5.1 The basic principles of modelling

Our main focus in this thesis is on disease models. OR models can be used not only for the more traditional organisational issues described above but also for representing disease processes. Today, disease management is an effective way to control healthcare costs and to improve quality of care for patients who especially suffer from chronic diseases. The natural history of chronic diseases such as cancer, HIV/AIDS and depression is important for clinical reasons. It is, therefore, important to quantify a patient's progression through different stages of chronic disease for early detection, prevention and treatment purposes (Shih et al, 2007). The progression of a chronic disease can be modelled and studied through a multi-state model (Chan et al, 1996).

Generally, building a disease model requires knowledge about the disease itself, together with the care required and the treatments available in each stage of illness. Modelling requires detailed knowledge of the range of outcomes of any treatment, including any side-effects. It requires a good understanding of the clinical conditions that cause distress, dysfunction or difficulty to the person afflicted. A disease model can also incorporate other features, such as biological or physiological markers, psychological or sociological aspects. OR modelling can help understand disease processes better and form knowledgeable preferences of medical actions. The healthcare community has become increasingly more willing to apply modelling techniques as a common means to reach well-informed clinical, policy and treatment decisions in real world settings (Lagergren, 1998).

Despite this, and despite the large academic literature on this topic, the impact of OR simulation models has been somewhat limited in healthcare settings (Lowery et al 1994, Brailsford 2005 and Proudlove et al 2007). Many reasons have been put forward for this disappointing impact of OR simulation models in healthcare. Authors such as Butler (1995), Robinson and Pidd (1998), Harper and Pitt (2004), Brailsford (2005), Proudlove et al (2007) and Eldabi (2009) all tried to shed light on this issue. Some of the reasons are listed below:

- conflict of interests between modellers and stakeholders
- low levels of mathematical skill within health service professionals
- overcomplicated models published in OR academic journals
- lack of OR papers specifically aimed at healthcare professionals
- lack of awareness with the process and jargon of OR
- lack of in-house OR capability
- engaging external OR consultants is costly
- translating theory into practice is never easy

- cultural unwillingness to embrace computer-based models of patient care processes
- natural complexity and multiple interactions in healthcare environment
- rapidly changing national policies
- lack of involvement of stakeholders in the modelling process and difficulty keeping stakeholders on board continually.

In the view of Jun et al (2009), the role of OR modelling in healthcare would have been even more extensive had the healthcare community not lacked knowledge of a wide variety of approaches. However, Jun et al (2009) argue that clear guidelines can help the healthcare community to use various efficient modelling approaches acceptably and fruitfully, even though people may find an approach unfamiliar to start with. Lowery et al (1994) and Brailsford (2005) provide excellent discussions on the barriers (e.g. methodological suitability, stakeholder issues and data problems) of putting OR simulation models into practice in healthcare, and proposed a number of potential ways that the healthcare community can generate more uptake.

Nevertheless, disease modelling has been used very successfully in healthcare to evaluate treatments, screening and other interventions (Davies and Davies, 1994; Russell et al, 1996; Mauskopf, 1998). Such models permit experimentation with different policies, which may be too expensive, too time-consuming (or even unethical!) to test in reality with a randomised controlled clinical trial. The aim of the model is to act as a risk-free environment in which to compare options and gain deeper understanding. A typical structure for such a model is depicted in Figure 5.2 below. The boxes represent system states, which may, as in this case, simply be disease states. Depending on the modelling approach chosen, the objects in the system (i.e. patients) move individually or as a group from state to state through time. Different patients may spend different times in each state or follow different pathways through the system.

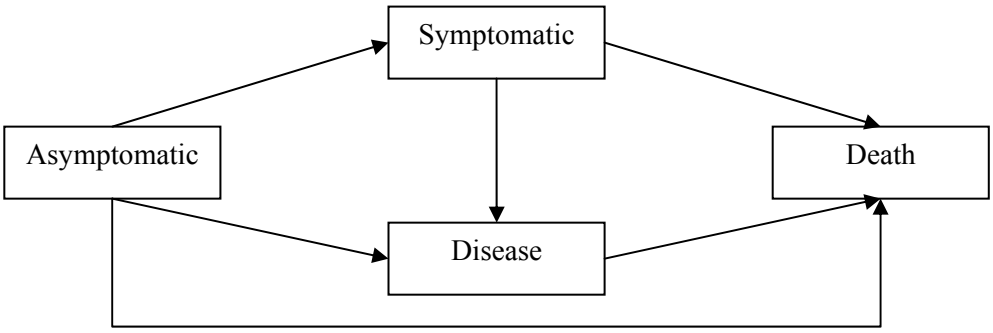


Figure 5.2 Transitions between different states in a disease model

5.2.1 OR models in chronic disease management

A chronic disease is a long-lasting, persistent or recurrent disease state. Examples include diseases such as diabetes, heart disease, asthma and (of course) mental disorders. Chronic diseases can be preventable and are often manageable by the patient themselves, although they tend to be very costly due to their long term and recurrent nature. Most chronic disease management models are concerned with cost-effectiveness analysis. Only a smaller body of work is dedicated to operational processes to improve patient flow.

5.2.2 Types of model in healthcare

The healthcare modelling literature contains examples of an extraordinarily wide range of mathematical OR approaches such as statistical analysis, simulation modelling, queuing theory, linear programming and integer programming. These methods are mainly used either to locate the optimal or a near optimal solution. In an extensive systematic review, Brailsford et al (2009) discovered a wide range of methods that are being used for healthcare modelling. They found statistical analysis as the most widely used method for healthcare modelling, while the second most widely used method was simulation modelling. Lagergren (1998) provided a broad range of bibliographic examples of operational applications of modelling, covering diseases such as HIV/AIDS, diabetes, cancer, heart disease and dementia. Fone et al (2003), Jun et al (1999), Wilson (1981) and Fries (1976) conducted literature reviews

of a wide range of OR modelling techniques that are applied to devise improved resource allocation decisions and healthcare planning.

5.3 Decision tree models in healthcare

Decision trees are widely used in disease models where clinical decisions have to be made and the long-term implications analysed. Decision trees are a family of analytic methods that include Chi-square Automatic Interaction Detector and Classification and Regression Trees. In principle, a decision tree can be used to represent any decision problem that complies with the assumption that all patients behave independently and do not compete for resources or interact with each other in any way (Barton et al, 2004). Decision trees or Markov models are often used to represent the flows of patients through various disease states. Many experts agree that decision tree models are inappropriate for modelling long-term health interventions and situations where events take place more than once (Sonnenberg et al, 1994; Karnon and Brown, 1998).

5.4 Markov models in healthcare

Markov models are used to represent a patient's stochastic progression through different health states over time. In a Markov model, a patient is assumed to be in one of a finite number of discrete health states (Sonnenberg and Beck, 1993). A Markov model is one of the most useful tools available to healthcare researchers to evaluate disease progression (Karnon, 2003). This approach offers exact, analytical outcomes for both time dependent progression and the steady state of the system. At each step, a patient may remain in his/her present state, go back to a previous state or move to totally a new state with a known probability distribution or rate of transition from one health state to another. The changes in state are called transitions, and during a time step or cycle, only a single state transition is allowed. The transitions between Markov states can be shown in a diagram involving all clinically significant events (see Figure 5.3):

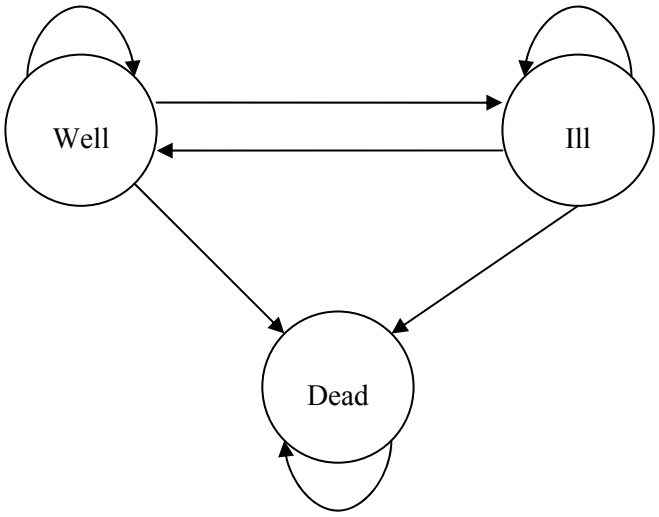


Figure 5.3 Example of state transitions within a Markov model

Markov states can be structured in a state transition matrix, which facilitates numerical evaluations. A state transition matrix is a square matrix that represents the probabilities of moving from one state to another. Each row of a transition matrix sums to one, since each row contains the probabilities of all possible transitions from the current state. It is more convenient to model prognosis for chronic conditions by means of Markov models (Sonnenberg and Beck, 1993). However, using a Markov model, it may not be easy to follow a patient's disease history accurately, due to the Markov "memoryless" property, which can formally be written as:

$$\Pr\left(X_{t+1} = x_{t+1} \middle| X_1 \ldots X_t = x_1 \ldots x_t\right) = \Pr\left(X_{t+1} = x_{t+1} \middle| X_t = x_t\right), \qquad (5.1)$$

where the state of the system at time (t + 1) depends only on the state of the system at time (t). This is not an issue when disease progression is independent of a patient's past medical history. However, this can be a limiting assumption since past medical history evidently plays a role in the progression of many diseases. This problem can be overcome by incorporating multiple health states, to capture small changes so that the disease pathway to a particular state is considered (Karnon and Brown, 1998; Lay et al, 2006). However, the use of too many health states in a Markov model may lead to a combinatorial "explosion" – known as the *curse of dimensionality* – making it very difficult or even impossible to solve the model analytically. In a situation like

this, simulation techniques can efficiently be used to model large Markov processes, and may often be the only practicable alternative (Klein et al, 1993).

A Markov model can be solved as a cohort simulation (one replication with many patients), as a Monte Carlo simulation (many replications, each for a single patient), or – if the number of states is tractable – by standard matrix algebra (Sonnenberg and Beck, 1993). Simulation is a relatively simple and flexible way to solve the model. A spreadsheet package can be used to model a deterministic Markov process. However, to model a stochastic Markov process, the spreadsheet would require use of a risk analysis add-in such as @Risk (Palisade, 2008) or Crystal Ball (Oracle, 2010) to solve the model stochastically, using randomly chosen input parameters from specified probability distributions. It is easier to represent time-dependent and recurrent events through Markov models, but they cannot model interactions between individuals (Barton et al, 2004). To allow interactions between individuals, which may occur in an infectious disease, discrete-event simulation models should be used instead.

Karnon (2003) used both Markov and discrete-event simulation models to assess the cost-effectiveness of the treatment of early breast cancer therapies. Their intention was to compare the outputs so that they can recommend an appropriate modelling technique. The outputs obtained from both models were fairly similar, but the long run time of the discrete-event simulation model was a notable disadvantage when compared to the run time of the Markov process. Furthermore, the discrete-event simulation model required to include more inputs. The author recommended the Markov process as an optimal method for their studied case study.

5.5 Simulation modelling in healthcare

A philosophy of approximation is expressed as “it is better to be approximately right, than precisely wrong”, which is also the philosophy of simulation modelling. Philippatos (1973) provides a useful definition of simulation: an artificial setting within which to assess the behaviour of the real-world system. Table 5.1 shows the use of simulation modelling in various sectors:

"Simulation Modelling" AND	Google Hits
Environment	136,000
Health	87,000
Transport	57,000
Agriculture	56,000
Defence	33,000

Table 5.1 Use of simulation modelling by sector. Source: Royston (2005)

In recent years, the use of simulation modelling for healthcare issues around the world has been recognised, since it is highly flexible and provides a risk-free setting in which to experiment. The degree to which simulation models could be applied in a range of healthcare areas varies from administrative or operational to medical or clinical issues (Pritsker, 1986). Mielczarek (2004) showed five key healthcare areas such as epidemiology, health care system operations, health and care systems design, medical decision making and crisis management, which have been supported by the simulation models. Lagergren (1998), in his paper, reviewed simulation models that have been applied to predict future incidence, prevalence and mortality for a wide range of diseases, as well as to plan disease intervention policies and assess various screening schemes.

The use of simulation in healthcare began in the 1960s (Brailsford, 2005), but the recognition of applying simulation methods for operational and clinical decision making has not been appreciated until recently (Fone et al, 2003). The periodic reviews of the use of simulations in healthcare conducted by Wilson (1981), Davies (1985), Lehaney and Hlupic (1995), Jun et al (1999) and Fone et al (2003) cover many healthcare related issues such as clinical decision-making, scheduling, resource allocation, screening and diseases. Jun et al's (1999) review of simulation studies found 8 studies from 1973 to 1977 and 28 studies from 1993 to 1997, and this provides an indication of increased acceptance of the use of simulation. Using Google

Scholar, Royston (2005) showed the increased trend of the number of citations of simulation in healthcare since 2000 (see Figure 5.5):

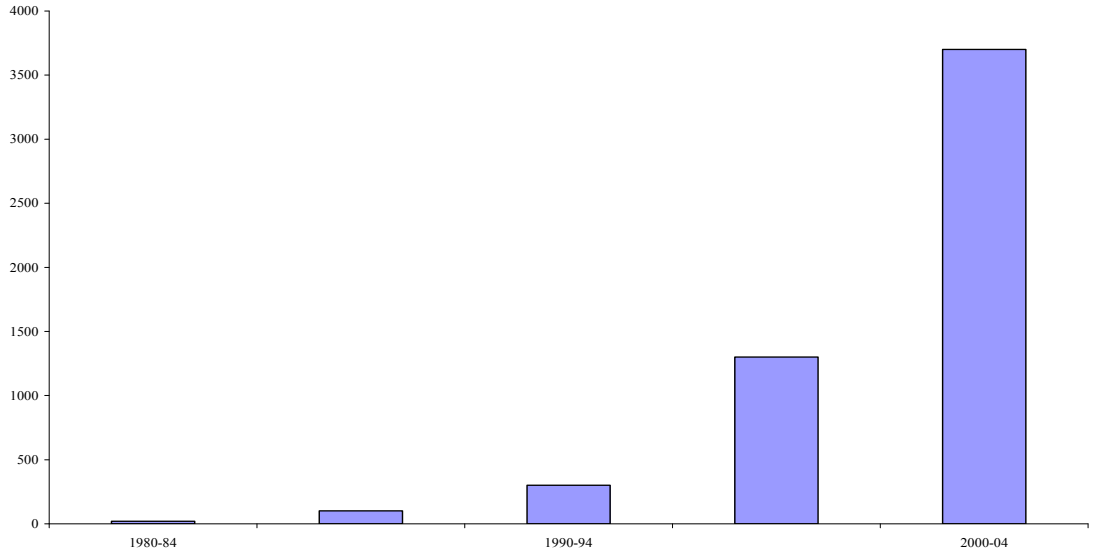


Figure 5.4 Growth in use of simulation in healthcare. Source: Royston (2005)

Simulation models have been applied for out-patient scheduling appointments (Worthington and Brahimi, 1993), capacity planning (Ridge et al, 1998), surgical bed planning (Millard et al, 2000), patient flow and resource utilisation (Rotondi et al, 1997) and waiting list management (Benneyan, 1997). Simulation has been massively applied within emergency departments. Brown et al (1999), Lane et al (2000), Coats and Michalis (2001), Eldabi et al (2002), Brailsford et al (2004) and Lattimer et al (2004) are amongst the many who have implemented a range of simulation techniques to improve aspects of healthcare within emergency departments.

5.5.1 Discrete-event simulation models in healthcare

Through the use of statistical distributions, discrete-event simulation (DES) models replicate the stochastic behaviour of a healthcare system over time very well. One of the key advantages of this method is its ability to model complex scenarios. DES can capture a vast amount of detail complexity. It is a flexible method which can describe health policy issues involving large population groups, but often specialist analytic knowledge is required to achieve greater flexibility (Davies et al, 2003).

Within a DES model, entities (patients) move from one event to another in sequential order. This type of model can capture individual attributes or characteristics. It is possible to assign significant risk factors to each individual, such as sex, age, disease history and attitude towards treatment, which may influence the path taken and the time between events (Davies and Davies, 1994). The pathways and times of patients' future events are typically sampled from parametric or empirical distributions derived from clinical data (Davies et al, 2003).

Heeg et al (2005) and Lay et al (2006) amongst many others argue that DES models are useful when it is necessary to consider a participant's past history to observe how it affects the course of the illness over time. Furthermore, unlike a decision tree or a Markov model, a DES model can incorporate interaction between individuals. Therefore, DES models allow more complex depictions of a system than Markov models (Campbell et al, 2001). The flow of patients through the healthcare process can be captured by such a model, and DES is clearly an ideal tool to describe the performance of a system in terms of patient arrival time, waiting time, service time, length of stay and capacity within an overall treatment network. However, DES modelling may be very time-consuming since such an individual-based model requires many repetitions to obtain statistically reliable approximations of the outcomes. In addition, it may require expert knowledge before developing and verifying such models (Davies et al, 2003).

Discrete-event simulations have been applied in a wide range of disease areas such as major depression (Lay et al, 2006), drug misuse (Zarkin et al, 2005), renal diseases (Huybrechts et al, 2005), mother-to-child HIV transmission (Rauner et al, 2005), gastric cancer and peptic ulcer (Roderick et al, 2003), liver transplantation (Ratcliffe et al, 2001), coronary artery disease treatment (Cooper et al, 2002) emergency admissions (Bagust et al, 1999) and outpatient clinics (Harper and Gamlin, 2003) and early breast cancer (Brown et al, 1999). Although discrete-event simulation models have successfully been applied at operational levels, its use to improve health policy has not yet been practiced (Gunal and Pidd, 2005).

An OR analyst may need to take account of human behaviour elements of the people involved to provide effective treatment. For example, distinction in human behaviour elements such as sleeping, change in appetite and talkativeness may indicate a treatment plan for depression. It is possible to include parameters that depend on human behaviour elements into OR models, and to this end, Brailsford and Schmidt (2003) describes the feasibility of incorporating human behaviour elements in a discrete-event simulation model to provide solutions for clinical decisions.

5.5.2 Monte Carlo simulation in healthcare

We noted in section 5.4 that one of the drawbacks of both decision trees and Markov models is the combinatorial explosion in the state-space as more and more complex disease states are introduced. A common approach for dealing with this problem is to use Monte Carlo simulation. Such models can handle this complex structure without growing unmanageably large (Barton et al, 2004). Monte Carlo simulation is a recognised approach in the field of healthcare, but it requires the analyst to set up a mathematical model of the process.

Put simply, Monte Carlo simulation uses the principle of random sampling to estimate outcomes of interest, which are too complex (or impossible) to be calculated using a mathematical formula. The values of the variables are sampled from probability distributions using simulation. Monte Carlo simulation methods have been widely applied in the construction industry, for assessing risks of infrastructural projects, calculating electricity options and forecasting the future contribution of biotechnology industry.

Using Monte Carlo simulation, like DES, it is possible to assign attributes to the entities within a model, which determine how that entity will behave. Moreover, the model can update these attributes while it is running. It is, therefore, possible to track individual entities over time using the output from a random number generator. Typical applications include cost calculations, inventory control calculations, demand projections, disease progressions and analyses of life years and quality-adjusted life years.

Monte Carlo simulations can be conducted using spreadsheet software such as Microsoft Excel. The additional functionality required to run multiple repetitions using Monte Carlo simulations is provided by add-ins such as @RISK, which can analyse the effect of varying inputs or outputs of the system being modelled. Through a Monte Carlo simulation approach, it is possible to run a highly complex scenario and determine the probability of a particular outcome. Moreover, crucially, this is not a point estimate: the whole range of likely outcomes can also be obtained. Furthermore, it is easier to test and improve model assumptions to forecast risk more accurately.

Economic evaluations of healthcare interventions can be performed using Monte Carlo disease simulations. Monte Carlo disease simulation works on an individual patient level, explicitly capturing the effect of variability between patients in both disease progression and response to treatments (Richter and Mauskopf, 1998). In addition, the question of risk related with individuals for whom the aggregate solutions produced by decision trees or Markov models are inappropriate can be addressed through a Monte Carlo disease simulation (Richter and Mauskopf, 1998).

6. OR Models in Mental Health

In the preceding chapter, we reviewed a range of Operational Research (OR) approaches that could be applied for various diseases. This chapter focuses on the use of OR modelling and simulation approaches in mental healthcare. We begin, however, by reflecting briefly on the philosophical nature of health and healthcare, which is of particular relevance in mental health.

6.1 Introduction

Mental health, which is a key public health issue, can be defined as "*a state of balance between the individual and the surrounding world, a state of harmony between oneself and others, a coexistence between the realities of the self, that of other people and that of the environment*" (Sartorius, 1983). Simply put, this is about how one thinks, feels and behaves. Shah (1982) states that mental health is "*the most essential and inseparable component of health… An integrated component of public health and social welfare programmes"*.

Mental health indisputably needs fostering, promoting and protecting to ensure that people with mental health problems can nevertheless enjoy a good quality of life, productive work, and meaningful interactions with others. Mental health care needs can be delivered in hospital and primary care settings, social services and self-help groups. Mental healthcare services are not merely confined to the treatment of affected people, but are closely related to all other health activities.

6.2 Biomedical and social models of mental health

Traditionally, medical care is usually defined by the "biomedical" model, which sees health as the absence of disease. In diagnosing diseases, physicians have been using the biomedical model of medicine as a conceptual model of illness for many years. Excluding the role of social factors or individual subjectivity, the focus of the

biomedical model of a disease is simply on physical or biological factors such as the pathology, the biochemistry and the physiology so as to comprehend a person's medical illness or disorder. The biomedical model also neglects the fact that a diagnosis is a result of a social mediation between doctor and patient (Annandale, 1998). Diseases such as mental disorders, heart disease and diabetes are very much dependent on a person's actions and behaviours. Therefore, it is restrictive not to take social factors or individual behaviour into account since, for example, it omits the idea of disease prevention.

Of course, the biomedical model has been undeniably useful for reducing morbidity and premature mortality. In mental health services, the biomedical model has been the dominant model since the main purpose behind the work of medically trained psychiatrists is the diagnosis and treatment of mental illness or disorders. However, mental health is a prime example in which the biomedical model has clear limitations. Therefore, there is a need for a much broader model of understanding mental health than the biomedical one alone to influence the work of psychiatrists, since mental health is a complex area in which one must focus not only on the disease alone, also on the role of way of life in mental health.

Psychiatrists are typically driven by the idea that medical science must alleviate mental diseases in order to return people to a state of health. This view regards health as an attribute that can be determined by the presence or absence of disease. The emphasis on the absence of disease is a significant indicator of good health, but ignores the effect of other important influences. The WHO states that "*Health is a state of complete physical, social and mental well-being and not merely the absence of disease or infirmity. The enjoyment of the highest attainable standard of health is one of the fundamental right of every human being, without distinction of race, religion, political beliefs or economic and social conditions*" (Constitution of the WHO).

We have seen that mental illness is usually caused by a mixture of individual, social and environmental factors. Thus, the social model can be effective in the

prevention of mental illness since it considers wider determinants (e.g. educational, cultural, social, spiritual, socio-economical and environmental), which may have an impact on people's health. Dalgren and Whitehead (1991) developed a model to show how a variety of factors influence health positively or negatively by trying to map the relationship between the individual, their environment and disease. The social model searches to educate people and allow them to recuperate control of their lives.

Duggan et al (2002) argue that the progressive social model in mental health should have the following key characteristics:

- it should embody the intricacy of human health and well-being
- it should pay attention to the internal and the external worlds of individuals
- it should facilitate interaction between social factors and biology
- it should commit to the development of theory and practice
- it should encompass the critical evaluation of process and outcome.

6.3 OR modelling for mental illness

The main purpose of this chapter is to review the current literature in relation to the use of OR techniques for mental illness, specifically focusing on the application of simulation for bipolar disorder. The English language academic and clinical literature was searched for documents in which the key focus was OR modelling and simulation in mental healthcare. We searched the electronic databases Medline and Web of Science, as well as the internet search engine Google Scholar and the web library catalogue TDNet.

It was clear from the literature review that the use of OR methods in mental healthcare is mainly limited to service system planning and cost-effectiveness analysis rather than disease modelling. For instance, Pagel et al (2008) applied OR techniques to reconfigure services with a view to improve access to a generic mental healthcare service. There are many reasons for this paucity of OR models to describe chronic psychiatric disease processes:

- unlike most physical diseases, there are few recognised staging or classification systems for mental disorders which lend themselves to modelling. The classification of mental illnesses is a crucial issue. The international classification of diseases (ICD-10, 1993) and the diagnostic and statistical manual for mental disorders (DSM-IV-TR, 2000) are currently the two main recognised diagnostic classification systems used to classify mental healthcare. However, there is a considerable scientific debate about the different kinds of categorisation (Manning, 2006) and psychiatrists still make diagnoses mainly by talking to the patient, using subjective judgments and their experience
- there is no single cause of any mental illness: the reasons for mental health problems are complex, and this complexity limits the practicality of modelling
- fear, misunderstanding and stigma are commonly associated with mental illness
- in mental illnesses it is difficult to quantify the risk of specific outcomes or behaviours. For example, one study shows that mental health professionals are mistaken 95% of the time when predicting aggressive behaviour (Ennis, 1972)
- the ethical aspects of research in this field can be very challenging
- it is extremely difficult to identify and recruit people who are at risk for mental disorders for prediction or prodromal behaviour studies (Heinssen et al, 2003)
- it is difficult to recruit sufficiently large samples to provide adequate statistical power to test hypotheses concerning illness onset and progression
- at times, it is difficult to persuade clinicians to collect relevant data
- in the UK, there is no national classification of mental health interventions except for children and adolescents (Elphic, 2007)
- it is difficult for professionals to arrive at an agreed approach in psychiatry. Menninger, in the 1960s, argued that diagnostic classification should be completely discarded (Farmer, 1997)

- lack of information technology and large databases
- reluctance of psychiatric patients to take part into longitudinal research studies (Whybrow et al, 2003)
- lack of research funding for mental health ("Cinderella diseases"): broadly speaking, mental health research is expensive
- lack of appreciation of the severity of mental illness
- many mental disorders such as depression and mania require person-specific assessment because different patients have different symptoms
- lack of uniformity in research findings and prevalence rates (Mee-Lee, 2006)
- it is impossible to define "abnormality" – any statistically-based definition would categorise people with extraordinary talents as abnormal too
- changing clinical processes as a result of modelling is particularly challenging because it requires changing clinicians' behaviour in a less "evidence-based" way than for physical disease.

6.4 Decision tree models for mental illness

Although decision trees can help identify the presence or absence of psychiatric symptoms in patients, the tools have been very rarely employed in the area of mental health (Batterham et al, 2009). This is probably due to the fact that predicting mental risk is more complex than predicting, for example, the risk of cardiovascular disease. However, decision trees have been used to predict suicide attempts in major psychiatric disorders (Mann et al, 2008), the relationship between neuroticism and depression (Schmitz et al, 2003) and quality of life in multiple sclerosis (D'Alisa et al, 2006).

6.4.1 Decision tree models to predict major depressive disorder

In a 4-year follow-up cohort study in Australia, Batterham et al (2009) investigated the likely effects of a reduction of risk factors that predict the presence of major

depressive disorder over time. In doing this, they employed a decision tree approach, which was compared with logistic regression analysis. Each cohort was followed up for a total period of 20 years. The results show that earlier depressive symptoms were significantly predictive after four years. Using the same predictors, the decision tree was found to have better sensitivity and specificity than the logistic regression. Their method was found to be useful in distinguishing and delineating a broad range of risk profiles over four years across different age groups.

The decision tree methodology was appropriate here, since it provided risk estimates by classifying patients into meaningful and significant categories. Furthermore, it demonstrated potential in assessing mental health risk. However, the range of risk indicators used to develop the risk model did not include the sleeping pattern, which is known to be a significant risk indicator in depression. A cohort approach reduces the possible selection bias, but validating a cohort model is not easy and it can be difficult to apply the results to a wider population. As such, further validation is required in other samples. Moreover, the study did not use a full clinical interview, and this may not have captured depressive episodes that may have occurred within a 4-year period. Over-reliance on patient reported data may also be a point of concern in the study.

6.5 Markov modelling for mental illnesses

Markov models divide the population into various health states, and the transitions between these health states occur according to assigned probabilities (Sonnenberg and Beck, 1993). Broadly speaking, Markov modelling is useful for providing a methodological framework within which to identify the presence or absence of psychiatric symptoms, or quantify the health benefits and healthcare cost savings.

Mentally ill patients visit their psychiatrist at differing intervals between episodes of a given type. Fisher and Knesper (1983) developed a Markov model to predict the operation of psychiatric services for such patients. Patten and Lee (2005) showed how modelling offers a methodological framework for incorporating

psychiatric epidemiological data. Besides, the prospect of using the computer to link different levels of healthcare as patients move through a system makes Markov modelling a potentially attractive approach in the design of care programs. The following discussion shows how Markov models can be applied in various aspects of psychiatric healthcare research.

6.5.1 Markov modelling for longitudinal data analysis

The "natural history" of major depression can be characterised by incidence, recurrence and episode duration. Incidence and episode duration are both important variables in understanding the dynamics of episodic chronic diseases such as major depression, as they can be used to estimate the prevalence of such diseases in the general population. However, it is not easy to estimate these variables without a full set of detailed long-term data. Patten and Lee (2004) applied a Markov approach to decrease the bias embedded in estimates obtained from data sources of the National Population Health Survey (NPHS) in Canada to underline the effectiveness of Markov models for long-term data analysis.

Patten and Lee developed a Markov model to concurrently simulate the associations among incidence and the incidence estimation and episode duration and the number of depressed weeks reported in the preceding year. A series of Monte Carlo simulations were applied over a 2-year time period to the Markov process, in order to identify the incidence and recovery probabilities resulting in the observed incidence estimations and numbers of depressed weeks. The annual incidence was calculated to be 3.1%, while episode duration was 17.1 weeks. Refined estimates using a Monte Carlo Markov model were found to be less subject to bias than the simple point estimates obtained by Kessler et al (1998) in the NPHS study for major depression. Therefore, Patten and Lee were able to get better estimates through modelling than those obtained without modelling.

However, in their approach, the non-depressed state was chosen to be the starting point for all simulated subjects in order to estimate the incidence. During the simulation interval, some depressed subjects at baseline may possibly get better and

subsequently undergo a recurrence, and this possibility was not considered in the incidence estimation. A health state such as "moderately depressed" could have been incorporated in the model to refine the severity of the illness so as to make the model even more useful for public health applications. Although the estimates provided are of assistance for health system priority setting and planning, connecting these epidemiological notions to clinical concerns such as risk and prognosis could offer more impact on the accountability of major depression in the population as a whole.

6.5.2 Markov modelling to enhance clinical decisions

There is a clear need to integrate epidemiological prevalence data into clinical decision making, and a Markov model may provide a useful methodology for doing this. Patten (2005) applied Markov modelling to describe incidence, prevalence and recovery from major depressive episodes to enhance clinical decisions. This analysis used data from a series of Canadian survey studies. Patten amalgamated epidemiological data from three different sources to describe incidence and recovery from unipolar major depressive episodes involving two health states: depressed and non-depressed. Monte Carlo simulation over a series of one week time steps was applied to fit model parameters to the epidemiological data, using incidence-related transition probabilities and weekly recovery probabilities obtained from a previous study conducted by Patten and Lee (2004).

A higher incidence in females confirmed the correlation of gender with major depression, while no evidence of correlation with educational level was reported. The findings from the Markov models developed by Patten (2005) are mostly consistent with the findings from Eaton et al (1997), who ran a follow-up study to monitor a similar prognosis in major depression. However, Eaton et al (1997) reported no significant correlation between marital status and episode duration, but Patten (2005) "found that an effect of unmarried status on prevalence was due to an impact of this variable on episode duration". Both Patten (2005) and Eaton et al (1997) observed decreasing incidence of major depression with age. The recovery pattern described by

Patten (2005) was similar to that found in other related studies carried out by Melartin et al (2004) and Vos et al (2004).

One source of data in Patten's (2005) study was obtained from a cross-sectional study (a "snapshot"). Prevalence estimates for major depression have commonly been assessed through a number of cross-sectional studies in recent decades. A cross-sectional study can be used to estimate various outcomes and risk factors, but it is not easy to make deductions about changes over time. The standard purpose of a cross-sectional study is to determine the prevalence of a particular condition in the chosen population at a given point of time. As such, different results may well be found at another point of time. A cross-sectional study is subject to the prevalence-incidence bias, especially in the case of a chronic disease like depression, because any risk factor that results in death will be under-represented among those with the disease.

6.5.3 Markov modelling to describe a longitudinal course

The benefits of using Markov models with epidemiological data of major depression to obtain a range of parameters such as incidence, episode duration and recovery have been discussed above. The longitudinal course of major depression can also be represented through a Markov model, and this was explored by Patten and Lee (2005). They amalgamated data from three surveys carried out by Statistics Canada in a Markov model to represent the long-term path of major depression. They separated the subjects into low, moderate and high recurrence categories. The values for the weekly transition probabilities in the model were 0.00028 for low risk, 0.0010 for moderate risk and 0.00575 for high risk to describe incidence, recurrence and recovery of major depression. Based on these values, they expressed the annual incidence of the subjects in low, moderate and high recurrence groups for major depression as 1.4%, 5.1% and 25.9% respectively using the following formula:

$$\textit{annual cumulative incidence} = 1 - (1 - \textit{weekly transition probability})^{52} \qquad (6.1)$$

To decrease random variation in the output, they used Monte Carlo simulation with 50,000 replications over the six years simulation time period. The simulation output yielded 16.8 weeks as the mean number of weeks for patients being depressed. The use of Markov modelling in Patten and Lee's (2005) study to describe the longitudinal course of major depression enhances the possible advantage of modelling in characterising the condition. And their study illustrated the possibility of integrating several data sources in epidemiological modelling to describe a rational epidemiological representation. However, their model did not include mortality, which may have an effect on the dynamics of epidemiology. Furthermore, the results of their study may be subject to bias due to any minor difference between the three surveys. However, they have clearly shown the possibility of obtaining dependable descriptions of major depression through modelling.

6.5.4 Markov modelling for cost-effectiveness analysis in depression

Depression has a significant bearing on the individual and society. It may grow as a recurrent illness in many patients, and the higher the number of previous episodes of depression, the higher the risk of recurrences (Solomon et al, 2000). The illness is associated with a high cost for social and health care. In 2000, the annual cost of managing the illness in USA was found to be $83.1 billion, of which 31% for therapeutic costs, 7% for suicidal mortality costs and 62% for costs related to missing output due to the illness (Greenberg et al, 2003). In Europe, the annual prevalence of major depression was estimated by Paykel et al (2005) at about 5%. The prevention of recurrences is associated with effective treatment plans. A longitudinal outlook in the treatment of patients with the illness can be beneficial.

Sobocki et al (2008) combined the results of several studies with a modified Markov model to examine the economic evaluation of longitudinal maintenance treatment with venlafaxine in recurrent unipolar major depressive patients from the Swedish healthcare setting. In a cost-utility study, Sobocki et al (2008) used the following standard "incremental cost-effectiveness ratio (*ICER*)":

$$ICER = \Delta C/\Delta E = (C1 - C0)/(E1 - E0)\ , \qquad (6.2)$$

where ΔC and ΔE represent the difference between intervention and no intervention 'in total cost' and 'in effectiveness' respectively.

Their model included high risk patients with recurrent depressive episodes from the population studied in the clinical trial led by Keller et al (2006). Sobocki et al (2008) estimated a Weibull survival function based on survival data from the above clinical trial to calculate time to relapse comparing venlafaxine with placebo. Data on direct costs, quality of life and transition probabilities were drawn from several naturalistic long-term observational studies, while the mortality risks data were obtained from the literature. Stochastic analyses were executed using Monte Carlo simulations with 10,000 runs to describe the uncertainty of key parameters. Sensitivity analyses were performed to estimate the variation of underlying parameters used in their model.

Their approach of using meta-analysis in a Markov model may well be less subject to bias. However, combining results of several studies may not reach well-founded conclusions since there remain some qualitative differences between the subjects of studies. In Sobocki et al's (2008) model, remission was chosen as the starting point for all simulated subjects who were at high risk of recurrent depressive episodes. Hence, the results may not be totally free of bias since it ignored patients who were in episodes at baseline. It is also unknown if their method would yield the same outputs, had it considered both high and low risk patients.

Sobocki et al (2008) showed that two years' maintenance treatment with venlafaxine in recurrent depression is cost-effective. They compared venlafaxine with placebo, but they did not discuss if venlafaxine is more cost-effective than other anti-depressants available. "*Psychotherapy, alone or combined with medication, has been shown to be effective in preventing further episodes of depression*" (Nierenberg et al, 2003). According to Burgess et al (2003), it is possible for a depressive patient to suffer high recurrence rates despite effective treatment, so treatment alone may not be the only solution. Lam et al (2000 and 2003) reported that depressive recurrence can

be delayed effectively through cognitive behaviour therapy. Adding both self-help and psychological components in addition to an anti-depressant may be even more cost-effective in the long run.

6.6 Discrete-event simulation modelling for mental illnesses

This approach has been rarely used for mental disease modelling. Examples of how discrete-event simulations can be applied in various decision-making processes are discussed below.

Kuno et al (2005) described a discrete-event simulation model to convince decision-makers to consider exercising model-based decision support tools for health service system planning for individuals with serious mental disorders. Their case study was based on the Philadelphia health system of psychiatric care. They depicted the existing psychiatric health care system and the nature of client flow within that system. It was a good choice of modelling since the decision-makers could use the model to visualise different policy scenarios in order to improve the system.

Klok et al (2007) investigated the cost-effectiveness of quetiapine, a treatment during acute mania in bipolar I disorder, evaluated with other substitutes such as lithium, olanzapine and risperidone. They developed a discrete-event simulation model to investigate the effectiveness of quetiapine specifically focusing on severe side-effects. They reported a better-quality combination of lithium with quetiapine than lithium with olanzapine or risperidone in terms of reduced possibility of picking a serious side-effect. The study was carried out in Netherlands.

Recently, Heeg et al (2005) developed a discrete-event simulation model to reflect the treatment for British schizophrenia patients with multiple psychotic episodes. Over a 5-year time period, the model considered the dependencies among an extensive array of factors into the decision making process, including gender, disease severity, symptoms, psychiatrist visits, treatment and treatment location, occurrence of psychotic relapses, compliance level, quality-adjusted life-years and side-effects. The model was used to estimate the economic evaluation of atypical

relative against conventional antipsychotics. Their method was applicable to schizophrenia since it required taking patients' histories into account.

6.6.1 Discrete-event simulation for major depression

Assessment of episode duration is useful since increased episode duration is linked with various causes such as physical illness, reduced social support and severity of the illness (Spijker et al, 2004). There is no association of increased episode duration with demographic variables (Spijker et al, 2002). However, Spijker et al (2002) and a research study carried out by Keller et al (1992) have shown that the chance of recovery from a major depressive episode decreases if an episode duration increases, and thus epidemiological data on the prediction of episodes may have important implications for clinical practice.

Patten (2006) predicted the probability of recovery during a specific time interval. In doing this, he first used cross-sectional Canadian psychiatric epidemiological data to fit a Weibull distribution to the duration of episodes. Using the longitudinal data from a sample from Canadian general population, Patten then developed a discrete-event simulation model and standardised the model by considering incidence and the Weibull distribution as inputs to the model. The resulting estimates from the model were used to calculate the probability of recovery. After six months, the probability of recovery during a subsequent week for an individual with major depression was found to be less than 1%, but the chance of recovery was estimated to be about 20% in the first week of illness.

The use of discrete-event simulation was a good choice in this study since the calculated episode duration was time-dependent, and the choice of the Weibull model was also good. However, as the model depended on data reported by patients, recording an incident incorrectly might have led to inaccuracy in the model calculations. Since the model does not specify the severity of a patient's state and since treatment may not be required for all major depressive episodes, it may not be easy for a clinician to identify an appropriate level of active treatment based only on

this model. Effects such as social changes and negative life events taking place over time as episodes progress may reduce the recovery rate as episodes grow longer.

6.6.2 Discrete-event simulation to describe depression prognosis

Lay et al (2006) developed a model in order to show the advantages of discrete-event simulation modelling over Markov modelling within the area of major depression. The model attempted to describe the prognosis of major depression. Although Markov modelling was adequate to model this disease progression, the authors believed that DES models could be even better in this regard, because of their flexibility in allowing patients with varying attributes to move from state to state in time order by taking into account attributes such as age, sex, disease history, treatment, suicide attempt and adverse events.

According to the authors, the main drawback of using a Markov model is the Markov "memoryless" property. They claimed that incorporating multiple patient-specific socio-demographic characteristics would only be possible by adding more health states. Therefore, they proposed DES modelling as a particularly useful technique, compared with Markov modelling techniques.

To argue with Lay et al's (2006) recommendation, it can be said that it takes a long time to develop a discrete-event simulation model owing to the need to include a large amount of detailed data in the model, and it is not always easy to find appropriate and good quality data (Davies et al, 2003). The price that must need to be paid to model such complex scenarios using a DES model is lengthy run times (Davies et al, 2003; Barton et al, 2004). Specialist software or programming skills are required to build DES models (Barton et al, 2004). Furthermore, discrete-event simulation models are complex to understand and interpreting the outcomes of such models require good statistical knowledge, and some may view this as a disadvantage (Davies et al, 2003). In general, simplicity in model buildings is beneficial (Ward, 1989).

The need to take various attributes into account to depict disease progression was not fully justified by Lay et al (2006). More research is required to more fully clarify the effectiveness of taking a variety of factors into account and to analyse the empirical value. For example, Patten (2005) reported that education had no effect in unipolar major depressive disorder, while Eaton et al (1997) reported no effect of marital status for the same disorder. In addition, a range of other factors may well be a crude pointer in depicting the progression of major depression in the study. Lay et al (2006) themselves pointed out that discrete-event simulation "may induce over-specification, whereby possible patient pathways become more complex than necessary, thus implying an increase in data requirements". Probabilistic sensitivity analysis can offer a broader representation of any associated uncertainty in notifying policy decisions.

6.7 Summary: OR modelling for bipolar disorder

We have presented the current literature on the use of OR modelling for psychiatric disorders and have set up a theoretical framework for our research. In general, there are very few examples of OR models for mental healthcare in the literature. In particular, there are no examples in the OR literature for models of the full spectrum of bipolar disorder. We found models for the depressive aspects of BD, but no single OR model which describes both aspects of the disorder. Most statistical studies of unipolar major depression have estimated characteristics that are applicable for public health applications, but not as markedly useful for informing clinical practice. There is an obvious gap in the literature and a real practical need for OR models to deal with both aspects of bipolar disorder in a robust manner, and are therefore applicable to clinical practice.

The literature suggests that the majority of mentally affected patients can be freed from requiring long-term or unnecessary hospitalisation if alternative home care and support services are obtainable. To this end, OR modelling can surely play a big role. Marneros and Brieger (2002) state that main features of studying the longitudinal course of any mental disorder are: onset of the disorder (type of onset,

age of onset), episodes (type, length, number), cycles (length, frequency) and outcome (end of follow-up in a defined period of time). An OR model can easily include these features, but it may not be possible for a model to predict any outcome with full precision. Having said this, a model may nevertheless be applied to make practicable and acceptable decisions.

Before choosing an appropriate model, a modeller must consider many aspects of the approach, including methodology, treatment strategies, key features of a disease progress, risk factors, timing, costing and expertise. It can cautiously be said that both Markov and discrete-event simulation models are efficient approaches in psychiatric healthcare research. Although validating such a model may involve a major research attempt, it may be commendable as the validated model for the different aspects of bipolar disorder could then be applied effectively. The use of OR modelling techniques in providing early detection of transitions between the normal, manic and depressed stages of both aspects of bipolar disorder may render very good reasons for future research, but it would require adequate understanding of the disease process from the modellers.

7. Methodology

This chapter describes the health state transition model that we have developed in order to help in the design of the PAM system. The modelling methodology, choice of approach and the model itself are all presented. The rationale for modelling decisions is explained.

7.1 Introduction

It is more difficult to apply OR methods to model mental diseases than physical diseases. Guidelines for prediction and prevention of mental diseases are plagued by the lack of evidence about the collective prognostic effect of identified risk factors. Section 6.4 of Chapter 6 provided a list of various reasons of underlying difficulties. To overcome these difficulties, an integrative research approach may be beneficial to better understand the complex procedures of mental illness.

The literature search did not reveal a single OR modelling approach that combines technology and modelling to describe the dynamic behaviour of bipolar patients to promote efficient monitoring and healthcare networking. Building models for bipolar disorder is clearly difficult, but we have attempted to build a multi-state transition model that is combined with computer simulation. This is the first example of an OR model to analyse the behavioural activities of bipolar patients. The model parameters were derived from the literature.

7.2 Choice of OR modelling approach

We decided to employ a very simple Excel-based Markov state transition model for the basic disease process, combined with Monte Carlo simulation (using @Risk) for generating the stochastic behaviour (in terms of daily activities) of our simulated individuals, and the corresponding stochastic data collected by the PAM system under a range of different scenarios. Thus, we would be able to test the performance

of the PAM system. For bipolar disorder, we clearly need to use a dynamic approach since the disorder evolves over time. The first step is to study the clinical literature to understand the natural history of BD, and thus define the clinical states required for the Markov model. The next stage was to embed this in a spreadsheet model which represented the activity patterns of hypothetical patients, and then somehow to model the collection of data from different configurations of sensors and the subsequent analysis and interpretation of these data by the PAM algorithms.

The literature suggests that Markov modelling and Monte Carlo simulation would be appropriate, since interaction between patients is irrelevant, the time element needs to be modelled and events in the model are recurrent. Furthermore, this approach allows us to keep the model fairly simple, which is advantageous in terms of flexibility, data requirements, run speed and interpreting the outcomes. The use of Monte Carlo simulation enabled us to incorporate variability and uncertainty.

7.3 The natural history of bipolar disorder

Bipolar disorder fluctuates over time in severity of symptoms, polarity of episodes and cyclicity (Kalbag et al, 1999). In fact, different patients show different rhythm (Kalbag et al, 1999; Goodwin and Jamison, 1990; Slater, 1938). Goodwin and Jamison (1990) suggest that bipolar patients vary from each other on severity of depressive, manic and mixed states, polarity and duration of episodes, cyclicity, instability or rapidity of state changes, and treatment responsivity.

7.3.1 Episode polarity

Almost all bipolar patients experience both manic and depressive episodes; the majority experience no mixed episodes (Miller I et al, 2004; Judd et al, 2002). Bipolar patients who experience mixed episodes appear to be at higher risk for rapid cycling; and rapid cycling has usually been found to be brief in several studies (Angst and Sellaro, 2000). To avoid further complexity, one may classify the episode polarity as either depressed or manic, as proposed by Kalbag et al (1999). However, according to Kalbag et al (1999), it is still possible to inform if a patient experiences mixed

episode or not by using a set of relevant decision rules. Kalbag et al (1999) identified mixed episodes by oscillating 5-day durations of fully syndromal manic and depressive symptoms simultaneously or separately within the same day or week.

7.3.2 Asymptomatic and symptomatic periods

Judd et al (2002) report that almost all bipolar patients have both asymptomatic and symptomatic phases. They also report that the median proportion of time that patients were asymptomatic in their study was 62% and symptomatic 38% (at mild, moderate and severe levels of severity). Similarly, Miller I et al (2004) report the median time as 59% for asymptomatic and 41% for symptomatic. Miller I et al (2004) also report that about 90% of patients experienced some time in all three of the symptom levels (asymptomatic, partially symptomatic or fully symptomatic).

7.3.3 Duration of symptomatic episodes

The median duration of acute episodes in bipolar illness is 3 to 6 months in clinical studies, and 2 to 3 months in epidemiological studies (Angst and Sellaro, 2000). The overall median length of episodes has been reported by Angst and Preisig (1995) as 3 months in a follow-up study. According to the UK bipolar disorder treatment guidelines, the average lengths of manic and depressive episodes are 9 and 13 weeks respectively (NCCMH, 2006). Angst and Sellaro (2000) suggest that over the past 120 years, the natural length of episodes has changed little.

7.3.4 Intervals between episodes

Falret (1851, 1854) and Kraepelin (1913) conceded the existence of mild fluctuations between episodes. Slater (1938) followed up 116 patients diagnosed with bipolar illness by Kraepelin, to investigate the length of intervals between episodes, and observed the existence of symptom-free intervals between episodes for most of these patients. Kalbag et al (1999) propose nine bipolar subtypes, and they marked the classic bipolar I pattern with periods of remission between fully syndromal manic and depressive episodes. In fact, most of the bipolar subtypes proposed by Kalbag et al

(1999) concede inter-episode intervals. Very recently, Fajutrao et al (2009) assumed that bipolar patients use euthymia as a transit before moving to an acute mood event.

7.3.5 Changes in polarity

Dunner et al (1979) reported a frequency rate of 0.54 episodes per patient per year in an 11.4-year follow-up of 140 bipolar I patients. They reported no association between episode frequency and age of onset. Similarly, Judd et al (2002) reported 0.6 median changes per patient per year in a 13-year follow-up of 146 bipolar I patients. Over the 18 months cycle, Paykel et al (2006) also reported changes in polarity.

7.3.6 Cycle length

In Zurich follow-up study, Angst and Preisig (1995) reported the median cycle length of 18 months. Median time to remission of bipolar I patients has been reported as 5 months in Miller I et al's (2004) study and 6 months in Keller et al's study (1986). Since the median is a more representative statistic in this context, we have chosen an 18-month horizon in our model for representing a segment of the disease process.

7.3.7 Onset of episodes

Depressive symptoms are more common in the longitudinal course of illness than manic or hypomanic symptoms (Judd et al, 2002). At least 5 out of 10 patients start with depression as their first episode as opposed to mania (Perugi et al, 2000). However, Kinkelin (1954) observed 146 patients in an extended study and reported that almost 86% patients started with a depressive episode.

7.3.8 Ranges of manic and depressive measurement

Based on two observer-rated scales (Young Mania Rating Scale and the Hamilton Depression Rating Scale), Bauer et al (2005) used the following scale to rate the various phases of the illness: 0-19 (moderate to severe depression), 20-39 (mild depression), 40-60 (euthymia), 61-80 (hypomania) and 81-100 (moderate to severe mania). On the other hand, Kalbag et al (1999) used the Life Chart Program^{TM} that is

designed as 0.0 (asymptomatic), 2.5 (mild symptoms), 5.0 (moderate symptoms), 7.5 (severe symptoms), or 10.0 (extreme symptoms). Both of them classified the episode polarity as either manic or depressed.

Having considered the above scales, one may propose mild to severe depression at the bottom of a scale, euthymia (i.e. normality) at the middle and mild to severe mania at the top. Like depression and mania, normality also ranges between two points due to the existence of residual symptoms. Kraepelin (1913) and many others (Judd et al, 1998; Paykel et al, 2006; Bauer et al, 2007) argued the existence of mild remaining symptoms following recovery from the manic and depressive episodes. The symptom levels are commonly labelled as mild, moderate and severe, but no one can be certain about the order in which the symptom levels change on the scale over time.

7.4 Modelling framework

A diagrammatic representation of the computer modelling framework is depicted in Figure 7.1:

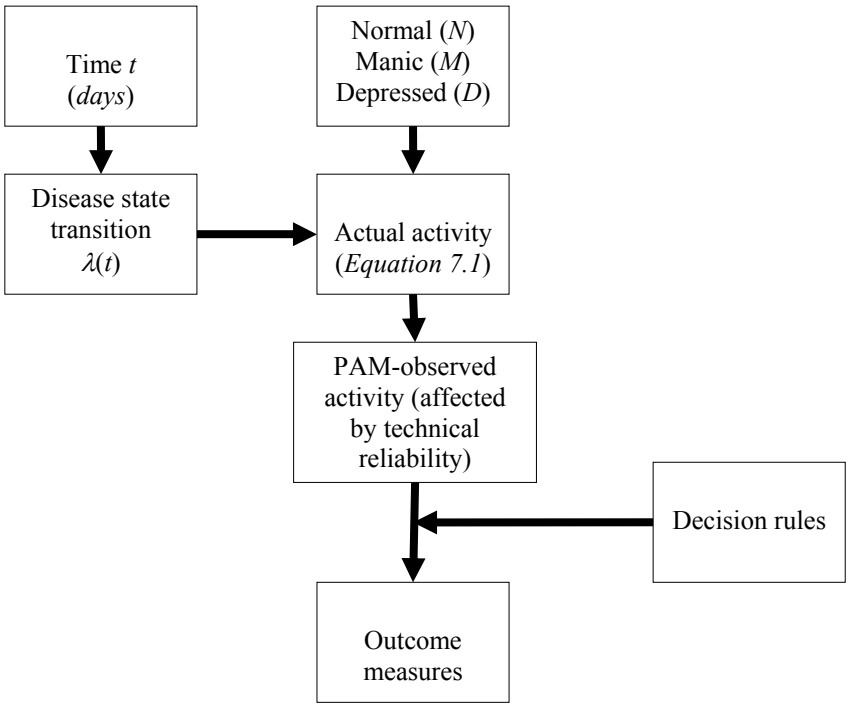


Figure 7.1 Modelling framework

A conceptual description of the above modelling framework is provided below:

- Step 1: At the start of each day t, the time-dependent health state value $\lambda(t)$ was generated stochastically, based on an archetypal patient trajectory (as depicted in Figure 7.2) through the "gross" health states *Normal – Depressed – Normal – Manic*, but with individual minor daily fluctuations within each gross state sampled using Monte Carlo simulation. Thus while all patients follow the same broad overall pattern of illness, no two patients will have exactly the same λ trajectory.

- Step 2: For each PAM-observable behavioural activity, for example hours of sleep, the actual, true value representing that activity on day t was calculated using Equation 7.1, which involves parameters $\lambda(t)$, N (the value for that activity for that person when normal), M (value when manic) and D (value when depressed). To account for natural individual daily variability unrelated to bipolar disorder, the actual values of N, M and D for day t used in Equation 7.1 were sampled from probability distributions describing the range of possible values for that parameter. To use the example in Figure 7.4, a person may sleep for $N = 8$ hours when fully normal. But even when completely healthy that person will not sleep for precisely 8 hours every night, so some natural random variation is added by sampling a value for N from a pre-determined interval representing the normal range for that individual, for example [7, 9]. The same is done for D and M, and then substituted in Equation 7.1 to yield the number of hours actually slept on day t: for example, 7.22.

- Step 3: In order to account for PAM technical reliability, the PAM-observed value of each activity was derived by sampling a value in a specified interval (small or wide, depending on the level of reliability) containing the true value. For the above example, if PAM is very reliable at detecting hours of sleep, the PAM-observed value would be sampled from within a small interval around 7.22, say [6.82, 7.62].

- Step 4: The decision rules of whether or not to send alerts were applied. For example, if the PAM-observed value of hours of sleep was say 6.84, this could mean the person was in the very early stages of a manic episode, but of course it could just be within the normal spectrum. So should the patient be alerted, or not?

- Step 5: The overall effectiveness of the PAM system (in terms of true/false positive and negative alerts) was calculated.

Based on the information from the clinical literature described in Section 7.3, we have represented the disease progression of bipolar I disorder in terms of a single parameter denoted by λ. The parameter λ can be conceptualised as a measurement of a person's mental health status, similar to the Young Mania Rating Scale and the Hamilton Depression Rating Scale described in Section 7.3.8. Although in reality this parameter is continuous, we have discretised it in the Excel model and thus we have a multi-state Markov model in which λ can take values between 0.00 and 1.00. The time-step is one day, and each day the value of λ either stays the same or is incremented or decremented, with a certain probability (shown in Table 7.3 below).

The parameter λ models the progression of individuals through the various disease states over a time horizon of 18 months, as described in Section 7.3.6. The λ values range between 0.00 and 1.00. Following Bauer et al (2005), values of λ between 0.00 and 0.19 represent severe to moderate depression; values between 0.20 and 0.39 represent mild depression; values between 0.40 and 0.60 represent normal health; values between 0.61 and 0.80 represent mild mania; and values between 0.81 and 1.00 represent moderate to severe mania. We then split the ranges 0.00 to 0.19 and 0.81 to 1.00 into a pair of two equal ranges (Kalbag et al, 1999): 0.00 to 0.09 and 0.10 to 0.19 represent severe and moderate levels of depression, while 0.81 to 0.90 and 0.91 to 1.00 represent moderate and severe levels of mania, respectively. These are summarised in Table 7.1:

Value of λ	Clinical state	Source
0.00 – 0.09	Severe depression	Kalbag et al (1999)
0.10 – 0.19	Moderate depression	Kalbag et al (1999)
0.20 – 0.39	Mild depression	Bauer et al (2005)
0.40 – 0.60	Normal health	Bauer et al (2005)
0.61 – 0.80	Mild (hypo) mania	Bauer et al (2005)
0.81 – 0.90	Moderate mania	Kalbag et al (1999)
0.91 – 1.00	Severe mania	Kalbag et al (1999)

Table 7.1 Values of the parameter λ and their clinical interpretation

In reality these transitions can be very subtle and gradual. An individual may move almost imperceptibly from the "normal" range to the depressed or manic range. The time spent making this transition will of course vary from individual to individual, and the boundaries between the "gross" states (Depressed, Normal and Manic) are blurred. To test the PAM system, we required a natural history model which represented an entire bipolar cycle, i.e. we needed to construct a complete and realistic trajectory of an "archetypal" BD patient including periods of depression, mania and normal health. The parameters given below are all derived from the literature in Section 7.3. In the Monte Carlo simulation, they are all subject to minor random fluctuation in order to create individual variability. Thus, although all patients follow the same general pattern, each individual has a slightly different trajectory.

Following Angst and Preisig (1995) we used a cycle length of 18 months (546 days), of which an average of 328 days (60%) are asymptomatic and an average of 218 days (40%) are symptomatic (Judd et al, 2002; Miller et al, 2004). Symptomatic periods include both depressive and manic episodes. On average, out of their total symptomatic time patients spend 130 days (60%) being depressed and 88 days (40%) being manic (Angst and Sellaro, 2000; Judd et al, 2002; NCCMH, 2006). We allowed symptom-free intervals between episodes (Slater, 1938; Kalbag et al, 1999), and in doing this so, we split the asymptomatic period in two equal halves (164 days each) to allow a change in polarity in the cycle length (Dunner et al, 1979; Judd et al, 2002; Paykel et al, 2006). It is not possible in our model to transition directly from depression to mania without passing through the asymptomatic state; this assumption was also made by Fajutrao et al (2009). Thus, we have not modelled mixed states.

Many patients start with depression as opposed to mania (Kinkelin, 1954; Kalbag et al, 1999; Perugi et al, 2000; Judd et al, 2002), and we have applied this in our model. We assigned mild, moderate and severe symptoms levels (Kalbag et al, 1999; Judd et al, 2002; Miller I et al, 2004; Bauer et al, 2005), and assumed that the symptom levels change over time from mild to moderate to severe. On average, out of the total 130 days of depression, patients spend 39 days in a mild depressive state (30%), 55 days in a moderate depressive state (42%) and 36 days in severe depression (28%). Similarly, out of the total 88 days of mania, 23 days (26%) are mild, 43 days (49%) are moderate and 22 days (25%) are severe (Judd et al, 2002). Figure 7.2 shows the overall sequence in which patients move from one state to another:

Normal States	→	***Depressive States***	→	***Normal States***	→	***Manic States***
0.4…0.6 (0.5 ⇒ *fully normal*)	→	0…0.39 (0 ⇒ *fully depressed*)	→	0.4…0.6 (0.5 ⇒ *fully normal*)	→	0.61…1 (1 ⇒ *fully manic*)
1 to 164 (*days*)	→	165 to 294 (*days*)	→	295 to 458 (*days*)	→	459 to 546 (*days*)

3 symptom levels:	165 to 203	204 to 239	240 to 294	*3 symptom levels*:	459 to 481	482 to 503	504 to 546
Ranges on the scale:	0.20 to 0.39	0.00 to 0.09	0.10 to 0.19	*Ranges on the scale*:	0.61 to 0.80	0.91 to 1.00	0.81 to 0.90

Figure 7.2 The archetypal values of λ and the associated disease states

In other words, our archetypal patient has a period of depression between days 165-294, a period of mania between days 459-546, and otherwise is normal. The simulation model then adds random noise within the ranges depicted in Figure 7.2 – as depicted in Figure 7.3 below:

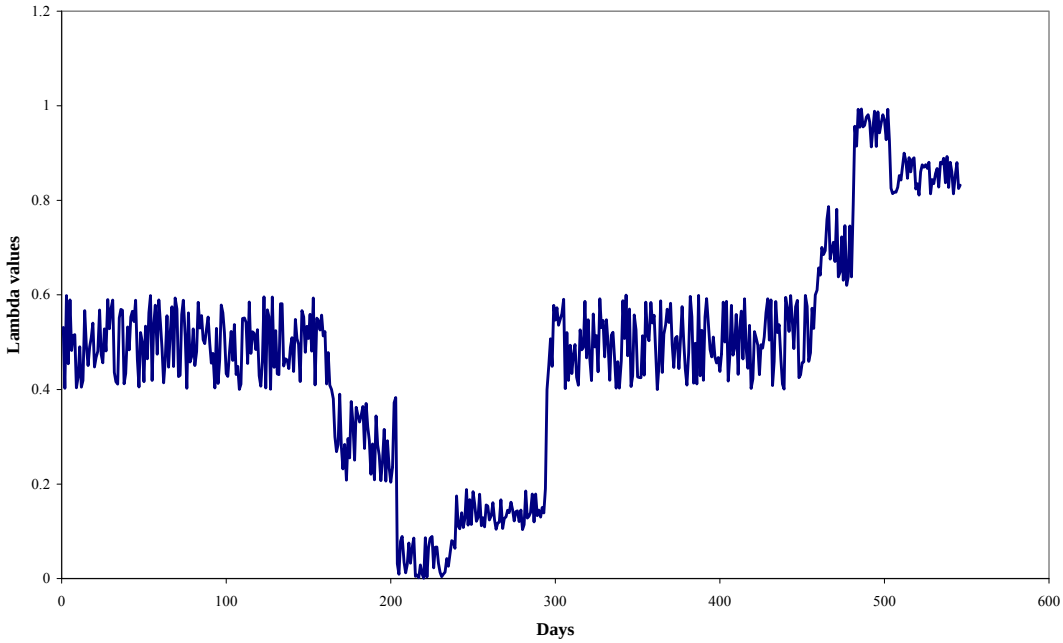


Figure 7.3 A sample trajectory of BD defined by small fluctuations of the parameter λ

The next step is to develop a model for a person's daily activity pattern, based on this natural history model. A key assumption of this model is that an individual's daily activity pattern is a function of two things: *a*) his/her mental health status (as defined by λ) and *b*) a random element totally unrelated to their health. A patient who passes from normal health into depression or mania may do so very gradually, with some good days and some bad days, and of course a healthy individual with no background of mental disorder can experience days on which their mood is either high or low, and their behaviour can reflect this, although they are in no way clinically ill. We have, therefore, constructed a function which maps λ onto a series of observed activities (sleeping, talking, watching TV etc) but also includes a random aspect – e.g. the person's TV may be broken, so they do not bother switching it on even though they are not depressed at all.

This mapping function is actually a slight simplification of the real-life PAM system as it omits part of the data processing that the real PAM performs. The sensors in PAM do not directly monitor a person's observable behaviours, but just collect a vast amount of raw data, such as the sound levels in decibels in a particular room at

10-second snap-shot intervals. By the use of intelligent "feature extraction" algorithms, PAM decodes this raw data into meaningful measures such as the average number of hours the person has slept in a 24-hour period. In our simulation model, we assume that this feature extraction has already occurred and that we can observe meaningful behaviours, which may have clinical significance in terms of bipolar disorder. In other words, the model assumes that we can measure and classify actual behaviours over a 24-hour period, using sensor data. It also assumes that we have calibrated these activity levels for our patient in the normal, manic and depressed states. This is not a restrictive assumption: most BD patients are aware of their own behaviour patterns in all three states. Moreover, in a practical setting the PAM system would be calibrated for a patient's normal activity before use "in earnest".

We have modelled this mapping function as follows. For a given individual, let N, D and M be the average levels of some particular variable (e.g. light levels in lux in the kitchen at 4.00 am) in the normal, extremely depressed and extremely manic mood states, respectively. We have devised the following function of the parameters λ, N, D and M to calculate the value of this variable across all possible mood states:

$$4\lambda(1-\lambda)N + (2\lambda^2 - 3\lambda + 1)D + \lambda\left(\lambda^2 + \frac{\lambda}{2} - \frac{1}{2}\right)M\,, \qquad (7.1)$$

where λ represents the multi-state-scale that ranges from 0 to 1 corresponding to the various mood states. When $\lambda = 0$ then the equation 7.1 yields D (the value when fully depressed), when $\lambda = 0.5$ we get N and when $\lambda = 1$ we get M. Equation 7.1 is similar in structure and in some detail to the expression utilised by Bauer et al (2005). As before, in the Monte Carlo simulation Equation 7.1 is not applied deterministically but is subject to small random variation.

Psychiatrists very commonly use various rating scales such as the Young Mania Rating Scale (Young et al, 1978), the Altman Self-Rating Mania Scale (Altman et al, 1997) and the Bipolar Affective Disorder Dimensional Scale (Craddock et al, 2004), to measure the severity of bipolar symptoms. These scales

rate the severity of mania and depression separately. It is this idea that initially led us to work out two separate equations: $(1 - \lambda)N + \lambda M$ and $(1 - \lambda)N + \lambda D$, in which λ ranges between 0 and 1 ($\lambda = 0$ implies N and $\lambda = 1$ implies either M or D), for our intended purpose. Our intention was to quantify the severity of mania through the former equation and quantify the severity of depression through the latter one.

The idea was virtually feasible, but we required having to create a common interface in which to combine the two aspects of the disorder at the end, which would have been rather time consuming. We then started to research the feasibility of using a single scale to rate the both aspects of the disorder, as proposed by Bauer et al (2005). Consequently, we constructed Equation 7.1 above for our simulation study for the purpose of quantifying the possible current activity across all possible mood states by assigning fully depressed state when $\lambda = 0$, normal state when $\lambda = 0.5$ and fully manic when $\lambda = 1$. Clearly, an individual cannot simultaneously be fully depressed and normal or fully manic and normal. We had to satisfy this by allowing the M and D disappear from the equation 7.1 when $\lambda = 0.5$. In the same way, we let M and N disappear when $\lambda = 0$, and D and N disappear when $\lambda = 1$. We also allowed the estimation of nonlinear effects among these interacting variables.

To make the above idea clearer, we have used a person's sleep pattern as a descriptive example. Suppose that over a 24-hour period, a person sleeps (on average) for 6 hours when they are in good health, for 10 hours when they are very depressed and for 4 hours when they are very manic. We can then use Equation 7.1 to work out the person's potential actual sleep hours over time, across any possible mood states. Figure 7.4 shows the mapping between the mood states and sleep hours and highlights the purpose of the equation even further. Approximate mental state is depicted on the x-axis, colour coded in an obvious way: green denotes normal, blue depressed and red, manic. We can clearly see from Figure 7.4 that sleep hours decrease nonlinearly from N to M as λ increases from 0.5 to 1, while sleep hours increase from N to D as λ decreases from 0.5 to 0. This is evidently in line with the fact that bipolar patients sleep less during mania and sleep more during depression.

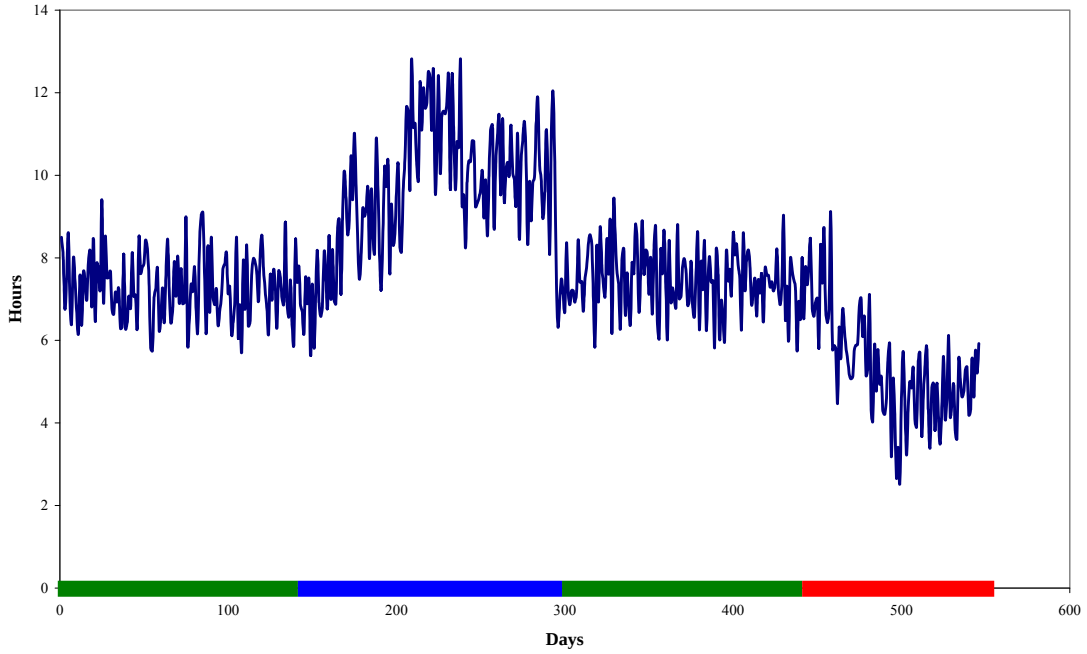


Figure 7.4 Time spent asleep across various days

In Figure 7.5, we have shown λ values on the x-axis and potential time spent in bed on the y-axis, to further show how λ affects sleeping pattern.

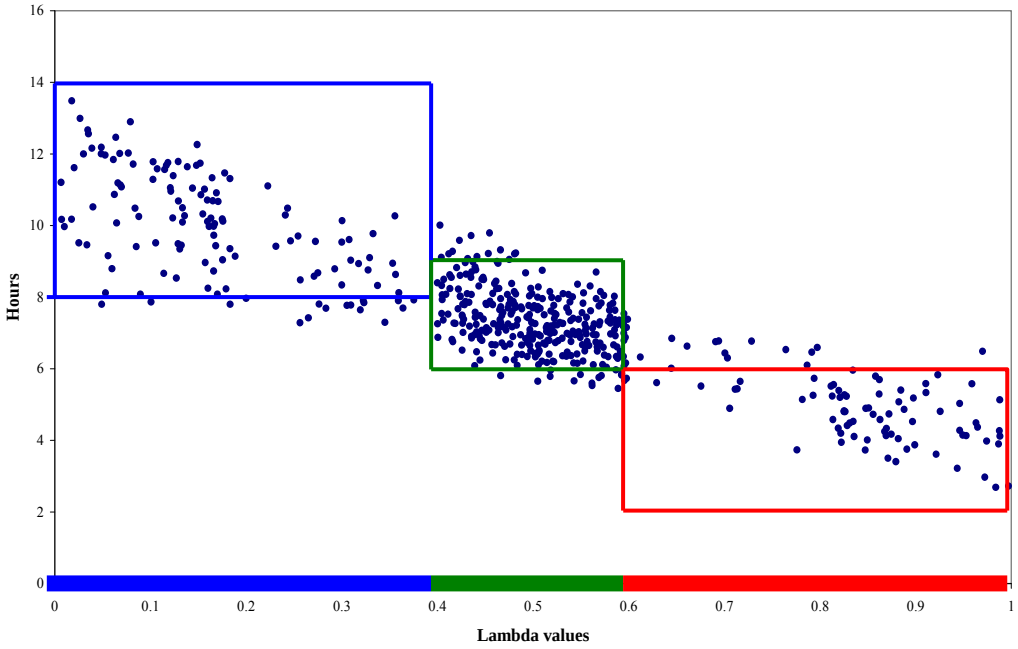


Figure 7.5 Time spent asleep across various mood states

Of course, observations such as disturbed sleep patterns (although a recognised symptom of BD) may naturally vary for reasons totally unrelated to mental health; thus we cannot use Equation 7.1 in a deterministic fashion. Normal healthy people can still find it difficult to sleep at times! In social and behavioural sciences, it is feasible to encompass nonlinear relations among variables (Moosbrugger et al, in press). We, therefore, allowed the estimation of nonlinear effects among interacting variables. In other words, we used *N*, *D* and *M* to obtain combined estimates.

7.4.1 Summary of model assumptions

During the first series of Steering Group meetings, we discussed the details of the model with a psychiatrist who is engaged with bipolar disorder. We have, thus, augmented the literature-based assumptions and parameters with expert opinion. Clearly, like any model this is an over-simplification of reality. We have had to make a number of assumptions and to estimate the values of some parameters. We have performed sensitivity analysis to test how robust the results are to these estimated parameters, so that areas of uncertainty are identified and the effect on the results noted, the level of confidence is increased and (in some cases) areas for further research identified. The following assumptions are taken into account in our analysis:

- it is clinically meaningful to use the parameter λ to represent a person's mental health state
- all patients start in the normal, healthy state at time zero
- the model time horizon is 18 months, since based on clinical data on average a patient will experience one episode of depression and one episode of mania in an 18-month cycle
- all patients experience roughly the same "archetypal" disease trajectory, as illustrated in Figure 7.4. Minor variation is introduced by small random changes

- direct transition from depression to mania is not allowed: all patients must pass (albeit briefly) through the asymptomatic state
- mixed states have not been modelled
- the average times spent in each state were derived from the literature and are as described in Section 7.4
- it is possible to devise a formula as a function of λ to represent a simulated person's daily activities, for a specific limited set of behaviours relevant for bipolar prodromes
- a patient's actual behaviours depend partly on their mental health state (i.e. on the parameter λ) and partly on a random factor
- it is possible to calibrate these activity levels for any patient in the normal, manic and depressed states.

7.4.2 Prodromes, behavioural and technical parameters

In addition to the disease-related parameters discussed above, the inputs to the model also include a selection of the most common bipolar prodromes, together with behavioural parameters and technical parameters relating to the choice of sensors and the reliability and accuracy of the PAM system.

Self-reporting of daily sleep, activity and mood fluctuations is an established clinical tool for the clinician to assess the severity of bipolar disorder (Bauer et al, 1991; Leverich and Post, 1996). The following list contains the most common bipolar prodromes, derived from the clinical literature (WHO, 1992; Morriss, 2004), which we mapped in the model to various observable behaviours:

- Activity levels
- Sleep
- Talkativeness
- Social energy
- Appetite

Other prodromal symptoms are described in the literature but were not included in PAM, either because they are hard to translate into observable activity, or because they are less common. These include 'feeling in another world' and anxiety, which may precede episodes of mania and depression respectively (Morriss, 2004). We also had to exclude another important prodrome – increased or decreased interest in sex – for obvious ethical and privacy reasons!

The following list contains the 14 PAM-observable behaviours, with units of measurement shown in parentheses, which were mapped in the model to the above five prodromes (see Figure 7.6 below):

A. daily activity (PAL). The PAL (Physical Activity Level) is a measurement commonly used to express a person's daily physical activity, and is used to approximate a person's total energy expenditure (UNU, 1994). For example, the PAL for an office worker getting little or no exercise fluctuate between 1.4 and 1.7

B. earliest time person leaves home in the morning (time of day)

C. latest time person gets back home in the evening (time of day)

D. total number of TV remote keypresses (number)

E. total time spent in bed in 24-hour period (hours)

F. average light level between 11pm and 7am (lux)

G. average noise level between 11pm and 7am (decibels)

H. total time spent talking on the telephone (minutes)

I. total number of daily phone calls (number)

J. total time spent outside the home between 5pm and 1am (hours)

K. cupboard doors usage (i.e. the total number of times the doors were opened)

L. fridge doors usage (ditto)

M. microwave door usage (ditto)

N. usual time person cooks the evening meal (time of day).

This choice of observable behaviours was based entirely on the capability of the PAM sensors selected in the real system. Other observable behaviours such as 'talking speed' or 'spending habits' could hypothetically have been considered in the model, but none of the PAM sensors can collect these types of information. Although observable behaviours such as 'time spent talking on the phone' and 'number of daily phone calls' can be used as proxies to indicate whether a person is talking more or less than usual, obviously, "talking speed" will not be captured. This hierarchy of clinical prodromes, observable behaviours and the sensor data is depicted in Figure 7.6 below. The prodromes defined by psychiatrists and cited in the literature are at the top level, with the observable behaviours at the next level down, and the sensor data at the bottom level.

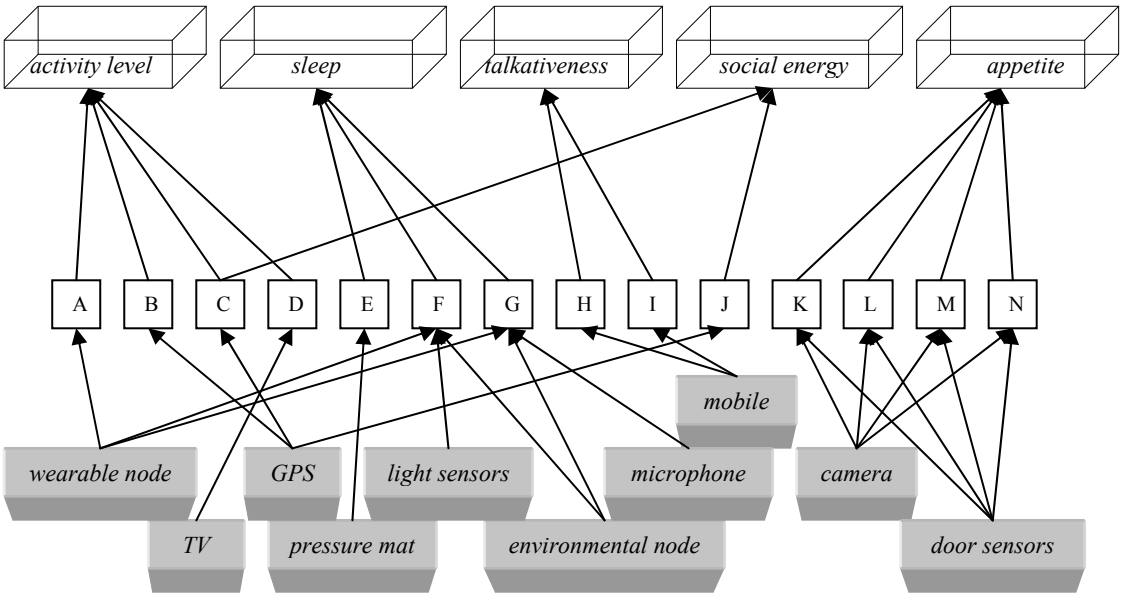


Figure 7.6 Mapping between prodromes, observable behaviours and sensors

Some of the observable behaviours such as 'time spent in bed' and 'daily activity' are generic (i.e. common to all people), while behaviours such as 'earliest time leaving home in the morning' and 'usual time for cooking' are variable depending on a patient's lifestyle, whether they live on their own, go out to work, have an active social life, cook for themselves, etc. We have assumed that our patients live alone, are employed and cook for themselves. Following discussion with the Steering Group, it was felt that the main use of PAM would be for patients who live alone, although in our technical trials of the equipment on members of the research team (all of whom lived with several other people) it was found to be possible to identify some data by individual. By definition, in practice the PAM system would be configured to suit the patient's particular lifestyle.

Table 7.2 depicts this mapping of prodromes indicating a possible change in mental health state onto observable behaviours:

Prodrome	**Observed behaviours**
Activity levels *(increased or decreased)*	Daily activity (A) Earliest time of leaving home (B) Latest time of getting back home (C) TV remote keypresses (D)
Sleep *(increased or decreased)*	Time spent in bed (E) Light level between 11pm to 7am (F) Noise level between 11pm to 7am (G)
Talkativeness *(more or less)*	Time spent talking on the phone (H) Number of daily phone calls (I)
Social energy *(increased or decreased)*	Latest time of getting back home (C) Time not at home between 5pm and 1am (J)
Appetite *(increased or decreased)*	Cupboard doors usage (K) Fridge doors usage (L) Microwave door usage (M) Usual time of cooking (N)

Table 7.2 Mapping between the prodromes and observable behaviours

Different people will have different sets of prodromes that may indicate the onset of an acute episode. In reality, people may have very personal and specific warning signs of an episode, which apply only to them. For example, our trial participant mentioned that she always knew she was about to "go off" (i.e. have an episode of mania) when she realised she was spending a lot of time in her spare bedroom, which normally she did not use. Other patients may know from their own personal experience that simultaneous changes in several different behaviours can indicate the onset of an episode. A patient may know that changes in both 'activity level' and 'sleep' mean that he/she is going to have an episode. Obviously, if this patient is not willing to have any sensors which monitor the observable behaviours of 'activity level' and 'sleep', then PAM will not work for him/her.

Patients will also differ in how they make the choice. Some people may object to a particular sensor rather than the activity it is intended to monitor. Figure 7.6 above shows that there are several ways in which a specific prodrome can be monitored. For example, sleep patterns could be monitored by a pressure mat placed in the bed which detects the presence or absence of a person in the bed, or by a pressure mat placed on the floor by the bed, or by light and/or sound levels in the bedroom. A patient might object to the pressure mat in the bed but be willing to have it on the floor. He may object to the sound level sensor but be happy about the light sensor (or vice versa). Another patient might object to having his sleep habits monitored at all.

The model considers 25 different patient types, defined on the basis of the prodromes they were willing to be monitored on rather that the individual sensors they were willing to use. This was a pragmatic choice since the potential number of combinations of sensors and different locations within a person's home is astronomically large. The 25 selected combinations of the selected prodromes are given below. Although the prodromes used in this research were selected on the basis of the clinical literature, this is obviously by no means an exhaustive set. However, the 25 patient types listed here are more than sufficient for the purposes of our

analysis. For each patient type, a different set of sensors were required to monitor each choice of prodromes, as shown in Appendix A11.

Patient type 1: *activity level + sleep*

Patient type 2: *activity level + talkativeness*

Patient type 3: *activity level + social energy*

Patient type 4: *activity level + appetite*

Patient type 5: *sleep + talkativeness*

Patient type 6: *sleep + social energy*

Patient type 7: *sleep + appetite*

Patient type 8: *talkativeness + social energy*

Patient type 9: *talkativeness + appetite*

Patient type 10: *social energy + appetite*

Patient type 11: *activity level + sleep + talkativeness*

Patient type 12: *activity level + sleep + social energy*

Patient type 13: *activity level + sleep + appetite*

Patient type 14: *activity level + talkativeness + social energy*

Patient type 15: *activity level + talkativeness + appetite*

Patient type 16: *activity level + social energy + appetite*

Patient type 17: *sleep + talkativeness + social energy*

Patient type 18: *sleep + talkativeness + appetite*

Patient type 19: *talkativeness + social energy + appetite*

Patient type 20: *activity level + sleep + talkativeness + social energy*

Patient type 21: *activity level + sleep + talkativeness + appetite*

Patient type 22: *activity level + sleep + social energy + appetite*

Patient type 23: *activity level + talkativeness + social energy + appetite*

Patient type 24: *sleep + talkativeness + social energy + appetite*

Patient type 25: *activity level + sleep + talkativeness + social energy + appetite*.

7.4.3 Incorporating individual variability and uncertainty

Equation 7.1 was used to generate the person's actual behaviours. The measurements are stochastic, based partly on the parameter λ and partly on the natural variability associated with the observed behaviours, just as in real life. The parameter λ has been generated randomly based on the "natural history" of the disease progression, as explained in Figure 7.2 above. The λ values were generated on the basis of the information provided in Table 7.3:

Day	Clinical state	Value of λ	Random variation
1 to 164	Normal health	0.40 – 0.60	RAND() × (0.6 – 0.4) + 0.4
165 to 203	Mild depression	0.20 – 0.39	RAND() × (0.39 – 0.2) + 0.2
204 to 239	Severe depression	0.00 – 0.09	RAND() × (0.09 – 0) + 0
240 to 294	Moderate depression	0.10 – 0.19	RAND() × (0.19 – 0.1) + 0.1
295 to 458	Normal health	0.40 – 0.60	RAND() × (0.6 – 0.4) + 0.4
459 to 481	Mild (hypo) mania	0.61 – 0.80	RAND() × (0.8 – 0.61) + 0.61
482 to 503	Severe mania	0.91 – 1.00	RAND() × (1 – 0.91) + 0.91
504 to 546	Moderate mania	0.81 – 0.90	RAND() × (0.9 – 0.81) + 0.81

Table 7.3 The archetypal λ values at various disease states

Thus, values of λ between 0.40 and 0.60 were generated randomly for days 1 - 164 and days 295 - 458, when the patient was in normal health. Similarly, for days 165 to 203, when the patient was mildly depressed, values of λ were generated randomly between 0.20 and 0.39; and between days 204-239, when the patient was severely depressed, values of λ were generated randomly between 0.0 and 0.19. Exactly the same approach was applied for mild and severe mania. In any of the "gross" states, e.g. mild depression, a person's mood and mental health status will not steadily decline, improve or stay static over the whole period: there are small daily fluctuations, and the patient will experience better days and worse days. This is also true for a healthy person in real life: everyone has good days and bad days within a "normal" spectrum.

Therefore, the parameter λ was allowed to fluctuate subtly and randomly over time (but always within the broad range for that gross state) with a daily time-step to represent the different clinical states. This daily random variation was achieved non-linearly through incorporating the random product *RAND*(), which yields real numbers between 0 and 1.

The random element of each behaviour, i.e. the part not dependent on mental health state but simply due to daily variability, was modelled by fitting triangular probability distributions. Clearly we did not have any empirical data to which to fit these distributions, other than common sense, practical experience and some clinical input. The triangular distribution was chosen as it is simple to parameterise and is widely used as a subjective description of a population for which there is only limited sample data, especially in cases where the relationship between variables is known but data is scarce. For example, if a person in the healthy state spends time in bed on average 7 hours a day ($N = 7$ in the terminology of Equation 7.1), a suitable triangular distribution might have min 6, mode 7 and max 9.

The next task then, was to specify these probability distributions for different simulated patients and all the behavioural observations (see Table 7.4 below for full set). This table is also provided in Appendix A3, which contains the list of references that the data were collected from.

(AA = Actual Activity)

PAM observed behaviour	When "normal" (*N*)	When "manic" (*M*)	When "depressed" (*D*)	Threshold level
Daily activity (PAL)	RiskTriang(1.7,1.9,2.0)	RiskTriang(1.9,2.2,2.5)	RiskTriang(1.0,1.4,1.7)	if AA > 2.0 OR AA < 1.7, then 1
Earliest time of leaving home (time)	RiskTriang(7,8,9)	RiskTriang(5,6,7)	RiskTriang(8,9,11)	if AA < 7 OR AA > 9, then 1
Latest time of getting back home (time)	RiskTriang(18,19,21)	RiskTriang(20,23,24)	RiskTriang(16,17,19)	if AA > 21 OR AA < 18, then 1
TV remote keypresses (number)	RiskTriang(20,40,60)	RiskTriang(40,75,100)	RiskTriang(0,15,25)	if AA > 60 OR AA < 20, then 1
Time spent in bed (hour)	RiskTriang(6,7,9)	RiskTriang(2,4,6)	RiskTriang(8,12,14)	if AA < 6 OR AA > 9, then 1
Light level between 11pm to 7am (lux)	RiskTriang(5,10,20)	RiskTriang(20,40,70)	RiskTriang(1,4,10)	if AA > 20 OR AA < 5, then 1
Noise level between 11pm to 7am (dB)	RiskTriang(15,20,25)	RiskTriang(25,30,40)	RiskTriang(5,10,20)	if AA > 25 OR AA < 15, then 1
Time spent talking on the phone (minute)	RiskTriang(10,25,50)	RiskTriang(40,70,120)	RiskTriang(0,10,20)	if AA > 50 OR AA < 10, then 1
Number of daily phone calls (number)	RiskTriang(2,5,7)	RiskTriang(7,10,16)	RiskTriang(0,1,3)	if AA > 7 OR AA < 2, then 1
Time being outside btwn 5pm & 1am (hour)	RiskTriang(1,2,4)	RiskTriang(3,5,8)	RiskTriang(0,0.5,1)	if AA > 3 OR AA < 1, then 1
Cupboard doors usage (number)	RiskTriang(8,10,14)	RiskTriang(12,18,22)	RiskTriang(2,4,10)	if AA > 14 OR AA < 8, then 1
Fridge doors usage (number)	RiskTriang(6,8,10)	RiskTriang(10,14,18)	RiskTriang(2,4,6)	if AA > 10 OR AA < 6, then 1
Microwave door usage (number)	RiskTriang(4,6,8)	RiskTriang(6,10,14)	RiskTriang(0,2,4)	if AA > 8 OR AA < 4, then 1
Usual time of cooking (time)	RiskTriang(18,19,21)	RiskTriang(20,22,24)	RiskTriang(16,18,19)	if AA > 20 OR AA < 18, then 1

Table 7.4 Dataset 1

Of course, the major source of uncertainty is the functionality of the PAM system itself. Indeed this was the prime motivation for the research in this thesis. Ambient data collection is inherently unreliable. The sensors may malfunction or break down completely, there may be a power loss, the patient may accidentally (or deliberately) switch off the PC, or simply forget to recharge the wearable device or the mobile phone. The patient may lose the wearable device or the mobile – our trial participant once accidently left the wearable in a friend's car. The patient may damage, lose or switch off any of the sensors (or the whole system) either accidentally or deliberately. There may be software problems, local or remote, with the PC and the onward transmission of data.

In these circumstances, PAM may report a change in behaviour which has not taken place (a false positive) or miss a change which has taken place (a false

negative). Both of these are undesirable: clearly failing to issue an alert if a genuine change in mental health state has occurred would render the whole PAM system pointless, but on the other hand if the system keeps issuing alerts when nothing is wrong then the patient will quickly become disillusioned with PAM and will stop using it. Data errors caused by technical malfunction were modelled by randomly modifying the relevant observed behavioural parameter upwards or downwards by an amount based on a combination of suggestions from the other members of the PAM team, and common sense, as shown in Table 7.5:

Observed behavioural activity (unit)	***Technical variation* (±)**
Daily activity (PAL)	0.2
Earliest time of leaving home (time of day)	0.5
Latest time of getting back home (time of day)	0.5
TV remote keypresses (number)	10
Time spent in bed (hours)	0.5
Light level between 11pm to 7am (lux)	2
Noise level between 11pm to 7am (decibels)	2
Time spent talking on the phone (minutes)	5
Number of daily phone calls (number)	1
Time being outside between 5pm and 1am (hours)	0.5
Cupboard doors usage (ditto)	2
Fridge doors usage (ditto)	2
Microwave door usage (ditto)	2
Usual time of cooking (time of day)	0.5

Table 7.5 Modelling technical variation

The data shown in the above table was then incorporated in the model using the function: *Uniform((actual activity – error), (actual activity + error))*. To give an illustrative example, the sampled value of 'Time spent in bed' was varied uniformly by ±0.5 hours (i.e. ±30 minutes). Thus, if the actual daily value for 'Time spent in bed' was 6 hours, then the PAM-detected corresponding value will be a randomly chosen value between 5.5 and 6.5 hours. In this case, we of course do not know if the value 0.5 is scientifically legitimate or not, but it is plausible, given the huge scope for error in the real PAM system. The simulation runs incorporated these plausible random effects to sample the corresponding actual values. Although the choice of these error margins was subjective and the values illustrative, it does not affect the underlying credibility of the model. Of course, we may not have dealt with

this problem justifiably, but we also considered a greater range of variability in the PAM-detected estimates to see the effects of more serious technical malfunctions.

7.4.4 Decision rules and threshold levels

The next step was to define the decision rules for identifying whether a significant change in behaviour had occurred so that PAM would issue an alert (i.e. a text message) to the patient. Although there was some guidance on this in the literature, as in the case of the behaviours the main aim was to produce rules that were credible and practicable. The decision rules and threshold levels were chosen using ideas from the literature together with common-sense judgement, in order to address the need for timely and accurate evaluation of bipolar relapses. Morriss (2004) used the occurrence of at least four out of a total of six prodromes to define a danger level of relapse, with two or three as indicating a warning level. We adopted a similar approach, assuming that the simultaneous presence of any combination of two or more prodromal symptoms (e.g. 'sleep' + 'talkativeness', 'activity level' + 'appetite', 'activity level' + 'sleep' + 'social energy', or etc) may trigger an alert.

However, we also assumed that not all the corresponding observed behaviours need to occur in order to indicate a prodrome. For example, when a person's 'activity level' is identified as high, then clearly that person could be highly active at home, or outside the home. However, it is impossible for the person to be highly active in two places at once! Similarly, the time a person spends outside the home is not just associated with that person's 'activity level', but also with 'sleep' and 'social energy'. The existence of any two or more prodromes may be sufficient to indicate a potential relapse, and thus we set a certain number of observed behaviours to be occurred at a time to imitate its associated prodromes (see Appendix A2).

We assigned a value of 1 (= 'yes') when an observed behaviour exceeded its specified threshold levels, and 0 (= 'no') otherwise. Hence, the PAM scoring system ranged from 0 to 14 since there are 14 observable behaviours. Hirschfeld et al (2000) used a similar type of scoring system in developing the Mood Disorder

Questionnaire. To be screened positive for a potential relapse, it is not mandatory to score the maximum 14 points. Different values were tested in the simulation.

The question remains how long a person should persist with the prodromal symptoms before receiving an alert, to minimise the number of false alerts. Again we used information from the clinical literature to guide our choice. Keane (2010) reported that a manic patient hardly slept for 4 successive days. In our first set of tests, PAM sent an alert if the prodromal symptoms persisted for 3 out of 5 successive days. Table 7.6 (where A, B, …, G represent the observed behaviours) shows a descriptive example of how such a decision rule would work for a hypothetical patient who had chosen to be monitored on 'activity level' and 'sleep' patterns:

Day	A	B	C	D	E	F	G	If A+B+C+D >= 2 and E+F+G >=2	PAM alert
1	0	0	1	0	0	0	1	*No*	*no*
2	1	0	0	1	0	1	0	*No*	*no*
3	0	1	0	1	1	0	1	*Yes*	*no*
4	1	0	1	0	1	1	0	*Yes*	*no*
5	0	1	1	1	0	0	1	*No*	*no*
6	1	1	0	0	1	1	1	*Yes*	*yes*
7	1	1	1	1	1	1	1	*Yes*	*yes*

Table 7.6 An example of how the decision rules work

Of course, the form of the decision rules and the threshold levels for deciding whether PAM should alert a patient or his/her carer may be too sensitive or not sensitive enough. This was the whole reason for undertaking the simulation modelling, to test out values and settings for these rules in an artificial environment on synthetic patients.

7.5 The computer model

The model was implemented in Microsoft Excel using the risk analysis add-in "@Risk" software (Palisade, 2008), which offers a wide choice of input probability distributions to test various monitoring scenarios. Our computer model is based on the theoretical framework of stochastic process to predict the risk of any outcome arising from the underlying multi-state process with the inclusion of individual traits.

Basically the model is driven by the mental health status (as determined by the parameter λ) of a single patient, varying stochastically over an 18-month time horizon. Each row in the spreadsheet represents one day, and each iteration of the model corresponds to one potential realisation of that patient's mental health trajectory over 18 months. Each day, the λ is used to generate a set of behaviours, and then according to the set of sensors that that person has chosen, that behaviour may (or may not) be observed by PAM, as described in Section 7.4.3. Finally, the chosen decision rule is applied for that day and an alert issued (or not). Since the true value of λ is known in the model, it is thus possible to determine whether an alert (if issued) was a true positive or a false positive, and otherwise, if no alert was issued, whether this was a true negative or a false negative.

7.5.1 Excel implementation

The model was implemented in Excel and coded in Visual Basic for Applications (see Appendices A.6.1, A.6.2 and A.6.3), which can be viewed using the memory stick enclosed in Appendix A8, but the "@Risk" software is required to view and run the model. It is driven by the *UserInterface* worksheet (see Figure 7.7), in which the various buttons correspond to the observed behaviours, and are used to store patient-specific data within specific locations of the *Data* worksheet, i.e. the values of *N*, *D* and *M* for each behaviour, as described in Section 7.4 (see Figure 7.8 below).

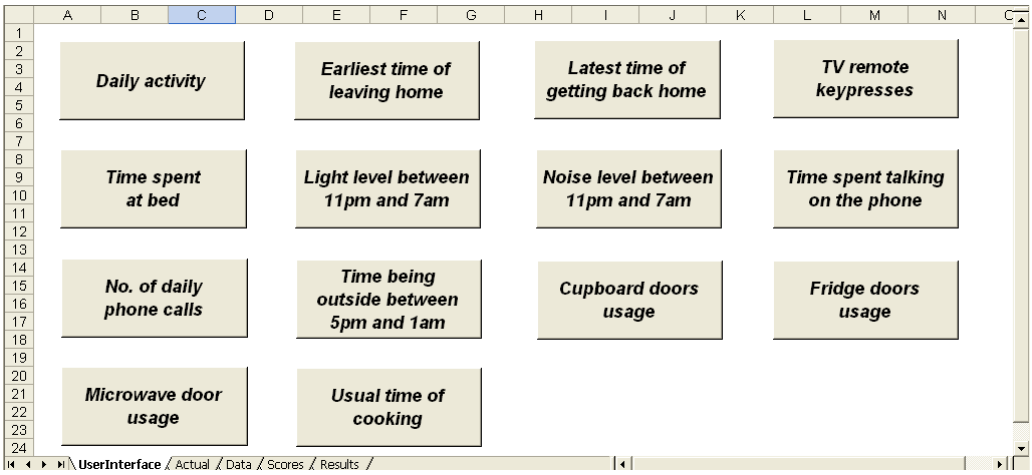
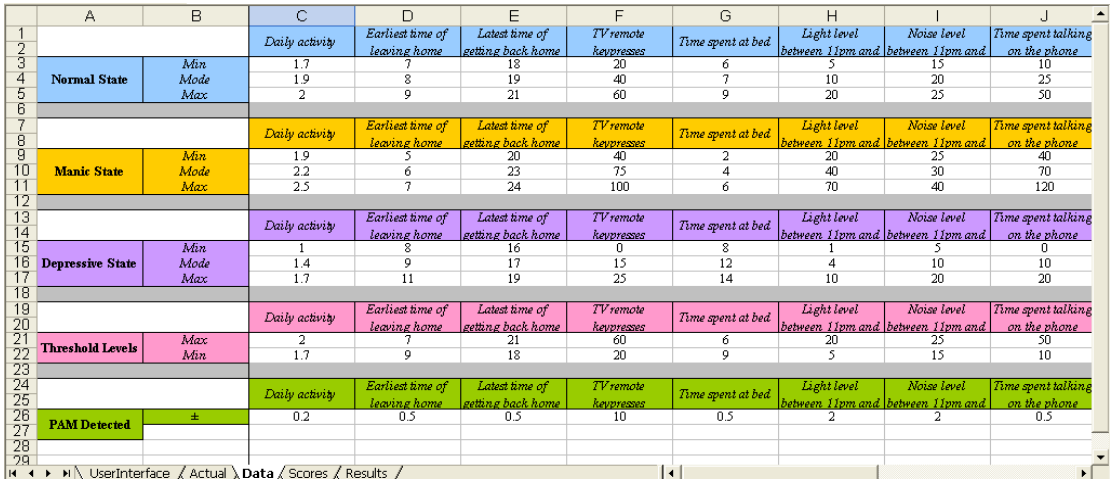


Figure 7.7 Screenshot of UserInterface worksheet

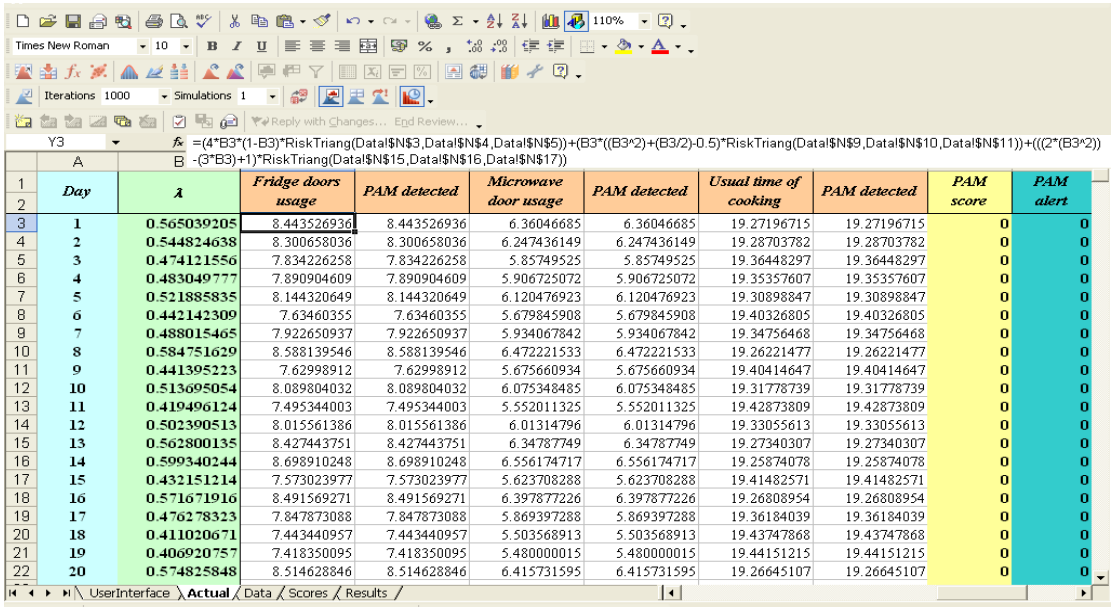


		Daily activity	Earliest time of leaving home	Latest time of getting back home	TV remote keypresses	Time spent at bed	Light level between 11pm and	Noise level between 11pm and	Time spent talking on the phone
Normal State	Min	1.7	7	18	20	6	5	15	10
	Mode	1.9	8	19	40	7	10	20	25
	Max	2	9	21	60	9	20	25	50
		Daily activity	Earliest time of leaving home	Latest time of getting back home	TV remote keypresses	Time spent at bed	Light level between 11pm and	Noise level between 11pm and	Time spent talking on the phone
Manic State	Min	1.9	5	20	40	2	20	25	40
	Mode	2.2	6	23	75	4	40	30	70
	Max	2.5	7	24	100	6	70	40	120
		Daily activity	Earliest time of leaving home	Latest time of getting back home	TV remote keypresses	Time spent at bed	Light level between 11pm and	Noise level between 11pm and	Time spent talking on the phone
Depressive State	Min	1	8	16	0	8	1	5	0
	Mode	1.4	9	17	15	12	4	10	10
	Max	1.7	11	19	25	14	10	20	20
		Daily activity	Earliest time of leaving home	Latest time of getting back home	TV remote keypresses	Time spent at bed	Light level between 11pm and	Noise level between 11pm and	Time spent talking on the phone
Threshold Levels	Max	2	7	21	60	6	20	25	50
	Min	1.7	9	18	20	9	5	15	10
		Daily activity	Earliest time of leaving home	Latest time of getting back home	TV remote keypresses	Time spent at bed	Light level between 11pm and	Noise level between 11pm and	Time spent talking on the phone
PAM Detected	±	0.2	0.5	0.5	10	0.5	2	2	0.5

UserInterface / Actual / Data / Scores / Results /

Figure 7.8 Screenshot of Data worksheet

At runtime, the *Actual* worksheet (see Figure 7.9) reads in the data from the *Data* worksheet and (using the @Risk functionality) samples that day's actual value for each behaviour from the appropriate triangular distribution. Next, @Risk is used to determine whether that actual activity was detected by PAM. Another column in the *Actual* worksheet is used to represent the PAM-detected behaviour. Finally, the model checks the decision rules and compares the person's PAM-observed behaviour with his/her normal behaviour to check if the threshold levels are exceeded. The *Scores* worksheet is used to calculate the daily PAM score, which is of course based on the threshold levels.



Y3 =(4*B3*(1-B3)*RiskTriang(Data!$N$3,Data!$N$4,Data!$N$5))+(B3*((B3^2)+(B3/2)-0.5)*RiskTriang(Data!$N$9,Data!$N$10,Data!$N$11))+(((2*(B3^2))-(3*B3)+1)*RiskTriang(Data!$N$15,Data!$N$16,Data!$N$17))

Day	λ	Fridge doors usage	PAM detected	Microwave door usage	PAM detected	Usual time of cooking	PAM detected	PAM score	PAM alert
1	0.565039205	8.443526936	8.443526936	6.36046685	6.36046685	19.27196715	19.27196715	0	0
2	0.544824638	8.300658036	8.300658036	6.247436149	6.247436149	19.28703782	19.28703782	0	0
3	0.474121556	7.834226258	7.834226258	5.85749525	5.85749525	19.36448297	19.36448297	0	0
4	0.483049777	7.890904609	7.890904609	5.906725072	5.906725072	19.35357607	19.35357607	0	0
5	0.521885835	8.144320649	8.144320649	6.120476923	6.120476923	19.30898847	19.30898847	0	0
6	0.442142309	7.63460355	7.63460355	5.679845908	5.679845908	19.40326805	19.40326805	0	0
7	0.488015465	7.922650937	7.922650937	5.934067842	5.934067842	19.34756468	19.34756468	0	0
8	0.584751629	8.588139546	8.588139546	6.472221533	6.472221533	19.26221477	19.26221477	0	0
9	0.441395223	7.62998912	7.62998912	5.675660934	5.675660934	19.40414647	19.40414647	0	0
10	0.513695054	8.089804032	8.089804032	6.075348485	6.075348485	19.31778739	19.31778739	0	0
11	0.419496124	7.495344003	7.495344003	5.552011325	5.552011325	19.42873809	19.42873809	0	0
12	0.502390513	8.015561386	8.015561386	6.01314796	6.01314796	19.33055613	19.33055613	0	0
13	0.562800135	8.427443751	8.427443751	6.34787749	6.34787749	19.27340307	19.27340307	0	0
14	0.599340244	8.698910248	8.698910248	6.556174717	6.556174717	19.25874078	19.25874078	0	0
15	0.432151214	7.573023977	7.573023977	5.623708288	5.623708288	19.41482571	19.41482571	0	0
16	0.571671916	8.491569271	8.491569271	6.397877226	6.397877226	19.26808954	19.26808954	0	0
17	0.476278323	7.847873088	7.847873088	5.869397288	5.869397288	19.36184039	19.36184039	0	0
18	0.411020671	7.443440957	7.443440957	5.503568913	5.503568913	19.43747868	19.43747868	0	0
19	0.406920757	7.418350095	7.418350095	5.480000015	5.480000015	19.44151215	19.44151215	0	0
20	0.574825848	8.514628846	8.514628846	6.415731595	6.415731595	19.26645107	19.26645107	0	0

UserInterface / Actual / Data / Scores / Results /

Figure 7.9 Screenshot of Actual worksheet

The numbers of true/false positive alerts and true/false negatives are recorded in the *Results* worksheet (see Figure 7.10), and form the primary output of the model, enabling the performance of the PAM system to be evaluated for that particular combination of patient, behaviours, sensors and decision rules. The model also outputs the number of days before the first depressive alert (denoted by ODE) and the number of days before the first manic alert (OME). Again, since the true (approximate date) of onset of an episode is known, we can determine how long PAM takes to detect this, and this was achieved through VBA coding (see Appendix A6.4). On average, the model took about 20 minutes to run each combination of prodromes on a PC with 1.18 GHz and 1.99 GB of RAM.

	A	B	C	D	E	F
1	***No. of days PAM took to detect the onset of depressive episode (ODE)***	***No. of days PAM took to detect the onset of manic episode (OME)***	***True Positives (TP)***	***False Positives (FP)***	***True Negatives (TN)***	***False Negatives (FN)***
2						
3	5.90	2.64	190.71	6.94	321.06	27.29

Figure 7.10 Screenshot of Results worksheet

8. Experimentation and Results

In this chapter, we first show that our OR model is valid and suitable for its intended purpose. We then describe all the different scenarios that were run for various hypothetical patient types in terms of their various preferences for sensors and their 'activity signatures'.

8.1 Model validation

We have used a continuous-time multiple state model, because it allowed us to illustrate the underlying process comprehensively. We estimated a large number of transition intensities, for which satisfactory data did not exist. We attempted to properly verify the computer simulation model by incorporating the comprehensive understanding of the actual operation to ensure that it is fit for purpose. We have performed a validation study in order to make sure that the model works well. In conducting the analyses, we essentially concentrated on some widely applied validation characteristics, including accuracy, repeatability, uncertainty, nonlinearity, range and clinical robustness. We considered questions such as did we proceed in the right direction, do the results make sense and were our scenarios appropriate?

The model plots graphs for all the observed behaviours. In order to do this, it was necessary to devise a quick way to measure all the PAM-observed behaviours for each day. Basically, we sampled the normal (N), manic (M) and depressive (D) data from triangular distribution for each day, and used these along with the value of λ into Equation 7.1 (see Page 103), which was then used to measure the actual behaviours during various mood states. An example of this is shown in the Excel function bar (see Appendix A10). Figures 8.1 to 8.4 below show the viability of using the modelling framework. The graphs for the ten remaining behaviours are included in Appendix A9. In these diagrams, approximate mental state is colour-coded on the x-axis. The normal state is shown in green, the depressed state in blue and the manic

state in red. Of course, in reality the transition from one state to the other is normally gradual, and not as abrupt as this colour-coding suggests.

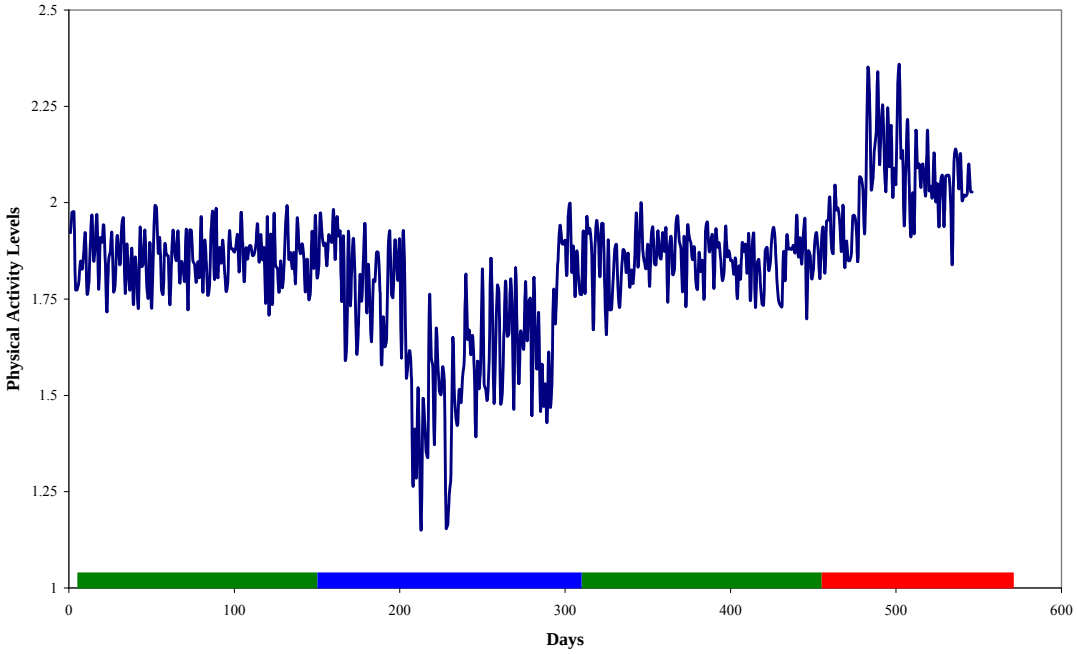


Figure 8.1 Physical activity levels during various mood states

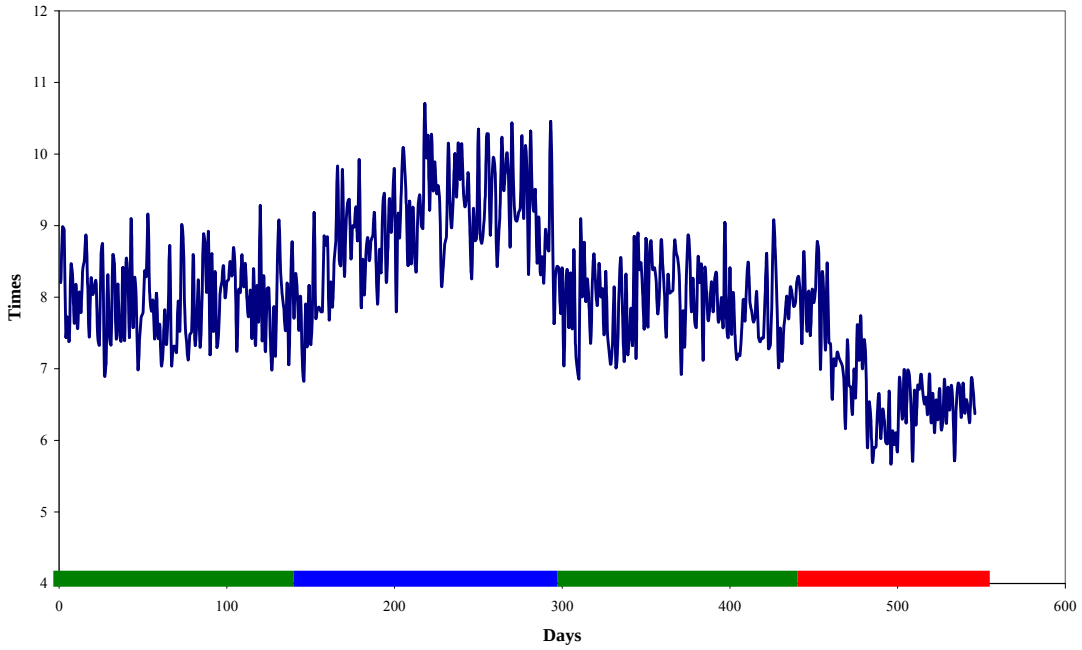


Figure 8.2 Earliest time of leaving home in the morning, during various mood states

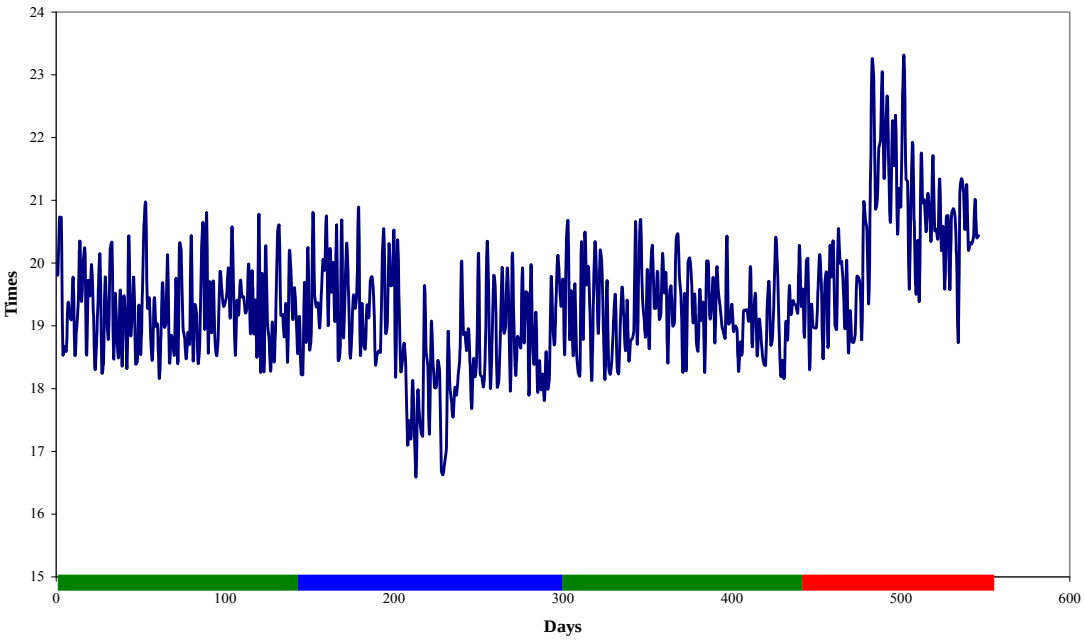


Figure 8.3 Latest time of getting back home during various mood states

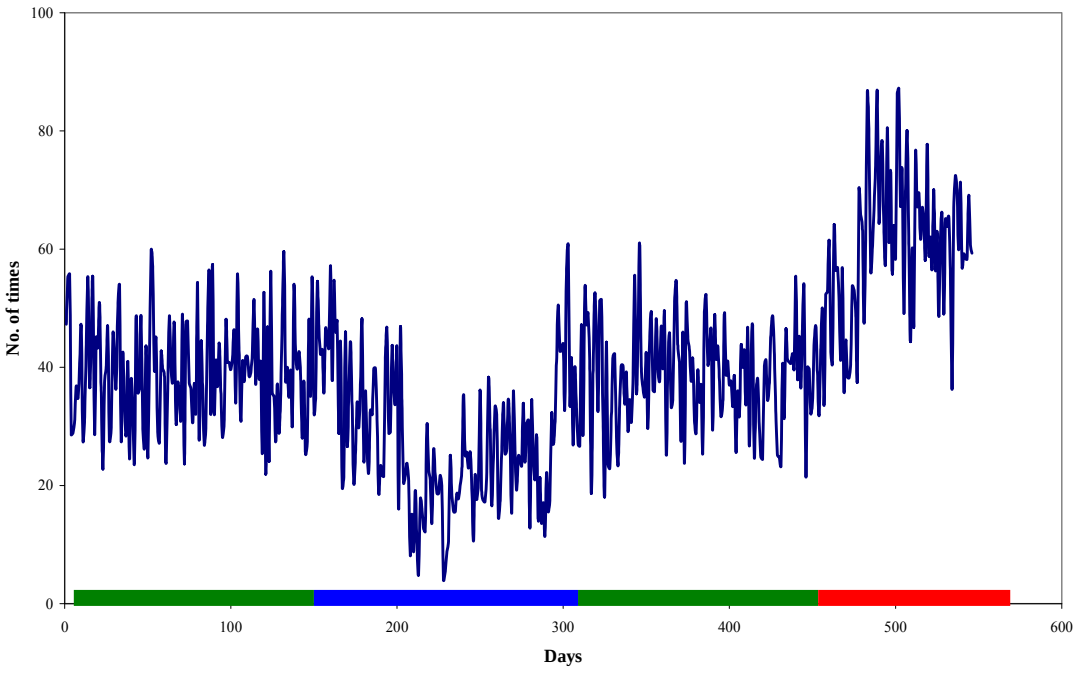


Figure 8.4 Daily no. of TV remote keypresses during various mood states

It can be seen that all the daily behavioural activity patterns illustrated in Figures 8.1 – 8.4 are in line with common sense and clinical experience. There is clearly daily fluctuation around an average value, but distinct changes can be seen correlated to mental health status in a logical way. For example, Figure 8.1 shows that the model has calculated the physical activity levels of bipolar individuals to be high during mania and to be low during depression, compared with their normal activity levels. Similarly, Figures A9.4 and A9.5 both show that the model has correctly indicated that bipolar individuals will talk more during mania and talk less during depression. These findings are in line with the clinical literature such as Lam et al (2001) and Morriss (2004). It can also be seen that the evidence of subtle symptoms change at every stage of the illness suggests that the moods are clearly never constant over a period of time.

We have shown that the model created is actually a very good representation of the reality. In order to model the potential technological uncertainties of the PAM system, random variation was added to the actual observations. An illustrative example of which has been shown in Figure 8.5:

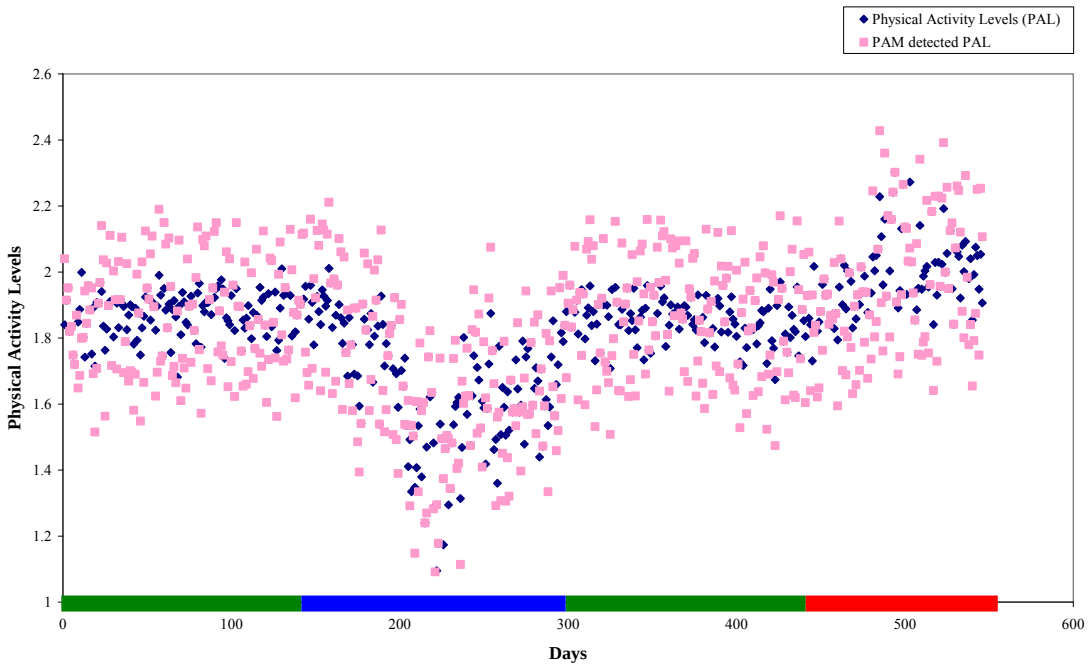


Figure 8.5 PAM detected physical activity levels during various mood states

8.2 Model outputs

To assess the usefulness of PAM, two datasets were used, representing different patient groups (roughly) corresponding to bipolar I disorder. *Dataset 1* (see Appendix A3) relates to bipolar people who typically show marked ups and downs during mania and depression with a minimal overlap with normality, whereas *Dataset 2* (see Appendix A4) contains data that overlap noticeably with normality. Intuitively, one would expect PAM to work better for *Dataset 1* than for *Dataset 2*. For both cases, the model was run for 1,000 iterations for each of the 25 hypothetical patient types, thus simulating the disease trajectories of 1,000 different patients of each type.

Each patient type was defined by the combinations of prodromes they selected, as described in Chapter 7 (see Page 110 and 111). Patient types 1 to 10 chose a selection of two different prodromes, patient types 11 to 19 chose a selection of three different prodromes, patient types 20 to 24 chose a selection of four different prodromes, and patient type 25 chose all five prodromes.

8.2.1 Running the simulation model with Dataset 1

The model was first run using the decision rules that if the selected prodromes were observed by PAM for 3 out of 5 successive days, then an alert would be sent. These were called *Decision rules 1*. An illustrative example of these rules was previously provided in Table 7.5 of Chapter 7 (see Page 113). The model computed the following metrics, which were all averaged over 1,000 iterations:

- TP (true positive count): the average number of days when patients were unwell, i.e. had mild, moderate or severe mania/depression, and PAM correctly sent an alert
- FP (false positive count): the average number of days when patients were in normal good health yet PAM sent an alert
- FN (false negative count): the average number of days when patients were unwell yet PAM did nothing

- TN (true negative count): the average number of days when patients were in normal health and PAM did nothing.

In addition, the model computed the average number of days that PAM took to detect the onset of a depressive episode (denoted ODE) and the average number of days that PAM took to detect the onset of a manic episode (OME). Table 8.1 shows the summary outputs, while Appendix A5.1 shows more detailed outputs:

Patient type	ODE	OME	TP	FP	FN	TN
Type 1	30.72	15.58	143.03	0.88	74.97	327.12
Type 2	40.26	18.67	102.87	0.18	115.13	327.82
Type 3	41.63	19.86	97.34	0.10	120.66	327.91
Type 4	33.23	21.14	124.08	0.86	93.93	327.14
Type 5	37.92	11.18	116.96	0.25	101.04	327.75
Type 6	38.40	07.53	125.90	0.32	92.10	327.68
Type 7	26.29	14.96	144.25	0.84	73.75	327.16
Type 8	42.61	13.31	97.64	0.05	120.36	327.95
Type 9	41.06	19.83	101.97	0.11	116.03	327.89
Type 10	40.62	17.86	107.89	0.22	110.11	327.78
Type 11	18.82	06.33	165.66	1.47	52.34	326.54
Type 12	20.83	04.81	166.60	1.47	51.40	326.53
Type 13	10.99	05.63	171.10	2.30	46.90	325.70
Type 14	33.35	07.15	137.02	0.67	80.98	327.33
Type 15	21.08	09.71	153.48	1.83	64.52	326.17
Type 16	32.84	20.85	124.88	0.81	93.12	327.19
Type 17	27.31	04.28	157.05	1.05	60.95	326.95
Type 18	16.73	06.54	165.20	1.39	52.80	326.62
Type 19	33.03	07.06	140.56	1.00	77.44	327.24
Type 20	08.10	03.12	185.77	2.79	32.23	325.21
Type 21	08.37	04.01	182.36	3.45	35.64	324.55
Type 22	08.17	03.54	183.47	4.26	34.53	323.74
Type 23	15.97	04.70	168.37	3.90	49.63	324.10
Type 24	11.14	03.43	179.93	3.02	38.07	324.98
Type 25	05.90	02.64	190.71	6.94	27.29	321.06

Table 8.1 Performance of PAM for all the different selections of prodromes (Dataset 1)

The performance of PAM for all 25 hypothetical patient types and their various prodromal choices is represented graphically in Figure 8.6 below. Clearly, the average TP is fairly low for patients who selected different combination of two prodromes only (see Figures 8.6 and 8.7), but considerably higher for those patients who chose three or more prodromes (see Figures 8.6, 8.8 and 8.9). As one would expect, as the number of prodromal choices increases, the average ODE and OME

correspondingly reduces (see Figures 8.7, 8.8 and 8.9). However, as the number of prodromal choices increases, so does the average FP (see Figure 8.6).

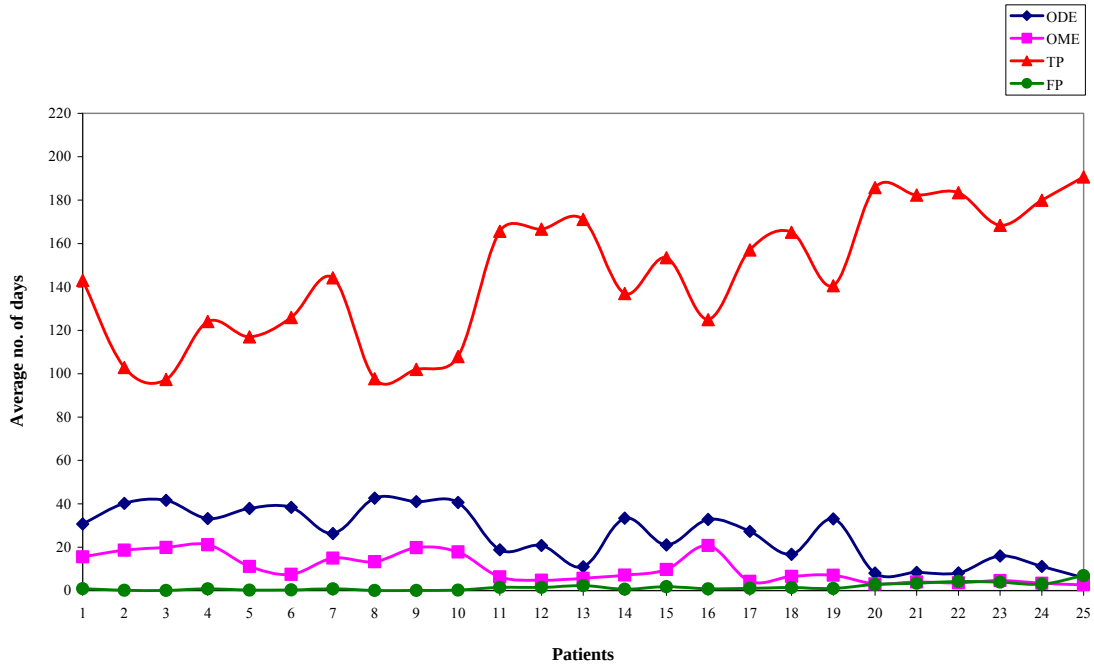


Figure 8.6 Performance of PAM for various selections of prodromes (Dataset 1)

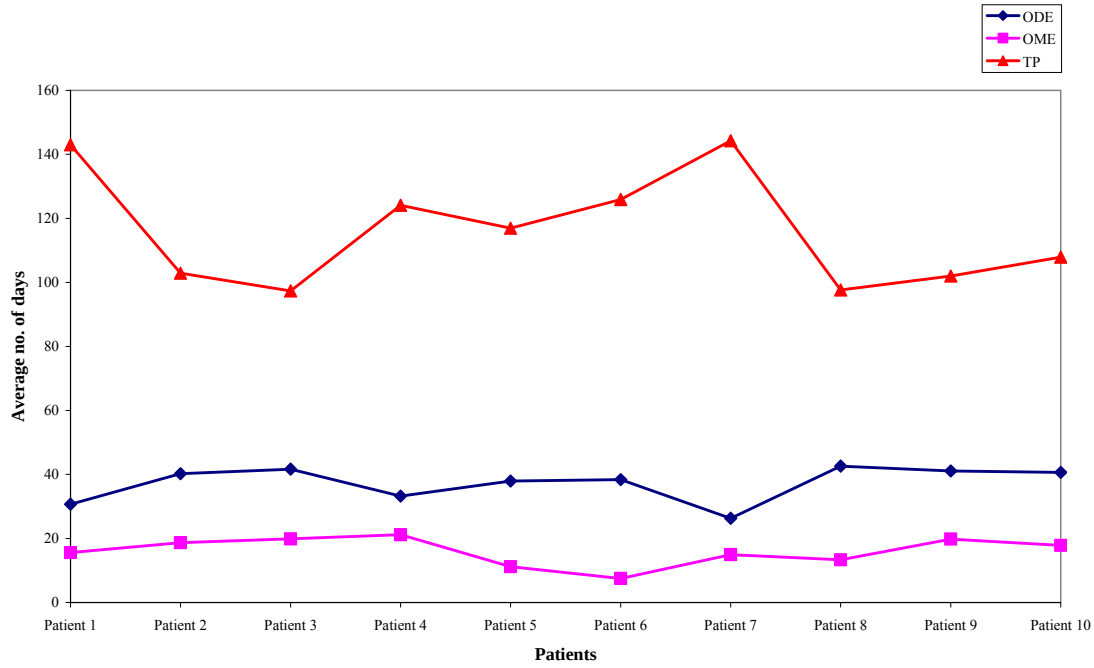


Figure 8.7 Performance of PAM for various selections of two prodromes (Dataset 1)

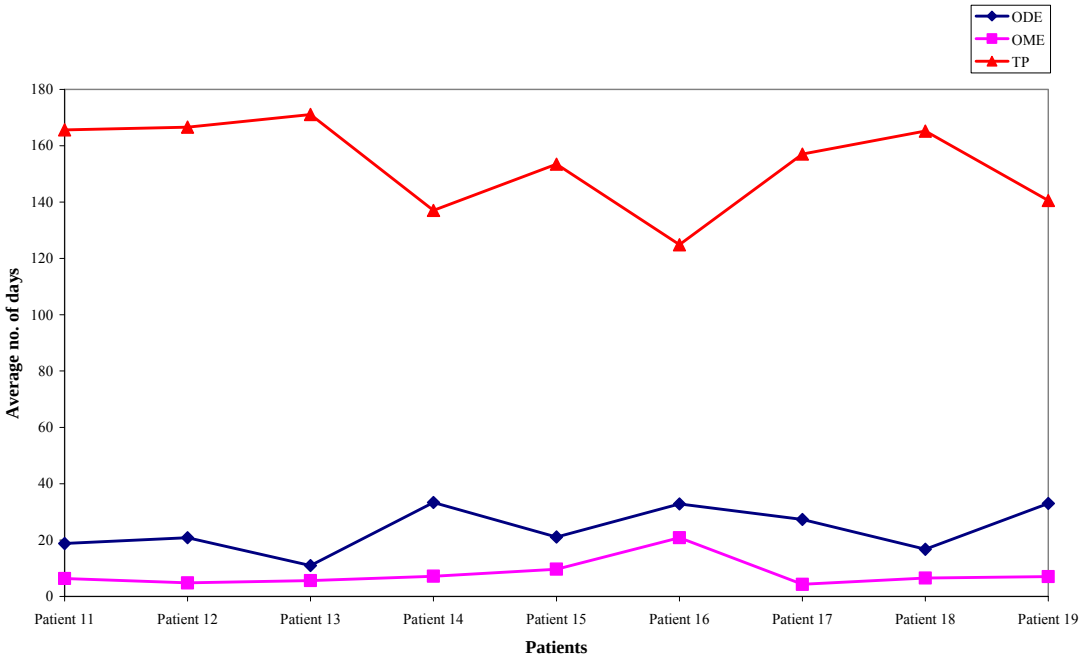


Figure 8.8 Performance of PAM for various selections of three prodromes (Dataset 1)

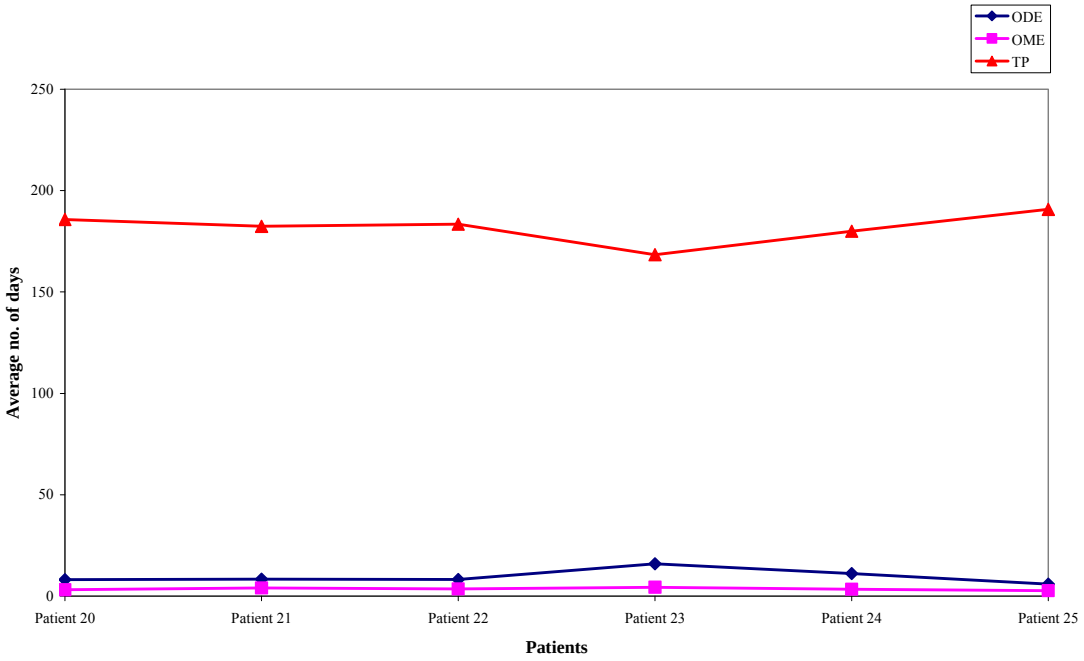


Figure 8.9 Performance of PAM for various selections of four/five prodromes (Dataset 1)

Recall that patient type 1 selected two prodromes ('activity level' and 'sleep'), patient type 11 selected three prodromes ('activity level', 'sleep' and 'talkativeness') and patient type 20 selected four prodromes ('activity level', 'sleep', 'talkativeness'

and ‘social energy’), on which to be observed. These three patient types have at least two prodromes in common, but type 11 has one more prodrome than type 1 and type 20 has one more prodrome than type 11. Likewise, patient types 2, 12 and 21 form another such subgroup. Figures 8.10, 8.11, 8.12 and 8.13 compare the values of TP, FP, ODE and OME within these various subgroups.

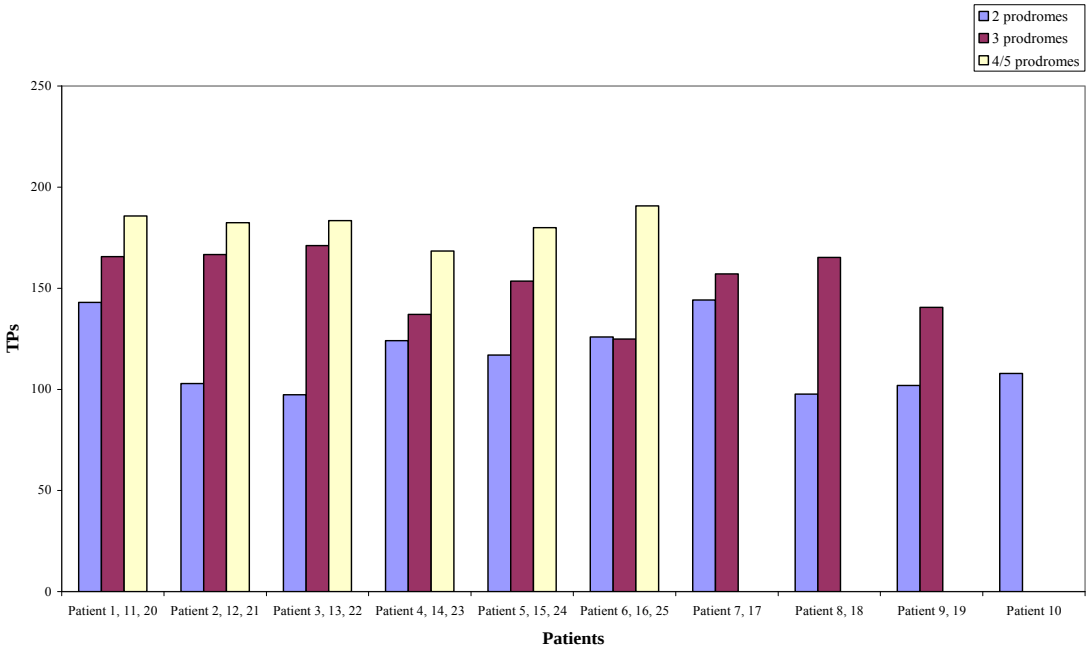


Figure 8.10 Comparison of TPs for various selections of prodromes (Dataset 1)

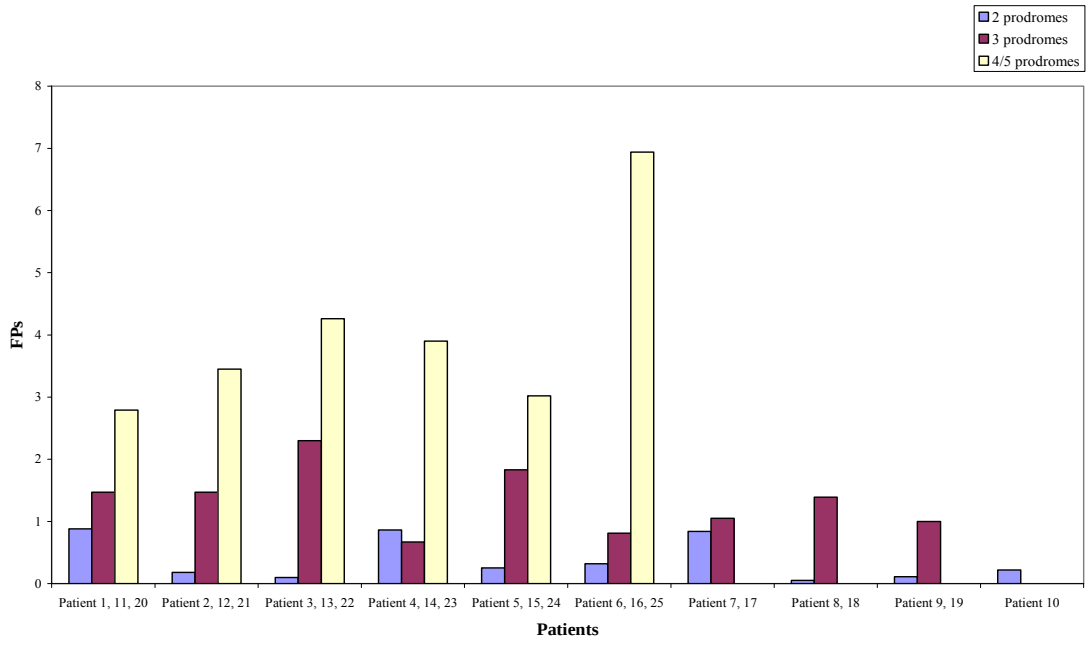


Figure 8.11 Comparison of FPs for various selections of prodromes (Dataset 1)

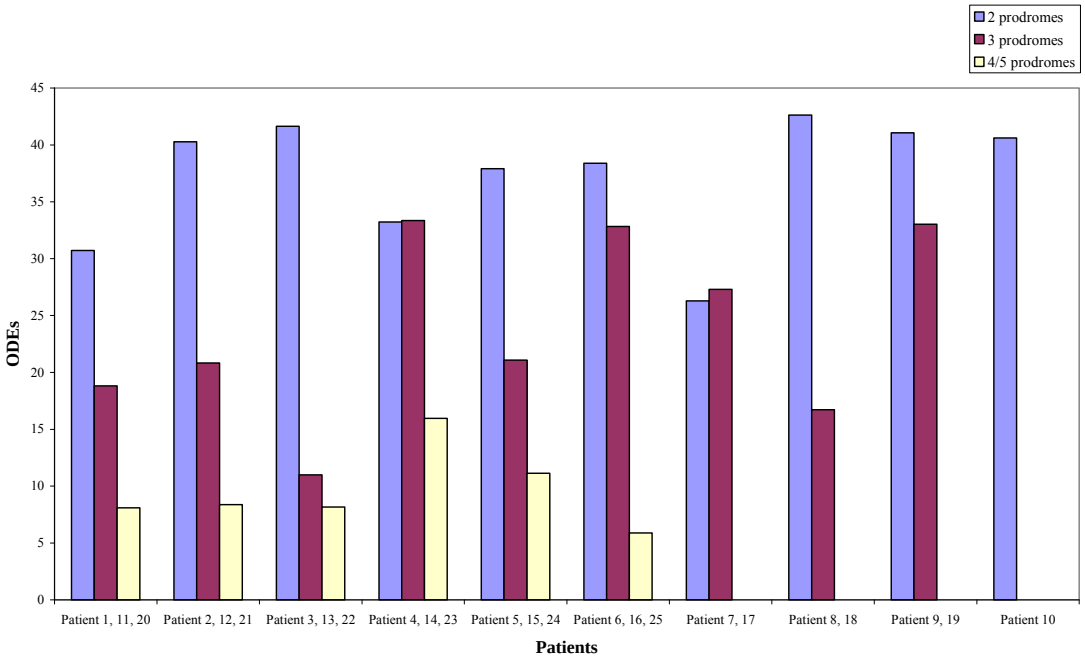


Figure 8.12 Comparison of ODEs for various choices of prodromes (Dataset 1)

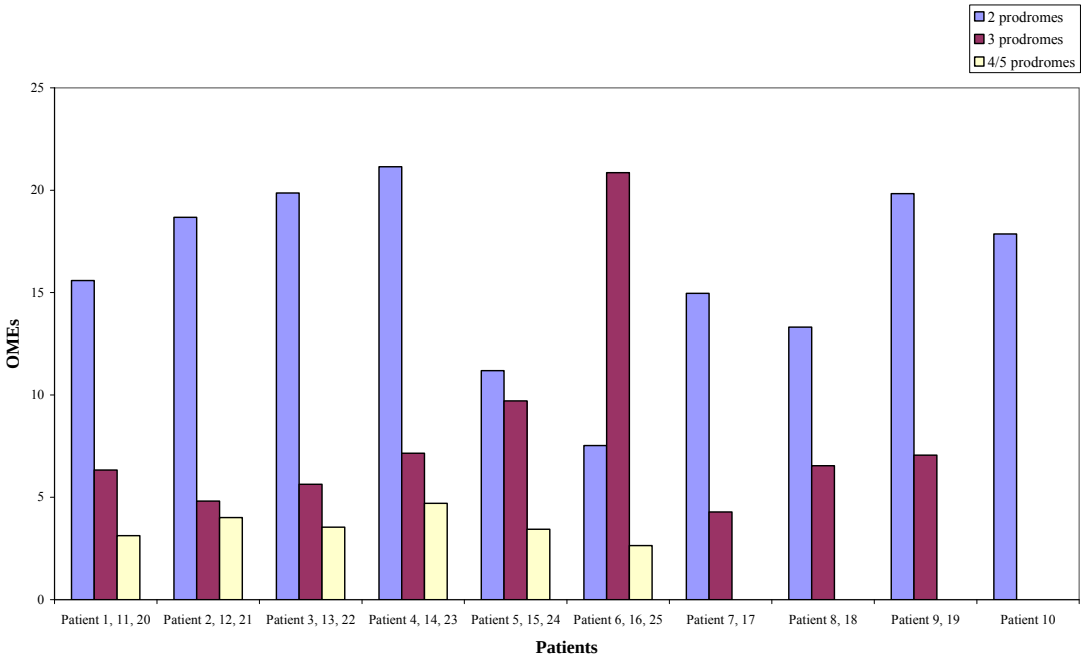


Figure 8.13 Comparison of OMEs for various choices of prodromes (Dataset 1)

Out of all the personalised choices of prodromes discussed above, it appears that the PAM system is sending reliable alerts for those individuals (patient types 11 to 25) who had selected three or more different choices of prodromes. However, PAM's performance is significantly better for individuals who were being observed for four/five different choices of prodromes than for those who only chose three prodromes. Clearly, in this case the additional prodrome provides additional information about the severity of illness, and in a real-life situation, this could be used to enable such a patient to make a better informed choice about the value of PAM. The patient would need to trade off the benefit of more reliable PAM performance against the drawback (for them) of having extra sensors to measure these additional prodromes: sensors which they would probably not, in an ideal world, have chosen.

A further potential disadvantage of adding sensors can be seen from the fact that once again, as the number of choices of prodromes increases, so do the FPs. Clearly, monitoring additional behaviours helps to increase the TP rate but also increases the FP rate. The three highest average FPs for patients selecting three or more prodromes were 25 were 6.94 days (patient type 25), 4.26 days (type 22) and 3.90 days (type 23). Nevertheless, these are not excessive out of a total of 546 days. Thus, the PAM system is not only capable of keeping false alarms to a reasonably low level, but is also capable of providing genuine alerts a high proportion of the time: the average true positive rates were 87.5% (type 25), 84.16% (type 22) and 77.23% (type 23).

However, for those individuals who had only chosen two prodromes (patient types 1 to 10), the results show that PAM may not be robust enough to send reliable alerts for all of these individuals. In particular, it can be seen from Table 8.1 and from Table 8.2 below that the overall performance for patient types 2, 3, 8, 9 and 10 is not as good as for the others. Nevertheless, the results for patient types 1, 4, 6 and 7 are more encouraging. The TP rates of these four particular patient types were respectively found to be 65.61%, 56.92%, 57.75% and 66.17%, while the FP rates are very low indeed: 0.27%, 0.26%, 0.10% and 0.26%.

The other performance measure was the time taken to detect the onset of depressive and manic episodes, i.e. the metrics ODE and OME. The model yielded satisfactory average ODEs and OMEs for nearly all patient types. Unsurprisingly, the results suggest that the more prodromes that are chosen, the better the outcomes. The best results were for type 25, with all five prodromes; in this case PAM would on average take 5.90 and 2.64 days to detect the onsets of depressive and manic episodes, respectively. Of those patient types who chose four prodromes, 20 to 24, the best results were for type 22. In this case PAM would respectively take 8.17 and 3.54 days to detect the onsets of depressive and manic episodes. The worst case was patient type 23, with 15.97 and 4.70 days respectively. For the patients who chose three prodromes, types 11 to 19, the ODEs and OMEs are respectively 10.99 and 7.26 days in the best case scenario (type 13), and 32.84 and 20.85 days in the worst case scenario (type 16).

On the other hand, the ODEs and OMEs for individuals with two prodromes whose choices are earlier described as encouraging in terms of TPs and FPs are shown in Table 8.2 below:

Patient type (choices of two prodromes)	ODE (days)	OME (days)	TP (%)	FP (%)
Type 1	30.72	15.58	65.61	0.27
Type 4	33.23	21.14	56.92	0.26
Type 6	38.40	07.53	57.75	0.10
Type 7	26.29	14.96	66.17	0.26

Table 8.2 Promising combinations of two prodromes (Dataset 1)

It can be seen from the results that the ODE is generally longer than the OME. This is in line with the clinical literature and provides further evidence that the model is producing valid results. The average duration of a depressive episode (128 days) has been found to be longer, and more gradual in onset, than the average duration of a manic episode (90 days). However, the main reason for the difference between the ODE and the OME is that the characteristics of manic behaviour are more distinctive and thus relatively easier to recognise than those of depressive behaviour. Many

people with no mental health problems at all will experience some of the symptoms of mild depression from time to time, whereas even mildly manic behaviour is much less common in the general population. Chapter 9 contains further discussion on this matter. Finally, of course, the underlying structure of the decision rules may also have been a causal factor. The model requires a threshold to have been exceeded for three out of the previous five days before sending an alert. Since the symptoms of mild depression are subtle and intermittent in nature, we would expect to have to wait longer to get a consistent result of 3 out of 5 days.

Therefore, even though the TPs and FPs appear reasonable for patient types 1, 4, 6 and 7 (see Table 8.2 above), the reliability of the PAM system is still unclear for these patients, especially for detecting the onset of depressive episodes. Thus, further experimentation is required in order to evaluate the benefit of PAM for these four patients.

8.2.2 Running the simulation model with Dataset 2

To further assess the usefulness of the PAM system, we next ran the model using the same decision rules for 1,000 iterations with *Dataset 2* for an 18-month period for each of the same 25 hypothetical patient choice types. Recall that *Dataset 2* refers to patients with generally milder symptoms. Various model outputs are shown in Table 8.3, Figure 8.14 and Appendix A5.2:

Patient type	ODE	OME	TP	FP	FN	TN
Type 1	34.15	22.23	128.23	0.68	89.77	327.32
Type 2	42.04	21.17	87.22	0.09	130.78	327.91
Type 3	46.13	25.37	49.71	0.01	168.29	327.99
Type 4	38.45	23.46	104.23	0.40	113.77	327.60
Type 5	40.98	20.74	96.53	0.13	121.47	327.87
Type 6	42.00	24.69	79.00	0.06	138.99	327.94
Type 7	36.67	23.65	119.60	0.49	98.40	327.51
Type 8	45.66	24.82	62.17	0.01	155.83	327.99
Type 9	43.31	23.66	81.76	0.04	138.09	327.96
Type 10	42.82	25.06	67.29	0.01	150.71	327.98
Type 11	25.13	12.24	153.25	1.15	64.75	326.85
Type 12	31.12	19.21	142.72	0.90	75.28	327.10
Type 13	20.73	15.76	156.32	1.45	61.68	326.55
Type 14	39.01	17.85	110.09	0.22	107.91	327.79
Type 15	32.02	15.16	138.52	0.94	81.48	327.06
Type 16	38.43	23.55	103.78	0.37	114.22	327.63
Type 17	36.35	17.83	125.44	0.34	92.56	327.60
Type 18	29.04	15.20	145.81	0.90	72.19	327.10
Type 19	40.23	21.49	109.59	0.18	108.41	327.82
Type 20	12.77	07.58	168.50	1.66	49.50	326.34
Type 21	14.45	08.37	166.88	1.97	51.12	326.03
Type 22	16.77	13.43	160.88	1.64	57.12	326.36
Type 23	27.21	12.43	145.87	1.19	72.13	326.81
Type 24	21.42	12.74	154.30	1.13	63.70	326.87
Type 25	11.33	07.49	170.46	2.28	47.54	325.72

Table 8.3 Performance of PAM for different choices of prodromes (Dataset 2)

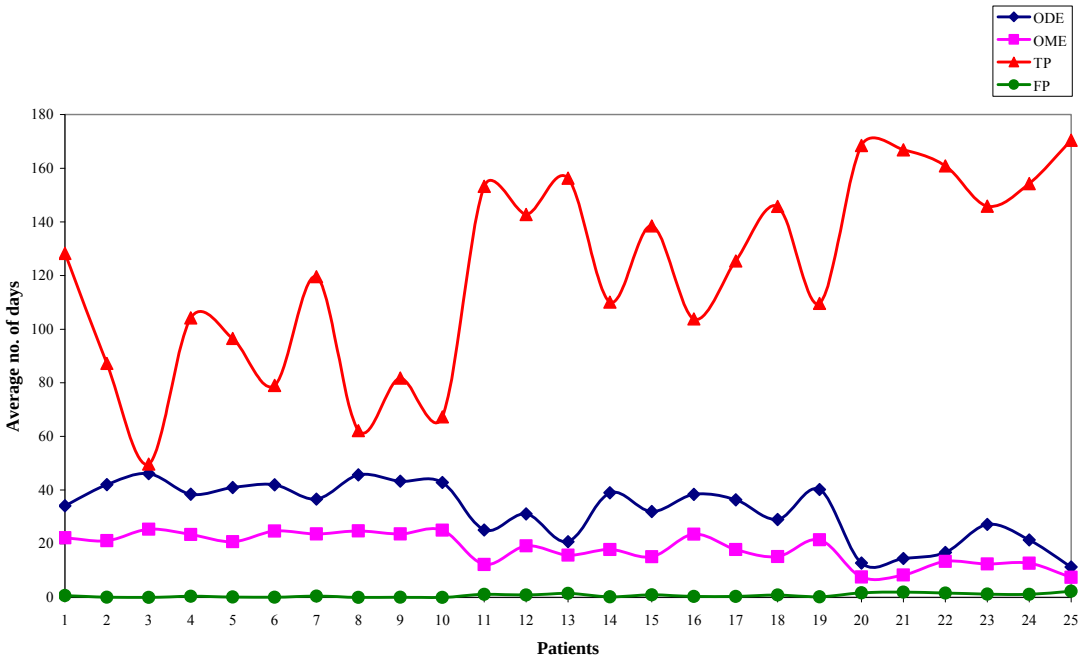


Figure 8.14 Performance of PAM for various choices of prodromes (Dataset 2)

The overall model outputs for *Dataset 2* suggest that the PAM system is most reliable for patient types 20 to 25, who were being observed for four/five different choices of prodromes. PAM is less reliable for patient types 11 to 19, who had only opted for three prodromes. Recall that *Dataset 2* relates to patients with less severe forms of BD, in which depressive and manic data overlap considerably with the normal range of data. In comparison with *Dataset 1*, there will indisputably be less variability in *Dataset 2* between normal and abnormal data. Thus, we would expect that it would be more difficult to detect the onset of acute episodes with such patients. We would not only expect fewer true and false alerts (TPs and FPs), but also greater delays in detecting the onset of depressive and manic episodes (ODEs and OMEs). This can indeed be seen from the model results (see Figures 8.15, 8.16, 8.17 and 8.18) and provides further confirmation that the model is realistic.

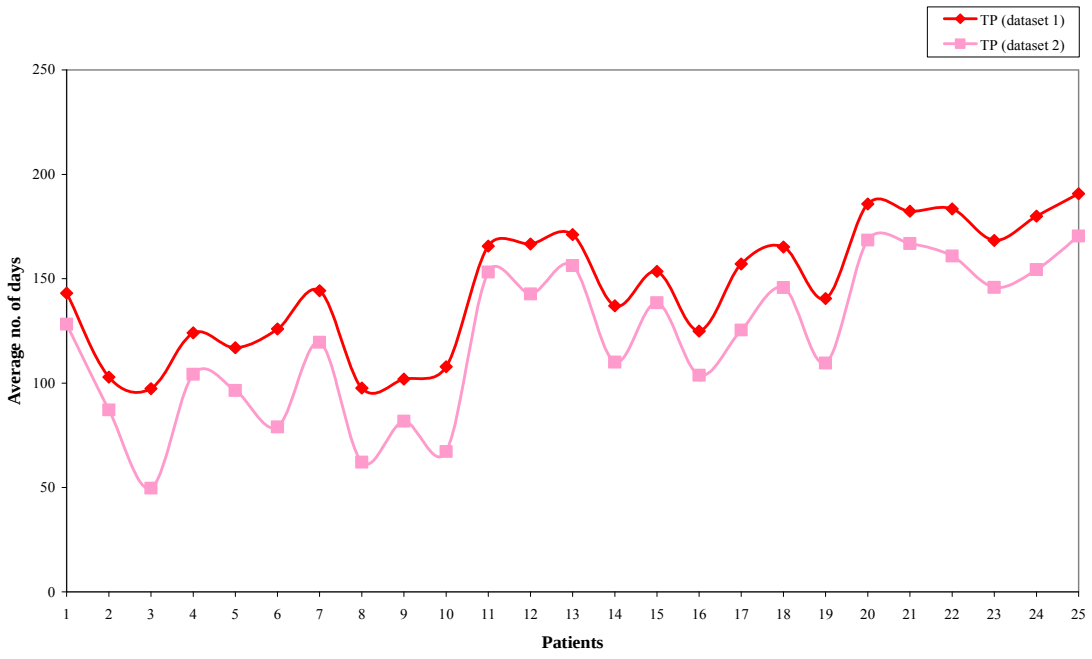


Figure 8.15 Comparison of TPs between Dataset 1 and Dataset 2

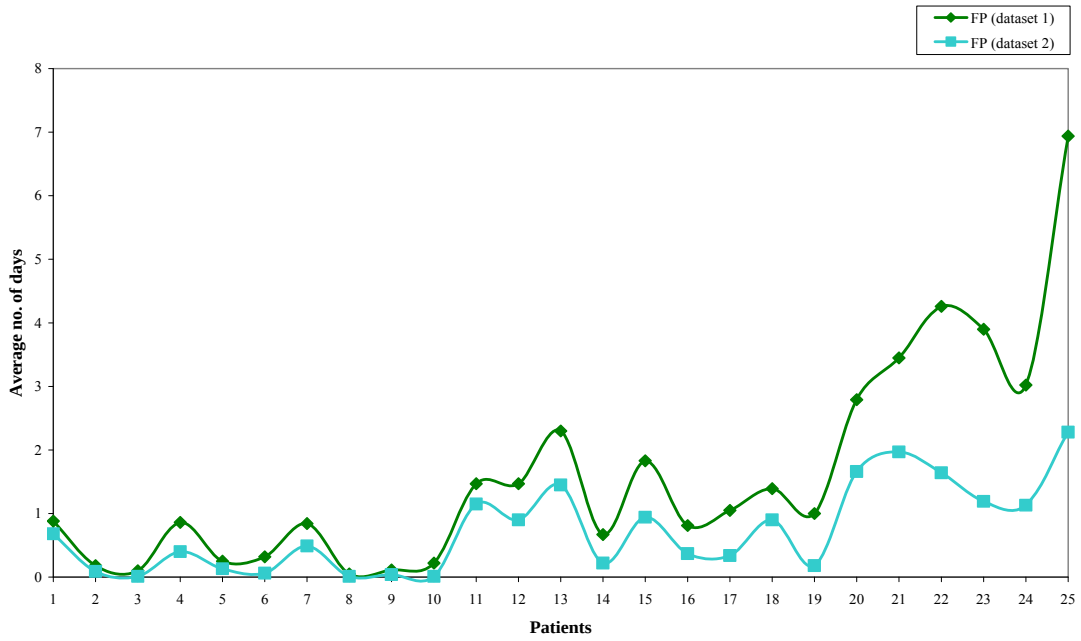


Figure 8.16 Comparison of FPs between Dataset 1 and Dataset 2

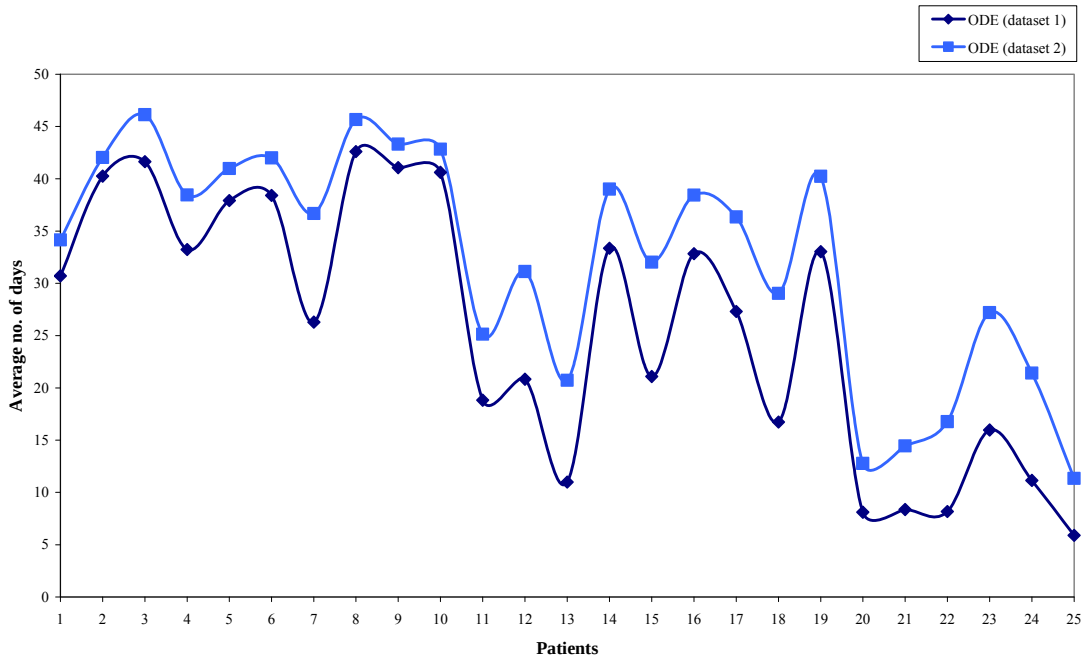


Figure 8.17 Comparison of ODEs between Dataset 1 and Dataset 2

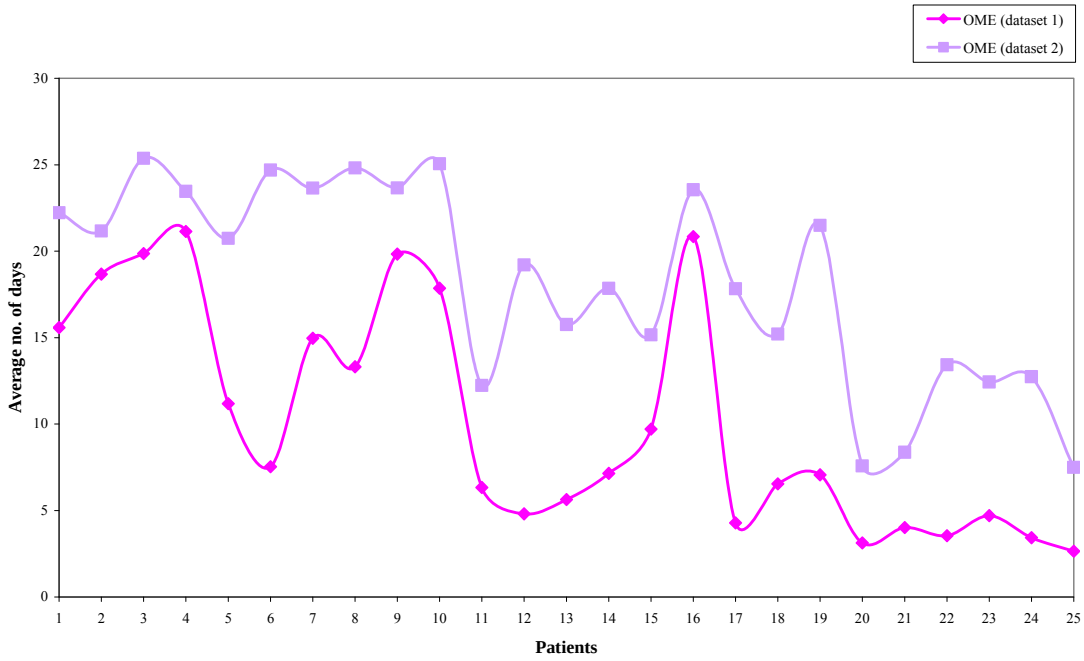


Figure 8.18 Comparison of OMEs between Dataset 1 and Dataset 2

However, even for individuals who chose three or more prodromes, there seems to be some poor outcomes in some cases, compared with the outputs of *Dataset 1*. The comparisons are shown in Tables 8.4 and 8.5 below:

Patient type (choices of three prodromes)	ODE (days)	OME (days)	TP (%)	FP (%)
Type 14	39.01	17.85	50.50	0.07
Type 16	38.43	23.55	47.61	0.11
Type 17	36.35	17.83	57.54	0.12
Type 19	40.23	21.49	50.27	0.05

Table 8.4 Poor outcomes for three prodromal choices (Dataset 2)

Patient type (choices of three prodromes)	ODE (days)	OME (days)	TP (%)	FP (%)
Type 14	33.35	07.15	62.85	0.20
Type 16	32.84	20.85	57.28	0.25
Type 17	27.31	04.28	72.04	0.32
Type 19	33.03	07.06	64.48	0.30

Table 8.5 Poor outcomes for three prodromal choices (Dataset 1)

The system performance in terms of TPs, ODEs and OMEs is particularly inadequate for the patient types listed in Table 8.4. Similarly, the overall results for patients with two personalised prodromal choices are also very poor. Nevertheless, the average results for patient type 1 (TP: 58.82%, FP: 0.21%, ODE: 34.15 days and OME: 22.23 days) are possibly acceptable (see Table 8.3).

8.2.3 Varying the decision rules

To investigate how the choice of decision rules affected the results, the model was run for 1,000 iterations with *Dataset 1* for an 18-month period for every patient, but this time an alert was sent if the threshold was exceeded for two successive days (*Decision rules 2*). So far, the PAM system's overall performance for patients with only two prodromal choices has been unsatisfactory. Therefore, we were particularly interested to see if the change in decision rules improved the overall system's performance for these patient types. Figure 8.19 and Table 8.6 below, and Appendix A5.3 show various model outputs:

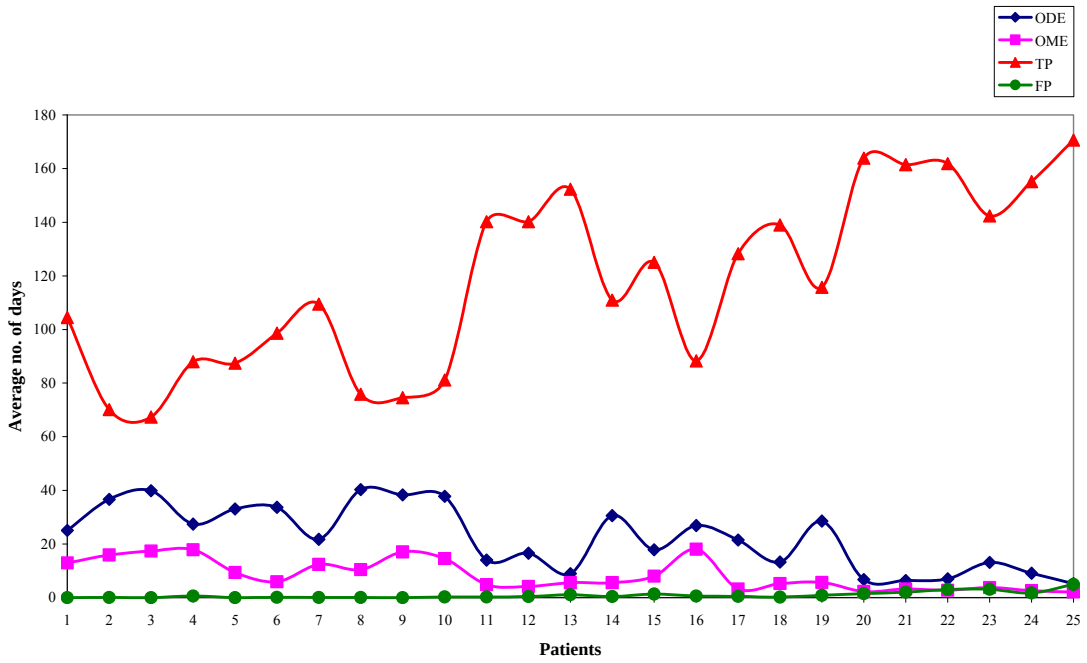


Figure 8.19 Performance of PAM for all patient types (Decision rules 2)

Patient type	ODE	OME	TP	FP	FN	TN
Type 1	25.04	12.93	104.48	0.02	113.52	327.98
Type 2	36.64	15.92	70.15	0.10	147.85	327.90
Type 3	39.89	17.37	67.35	0.03	150.65	327.97
Type 4	27.40	17.83	88.01	0.62	129.99	327.38
Type 5	33.10	09.35	87.45	0.01	128.55	327.99
Type 6	33.65	05.96	98.65	0.14	119.35	327.86
Type 7	21.74	12.41	109.47	0.04	108.54	327.97
Type 8	40.73	10.46	75.88	0.06	142.13	327.94
Type 9	38.31	17.07	74.57	0.02	143.43	327.98
Type 10	37.79	14.57	81.15	0.24	136.85	327.76
Type 11	14.03	04.85	140.18	0.26	77.82	327.74
Type 12	16.59	04.16	140.17	0.39	77.83	327.61
Type 13	08.93	05.58	152.28	1.07	65.72	326.93
Type 14	30.56	05.60	110.98	0.41	107.02	327.59
Type 15	17.86	08.04	125.00	1.37	93.00	326.63
Type 16	26.88	18.06	88.29	0.60	129.71	327.40
Type 17	21.50	03.22	128.24	0.45	89.76	327.55
Type 18	13.27	05.24	138.99	0.20	79.02	327.81
Type 19	28.58	05.66	115.73	0.86	102.22	327.14
Type 20	06.76	02.29	163.89	1.45	54.11	326.55
Type 21	06.43	03.26	161.50	2.05	56.50	325.95
Type 22	06.98	02.75	161.85	3.00	56.15	325.00
Type 23	13.12	03.80	142.36	3.10	75.64	324.90
Type 24	09.11	02.61	155.12	1.76	62.88	326.24
Type 25	05.18	02.08	170.64	4.93	47.36	323.07

Table 8.6 Performance of PAM for all patient types (Decision rules 2)

Unfortunately, the change in decision rules did not improve the system's performance for individuals with two prodromal symptoms. However, for patients with three or more prodromal choices, the system was again found to be reliable enough in terms of the TPs found with *Dataset 1* for both *Decision rules 1* and *Decision rules 2* (see Figure 8.20). Overall, the TPs found with *Decision rules 2* were consistently low (see Figure 8.20). However, one would to some extent expect this given the nature of the decision rules. For example, the system would send an alert for a pattern such as 1, 0, 1, 0, 1 using *Decision rule 1*, but not with *Decision rule 2* since prodromal symptoms in the pattern do not appear for two successive days.

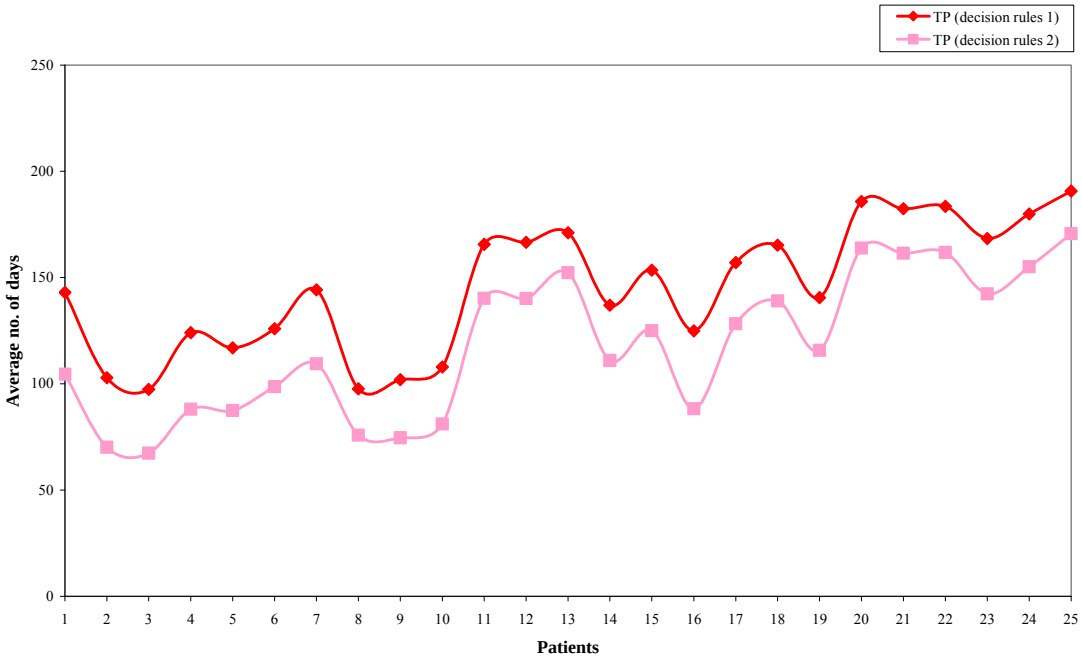


Figure 8.20 Comparison of TPs between the two different Decision rules (Dataset 1)

The outputs found here also suggest that there is no need to test the system with stricter criteria, for example continuance of prodromal symptoms for three or more successive days, since this would only further reduce the TP rates. On the other hand, the FP rate would be unacceptably high if only a single day's symptoms were used. Thus, *Decision rules 1* are clearly more effective in every respect.

8.2.4 Other sources of variation in PAM-detected output

Problems caused by sparse or missing data are very likely in the real PAM system, due not only to local technical problems with the sensors themselves (or the patient's deliberate or accidental actions, as described earlier) but also due to issues relating to consistent quality of service from the communications infrastructure. Therefore, the real-world collected data is likely at times to be sparse and incomplete, and the effectiveness of PAM in these circumstances was also tested through simulation modelling.

We have thus far allowed the actual estimations of observed behaviours to deviate by a relatively small amount, according to the numbers shown in the *Technical variation 1* column of Table 8.7, to deal with the potential technological uncertainties. We now assume a greater range of variability in the PAM-observed estimates for these behaviours, than the baseline *Technical variation 1* (as earlier shown in Table 7.4 in Page 111) due to more serious technical malfunctions.

Observed behavioural activity (unit)	***Technical variation 1* (±)**	***Technical variation 2* (±)**
Daily activity (PAL)	0.2	0.2 + 0.1
Earliest time of leaving home (time of day)	0.5	0.5 + 0.1
Latest time of getting back home (time of day)	0.5	0.5 + 0.1
TV remote keypresses (number)	10	10 + 5
Time spent in bed (hours)	0.5	0.5 + 0.2
Light level between 11pm to 7am (lux)	2	2 + 1
Noise level between 11pm to 7am (decibels)	2	2 + 1
Time spent talking on the phone (minutes)	5	5 + 3
Number of daily phone calls (number)	1	1 + 0
Time being outside between 5pm and 1am (hours)	0.5	0.5 + 0.1
Cupboard doors usage (ditto)	2	2 + 1
Fridge doors usage (ditto)	2	2 + 1
Microwave door usage (ditto)	2	2 + 0.5
Usual time of cooking (time of day)	0.5	0.5 + 0.1

Table 8.7 Further technical variations

The model was run for 1,000 iterations with *Dataset 1* and *Decision rules 1*, but with the new technical variations. Table 8.8 and Figure 8.21 below show the summary outputs, while Appendix A5.4 shows more detailed outputs:

Patient type	ODE	OME	TP	FP	FN	TN
Type 1	37.13	21.57	143.41	1.01	74.59	326.99
Type 2	36.80	15.02	106.13	0.46	111.88	327.54
Type 3	41.35	16.66	96.94	0.15	121.06	327.85
Type 4	19.44	15.34	125.25	7.43	92.76	320.57
Type 5	36.55	11.30	119.50	0.31	98.50	327.69
Type 6	32.17	06.30	97.73	0.23	120.27	327.77
Type 7	17.30	10.68	104.35	0.29	113.65	327.71
Type 8	42.28	12.84	98.73	0.07	119.27	327.93
Type 9	39.67	17.99	101.72	0.15	116.28	327.85
Type 10	39.32	15.36	106.38	1.00	111.62	327.00
Type 11	12.80	05.11	171.93	2.19	46.07	325.81
Type 12	13.59	04.24	171.34	2.21	46.66	325.79
Type 13	06.27	04.26	182.37	12.11	35.61	315.89
Type 14	28.62	05.95	142.75	1.62	75.25	326.38
Type 15	11.09	05.92	162.61	13.23	55.39	314.77
Type 16	20.42	14.95	125.18	7.03	92.82	320.97
Type 17	24.03	04.26	159.85	1.37	58.15	326.63
Type 18	12.28	05.47	169.10	1.80	48.90	326.20
Type 19	29.99	06.08	142.00	3.01	76.00	324.99
Type 20	06.27	02.83	190.05	5.67	27.95	322.31
Type 21	04.40	02.69	194.08	20.8	23.92	307.20
Type 22	04.49	02.45	193.60	25.63	24.40	304.27
Type 23	07.60	03.10	179.06	27.10	38.94	302.90
Type 24	08.07	02.88	184.22	08.08	33.78	319.92
Type 25	03.83	02.08	198.55	25.21	19.45	302.79

Table 8.8 Performance of PAM with Technical variation 2 (Dataset 1; Decision rules 1)

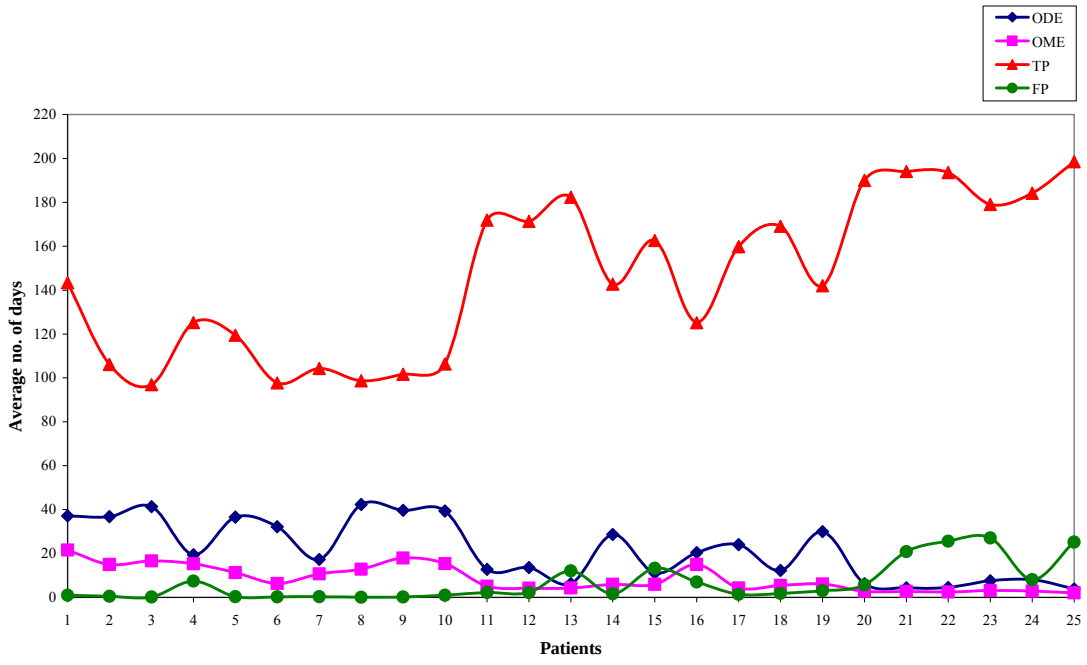


Figure 8.21 Performance of PAM for various choices of prodromes (Technical variation 2)

The outputs again tend to be better for patients with 3 or more prodromes than for those with 2 prodromes. Comparing the TPs (Figure 8.22), FPs (Figure 8.23), ODEs (Figure 8.24) and OMEs (Figure 8.25) for the two sets of technical variation, in many cases the change in technical variation did not appear to make any real significant difference. However, there are some notable differences. For example, the FP results for patient types 4, 13, 15, 16, 21, 22, 23 and 25, are strikingly different (see Figure 8.23). These patients all have two specific prodromes ('activity level' and 'appetite') in common. These two prodromes are based on eight observable behaviours, and thus the overall results differ more widely than for prodromes based on a smaller number of observable behaviours. Likewise, there are some differences in the TP results for patient types 7, 21 and 23 (see Figure 8.22 below), and in the ODE and OME results for patient types 4, 7, 15, 16 and 23 (see Figures 8.24 and 8.25). The following Table 8.9 looks more closely at these patient types:

Patient type	Prodromal choices	Total observed behaviours
4	Activity level Appetite	8
7	Sleep Appetite	7
13	Activity level Sleep, Appetite	11
15	Activity level Talkativeness Appetite	10
16	Activity level Social energy Appetite	9
21	Activity level Sleep Talkativeness Appetite	13
22	Activity level Sleep Social energy Appetite	12
23	Activity level Talkativeness Social energy Appetite	11
25	Activity level Sleep Talkativeness Social energy Appetite	14

Table 8.9 Patient types with specific prodromes

It can be seen from Table 8.9 that prodromes such as ‘activity level’ and ‘appetite’ (each with four observed behaviours) and ‘sleep’ (with three observed behaviours) are rather more informative than prodromes such as ‘talkativeness’ and ‘social energy’ both of which only have two observed behaviours.

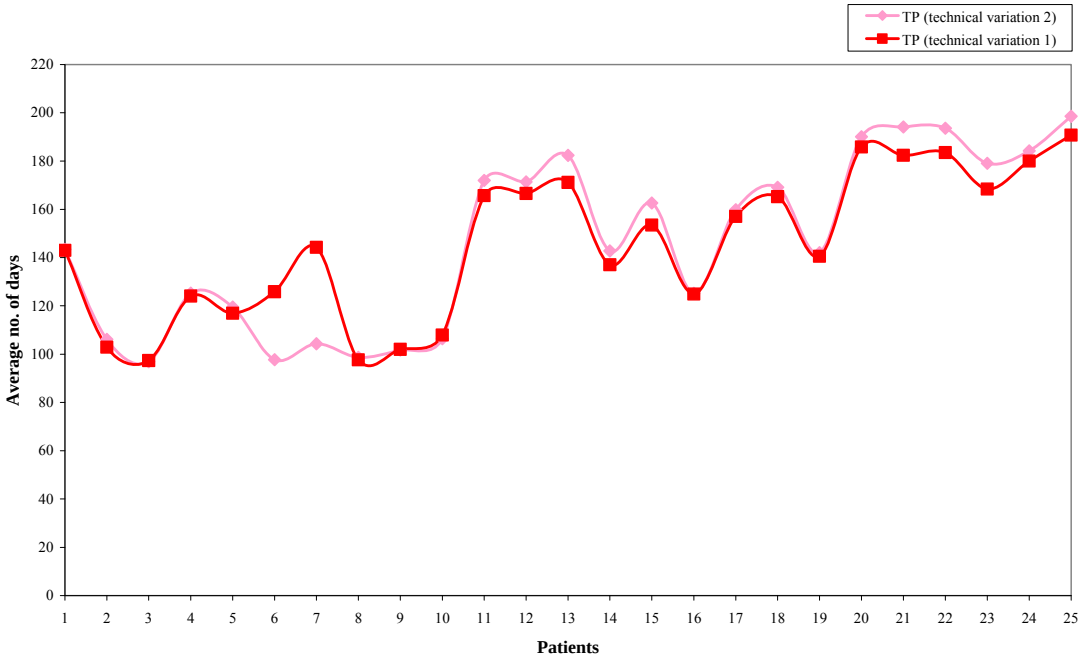


Figure 8.22 Comparison of TPs between the two sets of technical variation (Dataset 1)

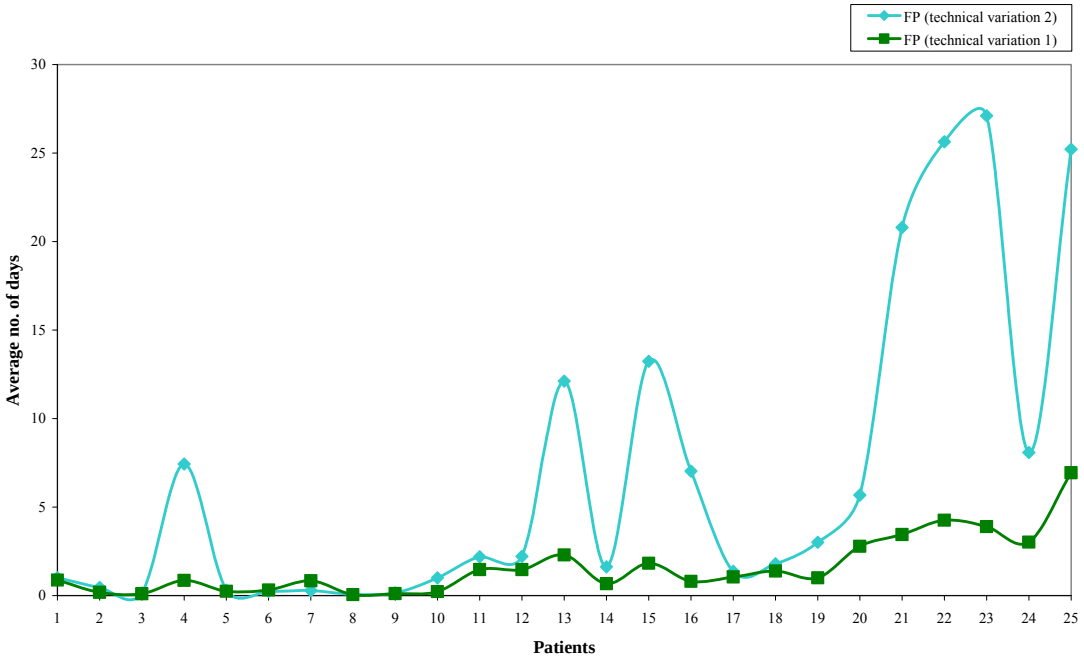


Figure 8.23 Comparison of FPs between the two sets of technical variation (Dataset 1)

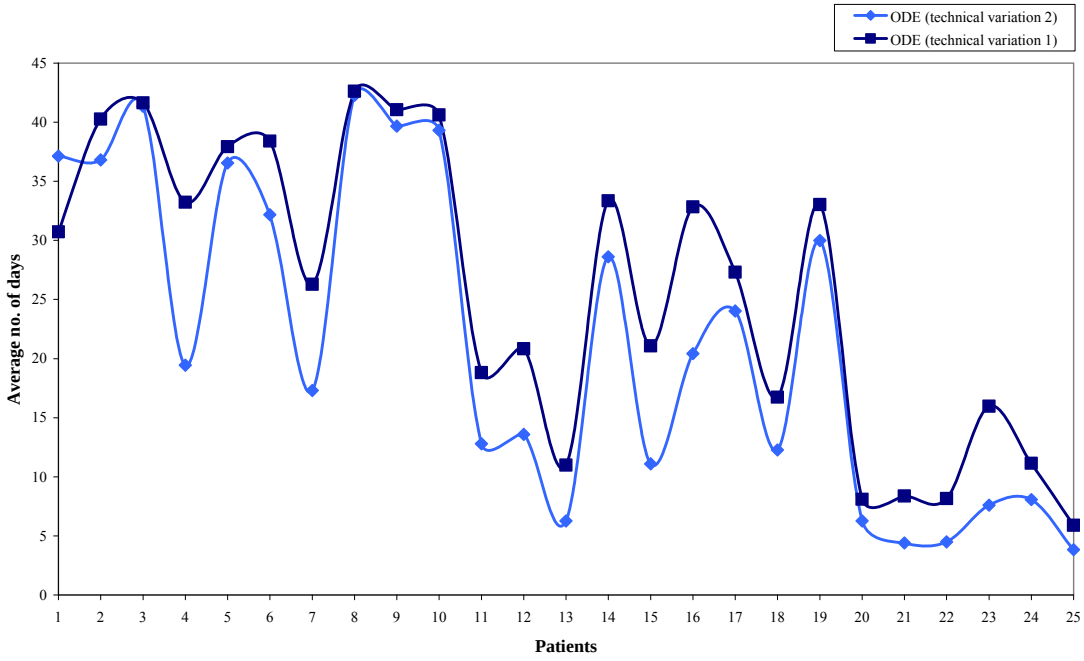


Figure 8.24 Comparison of ODEs between the two sets of technical variation (Dataset 1)

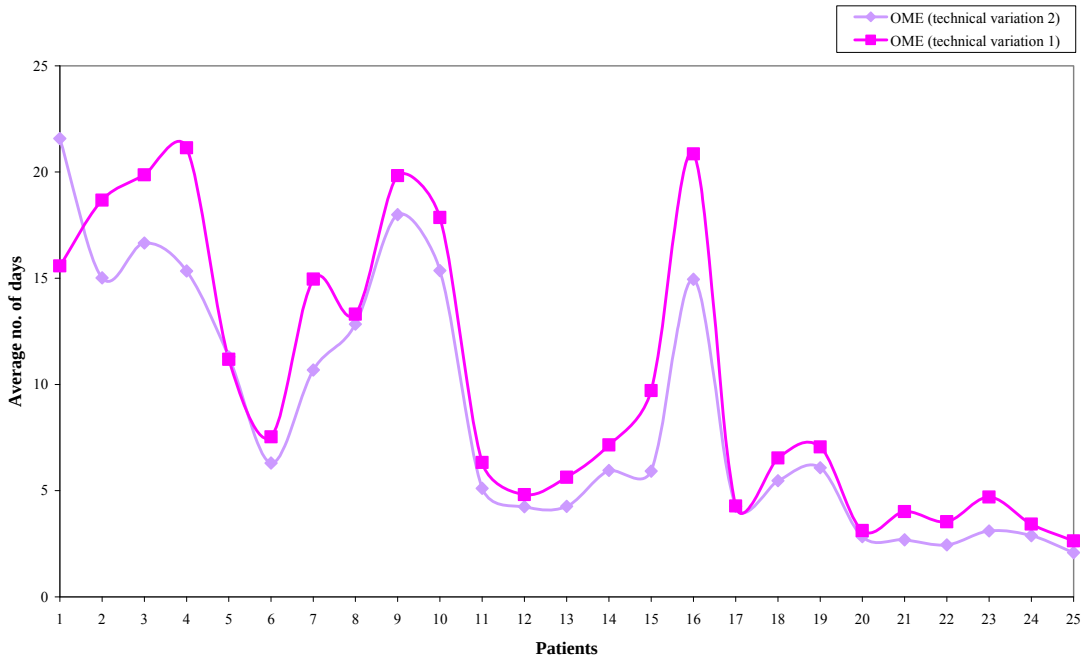


Figure 8.25 Comparison of OMEs between the two sets of technical variation (Dataset 1)

8.3 Minimum sets of sensors for personalised choices

The aim of the trials conducted so far was to evaluate the performance of PAM under various sensor configurations, corresponding to an individual patient's willingness to be monitored for different combinations of prodromes. Conversely, it is also of interest to investigate what sensors would be required in order to provide information of a given quality to the patient. In other words, the model results can be used "backwards": the required performance criteria are now defined in advance, and the model used to determine which sensor configurations meet these criteria. Following discussions with the rest of the PAM team, it was decided to set the following target performance criteria for PAM: a minimum TP rate of 70%, a maximum FP rate of 3%, a maximum ODE of 3 weeks and a maximum OME of 2 weeks.

Using the simulation results already obtained, the following three Tables (8.10, 8.11 and 8.12) were constructed. Each Table shows the smallest set of sensors acceptable to patients in each category, in descending order of TP values, which meet these target criteria for different combinations of *Dataset* and *Decision rules*.

Patient type	Prodromal choice	Minimum no. of sensors requirement	ODE (days)	OME (days)	TP (%)	FP (%)
Type 25	*Activity level* *Sleep* *Talkativeness* *Social energy* *Appetite*	*Accelerometer; GPS; TV usage sensor; Pressure mat; Light sensor; Microphone; Phone sensor; Camera; Cupboard door sensors*	05.90	02.64	87.48	2.12
Type 20	*Activity level* *Sleep* *Talkativeness* *Social energy*	*Accelerometer; GPS; TV usage sensor; Pressure mat; Light sensor; Microphone; Phone sensor*	08.10	03.12	85.22	0.85
Type 22	*Activity level* *Sleep* *Social energy* *Appetite*	*Accelerometer; GPS; TV usage sensor; Pressure mat; Light sensor; Microphone; Camera; Cupboard door sensors*	08.17	03.54	84.16	1.30
Type 21	*Activity level* *Sleep* *Talkativeness* *Appetite*	*Accelerometer; GPS; TV usage sensor; Pressure mat; Light sensor; Microphone; Phone sensor; Camera; Cupboard door sensors*	08.37	04.01	83.65	1.05

Type 24	*Sleep* *Talkativeness* *Social energy* *Appetite*	*GPS; Pressure mat; Light sensor; Microphone; Phone sensor; Camera; Cupboard door sensors*	11.14	03.43	82.54	0.92
Type 13	*Activity level* *Sleep* *Appetite*	*Accelerometer; GPS; TV usage sensor; Pressure mat; Light sensor; Microphone; Camera; Cupboard door sensors*	10.99	05.63	78.49	0.70
Type 23	*Activity level* *Talkativeness* *Social energy* *Appetite*	*Accelerometer; GPS; TV usage sensor; Phone sensor; Camera; Cupboard door sensors*	15.97	04.70	77.23	1.19
Type 12	*Activity level* *Sleep* *Social energy*	*Accelerometer; GPS; TV usage sensor; Pressure mat; Light sensor; Microphone*	20.83	04.81	76.42	0.45
Type 11	*Activity level* *Sleep* *Talkativeness*	*Accelerometer; GPS; TV usage sensor; Pressure mat; Light sensor; Microphone; Phone sensor*	18.82	06.33	75.99	0.45
Type 18	*Sleep* *Talkativeness* *Appetite*	*Pressure mat; Light sensor; Microphone; Phone sensor; Camera; Cupboard door sensors*	16.73	06.54	75.78	0.42
Type 15	*Activity level* *Talkativeness* *Appetite*	*Accelerometer; GPS; TV usage sensor; Camera; Cupboard door sensors*	21.08	09.71	70.40	0.56

Table 8.10 Acceptable choices in descending order of TPs (Dataset 1; Decision rules 1)

Figure 8.26 below shows the acceptable choices of prodromes (for *Dataset 1* and *Decision rules 1*) listed in the above table along with their associated number of sensors. The minimum number of sensors required is found to be 5 (out of a total 9 available) for patient type 15. However, with 7 sensors, the PAM system is capable of offering up to 4 personalised choices of prodromes (patient types 11, 20, 23 and 24). Clearly, patient type 21 would be better advised to choose the same set of prodromes as patient type 25, because both these types require exactly the same number of sensors but Type 21 is only monitored on 4 prodromes whereas Type 25 is monitored on 5. Similarly, it would be better for patient types 11 and 13 to choose the same prodromes as patient types 20 and 22, respectively.

The two trend-lines in Figure 8.26 show that the choices of prodromes which are derived from more sensors are generally more robust. However, this is not always true, for example compare patient type 20 (with a TP of 85.22% from 7 sensors) with patient type 21 (with a TP of 83.65% from 9 sensors). This apparently counterintuitive result occurs because 'activity level' and 'social energy' share common sensors, and both of these prodromes belong to patient type 20.

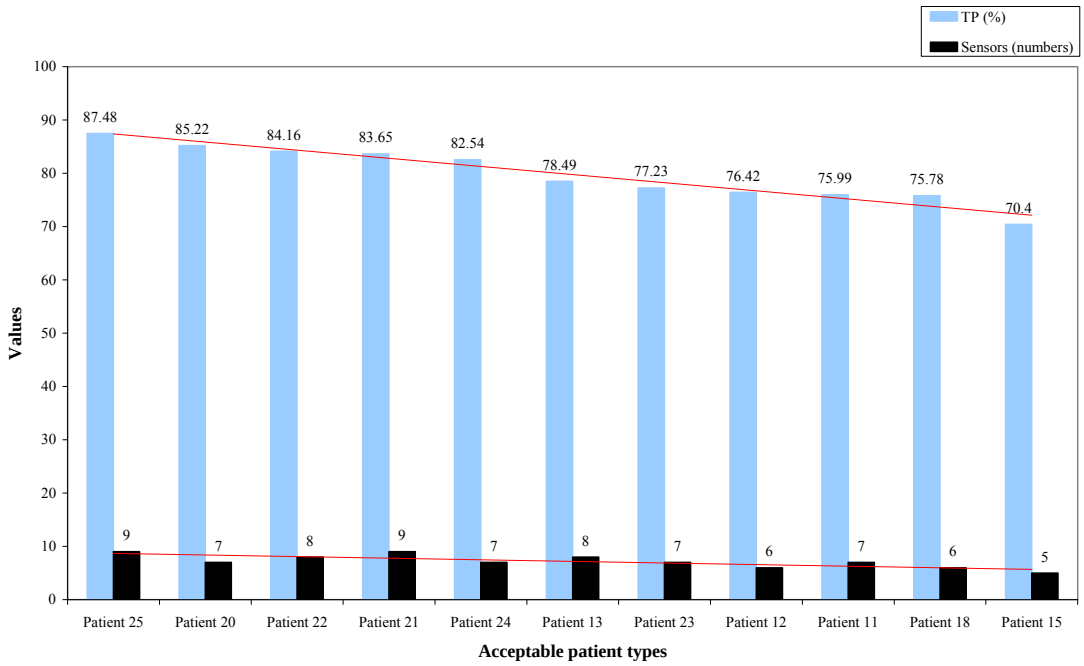


Figure 8.26 Acceptable choices of prodromes and their sensors requirement

Out of all the 11 acceptable choices of prodromes (see Table 8.10 above), 'activity level' appeared 9 times, 'sleep' 9 times, 'talkativeness' 8 times, 'social energy' 6 times and 'appetite' 8 times. This indicates that 'social energy' is clearly less informative compared with other prodromes. This is not surprising since the two observed behaviours that mapped 'social energy' were monitored only via one sensor (i.e. GPS), and moreover, one of these two observed behaviours is associated with 'activity level'. Moreover, it is tricky, in reality, to monitor specific behaviours that may be mapped to 'social energy', because a person can be socially energetic in their own home as well as outside it. Clearly, there is no simple way to monitor this prodrome, and thus to be able to detect this prodrome efficiently, the number of

required sensors would be quite high, which will of course increase the cost. The only reason for deciding to monitor 'social energy' was that the GPS was required for 'activity level', and therefore this sensor could also be used, at no extra cost, for 'social energy'.

Table 8.11 represents the acceptable choices of prodromes in descending order of TPs (for *Dataset 1* and *Decision rules 2*), while Table 8.12 represents the acceptable choices of prodromes in descending order of TPs (for *Dataset 2* and *Decision rules 1*). The patient types listed in these two tables are subsets of patient types listed in Table 8.10 above.

Patient type	Prodromal choice	Minimum no. of sensors requirement	ODE (days)	OME (days)	TP (%)	FP (%)
Patient 25	*Activity level* *Sleep* *Talkativeness* *Social energy* *Appetite*	*Accelerometer; GPS; TV usage sensor; Pressure mat; Light sensor; Microphone; Phone sensor; Camera; Cupboard door sensors*	05.18	02.08	78.28	1.50
Patient 20	*Activity level* *Sleep* *Talkativeness* *Social energy*	*Accelerometer; GPS; TV usage sensor; Pressure mat; Light sensor; Microphone; Phone sensor*	06.76	02.29	75.18	0.44
Patient 22	*Activity level* *Sleep* *Social energy* *Appetite*	*Accelerometer; GPS; TV usage sensor; Pressure mat; Light sensor; Microphone; Camera; Cupboard door sensors*	06.98	02.75	74.24	0.91
Patient 21	*Activity level* *Sleep* *Talkativeness* *Appetite*	*Accelerometer; GPS; TV usage sensor; Pressure mat; Light sensor; Microphone; Phone sensor; Camera; Cupboard door sensors*	06.43	03.26	74.08	0.63
Patient 24	*Sleep* *Talkativeness* *Social energy* *Appetite*	*GPS; Pressure mat; Light sensor; Microphone; Phone sensor; Camera; Cupboard door sensors*	09.11	02.61	71.16	0.54
Patient 13	*Activity level* *Sleep* *Appetite*	*Accelerometer; GPS; TV usage sensor; Pressure mat; Light sensor; Microphone; Camera; Cupboard door sensors*	08.93	05.58	69.85	0.33

Table 8.11 Acceptable choices in descending order of TPs (Dataset 1; Decision rules 2)

Patient type	Prodromal choice	Minimum no. of sensors requirement	ODE (days)	OME (days)	TP (%)	FP (%)
Patient 25	*Activity level Sleep Talkativeness Social energy Appetite*	*Accelerometer; GPS; TV usage sensor; Pressure mat; Light sensor; Microphone; Phone sensor; Camera; Cupboard door sensors*	11.33	07.49	78.19	0.70
Patient 20	*Activity level Sleep Talkativeness Social energy*	*Accelerometer; GPS; TV usage sensor; Pressure mat; Light sensor; Microphone; Phone sensor*	12.77	07.58	77.29	0.51
Patient 21	*Activity level Sleep Talkativeness Appetite*	*Accelerometer; GPS; TV usage sensor; Pressure mat; Light sensor; Microphone; Phone sensor; Camera; Cupboard door sensors*	14.45	08.37	76.55	0.60
Patient 22	*Activity level Sleep Social energy Appetite*	*Accelerometer; GPS; TV usage sensor; Pressure mat; Light sensor; Microphone; Camera; Cupboard door sensors*	16.77	13.43	73.80	0.50
Patient 13	*Activity level Sleep Appetite*	*Accelerometer; GPS; TV usage sensor; Pressure mat; Light sensor; Microphone; Camera; Cupboard door sensors*	20.73	15.76	71.71	0.44
Patient 24	*Sleep Talkativeness Social energy Appetite*	*GPS; Pressure mat; Light sensor; Microphone; Phone sensor; Camera; Cupboard door sensors*	21.42	12.74	70.78	0.34

Table 8.12 Acceptable choices in descending order of TPs (Dataset 2; Decision rules 1)

More challenging performance criteria, for example a minimum TP rate of 75%, a maximum FP rate of 1%, 2 weeks for ODE and one week for OME, are still achievable although not in quite so many cases.

9. Discussion and conclusion

This chapter discusses the achievements of this research, the model results and their implications, and discusses its limitations and future research ideas. We end with some final thoughts and conclusions. As mentioned earlier, we clearly distinguish between the objectives, achievements, scientific contributions and limitations of the PAM project as a whole, and those of the specific research described in this thesis.

9.1 Achievements of this thesis

We recall the original research questions, which underpinned the specific modelling work described in this thesis, presented in Section 1.8.

- Can a "natural history" model be developed for BD, using clinical data from the medical literature?
- What type of modelling approach is best suited for such a model?
- Is adequate clinical data available, and if not, can expert opinion be used instead?
- Can simulation be combined with this natural history model to capture the inherent variability and uncertainty in the real-world system?
- Can this model then be used to compare and evaluate different sets of indicator responses, from different configurations of sensors, in order to describe an individual's 'activity signature'?
- Can the model assist the design of PAM by minimising the number of sensors required, for a given individual?

These questions are addressed one by one, and represent the scientific contribution of the research carried out in this thesis.

- **Can a "natural history" model be developed for BD, using clinical data from the medical literature?**

This was possibly the greatest challenge in this research. BD manifests itself in various forms, and it is extremely difficult to develop a "natural history" model due to the illness's characteristics. There are no easily identifiable or measurable diagnostic criteria: diagnosis is a very individual and subjective process, based mainly on self-reports by patients on their moods or behaviour, combined with the observation, experience and judgment of the psychiatrist. No universally accepted staging models for bipolar disorder are to be been found in the medical literature, such as are common for physical diseases such as cancer. There are only a few examples of OR models for mental illness in general, compared with physical disease. No examples were found in the OR literature for modelling bipolar disorder, although Markov modelling has been applied to unipolar major depression, e.g. Patten and Lee (2005).

We developed a multi-state Markov disease state transition model based on transition parameters derived from the clinical literature, and this provided an "archetypal" clinical trajectory for bipolar disorder. In this model, the cycle length is 18 months. The patient starts in the normal state, and undergoes a lengthy period of depression followed later by a shorter period of mania. Through the use of Monte Carlo simulation, individual variability was incorporated this model by adding random "noise" to the current state.

- **What type of modelling approach is best suited for such a model?**

The choice of a Markov state transition model in combination with Monte Carlo simulation was made following the review of the literature and consideration of the objectives and purpose of our model. A Markov state transition model was simple yet sufficiently realistic for our purposes. This approach was the most commonly used in the literature for mental illness modelling. Combining the Markov model within a Monte Carlo experimental framework enabled us to capture stochastic and dynamic behaviour. We did consider discrete-event simulation, but discarded this for practical

reasons. A DES model would have required a great deal of detailed patient data, which were not available. We were able to perform multiple replications and sensitivity analyses in a much shorter time than would have been possible using a DES model. The other option we considered was a decision tree, but this would not have been feasible due to the unmanageable number of potential branches in the tree.

- **Is adequate clinical data available, and if not, can expert opinion be used instead?**

Some data regarding the natural history of BD were available from the clinical literature, and were used in the model (see Section 7.5). For some state transition probabilities and a general discussion about the disease model structure, we also took advice from the expert members of the PAM Steering Group. For the technical variables concerning the reliability of the PAM sensors, we consulted with the engineers in the PAM research team. However, for values of some of the behavioural parameters, for example the average number of phone calls a person might make each day or the normal time they leave home in the morning or cook their evening meal, we had to resort to common sense. While there was evidence in the literature about the importance of these behaviours, we had no secondary or primary data on which to populate the model.

- **Can simulation be combined with this natural history model to capture the inherent variability and uncertainty in the real-world system?**

The model was implemented in Microsoft Excel, coded in VBA, and using the @Risk simulation software add-in (Palisade, 2008). The Monte Carlo experimentation revealed that the stochastic Markov model is able to capture dynamic behaviour and detect a patient's true health state over time. Although it was important to be able to assess the model's reliability in practice, the process of model verification is never easy. The computer simulation model was verified by incorporating a comprehensive understanding of the real-world system. The results it produced were realistic and

plausible. In Section 8.2 of Chapter 8, we have shown that the model created is actually a very good representation of reality.

- **Can this model then be used to compare and evaluate different sets of indicator responses, from different configurations of sensors, in order to describe an individual's 'activity signature'?**

The answer to this question is yes, although we went a slightly different route than originally planned due to delays with the NHS ethics, and used illustrative rather than real-life data from the patient trial. However, the value of the thesis from a mathematical OR viewpoint has not been affected. The simulation model was used to perform a very large number of hypothetical trials, using *a*) two different (realistic) clinical patterns of bipolar disorder; *b*) two different decision rules about when to send the patient an alert about a potential change on their behaviour; *c*) 25 different combinations of up to five behavioural prodromes for which patients were willing to be monitored (this was a proxy for the actual personalised selection of sensors and locations in the home, of which there would be unfeasibly many); *d*) two different levels of technical reliability of the PAM system itself.

The PAM system was tested with two types of clinical data: *Dataset 1* and *Dataset 2*. Patients in the first group have more extreme mood swings and associated abnormal behaviour than in the second. These datasets are presented in Appendices A.3 and A.4. The results showed that the PAM system can offer a wider set of personalised prodromal choices to patients who fall into *Dataset 1* than into *Dataset 2*. In order to determine whether or not to send alerts to patients, two different groups of decision rules were tested, *Decision rules 1* and *Decision rules 2*. The former required prodromal thresholds to be exceeded (i.e. abnormal behaviour observed) on 3 out of 5 successive days. The latter required prodromal thresholds to be exceeded on two successive days. *Decision rules 1* were found to be more effective in enhancing the performance of the system, and were used in all further experimentation.

Output from the model included the four most common healthcare technology evaluators, i.e. true positive alerts (TP), false positive alerts (FP), true negatives (TN) and false negatives (FN). In addition the model computed the average number of days that the PAM system took to detect the onset of a depressive episode (ODE) and the onset of a manic episode (OME). The ideal would be a very low FP, a very high TP, and very low ODE and OME. Although the PAM system did send some false alerts, these were minimal in all cases. On the other hand, the TP rate did not exceed 90% for any of the personalised prodromal choices examined. This shows that the model is not biased towards keeping the FP values low.

The PAM system was able to detect both aspects of BD, but was more efficient in detecting the onset of manic relapse than depressive relapse. This is in accordance with clinical reality (see Section 8.2.1). Nevertheless, PAM was still able to detect the onset of depressive relapse early enough for various personalised prodromal choices.

- **Can the model assist the design of PAM by minimising the number of sensors required, for a given individual?**

The overall performance of the system was found to be inadequate for almost all the personalised choices of two prodromes only. This is not surprising, because the clinical literature suggests it is difficult to confirm a relapse signature with the appearance of only two prodromal symptoms. However, the outputs presented in Table 8.10 of Chapter 8 (see Page 145) confirms that the PAM system is efficient enough to provide its intended services to individuals who opt to be monitored for four or five personalised prodromal choices.

The performance of the system was also found to be efficient for various personalised choices of three prodromes. However, the system was found to be less effective for some combinations of personalised prodromal choices, for example ‘sleep’, ‘talkativeness’ and ‘social energy’, or ‘talkativeness’, ‘social energy’ and ‘appetite’. To be able to effectively offer choices such as these, the system may need

to increase the number of their associated observable behaviours. This will not only improve the performance of these particular choices, but will also improve the performance of other choices.

9.2 Achievements of the PAM project

We begin by recalling the original research questions, which underpinned the PAM project as a whole.

- Is it possible to obtain, in an automatic, ambient and unobtrusive manner, 'activity signatures' from mental health patients that provide information about the trajectory of their health status?
- If this is so, can this information be used to assist their healthcare?

The first of these two research questions has been answered affirmatively. The discreet sensor system built by the Nottingham group was indeed capable of deriving data which could be analysed and transmitted onwards using IT systems developed by the Stirling group, from which different activity patterns could clearly be identified through the use of feature extraction algorithms developed by the ISVR group. The PAM system was tested on healthy volunteers (the team members themselves) and also on a bipolar patient. The second question, however, still remains partly open: in order to test the practical clinical benefit, a large-scale randomised controlled trial will be required. However, the contribution of OR modelling and the findings from the simulation modelling reported in this thesis suggest strongly that clinically useful information can indeed be obtained through PAM.

9.3 Limitations

We were unable to recruit four bipolar patients for our original small field study, due to delays in receiving NHS ethical approval (in March 2010) and the inevitable bureaucratic processes required by NHS ethics in order to find suitable patients to take part in our study. Although we did recruit one participant, we did not receive her data in time for inclusion in any of the doctoral research which was part of PAM.

Thus, the OR model (and indeed, the algorithms and experimental hypotheses in the other three PhD theses arising from the PAM project) could not be tested on real patient data. Although most of the data used in our simulation model was obtained from the medical literature, this must nevertheless be seen as a limitation of the study. However, it is probably less of a limitation for the simulation modelling than it was for the engineering and computer science students.

Bipolar disorder is a multi-dimensional and extremely complex illness, and clearly even a multi-state Markov model based on a single parameter λ is a huge oversimplification. For example, BD is now clearly understood by psychiatrists to have mixed episodes as well as the simple one-dimensional spectrum from depression to mania. Moreover, clinical evidence suggests there are as many different patterns of BD as there are humans suffering from it, and to assume that this can be modelled by a single "archetypal" disease trajectory (albeit with some random variation in timing, duration and intensity of episodes) is arguably our most limiting assumption. We did attempt to mitigate this by modelling the two different clinical datasets: *Dataset 1* and *Dataset 2*. However, we believe that our disease state transition model is nevertheless fit for the intended purpose of this research, in the sense of providing an adequate description of the natural history of BD within which to test the PAM system.

Many other model assumptions, such as the choice of triangular distributions for the activity patterns, could have been made more realistic had empirical data been available and other distributions fitted. One obvious drawback of the triangular distribution is that it does not allow the sampling of extreme values. However, this was not felt to be a severely limiting assumption for the behavioural variables. The derivation of equation 7.1, and the mapping from λ to the generated behaviour using the parameters *N*, *M* and *D*, was again chosen somewhat arbitrarily and clearly other functions could have been used. Once again, we were restricted by the absence of empirical data and the total absence in the literature of any kind of similar research. As a first effort therefore, equation 7.1 and the subsequent addition of random noise to the generated values was considered adequate for its intended purpose. The

mapping from λ to a behaviour value matched both clinical experience and common sense.

9.4 Further research

Once again, some potential future work relates to the PAM project as a whole, and other work relates to further development or application of the simulation model described in this thesis. Clearly, the next stage in the development of PAM would be a larger scale clinical trial over a longer period (18 months to 2 years) involving a much larger set of patients. Patients could be randomised so that some use PAM and others use traditional manual diary-based methods. This would require the development of more robust sensors, software and IT communications network. For the small trial performed in this project, the devices were "home-made" and the network very much an experimental effort, rather than a production-type system. In addition, the level of user support would need to be fairly high initially, as inevitably there would be some technical issues. We had to make several home visits to our trial participant.

One key issue was not really explored in the PAM project due to lack of time, but which such a clinical trial could evaluate, was the acceptability of PAM from the perspectives of both patients and caregivers. The PAM team members tested the system on healthy volunteers (i.e. themselves), and showed that the system works technically and generates detectable repeated daily activity patterns from different sensors. It seemed from our discussions with self-help bipolar groups that in general, acceptability was unlikely to be a problem. However, we did not have the opportunity to assess for ourselves in practice whether bipolar sufferers would actually accept remote monitoring using PAM. This research question underpinned the practical trial with real participants, and unfortunately we were only able in the end to recruit one participant. Therefore, this remains a topic for further research. It is possible that this could be partially addressed by a patient survey, but we believe that people really need to experience monitoring for themselves before making a reliable judgment. The team members who tested PAM on themselves found that the process of being

monitored was in some ways quite different to what they had expected. Some aspects of monitoring were genuinely ignorable, but the researchers remained conscious of some sensors throughout the trial.

The simulation model showed the feasibility of inferring the mental state of bipolar patients from their behavioural activity patterns. The PAM technology could potentially be applied to the management of other chronic mental diseases in which the early onset of acute episodes is characterised by changes in behaviour, for example schizophrenia.

A common use of operational research and statistical modelling has been to support economic evaluation of medical interventions. For example, NICE (the National Institute of Health and Clinical Excellence) requires that a cost-effectiveness analysis is performed before any new treatment (device or drug) can be prescribed on the NHS. Modelling plays a key role in this (see Section 5.1). Thus, it would be very interesting to explore the potential use of the existing simulation model for an economic analysis. This could include cost analysis, cost effectiveness analysis, cost utility analysis and/or cost benefit analysis. It would be necessary to determine the operating costs of PAM once it had become established in production mode, as well as the intangible costs to the patient of being monitored: for example, the time spent charging portable devices or the potential "embarrassment factor" of having to explain the system to visitors.

It would also be necessary to estimate the financial benefits (e.g. the costs of averted hospital admissions or medications) and the intangible benefits such as increased psychological wellbeing and sense of control over one's condition. Of course, such an analysis would be hugely challenging and obtaining such data very difficult.

9.5 Conclusion

We have shown that bipolar disorder is a major and severely disabling chronic condition which can be alleviated by self-management. Through monitoring daily

activities, patients and caregivers develop a greater awareness of the context in which acute episodes occur, and can take pre-emptive action. Manual self-monitoring systems have been shown to be unreliable and in particular, ineffective at picking up the early signs of depression. The PAM system, consisting of a network of wearable and environmental sensors, is able to analyse activity patterns and detect small changes in behaviour which may indicate the onset of an acute episode.

Operational Research modelling was crucial to test the robustness of the PAM process. A dynamic disease state transition model was developed which represented the progression of bipolar disorder. This was then embedded in an Excel-based modelling framework which represented significant activity patterns for BD patients, and modelled the detection of these patterns using the PAM system. Technological considerations, risk and uncertainty inherent in monitoring patients using PAM were incorporated into this model through the use of Monte Carlo simulation using the add-in @Risk. The model showed that an automated ambient self-monitoring system like PAM can be adjusted and personalised, and can be offered as a direct motivator for behavioural change in bipolar patients.

In this novel approach, technology and modelling were combined to describe the dynamic behaviour of bipolar patients. We believe that this approach will be of great interest to the OR healthcare modelling community. The model tested the capability of the PAM system to produce reliable results in a real-life situation from a limited set of sensors. The model was used to specify the minimum best set of sensors for adequate monitoring. The overall performance of PAM was found to be good enough to support the need for further trialling.

The PAM system has been found to be capable to capture mood variability in an automatic, ambient and unobtrusive manner. Through PAM, it is possible to provide useful information about a patient's mental health status. The modelling component of the PAM project showed that it is possible to send timely alerts of an imminent bipolar episode through integrating behavioural signatures into a patient's healthcare plan. The system could therefore provide healthcare professionals with

additional clinical information to benefit bipolar patients. Hopefully, in future healthcare communities, patients and their families will find PAM reliable, simple to use and an effective method to improve the quality of care while also reducing costs. It is also possible that the same technology may be applicable to other patient groups with very little changes required.

References

Agid O, et al (1999). *Environment and vulnerability to major psychiatric illness: a case control study of early parental loss in major depression, bipolar disorder and schizophrenia.* Molecular Psychiatry 4: 163-72.

Ågren H and Backlund L (2007). *Bipolar disorder: balancing mood states early in course of illness effects long-term prognosis.* Physiology & Behavior 92: 199-202.

Akiskal H, et al (1979). *Cyclothymic temperamental disorders.* Psychiatric Clinics of North America 2: 527-54.

Akiskal H, et al. (2000). *Re-evaluating the prevalence of and diagnostic composition within the broad clinical spectrum of bipolar disorders.* Journal of Affective Disorders 59: 5-30.

Alloy L, et al (1999). *Cognitive styles and life events in subsyndromal unipolar and bipolar disorders: Stability and prospective prediction of depressive and hypomanic mood swings.* Journal of Cognitive Psychotherapy 13: 21-40.

Altman E, et al (1992). *Prodromal symptoms and signs of bipolar relapse: a report based on prospectively collected data.* Psychiatry Research, 41:18.

Altman E, et al (1997). *The Altman self-rating mania scale.* Biological Psychiatry 42(10): 948-55.

Altshuler L, et al (1995). *Antidepressant-induced mania and cycle acceleration: a controversy revisited.* American Journal of Psychiatry 152: 1130- 38.

Amor J and James C (2008). *Personalised ambient monitoring: accelerometry for activity level classification.* 4th European Congress for Medical and Biomedical Engineering (22-27 Nov).

Angst J (1998). *The emerging epidemiology of hypomania and bipolar II disorder.* Journal of Affective Disorders 50: 143-151.

Angst J and Gamma A (2002). *Update on maintenance treatments for bipolar disorder: a data given approach.* Barcelona, Spain, pp. 4-5.

Angst J and Preisig M (1995). *Course of a clinical cohort of unipolar, bipolar and schizoaffective patients. Results of a prospective study from 1959 to 1985.* Schweiz Arch Neurol Psychiatr 146(1): 5-16.

Angst J and Sellaro R (2000). *Historical perspectives and natural history of bipolar disorder.* Biological Psychiatry 48: 445-57.

Angst J, et al (2003). *Toward a re-definition of subthreshold bipolarity: epidemiology and proposed criteria for bipolar-II, minor bipolar disorders and hypomania.* Journal of Affective Disorders 73: 133-46.

Angst J, et al (2005). *The HCL-32: towards a self-assessment tool for hypomanic symptoms in outpatients.* Journal of Affective Disorders 88: 217-33.

Annandale E (1998). *The sociology of health and medicine: a critical introduction.* Polity Press.

APA - American Psychiatric Association **(2010).** Available at www.psych.org (accessed on 02.09.2010).

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

APA - American Psychiatric Association (2002). *Practice guideline for the treatment of patients with bipolar disorder (revision).* American Journal of Psychiatry 159: 1-50.

AstraZeneca (2010). Available at www.astrazeneca.co.uk (accessed on 11.02.2010).

AstraZeneca (2010). *New research shows that hospital stays are the main NHS cost of bipolar disorder (press releases).* Available at www.astrazeneca.co.uk/Media/latest-press-releases/2010/11446394?itemId=11446394 (accessed on 31.12.2010).

Bagust A, et al (1999). *Dynamics of bed use in accommodating emergency admissions: stochastic simulation model.* British Medical Journal 319(7203): 155-58.

Baldassano C (2005). *Assessment tools for screening and monitoring bipolar disorder.* Bipolar Disorders 7(1): 8-15.

Baldassano C, et al (2005). *Gender differences in bipolar disorder: Retrospective data from the first 500 STEP- BD participants.* Bipolar Disorders 7(5): 465-70.

Baldessarini R and Tondo L (2003). *Suicide risk and treatments for patients with bipolar disorder.* Journal of the American Medical Association 290(11): 1517-19.

Barton P, et al (2004). *Modelling in the economic evaluation of health care: selecting the appropriate approach.* Journal of Health Services Research & Policy 9(2): 110-18.

Batterham P, et al (2009). *Modifiable risk factors predicting major depressive disorder at four year follow-up: a decision tree approach.* BMC Psychiatry 9: 75.

Bauer M, et al (1991). *Independent assessment of manic and depressive symptoms by self-rating. Scale characteristics and implications for the study of mania.* Arch. of Gen. Psy. 48: 807-12.

Bauer M, et al (2000). *The Internal State Scale: replication of its discriminating abilities in a multisite, public sector sample.* Bipolar Disorders 2(4): 340-46.

Bauer M, et al (2004). *Using technology to improve longitudinal studies: self-reporting with ChronoRecord in bipolar disorder.* Bipolar Disorders 6(1): 67-74.

Bauer M, et al (2005). *Mood changes related to antidepressants: a longitudinal study of patients with bipolar disorder in a naturalistic setting.* Psychiatry Research 133(1): 73-80.

Bauer M, et al (2007). *Self-reported data from patients with bipolar disorder: frequency of brief depression.* Journal of Affective Disorders 101: 227-33.

Bebbington P, et al (1981). *Psychiatric disorders in selected immigrant groups in Camberwell.* Social Psychiatry 16: 43-51.

Belmaker R (2004). *Bipolar disorder.* N Engl J Med 351: 476-86.

Benazzi F (1997). *Prevalence of bipolar II disorder in outpatient depression: a 203-case study in private practice.* Journal of Affective Disorders 43(2): 163-66.

Benazzi F (2000). *Depressive mixed states: unipolar and bipolar II.* European Archives of Psychiatry and Clinical Neuroscience 250: 249-53.

Benazzi F (2002). *The mood disorder questionnaire for assessing bipolar spectrum disorder frequency.* Canadian Journal of Psychiatry 47(4): 386-87.

Benazzi F (2003). *Diagnosis of bipolar II disorder: a comparison of structured versus semistructured interviews.* Prog. in Neuro-Psychopharmacology and Biol. Psy. 27(6): 985-91.

Benneyan J (1997). *An introduction to using computer simulation in healthcare: patient wait case study.* Journal of Social Health Systems 5(3): 1-15.

BIME - Bath Institute of Medical Engineering **(2010).** *BIME - designing for disability and healthcare.* Available at www.bath.ac.uk/bime (accessed on 06.10.2010).

Blum J and Magill E (2008). *M-Psychiatry: sensor networks for psychiatric health monitoring.* Available at http://www.cms.livjm.ac.uk/pgnet2008/Proceeedings/ Papers/2008028.pdf

BOHRF - British Occupational Health Research Foundation **(2005).** *Workplace interventions for people with common mental health problems: evidence review and recommendations.* Available at www.bohrf.org.uk/downloads/cmh_rev.pdf (accessed on 11.12.2009).

Bowden C (2001). *Strategies to reduce misdiagnosis of bipolar depression.* Psychiatric Services 52(1): 51-55.

Box G and Draper N (1987). *Empirical model-building and response surfaces.* John Wiley & Sons: New York.

BPS - British Psychological Society (2010). Available at www.bps.org.uk (accessed on 17.07.2010).

Brailsford S (2005). *Overcoming barriers to the implementation of OR simulation models in healthcare.* The Journal of Clinical Investigative Medicine 28(6): 312-15.

Brailsford S and Schmidt B (2003). *Towards incorporating human behaviour in models of health care systems: an approach using discrete event simulation.* European Journal of Operational Research 150(1): 19-31.

Brailsford S, et al (2004). *Emergency and on-demand health care: modelling a large complex system.* Journal of the Operational Research Society 55(1): 34-42.

Brailsford S, et al (2009). *An analysis of the academic literature on simulation and modelling in healthcare.* Journal of Simulation 3: 130-40.

Brown G and Harris T (1982). *Fall-off in the reporting of life events.* Social Psy. 17(1): 23-28.

Brown J, et al (1999). *Using modelling in a phased approach to the economic evaluation of adjuvant therapy for early breast cancer.* Crit. Rev. in Oncology/Hamatology 32(2): 95-103.

BSDS - Bipolar Spectrum Diagnostic Scale **(2002a).** Available at http://imaging.ubmmedica.com/CME/pt/PDF/BSDSScale.pdf (accessed on 02.03.2008).

BSDS - Bipolar Spectrum Diagnostic Scale **(2002b).** Available at http://imaging.ubmmedica.com/CME/pt/PDF/BSDSScoring.pdf (accessed on 02.03.2008).

Burgess S, et al (2003). *Lithium for maintenance treatment of mood disorders.* In: The Cochrane Library 2. Oxford: Update Software.

Burgess W (2006). *The bipolar handbook: real-life questions with up-to-date answers.* Penguin.

Butler T (1995). *Management science/operations research projects in health care: the administrator's perspective.* Health Care Management Review 20: 19-25.

Calabrese J, et al (2003). *Impact of bipolar disorder on a US community sample.* Journal of Clinical Psychiatry 64: 425-32.

Campbell H, et al (2001). *Cost analysis of a hospital-at-home initiative using discrete event simulation.* Health Services Research Policy 5(5): 14-22.

Carter M (2002). *Diagnosis: mismanagement of resources. OR/MS today: health care management.* Avai. at http://www.lionhrtpub.com/orms/orms-4-02/frmismanagement.html

Chakravorty R (2006). *MobiCare: a programmable service architecture for mobile medical care.* Proc. of UbiCare 2006: the First Workshop on Ubiquitous and Pervasive Healthcare.

CIDI - Composite International Diagnostic Interview **(2001).** Available at www.cqaimh.org/pdf/tool_cidi.pdf (accessed on 10.11.2007).

Coats T and Michalis S (2001). *Mathematical modelling of patient flow through an accident and emergency department.* Emergency Medicine Journal 18: 190-92.

Colombo C, et al (1999). *Rate of switch from depression into mania after therapeutic sleep deprivation in bipolar depression.* Psychiatry Research 86: 267-70.

Componation P, et al (2003). *Benefits of modeling and simulation in implementing a shared component build strategy.* Systems Engineering 6(2): 63-74.

Cooper K, et al (2002). *The development of a simulation model of the treatment of coronary heart disease.* Health Care Management Science 5(4): 259-67.

Cooper K, et al (2007). *Choice of modelling technique for evaluating health care interventions.* Journal of the Operational Research Society 58: 168-76.

Coryell W, et al (1993). *The enduring psychosocial consequences of mania and depression.* American Journal of Psychiatry 150: 720-27.

CR - Counselling Resource **(2008).** *NIMH information about bipolar disorder.* Available at http://counsellingresource.com/distress/mood-disorders/bipolar-nimh.html (accessed on 10.10.2007).

Craddock N, et al (2004). *The Bipolar Affective Disorder Dimension Scale (BADDS) – a dimensional scale for rating lifetime psychopathology in bipolar spectrum disorders.* BMC Psychiatry 4: 19.

Crane A (2007). *Management engineers: a scientific approach to pinpoint a hospital's problems and find common-sense solutions.* Hospital & Health Networks 81: 50-4.

CST - Council for Science and Technology **(2010).** Available at www.bis.gov.uk/cst (accessed on 11.08.2010).

D'Alisa S, et al (2006). *Depression is the main determinant of quality of life in multiple sclerosis: a classification-regression (CART) study.* Disability and Rehabilitation 28(5): 307-14.

Davies R (1985). *An assessment of models of a health system.* Journal of the Operational Research Society 36(8): 679-86.

Davies R and Davies H (1994). *Modelling patient flows and resource provision in health systems.* Omega International Journal of Management Science 22: 123-31.

Davies R, et al (2003). *The evaluation of disease prevention and treatment using simulation models.* European Journal of Operational Research 150: 53-66.

De Jager C, et al (2003). *Sensitivity and specificity of neuropsychological tests for mild cognitive impairment, vascular cognitive impairment and alzheimer's disease.* Psychological Medicine 33: 1039-50.

Denicoff K, et al (2000). *Validation of the prospective NIMH-life-chart method (NIMH-LCM-p) for longitudinal assessment of bipolar illness.* Psychological Medicine 30: 1391-97.

Dennehy E, et al (2004). *Development of the brief bipolar disorder symptom scale for patients with bipolar disorder.* Psychiatry Research 127: 137-45.

Der G and Bebbington P (1987). *Depression in inner London: a register study.* Social Psychiatry 22: 73-84.

Dickson W and Kendell R (1986). *Lithium and the changing incidence of mania.* Psychological Medicine 16: 521-30.

Dienes K, et al. (2006). *The stress sensitization hypothesis: understanding the course of bipolar disorder.* Journal of Affective Disorders 95(1-3): 43-50.

DoH - Department of Health **(1992).** *The Health of the Nation.* HMSO: London.

DoH - Department of Health **(1994).** *Health of the Nation Mental Illness Key Area Handbook.* HMSO: London.

DoH - Department of Health **(1999).** *Our Healthier Nation.* TSO: London.

DoH - Department of Health **(2001).** *National Service Framework for Mental Health.* TSO: London.

DoH - Department of Health (2006). *Department of Health: departmental report 2006.* Available at http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/AnnualReports/DH_4134613 (accessed on 23.03.2009).

Dore G and Romans S (2001). *Impact of bipolar affective disorder on family and partners.* Journal of Affective Disorders 67(1-3): 147-58.

DSM-IV-TR (2000). *Diagnostic and Statistical Manual of Mental Disorders (Fourth Edition - Text Revision).* American Psychiatric Association, Washington, DC.

Duggan M, et al (2002). *Modernising the social model in mental health: a discussion paper.* London, SPN and TOPPS.

Dunner D, et al (1979). *Episode frequency prior to lithium treatment in bipolar manic-depressive patients.* Comprehensive Psychiatry 20: 511-15.

Dunner L (2003). *Clinical consequences of under-recognized bipolar spectrum disorder.* Bipolar Disorder 5: 456-63.

Eaton W, et al (1997). *Natural history of diagnostic interview schedule/DSM-IV major depression. The Baltimore epidemiological catchment area follow-up.* Archives of General Psychiatry 54(11): 993-99.

Eckblad M and Chapman L (1986). *Development and validation of a scale for hypomanic personality.* Journal of Abnormal Psychology 95: 214-22.

Eldabi T (2009). *Implementation Issues of Modeling Healthcare Problems: Misconceptions and Lessons.* Proceedings of the 2009 Winter Simulation Conference.

Eldabi T, et al (2002). *A proposed approach for modelling health-care systems for understanding.* Journal of Management in Medicine 16(4): 170-87.

Eldabi T, et al (2007). *Simulation modelling in healthcare: reviewing legacies and investigating futures.* Journal of the Operational Research Society 58: 262-70.

Eldabi T, et al (2008). *A survey of simulation techniques in commerce and defence.* Proceedings of the OR Society, 4th Simulation Workshop, Worcestershire, UK.

Ellicot A, et al (1990). *Life events and the course of bipolar disorder.* American Journal of Psychiatry 147: 1194-98.

Elphic M (2007). *Information-based management of mental health services: a two-stage model for improving funding mechanisms and clinical governance.* Psychiatric Bulletin 31: 44-48.

Ennis B (1972). *Prisoners of psychiatry.* New York: Harcourt Brace Jovanovich.

EQUAL - Extending Quality Life **(2004).** *The EPSRC EQUAL research network. (annual report).* Available at www.extra.rdg.ac.uk/equal/Archives/Annual%20Report03-04.pdf (accessed on 30.10.2007).

Fahrenberg J (2006). *Assessment in daily life. A review of computer-assisted methodologies and applications in psychology and psychophysiology, years 2000-2005.* Available at http://www. ambulatory-assessment.org/

Fajutrao L, et al (2009). *Cost-effectiveness of quetiapine plus mood stabilizers compared with mood stabilizers alone in the maintenance therapy of bipolar I disorder: results of a Markov model analysis.* Clinical Therapeutics 31, Theme Issue.

Falret J (1851). *Marche de la folie.* Gazette Hopitaux 24: 18-19.

Falret J (1854). *De la folie circulaire ou forme de maladie mentale caractérisée par l'alternative régulière de la manie et de la mélancholie.* Bull Acad Med (Paris) 19: 382.

Faravelli C, et al (1990). *Epidemiology of mood disorders: a community survey in Florence.* Journal of Affective Disorders 20 (2): 135-41.

Farmer A (1997). *Current approaches to classification. In: Murray R, et al (editors). The essentials of postgraduate psychiatry.* Cambridge: Cambridge University Press.

Fava G, et al (1984). *Health attitudes and psychological distress in patients attending a lithium clinic.* Acta Psychiatrica Scandinavica 70: 591-93.

Fava G, et al (1987). *Hostility and irritable mood during lithium prophylaxis.* Medical Science Research 15: 901-02.

Ferrier I and Thompson J (2002). *Cognitive impairment in bipolar affective disorder: implications for the bipolar diathesis.* British Journal of Psychiatry 180: 293-95.

First M, et al (1997). *Structured Clinical Interview for DSM-IV axis I disorders -Clinician Version (SCID-CV).* Washington DC, USA: American Psychiatric Press.

Fisher D and Knesper D (1983). *Markov models and the utilization of mental health services: a study of endogenously depressed patients.* Socio-economic Planning Sciences 17(1): 21-31.

Flagle C (2002). *Some origins of operations research in the health services.* Operations Research 50: 52-60.

Fone D, et al (2003). *Systematic review of the use and value of computer simulation modelling in population health and health care delivery.* Journal of public health medicine 25: 325-35.

Fries B (1976). *Bibliography on operations research in health-care systems.* Operations Research 24(5): 801-14.

Ghaemi S, et al (1999). *Is bipolar disorder still underdiagnosed? Are antidepressants overutilized?* Journal of Affective Disorders 52: 135-44.

Ghaemi S, et al (2002). *Cade's disease' and beyond: misdiagnosis, antidepressant use and a proposed definition for bipolar spectrum disorder.* Canadian Journal of Psychiatry 47:125-34.

Ghaemi S, et al (2005). *Sensitivity and specificity of a new bipolar spectrum diagnostic scale.* Journal of Affective Disorders 84: 273-77.

Gitlin M, et al (1995). *Relapse and impairment in bipolar disorder.* American Journal of Psychiatry 152: 1635-40.

Goetzel R, et al (2003). *The health and productivity cost burden of the "top 10" physical and mental health conditions affecting six large US employers in 1999.* Journal of Occupational and Environmental Medicine 45: 5-14.

Goldney R, et al (2005). *Bipolar I and II disorders in a random and representative Australian population.* Australian and New Zealand Journal of Psychiatry 39(8): 726-29.

Goodwin F and Jamison K (1990). *Manic-depressive illness.* New York: Oxford Uni. Press.

Gordon, S (2001). *Economics of simulation task force.* Proceedings of SPIE - The International Society for Optical Engineering 4367: 281-287.

Graves R, et al (2007). *Diagnosing bipolar disorder in trauma exposed primary care patients.* Bipolar Disorder 9: 318-23.

Gray P (1999). *Mental health in the workplace: tackling the effects of stress.* London: Mental Health Foundation.

Greenberg P, et al (2003). *The economic burden of depression in the United States: how did it change between 1990 and 2000?* Journal of Clinical Psychiatry 64(12): 1465-75.

Gunal M and Pidd M (2005). *Simulation modelling for performance measurement in healthcare.* Proceedings of the 2005 Winter Simulation Conference.

Gupta R and Guest J (2002). *Annual cost of bipolar disorder to UK society.* British Journal of Psychiatry, Vol. 180, pp. 227-33.

Hardoy M, et al (2005). *Validation of the Italian version of the Mood Disorder Questionnaire for the screening of bipolar disorders.* Clinical Practice and Epidemiology in Mental Health 1: 8.

Harper P and Gamlin H (2003). *Reduced outpatient waiting times with improved appointment scheduling: a simulation modelling approach.* OR Spectrum 25(2): 207-22.

Harper P and Pitt M (2004). *On the challenges of healthcare modelling and a proposed project life-cycle for success-ful implementation.* Journal of the OR Society 55: 657-61.

Heeg B, et al (2005). *Modelling the treated course of schizophrenia: development of a discrete event simulation model.* PharmacoEconomics 23(1): 17-33.

Heinssen R, et al (2003). *Overcoming barriers to research in early serious mental illness: issues for future collaboration.* Schizophrenia Bulletin 29(4): 737-45.

Hill R, et al (1996). *A survey of the manic depression fellowship.* London: Sainsbury Centre for Mental Health.

Hilty D, et al (1999). *A review of bipolar disorder among adults.* Psychiatric Services 50: 201-13.

Himmelhoch J, et al (1991). *Tranylcypromine versus imipramine in anergic bipolar depression.* American Journal of Psychiatry 148: 910-16.

Hirschfeld R, et al (2000). *Development and validation of a screening instrument for bipolar spectrum disorder: the Mood Disorder Questionnaire.* Am. Jour. of Psychiatry 157: 1873-75.

Hirschfeld R, et al (2003). *Validity of the Mood Disorder Questionnaire: a general population study.* American Journal of Psychiatry 160: 178-80.

Hirschfeld R, et al (2005). *Screening for bipolar disorder in patients treated for depression in a family medicine clinic.* Journal of the American Board of the Family Practice 18: 233-39.

Hunter E, et al (2000). *Development and validation of a mania scale for the Symptom Checklist 90.* Journal of Nervous and Mental Disease 188: 176-79.

Huybrechts K, et al (2005). *Health and economic consequences of sevelamer use for hyperphosphatemia in patients on hemodialysis.* Value in Health 8(5): 549-61.

Hyman S (1999). *Introduction to the complex genetics of mental disorders.* Biological Psychiatry 45(5): 518-21.

Hyman S (2000). *Testimony on suicide awareness and prevention.* National Institute of Mental Health. The Senate Committee on Appropriations, Subcommittee on Labor, Health and Human Services, Education.

Hyman S and Rudorfer M (2000). *Depressive and bipolar mood disorders.* In: Dale D and Federman D (editors). Scientific American Medicine, Vol. 3, New York: Healtheon/WebMD Corporation, Section 13, Subsection II, p. 1.

ICD-10 - International Classification of Diseases **(1993).** *The ICD-10 classification of mental and behavioural disorders. Diagnostic criteria for research.* Geneva: World Health Organization.

Jahangirian M, et al (2010). *Simulation in manufacturing and business: a review.* European Journal of Operational Research 203: 1-13.

James C, et al (2008). *Personalised ambient monitoring of the mentally ill.* 4th European Congress for Medical and Biomedical Engineering (22-27 Nov).

Joffe R, et al (2002). *Induction of mania and cycle acceleration in bipolar disorder, effect on different classes of antidepressant.* Acta Psychiatrica Scandinavica 105: 427-40.

Johnson S (2005). *Life events in bipolar disorder: towards more specific models.* Clinical Psychology Review 25: 1008-27.

Johnson S and Leahy R (2004). *Psychological treatment of bipolar disorder (editors).* New York: Guilford Press.

Johnson S and Miller I (1997). *Negative life events and recovery from episodes of bipolar disorder.* Journal of Abnormal Psychology 106: 449-57.

Johnson S, et al (2000). *Increases in manic symptoms following life events involving goal-attainment.* Journal of Abnormal Psychology 109: 721-27.

Johnson S, et al (2004). *Negative and goal-attainment life events: Predicting the course of mania and depression.* Manuscript under review.

Judd L and Akiskal H (2003). *The prevalence and disability of bipolar spectrum disorders in the US population: re-analysis of the ECA database taking into account subthreshold cases.* Journal of Affective Disorders 73: 123-31.

Judd L, et al (1998). *A prospective 12-year study of subsyndromal and syndromal depressive symptoms in unipolar major depressive disorders.* Arch. of General Psychiatry 55: 694-700.

Judd L, et al (2002). *The long-term natural history of the weekly symptomatic status of bipolar I disorder.* Archives of General Psychiatry 59(6): 530-37.

Judd L, et al (2003). *A prospective investigation of the natural history of the long-term weekly symptomatic status of bipolar II disorder.* Archives of General Psychiatry 60(3): 261-69.

Jun G, et al (2009). *Health care process modelling: whichmethod when?* International Journal for Quality in Health Care 21(3): 214-24.

Jun J, et al (1999). *Application of discrete event simulation in healthcare clinics: a survey.* Journal of the Operational Research Society 50: 109-23.

Kalbag A, et al (1999). *A method for classifying the course of bipolar I disorder.* Behaviortherapy 30: 355-72.

Karnon J (2003). *Alternative decision modelling techniques for the evaluation of health care technologies: Markov processes versus discrete event simulation.* Health Eco. 12: 837-48.

Karnon J and Brown J (1998). *Selecting a decision model for economic evaluation: a case study and review.* Health Care Management Science 1: 133-40.

Keane F (2010). *Why I sectioned my wife over her bipolar disorder.* Taking a stand was on BBC Radio 4 Tuesday 16 February at 0900 and 2130 GMT.

Keller M, et al (1986). *Differential outcome of pure manic, mixed/cycling, and pure depressive episodes in patients with bipolar illness.* Journal of American Medical Assoc. 255: 3138-42.

Keller M, et al (1992). *Time to recovery, chronicity, and levels of psychopathology in major depression. A 5-year prospective follow-up of 431 subjects.* Arch. of Gen. Psy. 49: 809-16.

Keller M, et al (2006). *Recurrence prevention: efficacy of two years of maintenance treatment with venxalafine XR in patients with recurrent unipolar major depression.* Toronto, ON: American Psychiatric Association Meeting (20-25 May).

Kemp D, et al (2008). *Screening for bipolar disorder in a county jail at the time of criminal arrest.* Journal of Psychiatry Research 42: 778-86.

Kessler R, et al (1997). *Childhood adversity and adult psychiatric disorder in the US national comorbidity survey.* Psychological Medicine 27: 1101-19.

Kessler R, et al (1998). *The World Health Organization Composite International Diagnostic Interview Short-Form (CIDI-SF).* Int. Journal of Methods in Psychiatric Research 7: 171-85.

Kessler R, et al (2005). *Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication.* Arch. of General Psychiatry 62(6): 593-602.

Kessler R, et al (2006). *The prevalence and correlates of adult ADHD in the United States: results from the National Comorbidity Survey Replication.* Am J of Psy 163(4): 716-23.

Kinkelin M (1954). *Verlauf und prognose des manisch-depressiven irreseins. Schweiz* Arch Neurol Neurochir Psychiatr 73: 100-46.

Klein D, et al (1996). *Hypomanic personality traits in a community sample of adolescents.* Journal of Affective Disorders 38: 135-43.

Klein R, et al (1993). *Simulation modeling and health-care decision making.* Medical Decision Making 13(4): 347-53.

Klok R, et al (2007). *Pharmacoeconomics of quetiapine for the management of acute mania in bipolar I disorder.* Expert Review of Pharmacoeconomics Outcomes Research 7(5): 459-67.

Kobak K, et al (2001). *New technologies to improve clinical trials.* Journal of Clinical Psychopharmacology 21: 255-56.

Konuk N, et al (2007). *Validation of the Turkish version of the mood disorder questionnaire for screening bipolar disorders.* Turkish Journal of Psychiatry 18: 147-54.

Kraepelin E (1913). *Psychiatrie ein lehrbuch für studierende und aerzte (8th edition).* Vol. III, Leipzig, Germany: Barth.

Kukopulos A, et al (1980). *Course of the manic-depressive cycle and changes caused by treatment.* Pharmakopsychiatrie Neuro-Psychopharmakologie 13: 156-67.

Kuno E, et al (2005). *A service system planning model for individuals with serious mental illness.* Mental Health Services Research 7(3): 135-44.

Kupfer D, et al (1988). *Possible role of antidepressants in precipitating mania and hypomania in recurrent depression.* American Journal of Psychiatry 145: 804-08.

Lagergren M (1998). *What is the role and contribution of models to management and research in the health services? A view from Europe.* European Journal of OR 105(2): 257-66.

Laje G, et al (2002). *The rating for depression as assessed by clinicians, patients, and informants (abstract).* In: Annual Meeting of the American Psychiatric Association, Philadelphia, PA.

Lam D and Wong G (1997). *Prodromes, coping strategies, insight and social functioning in bipolar affective disorders.* Psychological Medicine 27: 1091-1100.

Lam D, et al (2003). *A randomised controlled study of cognitive therapy for relapse prevention for bipolar affective disorder: outcome of the first year.* Arch. of Gen. Psy. 60(2): 145-52.

Lam D, et al (2000). *Cognitive therapy for bipolar illness - a pilot study of relapse prevention.* Cognitive Therapy and Research 24: 503-20.

Lam D, et al (2001). *Prodromes coping strategies and course of illness in bipolar affective disorder - a naturalistic study.* Psychological Medicine 31: 1397-1402.

Lane D, et al (2000). *Looking in the wrong place for healthcare improvements: a system dynamics study of an accident and emergency department.* Journal of the Operational Research society 51(5): 518-31.

Lane S, et al (2006). *A review of randomized controlled trials comparing the effectiveness of hand held computers with paper methods for data collection.* BMC Medical Informatics and Decision Making, vol. 6.

Lattimer V, et al (2004). *Reviewing emergency care systems 1: insights from system dynamics modelling.* Emergency Medicine Journal 21: 685-91.

Lay A, et al (2006). *Can discrete event simulation be of use in modelling major depression?* Cost Effectiveness and Resource Allocation 4: 19.

Leff J, et al (1976). *A cross-national epidemiological study of mania.* British Journal of Psychiatry 129: 428-42.

Lehaney B and Hlupic V (1995). *Simulation modelling for resource allocation and planning in the health sector.* Journal Royal Society of Health 115(6): 382-85.

Leibenluft E, et al (1996). *Relationship between sleep and mood in patients with rapid-cycling bipolar disorder.* Psychiatry Research 63: 161-68.

Leverich G and Post R (1996). *Life charting the course of bipolar disorder.* Curr Rev Mood Anxiety Disorder 1: 48-61.

Leverich G and Post R (2003). *Charting the course of bipolar illness and its response to treatment.* Medscape Psychiatry & Mental Health eJournal Available at: http://www.medscape.com

Li J, et al (2002). *Cost of treating bipolar disorder in the California Medicaid (Medi-Cal) program.* Journal of Affective Disorders 71(1-3): 131-39.

Lish J, et al (1994). *The National Depressive and Manic depressive Association (DMDA) survey of bipolar members.* Journal of Affective Disorders 31(4): 281-94.

Lloyd T, et al (2005). *Incidence of bipolar affective disorder in three UK cities Results from the ESOP study.* The British Journal of Psychiatry 186: 126-31.

Lowery J, et al (1994). *Barriers to implementing simulation in health care.* Proceedings of the 26th Winter Simulation Conference.

MacQueen G and Young L (2001). *Bipolar II disorder: symptoms, course, and response to treatment.* Psychiatric Services 52: 358-61.

Malan D, et al (2004). *CodeBlue: an adhoc sensor network infrastructure for emergency medical care.* Proceedings of the Int. Workshop on Wearable and Implantable Body Sensor Networks.

Mann J, et al (2008). *Classification trees distinguish suicide attempters in major psychiatric disorders: a model of clinical decision making.* Journal of clinical psychiatry 69(1): 23-31.

Manning N (2006). *DSM-IV and dangerous and severe personality disorder—an essay.* Social Science & Medicine 63: 1960-71.

Marneros A and Brieger P (2002). *Prognosis of bipolar disorders: a review.* In Maj M, et al (editors). Bipolar Disorder, Vol. 5, John Wiley and Sons Ltd.

Martinez-Arán A, et al (2004). *Cognitive function across manic or hypomanic, depressed, and euthymic states in bipolar disorder.* American Journal of Psychiatry 161: 262-70.

MATCH - Mobilising Advanced Technologies for Care at Home **(2009).** Available at www.cs.stir.ac.uk/~kjt/research/match/resources/documents/match-information.pdf (accessed on 11.02.2010).

Mauskopf J (1998). *Prevalence based economic evaluation.* Value in Health 1: 251-59.

Mayo Clinic (2010). *Bipolar disorder.* Available at www.mayoclinic.com/health/bipolar-disorder /DS00356 (accessed on 05.11.2010).

McCombs J, et al (2007). *The impact of unrecognized bipolar disorders among patients treated for depression with antidepressants in the fee-for-services California Medicaid (Medi-Cal) program: a 6-year retrospective analysis.* Journal of Affective Disorders 97: 171-79.

McElroy S, et al (2001). *Axis I psychiatric comorbidity and its relationship to historical illness variables in 288 patients with bipolar disorder.* American Journal of Psychiatry 158: 420-26.

MDQ - Mood Disorder Questionnaire **(2000).** Available at www.cqaimh.org/pdf/tool_mdq.pdf (accessed on 13.01.2008).

Mee-Lee D (2006). *Dual diagnosis (co-occurring mental and substance use disorders): assessment dilemmas.* Available at www.DMLMD.com (Accessed: 01 June 2010).

Melartin T, et al (2004). *Severity and comorbidity predict episode duration and recurrence of DSM-IV major depressive disorder.* Journal of Clinical Psychiatry 65: 810-19.

Meyer T and Hautzinger M (2003). *Screening for bipolar disorders using the Hypomanic Personality Scale.* Journal of Affective Disorders 75(2): 149-54.

Meyer T, et al (1998). *The Hypomanic Personality Scale as predictor of depression and Anxiety.* 13th Annual Meeting of the Society for Research in Psychopathology, Cambridge, MA, USA, Harvard University Press.

Michalak E, et al (2007). *Quality of life in bipolar disorder: a review of the literature.* Focus, Vol. 5, pp. 117-34.

Mielczarek B (2004). *Symulacja Komputerowa jako Narzędzie Wspomagania Decyzji w Ochronie Zdrowia.* In: Symulacja systemów gospodarczych. Balcerak A and Kwaśnicki W (editors) Poland, Wrocław, pp. 159-81.

MIFA - Mental Illness Fellowship of Australia **(2008).** *Bipolar disorder and alcohol and drug use.* Available at www.mifa.org.au/sites/www.mifa.org.au/files/documents/BipolarDisorder AlcoholandDrugUse.pdf (accessed on 09.03.2009).

Millard P, et al (2000). *Measuring and modelling surgical bed usage.* Annals of the Royal College of Surgeons England 82: 75-82.

Miller C, et al (2004). *Sensitivity and specificity of the Mood Disorder Questionnaire for detecting bipolar disorder.* Journal of Affective Disorders 81(2): 167-71.

Miller I, et al (2004). *Longitudinal course of bipolar I disorder.* Comprehensive Psychiatry 45(6): 431-40.

Molnar G, et al (1988). *Duration and symptoms of bipolar prodromes.* American Journal of Psychiatry 145: 1576-78.

Morgan V, et al (2005). *The epidemiology of bipolar disorder: sociodemographic, disability and service utilisation data from the Australian National Study of Low Prevalence (Psychotic) Disorder.* Bipolar Disorder 7(4): 326-37.

Morriss R (2004). *The early warning symptom intervention for patients with bipolar affective disorder.* Advances in Psychiatric Treatment, vol. 10, pp. 18-26.

Morriss R, et al (2004). *Interventions for helping people recognise early signs of recurrence in bipolar disorder.* The Cochrane Database of Systematic Reviews, Issue 3.

Mortensen P, et al (2003). *Individual and familial risk factors for bipolar affective disorders in Denmark.* Archives of General Psychiatry 60: 1209-15.

Müller-Oerlinghausen B, et al (2002). *Bipolar disorder.* Lancet 359: 241-47.

Naseer A, et al (2008). *Potential applications of simulation modelling techniques in healthcare: Lessons learned from aerospace & military.* Proceedings of European and Mediterranean Conference on Information Systems 2008, Dubai (25-26 May).

NCCMH - National Collaborating Centre for Mental Health **(2006).** *The management of bipolar disorder in adults, children and adolescents in primary and secondary care.* Witney, UK: Alden Press.

Nierenberg A, et al (2003). *Prevention of relapse and recurrence in depression: the role of long-term pharmacotherapy and psychotherapy.* Journal of Clinical Psychiatry 64(15): 13-7.

NIMH - National Institute of Mental Health **(2009).** Available at www.nimh.nih.gov/health/publications/bipolar-disorder/what-are-the-symptoms-of-bipolar-disorder.shtml (accessed on 07.06.2009).

NIMH - National Institute of Mental Health **(2010).** Available at www.nimh.nih.gov/health/publications/bipolar-disorder/complete-index.shtml (accessed on 14.05.2010).

NIMHGW - National Institute of Mental Health Genetics Workgroup **(1998).** *Genetics and mental disorders.* NIH Publication No. 98-4268, Rockville.

Oracle (2010). *Oracle Crystal Ball.* Available at http://landingpad.oracle.com/webapps/dialogue/ns/dlgwelcome.jsp?p_ext=Y&p_dlg_id=9580446&src=7054584&Act=39&sckw=WWMK10058778MPP002.GCM.1003.170 (accessed on 27.12.2010).

Osby U, et al (2001). *Excess mortality in bipolar and unipolar disorder in Sweden.* Archives of General Psychiatry 58: 844-50.

Pagel C, et al (2008). *Working above capacity: modelling a generic mental health process (abstract).* 34th Annual ORAHS Conference, Toronto, Canada.

Palisade (2008). *@Risk a new standard in risk analysis.* Available at http://www.palisade.com/risk/?gclid=CKm9keL8oqYCFQJO4QodyUqrZw (accessed on 15.01.2008).

Papadimitriou G, et al (2005). *Rapid cycling bipolar disorder: Biology and pathogenesis.* International Journal of Neuropsychopharmacology 8(2): 281-92.

Passik S, et al (2001). *An attempt to employ Zung self-rating depression scale as a 'lab test' to trigger follow-up in ambulatory oncology clinics: criterion validity and detection.* Journal of Pain and Symptom Management 21(4): 273-81.

Patten S (2005). *Markov models of major depression for linking psychiatric epidemiology to clinical practice.* Clinical Practice and Epidemiology in Mental Health 1: 2.

Patten S (2006). *A major depression prognosis calculator based on episode duration.* Clinical Practice and Epidemiology in Mental Health 2: 13.

Patten S and Lee R (2004). *Refining estimates of major depression incidence and episode duration in Canada using a Monte Carlo Markov model.* Med. Decision Making 24: 351-58.

Patten S and Lee R (2005). *Describing the longitudinal course of major depression using Markov models: data integration across three national surveys.* Popul. Health Metrics 3: 11.

Paykel E (1997). *The interview for recent life events.* Psychological Medicine 27: 301-10.

Paykel E, et al (2005). *Size and burden of depressive disorders in Europe.* European Neuropsychopharmacology 15(4): 411-23.

Paykel E, et al (2006). *Sub-syndromal and syndromal symptoms in the longitudinal course of bipolar disorder.* British Journal of Psychiatry 189: 118-23.

Peet M (1994). *Induction of mania with selective serotonin reuptake inhibitors and tricyclic antidepressants.* British Journal of Psychiatry 164: 549-50.

Perry A, et al (1999). *Randomised controlled trial of efficacy of teaching patients with bipolar disorder to identify early symptoms of relapse and obtain treatment.* British Medical Journal 318(7177): 149-53.

Perugi G, et al (2000). *Polarity of the first episode, clinical characteristics, and course of manic depressive illness: a systematic retrospective investigation of 320 bipolar I patients.* Comprehensive Psychiatry 41: 13-18.

Philippatos G (1973). *Financial management: theory and techniques.* San Francis., Holden Day.

Pidd M (2010). *Why modelling and model use matter.* Journal of the Operational Research Society 61: 14-24.

Pini S, et al (2005). *Prevalence and burden of bipolar disorders in European countries.* European Neuropsychopharmacology 15: 425-34.

Post R, et al (2003). *Morbidity in 258 bipolar outpatients followed for 1 year with daily prospective ratings on the NIMH life chart method.* Journal of Clin. Psychiatry 64(6): 680-90.

Pritsker A (1986). *Introduction to simulation and slam ZZ.* Systems Publishing Corporation.

Proudlove N, et al (2007). *OR and the challenge to improve the NHS: modelling for insight and improvement in -patient flows.* Journal of the Operational Research Society 58: 145-58.

Ratcliffe J, et al (2001). *A simulation modelling approach to evaluating alternative policies for the management of the waiting list for liver transplantation.* Health Care Management Science 4(2): 117-24.

Rauner M, et al (2005). *Use of discrete-event simulation to evaluate strategies for the prevention of mother-to-child transmission of HIV in developing countries.* Journal of the Operational Research Society 56: 222-33.

Rawlins M and Littlejohns P (2004). *Delivering quality in the NHS 2004.* Radcliffe Publishing: Oxford.

Regeer E, et al (2004). *Prevalence of bipolar disorder in the general population: a reappraisal study of the Netherlands mental health survey and incidence study.* Acta Psychiatrica Scandinavica 110(5): 374-82.

Richter A and Mauskopf J (1998). *Monte Carlo simulation in health care models.* Value in Health 1(1): 84-85.

Ridge J, et al (1998). *Capacity planning for intensive care units.* European Journal of Operational Research 105: 346-55.

Robinson S and Pidd M (1998). *Provider and customer expectations of successful simulation projects.* Journal of the Operational Research Society 49: 200-09.

Roderick P, et al (2003). *The cost-effectiveness of screening for Helicobacter pylori to reduce mortality and morbidity from gastric cancer and peptic ulcer disease: a discrete event simulation model.* Health Technology Assessment 7(6): 1-86.

Rotondi A, et al (1997). *Benchmarking the perioperative process. I. Patient routing systems: a method for continual improvement of patient flow and resource utilization.* Journal of Clinical Anesthesia 9(2): 159-69.

Royston G (2005). *Modelling and simulation in health - potential, achievement and challenge.* Presentation for MASHnet launch. Available at http://www.pms.ac.uk/mashnet/index.php?section=sub&act=MASHnetLaunchReport20Sep05#RoystonTalk (accessed 10 Jan 2008).

Royston G (2009). *One hundred years of operational research in health—UK 1948–2048.* Journal of the Operational Research Society 60: S169-79.

Russell L, et al (1996). *The role of cost-effectiveness analysis in health and medicine. Panel on Cost-Effectiveness in Health and Medicine.* Journal of American Med. Assoc. 276: 1172-77.

Russell S and Browne J (2005). *Staying well with bipolar disorder.* Australian and New Zealand Journal of Psychiatry 39(3): 187-93.

Sachs G, et al (2000). *The expert consensus guideline series: medication treatment of bipolar disorder 2000.* Postgraduate Medicine, Spec No: 1-104.

Sachs G, et al (2002). *A clinical monitoring form for mood disorders.* Bipolar Disord. 4: 323-27.

SAPHE - Smart and Aware Pervasive Healthcare Environments **(2006).** *Vision statement and executive summary.* Available at http://ubimon.doc.ic.ac.uk/saphe/public/SAPHE-vision-statement-pub.pdf (accessed on 09.11.2007).

Sartorius N (1983). *Mental health in the early 1980s: some perspectives.* Bulletin of the World Health Organization 61.

Schmitz N, et al (2003). *On the relation between neuroticism, self-esteem, and depression: results from the national comorbidity survey.* Comprehensive psychiatry 44(3): 169-76.

Schneck C, et al (2004). *Phenomenology of rapid cycling bipolar disorder: data from the first 500 participants in the Systematic Treatment Enhancement Program.* American Journal of Psychiatry 161: 1902-08

Scott J (2002). *Overcoming mood swings.* Constable Robinson, London, pp. 21-36.

Shah A (1982). *Integration of mental health.* Indian Journal of Psychiatry 24: 3-7.

Shih H, et al (2007). *Estimation of progression of multi-state chronic disease using the Markov model and prevalence pool concept.* BMC Medical Informatics and Decision Making 7: 34.

Shugar G, et al (1992). *Development, use, and factor analysis of a self-report inventory for mania.* Comprehensive Psychiatry 33(5): 325-31.

Silverstone T, et al (1998). *How effective is lithium in the prevention of relapse in bipolar disorder? A prospective naturalistic follow-up study.* Australian and New Zealand Journal of Psychiatry 32(1): 61-66.

Simon G (2003). *Social and economic burden of mood disorders.* Biol. Psychiatry 54: 208-15.

Singleton N, et al (2001). *Office for national statistics: psychiatric morbidity among adults living in private households.* London: Stationery Office.

Slater E (1938). *Zur periodik des manisch-depressiven irreseins.* Z Ges Neurol Psychiatr 162: 794-801.

Soares J and Mann J (1997). *The anatomy of mood disorders—review of structural neuroimaging studies.* Biological Psychiatry 41(1): 86-106.

Sobocki P, et al (2008). *The cost-utility of maintenance treatment with venlafaxine in patients with recurrent major depressive disorder.* Int. Journal of Clinical Practices 62(4): 623-32.

Solomon D, et al (1995). *Course of illness and maintenance treatments for patients with bipolar disorder.* Journal of Clinical Psychiatry 56: 5-13.

Solomon D, et al (2000). *Multiple recurrences of major depressive disorder.* American Journal of Psychiatry 157: 229-33.

Sonnenberg F and Beck J (1993). *Markov models in medical decision making: a practical guide.* Medical Decision Making 13: 322-38.

Sonnenberg F, et al (1994). *Toward a peer review process for medical decision analysis models.* Med Care 32(7): JS52-64.

SPARC - Strategic Promotion of Ageing Research Capacity **(2007).** Life in the home. Available at www.sparc.ac.uk/media/downloads/SPARC_information_sheet%201.pdf (accessed on 30.10.2007).

Spijker J, et al (2002). *Duration of major depressive episodes in the general population: results from the Netherlands Mental Health Survey and Incidence Study (NEMESIS).* British Journal of Psychiatry 181: 208-13.

Spijker J, et al (2004). *Determinants of persistence of major depressive episodes in the general population. Results from the Netherlands Mental Health Survey and Incidence Study (NEMESIS).* Journal of Affective Disorders 81: 231-40.

Strakowski S and DelBello M (2000). *The co-occurrence of bipolar and substance use disorders.* Clinical Psychology Review 20(2): 191-206.

Swann A (1995). Mixed or dysphoric manic states: Psychopathology and treatment. Journal of Clinical Psychiatry 56(3): 6-10.

Symonds R and Williams P (1981). *Lithium and the changing incidence of mania.* Psychological Medicine 11: 193-96.

Szadoczky E, et al (1998). *Prevalence of major depressive and bipolar disorder in Hungary.* Journal of Affective Disorders 50(2-3): 155-62.

Taylor S and Robinson S (2006). *So where to next? A survey of the future for discrete-event simulation.* Journal of Simulation 1(1): 1-6.

ten Have M, et al (2002). *Bipolar disorder in the general population in the Netherlands (prevalence, consequences and care utilisation): results from the Netherlands mental health survey and incidence study.* Journal of Affective Disorders 68: 203-13.

Tohen M, et al (1990). *Outcome in mania: a 4 year prospective follow-up of 75 patients utilising survival analysis.* Archives of General Psychiatry 47: 1106-11.

Tohen M, et al (2005). *Olanzapine versus lithium in the maintenance treatment of bipolar disorder: a 12-month, randomized, double-llind, controlled clinical trial.* American Journal of Psychiatry 162: 1281-90.

Truman C, et al (2002). *A comparison of self-ratings and observer ratings in bipolar patients (abstract).* In: Annual Meeting of the American Psychiatric Association, Philadelphia, PA.

Tse S and Walsh A (2001). *How does work work for people with bipolar affective disorder?* Occupational Therapy International 8: 210-25.

Twiss J, et al (2008). *Validation of the mood disorder questionnaire for screening for bipolar disorder in a UK sample.* Journal of Affective Disorders 110(1): 180-84.

UbiCare - Ubiquitous Computing for Healthcare in the Community **(2003).** Available at www-dse.doc.ic.ac.uk/Projects/ubicare/index.shtml (accessed on 15.11.2007).

UNU - United Nations University **(1994)**. *Total energy expenditure (TEE) and physical activity levels (PAL) in adults: doubly-labelled water data.* Energy and protein requirements, Proceedings of an IDECG workshop.

Van Laerhoven K, et al (2004). *Medical healthcare monitoring with wearable and implantable sensors.* Proc. of the 3rd Int. Workshop on Ubiquitous Computing for Healthcare Apps.

Vieta E, et al (2000). *Bipolar II disorder and comorbidity.* Comprehensive Psychiatry 41:339-43.

Vieta E, et al (2007). *Cross validation with the mood disorder questionnaire (MDQ) of an instrument for the detection of hypomania in Spanish: the 32 item hypomania symptom check list (HCL-32).* Journal of Affective Disorders 101(1-3): 43-55.

Vojta C, et al (2001). *Self-reported quality of life across mood states in bipolar disorder.* Comprehensive Psychiatry 42: 190-195.

Vos T, et al (2004). *The burden of major depression avoidable by longer-term treatment strategies.* Archives of General Psychiatry 61: 1097-1103.

Ward S (1989). *Arguments for constructively simple models.* Journal of the Operational Research Society 40: 141-53.

Warner J (2004). *Clinicians' guide to evaluating diagnostic and screening tests in psychiatry.* Advances in Psychiatric Treatment, vol. 10, 446-54.

Weber R, et al (2005). *Screening for bipolar disorders using a French version of the Mood Disorder Questionnaire (MDQ).* Journal of Affective Disorders 88(1): 103-08.

Wehr T and Goodwin F (1987). *Can antidepressants cause mania and worsen the course of affective illness?* American Journal of Psychiatry 144: 1403-11.

Weissman M, et al (1996). *Cross-national epidemiology of major depression and bipolar disorder.* Journal of the American Medical Association, Vol. 276 No. 4.

WHO - World Health Organization **(2001).** *Ministerial round tables: mental health.* 54th World Health Assembly, A54/DIV/4. Available at http://apps.who.int/gb/archive/pdf_files/WHA54/ea54d4.pdf (accessed on 14.01.2008).

Whybrow P, et al (2003). *The electronic assessment of the longitudinal course of bipolar disorder: the ChronoRecord software.* Pharmacopsychiatry 36(3): S244-49.

Wilson J (1981). *Implementation of computer simulation projects in health care.* Journal of the Operational Research Society 32(9): 825-32.

Wilson J and Junger G (1968). *Principles and practice of screening for disease.* Public health papers no. 34. Geneva: World Health Organization.

Wood A, et al (2006). *ALARM-NET: wireless sensor networks for assisted-living and residential monitoring.* Technical Report, Department of Computer Science, University of Virginia.

Worthington D and Brahimi M (1993). Improving out-patient appointment systems. Int. Journal of Health Care Quality Assurance 6(1): 18-23.

Young R, et al (1978). *A rating scale for mania: reliability, validity and sensitivity.* The British Journal of Psychiatry 133: 429-35.

Zarkin G, et al (2005). *Benefits and costs of methadone treatment: results from a lifetime simulation model.* Health Economics 14(11): 1133-50.

Appendix A1 PAM Steering Group

Academics:

- **Professor Christopher James**, formerly of the Institute of Sound and Vibration Research, University of Southampton, now at the University of Warwick (Principal Investigator)
- **Professor Sally Brailsford**, School of Management, University of Southampton
- **Professor John Crowe**, School of Electrical & Electronic Engineering, University of Nottingham
- **Professor Evan Magill**, Department of Computing Science and Mathematics, University of Stirling

PhD students:

- **Mr James Amor** (ISVR, Southampton)
- **Mr Syed Mohiuddin** (Management, Southampton)
- **Mr Pawel Prociow** (Nottingham)
- **Mr Jesse Blum** (Stirling)

External members:

- **Ms Sylvia Wyatt**, Future Healthcare Network, NHS Scotland
- **Dr Amy Drahota**, Research Fellow, University of Portsmouth
- **Mr Peter Jones, C**ommunity mental health nurse, Lancashire Care NHS Trust
- **Dr Paul Courtney**, Consultant Psychiatrist, Hampshire Partnership Trust
- **Mr Richard Barritt**, Chief Executive, Solent MIND
- **Mr James Stubbs**, Service User Representative

Appendix A2 Personalised Prodromal Choices

Prodromal grouping	PAM observed behaviour	Trigger if and only if
Activity level Sleep	Daily activity (A) Earliest time of leaving home (B) Latest time of getting back home (C) TV remote keypresses (D) Time spent in bed (E) Light level between 11pm to 7am (F) Noise level between 11pm to 7am (G)	[If A+B+C+D >= 2 AND If E+F+G >= 2] **OR** [If A+E = 2 AND If B+C+D+F+G >= 2] **OR** [If A = 1 AND If E+F+G >= 2] **OR** [If E = 1 AND If A+B+C+D >= 2] **OR** [If A+D+E = 3] **OR** [If A+B+C+D+E+F+G >= 5]
Activity level Talkativeness	Daily activity (A) Earliest time of leaving home (B) Latest time of getting back home (C) TV remote keypresses (D) Time spent talking on the phone (H) Number of daily phone calls (I)	[If A+B+C+D >= 2 AND If H+I >= 1] **OR** [If A+H = 2 AND If B+C+D+I >= 2] **OR** [If A+I = 2 AND If B+C+D+H >= 2] **OR** [If A = 1 AND If H+I >= 1] **OR** [If A+B+C+D+H+I >= 5]
Activity level Social energy [1]	Daily activity (A) Earliest time of leaving home (B) Latest time of getting back home (C) TV remote keypresses (D) Time being outside between 5pm and 1am (J) Latest time of getting back home (C)	[If A+B+C+D >= 2 AND If J = 1] **OR** [If A+C+J = 3] **OR** [If A+J = 2 AND

[1] Social energy comes from spending time in a stimulating environment with other people. Source: http://www.scotthyoung.com/blog/2007/04/10/social-energy/ (accessed on 20.01.2010)

		If B+C+D >= 2] **OR** [If **C**+J = 2 AND If A+B+D >= 2] **OR** [If A+B+C+D+J >= 4]
Activity level Appetite [2]	Daily activity (A) Earliest time of leaving home (B) Latest time of getting back home (C) TV remote keypresses (D) Cupboard doors usage (K) Fridge doors usage (L) Microwave door usage (M) Usual time of cooking (N)	[If A+B+C+D >= 2 AND If K+L+M+N >= 2] **OR** [If A+N = 2 AND If B+C+D+K+L+M >= 2] **OR** [If A = 1 AND If K+L+M+N >= 2] **OR** [If N = 1 AND If A+B+C+D >= 2] **OR** [If A+M+N = 3] **OR** [If B+C+D+M+N = 5] **OR** [If A+B+C+D+K+L+M+N >= 6]
Sleep Talkativeness	Time spent in bed (E) Light level between 11pm to 7am (F) Noise level between 11pm to 7am (G) Time spent talking on the phone (H) Number of daily phone calls (I)	[If E+F+G >= 2 AND If H+I >= 1] **OR** [If E+H = 2 AND If F+G+I >= 1] **OR** [If E+I = 2 AND If F+G+H >= 1] **OR** [If E = 1 AND If H+I >= 1] **OR** [If E+F+G+H+I >= 4]
Sleep Social energy	Time spent in bed (E) Light level between 11pm to 7am (F) Noise level between 11pm to 7am (G) Time being outside between 5pm and 1am (J) Latest time of getting back home (C)	[If E+F+G >= 2 AND If J+C >= 1] **OR** [If E+J = 2 AND If F+G+C# >= 1] **OR** [If E+C = 2 AND If F+G+J >= 1] **OR**

[2] Appetite is the desire to eat and drink. A decreased appetite refers to when someone has a reduced desire to eat or drink despite the body's basic energy needs. Source: Webster's Dictionary.

		[If E = 1 AND If J+C >= 1] **OR** [If E+F+G+J+C >= 4]
Sleep Appetite	Time spent in bed (E) Light level between 11pm to 7am (F) Noise level between 11pm to 7am (G) Cupboard doors usage (K) Fridge doors usage (L) Microwave door usage (M) Usual time of cooking (N)	[If E+F+G >= 2 AND If K+L+M+N >= 2] **OR** [If E+N = 2 AND If F+G+K+L+M >= 2] **OR** [If E = 1 AND If K+L+M+N >= 2] **OR** [If N = 1 AND If E+F+G >= 2] **OR** [If E+M+N = 3] **OR** [If F+G+M+N = 4] **OR** [If E+F+G+K+L+M+N >= 5]
Talkativeness Social energy	Time spent talking on the phone (H) Number of daily phone calls (I) Time being outside between 5pm and 1am (J) Latest time of getting back home (C)	[If H+I >= 1 AND If J+C >= 1] **OR** [If H = 1 AND If J+B >= 1] **OR** [If H+J = 2 AND If I+C >= 1] **OR** [If H+C = 2 AND If I+J >= 1] **OR** [If H+I+J+C >= 3]
Talkativeness Appetite	Time spent talking on the phone (H) Number of daily phone calls (I) Cupboard doors usage (K) Fridge doors usage (L) Microwave door usage (M) Usual time of cooking (N)	[If H+I >= 1 AND If K+L+M+N >= 2] **OR** [If H+N = 2 AND If I+K+L+M >= 2] **OR** [If I+N = 2 AND If H+K+L+M >= 2] **OR** [If N = 1 AND If H+I >= 1] **OR** [If H+I+N = 3] **OR**

		[If H+M+N = 3] **OR** [If I+M+N = 3] **OR** [If H+I+K+L+M+N >= 5]
Social energy Appetite	Time being outside between 5pm and 1am (J) Latest time of getting back home (C) Cupboard doors usage (K) Fridge doors usage (L) Microwave door usage (M) Usual time of cooking (N)	[If J+C >= 1 AND If K+L+M+N >= 2] **OR** [If J+N = 2 AND If C+K+L+M >= 2] **OR** [If C+N = 2 AND If J+K+L+M >= 2] **OR** [If J+M+N = 3] **OR** [If C+M+N = 3] **OR** [If J+C+K+L+M+N >= 5]
Activity level Sleep Talkativeness	Daily activity (A) Earliest time of leaving home (B) Latest time of getting back home (C) TV remote keypresses (D) Time spent in bed (E) Light level between 11pm to 7am (F) Noise level between 11pm to 7am (G) Time spent talking on the phone (H) Number of daily phone calls (I)	[Activity level Sleep] **OR** [Activity level Talkativeness] **OR** [Sleep Talkativeness] *(see above)*
Activity level Sleep Social energy	Daily activity (A) Earliest time of leaving home (B) Latest time of getting back home (C) TV remote keypresses (D) Time spent in bed (E) Light level between 11pm to 7am (F) Noise level between 11pm to 7am (G) Time being outside between 5pm and 1am (J) Latest time of getting back home (C)	[Activity level Sleep] **OR** [Activity level Social energy] **OR** [Sleep Social energy] *(see above)*
Activity level Sleep Appetite	Daily activity (A) Earliest time of leaving home (B) Latest time of getting back home (C) TV remote keypresses (D) Time spent in bed (E) Light level between 11pm to 7am (F) Noise level between 11pm to 7am (G) Cupboard doors usage (K) Fridge doors usage (L) Microwave door usage (M) Usual time of cooking (N)	[Activity level Sleep] **OR** [Activity level Appetite] **OR** [Sleep Appetite] *(see above)*
Activity level Talkativeness Social energy	Daily activity (A) Earliest time of leaving home (B) Latest time of getting back home (C) TV remote keypresses (D) Time spent talking on the phone (H) Number of daily phone calls (I) Time being outside between 5pm and 1am (J) Latest time of getting back home (C)	[Activity level Talkativeness] **OR** [Activity level Social energy] **OR** [Talkativeness Social energy]

		(see above)
Activity level Talkativeness Appetite	Daily activity (A) Earliest time of leaving home (B) Latest time of getting back home (C) TV remote keypresses (D) Time spent talking on the phone (H) Number of daily phone calls (I) Cupboard doors usage (K) Fridge doors usage (L) Microwave door usage (M) Usual time of cooking (N)	[Activity level Talkativeness] **OR** [Activity level Appetite] **OR** [Talkativeness Appetite] *(see above)*
Activity level Social energy Appetite	Daily activity (A) Earliest time of leaving home (B) Latest time of getting back home (C) TV remote keypresses (D) Time being outside between 5pm and 1am (J) Latest time of getting back home (C) Cupboard doors usage (K) Fridge doors usage (L) Microwave door usage (M) Usual time of cooking (N)	[Activity level Social energy] **OR** [Activity level Appetite] **OR** [Social energy Appetite] *(see above)*
Sleep Talkativeness Social energy	Time spent in bed (E) Light level between 11pm to 7am (F) Noise level between 11pm to 7am (G) Time spent talking on the phone (H) Number of daily phone calls (I) Time being outside between 5pm and 1am (J) Latest time of getting back home (C)	[Sleep Talkativeness] **OR** [Sleep Social energy] **OR** [Talkativeness Social energy] *(see above)*
Sleep Talkativeness Appetite	Time spent in bed (E) Light level between 11pm to 7am (F) Noise level between 11pm to 7am (G) Time spent talking on the phone (H) Number of daily phone calls (I) Cupboard doors usage (K) Fridge doors usage (L) Microwave door usage (M) Usual time of cooking (N)	[Sleep Talkativeness] **OR** [Sleep Appetite] **OR** [Talkativeness Appetite] *(see above)*
Talkativeness Social energy Appetite	Time spent talking on the phone (H) Number of daily phone calls (I) Time being outside between 5pm and 1am (J) Latest time of getting back home (C) Cupboard doors usage (K) Fridge doors usage (L) Microwave door usage (M) Usual time of cooking (N)	[Talkativeness Social energy] **OR** [Talkativeness Appetite] **OR** [Social energy Appetite] *(see above)*
Activity level Sleep Talkativeness Social energy	Daily activity (A) Earliest time of leaving home (B) Latest time of getting back home (C) TV remote keypresses (D) Time spent in bed (E) Light level between 11pm to 7am (F) Noise level between 11pm to 7am (G) Time spent talking on the phone (H) Number of daily phone calls (I)	[Activity level Sleep] **OR** [Activity level Talkativeness] **OR** [Activity level Social energy] **OR**

	Time being outside between 5pm and 1am (J) Latest time of getting back home (C)	[Sleep Talkativeness] **OR** [Sleep Social energy] **OR** [Talkativeness Social energy] *(see above)*
Activity level Sleep Talkativeness Appetite	Daily activity (A) Earliest time of leaving home (B) Latest time of getting back home (C) TV remote keypresses (D) Time spent in bed (E) Light level between 11pm to 7am (F) Noise level between 11pm to 7am (G) Time spent talking on the phone (H) Number of daily phone calls (I) Cupboard doors usage (K) Fridge doors usage (L) Microwave door usage (M) Usual time of cooking (N)	[Activity level Sleep] **OR** [Activity level Talkativeness] **OR** [Activity level Appetite] **OR** [Sleep Talkativeness] **OR** [Sleep Appetite] **OR** [Talkativeness Appetite] *(see above)*
Activity level Sleep Social energy Appetite	Daily activity (A) Earliest time of leaving home (B) Latest time of getting back home (C) TV remote keypresses (D) Time spent in bed (E) Light level between 11pm to 7am (F) Noise level between 11pm to 7am (G) Time being outside between 5pm and 1am (J) Latest time of getting back home (C) Cupboard doors usage (K) Fridge doors usage (L) Microwave door usage (M) Usual time of cooking (N)	[Activity level Sleep] **OR** [Activity level Social energy] **OR** [Activity level Appetite] **OR** [Sleep Social energy] **OR** [Sleep Appetite] **OR** [Social energy Appetite] *(see above)*
Activity level Talkativeness Social energy Appetite	Daily activity (A) Earliest time of leaving home (B) Latest time of getting back home (C) TV remote keypresses (D) Time spent talking on the phone (H) Number of daily phone calls (I) Time being outside between 5pm and 1am (J) Latest time of getting back home (C) Cupboard doors usage (K) Fridge doors usage (L) Microwave door usage (M) Usual time of cooking (N)	[Activity level Talkativeness] **OR** [Activity level Social energy] **OR** [Activity level Appetite] **OR** [Talkativeness Social energy] **OR** [Talkativeness

		Appetite] **OR** [Social energy Appetite] *(see above)*
Sleep Talkativeness Social energy Appetite	Time spent in bed (E) Light level between 11pm to 7am (F) Noise level between 11pm to 7am (G) Time spent talking on the phone (H) Number of daily phone calls (I) Time being outside between 5pm and 1am (J) Latest time of getting back home (C) Cupboard doors usage (K) Fridge doors usage (L) Microwave door usage (M) Usual time of cooking (N)	[Sleep Talkativeness] **OR** [Sleep Social energy] **OR** [Sleep Appetite] **OR** [Talkativeness Social energy] **OR** [Talkativeness Appetite] **OR** [Social energy Appetite] *(see above)*
Activity level Sleep Talkativeness Social energy Appetite	Daily activity (A) Earliest time of leaving home (B) Latest time of getting back home (C) TV remote keypresses (D) Time spent in bed (E) Light level between 11pm to 7am (F) Noise level between 11pm to 7am (G) Time spent talking on the phone (H) Number of daily phone calls (I) Time being outside between 5pm and 1am (J) Latest time of getting back home (C) Cupboard doors usage (K) Fridge doors usage (L) Microwave door usage (M) Usual time of cooking (N)	[Activity level Sleep] **OR** [Activity level Talkativeness] **OR** [Activity level Social energy] **OR** [Activity level Appetite] **OR** [Sleep Talkativeness] **OR** [Sleep Social energy] **OR** [Sleep Appetite] **OR** [Talkativeness Social energy] **OR** [Talkativeness Appetite] **OR** [Social energy Appetite] *(see above)*

Appendix A3 Dataset 1

(AA = Actual Activity)

PAM observed behaviour	When "normal" (*N*)	When "manic" (*M*)	When "depressed" (*D*)	Threshold level
Daily activity (PAL)[3]	RiskTriang(1.7,1.9,2.0)	RiskTriang(1.9,2.2,2.5)	RiskTriang(1.0,1.4,1.7)	if AA > 2.0 OR AA < 1.7, then 1
Earliest time of leaving home (time)	RiskTriang(7,8,9)	RiskTriang(5,6,7)	RiskTriang(8,9,11)	if AA < 7 OR AA > 9, then 1
Latest time of getting back home (time)	RiskTriang(18,19,21)	RiskTriang(20,23,24)	RiskTriang(16,17,19)	if AA > 21 OR AA < 18, then 1
TV remote keypresses (number)	RiskTriang(20,40,60)	RiskTriang(40,75,100)	RiskTriang(0,15,25)	if AA > 60 OR AA < 20, then 1
Time spent in bed (hour) [4]	RiskTriang(6,7,9)	RiskTriang(2,4,6)	RiskTriang(8,12,14)	if AA < 6 OR AA > 9, then 1
Light level between 11pm to 7am (lux) [5, 6]	RiskTriang(5,10,20)	RiskTriang(20,40,70)	RiskTriang(1,4,10)	if AA > 20 OR AA < 5, then 1
Noise level between 11pm to 7am (dB) [7]	RiskTriang(15,20,25)	RiskTriang(25,30,40)	RiskTriang(5,10,20)	if AA > 25 OR AA < 15, then 1
Time spent talking on the phone (minute)	RiskTriang(10,25,50)	RiskTriang(40,70,120)	RiskTriang(0,10,20)	if AA > 50 OR AA < 10, then 1
Number of daily phone calls (number)	RiskTriang(2,5,7)	RiskTriang(7,10,16)	RiskTriang(0,1,3)	if AA > 7 OR AA < 2, then 1
Time being outside btwn 5pm & 1am (hour)	RiskTriang(1,2,4)	RiskTriang(3,5,8)	RiskTriang(0,0.5,1)	if AA > 3 OR AA < 1, then 1
Cupboard doors usage (number)	RiskTriang(8,10,14)	RiskTriang(12,18,22)	RiskTriang(2,4,10)	if AA > 14 OR AA < 8, then 1
Fridge doors usage (number)	RiskTriang(6,8,10)	RiskTriang(10,14,18)	RiskTriang(2,4,6)	if AA > 10 OR AA < 6, then 1
Microwave door usage (number)	RiskTriang(4,6,8)	RiskTriang(6,10,14)	RiskTriang(0,2,4)	if AA > 8 OR AA < 4, then 1
Usual time of cooking (time)	RiskTriang(18,19,21)	RiskTriang(20,22,24)	RiskTriang(16,18,19)	if AA > 20 OR AA < 18, then 1

[3] Sedentary lifestyle (1.40-1.69 PAL), moderately active lifestyle (1.70-1.99 PAL), vigorously active lifestyle (2.00-2.40 PAL) and extremely active lifestyle (>2.40 PAL). Source: www.fao.org/docrep/007/y5686e/y5686e07.htm (accessed on 15.01.2010).

[4] A general guideline for adults is 7 to 9 hours of sleep each night. Source: the National Institutes of Health.

[5] Pollard B, et al (2008). *Breast mass detection under increased ambient lighting.* Proceedings of the 9th international workshop on Digital Mammography.

[6] Campbell S and Dawson D (1990). *Enhancement of nighttime alertness and performance with bright ambient light.* Physiology and Behavior 48(2): 317-20.

[7] Bedroom - sleep disturbance and annoyance if >30 dB. Source: www.engineeringtoolbox.com/decibel-dba-levels-d_728.html (accessed on 16.01.2010).

Appendix A4 Dataset 2

(AA = Actual Activity)

PAM observed behaviour	When "normal" (*N*)	When "manic" (*M*)	When "depressed" (*D*)	Threshold level
Daily activity (PAL)	RiskTriang(1.7,1.9,2.0)	RiskTriang(1.9,2.2,2.5)	RiskTriang(1.0,1.4,1.8)	if AA > 2.0 OR AA < 1.7, then 1
Earliest time of leaving home (time)	RiskTriang(7,8,9)	RiskTriang(5,6,8)	RiskTriang(8,9,11)	if AA < 7 OR AA > 9, then 1
Latest time of getting back home (time)	RiskTriang(18,19,21)	RiskTriang(20,23,24)	RiskTriang(16,17,19)	if AA > 21 OR AA < 18, then 1
TV remote keypresses (number)	RiskTriang(20,40,60)	RiskTriang(40,75,100)	RiskTriang(0,15,30)	if AA > 60 OR AA < 20, then 1
Time spent in bed (hour)	RiskTriang(5,7,9)	RiskTriang(2,4,6)	RiskTriang(8,12,14)	if AA < 5 OR AA > 9, then 1
Light level between 11pm to 7am (lux)	RiskTriang(5,10,20)	RiskTriang(15,40,70)	RiskTriang(1,4,10)	if AA > 20 OR AA < 5, then 1
Noise level between 11pm to 7am (dB)	RiskTriang(15,20,25)	RiskTriang(20,30,40)	RiskTriang(5,10,20)	if AA > 25 OR AA < 15, then 1
Time spent talking on the phone (minute)	RiskTriang(10,25,50)	RiskTriang(35,70,100)	RiskTriang(0,10,20)	if AA > 50 OR AA < 10, then 1
Number of daily phone calls (number)	RiskTriang(2,5,7)	RiskTriang(5,10,16)	RiskTriang(0,1,4)	if AA > 7 OR AA < 2, then 1
Time being outside btwn 5pm & 1am (hour)	RiskTriang(1,2,4)	RiskTriang(3,5,8)	RiskTriang(0,1,2)	if AA > 4 OR AA < 1, then 1
Cupboard doors usage (number)	RiskTriang(8,10,14)	RiskTriang(10,16,22)	RiskTriang(2,4,10)	if AA > 14 OR AA < 8, then 1
Fridge doors usage (number)	RiskTriang(6,8,10)	RiskTriang(8,14,18)	RiskTriang(2,4,8)	if AA > 10 OR AA < 6, then 1
Microwave door usage (number)	RiskTriang(4,6,8)	RiskTriang(6,10,14)	RiskTriang(0,2,6)	if AA > 8 OR AA < 4, then 1
Usual time of cooking (time)	RiskTriang(18,19,20)	RiskTriang(19,22,24)	RiskTriang(16,18,19)	if AA > 20 OR AA < 18, then 1

Appendix A5 Model outputs

A5.1 Dataset 1, Decision rules 1 and Technical variation 1

Patient type	Descriptive Statistics	ODE	OME	TP	FP	FN	TN
Patient 1	*Min*	2	2	119	0	52	325
	Max	46	27	166	3	99	328
	Mean	30.72	15.58	143.03	0.88	74.97	327.12
	SD	13.45	8.14	7.37	0.83	7.37	0.83
Patient 2	*Min*	0	1	80	0	93	324
	Max	57	27	125	4	138	328
	Mean	40.26	18.67	102.87	0.18	118.01	327.82
	SD	6.59	7.75	7.92	0.53	7.92	0.53
Patient 3	*Min*	7	1	73	0	96	325
	Max	55	27	122	3	145	328
	Mean	41.63	19.86	97.34	0.10	120.66	327.91
	SD	3.09	7.26	7.42	0.24	7.42	0.24
Patient 4	*Min*	0	0	95	0	59	320
	Max	45	26	159	8	123	328
	Mean	33.23	21.14	124.08	0.86	93.93	327.14
	SD	11.81	6.63	8.65	1.17	8.65	1.17
Patient 5	*Min*	2	2	90	0	71	325
	Max	53	26	147	3	128	328
	Mean	37.92	11.18	116.96	0.25	101.04	327.75
	SD	9.66	7.35	8.39	0.58	8.39	0.58
Patient 6	*Min*	3	1	97	0	65	325
	Max	48	25	153	3	121	328
	Mean	38.40	7.53	125.90	0.32	92.10	327.68
	SD	8.60	5.57	8.50	0.68	8.50	0.68
Patient 7	*Min*	2	0	121	0	53	325
	Max	45	26	165	3	97	328
	Mean	26.29	14.96	144.25	0.84	73.75	327.16
	SD	14.05	8.13	7.43	0.85	7.43	0.85
Patient 8	*Min*	5	1	74	0	97	324
	Max	55	25	131	4	144	328
	Mean	42.61	13.31	97.64	0.05	120.36	327.95
	SD	2.58	7.94	6.55	0.28	6.55	0.28
Patient 9	*Min*	4	1	81	0	92	324
	Max	62	26	126	4	137	328
	Mean	41.06	19.83	101.97	0.11	116.03	327.89
	SD	5.50	7.21	6.73	0.43	6.73	0.43
Patient 10	*Min*	4	0	89	0	86	321
	Max	51	26	132	7	129	328
	Mean	40.62	17.86	107.89	0.22	110.11	327.78
	SD	4.65	7.99	6.66	0.63	6.66	0.63
Patient 11	*Min*	0	0	145	0	34	323
	Max	43	25	184	5	73	328
	Mean	18.82	6.33	165.66	1.47	52.34	326.53
	SD	13.03	4.78	6.57	0.91	6.57	0.91
Patient 12	*Min*	0	0	148	0	32	322
	Max	43	24	186	6	70	328
	Mean	20.83	4.81	166.60	1.47	51.40	326.53
	SD	13.49	3.39	6.67	0.97	6.67	0.97

Patient 13	*Min*	0	0	153	0	25	316
	Max	41	25	193	12	65	328
	Mean	10.99	7.26	171.10	2.3	46.90	325.70
	SD	9.03	5.63	7.03	1.41	7.03	1.41
Patient 14	*Min*	0	0	115	0	57	322
	Max	45	25	161	6	103	328
	Mean	33.35	7.15	137.02	0.67	80.98	327.33
	SD	10.91	5.49	8.65	0.99	8.65	0.99
Patient 15	*Min*	0	0	127	0	40	319
	Max	43	25	178	9	91	328
	Mean	21.88	9.71	153.48	1.83	64.52	326.17
	SD	14.12	6.99	7.36	1.64	7.36	1.64
Patient 16	*Min*	0	0	98	0	67	322
	Max	46	26	151	6	120	328
	Mean	32.84	20.85	124.88	0.81	93.12	327.19
	SD	12.50	6.96	8.89	1.11	8.89	1.11
Patient 17	*Min*	0	0	133	0	39	322
	Max	43	25	179	6	85	328
	Mean	27.31	4.28	157.05	1.05	60.95	326.95
	SD	13.86	2.90	7.30	1.08	7.30	1.08
Patient 18	*Min*	1	0	139	0	30	323
	Max	43	25	188	5	79	328
	Mean	16.73	6.54	165.20	1.39	52.80	326.62
	SD	12.00	4.85	7.02	0.90	7.02	0.90
Patient 19	*Min*	0	0	113	0	41	319
	Max	3	25	177	9	105	328
	Mean	33.03	7.06	140.56	1.00	77.44	327.00
	SD	12.50	5.21	8.58	1.24	8.58	1.24
Patient 20	*Min*	0	0	162	0	13	313
	Max	40	13	206	15	56	328
	Mean	8.10	3.12	185.77	2.79	32.23	325.21
	SD	6.73	1.80	6.65	1.59	6.65	1.59
Patient 21	*Min*	0	0	164	0	16	313
	Max	40	17	202	15	54	328
	Mean	8.37	4.01	182.36	3.45	35.64	324.55
	SD	7.12	2.65	7.09	2.02	7.09	2.02
Patient 22	*Min*	0	0	165	0	14	313
	Max	41	21	204	15	53	328
	Mean	8.17	3.54	183.47	4.26	34.53	323.74
	SD	6.82	2.40	6.81	2.54	6.81	2.54
Patient 23	*Min*	0	0	147	0	26	305
	Max	41	25	192	23	71	328
	Mean	15.97	4.70	168.37	3.90	49.63	324.10
	SD	12.33	3.65	6.85	2.68	6.85	2.68
Patient 24	*Min*	0	0	160	0	11	316
	Max	41	16	207	12	58	328
	Mean	11.14	3.43	179.93	3.02	38.07	324.98
	SD	9.33	2.16	6.53	1.97	6.53	1.97
Patient 25	*Min*	0	0	179	0	2	258
	Max	24	8	216	70	39	328
	Mean	5.90	2.64	190.71	6.94	27.29	321.06
	SD	3.34	1.50	6.71	9.26	6.71	9.26

A5.2 Dataset 2, Decision rules 1 and Technical variation 1

Patient type	Descriptive Statistics	ODE	OME	TP	FP	FN	TN
Patient 1	*Min*	2	2	101	0	65	326
	Max	46	28	153	2	117	328
	Mean	34.15	22.23	128.23	0.68	89.77	327.32
	SD	11.82	5.66	8.40	0.79	8.40	0.79
Patient 2	*Min*	1	0	52	0	105	326
	Max	58	28	113	2	166	328
	Mean	42.04	21.17	87.22	0.09	130.78	327.91
	SD	5.71	6.67	8.40	0.34	8.40	0.34
Patient 3	*Min*	5	3	24	0	136	326
	Max	72	31	82	2	194	328
	Mean	46.13	25.37	49.71	0.01	168.29	327.99
	SD	5.88	1.70	8.70	0.14	8.70	0.14
Patient 4	*Min*	1	0	79	0	74	321
	Max	48	27	144	7	139	328
	Mean	38.45	23.46	104.23	0.40	113.77	327.60
	SD	8.33	4.60	8.91	0.75	8.91	0.75
Patient 5	*Min*	3	2	70	0	98	326
	Max	55	30	120	2	148	328
	Mean	40.98	20.74	96.53	0.13	121.47	327.87
	SD	6.98	6.75	8.20	0.40	8.20	0.40
Patient 6	*Min*	4	4	53	0	113	326
	Max	52	32	105	2	165	328
	Mean	42	24.69	79	0.06	138.99	327.94
	SD	3.43	2.42	8.33	0.29	8.33	0.29
Patient 7	*Min*	2	2	87	0	74	325
	Max	47	27	144	3	131	328
	Mean	36.67	23.65	119.60	0.49	98.40	327.51
	SD	10.10	4.03	8.92	0.75	8.92	0.75
Patient 8	*Min*	39	4	32	0	130	326
	Max	130	31	88	2	186	328
	Mean	45.66	24.82	2.17	0.01	155.83	327.99
	SD	5.71	1.93	8.44	0.03	8.44	0.03
Patient 9	*Min*	2	2	60	0	112	325
	Max	68	29	111	3	162	328
	Mean	43.31	23.66	81.76	0.04	138.09	327.96
	SD	3.58	4.04	8.21	0.23	8.21	0.23
Patient 10	*Min*	11	8	44	0	121	326
	Max	54	30	97	2	174	328
	Mean	42.82	25.06	67.29	0.01	150.71	327.98
	SD	2.55	1.22	8.48	0.15	8.48	0.15
Patient 11	*Min*	1	1	131	0	46	323
	Max	43	26	172	5	87	328
	Mean	25.13	12.24	153.25	1.15	64.75	326.85
	SD	13.97	7.76	6.43	0.91	6.43	0.91
Patient 12	*Min*	2	2	123	0	56	325
	Max	43	27	162	3	95	328
	Mean	31.12	19.21	142.72	0.90	75.28	327.10
	SD	12.58	7.63	6.69	0.84	6.69	0.84
Patient 13	*Min*	1	1	136	0	43	321
	Max	42	26	175	7	82	328
	Mean	20.73	15.76	156.32	1.45	61.68	326.55
	SD	13.50	7.97	5.73	0.84	5.73	0.95
Patient 14	*Min*	3	1	86	0	84	324
	Max	50	26	134	4	132	328
	Mean	39.01	17.85	110.09	0.22	107.91	327.79
	SD	8.07	7.77	7.67	0.57	7.67	0.57

Patient 15	*Min*	0	0	112	0	56	322
	Max	6	25	162	8	106	328
	Mean	32.02	15.16	138.52	0.94	81.48	327.06
	SD	12.69	8.16	8.17	1.15	8.16	1.15
Patient 16	*Min*	1	1	73	0	84	323
	Max	47	27	134	5	145	328
	Mean	38.43	23.55	103.78	0.37	114.22	327.63
	SD	8.26	4.32	9.58	0.74	9.58	0.74
Patient 17	*Min*	2	2	100	0	6	325
	Max	46	26	156	3	118	328
	Mean	36.35	17.83	125.44	0.40	92.56	327.60
	SD	10.21	7.72	8.00	0.69	8.00	0.69
Patient 18	*Min*	2	1	122	0	48	325
	Max	43	26	170	3	96	328
	Mean	29.04	15.20	145.81	0.90	72.19	327.10
	SD	13.75	8.16	7.14	0.87	7.14	0.87
Patient 19	*Min*	2	1	89	0	87	325
	Max	47	26	131	3	129	328
	Mean	40.23	21.49	109.59	0.18	100.41	327.82
	SD	5.28	6.09	6.60	0.52	6.60	0.52
Patient 20	*Min*	0	0	148	0	27	323
	Max	42	25	191	5	70	328
	Mean	12.77	7.58	168.50	1.66	49.50	326.34
	SD	10.05	5.62	6.47	0.86	6.47	0.86
Patient 21	*Min*	0	0	148	0	27	320
	Max	42	25	191	10	70	328
	Mean	14.45	8.37	166.88	1.97	51.12	326.03
	SD	11.13	6.15	6.69	1.18	6.69	1.18
Patient 22	*Min*	0	0	145	0	37	321
	Max	41	25	181	7	73	328
	Mean	16.77	13.43	160.88	1.64	57.12	326.36
	SD	12.01	7.89	5.52	0.87	5.52	0.87
Patient 23	*Min*	0	0	121	0	47	322
	Max	43	25	171	6	97	328
	Mean	27.21	12.43	145.87	1.19	72.13	326.01
	SD	11.11	7.89	7.64	1.20	7.64	1.20
Patient 24	*Min*	1	0	135	0	46	324
	Max	42	25	176	4	87	328
	Mean	21.42	12.74	154.30	1.13	63.70	326.87
	SD	13.95	7.75	6.37	0.88	6.37	0.88
Patient 25	*Min*	0	0	153	0	22	317
	Max	41	25	196	11	65	328
	Mean	11.33	7.49	170.46	2.28	47.54	325.72
	SD	9.41	5.82	6.82	1.35	6.82	1.35

A5.3 Dataset 1, Decision rules 2 and Technical variation 1

Patient type	Descriptive Statistics	ODE	OME	TP	FP	FN	TN
Patient 1	*Min*	0	0	82	0	90	327
	Max	44	26	128	1	136	328
	Mean	25.04	12.93	104.48	0.02	113.53	327.98
	SD	14.44	8.46	7.60	0.13	7.60	0.13
Patient 2	*Min*	0	0	47	0	126	326
	Max	51	27	92	2	171	328
	Mean	36.64	15.92	70.15	0.10	147.85	327.90
	SD	10.28	8.43	6.90	0.32	6.9	0.32
Patient 3	*Min*	0	0	49	0	129	327
	Max	51	26	89	1	169	328
	Mean	39.89	17.37	67.35	0.03	150.65	327.97
	SD	6.03	7.92	6.44	0.17	6.44	0.17
Patient 4	*Min*	0	0	62	0	98	324
	Max	44	26	120	4	156	328
	Mean	27.40	17.83	88.01	0.62	150.65	327.38
	SD	14.19	7.87	7.14	0.81	6.44	0.81
Patient 5	*Min*	1	0	69	0	109	327
	Max	52	25	109	1	149	328
	Mean	33.10	9.35	87.45	0.01	128.55	327.99
	SD	13.01	7.44	6.36	0.01	6.36	0.01
Patient 6	*Min*	0	0	80	0	96	325
	Max	49	24	122	3	138	328
	Mean	33.65	5.96	98.65	0.14	119.35	327.86
	SD	11.96	5.52	6.60	0.38	6.60	0.38
Patient 7	*Min*	0	0	86	0	85	326
	Max	44	26	133	2	132	328
	Mean	21.74	12.41	109.47	0.04	108.54	327.86
	SD	14.42	8.3	7.21	0.19	7.21	0.38
Patient 8	*Min*	1	0	57	0	123	326
	Max	56	26	95	2	161	328
	Mean	40.73	10.46	75.88	0.06	142.13	327.94
	SD	5.56	7.81	5.81	0.25	5.81	0.25
Patient 9	*Min*	1	0	55	0	122	327
	Max	54	26	96	1	163	328
	Mean	38.31	17.07	74.57	0.02	143.43	327.98
	SD	8.58	8.17	6.28	0.12	6.28	0.12
Patient 10	*Min*	0	0	62	0	116	324
	Max	49	26	102	4	156	328
	Mean	37.79	14.57	81.15	0.24	136.85	327.76
	SD	8.47	8.75	6.43	0.53	6.43	0.53
Patient 11	*Min*	0	0	116	0	59	325
	Max	42	24	154	3	102	328
	Mean	14.03	4.85	140.18	0.26	77.82	327.74
	SD	11.72	4.62	6.34	0.53	6.34	0.53
Patient 12	*Min*	0	0	121	0	58	325
	Max	42	21	160	3	97	328
	Mean	16.59	4.16	140.17	0.39	77.83	327.61
	SD	12.80	3.64	6.49	0.61	6.49	0.61
Patient 13	*Min*	0	0	124	0	48	323
	Max	40	24	170	5	94	328
	Mean	8.93	5.58	152.28	1.07	65.72	326.93
	SD	8.28	5.23	6.12	1.09	6.12	1.04
Patient 14	*Min*	0	0	93	0	89	323
	Max	46	24	129	5	125	328
	Mean	30.56	6.00	110.98	0.41	107.02	327.59
	SD	13.39	5.43	6.08	0.68	6.08	0.68

Patient 15	*Min*	0	0	104	0	71	321
	Max	42	24	147	7	114	328
	Mean	17.86	8.04	125.00	1.37	93	326.63
	SD	13.57	6.87	6.49	1.17	6.49	1.17
Patient 16	*Min*	0	0	67	0	103	324
	Max	46	26	115	4	151	328
	Mean	26.88	18.06	88.29	0.60	129.71	327.40
	SD	14.41	8.13	7.10	0.78	7.10	0.78
Patient 17	*Min*	0	0	110	0	72	324
	Max	43	24	146	4	108	328
	Mean	21.50	3.22	128.24	0.45	89.76	327.55
	SD	14.33	3.20	6.40	0.68	6.40	0.68
Patient 18	*Min*	0	0	116	0	61	325
	Max	41	24	157	3	102	328
	Mean	13.27	5.24	138.99	0.20	79.02	327.81
	SD	11.42	4.66	6.54	0.46	6.54	0.46
Patient 19	*Min*	0	0	92	0	83	322
	Max	45	24	135	6	126	328
	Mean	28.58	5.66	115.73	0.86	102.27	327.14
	SD	13.87	5.23	6.00	0.98	6.00	0.98
Patient 20	*Min*	0	0	148	0	30	321
	Max	40	12	188	7	70	328
	Mean	6.76	2.29	163.89	1.45	54.11	326.55
	SD	6.04	2.01	6.08	1.30	6.08	1.30
Patient 21	*Min*	0	0	143	0	39	321
	Max	40	22	179	7	75	328
	Mean	6.43	3.26	161.50	2.05	56.50	325.95
	SD	6.41	3.04	6.06	1.49	6.06	1.49
Patient 22	*Min*	0	0	141	0	38	317
	Max	40	15	180	11	77	328
	Mean	6.98	2.75	161.85	3.00	56.15	325.00
	SD	6.94	2.45	5.86	1.90	5.86	1.90
Patient 23	*Min*	0	0	120	0	58	316
	Max	42	24	160	12	98	328
	Mean	13.12	3.80	142.36	3.10	75.64	324.90
	SD	11.40	3.62	6.22	1.94	6.22	1.94
Patient 24	*Min*	0	0	136	0	42	319
	Max	40	17	176	9	82	328
	Mean	9.11	2.61	155.12	1.76	62.88	326.24
	SD	8.61	2.33	6.23	1.49	6.23	1.49
Patient 25	*Min*	0	0	152	0	31	314
	Max	28	18	187	14	66	328
	Mean	5.18	2.08	170.64	4.93	47.36	323.07
	SD	4.98	1.99	5.60	2.44	5.60	2.44

A5.4 Dataset 1, Decision rules 1 and Technical variation 2

Patient type	Descriptive Statistics	ODE	OME	TP	FP	FN	TN
Patient 1	*Min*	2	2	119	0	53	325
	Max	43	27	165	3	99	328
	Mean	37.13	21.57	143.41	1.01	74.59	326.99
	SD	9.08	6.21	6.28	0.83	6.28	0.83
Patient 2	*Min*	0	0	81	0	86	323
	Max	51	27	132	5	137	328
	Mean	36.80	15.02	106.13	0.46	111.88	327.54
	SD	10.08	8.30	8.42	0.89	8.42	0.89
Patient 3	*Min*	3	0	70	0	91	325
	Max	57	27	127	3	148	328
	Mean	41.35	16.66	96.94	0.15	121.06	327.85
	SD	4.84	8.23	8.24	0.50	8.24	0.50
Patient 4	*Min*	0	0	97	0	66	297
	Max	43	27	152	31	121	328
	Mean	19.44	15.34	125.25	7.43	92.76	320.57
	SD	14.06	8.73	9.45	4.83	9.15	4.83
Patient 5	*Min*	2	1	92	0	70	324
	Max	52	25	148	4	126	328
	Mean	36.55	11.30	119.50	0.31	98.50	327.69
	SD	10.54	7.43	8.94	0.64	8.94	0.64
Patient 6	*Min*	0	0	77	0	103	325
	Max	50	24	115	3	141	328
	Mean	32.17	6.30	97.73	0.23	120.27	327.77
	SD	12.94	5.57	6.77	0.48	6.77	0.48
Patient 7	*Min*	0	0	81	0	90	325
	Max	45	24	128	3	137	328
	Mean	17.30	10.68	104.35	0.29	113.65	327.71
	SD	12.91	7.90	7.53	0.55	7.53	0.55
Patient 8	*Min*	6	0	78	0	95	322
	Max	57	26	123	6	140	328
	Mean	42.28	12.84	98.73	0.07	119.27	327.93
	SD	3.32	7.96	6.82	0.4	6.82	0.40
Patient 9	*Min*	1	0	75	0	90	322
	Max	58	26	128	6	143	328
	Mean	39.67	17.99	101.72	0.15	116.28	327.85
	SD	8.38	8.01	7.97	0.50	7.97	0.50
Patient 10	*Min*	0	0	85	0	90	318
	Max	51	26	128	10	133	328
	Mean	39.32	15.36	106.38	1.00	111.62	327
	SD	7.68	8.39	7.25	1.55	7.25	1.55
Patient 11	*Min*	0	0	148	0	26	318
	Max	42	23	192	10	70	328
	Mean	12.80	5.11	171.93	2.19	4.07	325.82
	SD	10.28	3.83	7.28	1.54	7.28	1.54
Patient 12	*Min*	0	0	146	0	27	315
	Max	42	24	191	13	72	328
	Mean	13.59	4.24	171.34	2.21	46.66	325.79
	SD	10.63	2.99	7.12	1.78	7.12	1.78
Patient 13	*Min*	0	0	159	1	12	282
	Max	39	24	206	46	59	327
	Mean	6.27	4.26	182.37	12.11	35.61	315.89
	SD	5.87	3.72	7.70	6.17	7.70	6.17
Patient 14	*Min*	0	0	111	0	46	318
	Max	45	25	172	10	107	328
	Mean	28.62	5.95	142.75	1.62	75.25	326.38
	SD	13.94	4.69	8.83	1.76	8.83	1.76

Patient 15	*Min*	0	0	134	1	28	289
	Max	45	25	190	39	84	327
	Mean	11.09	5.92	162.61	13.23	55.39	314.77
	SD	10.30	5.10	8.79	6.60	8.79	6.76
Patient 16	*Min*	0	0	88	0	64	300
	Max	48	29	154	28	130	328
	Mean	20.42	14.95	125.18	7.03	92.82	320.97
	SD	14.41	8.70	10.01	4.76	10.01	4.76
Patient 17	*Min*	0	0	131	0	36	319
	Max	43	22	182	9	87	328
	Mean	24.03	4.26	159.85	1.37	58.15	326.63
	SD	14.18	2.94	7.39	1.37	7.39	1.37
Patient 18	*Min*	0	0	147	0	20	320
	Max	42	25	198	8	71	328
	Mean	12.28	5.47	169.10	1.8	48.90	326.20
	SD	9.90	4.17	7.79	1.43	7.79	1.43
Patient 19	*Min*	0	0	118	0	52	311
	Max	45	25	166	17	100	328
	Mean	29.99	6.08	142	3.01	76	324.99
	SD	13.66	4.92	8.74	2.68	8.74	2.68
Patient 20	*Min*	0	0	168	0	7	306
	Max	39	17	211	22	50	328
	Mean	6.27	2.83	190.05	5.69	27.95	322.31
	SD	5.35	1.78	6.70	3.23	6.70	3.23
Patient 21	*Min*	0	0	172	3	4	280
	Max	32	21	214	48	46	325
	Mean	4.40	2.69	194.08	20.80	23.92	307.20
	SD	3.99	2.22	6.92	8.18	6.92	8.18
Patient 22	*Min*	0	0	166	6	4	266
	Max	32	13	214	62	52	322
	Mean	4.49	2.45	193.60	25.63	24.40	302.37
	SD	4.23	2.09	7.05	8.33	7.05	8.33
Patient 23	*Min*	0	0	156	5	16	269
	Max	41	22	202	59	62	323
	Mean	7.60	3.10	179.06	27.10	38.94	300.90
	SD	7.74	2.81	7.73	9.38	7.73	9.38
Patient 24	*Min*	0	0	164	0	12	296
	Max	41	13	206	32	54	328
	Mean	8.07	2.88	184.22	8.08	33.78	319.92
	SD	7.36	1.93	6.88	4.33	6.88	4.33
Patient 25	*Min*	0	0	177	5	2	269
	Max	25	11	216	59	41	323
	Mean	3.83	2.08	198.55	25.21	19.45	302.79
	SD	3.47	1.62	6.27	8.95	6.27	8.95

Appendix A6 VBA Pseudo-code

A6.1 Forming questions for “OK” command

```
Private Sub cmdOK_Click()
  Unload Me
End Sub

Private Sub UserForm_Initialize()
   ResetForm
End Sub

Public Function Ask(Optional ByVal Prompt As String, Optional ByVal Caption As String, _
Optional ByRef Cancel As Boolean) As String

   ResetForm
   Me.Caption = Caption
   lblPrompt.Caption = Prompt
   txtResponse.SetFocus
   Me.Show vbModal

   Cancel = mblnCancel
   If mblnCancel Then
      Ask = ""
   Else
      Ask = txtResponse.Text
   End If
End Function

Private Sub ResetForm()
   cmdOK.Caption = "OK"
   cmdCancel.Caption = "Cancel"
   Me.Caption = "Personalised Ambient Monitoring"
   lblPrompt.Caption = "Prompt:"
   mblnCancel = False
   With txtResponse
      .Text = ""
      .WordWrap = True
      .MultiLine = True
   End With
End Sub

Private Sub UserForm_QueryClose(Cancel As Integer, CloseMode As Integer)
  If (CloseMode = vbFormControlMenu) Then
      mblnCancel = True
   End If
End Sub
```

A6.2 Forming questions for “Cancel” command

```
Private mblnCancel As Boolean

Private Sub cmdCancel_Click()
   mblnCancel = True
   Unload Me
End Sub

Private Sub cmdOK_Click()
  Unload Me
End Sub

Private Sub UserForm_Initialize()
   ResetForm
End Sub

Public Function Ask(Optional ByVal Prompt As String, Optional ByVal Caption As String, _
Optional ByRef Cancel As Boolean) As String

   ResetForm
   Me.Caption = Caption
   lblPrompt.Caption = Prompt
   txtResponse.SetFocus
   Me.Show vbModal

   Cancel = mblnCancel
   If mblnCancel Then
      Ask = ""
   Else
      Ask = txtResponse.Text
   End If
End Function

Private Sub ResetForm()
   cmdOK.Caption = "OK"
   cmdCancel.Caption = "Cancel"
   Me.Caption = "Personalised Ambient Monitoring"
   lblPrompt.Caption = "Prompt:"
   mblnCancel = False
   With txtResponse
      .Text = ""
      .WordWrap = True
      .MultiLine = True
   End With
End Sub

Private Sub UserForm_QueryClose(Cancel As Integer, CloseMode As Integer)
  If (CloseMode = vbFormControlMenu) Then
      'End
      mblnCancel = True
   End If
End Sub
```

A6.3 Prodromal data storing

```
Sub Activity()
   Dim strInput_minNormal As String
   Dim strInput_modeNormal As String
   Dim strInput_maxNormal As String
   Dim strInput_minManic As String
   Dim strInput_modeManic As String
   Dim strInput_maxManic As String
   Dim strInput_minDepressed As String
   Dim strInput_modeDepressed As String
   Dim strInput_maxDepressed As String
   Dim strInput_thresholdMax As String
   Dim strInput_thresholdMin As String
   Dim blnCancel As Boolean
   Dim blnIsNumeric As Boolean

   ' MINIMUM ACTIVITY LEVEL WHEN NORMAL
   Do
      strInput_minNormal = frmQuestion.Ask("Minimum activity level when normal", "Min Normal", blnCancel)

      blnIsNumeric = IsNumeric(strInput_minNormal)
      Data.Cells(3, 3).Value = strInput_minNormal

      If Not blnCancel And Not blnIsNumeric Then
         MsgBox "You must provide a number", , "Min Normal"
      End If

      If blnCancel Then
         MsgBox "A default value # will be stored", , "Min Normal"
         Data.Cells(3, 3).Value = #
         Exit Do
      End If
   Loop Until blnIsNumeric Or blnCancel

   ' MAXIMUM ACTIVITY LEVEL WHEN NORMAL
   Do
      strInput_maxNormal = frmQuestion.Ask("Maximum activity level when normal", "Max Normal", blnCancel)

      blnIsNumeric = IsNumeric(strInput_maxNormal)
      Data.Cells(5, 3).Value = strInput_maxNormal

      If Not blnCancel And Not blnIsNumeric Then
         MsgBox "You must provide a number", , "Max Normal"
      End If

      If blnCancel Then
         MsgBox "A default value # will be stored", , "Max Normal"
         Data.Cells(5, 3).Value = #
         Exit Do
      End If
   Loop Until blnIsNumeric Or blnCancel
```

```
' MODE OF ACTIVITY LEVEL WHEN NORMAL
Do
   strInput_modeNormal = frmQuestion.Ask("Mode activity level when normal", "Mode Normal", blnCancel)

   blnIsNumeric = IsNumeric(strInput_modeNormal)
   Data.Cells(4, 3).Value = strInput_modeNormal

   If Not blnCancel And Not blnIsNumeric Then
      MsgBox "You must provide a number", , "Mode Normal"
   End If

   If blnCancel Then
      MsgBox "A default value # will be stored", , "Mode Normal"
      Data.Cells(4, 3).Value = #
      Exit Do
   End If
Loop Until blnIsNumeric Or blnCancel

' MINIMUM ACTIVITY LEVEL WHEN MANIC
Do
   strInput_minManic = frmQuestion.Ask("Minimum activity level when manic", "Min Manic", blnCancel)

   blnIsNumeric = IsNumeric(strInput_minManic)
   Data.Cells(9, 3).Value = strInput_minManic

   If Not blnCancel And Not blnIsNumeric Then
      MsgBox "You must provide a number", , "Min Manic"
   End If

   If blnCancel Then
      MsgBox "A default value # will be stored", , "Min Manic"
      Data.Cells(9, 3).Value = #
      Exit Do
   End If
Loop Until blnIsNumeric Or blnCancel

' MAXIMUM ACTIVITY LEVEL WHEN MANIC
Do
   strInput_maxManic = frmQuestion.Ask("Maximum activity level when manic", "Max Manic", blnCancel)

   blnIsNumeric = IsNumeric(strInput_maxManic)
   Data.Cells(11, 3).Value = strInput_maxManic

   If Not blnCancel And Not blnIsNumeric Then
      MsgBox "You must provide a number", , "Max Manic"
   End If

   If blnCancel Then
      MsgBox "A default value # will be stored", , "Max Manic"
      Data.Cells(11, 3).Value = #
      Exit Do
   End If
Loop Until blnIsNumeric Or blnCancel
```

```
' MODE ACTIVITY LEVEL WHEN MANIC
Do
   strInput_maxManic = frmQuestion.Ask("Mode activity level when manic", "Mode Manic", blnCancel)

   blnIsNumeric = IsNumeric(strInput_modeManic)
   Data.Cells(10, 3).Value = strInput_modeManic

   If Not blnCancel And Not blnIsNumeric Then
      MsgBox "You must provide a number", , "Mode Manic"
   End If

   If blnCancel Then
      MsgBox "A default value (2.2) will be stored", , "Mode Manic"
      Data.Cells(10, 3).Value = 2.2
      Exit Do
   End If
Loop Until blnIsNumeric Or blnCancel

' MINIMUM ACTIVITY LEVEL WHEN DEPRESSED
Do
   strInput_minDepressed = frmQuestion.Ask("Min. activity level when depressed", "Min Dep.", blnCancel)

   blnIsNumeric = IsNumeric(strInput_minDepressed)
   Data.Cells(15, 3).Value = strInput_minDepressed

   If Not blnCancel And Not blnIsNumeric Then
      MsgBox "You must provide a number", , "Min Depressed"
   End If

   If blnCancel Then
      MsgBox "A default value # will be stored", , "Min Depressed"
      Data.Cells(15, 3).Value = #
      Exit Do
   End If
Loop Until blnIsNumeric Or blnCancel

' MAXIMUM ACTIVITY LEVEL WHEN DEPRESSED
Do
   strInput_maxDepressed = frmQuestion.Ask("Max. activity level when depressed", "Max Dep.", blnCancel)

   blnIsNumeric = IsNumeric(strInput_maxDepressed)
   Data.Cells(17, 3).Value = strInput_maxDepressed

   If Not blnCancel And Not blnIsNumeric Then
      MsgBox "You must provide a number", , "Max Depressed"
   End If

   If blnCancel Then
      MsgBox "A default value # will be stored", , "Max Depressed"
      Data.Cells(17, 3).Value = #
      Exit Do
   End If
Loop Until blnIsNumeric Or blnCancel
```

```
  ' MODE ACTIVITY LEVEL WHEN DEPRESSED
  Do
    strInput_maxManic = frmQuestion.Ask("Mode activity level when depressed", "Mode Dep.", blnCancel)

    blnIsNumeric = IsNumeric(strInput_modeDepressed)
    Data.Cells(16, 3).Value = strInput_modeDepressed

    If Not blnCancel And Not blnIsNumeric Then
      MsgBox "You must provide a number", , "Mode Depressed"
    End If

    If blnCancel Then
      MsgBox "A default value # will be stored", , "Mode Depressed"
      Data.Cells(16, 3).Value = #
      Exit Do
    End If
  Loop Until blnIsNumeric Or blnCancel

End Sub
```

A6.4 Logic to work out the onset of bipolar episodes

```
Public Sub Decision_Rules()
   Dim i As Long
   Dim j As Integer

   Dim Count_ODE As Integer
   Dim Count_OME As Integer

   Count_ODE = 0
   Count_OME = 0

   'PAM SCORING SYSTEM COLUMN #
   j = 31

   'WORKING OUT THE FIRST DAY OF AN ONSET OF DEPRESSIVE EPISODE
   For i = 167 To 296
      If Not Actual.Cells(i, j + 1) = 0 Then
         Exit For
      End If

      If Actual.Cells(i, j + 1) = 0 Then
         Count_ODE = Count_ODE + 1
      End If
   Next i
   Results.Cells(3, 2).Value = "=RiskOutput() + " & Count_ODE

   'WORKING OUT THE FIRST DAY OF AN ONSET OF MANIC EPISODE
   For i = 461 To 548
      If Not Actual.Cells(i, j + 1) = 0 Then
         Exit For
      End If

      If Actual.Cells(i, j + 1) = 0 Then
         Count_OME = Count_OME + 1
      End If
   Next i
   Results.Cells(3, 3).Value = "=RiskOutput() + " & Count_OME
End Sub
```

Appendix A7 NHS Ethics Participant Information Sheet

PAM project

INFORMATION SHEET

What is PAM?

PAM - *Personalised Ambient Monitoring* - is a three year research project funded by the UK Engineering and Physical Sciences Research Council. PAM aims to provide new technology to help people with bipolar disorder to take control of their health care and lead more independent lives. Using a system of discreet and unobtrusive sensors, the aim is to help you identify the very early signs of an acute manic or depressive episode and thus hopefully avoid it. It is well known that early detection may be able to avoid both full-blown episodes and hospital admissions. For more detailed information, see the project website: www.pam-research.org

Who is doing this research?

This project builds on existing research in this area at the Universities of Southampton, Nottingham and Stirling. The PAM team consists of four senior academics: a biomedical signal processing engineer, Dr Christopher James, and a management scientist, Professor Sally Brailsford (both University of Southampton), a biomedical engineer, Professor John Crowe (University of Nottingham) and a computer scientist, Professor Evan Magill (University of Stirling). Each of these four academics has a PhD student funded as part of the project, making a multi-disciplinary team of eight in total. PAM also has a Steering Group which includes a service user, plus a clinical psychiatrist (Dr Paul Courtney of Hampshire Partnership Trust) and the Chief Executive of the mental health charity Solent MIND, Mr Richard Barritt.

How does PAM work?

PAM uses a system of unobtrusive small wearable and environmental sensors to monitor your personal daily behaviour patterns. PAM analyses the data from these to determine

your normal "activity signature", i.e. a picture (like a fingerprint) of your normal activity. PAM is not "Big Brother" – the type of data it collects is very basic and cannot identify individuals directly. Having established your normal baseline activity signature, the aim of PAM is to identify small changes, for example minor unexplained disruptions in your sleep or meal patterns, which you may not be aware of yourself but which may potentially herald the early signs of an acute episode. PAM will alert you to the change and then you can decide what to do: maybe nothing, or maybe contact your doctor, for example.

The project team wants to test the PAM system in a six-month trial involving volunteers with BD. The P in PAM stands for "Personalised", since the level of monitoring that each person will be comfortable with will be different. The system will allow you to adjust the monitoring to suit your individual preferences. You can switch individual sensors on or off, as you like, or even switch the whole system off.

How many participants are being recruited?

We are only planning to recruit four participants. This is not a clinical trial, but a feasibility study, to see if the PAM kit works and is acceptable to users.

What will happen if I agree to take part?

There are two stages to agreeing to participating in PAM. Initially, you would only be agreeing to a home visit by two members of the research team (a senior investigator and one of the PhD students). The aim of the home visit is to demonstrate the sensors and the rest of the PAM kit, and to answer all your questions about participation. There is absolutely no commitment to continue any further with the trial if at this point you decide PAM is not for you.

From the date you receive this information pack, you will be given two weeks to think it over, to obtain the letter from your GP and to return the consent form to us. If we do not hear back from you within two weeks of the postmark on this letter, we will make one further attempt to contact you, by email or telephone. If we cannot get hold of you, or if you tell us you have decided not to proceed, we will regard you as withdrawn from the study and we will delete all correspondence with you.

On receipt of the signed consent form and GP letter, the researchers will contact you by email or telephone and make an appointment for a home visit. At this first home visit, which could last between one and two hours, the research team will give a full verbal explanation of the nature of the involvement in PAM, and will demonstrate all the available sensors. They will answer any questions, and also explain that you can telephone number/email a member of the research team if you have further questions after this visit.

There will be a second potential exit point here if you now decide not to proceed. You will be given one week to think it over. After one week, the researchers will contact you again

by email or telephone, check if you wish to proceed, and if so make an appointment for a second home visit.

At the second home visit, which will be made by two or three members of the research team (one of the investigators and one or two of the PhD students), the main consent form will be signed (this is to PAM being installed). This will be followed by selection of the acceptable set of PAM sensors to be used, and then installation of the PAM kit and user training. This visit could take up to six hours. We will also conduct an entry interview at this time, to record your feelings about participating before it starts.

The daily routine for the next six months will involve charging and checking devices, putting on the wearable device in the morning and removing it in the evening, brief mood status checks (using the special mobile phone), and initiating uploading of data. This will take a maximum of 15 minutes per day. Reporting of faults in the system should they occur, plus any queries about the system or concerns about the sensors (for example, if you decide that you are no longer happy to use a particular sensor), may take slightly longer. You will be given a telephone number and email address which you should contact, and a response within 24 hours is guaranteed. There will also be a global "Off" switch which you can use at any stage, which will turn off all sensors. You will also have the right to withdraw at any stage during the study and have all your data removed and deleted.

At the end of the six months, or earlier if you decide to withdraw or has an acute bipolar episode, the PAM kit will be removed. This will be done at an agreed time and your house will be left in good order. There will be an exit interview at this point, when you will have an opportunity to comment on the experience.

Exactly what sensors can I choose?

PAM collects data from three types of source. Firstly, from sensors situated in the home that collect information on light levels, sound levels, movement information, and aspects of television usage. Secondly, from sensors that are worn by the individual that detect sound and light levels, movement, and position. Finally, using a mobile phone the system collects information from individuals on activities and mood.

We have minimised the level of identifiable information these sensors produce. The camera system does not record actual visual images but merely detects the presence or absence of moving objects, by comparing successive black and white images taken at 10-second intervals. The Passive Infra-Red (PIR) sensor works in a similar way, in exactly the same way that a domestic security light detects movement. Finally, the microphones do not record actual sounds or speech but merely record sound levels.

You can select any or all of the following list of sensors, and you will be shown how to turn them off at any point if you choose. All of these will be demonstrated and explained to you in plain language during the initial home visit.

- Wearable device with microphone (records sound features only not actual voice); also includes GPS, light sensor and accelerometer. This is comparable in size and weight to an iPod or mobile phone. Any of the sensors on it can be disabled.
- PIR (passive infra-red) devices: only record presence/absence of movement, like a household intruder detector
- Cameras (do not store images but merely presence/absence of moving objects)
- Ambient microphone (records sound features only not actual voice)
- Ambient light sensors: detect levels of daylight/artificial light
- Pressure mats (for detecting movements through doorways, thresholds, bedside mats etc): these will only be placed in locations where you give consent
- TV remote monitor (counts number of button presses only)
- Read switches (used on cupboard or fridge doors, etc, to detect when a door is opened or closed)
- Bluetooth encounters (a device on the PAM mobile phone to detect proximity to other devices using bluetooth protocol, e.g. mobile phones)

You will also be given a special mobile phone. You may continue to use your own mobile as well if you wish, but you could also choose to use the PAM phone as your sole mobile (at your own expense). The actual PAM services do not involve any cost to you, you would just pay for calls and other services e.g. internet. The PAM mobile is needed for storing and processing data from the wearable, and uploading it to the PC; and for brief mood and activity status requests. If you wish to use the PAM phone as your sole mobile, the research team will set this up for you and will ensure that it is working properly.

You will also be provided with a dedicated PC. This will require internet access. The PC will be provided and installed free of charge by the PAM team. It is needed for temporary local data storage and manipulation and onward transmission (in an encrypted format) to one of the participating universities for analysis.

Can I be personally identified as a result of taking part?

Absolutely not. You will be completely anonymous in any publications or reports produced in this study. If we need to refer to you individually, you will be assigned a letter, e.g. “Patient A”. If we wish to use a direct quotation from your interview, we will always ask your permission first and show you a draft of the text for your approval before we go ahead. Obviously the research team will know which data is yours. However there will be nothing transmitted or stored which would identify you personally to anyone outside the research team. Within six months after the end of the study, all identifiers (names, address, contact details) will be deleted. As experienced researchers, we are accustomed to dealing with confidential data and all our universities have facilities for secure storage of such data. All transmitted data will be encrypted, and the data will be kept on a secure password-protected server in a locked room.

Will I be able to find out the results of the trial?

At the end of the study, we plan to hold one final meeting with each participant, in a place which is mutually convenient (possibly your home, or one of the participating

universities) where we will report the findings of the study. At this stage you will have a final opportunity to reflect and comment on your experience.

What is the benefit to me in participating?

There is no direct benefit to you personally, but you would have the satisfaction of knowing that you had potentially helped to develop an intervention which could have a big impact on the quality of life of bipolar patients in the future. We also hope you would find the experience interesting.

What are the risks to me in participating?

There are no physical risks involved in using any of the PAM equipment. The clinician on our Steering Group has assured us that there is no medical reason why PAM should have an adverse effect on your mental health, but we would still like to be assured that your GP knows you are participating and is happy for you to do so. Your GP can contact one of the research team for more information if they like.

We do appreciate that different people will respond in different ways to being monitored, and therefore if you do find that you don't get on with PAM once you have started, we stress that you will be free to withdraw at any time, without needing to give any reason and without your legal rights being affected.

What happens if I should become ill during the study?

If you becomes so unwell during the study that you are unable to use the PAM equipment, you will be withdrawn from the study. PAM is not intended for use with patients who are actually suffering from acute manic or depressive episodes: it is intended to give people early warning of an impending episode. However clearly the data obtained prior to this episode would potentially be extremely valuable and so we would not wish to lose it. The possibility of this will be discussed with you in confidence at the outset and will form part of the consent form.

How can I find out more?

You could visit the PAM website www.pam-research.org. You could also contact one of the academic investigators, whose names, emails and telephone numbers are given below.

I want to participate. What should I do next?

You should sign and return the Stage 1 consent form enclosed with this Information Sheet using one of the stamped addressed envelopes provided. You should send (or give) the second copy of this information sheet to your GP and ask him/her to sign the GP consent form and either return it direct to us, or send it to you so that you can then forward it to us with your own Stage 1 consent form. Your GP can of course contact one of the research team for further information.

Is there an independent person I can contact if I am unhappy about the research team?

If you are unhappy about any aspect of the research and do not wish to raise it with the research team, you can contact the project sponsor, Dr Martina Prude, Head of Research Governance at the University of Southampton. Tel 023 8059 8848, email mad1@soton.ac.uk.

Researcher contact details

Professor Sally Brailsford	023 8059 3567	s.c.brailsford@soton.ac.uk
Dr Christopher James	023 8059 3043	c.james@soton.ac.uk
Professor John Crowe	0115 951 5590	john.crowe@nottingham.ac.uk
Professor Evan Magill	01786 46 7425	ehm@cs.stir.ac.uk

Appendix A8 Computer model

Appendix A9 Validation of behaviour

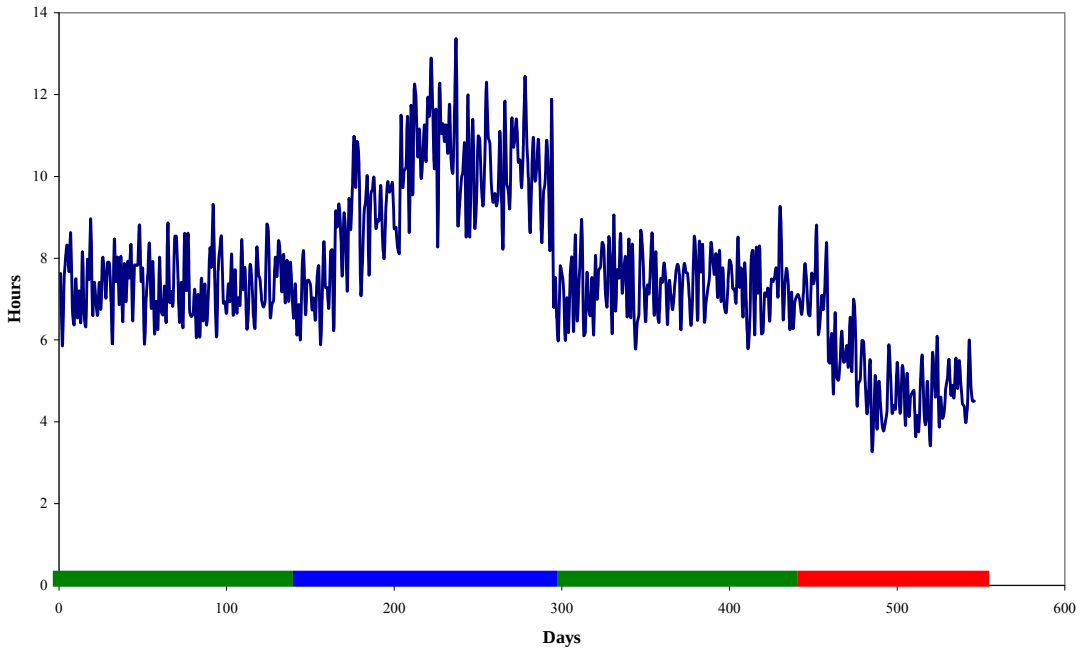


Figure A9.1 Time spent in bed during various mood states

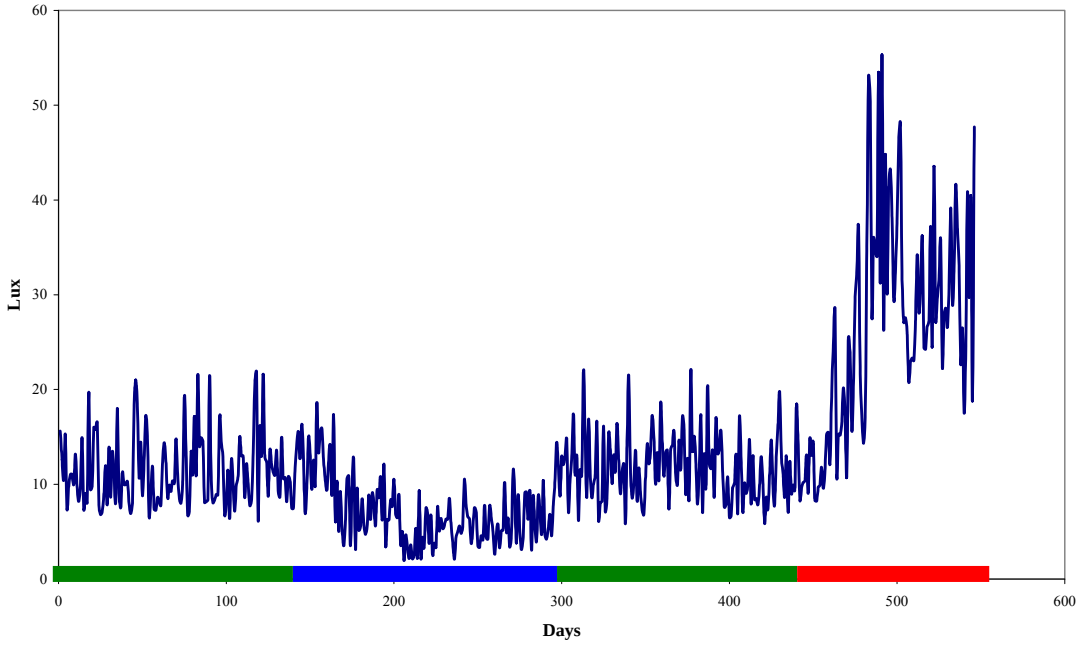


Figure A9.2 Light levels between 11pm and 7am during various mood states

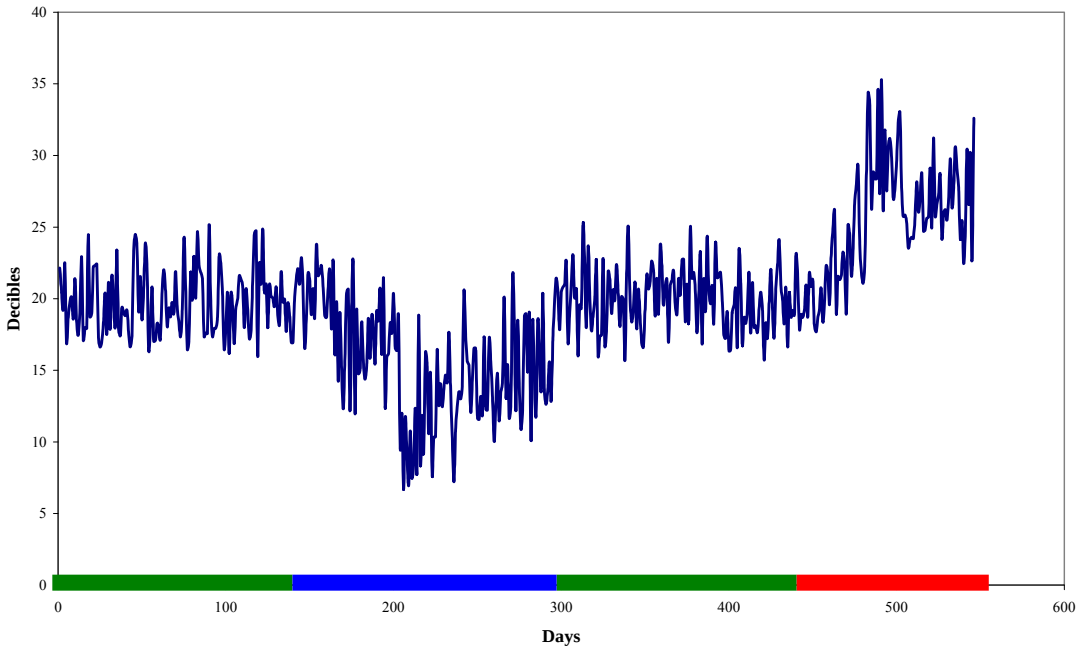


Figure A9.3 Noise levels between 11pm and 7am during various mood states

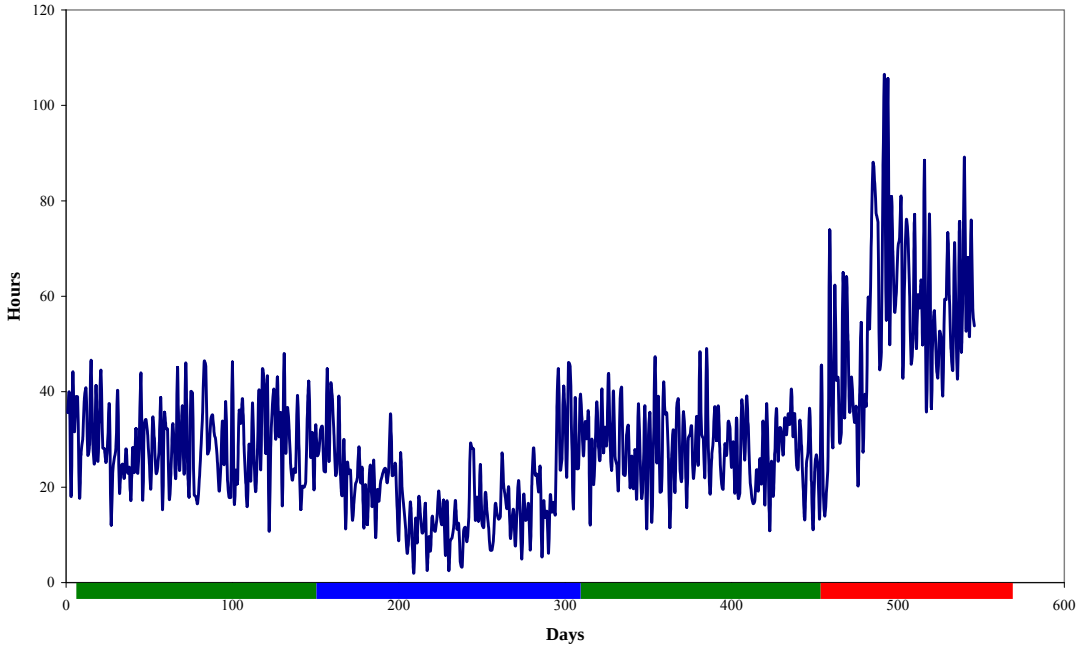


Figure A9.4 Time spent talking on the phone during various mood states

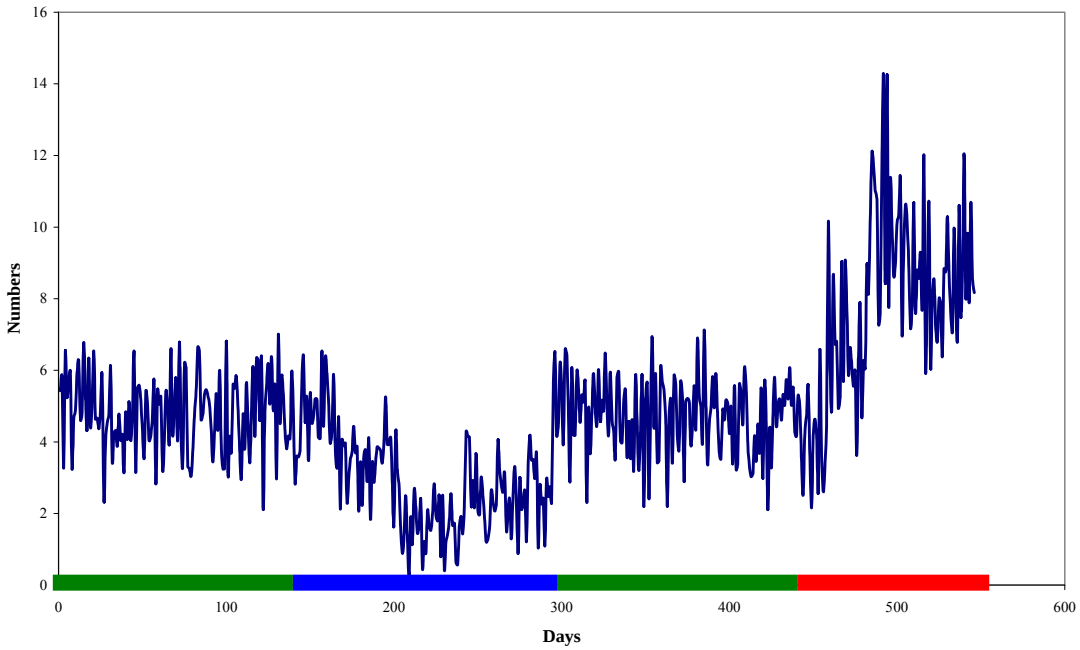


Figure A9.5 Number of daily phone calls during various mood states

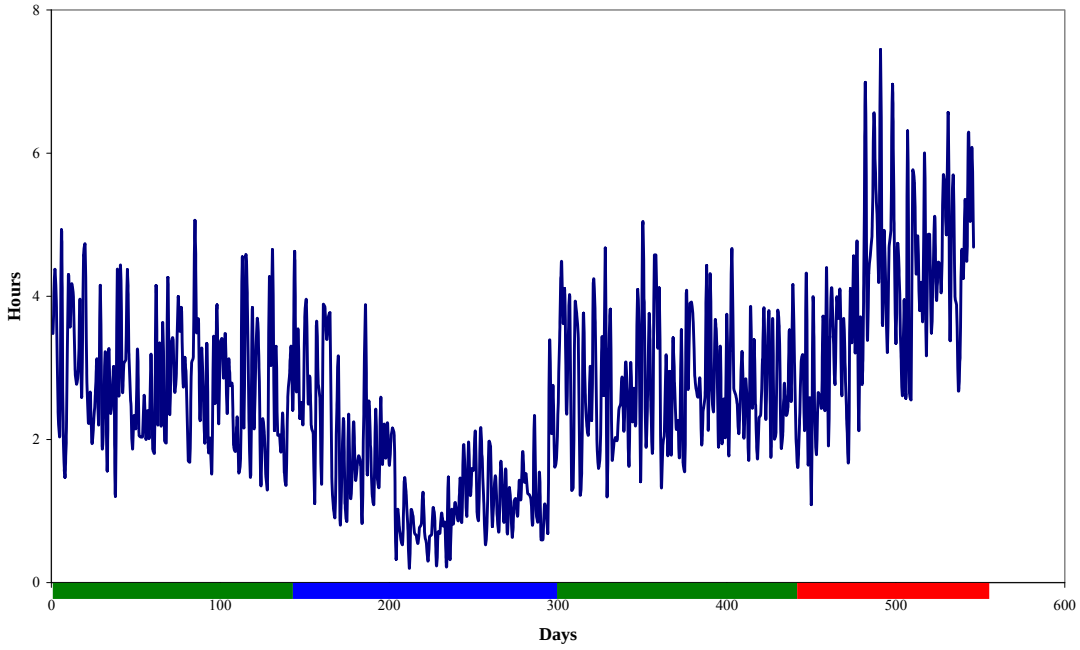


Figure A9.6 Time away from home between 5pm and 1am during various mood states

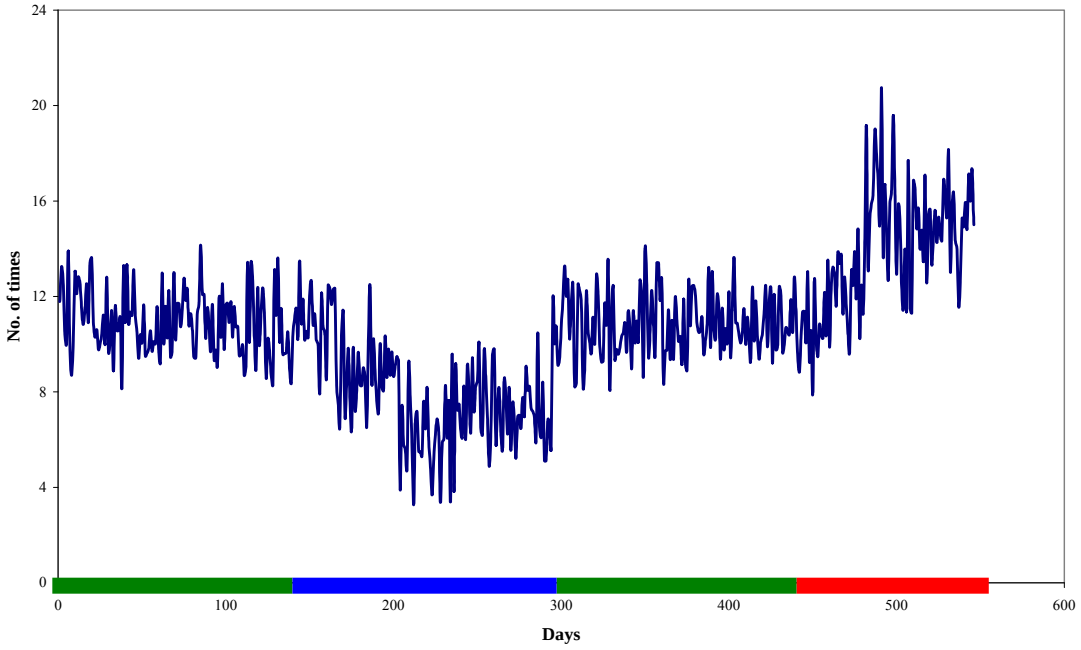


Figure A9.7 Cupboard door usage during various mood states

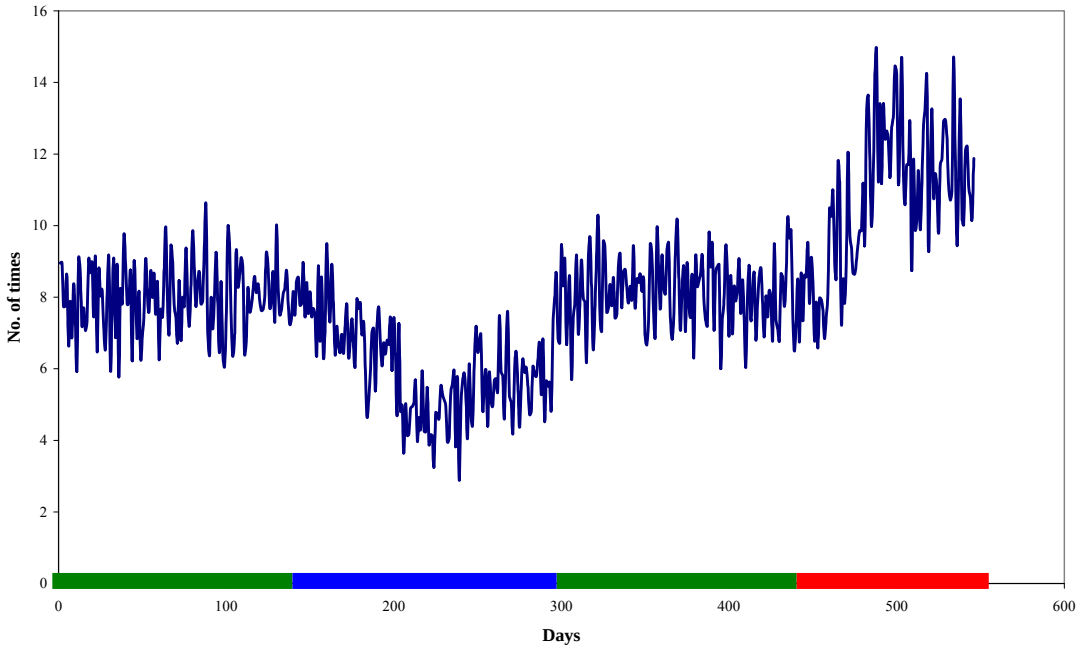


Figure A9.8 Fridge door usage during various mood states

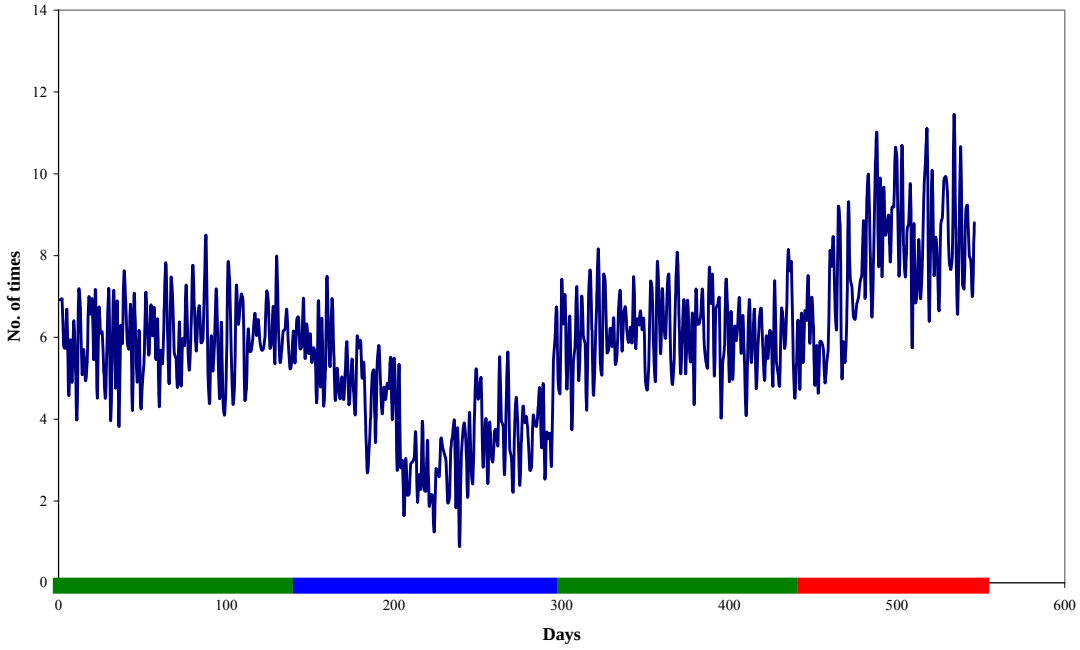


Figure A9.9 Microwave door usage during various mood states

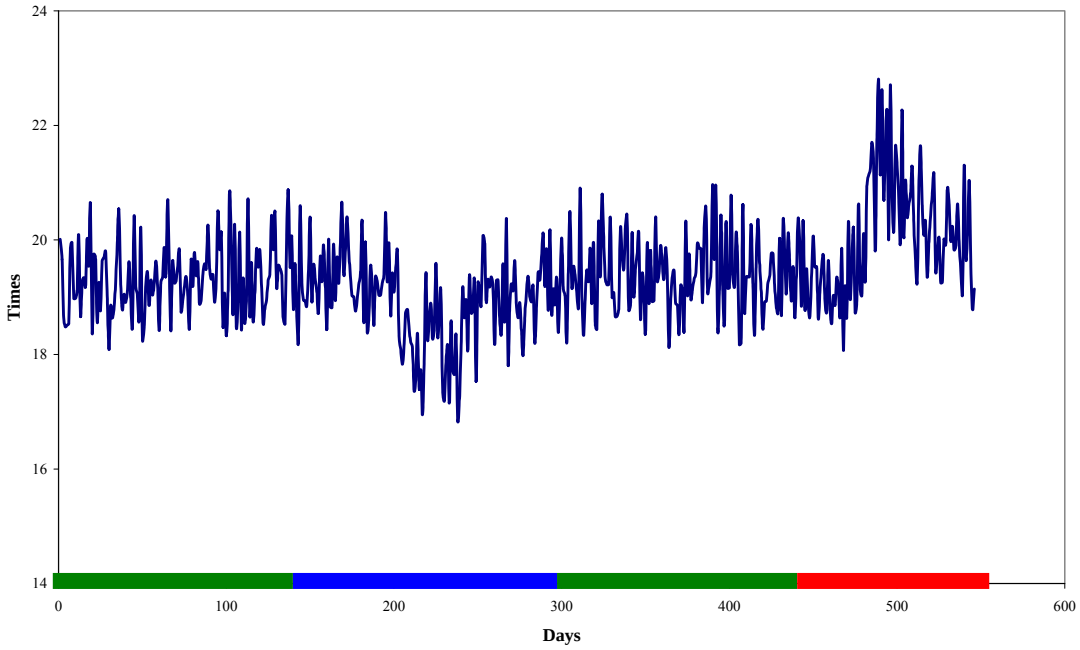
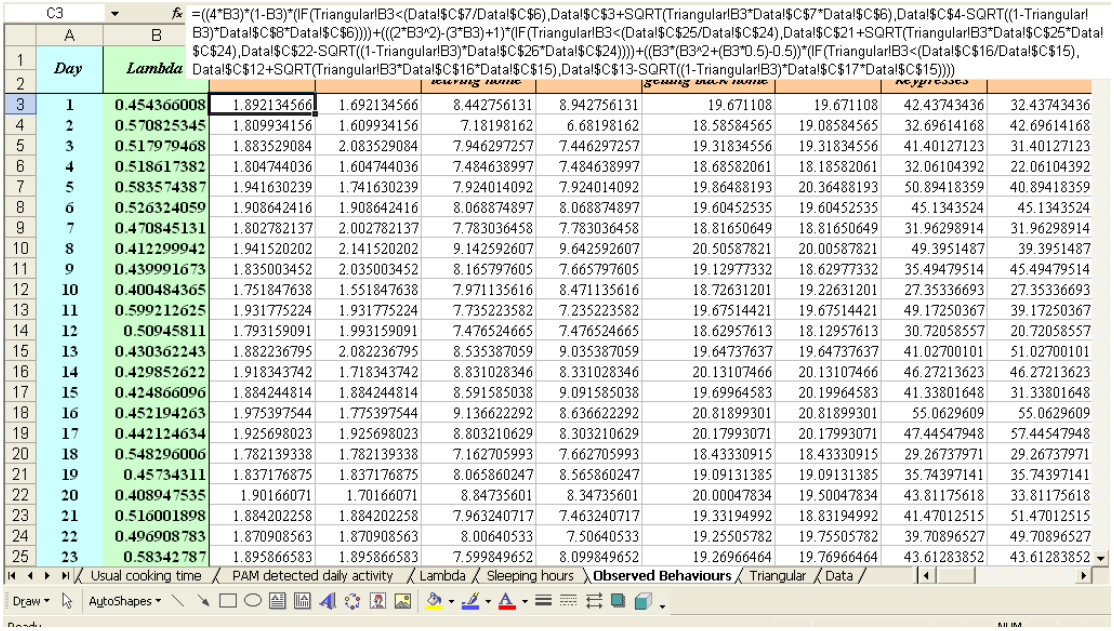


Figure A9.10 Usual time of cooking during various mood states

Appendix A10 Use of Excel Functions



C3 =((4*B3)*(1-B3)*(IF(TriangularIB3<(Data!$C$7/Data!$C$6),Data!$C$3+SQRT(TriangularIB3*Data!$C$7*Data!$C$6),Data!$C$4-SQRT((1-TriangularIB3)*Data!$C$8*Data!$C$6))))+(((2*B3^2)-(3*B3)+1)*(IF(TriangularIB3<(Data!$C$25/Data!$C$24),Data!$C$21+SQRT(TriangularIB3*Data!$C$25*Data!$C$24),Data!$C$22-SQRT((1-TriangularIB3)*Data!$C$26*Data!$C$24))))+((B3*(B3^2+(B3*0.5)-0.5))*(IF(TriangularIB3<(Data!$C$16/Data!$C$15),Data!$C$12+SQRT(TriangularIB3*Data!$C$16*Data!$C$15),Data!$C$13-SQRT((1-TriangularIB3)*Data!$C$17*Data!$C$15))))

	A	B								
1	*Day*	*Lambda*								
2										
3	1	0.454366008	1.892134566	1.692134566	8.442756131	8.942756131	19.671108	19.671108	42.43743436	32.43743436
4	2	0.570825345	1.809934156	1.609934156	7.18198162	6.68198162	18.58584565	19.08584565	32.69614168	42.69614168
5	3	0.517979468	1.883529084	2.083529084	7.946297257	7.446297257	19.31834556	19.31834556	41.40127123	31.40127123
6	4	0.518617382	1.804744036	1.604744036	7.484638997	7.484638997	18.68582061	18.18582061	32.06104392	22.06104392
7	5	0.583574387	1.941630239	1.741630239	7.924014092	7.924014092	19.86488193	20.36488193	50.89418359	40.89418359
8	6	0.526324059	1.908642416	1.908642416	8.068874897	8.068874897	19.60452535	19.60452535	45.1343524	45.1343524
9	7	0.470845131	1.802782137	2.002782137	7.783036458	7.783036458	18.81650649	18.81650649	31.96298914	31.96298914
10	8	0.412299942	1.941520202	2.141520202	9.142592607	9.642592607	20.50587821	20.00587821	49.3951487	39.3951487
11	9	0.439991673	1.835003452	2.035003452	8.165797605	7.665797605	19.12977332	18.62977332	35.49479514	45.49479514
12	10	0.400484365	1.751847638	1.551847638	7.971135616	8.471135616	18.72631201	19.22631201	27.35336693	27.35336693
13	11	0.599212625	1.931775224	1.931775224	7.735223582	7.235223582	19.67514421	19.67514421	49.17250367	39.17250367
14	12	0.50945811	1.793159091	1.993159091	7.476524665	7.476524665	18.62957613	18.12957613	30.72058557	20.72058557
15	13	0.430362243	1.882236795	2.082236795	8.535387059	9.035387059	19.64737637	19.64737637	41.02700101	51.02700101
16	14	0.429852622	1.918343742	1.718343742	8.831028346	8.331028346	20.13107466	20.13107466	46.27213623	46.27213623
17	15	0.424866096	1.884244814	1.884244814	8.591585038	9.091585038	19.69964583	20.19964583	41.33801648	31.33801648
18	16	0.452194263	1.975397544	1.775397544	9.136622292	8.636622292	20.81899301	20.81899301	55.0629609	55.0629609
19	17	0.442124634	1.925698023	1.925698023	8.803210629	8.303210629	20.17993071	20.17993071	47.44547948	57.44547948
20	18	0.548296006	1.782139338	1.782139338	7.162705993	7.662705993	18.43330915	18.43330915	29.26737971	29.26737971
21	19	0.45734311	1.837176875	1.837176875	8.065860247	8.565860247	19.09131385	19.09131385	35.74397141	35.74397141
22	20	0.408947535	1.90166071	1.70166071	8.84735601	8.34735601	20.00047834	19.50047834	43.81175618	33.81175618
23	21	0.516001898	1.884202258	1.884202258	7.963240717	7.463240717	19.33194992	18.83194992	41.47012515	51.47012515
24	22	0.496908783	1.870908563	1.870908563	8.00640533	7.50640533	19.25505782	19.75505782	39.70896527	49.70896527
25	23	0.58342787	1.895866583	1.895866583	7.599849652	8.099849652	19.26966464	18.76966464	43.61283852	43.61283852

Usual cooking time | PAM detected daily activity | Lambda | Sleeping hours | Observed Behaviours | Triangular | Data

Draw AutoShapes

Figure A10 Screenshot of Excel functions

Appendix A11 Patient types and their sensors requirement

Patient type	Prodromal choice	PAM-observed behaviour	Sensor requirement
1	Activity level Sleep	Daily activity Earliest time of leaving home Latest time of getting back home TV remote keypresses Time spent in bed Light level between 11pm to 7am Noise level between 11pm to 7am	Accelerometer GPS TV usage sensor Pressure mat Light sensor Microphone
2	Activity level Talkativeness	Daily activity Earliest time of leaving home Latest time of getting back home TV remote keypresses Time spent talking on the phone Number of daily phone calls	Accelerometer GPS TV usage sensor Phone sensor
3	Activity level Social energy	Daily activity Earliest time of leaving home Latest time of getting back home TV remote keypresses Time being outside between 5pm and 1am Latest time of getting back home	Accelerometer GPS TV usage sensor
4	Activity level Appetite	Daily activity Earliest time of leaving home Latest time of getting back home TV remote keypresses Cupboard doors usage Fridge doors usage Microwave door usage Usual time of cooking	Accelerometer GPS TV usage sensor Camera Cupboard door sensors
5	Sleep Talkativeness	Time spent in bed Light level between 11pm to 7am Noise level between 11pm to 7am Time spent talking on the phone Number of daily phone calls	Pressure mat Light sensor Microphone Phone sensor
6	Sleep Social energy	Time spent in bed Light level between 11pm to 7am Noise level between 11pm to 7am Time being outside between 5pm and 1am Latest time of getting back home	Pressure mat Light sensor Microphone GPS
7	Sleep Appetite	Time spent in bed Light level between 11pm to 7am Noise level between 11pm to 7am Cupboard doors usage Fridge doors usage Microwave door usage Usual time of cooking	Pressure mat Light sensor Microphone Camera Cupboard door sensors
8	Talkativeness Social energy	Time spent talking on the phone Number of daily phone calls Time being outside between 5pm and 1am Latest time of getting back home	Phone sensor GPS

9	Talkativeness Appetite	Time spent talking on the phone Number of daily phone calls Cupboard doors usage Fridge doors usage Microwave door usage Usual time of cooking	Phone sensor Camera Cupboard door sensors
10	Social energy Appetite	Time being outside between 5pm and 1am Latest time of getting back home Cupboard doors usage Fridge doors usage Microwave door usage Usual time of cooking	GPS Camera Cupboard door sensors
11	Activity level Sleep Talkativeness	Daily activity Earliest time of leaving home Latest time of getting back home TV remote keypresses Time spent in bed Light level between 11pm to 7am Noise level between 11pm to 7am Time spent talking on the phone Number of daily phone calls	Accelerometer GPS TV usage sensor Pressure mat Light sensor Microphone Phone sensor
12	Activity level Sleep Social energy	Daily activity Earliest time of leaving home Latest time of getting back home TV remote keypresses Time spent in bed Light level between 11pm to 7am Noise level between 11pm to 7am Time being outside between 5pm and 1am Latest time of getting back home	Accelerometer GPS TV usage sensor Pressure mat Light sensor Microphone
13	Activity level Sleep Appetite	Daily activity Earliest time of leaving home Latest time of getting back home TV remote keypresses Time spent in bed Light level between 11pm to 7am Noise level between 11pm to 7am Cupboard doors usage Fridge doors usage Microwave door usage Usual time of cooking	Accelerometer GPS TV usage sensor Pressure mat Light sensor Microphone Camera Cupboard door sensors
14	Activity level Talkativeness Social energy	Daily activity Earliest time of leaving home Latest time of getting back home TV remote keypresses Time spent talking on the phone Number of daily phone calls Time being outside between 5pm and 1am Latest time of getting back home	Accelerometer GPS TV usage sensor Phone sensor
15	Activity level Talkativeness Appetite	Daily activity Earliest time of leaving home Latest time of getting back home TV remote keypresses Time spent talking on the phone Number of daily phone calls Cupboard doors usage Fridge doors usage Microwave door usage Usual time of cooking	Accelerometer GPS TV usage sensor Phone sensor Camera Cupboard door sensors

16	Activity level Social energy Appetite	Daily activity Earliest time of leaving home Latest time of getting back home TV remote keypresses Time being outside between 5pm and 1am Latest time of getting back home Cupboard doors usage Fridge doors usage Microwave door usage Usual time of cooking	Accelerometer GPS TV usage sensor Camera Cupboard door sensors
17	Sleep Talkativeness Social energy	Time spent in bed Light level between 11pm to 7am Noise level between 11pm to 7am Time spent talking on the phone Number of daily phone calls Time being outside between 5pm and 1am Latest time of getting back home	Pressure mat Light sensor Microphone Phone sensor GPS
18	Sleep Talkativeness Appetite	Time spent in bed Light level between 11pm to 7am Noise level between 11pm to 7am Time spent talking on the phone Number of daily phone calls Cupboard doors usage Fridge doors usage Microwave door usage Usual time of cooking	Pressure mat Light sensor Microphone Phone sensor Camera Cupboard door sensors
19	Talkativeness Social energy Appetite	Time spent talking on the phone Number of daily phone calls Time being outside between 5pm and 1am Latest time of getting back home Cupboard doors usage Fridge doors usage Microwave door usage Usual time of cooking	Pressure mat Light sensor Microphone Phone sensor GPS Camera Cupboard door sensors
20	Activity level Sleep Talkativeness Social energy	Daily activity Earliest time of leaving home Latest time of getting back home TV remote keypresses Time spent in bed Light level between 11pm to 7am Noise level between 11pm to 7am Time spent talking on the phone Number of daily phone calls Time being outside between 5pm and 1am Latest time of getting back home	Accelerometer GPS TV usage sensor Pressure mat Light sensor Microphone Phone sensor
21	Activity level Sleep Talkativeness Appetite	Daily activity Earliest time of leaving home Latest time of getting back home TV remote keypresses Time spent in bed Light level between 11pm to 7am Noise level between 11pm to 7am Time spent talking on the phone Number of daily phone calls Cupboard doors usage Fridge doors usage Microwave door usage Usual time of cooking	Accelerometer GPS TV usage sensor Pressure mat Light sensor Microphone Phone sensor Camera Cupboard door sensors

22	Activity level Sleep Social energy Appetite	Daily activity Earliest time of leaving home Latest time of getting back home TV remote keypresses Time spent in bed Light level between 11pm to 7am Noise level between 11pm to 7am Time being outside between 5pm and 1am Latest time of getting back home Cupboard doors usage Fridge & microwave doors usage Usual time of cooking	Accelerometer GPS TV usage sensor Pressure mat Light sensor Microphone Camera Cupboard door sensors
23	Activity level Talkativeness Social energy Appetite	Daily activity Earliest time of leaving home Latest time of getting back home TV remote keypresses Time spent talking on the phone Number of daily phone calls Time being outside between 5pm and 1am Latest time of getting back home Cupboard doors usage Fridge doors usage Microwave door usage Usual time of cooking	Accelerometer GPS TV usage sensor Phone sensor Camera Cupboard door sensors
24	Sleep Talkativeness Social energy Appetite	Time spent in bed Light level between 11pm to 7am Noise level between 11pm to 7am Time spent talking on the phone Number of daily phone calls Time being outside between 5pm and 1am Latest time of getting back home Cupboard doors usage Fridge doors usage Microwave door usage Usual time of cooking	Pressure mat Light sensor Microphone Phone sensor Camera Cupboard door sensors
25	Activity level Sleep Talkativeness Social energy Appetite	Daily activity Earliest time of leaving home Latest time of getting back home TV remote keypresses Time spent in bed Light level between 11pm to 7am Noise level between 11pm to 7am Time spent talking on the phone Number of daily phone calls Time being outside between 5pm and 1am Latest time of getting back home Cupboard doors usage Fridge doors usage Microwave door usage Usual time of cooking	Accelerometer GPS TV usage sensor Pressure mat Light sensor Microphone Phone sensor Camera Cupboard door sensors