



Open

A multi-state model to improve the design of an automated system to monitor the activity patterns of patients with bipolar disorder

SG Mohiuddin¹, SC Brailsford^{2*}, CJ James³, JD Amor⁴, JM Blum⁵, JA Crowe⁶, EH Magill⁵ and PA Prociow⁶

¹Health Sciences Research Group, University of Manchester, Manchester, UK; ²School of Management, University of Southampton, Southampton, UK; ³Institute of Digital Healthcare, University of Warwick, Warwick, UK; ⁴Institute of Sound and Vibration Research, University of Southampton, Southampton, UK; ⁵University of Stirling, Stirling, UK; and ⁶Nottingham University, Nottingham, UK

This paper describes the role of mathematical modelling in the design and evaluation of an automated system of wearable and environmental sensors called PAM (Personalised Ambient Monitoring) to monitor the activity patterns of patients with bipolar disorder (BD). The modelling work was part of an EPSRC-funded project, also involving biomedical engineers and computer scientists, to develop a prototype PAM system. BD is a chronic, disabling mental illness associated with recurrent severe episodes of mania and depression, interspersed with periods of remission. Early detection of the onset of an acute episode is crucial for effective treatment and control. The aim of PAM is to enable patients with BD to self-manage their condition, by identifying the person's normal 'activity signature' and thus automatically detecting tiny changes in behaviour patterns which could herald the possible onset of an acute episode. PAM then alerts the patient to take appropriate action in time to prevent further deterioration and possible hospitalisation. A disease state transition model for BD was developed, using data from the clinical literature, and then used stochastically in a Monte Carlo simulation to test a wide range of monitoring scenarios. The minimum best set of sensors suitable to detect the onset of acute episodes (of both mania and depression) is identified, and the performance of the PAM system evaluated for a range of personalised choices of sensors.

Journal of the Operational Research Society (2013) **64**, 372–383. doi:10.1057/jors.2012.57

Published online 9 May 2012

Keywords: mental health; bipolar disorder; activity signatures; personalised ambient monitoring; Monte Carlo simulation

Introduction

Bipolar disorder (BD) is a severe, chronic form of mental illness associated with two types of recurrent episode, mania and depression, both of which drastically affect quality of life and the ability to function normally (Vojta *et al*, 2001; Michalak *et al*, 2007). Patients with severe BD often find it difficult to hold down a regular job or maintain personal relationships. BD can lead to significant psychological, functional, occupational and cognitive impairment. The illness is associated with high morbidity and mortality: 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 BD patient increases over time, and can differ from a few weeks to many months. In a 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. 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).

BD is associated with high social and health-care costs. Higher dependence on public assistance (Judd and Akiskal, 2003) and increased health-care use and costs (Judd and Akiskal, 2003; Simon, 2003) are found to be closely associated with BD. The costs to society are considerable. BD can lead to 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), and episodic antisocial behaviour (APA, 2000). The annual cost of managing BD in the UK NHS in 1997–1998 was estimated to be £199 million, of which £69 million was spent on hospital

*Correspondence: SC Brailsford, School of Management, University of Southampton, Southampton, Hants SO17 1BJ, UK.
E-mail: S.C.Brailsford@soton.ac.uk

admissions (Gupta and Guest, 2002). Based on a total of approximately 12 400 hospital episodes for BD in that year, this gives a rough average cost of over £5000 per admission at 1998 prices (Gupta and Guest, 2002). About 10% of the total NHS expenditure on BD is spent on medication and about 90% is spent on hospital admissions. Many drugs are available for the treatment of BD, 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 acute episodes, self-awareness and coping skills. Since drug treatment is only partially successful, psychosocial interventions are often combined with maintenance pharmacotherapy to target all aspects of the disorder and thus improve overall treatment outcome. Surveys have shown that many BD 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).

However, despite its severely disabling nature, BD can be managed effectively through self-monitoring. Many BD patients are reportedly keen to monitor their condition regularly to minimise the severity of their episodes. 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.*, 2007). The importance of analysing the early warning signs of relapse is therefore clear: if BD patients can recognise the signs early enough, actions can be taken to avert the progress of a full-blown episode. Equally, early symptoms of relapse are useful indicators to patients themselves, family members or clinicians, since extra support can then be provided to help prevent progression into a full-blown episode. Each episode usually begins with a similar pattern of symptoms (called *prodromes*) 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).

To date, most self-management interventions have been manual and diary-based. These are not only time-consuming and expensive, but are also unreliable. Moreover, they are less accurate in detecting the onset of depression (Perry *et al.*, 1999). 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). Therefore, automated ambient data collection to identify a BD patient's daily activity patterns may avoid the drawbacks of manual systems and moreover, may detect both aspects of the disorder.

The PAM project

Personalised Ambient Monitoring (PAM) is a multi-disciplinary EPSRC-funded project involving biomedical engineers, operational researchers and computer scientists at the Universities of Southampton, Nottingham, Stirling and Warwick. The aim of the project was to develop an automated system of unobtrusive sensors to monitor the behaviour patterns of patients with BD, and hopefully detect changes in these behaviour patterns that might 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 PAM system uses a system of unobtrusive small wearable and environmental sensors to monitor patients' personal daily behaviour patterns. PAM analyses the data from these to determine a normal 'activity signature', that is, a kind of fingerprint of normal daily activity. Having established a normal baseline activity signature, PAM can then identify small changes, for example minor unexplained disruptions in sleep or meal patterns, which patients may not be aware of themselves but which may potentially herald the early signs of an acute episode.

The key aspect (the 'P' of PAM) is that the level of monitoring that each person is comfortable with will be different. The PAM system allows patients to adjust the monitoring to suit their individual preferences. They can switch individual sensors on or off, as they like, or even switch the whole system off. PAM collects data from three types of source. First, from sensors situated in the home that collect information on light levels, sound levels, movement information, and aspects of television usage. Second, from sensors 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.

The available sensors consist of a wearable device with a microphone (which records sound features only, not actual voice); the wearable also includes a GPS, light sensor and accelerometer. It is comparable in size and weight to an iPod or mobile phone, and any of the sensors on it can be disabled. The 'environmental' sensors include PIR (passive infra-red) devices, which only record presence/absence of movement, like a household intruder detector; cameras (which do not record or store images but merely the presence/absence of moving objects); ambient microphones (which record sound features only, not actual voice); ambient light sensors (which detect levels of daylight/artificial light); pressure mats (for detecting movements through doorways, bedside mats etc); a TV remote monitor (counts number of button presses only); reed switches (used on cupboard or fridge doors, to detect when door is opened or closed); and bluetooth encounters (a device on the PAM mobile phone to detect proximity to other devices using bluetooth protocol, eg mobile phones).

The data from all these sensors are analysed on a dedicated PC in the patient's home. A prototype PAM system has been built by the engineers at the three partner universities. Although a very small-scale feasibility study was performed, and the researchers tested the technical performance of the prototype system by monitoring themselves, the PAM system is not yet sufficiently developed to carry out a proper clinical trial with real patients. The role of the Operational Research team was therefore to develop a model which would enable the PAM system to be tested 'in silico' for a wide range of potential choices of sensors.

A natural history model for BD

Clinical diagnosis of mental disorders is very challenging. Diagnosis is generally made on the basis of a conversation (or series of conversations) between the patient and an experienced clinical psychiatrist. No two patients will describe exactly the same symptoms and there are no universally accepted clinical staging models for mental disorders, based on objective clinical measurements such as tumour size, CD4 cell count or cholesterol levels, as there are for most physical diseases. Most disease models in the OR literature use recognised 'compartments' or stages which are defined by these clinical markers and have a clinical meaning. Thus it is far more difficult to develop mathematical models for the natural history of mental disorders than it is for physical diseases, and the literature reflects this. A literature search did not reveal a single model for BD, although there are a number of models for unipolar depression. One of the best-known examples is Patten and Lee (2004, 2005) who developed a Markov model to estimate the associations among incidence and the incidence estimation and episode duration and the number of depressed weeks reported in the preceding year.

An Excel-based Markov state transition model was developed for the basic disease process, combined with Monte Carlo simulation for generating the stochastic behaviour (in terms of daily activities) of the simulated individuals, and the corresponding stochastic data collected by the PAM system under a range of different scenarios. The first step was 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 model the collection of data from different configurations of sensors and the subsequent analysis and interpretation of these data by the PAM algorithms.

Based on the information found from the clinical literature, and following discussions with the clinical psychiatrist on the Steering Group, the progression of

BD is represented by a parameter λ which can be conceptualised as a measurement of a person's mental health status, similar to the Young Mania Rating Scale (Young *et al*, 1978) and the Hamilton Depression Rating Scale (Hamilton, 1960). Although in reality this parameter is continuous, it is discretised in the spreadsheet model so that λ takes values in steps of 0.01 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. We followed the clinical literature (Kalbag *et al*, 1999; Bauer *et al*, 2005) to assign values of λ to different mental health states, as depicted in Table 1.

In reality these transitions may 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, that is, we needed to construct a complete and realistic trajectory of an 'archetypal' BD patient including periods of depression, mania and normal health. 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, as depicted in Table 2 and based on the clinical literature as described below, each individual has a slightly different trajectory.

The model has a cycle length of 18 months (Angst and Preisig, 1995). Patients start in a healthy state and the first acute episode is depression (Kinkelin, 1954; Kalbag *et al*, 1999; Perugi *et al*, 2000; Judd *et al*, 2002). The durations of the episodes of depression and mania are taken from Angst and Sellaro (2000), Judd *et al* (2002) and NCCMH (2006). The durations of the symptom-free intervals between episodes and the changes in polarity are taken from Slater (1938), Kalbag *et al* (1999), Dunner *et al* (1979), Judd *et al* (2002), and Paykel *et al* (2006).

It can be seen that the archetypal patient has a period of depression between days 165 and 294, a period of mania

Table 1 Values of the parameter λ and their clinical interpretation

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

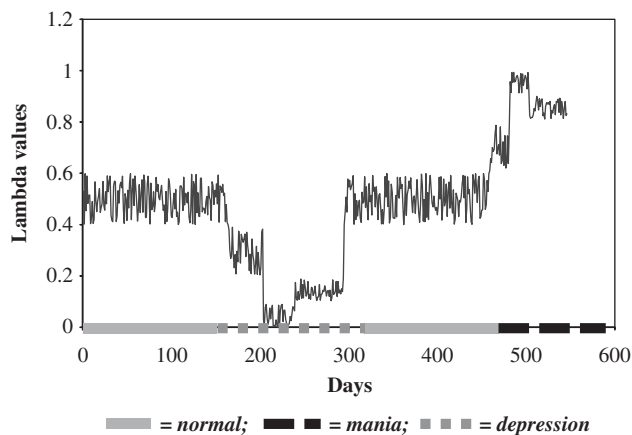


Figure 1 A sample trajectory of BD defined by small fluctuations of the parameter λ .

The following function of the parameters λ , N , D and M was devised to calculate the value of this variable across all possible mood states:

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

The form and parameters of this function are arbitrary, and there is no significance behind the choice of quadratic or cubic powers of λ . The function was chosen simply to provide face validity with the interpretation of λ , so that it had the following required properties: when $\lambda=0$, Equation (1) yields D (the value when fully depressed); $\lambda=0.5$ gives N and $\lambda=1$ gives M . Intermediate values of λ give a smooth curve with the desired 'mixed' values, corresponding to milder states of mania or depression, as shown in Figure 2. Equation (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 (1) is not applied deterministically but is subject to small random variation in the parameters N , D and M , implemented in the model by sampling from a uniform distribution. For example, a person who says they normally sleep 7 h a night may in practice sleep for anything between 6 and 8 h, irrespective of their mental health status.

By way of illustration, consider sleep pattern, which is known to be affected in BD (Morriss, 2004). Suppose that over a 24-h period, a person sleeps (on average) for 6 h when they are in good health, for 10 h when they are very depressed and for 4 h when they are very manic: thus $N=6$, $D=10$ and $M=4$. Figure 2 shows the mapping between the mood states and hours of sleep. It can clearly be seen from Figure 2 that the time spent asleep oscillates on a daily basis, but overall decreases nonlinearly from N to M as λ increases from 0.5 to 1.0, and increases from N to D as λ decreases from 0.5 to 0.0, as would be expected, since patients tend to sleep longer when depressed and less when manic (Morriss, 2004).

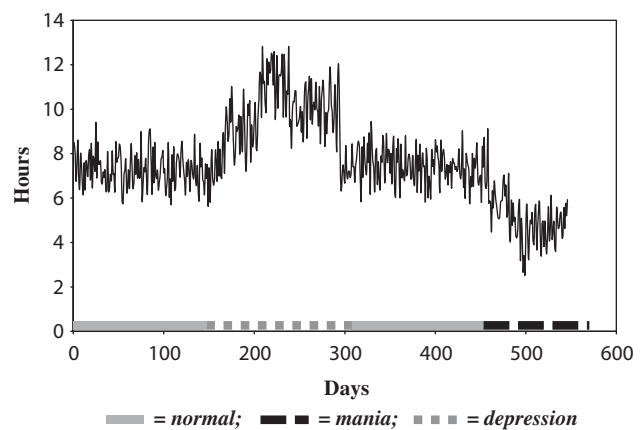


Figure 2 Time spent asleep in different mood states.

Of course, observations such as disturbed sleep patterns (although a recognised symptom in BD) may naturally vary for reasons totally unrelated to mental health; thus we cannot use the equation in a simplistic fashion to predict or diagnose BD. Normal healthy people can still find it difficult to sleep at times! Therefore, the PAM system does not use a single behavioural measure to infer anything about mental state, but rather, combinations of behaviours repeated over several days.

The details of the model were discussed with a clinical psychiatrist who treats many BD patients. Thus the literature-based assumptions and parameters of the model were augmented and validated with expert opinion. Clearly, like any model this is an over-simplification of reality. Sensitivity analysis was performed to test the robustness of the results to any estimated parameters, so that areas of uncertainty were identified and the effect on the results noted.

Modelling PAM-observable behaviours

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 BD (Bauer *et al*, 1991; Leverich and Post, 1996). The five most common bipolar prodromes, derived from the clinical literature (WHO, 2001; Morriss, 2004), were mapped in the model to various observable behaviours: these prodromes are *activity levels*, *sleep*, *talkativeness*, *social energy*, and *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 (Morris, 2004). Adherence to medication is also clearly important but this could not be monitored by the PAM system, which merely records observable activity. Even putting a sensor on the lid of a pill box, to record whether it had been opened, would not necessarily guarantee that the patient had then swallowed the tablets. We also had to exclude another important prodrome—increased or decreased ‘interest in sex’—for obvious ethical and privacy reasons!

The 14 PAM-observable behaviours, with units of measurement shown in parentheses, are shown below. These were mapped in the model to the above five prodromes (see Figure 3).

- (a) Daily activity (PAL). The PAL (Physical Activity Level) is 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 a 24-h period (hours)
- (f) Average light level between 11 pm and 7 am (lux)
- (g) Average noise level between 11 pm and 7 am (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 5 pm and 1 am (hours)
- (k) Cupboard door usage (ie the total number of times the doors were opened)
- (l) Fridge door usage (ditto)
- (m) Microwave door usage (ditto)

- (n) Usual time the person cooks the evening meal (time of day).

This choice of 14 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 3. The five prodromes defined by psychiatrists and cited in the literature are at the top level, with the 14 observable behaviours at the next level down, and the sensor data at the bottom level.

Some of the observable behaviours such as ‘time spent in bed’ and ‘daily activity’ are generic (ie 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, and so on. Following discussion with the Steering Group, it was felt that the main use of PAM would be for patients who live alone; although in the 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, and the activity patterns for an unemployed person would obviously be different from the cases considered here.

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

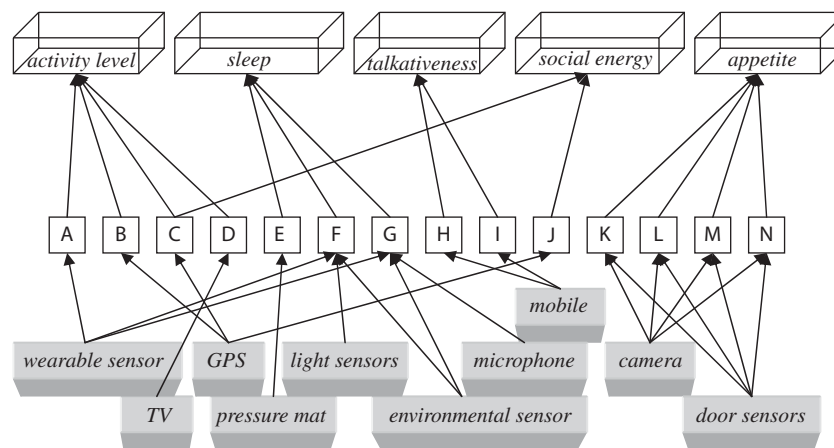


Figure 3 Mapping between prodromes, observable behaviours and sensors.

of an episode, which apply only to them. 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. Some people may object to a particular sensor rather than the activity it is intended to monitor. Figure 3 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/she 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 than 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. 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, 25 patient types are more than sufficient for the purposes of this analysis. Patient types 1–10 chose a selection of two different prodromes, patient types 11–19 chose a selection of three different prodromes, patient types 20–24 chose a selection of four different prodromes, and patient type 25 chose all five prodromes.

The random element of each behaviour, that is the part not dependent on mental health state but simply due to daily variability, was modelled by fitting triangular probability distributions. There were no empirical data to which to fit these distributions, so they were determined by a combination of 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 are scarce. 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.

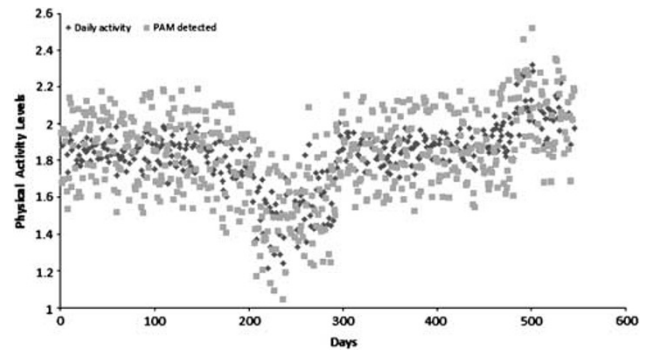


Figure 4 PAM detected physical activity levels during various mood states.

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 paper. 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 damage, lose or switch off any of the sensors. There may be software problems with the PC. 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 technical members of the PAM team, and common sense. To give an illustrative example, the sampled value of 'Time spent in bed' was varied uniformly by ± 0.5 h (ie ± 30 min). Thus, if the actual sampled value for 'Time spent in bed' on some given day was 6 h, then the PAM-detected corresponding value would be a randomly chosen value between 5.5 and 6.5 h. An example of this, for PAL, is shown in Figure 4.

PAM decision rules

A key purpose of the model was to define effective decision rules for identifying whether a significant change in behaviour had occurred, so that PAM would issue an alert (ie send a text message) to the patient. Although there was some guidance on this in the literature, as in the case

Table 3 An example of how the decision rules work

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

of the behaviours the main aim was to produce rules that were simple, 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 of the five prodromal symptoms 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, 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.

A value of 1 (=yes) was assigned when an observed behaviour exceeded its specified threshold levels, and 0 (=no) otherwise. Hence, the 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 clearly 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, in order to minimise the number of false alerts. Again, information from the clinical literature was used to guide the choice. For example, Keane (2010) reported that a manic patient did not sleep for four successive days. In the first set of experiments, PAM sent an alert if the prodromal symptoms persisted for 3 out of 5 successive days. Table 3 (where A, B, ..., G represent the observed behaviours) shows an illustrative 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.

The model was implemented in Microsoft Excel using the Monte Carlo simulation add-in *@Risk* (Palisade, 2008). Since the true value of λ is known in the model, it is therefore 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.

Experimentation and results

Two data sets were used for experimentation, representing different patient groups corresponding roughly to the clinical categories bipolar type I and bipolar type II. Bipolar type I is more common, and involves severe mania and depression, whereas type II involves severe depression but only mild mania. *Data set 1* contains patients who typically show marked mood swings with a minimal overlap with normality, whereas *Data set 2* contains data that overlap noticeably with normality. Intuitively, one would expect PAM to work better for *Data set 1* than for *Data set 2*. For both cases, the model was run for 1000 iterations for each of the 25 hypothetical patient types, thus simulating the disease trajectories of 1000 different patients of each type.

Output from the model included the four most common health-care technology evaluators, that is 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 PAM system is not biased towards keeping the FP values low.

The aim of the trials was to evaluate the performance of PAM under various sensor configurations, corresponding to an individual patient's willingness to be monitored for

Table 4 Acceptable choices in descending order of TPs (Data set 1)

<i>Patient types</i>	<i>Choices of prodromes</i>	<i>Minimum no. of sensors required</i>	<i>ODE (days)</i>	<i>OME (days)</i>	<i>TP (%)</i>	<i>FP (%)</i>
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	6.73	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 5 Acceptable choices in descending order of TPs (Data set 2)

<i>Patient types</i>	<i>Prodromal symptoms</i>	<i>Minimum no. of sensors required</i>	<i>ODE (days)</i>	<i>OME (days)</i>	<i>TP (%)</i>	<i>FP (%)</i>
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	11.33	07.49	78.19	0.70
Type 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
Type 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
Type 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
Type 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
Type 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

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. Tables 4 and 5 show the smallest set of sensors acceptable to patients in each category, in descending order of TP values, which meet these target criteria.

The results showed that the PAM system can offer a wider set of personalised prodromal choices to patients who fall into *Data set 1* than into *Data set 2*. 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 experience. Nevertheless, PAM was still able to detect the onset of depressive relapse early enough for various personalised prodromal choices. 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.

In comparison with *Data set 1*, there will indisputably be less variability in *Data set 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 5–8).

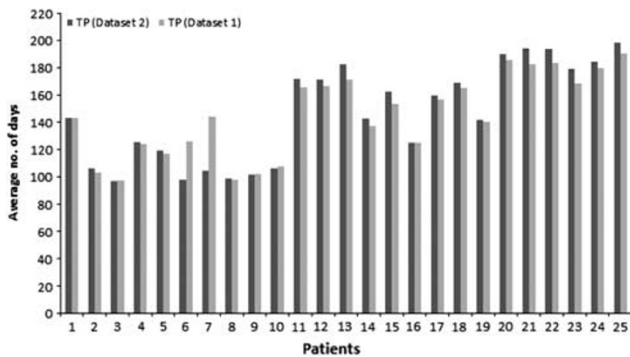


Figure 5 Comparison of true positives (TP) between Figure Data set 1 and Data set 2.

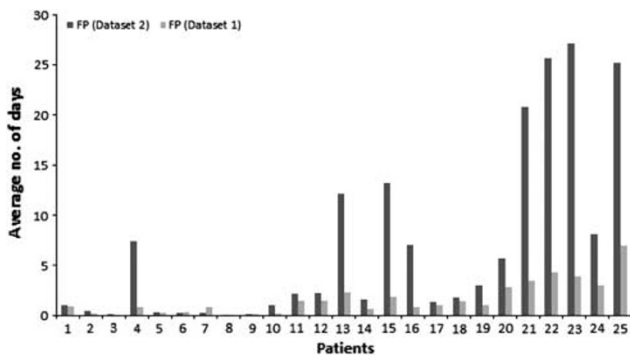


Figure 6 Comparison of false positives (FP) between Data set 1 and Data set 2.

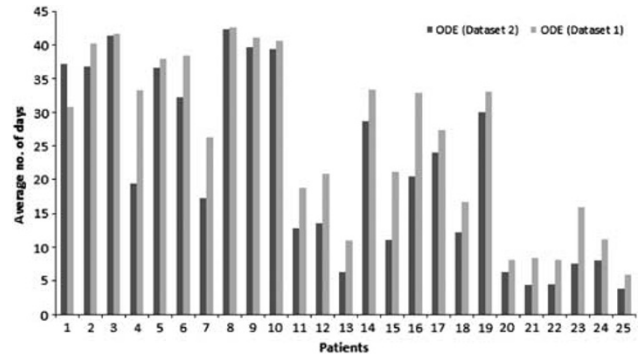


Figure 7 Comparative time to onset of a depressive episode (ODE) between Data sets 1 and 2.

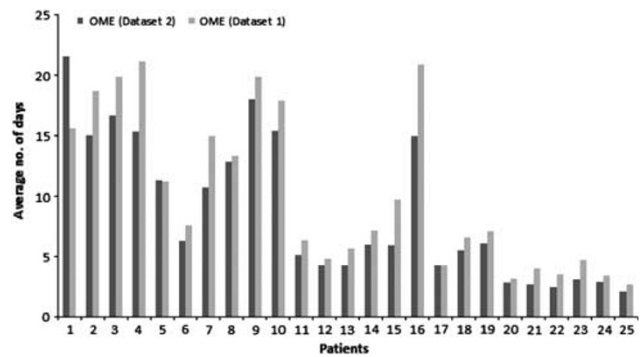


Figure 8 Comparative time to onset of a manic episode (OME) between Data sets 1 and 2.

Discussion

The overall performance of the PAM system was found to be inadequate for almost all the personalised choices of two prodromes only. This was not surprising, because the clinical literature suggests it is difficult to confirm a relapse with the appearance of only two prodromal symptoms. The performance of the system was found to be efficient for various personalised choices of three prodromes. However, the system was found to be less effective for a few specific combinations of personalised prodromal choices, for example ‘sleep’, ‘talkativeness’ and ‘social energy’, or ‘talkativeness’, ‘social energy’ and ‘appetite’, because these prodromes were associated with relatively few observable behaviours. To be able to effectively offer choices such as these, the PAM system would 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.

BD is a multi-dimensional and extremely complex illness, and clearly this model 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 a limiting assumption. We did attempt to mitigate this by modelling the two different clinical data sets. Nevertheless, the disease state transition model is 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 (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. Equation (1) and the subsequent addition of random noise to the generated values were adequate for their intended purpose, and the mapping from λ to a behaviour value matched both clinical experience and common sense.

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. 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 health-care plan. 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 system could therefore provide health-care professionals with additional clinical information to benefit bipolar patients. The overall performance of PAM was found to be good enough to support the need for further trialling, and the next steps are clearly to proceed to a larger-scale clinical trial.

Hopefully, in future BD patients and their families will find PAM reliable, simple to use and effective in improving quality of life while at the same time reducing costs by preventing unnecessary hospital admissions. The same technology could also be applicable to other patient groups, for example schizophrenia or dementia. The modelling approach could also be used in many other situations where remote monitoring is required.

Acknowledgements—The authors would like to thank the two anonymous referees for their comments, which greatly improved the paper. This research was supported by EPSRC grant EP/F005091/1.

References

- Amor, J.D. and James and C.J (2008). Personalized ambient monitoring: Accelerometry for activity level classification. In: van der Sloten J, Verdonck P, Nyssen M and Hauelsen J (eds). *4th European Conference of the International Federation for Medical and Biological Engineering (IFMBE Proceedings 22)*. Springer, Berlin, pp 866–870.
- 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. *Schweizer Archiv für Neurologie und Psychiatrie* **146**(1): 5–16.
- Angst J and Sellaro R (2000). Historical perspectives and natural history of bipolar disorder. *Biological Psychiatry* **48**(6): 445–457.
- APA—American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR)*. American Psychiatric Publishing: Washington, DC.
- Bauer M *et al* (1991). Independent assessment of manic and depressive symptoms by self-rating. Scale characteristics and implications for the study of mania. *Archives of General Psychiatry* **48**(9): 807–812.
- 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.
- Belmaker R (2004). Bipolar disorder. *New England Journal of Medicine* **351**(5): 476–486.
- Dienes KA, Hammen C, Henry RM, Cohen AN and Daley SE (2006). The stress sensitization hypothesis: Understanding the course of bipolar disorder. *Journal of Affective Disorders* **95**(1–3): 43–50.
- Dunner DL, Murphy D, Stallone F and Fieve RR (1979). Episode frequency prior to lithium treatment in bipolar manic-depressive patients. *Comprehensive Psychiatry* **e20**(6): 511–515.
- Gitlin MJ, Swendsen J, Heller TL and Hammen C (1995). Relapse and impairment in bipolar disorder. *American Journal of Psychiatry* **152**(11): 1635–1640.
- Goetzel RZ, Hawkins K, Ozminkowski RJ and Wang S (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**(1): 5–14.
- Gupta R and Guest J (2002). Annual cost of bipolar disorder to UK society. *British Journal of Psychiatry* **180**(3): 227–233.
- Hamilton M (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery & Psychiatry* **23**(1): 56–62.
- Hirschfield R *et al* (2000). Development and validation of a screening instrument for bipolar spectrum disorder: The mood disorder questionnaire. *American Journal of Psychiatry* **157**(11): 1873–1875.
- Hill R, Hardy P and Shepherd G (1996). *A Survey of the Manic Depression Fellowship*. Sainsbury Centre for Mental Health: London.
- 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**(2): 123–131.
- 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): 530–537.

- Kalbag AS, Miklowitz DJ and Richards JA (1999). A method for classifying the course of bipolar I disorder. *Behavior Therapy* **30**(2): 355–372.
- Keane F (2010). Why I sectioned my wife over her bipolar disorder. *Taking a Stand*, BBC Radio 4, 16 February at 0900 and 2130 GMT.
- Kinkelin M (1954). Verlauf und prognose des manisch-depressiven irreseins. *Schweizer Archiv für Neurologie und Psychiatrie* **73**(2): 100–146.
- Kobak KA, Greist JH, Jefferson JW, Katzelnick DJ and Mundt JC (2001). New technologies to improve clinical trials. *Journal of Clinical Psychopharmacology* **21**(3): 255–256.
- Lam D, Wong G and Sham P (2001). Prodromes coping strategies and course of illness in bipolar affective disorder—A naturalistic study. *Psychological Medicine* **27**(5): 1397–1402.
- Leverich G and Post R (1996). Life charting the course of bipolar disorder. *Current Reviews of Mood and Anxiety Disorders* **1**(1): 48–61.
- Lish JD, Dime-Meenan S, Whybrow PC, Price RA and Hirschfeld RM (1994). The national depressive and manic depressive association (DMDA) survey of bipolar members. *Journal of Affective Disorders* **31**(4): 281–294.
- Michalak EE, Yatham LN and Lam RW (2007). Quality of life in bipolar disorder: A review of the literature. *Focus* **5**(1): 117–134.
- Morriss R (2004). The early warning symptom intervention for patients with bipolar affective disorder. *Advances in Psychiatric Treatment* **10**(1): 18–26.
- Morriss R, Mohammed F, Bolton C, McCarthy J, Williamson P and Jones A (2007). Interventions for helping people recognise early signs of recurrence in bipolar disorder: Full review. *Cochrane Collaboration* **24**: CD004854.
- Müller-Oerlinghausen B, Berghöfer A and Bauer M (2002). Bipolar disorder. *Lancet*. **359**: 241–247.
- NCCMH—National Collaborating Centre for Mental Health (2006). *The Management of Bipolar Disorder in Adults, Children and Adolescents in Primary and Secondary Care*. Alden Press: Witney, UK.
- Palisade (2008). @Risk software for risk analysis, <http://www.palisade.com/risk/>, accessed 15 January 2008.
- Patten S and Lee R (2004). Refining estimates of major depression incidence and episode duration in Canada using a Monte Carlo Markov model. *Medical Decision Making* **24**(4): 351–358.
- Patten S and Lee R (2005). Describing the longitudinal course of major depression using Markov models: Data integration across three national surveys. *Population Health Metrics* **3**: 11.
- Paykel ES, Abbott R, Morriss R, Hayhurst H and Scott J (2006). Sub-syndromal and syndromal symptoms in the longitudinal course of bipolar disorder. *British Journal of Psychiatry* **189**(2): 118–123.
- Perry A, Tarrrier N, Morriss R, McCarthy E and Limb K (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): 149–153.
- 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**(1): 13–18.
- Simon G (2003). Social and economic burden of mood disorders. *Biological Psychiatry* **54**(3): 208–215.
- Slater E (1938). Zur Periodik des manisch-depressiven Irreseins. *Zeitschrift für die Gesamte Neurologie und Psychiatrie* **162**(1): 794–801.
- Solomon RL, Keitner GI, Miller IW, Shea MT and Keller MB (1995). Course of illness and maintenance treatments for patients with bipolar disorder. *Journal of Clinical Psychiatry* **56**(1): 5–13.
- Tohen M *et al* (2005). Olanzapine versus lithium in the maintenance treatment of bipolar disorder: A 12-month, randomized, double-blind, controlled clinical trial. *American Journal of Psychiatry* **162**(7): 1281–1290.
- Tse S and Walsh A (2001). How does work work for people with bipolar affective disorder? *Occupational Therapy International* **8**(3): 210–225.
- 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*.
- Vojta C, Kinoshian B, Glick H, Altshuler L and Bauer MS (2001). Self-reported quality of life across mood states in bipolar disorder. *Comprehensive Psychiatry* **42**(3): 190–195.
- WHO—World Health Organization (2001). Ministerial round tables: Mental health. 54th World Health Assembly, A54/DIV/4, http://apps.who.int/gb/archive/pdf_files/WHA54/ea54d4.pdf, accessed 14 January 2008.
- Young RC, Biggs JT, Ziegler VE and Meyer DA (1978). A rating scale for mania: Reliability, validity and sensitivity. *The British Journal of Psychiatry* **133**(5): 429–435.

Received March 2011;

accepted March 2012 after one revision



This work is licensed under a Creative Commons Attribution 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/3.0/>