



Individual differences in the temporal variability of medically unexplained symptom reporting

Eamonn Ferguson*, Helen J. Cassaday, James Erskind and Gemma Delahaye

School of Psychology, University of Nottingham, UK

Theory: Non-specific or medically unexplained symptoms account for up to 35% of outpatient referrals. In contrast to the literature on affect, little is known about how people frame daily symptoms and how these patterns are influenced by individual differences.

Method: Three fixed occasion diary studies over an 8-day period (one on ill veterans of the Gulf War and two on healthy undergraduates) and a single cross-sectional survey (on ill Gulf War veterans) were conducted. The severity and frequency of daily symptoms were assessed in all daily diary samples, together with the Big 5 personality domains in one of the undergraduate samples and in the survey of veterans. The data were analysed using auto-correlations and hierarchical multivariate linear modelling.

Results: In all the chronically ill and healthy samples, the data suggested potential cyclical patterns for symptom severity reporting. With regard to the frequency of symptom reporting, the veterans showed a pattern of constant symptom reporting. Finally, the relationship between the reported severity of symptoms was attenuated by intellect and surgency.

Discussion: There is evidence that daily experiences of symptom severity are framed relative to each other and this relationship is influenced by personality. The practical implications of these findings are discussed.

Medically unexplained symptoms (MUS)¹ include headaches and joint pain and are not in themselves diagnostic of any illness (e.g. Deary, 1996). MUS are reported by patients presenting with a number of chronic unexplained medical conditions accounting for up to 35% of outpatient referrals in the UK (Wessely, Nimnuan, & Sharpe, 1999). They are also common in the general population (e.g. Addington, Gallo, Ford, & Eaton, 2001; Pennebaker, 1982; Rief, Hessel, & Braehler, 2001) where they may or may not lead the individual to seek medical attention (Kroenke, 2001). Thus MUS are a major concern for the health services. It is surprising, therefore, to

* Correspondence should be addressed to Eamonn Ferguson, School of Psychology, University of Nottingham, University Park, Nottingham, NG7 2RD 7, UK (e-mail: eamonn.ferguson@nottingham.ac.uk).

¹ To be consistent with much of the literature, the symptoms described in this study are referred to as medically unexplained; they are also known as 'functional', 'non-specific' or 'ideopathic' symptoms.

note how little is known about individuals' day-to-day experiences of MUS. Most work on the psychology of symptom reporting has not examined their temporal aspect (Pennebaker, 1982). Kroenke (2001) highlights temporal factors (or episodicity) in symptom reporting as an important topic in research on MUS, arguing that a better understanding of these will inform the design and evaluation of interventions, as well as models about the causes of the natural history of symptom reporting (*cf.* Gable, Reis, & Elliot, 2000). For example, is it the case that patients with chronic conditions have good and bad days (see Ferguson & Cassaday, 2001)? There is a growing body of research exploring such temporal variability in daily affect (e.g. Gable *et al.*, 2000), but none in symptom reporting. Also, while there is evidence from the literature on daily affect showing that individual differences modify how strongly mood is entrained to weekly cycles (see Larsen & Kasimatis, 1990), no studies have examined the role of individual differences in relation to variability in MUS reporting. From a practical perspective, understanding the temporal variability in symptom reporting has particularly important implications for the management of symptoms in patients with chronic conditions (*cf.* McMillan & Pihl, 1987).

Assessing MUS (daily frequency and severity)

Within the diary literature, MUS have been analysed in terms of their frequency (how many symptoms each day: e.g. Brown & Moskowitz, 1997) or their severity (how severe was each symptom each day: e.g. Hazlett & Haynes, 1992). These two indices are not necessarily interchangeable. Indeed, Kroenke (2001) has argued that frequency and severity are two threshold factors people use when defining their symptoms as an illness. For example, one individual may report few symptoms but perceive them as severe, whereas another may report many symptoms but perceive them to be of low severity. However, the frequency and severity of symptom reports can be better understood in terms of the temporal dynamics of daily symptom reports. For example, a person may report the same symptoms each day (constant frequency) but perceive their severity to vary. Therefore, analyses of the temporal variability in both the daily frequency and severity of symptoms are required to understand individual variability in symptom reporting.

Temporal patterning and MUS

There is a growing body of evidence demonstrating that psychosocial factors can predict both the severity (e.g. Dancey, Taghav, & Fox, 1998; Hazlett & Haynes, 1992; Hokanson, Tate, Niu, Stader, & Flynn, 1994) and frequency (Brown & Moskowitz, 1997) of subsequent days' MUS. However, little is known concerning the predictors of temporal variability in MUS. Temporal variability refers to the cycles with which symptoms are reported. For example, some people may report having severe symptoms on one day, less severe symptoms the next day, followed by more severe symptoms on the subsequent day and so on. There are few studies exploring predictors of MUS reporting cycles. One previous study tested the extent to which daily stress predicted fluctuations in symptom reporting in patients with Systemic Lupus Erythematosus (SLE) by partialling out the previous day's symptoms (Adams, Dammers, Saia, Brantley, & Gaydos, 1994). However, these authors did not report beta coefficients and, therefore, the impact of stress and the association between symptoms across days is difficult to assess.

When defining the temporal variability in symptom reporting, three parameters need to be considered:

- (1) the duration of the variability;
- (2) the direction of the variability; and
- (3) the size of the variability (*cf.* Kroenke, 2001).

Using these parameters, temporal patterning may be defined in terms of the auto-correlation functions (ACFs) underlying daily symptom reports. ACFs describe how a variable co-varies with itself over time, and as such provide an index of how well a variable is able to predict itself (e.g. McDowall, McCleary, Meidinger, & Hay, 1987). The available evidence on the duration, direction and size of ACFs for MUS is discussed below.

Duration, direction and size of variation

To the authors' knowledge, to date only three studies have reported data on ACFs for MUS (Brown & Moskowitz, 1997; Dancey *et al.*, 1998; DeLongis, Folkman, & Lazarus, 1988). Two of these studies have examined the reported frequency of MUS (Brown & Moskowitz, 1997; DeLongis *et al.*, 1988) and one the reported severity (Dancey *et al.*, 1998). DeLongis *et al.*, reported a 1-day lag in symptom reporting, with 37% of the variance in the subsequent day's symptoms predicted by the previous day's symptoms. However, where individual variability has been reported, the evidence suggests that

- (1) significant ACFs do not emerge for all participants (Dancey *et al.*, 1998), and
- (2) negative ACFs are observed for some participants (Brown & Moskowitz, 1997).

Negative ACFs would suggest potential cyclical process (see Sutton, 1994). Sutton (1994) argues that a negative association between past and future behaviour potentially indicates a periodic or cyclic process. While evidence is sparse, there is some (frequency-based) data to suggest that people report new symptoms every two to four days (see Katon, Sullivan, & Walker, 2001; White, Williams, & Greenberg, 1961). Therefore, it can be hypothesized that variability in both severity and frequency ratings should fall within these temporal parameters. As with mood ratings, people may frame the current severity of their symptoms relative to how they felt in the near past (*cf.* Gable *et al.*, 2000).

While individual variability is, therefore, apparent in MUS reporting, there are no data exploring predictors of this variability. Therefore, work examining individual temporal variability in terms of both the frequency and severity of MUS is needed, as is an exploration of predictors of any such variability (e.g. illness chronicity or personality).

Individual differences in MUS reporting temporal variation

The temporal variation or natural rhythm of MUS reporting could show a number of individual differences. For example, is the pattern the same for those with a chronic condition compared to those without such a condition? Is the pattern of temporal variability moderated by personality?

Illness chronicity and the temporal variability of MUS

It is to be expected that individuals with a chronic condition will report more symptoms, report them more severely and/or have been doing so for longer, relative to a healthy group. Are these different experiences of illness manifest in different patterns of

symptom reporting? To explore this issue, a sample of ill Gulf War veterans were compared to healthy undergraduate samples. For both groups, it was predicted that temporal variability should be present for the severity but not the frequency of daily MUS. Gulf War veterans experience many symptoms at a fairly constant level (e.g. McCauley, Joos, Lasarev, Storzbach, & Bourdette, 1999; Unwin *et al.*, 1999), but they are likely to vary with respect to the severity of those symptoms (Ferguson & Cassaday, 2001; Unwin *et al.*, 1999). By contrast, healthy undergraduates are likely to experience relatively few symptoms, producing a low variability in symptom frequency reporting. However, even though the frequency is low, judgments of severity may still vary.

Personality and the temporal variability of MUS

Larsen and Kasimatis (1990) have shown that self-reports of mood are entrained to a weekly 7-day cycle. Further they showed that this entrainment is modified by personality. Could the same be true for symptom reporting? Is there a pattern that underlies symptom reporting which is modified by personality? An extensive body of correlational (Deary, Scott, & Wilson, 1997; Ferguson, 2000; Wolfenstien & Trull, 1997), experimental (Feldman, Cohen, Doyle, Skoner, & Gwaltney, 1999) and meta-analytic (DeNeve & Cooper, 1998) data supports a link between normal personality and symptom reporting. However, to the authors' knowledge there is little if any work on how normal personality predicts temporal variability in MUS. A question that remains unanswered, therefore, is whether personality predicts the pattern of change seen in the reported levels of MUS over a number of days.

The Big 5

The Big 5 has been proposed to provide a theoretical framework for the role of personality in relation to health and was, therefore, used in the studies reported here (Marshall, Wortman, Vickers, Kusulas, & Hervig, 1994). All domains of the Big 5 are known to correlate with general subjective well being, with emotional stability (ES) and conscientiousness (C) showing the strongest degree of association, and intellect (I) the weakest (see DeNeve & Cooper, 1998). However, less is known about which personality domains predict temporal variation in MUS reporting.

Hypotheses concerning how personality should moderate the temporal variability in symptom reports can be based on:

- (1) the Dynamic Equilibrium Model (DEM);
- (2) the likely role of selectivity in learning; and
- (3) the predictability of behaviour (in which selective learning could play some role).

The DEM of subjective well being, while not derived especially to explain MUS reporting, suggests that the personality domains of ES, surgency (S) and intellect (I) moderate short-term subjective well-being (Heady & Wearing, 1989). It is suggested by Heady and Wearing that individuals contrast recent events with their equilibrium point. This comparison process is consistent with a role for selective attention, perhaps mediated by effects on latent inhibition.

Latent inhibition (LI) refers to the later reduced ability to learn associations to a stimulus previously presented without consequences. Recent evidence suggests that S and I are inversely related to LI (Peterson & Carson, 2000). That is, those scoring high on either S or I show an attenuated LI effect. Such individuals are more likely to remain

aware of both internal and external stimuli (i.e. they have not learned to ignore contextual stimuli) and as such to be more aware of the experience of MUS and be more likely to form associations between environmental stimuli and MUS. On this account, therefore, the temporal relationship between symptom reports would be reduced for those scoring high on S and/or I.

Larsen and Kasimatis (1990) showed (consistent with a role for LI) that the entrainment of mood to a 7-day weekly cycle was reduced for those scoring high on extraversion (i.e. surgency). One suggestion they offered for this was that extraverts were less routine and predictable in their everyday life. Therefore, using Larsen and Kasimatis (1990), it can be predicted that those scoring high on S will show an attenuated temporal relationship between daily symptom reports. If this problem with predictability derives (in part), from a difficulty 'tuning out' irrelevant stimuli, then the same argument should follow for I (that also predicts reduced LI; Peterson & Carson, 2000).

Therefore, it is predicted that ES and C should be the strongest predictors of mean daily reports of MUS and that S and I should predict relatively reduced temporal fluctuation in so far as this has an associative basis (see DeNeve & Cooper, 1998; Peterson & Carson, 2000).

Overview of current studies

Four studies are reported here. The first is a cross-sectional survey of ill Gulf War veterans and is used to validate the measure of Gulf War MUS used. The second explores temporal variability in symptom reporting in a sample of veterans of the Gulf War (chronic illness group). The third replicates Study 2 in a sample of undergraduates. The fourth explores symptom reporting in healthy undergraduates with personality as a moderator of temporal variability.

Three main questions are addressed. First, are there statistically reliable patterns of temporal variability in the daily reports of symptoms? Secondly, are such patterns similar in those with a chronic condition (Gulf War syndrome) and those without any diagnosed illness (healthy undergraduates)? Finally, are there any individual differences with respect to personality domains (the Big 5)?

STUDY 1: ASSESSING MUS IN GULF WAR VETERANS

There is evidence that MUS can be accounted for by a single factor and that the different measures of MUS all tap the same general constructs (see Barsky & Borus, 1999; Deary, 1996; Ferguson & Cassaday, 1999; Wessley *et al.*, 1999). In relation to Gulf War syndrome, Haley, Kurt, and Hom (1997) provided a description of the main symptom clusters observed in Gulf War syndrome. These clusters were used to derive a set of 19 key symptoms for Gulf War veterans. The aim of this first study was to establish the validity for the MUS index in use by showing that it is correlated with a standard measure of general MUS and the Big 5 personality domains.

Method

Sample

Gulf War veterans were sampled from delegates at The National Gulf Veterans and Families Association's annual meeting in 2003. Fifty-five veterans returned questionnaires

(response rate = 70%). The sample of veterans had a mean age of 42.2 years ($SD = 8.6$). They were 90% male, had served on average 157 days in the Gulf ($SD = 281$ days) and 79% were smokers or had smoked.

Measures

The following measures were taken.

Gulf War-related MUS

Veterans indicated, in general, how severely they experienced each of 19 symptoms with respect to the course of their illness. Judgments were made using 6-point Likert-type scales ranging from 0 = never experienced the symptom to 5 = experienced very severely. These 19 symptoms were derived from the six cluster definitions provided by Haley *et al.* (1997) and covered the common symptoms reported by Gulf War veterans. Example symptoms used to map Haley *et al.*'s Syndrome Cluster 1 of impaired cognition were 'problems remembering things', 'feeling sleepy in the day time' and 'migraine type headaches'. The following are example symptoms designed to index the remaining five symptom groups: Cluster 2 (confusion ataxia) included 'problems stumbling or losing things' and 'sexual problems'; Cluster 3 (arthro-myo-neuropathy) included 'muscle fatigue after exertion' and 'tingling in the hands, feet, arms or legs'; Cluster 4 (phobia-apraxia) included 'trembling during activities like cleaning your teeth'; Cluster 5 (fever-adenopathy) included 'swollen gland in any bodily location'; and Cluster 6 (weakness-incontinence) included 'problems controlling your bowels or bladder' and 'problems standing from a chair'. The above breakdown sets out the full range of symptoms sampled. However, recent evidence suggests that a single factor can account for the covariation among MUS, so the 19 symptoms were summed to give a single symptom index.

General MUS

Veterans also completed a general validated index of MUS. They indicated how severely they had experienced each of 11 symptoms derived from Issac *et al.* (1995) and the Symptom Interpretation Questionnaire (SIQ; Robbins & Kirmayer, 1991). Symptoms included dizziness, breathlessness and palpitations. Judgments were made using 6-point Likert-type scales ranging from 0 = never experienced the symptom to 5 = experienced very severely.

Both measures were also dichotomized to create frequency-based measures. Where veterans scored 0 on each symptom severity scale, this indicated the absence of the symptom (coded as 0). Where veterans scored 1 or greater on the severity scale, this indicated that the symptom had been experienced (recoded as 1).

The Big 5

The domains of the Big 5 were assessed using Goldberg's (1992) bipolar markers. These consist of 35 bipolar adjective markers, with seven per domain. Each adjective is scored on a 9-point Likert-type scale. All domains are scored such that high scores correspond to the domain titles (emotional stability, surgency, intellect, agreeableness, and conscientiousness).

Results and discussion

The descriptive statistics, internal reliabilities and correlations among the measures are shown in Table 1.

As can be seen from the table, the two indices of MUS (both in terms of severity and frequency) show a high degree of correspondence. Furthermore, both indices of MUS show patterns of association with the Big 5 that are consistent with the general literature (see DeNeve & Cooper, 1998). Correlations were also calculated for the four indices of MUS with age and gender. There was one significant negative association between age and frequency-based assessment of general MUS ($r = -.26, p = .036$) and there were no significant associations with gender.²

The above data show that the measure of Gulf War syndrome MUS derived from Haley *et al.*'s (1997) work has validity with respect to an existing measure of MUS and personality correlates. As the Gulf War MUS index contains symptoms that are specific to veterans (sexual and cognitive factors) such as 'sexual pain in your partner' and 'problems with attention', it was decided that this index should be used with the veterans groups studied.

STUDY 2: TEMPORAL VARIABILITY IN MUS SEEN IN GULF WAR VETERANS

This study explores the pattern of temporal variability for MUS in a sample with chronic levels of MUS: ill veterans of the Gulf War.

Method

Sampling and participants

Gulf War veterans were sampled through two veterans support charities (The National Gulf Veterans and Families Association, and The Gulf Veterans Association). Adverts were placed in their newsletters and magazines and 21 veterans agreed to participate. These studies are about exploring temporal patterns in the data; as such the primary units of analyses are time intervals and so relatively small sample sizes are appropriate (e.g. 12 and 24; see Hazlett & Haynes, 1992; Steiger *et al.*, 1999). However, it is important to show the sample to be representative. The sample of veterans had a mean age of 43.3 years ($SD = 9.1$). They were 95% male, had served on average 111 days in the Gulf ($SD = 44$ days) and 56% were smokers or had smoked. The demographics for the veterans' sample are representative of those reported in large-scale epidemiological studies (e.g. Coker, Bhatt, & Graham, 1999; Haley *et al.*, 1997; Joseph, 1997; Macfarlane, Thomas, & Cherry, 2000; McCauley *et al.*, 1999; Unwin *et al.*, 1999).

Design

An 8-day fixed occasion diary study (Monday to the following Monday) was conducted following the same procedures used by David, Green, Martin, and Suls (1997; see also David & Suls, 1999). The assessment of the Level-1 variables conformed to a fixed

² Similarly across the three diary studies, age and gender were uncorrelated with daily symptom report. Of the 80 potential associations with severity and frequency; only three associations were significant, in line with the four significant associations expected by chance.

Table 1. Means, (SDs), internal reliabilities (on the diagonal) and bivariate correlations between the measures of MUS and the Big 5

	M	SD	1	2	3	4	5	6	7	8	9
GWS severity (1)	56.7	18.6	(.90)								
GWS frequency (2)	15.4	3.8	.85**	(.85)							
General MUS severity (3)	38.6	9.8	.81***	.66***	(.87)						
General MUS frequency (4)	10.1	1.6	.58***	.77**	.77***	(.77)					
Emotional stability (5)	27.4	13.3	-.43**	-.41**	-.34*	-.35*	(.90)				
Surgeency (6)	30.61	12.1	-.53**	-.33*	-.43**	-.41**	.63***	(.87)			
Intellect (7)	43.8	11.5	-.11	.01	-.04	-.06	.50***	.61***	(.91)		
Agreeableness (8)	38.0	15.2	-.30*	-.15	-.17	-.11	.70***	.63***	.73***	(.93)	
Conscientiousness (9)	38.8	13.6	-.44**	-.31**	-.18	-.15	.74***	.66***	.84***	.86***	(.88)

* $p < .05$; ** $p < .01$; *** $p < .001$.

Notes: Ns range 33 to 44. GWS = MUS scale derived for veterans. General = non-specific MUS measures.

occasion design (see Snijders & Bosker, 1999). That is, there was a fixed set of $t = 1, \dots, m$ measurement occasions. These consisted of daily reports, over an 8-day period, of the rated severity of 19 symptoms (derived from Haley *et al.* (1997) and summed in the analyses to form a single daily index).

Daily assessments were made at the end of each day (David *et al.*, 1997). End-of-day assessments are useful as they represent for people a natural time period for segmenting activities (see Gable *et al.*, 2000; Ries & Gable, 2000). Furthermore, the existing literature (see Stone, Kessler, & Haythornthwaite, 1991; West & Hepworth, 1991) suggests that symptoms and their predictors can vary over this interval (see Dancy *et al.*, 1998; Hokanson *et al.*, 1994).

While end-of-day assessments over eight days have been used before (see David *et al.*, 1997), there are two other justifications for this time frame in the current study. First, initial pilot data ($N = 6$), using a 14-day fixed occasion diary with veterans, resulted in a zero return rate (so it was inadvisable to exceed the minimum required). Secondly, results reported by Gable *et al.* (2000) showed that end-of-day data recorded over a 7-day period produced the same pattern of results as that obtained over a 14-day period. Therefore, based on this evidence, an 8-day assessment period was both practical for use in this population and sufficient to produce reliable results.

The requirements of the diary study were discussed with workers from the veterans' charities, so that they could explain the procedure to the participants.

Measures

Daily symptoms

These were measured using a 19-symptom checklist. Participants indicated, at the end of each day, how severely they had experienced each separate symptom that day. Symptoms were scored on a 6-point Likert-type scale ranging from 0 = did not experience the symptom to 5 = experienced very severely. Coefficient alphas showed that the sum of these 19 symptoms was reliable for each day of assessment (mean $\alpha = .91$). Recent theoretical (Barsky & Borus, 1999; Wessley *et al.*, 1999) and empirical (Deary, 1996; Ferguson & Cassaday, 1999) work has suggested that these diverse symptoms are best explained by a single higher order factor and as such a single score was returned for each day for each participant.

A symptom frequency index was calculated from the severity ratings of the above 19 symptoms. This was achieved by deriving a dichotomous score for each symptom. Where veterans scored 0 on each symptom severity scale, this indicated the absence of the symptom (coded as 0). Where veterans scored 1 or greater on the severity scale, this indicated that the symptom had been experienced (recoded as 1). Thus for each of the 19 symptoms each veteran's severity score was recoded as either a 0 (symptom absent) or 1 (symptom present). The 19 dichotomous items were summed to give an overall daily frequency score (mean $KR - 20 = .86$).

General data analysis strategy

The basic data analysis strategy across the three diary studies uses auto-correlations (see Catalano, Dooley, & Jackson, 1983; Chatfield, 1995; Jaccard & Wan, 1993; McDowall *et al.*, 1987; West & Hepworth, 1991). Correlograms for the ACFs were examined to identify trends and serial dependency in the data. The Z test for a proportion

was used to establish whether the number of participants exhibiting significant auto-correlation effects was significantly different from chance (*cf.* West & Hepworth, 1991).

As there are only eight days of data, the number of data points is too small to apply more formal Box-Jenkins auto-regressive-integrated-moving-average (ARIMA) models (West & Hepworth, 1991). Therefore, the data analyses are necessarily more descriptive and exploratory in nature. When the amount of temporal data is small (as in this case) it is necessary to show that any observed patterns replicate in other samples (see Ferguson, *in press*) and with other statistical techniques. West and Hepworth (1991) have argued that when the extent of the temporal data is small but the number of participants is moderate to large, then regression modelling approaches are more appropriate. Therefore, in the present study, multi-level modelling is used in Study 4 (see below) to see if patterns observed in the other two diary studies replicate using this regression approach.

Results and discussion

Data pertaining to the frequency of MUS reporting are presented first, followed by the severity data. The group level descriptive statistics for the eight days of symptoms severity and frequency reporting is shown in Table 2.

Table 2. Means (SDs) for the severity and frequency of MUS in Studies 2 and 3

	Study 2—Gulf War veterans		Study 3—undergraduates	
	Severity	Frequency	Severity	Frequency
Monday	38.2 (20.9)	11.9 (4.7)	7.8 (6.5)	4.1 (2.7)
Tuesday	36.6 (19.4)	11.8 (4.9)	6.2 (5.7)	3.3 (2.5)
Wednesday	36.2 (19.3)	11.9 (4.9)	7.4 (7.1)	3.9 (2.5)
Thursday	38.1 (10.1)	12.3 (4.4)	7.6 (6.7)	3.9 (2.8)
Friday	37.1 (20.1)	11.9 (4.5)	5.3 (5.2)	2.9 (2.5)
Saturday	39.0 (21.5)	12.2 (4.8)	5.9 (6.1)	3.4 (2.9)
Sunday	36.4 (19.8)	11.2 (5.1)	5.1 (5.2)	3.1 (2.4)
Monday	37.9 (20.1)	11.6 (4.6)	5.8 (6.5)	3.0 (2.8)

Temporal variability in MUS frequency

Of the 21 veterans, 20 provided usable data (one failed to provide sufficient daily symptom data). The number of MUS reported between veterans across days ranged from 1 to 19 symptoms. The within-veteran changes in reported daily MUS varied from 1 to 5 symptoms.

There were two veterans with significant ACFs: one was positive (.62) and one was negative (−.75); both were at a one-day lag. If an alpha level of .05 is assumed then it would be expected that, with 20 participants, one might randomly show a significant effect. The *Z* test for a proportion indicated that two veterans with significant ACFs were not significantly different from the one expected by chance. Of the 20 veterans, seven had a constant symptom frequency profile, indicating the same number of symptoms each day for the eight days (range 9–19 symptoms per day). This proportion of veterans with a constant profile was significantly different from chance ($Z = 6.2, p < .01$).

Temporal variability in MUS severity

Of the 21 veterans, 20 provided usable data. MUS severity scores between veterans across days ranged from 1 to 83. The within-veteran changes in daily NSS severity scores ranged from 1 to 46.

There were eight participants with significant ACFs and this proportion was significantly different from chance ($Z = 7.3, p < .01$). Six of the eight ACFs were negative (range $-.47$ to $-.75$) and two were positive (both $.45$). One was at a one-day lag, three were at a 2-day lag, two were at a 4-day lag, with the remainder at 5- and 6-day lags respectively.

Of the seven veterans reporting a constant number of symptoms each day, three showed significant ACFs for symptom severity ($-.47, -.63$ and $-.75$). Therefore, even when the frequency of symptoms remained constant, there was still variability in symptom reporting.

These data suggest that, for veterans, there is no significant variation in the frequency of symptom reporting—in fact the opposite in that they showed consistency in their symptom frequency reports. However, there were significant individual differences in the reported symptom severity (*cf.* Dancey *et al.*, 1998). The severity of symptoms showed a potential cyclical pattern.

Confounds

One major methodological problem discussed in relation to daily diary studies is that of ‘reactivity’ (Affleck, Urrows, Tennen, Higgins, & Abeles, 1996; Feldman, Downey, & Schaffer-Neitz, 1999). That is, significant effects observed in the data may result from practice effects attributable to completing the diary each day, as the participants become more aware of their symptoms. If participants report higher levels of symptoms over the eight-day diary period, any lagged associations may be attributable to this general trend in the data. To examine this, two one-way repeated measures ANOVAs, with days as the within-participant factor, were conducted on the severity and frequency data. There were no significant effects for severity ($F = 1.7, p = .79$) or the frequency ($F = 0.90, p = .56$) of reported symptoms across days.

STUDY 3: REPLICATION OF STUDY 2 WITH UNDERGRADUATES

Study 3 was a replication of Study 2 in a sample of university undergraduates. The same materials and methods were used as used in Study 2.

Method

Sampling and participants

Undergraduates were recruited through lectures and seminars to a total of 36 participants (mean age = 21.2 yrs, $SD = 1.2$, 39% male). Of these 36 participants, 33 provided useable data.

Design and measures

As with Study 2, an 8-day fixed occasion diary study (Monday to the following Monday) was conducted. Daily symptoms were measured using the same 19-symptom checklist as used in Study 2 and scored in the same manner (mean α for the severity rating = $.79$; and mean KR-20 for the frequency ratings = $.74$).

Results

Data pertaining to the frequency of MUS reporting are presented first, followed by the severity data. Group level descriptive data are shown in Table 2.

Temporal variability in MUS frequency

The number of MUS reported between participants across days ranged from 0 to 15 symptoms. The within-participant changes in daily MUS reports varied from 1 to 5 symptoms.

There were six participants with significant ACFs, and this proportion is significantly different from chance ($Z = 3.4, p < .05$). Five of these had single ACFs and all were negative (range $-.39$ to $-.54$), and one had two significant ACFs. Both ACFs ($-.42$ and $-.40$) were negative. Examining the ACFs, three were at a three-day lag, one at a four-day lag and three at a five-day lag.

Temporal variability in MUS severity

MUS severity scores between participants across days ranged from 0 to 41. The within-participant changes in daily MUS severity scores ranged from 1 to 11.

Again there were six participants with significant ACFs and this proportion was significantly different from chance ($Z = 3.4, p < .05$). Five of these participants had single ACFs and all were negative (range $-.36$ to $-.77$), and one had more than one significant ACF. In this case, there were three ACFs ($.61, -.42$ and $-.40$), two of which were negative. For the single ACFs, two were at a 1-day lag, two were at a 3-day lag and one was at a 5-day lag. The multiple ACFs were at lags 1, 5 and 6.

These results show a degree of replication with Study 2, with a significant number of participants showing temporal variability in the reporting of symptom severity. However, with respect to the frequency of symptom reporting, chronically ill veterans were more likely to report a constant number of symptoms whereas undergraduates showed variability. This difference may reflect the difference in illness chronicity between the veterans and the undergraduates.

Confounds

Two one-way repeated measures ANOVAs, with days as the within-participant factor, were conducted on the severity and frequency data. There were no significant effects that would indicate reactivity for either severity ($F = 1.4, p = .26$) or the frequency ($F = 2.2, p = .07$) of reported symptoms across days.

STUDY 4: PERSONALITY AND TEMPORAL VARIABILITY IN MUS

Studies 2 and 3 showed, using different samples (ill Gulf War veterans and healthy undergraduates), that a statistically reliable number of both veterans and undergraduates showed temporal variability in MUS severity reporting. However, whereas a statistically significant proportion of the veterans showed consistency in the number of symptoms that they reported, a significant proportion of undergraduates showed variability. To test the reproducibility of the findings, this study explored the same question as Studies 2 and 3, again using an undergraduate sample, but using the measure of MUS that was used in Study 1, designed for use in general healthy samples. In addition, Study 4

explores the role of the Big 5 domains in modifying the temporal variability in the reporting of MUS. Finally, the data were also analysed using multi-level modelling (MLM), a technique that is designed to deal with temporal hierarchical data of this kind. The adoption of an additional statistical approach can be used to provide cross-validation with respect to the results reported in Studies 2 and 3.

Method

Design

This used the same 8-day fixed occasion diary study as used in Studies 2 and 3. The diary study followed a two-level hierarchical design with Big 5 personality domains as the Level 2 variables, and daily symptoms (11 symptoms, summed to form a single daily index) as the Level 1 variable. Prior to completing their 8-day diaries, volunteers completed Goldberg's bipolar markers for the Big 5 (Goldberg, 1992).

Participants and sampling

Students responded to notices in lectures and seminars. Fifty diaries were distributed and 23 (46%) diaries were returned (mean age = 20.7, $SD = 0.82$, 61% female). A total of 183 days worth of usable data were recorded.

Measures

Daily MUS was assessed using 11 symptoms derived from Issac *et al.* (1995) and the Symptom Interpretation Questionnaire (SIQ; Robbins & Kirmayer, 1991). This 11-item index was used to measure the main types of symptoms reported in this healthy population. Each day, participants used these 11 symptoms to indicate on a scale of 1–10 the severity of the MUS they experienced (1 = did not experience the symptom, 10 = experienced the symptom very severely). The 11 symptom ratings from each day were summed to form a single index of symptom severity (mean $\alpha = .88$). As before, a frequency measure was derived by dichotomizing the severity scale. A score of 1 (symptom absent) was recoded as 0 and any score from 2 to 10 was recoded as 1 (symptom present). These were summed to provide an index of daily symptom frequency (mean $KR - 20 = .64$).

The domains of the Big 5 were assessed using Goldberg's (1992) bipolar markers. The following descriptive statistics and reliability estimates apply in each of the five domains: $E(M = 42.8, SD = 9.1, \alpha = .88)$; $S(M = 45.0, SD = 6.6, \alpha = .82)$, $I(M = 49.2, SD = 5.3, \alpha = .70)$, $A(M = 49.3, SD = 5.1, \alpha = .71)$, and $C(M = 43.0, SD = 8.8, \alpha = .86)$.

Data analysis

The data were analysed using ACFs and hierarchical multivariate linear modelling (HMLM; see Bryk & Raudenbush, 1992).

Hierarchical multivariate linear modelling (HMLM)

The hierarchical patterning in the diary data was explored by HMLM using HLM5 (Raudenbush, Bryk, Cheong, & Congdon, 2000). Snijders and Bosker (1999) and Raudenbush *et al.* (2000) argue that this is a suitable modelling strategy for a fixed

occasion design, as it allows various models of the temporal variance/covariance structure to be assessed. In all the models tested here, those that assumed unrestricted Level 1 variance/covariance structure were the best fits to the data and the parameters from these models are reported below (see also Raudenbush & Bryk, 2002).

The data hierarchy in this case consists of two levels. Level 1 assesses the within-person variation (i.e. daily patterns of symptom reporting) and Level 2 the between-person variability (i.e. personality). The Level 1 predictor variables were group mean centred and the Level 2 variable was centred around the grand mean (Bryk & Raudenbush, 1992; Nezlek, 2001).

Data exclusions

Participants were excluded from the HMLM analyses if they had missing data. If only one or two days' ratings were missing, the remaining data could be included.

Results and discussion

The ACF analyses for the frequency and severity data are presented first, followed by the HMLM analyses. The descriptive statistics for the severity and frequency symptom reports can be found in Table 3.

Table 3. Means (SDs) for MUS reports from Study 4

	Severity	Frequency
Monday	18.8 (7.9)	2.6 (1.8)
Tuesday	19.4 (8.6)	2.6 (2.2)
Wednesday	17.2 (4.6)	2.2 (1.6)
Thursday	22.4 (10.5)	3.0 (2.4)
Friday	16.6 (5.7)	2.0 (2.0)
Saturday	19.2 (7.7)	2.4 (2.2)
Sunday	21.0 (9.8)	3.3 (2.4)
Monday	17.6 (6.6)	2.1 (1.8)

Temporal variability in MUS frequency

Between participants, the number of MUS reported across days ranged from 0 to 9 symptoms. The within-participant changes in daily reported MUS varied from 1 to 5 symptoms.

Three participants had significant ACFs, and this is not significantly different from chance ($Z = 1.77, p > .05$).

Temporal variability in MUS severity

Between participants, the MUS severity scores across days ranged from 11 to 49. The within-participant changes in daily MUS severity scores ranged from 1 to 35.

Of the 23 participants, seven had significant ACFs. This is significantly different from chance ($Z = 5.6, p < .05$). Six of these seven participants had a single significant ACF and one participant had two significant ACFs. Five of the six participants with a single ACF had negative functions (range $-.44$ to $-.64$), and one was positive (.60). Three of the six single ACFs had a 2-day lag, two had a 1-day lag and one had a 4-day lag.

For the single participant who had two significant ACFs, the first function was at a 1-day lag and was negative ($-.60$) and the second, at a 2-day lag, was positive ($.58$).

Therefore, consistent with Studies 2 and 3, a statistically significant number of undergraduates showed temporal variability in their symptom severity reporting. Consistent with Study 2, there was no systematic variability in the frequency scores. Furthermore, the majority of the ACFs were again negative.

HMLM

The overall picture emerging from the diary studies reported here is that there is a degree of consensus that there is a negative lagged ACF for reports of symptom severity. The lags range from 1 to 5. The majority of these lags were at 1 or 2 days. To further explore these lagged effects, three separate initial HMLMs were specified containing:

- (1) only a 1-day lag;
- (2) only a 2-day lag; and
- (3) both a 1-day and a 2-day lag.

The equations expressing the model containing both the 1-day and 2 day lags are as follows:

Level 1 equation:

$$[Y_{ij} = \beta_{0j} + \beta_1(\text{symptoms})_{t-1} + \beta_2(\text{symptoms})_{t-2} + r_{ij}]$$

Level 2 equations:

$$\beta_0 = \gamma_{00}$$

$$\beta_1 = \gamma_{10}$$

$$\beta_2 = \gamma_{20}$$

where: Y_{ij} is the within-person variation in daily symptoms; β_{0j} is the intercept term; β_1 provides the slope estimates for symptom scores lagged one-day in time and the current day's symptoms and β_2 provides the slope estimates for symptom scores lagged two days in time and the current day's symptoms; r_{ij} is the error term; and γ is the coefficient expressing the Level 1 parameters. That is, γ_{00} provides the estimate for the average number of symptoms over the time series; γ_{10} provides the estimate for the average lagged association between the current day's levels of symptoms and the previous day's level of symptoms; and γ_{20} provides the estimate for the average lagged association between the current day's levels of symptoms and the level of symptoms reported two days previous. The models expressing the 1-day and 2-day lags separately were as above but only contained one predictive term each.

The next model tested whether associations observed at Level 1 were conditional on (modified by) Level 2 variables. The following equations were used to specify the Level 2 between-participant effects, for the intercept β_0 and slope components specified as β_1 and β_2 .

$$\begin{aligned} &[\beta_0 = \gamma_{00} + \gamma_{01}(ES) + \gamma_{02}(S) + \gamma_{03}(I) + \gamma_{04}(A) + \gamma_{05}(C)] \\ &[\beta_1 = \gamma_{10} + \gamma_{11}(ES) + \gamma_{12}(S) + \gamma_{13}(I) + \gamma_{14}(A) + \gamma_{15}(C)] \\ &[\beta_2 = \gamma_{20} + \gamma_{21}(ES) + \gamma_{22}(S) + \gamma_{23}(I) + \gamma_{24}(A) + \gamma_{25}(C)] \end{aligned}$$

where γ_{00} provides the average for the symptoms over the time series and γ_{10} and γ_{20} provide the average lagged associations at 1 day and 2 days, respectively. Further to this, the model contains a series of terms expressing the influence of the Big 5 on the mean level of symptoms (γ_{01} to γ_{05}), the one-day lagged associations (γ_{11} to γ_{15}) and the two-day lagged associations (γ_{21} to γ_{25}). ES refers to the effect of emotional stability, S to the effect of surgency, I to the effect of intellect, A to the effect of agreeableness and C to the effect of conscientiousness. For example the effect $\gamma_{01}(\text{ES})$ would indicate the extent to which mean severity of daily reported symptoms varied as a function of emotional stability. A negative coefficient would indicate that those with low ES report on average more severe symptoms over the 8 days. The term $\gamma_{11}(\text{ES})$ refers to the extent to which the strength of the lagged association between symptoms over a single day was moderated by emotional stability.

The deviance coefficients for each of the three initial models were compared to see if a model with a single 1-day lag, a single 2-day lag or a model containing both a 1-day and a 2-day lag was the best fit to the data. The deviance for the 1-day lag model was 1037.05 (d.f. = 30), the deviance for the 2-day lag model was 903.26 (d.f. = 23) and the deviance for the 1-day and 2-day lagged model was 893.43 (d.f. = 24). The single 2-day lag model was a better fit to the data than the single 1-day lagged model (deviance difference = 133.79, d.f. = 7). The single 2-day lagged model was a worse fit to the data than the model containing both 1-day and 2-day lagged effects (deviance difference = 9.83, d.f. = 1). Therefore, the model containing both one-and two-day lags was explored. The results for this best fitting initial model are presented in Table 4.

Table 4. Initial best fitting Level 1 model

	HMLM symbol	Coefficient	SE	p value
Intercept (mean symptom level)	$\gamma_{(00)}$	17.48	.89	<.001
One-day lag	$\gamma_{(10)}$	-0.71	.07	<.001
Two-day lag	$\gamma_{(20)}$	-0.25	.09	= .01

Notes: one-day lag = the relationship between the severity of the previous day's symptoms and the severity of the current day's symptoms; two-day lag = the relationship between the severity of symptoms two days previously and the severity of the current day's symptoms.

The results presented in Table 4 confirm the previous analyses demonstrating a significant negative association for symptom severity at the one-day and two-day lags. The first negative function indicates that high scores on one day were related to low scores on the subsequent day or vice versa, with the second negative function indicating that this effect extends over an additional day.

The results from the model containing the Level 2 personality variables are presented in Table 5. In terms of mean symptom levels there was an influence of C (i.e. $\gamma_{(05)}$). The coefficient for this effect was negative, indicating that those scoring lower on C reported more symptoms. This finding replicates the cross-sectional results observed in Study 1.

More interesting are the cross-level interactions with temporal variability. Cross-level interactions as predicted were present for S (i.e. $\gamma_{(12)}$) and I (i.e. $\gamma_{(13)}$) for the one-day lagged effect. In both cases the coefficients were positive. This indicates, that for those

Table 5. Model containing Level 2 personality variables

	HMLM symbol	Coefficient	SE	p value
Intercept (mean symptom level)	$\gamma_{(00)}$	19.51	0.83	< .001
ES	$\gamma_{(01)}$	0.005	0.11	= .966
S	$\gamma_{(02)}$	0.26	0.15	= .122
I	$\gamma_{(03)}$	-0.17	0.21	= .432
A	$\gamma_{(04)}$	-0.32	0.19	= .115
C	$\gamma_{(05)}$	-0.33	0.12	< .05
1-day lag	$\gamma_{(10)}$	-0.82	-0.05	< .001
ES	$\gamma_{(11)}$	0.006	0.007	= .421
S	$\gamma_{(12)}$	0.03	0.01	< .05
I	$\gamma_{(13)}$	0.04	0.016	< .05
A	$\gamma_{(14)}$	0.006	0.01	= .669
C	$\gamma_{(15)}$	0.002	0.009	= .820
2-day lag	$\gamma_{(20)}$	-0.50	0.08	< .001
ES	$\gamma_{(21)}$	-0.02	0.01	= .117
S	$\gamma_{(22)}$	0.009	0.01	= .448
I	$\gamma_{(23)}$	0.003	0.02	= .859
A	$\gamma_{(24)}$	-0.002	0.02	= .923
C	$\gamma_{(25)}$	-0.008	0.01	= .540

Notes: one-day lag = the relationship between the severity of the previous day's symptoms and the severity of the current day's symptoms; two-day lag = the relationship between the severity of symptoms two days previously and the severity of the current day's symptoms. ES = emotional stability, S = surgency, I = intellect, A = agreeableness, C = conscientiousness.

scoring higher on S and/or I the degree of temporal variability between subsequent days was less pronounced.

Confounds

Two one-way repeated measures ANOVAs, with days as the within-participant factor, showed no reactivity effect for symptom severity ($F = 1.2, p = .35$) or symptom frequency ($F = 1.6, p = .18$).

GENERAL DISCUSSION

Three main findings emerge from these studies. First, for MUS severity, temporal variability was observed for a significant proportion of both the Gulf War veterans and healthy undergraduates. The temporal variability in severity conformed to a cyclical pattern. Secondly, the domains of normal personality:

- (1) influenced the mean level of reported daily symptom severity; and
- (2) moderated the temporal variability in symptom reporting.

Thirdly, with a chronic illness, a significant number of veterans reported a constant number of MUS.

Temporal variability in symptom reporting

First, the patterning of temporal variability was different for frequency data and severity data. Secondly, there were individual variations with regard to these patterns.

For the Gulf War veterans, a significant proportion recorded a constant number of symptoms each day. Therefore, it appears that for participants with long-term chronic MUS, there was a significant proportion who reported the same number of symptoms each day. This subset of participants may be different qualitatively from those veterans who showed no pattern.

For the severity data, there was a significant proportion of veterans and undergraduates who showed temporal variability. This temporal variability was suggestive of a cyclical pattern. This provides evidence that the 'natural history' of symptom reporting involves fluctuation between periods of relatively more and less severe symptoms. Thus, it appears that although people may experience little variation in the number of symptoms they report, there is variability in their reported severity.

The time interval for this cycle was around one–two days. This time window is consistent with epidemiological data that suggest that the majority of people report at least one symptom every two–four days (see Katon *et al.*, 2001; White *et al.*, 1961). Therefore, in terms of our understanding of the psychological processes involved in symptom perception and reporting, there seems to be a finding of a two-day window such that people judge the severity of impact of their current symptom relative to how they felt one–two days previously.

Personality and symptom reporting

In terms of daily reports of symptom severity, C was the only personality domain to be associated with average daily levels of symptom reporting. The overall lack of association between personality and symptom reporting might at first glance appear at odds with the generally reported finding that the domains of the Big 5 are related to symptom reporting (DeNeve & Cooper, 1998). However, the studies that have reported such associations have tended to use cross-sectional retrospective analyses, and it has been argued that it is the retrospective nature of these studies that, in part, contributes to the observed associations (Brown & Moskowitz, 1997). The nature of the association with C was such that those scoring lower on C reported more severe symptoms. There is evidence that low levels of C are related to increased levels of health anxiety (Cox, Borger, Asmundson, & Taylor, 2000; Ferguson, 2000) and that first degree relative of hypochondriacal patients have lower levels of C (Noyes, Holt, Happel, Kathol, & Yagla, 1997). Thus the finding that lower levels of C were related to higher levels of self-reported symptom severity may be associated with (subclinical) health anxiety.

HMLM demonstrated that MUS in a healthy young population varies within a two-day time window. This pattern of results itself suggests that people may frame each day's experienced levels of symptom severity relative to the previous one–two days (*cf.* Heady & Wearing, 1989). Thus, for example, if someone feels ill on one day they are likely to report feeling slightly better on the subsequent day, and then worse on the following day and so on. The present data show that the degree of this variability is moderated by two of the domains of the Big 5—S, and I. Specifically, the degree of daily variability was less pronounced for individuals scoring high on S and/or I. The modifying effect of S finds parallel in the results reported by Larsen and Kasimatis (1990) with respect to the entrainment of mood to a weekly cycle. One suggestion for the attenuating effect of S offered by Larsen and Kasimatis (1990) is a behavioural one. They suggest that extraverts were less routine and predictable in their everyday life. This greater distractibility may reflect the inability to 'tune out' irrelevant environmental stimuli, consistent with the finding that those high in S and I show attenuated LI

(Peterson & Carson, 2000). Thus they would be more likely to form associations between environment stimuli and MUS.

Finally, the analyses reported here show that with respect to S and I, the formulations of the DEM have some validity (Heady & Wearing, 1989). The precise mechanisms by which S and I moderate the temporal variability in daily MUS reporting are unclear. It is suggested by Heady and Wearing that individuals contrast recent events with their equilibrium point.

Practical implications

The main practical implication to arise from the results reported in these studies is the potential use of diaries and time series analyses as an intervention strategy (*cf.* Hokanson *et al.*, 1994). That is, individuals could be asked to keep a daily diary and basic ACF analysis could be used to help the individuals to understand their patterns of symptom reporting. As the data reported here from these studies did not show any evidence of reactivity (i.e. the mere completion of the diary did not alter the responses), then just keeping a diary without such statistical analysis would not be beneficial. Furthermore, both the frequency and severity of MUS should be explored as the data reported here show that they provide different information about symptom reporting. Finally, daily diary data may also be used to differentiate between classes of individuals who show different patterns of symptom reporting and may benefit from different types of interventions. A good example of daily diary data used in this way in relation to depression is reported by McMillan and Pihl (1987). Diary methods could, therefore, form an important part of the health psychologist battery of therapeutic and investigative tools for helping people to understand and deal with the symptoms associated with chronic physical illness (see also Ferguson, *in press*).

Feinstein (1983) introduced the term 'clinimetrics' as a description of the science concerned with the measurement and understanding of physical symptoms and functional impairments. Health psychologists are in a very strong position to develop this area, as they have experience working with physical illness as well as methodological skills in field research and psychometrics. The temporal analyses of symptom data reported here could be seen as developing our clinimetric understanding of symptom reporting.

Conclusions and future work

The present data show that daily patterns in non-specific symptom reporting in both chronically ill and healthy samples demonstrate individual differences. Specifically, there are cyclical patterns for the reported severity of symptoms but not the frequency of symptom reporting, but this varied across individuals. Finally, cycles in the reported severity of symptoms were moderated by S and I from the Big 5.

The patterns of results observed in these data, especially the temporal variability, need to be explored further with longer time series and subject to a spectral analysis. This will help further to establish the cyclical nature of symptom severity reports. Furthermore, other patient and occupational samples, where MUS is and is not perceived as a problem, should be explored and patterns in their data compared. This will help to develop a better theoretical understanding of how symptoms are perceived and reported.

Acknowledgements

The authors would like to acknowledge the support of the Gulf War veterans' associations and Dr Peter Bibby for comments on an earlier draft of this study.

References

- Adams, S. G., Dammers, P. M., Saia, T. L., Brantley, P. J., & Gaydos, G. R. (1994). Stress, depression, and anxiety predict average symptom severity and daily symptom fluctuation in Systemic Lupus Erythematosus. *Journal of Behavioral Medicine, 17*, 459–477.
- Addington, A. M., Gallo, J. J., Ford, D. E., & Eaton, W. W. (2001). Epidemiology of unexplained fatigue and major depression in the community: The Baltimore ECA follow up, 1981–1994. *Psychological Medicine, 31*, 1037–1044.
- Affleck, G., Urrows, S., Tennen, H., Higgins, P., & Abeles, M. (1996). Sequential daily relations of sleep, pain intensity, and attention to pain among women with fibromyalgia. *Pain, 68*, 363–368.
- Barsky, A. J., & Borus, J. (1999). Functional somatic syndromes. *Annals of Internal Medicine, 130*, 910–921.
- Brown, K. W., & Moskowitz, D. S. (1997). Does unhappiness make you sick? The role of affect and neuroticism in the experience of common physical symptoms. *Journal of Personality and Social Psychology, 72*, 907–917.
- Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models*. London: Sage.
- Catalano, R. A., Dooley, D., & Jackson, R. (1983). Selecting a time-series strategy. *Psychological Bulletin, 94*, 506–523.
- Chatfield, C. (1995). *The analysis of time series: An introduction* (4th ed.). London: Chapman & Hall.
- Coker, W., Bhatt, B., & Graham, J. (1999). Clinical findings for the first 1000 Gulf war veterans in the Ministry of Defence's medical assessment programme. *British Medical Journal, 318*, 290–294.
- Cox, B. J., Borger, S. C., Asmundson, G. J. G., & Taylor, S. (2000). Dimensions of hypochondriasis and the five factor model of personality. *Personality and Individual Differences, 29*, 99–108.
- Dancey, C. P., Taghav, M., & Fox, R. J. (1998). The relationship between daily stress and symptoms of irritable bowel: A time-series approach. *Journal of Psychosomatic Research, 44*, 537–545.
- David, J. P., & Suls, J. (1999). Coping efforts in daily life: Role of the Big Five traits and problem appraisals. *Journal of Personality, 67*, 265–294.
- David, J. P., Green, P. J., Martin, R., & Suls, J. (1997). Differential roles of neuroticism, extraversion, and event desirability for mood in daily life: An integrative model of top-down and bottom-up influences. *Journal of Personality and Social Psychology, 73*, 149–159.
- Deary, I. J. (1996). A (latent) Big Five personality model in 1915? A reanalysis of Webb's data. *Journal of Personality and Social Psychology, 71*, 992–1005.
- Deary, I. J., Scott, S., & Wilson, J. A. (1997). Neuroticism, alexithymia and medically unexplained symptoms. *Personality and Individual Differences, 22*, 551–564.
- DeLongis, A., Folkman, S. M., & Lazarus, R. S. (1988). The impact of daily stress on health and mood: Psychological and social resources as mediators. *Journal of Personality and Social Psychology, 54*, 486–495.
- DeNeve, K. M., & Cooper, H. (1998). The happy personality: A meta-analysis of 137 personality traits and subjective well-being. *Psychological Bulletin, 124*, 197–229.
- Feinstein, A. R. (1983). An additional basic science for clinical medicine: IV. The development of clinimetrics. *Annals of Internal Medicine, 99*, 898–904.
- Feldman, P. J., Cohen, S., Doyle, W. J., Skoner, D. P., & Gwaltney, J. M., Jr. (1999). The impact of personality on the reporting of unfounded symptoms and illness. *Journal of Personality and Social Psychology, 77*, 370–378.

- Feldman, S. J., Downey, G., & Schaffer-Neitz, R. (1999). Pain, negative mood, and perceived support in chronic pain patients: A daily diary study of people with reflex sympathetic dystrophy syndrome. *Journal of Consulting and Clinical Psychology, 67*, 776-785.
- Ferguson, E. (2000). Hypochondriacal concerns and the five factor model of personality. *Journal of Personality, 68*, 705-724.
- Ferguson, E. (in press). The use of diary methodologies in health and clinical psychology. In J. Miles & P. Gilbert (Eds.), *A handbook of research methods in clinical and health psychology*. Oxford: Oxford University Press.
- Ferguson, E., & Cassaday, H. J. (1999). The Gulf War and illness by association. *British Journal of Psychology, 90*, 459-576.
- Ferguson, E., & Cassaday, H. J. (2001). Theoretical accounts of Gulf War syndrome: From environmental toxins to psychoneuroimmunology and neurodegeneration. *Behavioural Neurology, 13*, 133-147.
- Gable, S. L., Reis, H. T., & Elliot, A. J. (2000). Behavioral activation and inhibition in everyday life. *Journal of Personality and Social Psychology, 78*, 1135-1149.
- Goldberg, L. R. (1992). The development of markers for the big-five factor structure. *Psychological Assessment, 4*, 26-42.
- Haley, R. W., Kurt, T., & Hom, J. (1997). Is there a Gulf War syndrome? Searching for syndromes by factor analysis of symptoms. *Journal of the American Medical Association, 227*, 215-222.
- Hazlett, R. L., & Haynes, S. N. (1992). Fibromyalgia: A time-series analysis of the stressor-physical symptom association. *Journal of Behavioral Medicine, 15*, 541-559.
- Heady, B., & Wearing, A. (1989). Personality, life events and subjective well-being: Toward a dynamic equilibrium model. *Journal of Personality and Social Psychology, 57*, 731-739.
- Hokanson, J. E., Tate, R. L., Niu, X., Stader, S. R., & Flynn, H. A. (1994). Illustration of concomitant time series analyses in a case of somatoform disorder. *Cognitive Theory and Research, 18*, 413-437.
- Issac, M., Janca, A., Burke, K. C., Costa e Silva, J. A., Acuda, S. W., Altamura, A. C., Burke, J. D., Jnr., Chandrashekar, C. R., Miranda, C. T., & Tacchini, G. (1995). Medically unexplained somatic symptoms in different cultures. *Psychotherapy and Psychosomatics, 64*, 88-93.
- Jaccard, J., & Wan, C. K. (1993). Statistical analysis of temporal data with many observations: Issues for behavioral medicine data. *Annals of Behavioral Medicine, 15*, 41-50.
- Joseph, S. (1997). A comprehensive clinical evaluation of 20,000 Persian Gulf War veterans. *Military Medicine, 162*, 149-155.
- Katon, W., Sullivan, M., & Walker, E. (2001). Medical symptoms without pathology: Relationship to psychiatric disorders, childhood and adult trauma, and personality traits. *Annals of Internal Medicine, 134*, 917-925.
- Kroenke, K. (2001). Studying symptoms: Sampling and measurement issues. *Annals of Internal Medicine, 134*, 844-853.
- Larsen, R. J., & Kasimatis, M. (1990). Individual differences in entrainment of mood to the weekly calendar. *Journal of Personality and Social Psychology, 58*, 164-171.
- Macfarlane, G., Thomas, E., & Cherry, N. (2000). Mortality among UK Gulf War veterans. *British Medical Journal, 356*, 17-21.
- Marshall, G. N., Wortman, C. B., Vickers, R. R., Jnr., Kusulas, J. W., & Hervig, L. K. (1994). The five-factor model of personality as a framework for personality-health research. *Journal of Personality and Social Psychology, 67*, 278-286.
- McCaughey, L., Joos, S., Lasarev, M., Storzach, D., & Bourdette, D. (1999). Gulf War unexplained illnesses: Persistence and unexplained nature of self-report symptoms. *Environmental Research (Section A), 81*, 215-223.
- McDowall, D., McCleary, R., Meidinger, E. E., & Hay, R. A., Jnr. (1987). *Interrupted time series analysis*. London: Sage.
- McMillan, M. J., & Pihl, R. O. (1987). Premenstrual depression: A distinct entity. *Journal of Abnormal Psychology, 96*, 149-154.

- Nezlek, J. B. (2001). Multilevel random coefficients analyses of event- and interval-contingent data in social and personality psychology research. *Personality and Social Psychological Bulletin*, 27, 771–785.
- Noyes, R., Holt, C. S., Happel, R. L., Kathol, R. G., & Yagla, S. J. (1997). A family study of hypochondriasis. *Journal of Nervous and Mental Diseases*, 185, 223–232.
- Pennebaker, J. W. (1982). *The psychology of physical symptoms*. New York: Springer-Verlag.
- Peterson, J. B., & Carson, S. (2000). Latent inhibition and openness to experience in a high-achieving student population. *Personality and Individual Differences*, 28, 323–332.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models* (2nd ed.). London: Sage.
- Raudenbush, S., Bryk, A., Cheong, Y. F., & Congdon, R. (2000). *HLM5: Hierarchical linear and non-linear modelling*. Homewood, IL: Scientific Software International.
- Rief, W., Hessel, A., & Braehler, E. (2001). Somatization symptoms and hypochondriacal features in the general population. *Psychosomatic Medicine*, 63, 595–602.
- Ries, H. T., & Gable, S. L. (2000). Event sampling and other methods of studying daily experiences. In H. T. Ries, & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 190–222). New York: Cambridge University Press.
- Robbins, J., & Kirmayer, L. (1991). Attributions of common somatic symptoms. *Psychological Medicine*, 21, 1029–1045.
- Snijders, T., & Bosker, R. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modelling*. London: Sage.
- Steiger, H., Gauvin, L., Jabalpurwala, S., Seguin, J.R., & Stotland, S. (1999). Hypersensitivity to social interactions in bulimic syndromes: Relationship to binge eating. *Journal of Consulting and Clinical Psychology*, 67, 765–775.
- Stone, A. A., Kessler, R. C., & Haythornthwaite, J. A. (1991). Measuring daily events and experiences: Decisions for the researcher. *Journal of Personality*, 59, 575–607.
- Sutton, S. (1994). The past predicts the future: Interpreting behavior-behaviour relationships in social psychological models of health behaviour. In D. Rutter and L. Quine (Eds), *Social psychology and health: European perspectives* (pp. 71–88). Aldershot, UK: Avebury/Ashgate Publishing Company.
- Unwin, C., Blatchley, N., Coker, W., Ferry, S., Hotopf, M., Hull, L., Ismail, K., Palmer, I., David, A., & Wessely, S. (1999). Health of UK servicemen who served in the Persian Gulf War. *The Lancet*, 353, 169–178.
- Wessley, S., Nimmuan, C., & Sharpe, M. (1999). Functional somatic syndromes: One or many? *The Lancet*, 354, 936–939.
- West, S. G., & Hepworth, J. T. (1991). Statistical issues in the study of temporal data: Daily experiences. *Journal of Personality*, 59, 608–662.
- White, K., Williams, T., & Greenberg, B. (1961). The ecology of medical care. *New England Journal of Medicine*, 265, 885–892.
- Wolfenstien, M., & Trull, T. J. (1997). Depression and openness to experience. *Journal of Personality Assessment*, 69, 614–632.

Received 17 July 2002; revised version received 26 March 2003