THE UNIVERSITY
of EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

Measuring disease in dermatology: studies of objective and subjective methods

Caroline Siân Murray

Ph.D. thesis

The University of Edinburgh

2016

Declaration

I have composed this thesis. The studies are my own work. This thesis has not been submitted for any other degree or professional qualification.

Caroline Siân Murray

Abstract

Itch lies second only to disturbance of body image as a reported symptom in dermatology. This study started by concentrating on improving the measurement of itch. Itch has a paired physical response, scratch. The pairing can be exploited: preliminary work by this unit had validated the use of wrist-worn movement-measuring machines called 'accelerometers' to measure itch-related movement (scratch and rub). The first part of this research developed use of these machines. Simple accelerometers ('Actiwatch Plus') were used to observe the pattern of variation of itch over clusters of nights and in different conditions. The accelerometer scores were able to identify controls' scores from those with itchy disease. Considerable variation (56%) was discovered in objective score between subject and considerable variation was noted (46%) even within subject. More complex accelerometers, ('DigiTrac') which could potentially specifically identify itch-related movement on the basis of frequency of action derived from Fast Fourier Transform (FFT), were validated against the 'gold standard' measurement of itch-related movement, directly observed movement (via infra red video recording). It was necessary to characterise the 'frequency of action' of itch on video and, as an aside, the characteristics of human itch-related movement were compared to other mammals' itch-related movement 'frequency of action'. The 'frequency of action' and video data was used to enrich the DigiTrac readouts to improve specificity of itch-related movement detection.

During the accelerometer studies, an unexpected finding came to light: objective score of itch was not related to subjective score. To try to explain the lack of relationship, a 42 day longitudinal study of atopic dermatitis patients' subjective and objective scores was undertaken. The results demonstrated autocorrelation for subjective scores, but not for the objective scores but still did not fully explain the lack of relationship.

In an effort to explain the disconnect between subjective and objective scores a second tranche of experiments and the second part of this research interrogated whether the methods with which we measure *disease as a whole* in dermatology are robust. One study investigated whether the way patients are asked about subjective symptoms in general was resistant to the effects of focusing and framing bias. The results were reassuring as they suggested that the commonly used and recommended symptom scoring systems were robust in the face of bias. In order to assess whether perspective or perception of disease explained the disconnect, a study was designed in collaboration with the Edinburgh College of Art. A series of computer-generated images of different psoriasis severities were created and used to assess how doctors and patients assessed disease-extent. This study showed that, whilst each group had a naturally divergent opinion of extent of disease, by scoring disease using the models it was possible to unify the perspective and perception of extent. Finally, an exploratory study to reduce recall bias to a minimum, in case this had caused the disconnect between objective and subjective, was undertaken. This employed a novel questionnaire, the Day Reconstruction Method.

Lay summary

Itch lies second only to disturbance of body image as a reported symptom in dermatology. This study started by concentrating on improving the measurement of itch. Itch has a paired physical response, scratch, therefore, if you measure scratch-movements, you have a measure of itch. This unit had proved that wrist-worn movement-measuring machines called 'accelerometers' could be used to measure itch. The first part of this research further developed use of these machines. Simple accelerometers ('Actiwatch Plus') were used to observe the pattern of variation of itch over clusters of nights and in different conditions. The accelerometer scores were able to identify non-itchy people (controls) from those with itchy disease (subjects). Considerable variation was discovered in the score between subject and nearly as much variation was noted even within subject. More complex accelerometers, ('DigiTrac') were acquired which could potentially specifically identify itch-related movement (as opposed to generally moving about) on the basis of the rhythm of the movements. These were validated against the 'gold standard' measurement of itch-related movement, directly observed movement (on video recording). It was necessary to find out the 'frequency of action' of itch from videos and, as an aside, frequency of human scratch was compared to other mammals'. The 'frequency of action' and video data was used to refine the DigiTrac readouts to specifically measure itch-related movement.

An unexpected finding came to light: the accelerometer score (the objective measure), which we knew was accurate as it had been compared to video-studies and was not affected by conscious thought (as it was measured overnight) was not related to how itchy people felt and told us in questionnaires (subjective score). To try to explain the lack of relationship, a 42 day-long study of itchy eczema patients' subjective and objective scores was undertaken. The results further supported the lack of relationship and seemed to suggest that the

subjective scores rotated, day-by-day around a kind of personal thermostat of itch-sensation. The objective scores did not do this.

In a further effort to explain the disconnect between subjective and objective scores, a second tranche of experiments and the second part of this research looked into the we measure *disease as a whole* in dermatology. One study investigated whether the way patients are asked about subjective symptoms affected the answers given. The results were reassuring as they suggested that the commonly used questionnaires were robust in the face of bias. To see whether having a disease or being a doctor affected how you assess disease extent, a study was designed in collaboration with the Edinburgh College of Art. A series of computer-generated images of different psoriasis severities were created and used to assess how doctors and patients assessed disease-extent. This study showed that whilst each group had a naturally divergent opinion of extent of disease, by scoring disease using the models it was possible to unify the opinion. Finally, a questionnaire which improves the recall of events, the Day Reconstruction Method was trialled to measure subjective itch.

Acknowledgements

Firstly, I would like to thank Professor Andrew Finlay for allowing me to use and alter his DLQI for this study. Secondly, I am very grateful to Beverley Hood, Postgraduate Co-ordinator in Art & Design, Edinburgh College of Art for her help in designing the 3-D models used in Chapter 7.

I owe a debt of gratitude to Professor Jonathan Rees who indulged me in my request to do research, helped me acquire funding and then guided me.

I am very grateful to the British Skin Foundation and The Wellcome Trust for funding this research.

Finally, thank you to all the volunteers, as without them, none of this would have been possible.

Table of Contents

List of Tables

List of Figures

Abbreviations

ADLQI	Altered Dermatology Life Quality Index
ANOVA	Analysis of variance
BSA	Body Surface Area
CV	Coefficient of variance
DRM	Day Reconstruction Method
F	F ratio: ratio of two mean square values in ANOVA
FFT	Fast Fourier Transform
GHQ	Global Health Question
G	Gravity
G-s	Gravity x seconds
Hz	Hertz
KW	Kruskal-Wallis
HRQoL	Health Related Quality of Life (score)
LREC	Lothian Regional Ethics Committee
Min.s	Minutes
p	Probability
PASI	Psoriasis Area Severity Index
QI	Quality of life score
QoL	Quality of life
SCORAD	SCORe for Atopic Dermatitis extent
s	Seconds
SF-36 a.k.a. RAND SF-36	Short Form-36 health survey
	a.k.a Research And Development Short Form 36
VAS	Visual analogue scale
VCR	Video cassette recorder

Introduction

Itch is a common, distressing symptom. It is the second most common presenting complaint to the dermatologist (G. Krueger et al., 2001; Yosipovitch, Greaves, & Schmelz, 2003). It, classically, represents a primarily cutaneous pathology but can, also, be a complication of systemic disease such as uraemia, haematoproliferative or cholestatic disorders. Itch can be distressing and disabling, not least because it also disturbs sleep, and anyone, even some one who has never experienced itch (however unlikely) can empathise with this. That the sufferers of oncocerciasis or 'river blindness' state that the itch is worse than the blindness probably paints a clear enough picture of what living with itch can be like (Kale, 1998). Unfortunately, the number of effective treatments available for itch does not match the number of afflicted patients and the main excuse given for this shortfall is that the pathophysiology of itch is not well understood.

There have been a tremendous number of developments in the understanding of itch since I was first introduced to the subject in 2002 (Rees & Murray, 2005). A renaissance in itch research was triggered by a paper, published in 1997, which excited new enthusiasm since it indicated that there were specific itch fibres to conduct the sensation (Schmelz, Schmidt, Bickel, Handwerker, & Torebjörk, 1997). Right now, although there is not complete understanding, there is a much more explicit understanding of how itch works. As with most topics, knowing more, does not necessarily simplify the situation and whilst the idea of a single pathway would be tidy, research now does not seem to indicate that a single pathway does exist.

My interest in itch has always had a heavy clinical bias due to my background in clinical medicine and, specifically, as a dermatologist. I find myself fascinated by my patients when, in clinic, they report symptoms that 'explain' itch. For instance, patients with atopic

dermatitis complaining that they itch more after exercising and sweating (to do with acetylcholine and spinal disinhibition, most likely (Vogelsang, Heyer, & Hornstein, 1995). Patients with chronic itching tell me how the only relief they can get is by running their skin under cold water or by putting ice on their skin (spinal and central inhibition by cold (Atanassoff et al., 1999; Craig, Reiman, Evans, & Bushnell, 1996). Patients often seem quite relieved when I agree with them that their non-sedating antihistamines probably are not doing anything for the itch associated with their eczema – I am renowned for being the doctor that 'crosses off' antihistamines from the Drug Kardex, except, of course, if they are sedating ones for use overnight (Krause & Shuster, 1983; Rukwied, Lischetzki, Mcglone, Heyer, & Schmelz, 2000; Steinhoff et al., 2003). Sometimes, however, I am entirely flustered by my patients, even if they are the physical embodiment of the itch pathway: on my office shelf I have a series of stacking pots full of 'bugs.' There are, of course no bugs in the pots but fluff and dust. However, the patient who pressed these samples into my hand was quite convinced that they contained parasites and thus, was demonstrating the central processing of itch and psychogenic pruritus causing delusions of parasitosis. Trying to deal with this, as a dermatologist, really does take me to the limit of my comfort-zone.

My work in the field of itch has centred upon improving measurement of itch. The reason for this lies in the following thinking: "We do not really understand itch, therefore, we can not really treat it effectively, so we should do more research into itch. However, we can not accurately measure itch. How are we going to be able to rely on our results, then? We should design a good way of measuring itch, preferably something objective." This was my thought process. Unfortunately, as previously mentioned, whilst wishing to keep things simple, human biology and nature does not necessarily oblige and it became obvious that, whilst an objective measure of itch was logical, it was not capturing all the information and most specifically and importantly, the perception of itch. This led me to investigate the area of subjective disease measurement. If my world had been rocked by the discovery that

objective itch was not going to prove as useful as I had hoped, my universe was then, when I began to appreciate that some of the measurement tools, dogmatically upheld as 'validated' measures potentially provided inaccurate information. At this point I specifically have the Visual Analogue Scale in mind (Torrance, Feeny, & Furlong, 2001). To my mind the Visual Analogue Scale is a comfort-blanket: it is vehemently clutched and makes researchers feel better (they have some measurements to analyse) but it does not have much practical use. Trying to take that comfort-blanket away results in much wailing and gnashing of teeth.

When considering how to introduce this work, I really wanted to avoid using the word 'journey' as I find the term a little hackneyed (conjures up pictures as follow: my journey through weight loss, addiction etcetera). However, I do not know better how to describe this research other that the following journey: I started off trying to objectively measure itch. This was the holy grail for me. I developed the tools and capability to do this to the optimal extent and found that I was still missing a facet of quantifying itch. The objective measures did not marry up with subjective ones, so I began to review and develop measuring disease in other ways.

I: Chapter outlines

Since my research was an evolution, the layout of this thesis is slightly unconventional. Entire experiments including hypothesis, methodology, results and conclusions are confined to a particular chapter, as opposed to describing all the methods and results in separate chapters. I have tried to impart the flow of this research by using this layout.

Chapter 1: *Direct observation studies, meaning using the 'gold standard' of videoing to characterise itch.* Describes in detail the action of scratching, first in humans and then in animals.

Chapter 2: *Introduction and further validation of accelerometers already used in the literature, the Actiwatch Plus accelerometers.* Since night-to-night variation had been described in previous accelerometer research, this study sought to corroborate and explain this finding in larger night-to-night studies which were also validated against video recordings. Before the studies could take place, improvements in the handling of the accelerometer data had to be made and these are described too.

Chapter 3: *Introduction and validation of a newer accelerometer, the DigiTrac.* This chapter talks about my developing the newer accelerometer which promised better specific discrimination of itch-related movement. In this chapter, basic inter-device validation is presented. Data behind decisions that were made about the sensing axis to use in further studies are demonstrated.

Chapter 4: *Clinical experiments using the newer accelerometer, DigiTrac.* This chapter starts by integrating information about frequency of action of itch, gleaned in Chapter 1 with pilot-study data (acquired here) in order to optimise signal processing and specific itch-

related movement detection. It then describes a larger study, using the DigiTrac in atopic dermatitis patients.

Chapter 5: *Measuring itch over time in chronic disease.* This chapter contains published research. The thrust trying to disentangle the poor correlation between objective and subjective scores continually discovered. It does this is in a several week study comparing objective (accelerometer) scores to subjective ones. It also compares physician-assessed extent. A new questionnaire, not previously used in dermatology is introduced too, the Day Reconstruction Method.

Chapter 6: *The effect of bias on subjective disease scores.* This chapter contains published research. Since I had developed concerns about the possibility of bias playing a role in subjective measurement of disease in dermatology, commonly used subjective disease measures were tested for this in three experiments.

Chapter 7: *Effect of computer generated 3D models and photographs on self and third person measurement of disease extent in psoriasis.* This presents work from a very enjoyable collaboration with an artist who helped to develop models with a known body surface area coverage of psoriasis. It was hoped that these might unify the perspectives of disease extent between the specialist, non-specialist and patient. A short study investigates this and a longer one incorporates the 3D pictures plus self-photographs to see if perception of disease is altered.

II: Definition of itch

Attributed to Hafenfaffer, 1660, 'Unpleasant sensation which provokes the desire to scratch.' Whilst some have quibbled with this definition over the years, I still think this statement provides the most succinct accuracy.

III: Clinical classifications of itch

Firstly, (Yosipovitch et al., 2003):

- Acute itch: itchy condition lasting a short period (a few seconds up to one week). Elicited by substantial inflammation or injury, for instance, following an insect bite.
- Chronic itch: itchy condition which responds to treatment transiently then relapses
- Intractable itch: itchy condition unresponsive to therapy
- Alloknesis: itch due to an innocuous stimulus which does not normally cause itch (such as touch).

The merits of this classification lie in the fact that acute itch is more commonly histamine mediated and chronic not (thereby alluding to the fact that there may be two conduction mechanisms). It also raises the subject of alloknesis, the homologue of allodynia, but here representing spinal sensitization of itch. The downfall of this classification is that I can not really separate chronic from intractable itch unless the authors are aiming to allude to the fact that chronic itch is due to peripheral itch from inflammatory mediators and intractable itch being due to central causes of itch, such as μ-opioids in cholestasis.

Secondly, (Twycross et al., 2003):

- Pruritoceptive itch: due to peripheral inflammation, for example, from xerosis, urticarial, insect-bite.
- Neuropathic itch: due to pathology at any point of the afferent itch pathway, for example, post-herpetic, brachioradial pruritus.
- Neurogenic itch: originates centrally without evidence of pathology in the neural pathway, for example, cholestatic itch due to the effect of opioid neuropeptides on μ opioid receptors.
- Psychogenic itch: associated with psychological disease, for instance, delusions of parasitosis.

Any of the above may co-exist. This, I feel is a more helpful classification as it helps, to some extent, direct therapy.

IV: Literature review

This introduction's literature review starts by describing how itch 'works.' I aim to impart current understanding as accurately and as clearly as I can despite on-going queries. Then, since this thesis focusses on the quantification of itch, I will describe the development of this field. This section will be divided into the psychophysical measurement of itch, and moves into true objective measurement. Finally, I will describe the commonly-used subjective and disease extent measurement tools employed in the experimental section of this thesis.

V: Itch pathway(s)

To this simple dermatologist's mind, itch signalling is best understood as follows and the components highlighted in bold:

1) Entity to cause itch in the periphery: **chemical, mechanical, electrical**
2) Something in periphery to detect entity causing itch: **receptors**
3) Wiring to pass itch signal on from peripheral receptor: **primary afferent nerves**
4) Relay in wiring at spinal cord: **dorsal root ganglion**
5) Wiring to pass itch signal to brain: **lateral spinothalamic tract**
6) Parts of **brain** to make subject aware of itch sensation and other parts including motor cortex to provoke response with scratch.

Now I will expand each section of the process

1) *Entities to cause itch in the periphery:* ***chemical, mechanical, electrical***

It is generally accepted that itch developed, evolutionarily, to deal with parasites: if a bug is walking about in your hair/fur, the itch and ensuing reflex to scratch removes the offending creature. The movement of the bug and your hairs causes *mechanical stimulation.* On a slightly more complex level, parasites and other toxic entities are dealt with by mast cell degranulation. The products of mast cell degranulation are *chemicals.* These chemicals include the archetypal inducer of itch, histamine, plus a myriad of others such as serotonin, serine proteases and lipid mediators. More recently, and more complexly, it has become apparent that other cells in the skin can produce itch-inducing chemicals including T cells and keratinocytes. Finally, it is worth considering old fashioned itching powder (not the poor

impersonator that they sell in joke shops these days). The itching powder was actually made up of spicules from the pods of the bean of *Mucuna pruriens.* The spicules enter the skin to cause *mechanical* stimulation of itch but also, thanks to a brilliant bit of inspired basic science (boiling the spicules to denature protein), we know they also cause *chemical* mediation of itch from proteases on the surface of the spicules (Shelley & Arthur, 1955a).

Mechanical itch is elicited, obviously, by a mechanical force and detected by a pruritoceptor. This will be discussed in part 2 of the itch pathway.

As far as chemicals that cause itch are concerned, histamine has long been regarded as the classic itch-mediator. Histamine manifests itself by Lewis' triple response (T. Lewis, 1926), the hives or wheals that commonly afflict after type I hypersensitivity. Personally, however, a couple of clinical observations made me suspect, as others had, that it was not working or at least working alone in causing itch, in vivo. Firstly, if histamine caused itch in eczema, why were patients not covered in hives/wheals? The only explanation could be tachyphylaxis. Secondly, if histamine was causing the itch in chronic diseases other than urticaria, why were antihistamines so ineffective? Part of the bias towards using histamine as an itch-mediator is probably because it is easy and readily available for use. Confusing results have been thought to be due to experimental methods involving the injection of histamine: this is obviously painful as well as itchy. However, histamine can successfully and painlessly be iontophoresed and this has been the application of choice more recently. Unfortunately, and perhaps tellingly, it is also known that high concentrations of histamine can cause pain and so it may not be as 'clean' a pruritogen as would be desirable, again accounting for some of the confusing experimental findings over the years most particularly in trying to determine a specific itch pathway: it was histamine that was used to identify the 'itch-fibres' and these were not proven to be purely responsive to itch, they conducted pain too (Schmelz, 2001; 2003; Schmelz et al., 1997).

Other peripheral mediators which stimulate the 'itch fibres' include prostaglandin, particularly PGE2. This may work in a histamine independent way and certainly enhances histamine induced itch in the skin (Hägermark & Strandberg, 1977; Hägermark, Strandberg, & Hamberg, 1977). It has, more recently, been shown to be a histamine independent itch-mediator in the eye (Woodward et al., 1995). Serotonin can cause pain and itch in humans (Hägermark, 1992) and is a particularly potent mediator of itch in mice (Inagaki et al., 2001). It is of note also that different strains of mice respond in different ways to pruritogens (Inagaki et al., 2001). This is only one of the many limitations of using mouse models for itch.

As far back as 1955 and the Shelley and Arthur experiments, proteases (also called proteinases) were noted to be mediators of itch. They coined the term 'mucainain' for the protease which coats the spicules of cowhage, *Mucuna pruriens* (Arthur & Shelley, 1955; Shelley & Arthur, 1955b). Gradually, more endogenous proteases were reported as itch-mediators, specifically mast cell products such as rat cell chymase (Hägermark, Rajka, & Bergvist, 1972) and tryptase, whose effect was found to be mediated by PAR-2 receptor (Ui, Andoh, Lee, Nojima, & Kuraishi, 2006). Importantly, the itch of cowhage is not accompanied by wheal and flare – the significance of this finding and its relevance to separate itch pathways will be discussed later (Namer et al., 2008).

Acetylcholine has been reported to cause itch. It has been thought that it is this chemical which explains the fact that atopic dermatitis patients find sweat itchy rather than painful (Vogelsang et al., 1995). It is proposed that acetylcholine actually reduces the inhibition of itch (by pain) at a spinal level, thus causing the atopics to perceive itch rather than pain when sweating.

More recently, cytokines have been suggested as itch-mediators. T cell derived Il-31 (Singer et al., 2013) which appears to cause itch in T cell lymphoma. Epithelial cell derived thymic stromal lymphopoietin (TSLP), otherwise known as epithelial stromal lipoprotein (S. R. Wilson et al., 2013) has been implicated in atopic dermatitis itch.

2) *Something in periphery to detect entity causing itch:* ***receptors.***

Pruritoceptors sense the peripheral itch stimuli. Multiple receptor molecules are expressed which have one thing in common: they are G protein-coupled receptors (GPCR). Examples of these receptors include histamine receptors, H1 and H4 (Bell, McQueen, & Rees, 2004) (Rossbach et al., 2011), serotonin receptors (Kim et al., 2008) and PAR-2 (Reddy, Shimada, Sikand, LaMotte, & Lerner, 2010).

More recently, members of the Mas-related G protein-coupled receptor (Mrgpr) family were discovered which were activated by mast cell mediators and promoted histamine-independent itch (Q. Liu et al., 2009; S. R. Wilson et al., 2011). These receptors are activated by exogenous chemicals which cause itch, such as chloroquine (MrpgpA3) and by endogenous itch mediator BAM 8-22 (MrpgC11 (S. R. Wilson et al., 2011)).

Itch G protein-coupled receptors trigger G protein coupled signalling cascades via Phospholipase C and Gβσ (gamma) and ultimately mediate gating of TRP-ion channels (Imamachi et al., 2009; S. R. Wilson et al., 2011). TRP-ion channels include TRPV1 and TRPA1. TRPA1 is expressed in a subset of TRPV1 expressing neurons. These TRP channels are classically activated by pain producing compounds such as isothiocyanates which include mustard oil, brassicas, cinnamon and, of course, capsaicin (Costa et al., 2008). Cold

receptors are also TRP-ion channels (TRPM8) and whilst there is a reciprocal effect of cold, TRPM8 signalling on hot, TRPA1 (Takaishi et al., 2015) signalling, it's not, therefore, a huge leap of faith to appreciate that a similar mechanism may explain the counterirritant effect of heat, pain and cold on itch. These receptors help in the 'gating' of itch as described in section 4.

Apart from G-protein coupled receptors, there are non-G-protein coupled receptors involved in itch-mediation, and these are the receptors which appear to be activated by cytokines (Boulay, O'Shea, & Paul, 2003). The non-G-protein coupled receptors directly, or indirectly lead to signalling cascades which cause phosphorylation of multiple intracellular proteins. Examples of these receptors are the receptor for epithelial stromal lipoprotein or TSLP Thymic Stromal Lymphopoetin) receptor and Il-7Ra which is also activated by TSLP. These, as will be mentioned later have been found to be expressed in pruritoceptors in atopic dermatitis (S. R. Wilson et al., 2013). Also, IL-31Ra have been found to be expressed in a subset of sensory neurons which are activated by Il-31, the cytokine thought to be responsible for lymphoma-itch (Cevikbas et al., 2014).

Finally, receptor-wise, Toll-like receptors are thought be be likely to contribute to itch-mediation. Their expression in pruritoceptors has not been proven, however (Hoon, 2015).

With regard to the *specificity* of pruritoceptors for itch, this field is developing and actually, moleculogenetic and pharmacological studies are providing of specificity in itch pathways in a way that electrophysiological studies are unable to. Moleculogenetic and pharmacological strategies suggest coding-specificity in pruritoceptors: MrgprA3 receptor coding neurons have been found to be required to detect most itch. Ablation of these receptors reduced, but did not abolish itch and activation of the receptors, even to high intensity, resulted in itch, not pain (Han et al., 2012; Roberson et al., 2013).

3) Wiring to pass itch signal on from peripheral receptor: ***primary afferent nerves***

Clearly, I have described the interface between mediators and the nerves for conduction in the previous section. The 'wiring' deserves a special mention as it was because of a paper published in 1997 which described specific itch receptors in the skin, that interest in itch-research really resumed (Schmelz et al., 1997).

Itch researchers have long asked, 'Is there a specific itch pathway?' It was originally thought that itch was just less severe pain conducted along the same pathways, 'The Intensity Theory.' This was not without reason: histamine at low concentration causes itch and at high concentration, pain. It was not until an abstract was published in 1981 that specificity of itch away from pain seemed a possibility. In the reported study, low concentration algogens were shown to cause less intense pain, not itch. Not only this, but intraneural microstimulation in afferent nerves in humans caused either pain or itch (admittedly, itch less frequently). Importantly, there was no switch from pain to itch, or vice versa with intensity of stimulation (Torebjörk & Ochoa, 1981).

Due to the neurophysiological characteristics of the firing, it was evidenced that C-fibres had to be the wiring modality. These are unmyelinated, slow conducting fibres. These nociceptors, or C-fibres which conduct unpleasant sensations, were known to be 80% Chemo-Mechano-Heat (CMH) nociceptors, also called polymodal nociceptors, as they are stimulated by many modalities. 20% of nociceptors were considered 'silent' or sleeping nociceptors (Lynn, 1991). These 'silent' nerves are also called mechanoinsensitive (CMiHi). In 1974, it had been found that the majority of CMH neurons were insensitive or just weakly activated by histamine (Torebjörk, 1974).

Thanks to electrophysiological studies after iontophoresis of histamine, it was discovered that there were histamine-sensitive fibres amongst the CMiHi subset (Schmelz, 2001; Schmelz et al., 1997). The histamine sensitive fibres make up a fifth of the CMiHi or silent fibres, or 5% of C-fibres overall.

The characteristics of itch-conducting fibres is that they are: low conduction velocity, high transcutaneous electrical threshold, large innervation territory nerves. The latter observation explains demonstrated poor two-point discrimination of itch (Wahlgren & Ekblom, 1996).

When I came into the field of itch research, I was frustrated as it seemed that there was an obvious confounding factor in the research: all the neurophysiological studies, even ones which looked at how the fibres reacted to different itch mediators, initially identified the itch C-fibres by iontophoresing histamine. The C fibres were, therefore a self-selecting group. Thankfully, more recently, there has been a renaissance of the use of cowhage in studies and it is because of this that it is now appreciated that fibres which respond to proteases are a different type. Histamine sensitive fibres are mechano-unresponsive whilst protease sensitive fibres and those using the PAR-2 pathway are mechano-responsive (Namer et al., 2008). This finding does not concur with the concept of a single itch-pathway.

4) *Relay in wiring at spinal cord:* ***dorsal root ganglion***

In 2001, second order neurons were discovered in cats (Andrew & Craig, 2001). This followed evidence of similar neurons in rats but a criticism of this study had been that histamine had been injected which could have caused pain and which, of course could have confounded the readings (Jinks & Carstens, 1998; 2000). The cat study iontophoresed histamine and recordings were made from single spinothalamic neurons (Andrew & Craig,

2001). It was then discovered that cooling could reduce perceived itch intensity on a central level, presumably by reducing activity of primary afferents (Bromm, Scharein, Darsow, & Ring, 1995) although it is only thanks to more recent research that we appreciate that this is likely to be due to TRP-ion channels (Carstens & Akiyama, 2014).

Much in the same way that identification of the primary afferent C fibres had been limited by using histamine to hunt for them, the same happened at the second order level initially. Primate experiments have, since, identified different populations of spinothalamic neurons: some relay histamine itch and another relay cowhage mediated itch (Davidson et al., 2007).

As well as passing the itch message up centrally, it is at the cord level that other processing occurs. This would appear to be the pertinent moment to explain a feature of itch, alloknesis. Alloknesis is to itch what allodynia is to pain. Pain research had explained that the sensation has two facets: firstly there is acute pain sensation, the signal that leads to the central nervous system, helps us react (withdraw) and heads to the cortex to make us aware of the situation (perception). Secondly, there is sensitisation of secondary order neurons in the dorsal horn leading to 'hyperalgesia.' Hyperalgesia has two manifestations: allodynia and punctate hyperalgesia. Allodynia describes the phenomenon whereby normally painless touch is perceived as painful. Allodynia requires ongoing activity of the primary afferent nociceptor and is elicited by low threshold mechanorecptors, Aβ. Punctate hyperalgesia is that slightly painful pinprick like sensation. Punctate hyperalgesia does not require ongoing activity of primary afferent nociceptors and can persist for hours (LaMotte, Shain, Simone, & Tsai, 1991; Simone et al., 1991b). This sensation is elicited by Aδ fibres.

The itch equivalent is as follows:

a) Acute itch causes the reflex (scratch via spinal relay) and sensation of itch (perception at cortex). This is the message firing from the stimulus to the brain.

b) Furthermore, activation of second order dorsal horn neurons produces alloknesis, whereby touch evokes pruritus (Simone, Alreja, & LaMotte, 1991a). Alloknesis requires ongoing activity in the primary afferent. The sensation is elicited by low threshold Aβ mechanoreceptors.

c) Also, pricking 'hyperknesis' is produced. This was observed after histamine iontophoresis experiments, requires second order dorsal horn relay and is elicited by Aδ fibres (Atanassoff et al., 1999) This does not require ongoing primary afferent activity.

The primary mediator of itch between primary afferent and second order neuron is glutamate. This is evidenced by the fact that post synaptic input from itch sensory afferents can be blocked by inotropic glutamate ion channel antagonists and also by the knowledge that all sensory neurons possess the equipment to generate glutamate (Koga et al., 2011; Lagerström et al., 2010; Y. Liu et al., 2010). Hinting, again, at a second pathway, is the fact that eliminating glutamate from pruritoceptors does not abolish itch (Lagerström et al., 2010; Y. Liu et al., 2010). There was considerable excitement when in 2007, data was published suggesting that Gastrin Releasing Protein (GRP) may be another candidate for transmitter(Sun & Chen, 2007). The group had found the GRP receptor (GRPR) co-localised with TRPV1 in appropriate situations in the dorsal horn and then showed that mutant GRPR mice had an attenuated response to pruritogens. The fact that RNA for GRP is sparse in pruritoceptors has raised a question mark over its relevance more recently, however (Goswami et al., 2014).

So what other candidate might there be for spinal itch mediator? Natriuretic peptide b (Nppb) is the latest proposal (Mishra & Hoon, 2013). This chemical has been found to be co-expressed in with itch receptors and signalling components. Nppb is required to elicit an itch response in mice and if ablated in the spinal cord ablates itch behaviour. In fact, it may be that GRP and Nppb are both required, or work together: GRP deficiency (in mice), GRP antagonism and GRP ablation reduce the mouse response to histamine and to Nppb. However, blocking Nppb does not affect GRP induced itch behaviour. At this point I struggle for clarity but pharmacologists suggest that GRP is co-expressed with Nppb or that GRP is a secondary mediator (Hoon, 2015).

A final feature of spinal itch is 'gating.' This is highly relevant to this thesis as it is the method whereby scratch is linked to itch. Itch causes scratching, scratching causes pain, the pain sensation counteracts itch. This takes the discussion back to glutamate. Groups have genetically eliminated a glutamate transporter, VGlut2, from a subset of sensory neurons (Lagerström et al., 2010; Y. Liu et al., 2010) and discovered that the mice were much itchier. The mice demonstrated spontaneous itch and hyperreactivity to pruritic compounds. At the same time they were less sensitive to pain (algogens). The explanation is that there is reduction in the tonic inhibition of pain sensing neurons which usually silence itch.

Furthermore, mice lacking Bhlhb5 inhibitory neurons are also very itchy and exhibit the same behaviour as the Vglut2 deficient (Ross et al., 2010). Bhlhb5 is a neural-specific basic helix-loop-helix (bHLH) transcription factor. Bhlhb5 expressing neurons can be activated by capsaicin, mustard and menthol all of which are itch counter-irritants. The neurons which express Bhlhb5 also contain the itch relieving neuropeptide, dynorphin and it is proposed that this is the mechanism for menthol counteractivity of itch (Kardon et al., 2014).

Just a quick mention of opiates at this point: opiates were observed to cause analgesia after

epidural, but also sometimes intolerable segmental itch (Ballantyne, Loach, & Carr, 1988). Mast cell degranulation had been blamed for the problem but when this was negated, in microdialysis trials, and the itch prevailed, it was apparent that the opiates were having another effect (Blunk et al., 2004). Different types of opioids and their receptors have very different effects, but simply and for the benefit of this context, it is most salient to note that μ-opioids are good analgesics and so good that they reduce the tonal inhibition of itch at the spinal level (Andoh et al., 2008). Although known to be not directly mediated via opioid receptors, evidence for for tonal inhibition of pain by itch is derived from exacerbation of histamine induced itch by previous injection of local anaesthetic (Atanassoff et al., 1999).

5) *Wiring to pass itch signal to brain:* ***lateral spinothalamic tract***

I have alluded to the fact that the onward wiring pathway is via the lateral spinothalamic tract in section 4. It was Bickford who as long ago as 1938 identified this as the pathway by observing that disease and cordotomy involving this area abolished itch in humans (Bickford, 1938). Unlike other parts of the pathway this is not in question.

6) *Parts of* ***brain*** *to make subject aware of itch sensation and other parts including motor cortex to provoke response with scratch.*

These areas have been elucidated by imaging of the brain. First, Hsieh (Hsieh et al., 1994) injected histamine and PET scanning highlighted the contralateral anterior cingulate cortex and limbic motor cortex, thought to confer the urge to scratch, and the supplemental motor centres bilaterally. Due to the injection of histamine, the results were open to confounding by pain. Further studies have 'pricked' histamine and imaging initially by PET and later by functional MRI has confirmed that the motor and supplemental motor areas are activated by

itch. Involvement of 'unpleasantness' areas including the contralateral insula and primary somatosensory areas have been demonstrated and most importantly, the areas activated by itch are separate from those activated by pain (Darsow et al., 2000; Drzezga et al., 2001; Walter et al., 2005). The particular cortical areas activated, the anterior cingulate cortex and insular cortex, marry up with the spinal wiring inasmuch as that is where the ascending lamina I axons connect (Craig et al., 1996) and so, at least for this part of itch signalling, a single pathway separate to pain's seems likely.

VI: Quantification of itch.

Why measure itch?

Not having a good way of measuring an entity is a huge hurdle. Measuring something standardises it across the board. The problem of a lack of ways to measure disease in dermatology was considered significant enough for Professor Sam Shuster to comment in his Dowling Oration of 1966 (Shuster, 1966) that, "this metrophobia prevents us studying our problems in depth."

With a validated, objective measure for itch, within person measurements could monitor progress of treatment and thus effectiveness of treatment, or even forecast relapse. Measurement would be desirable in a clinical situation but could be argued to be vital in an experimental scenario.

How to measure itch.

Itch is a subjective symptom but it is, helpfully, coupled to a physical reaction, scratch-movement. As such, the options available with regard to measuring itch are:

a) to use the science and tools of **psychophysics** to convert the subjects' observations into recordable values, or
b) to **measure the movements that occur as a consequence of itch**, *viz* definition, scratch.

There are, of course, *quantitative* and *qualitative* features to itch, for instance, *intensity* would be the *quantitative* value and a *description* of the sensation, for instance, prickling, tickling, burning, would be *qualitative*. Of the two ways mentioned to measure itch, it would be hoped that subjective scores would provide measures of both intensity and quality, whilst objective measurement of scratch can only hope to provide a quantitative measure. Whilst an objective score is not as multi-faceted as a subjective one it is, at least, immune to the problems and bias of psychological overlay.

a) Using the science and tools of **Psychophysics** to convert the subjects' observations into recordable values.

Psychophysics is defined as 'the science of relations between psychological dimensions (mind) and physical dimensions (body), i.e., between the perceived sensation and its stimulus, dealing with the measurement of sensory perception (Wahlgren, 1995).' Itch is not the only sensory perception which has been required to be measured. Some of the simplest and earliest developed tools were scales, and their use is still prevalent today.

The scales can be described as follows (Wahlgren, 1995):

1) *Nominal scale* – categorises only, for example: itch versus no itch
2) *Ordinal scale* – provides a rank order although the steps are not equidistant, for example: mild, moderately or severely itchy.

3) *Interval scale* – categorises into a rank order, but with equidistant steps, for example the Celsius or Fahrenheit scale, zero is arbitrary, however.
4) *Ratio scale* – similar to the interval scale but with a true zero, for example Kelvin or even the Cardinal scale of numbers routinely used.

Following on from this background of scaling, itch-rating scales used have included:

1) *Ratio Estimation Scale* – an intensity of sensation is compared to another, for instance electrically stimulated itch on one arm compared to the other. This scale is obviously most useful in measuring experimental itch but can apply in the pre/post therapeutic scenario (Tuckett, 1982).
2) *Magnitude Estimation Scale* – the subject merely chooses a number to assign intensity, there is no limit to the number they can choose. The numbers are standardised by a conversion equation afterwards (Wahlgren, 1995).
3) *Graphic Rating Scale* - developed in the 1920's for 'subjective feelings.' A commonly known and used version is the visual analogue scale (VAS). This is a rating on a line of fixed length where only the extremes of the line are defined. The subject chooses a score which corresponds to their perceived itch intensity. The visual analogue (VAS) has been validated in the assessment of pain (Melzack, 1975). It is simple, sensitive and easily reproducible. It has been proven to be sensitive enough to record a dose-response in experimental itch provoked by different concentrations of histamine (Wahlgren, 1995).
4) *Numerical Rating Scales* – a digital assessment scale with fixed steps, for instance, 0-3, 0-10 with no descriptive terms except the end points. Again this scale has been validated in the evaluation of pain (Melzack, 1975; Wahlgren, 1995).

One issue with the entire area of subjective reporting of disease is recall bias (Robinson & Clore, 2002a; D. L. Thomas & Diener, 1990). Basically, the further a report is from when it

happened, the more likely ‘semantic,’ stereotyped symptoms are reported over what was actually experienced, ‘episodic’ memory. For example, you ask someone if they had a nice summer holiday when you meet up several months later and they report that they did: in actual fact it was not very nice, their luggage was lost on the way out and the room was not great. However, looking back at this point it was good: it was a holiday after all, it is expected that it should have been enjoyable. It has been shown in medical situations that recorded subjective opinion is influenced by the peak and most recently experienced sensation and is not a true reflection of the entire experience (Fredrickson & Kahneman, 1993; Kahneman, Fredrickson, Schreiber, & Redelmeier, 1993; Redelmeier & Kahneman, 1996). Tellingly, one of two similarly painful experiences can be perceived as less painful by reducing the pain experienced at the end of the procedure (Kahneman et al., 1993; Redelmeier & Kahneman, 1996).

As far back as 1989, a Swedish group (Wahlgren, Ekblom, & Hägermark, 1989) realised this potential for bias and attempted to tackle this by linking, first, a seven step graded fixed point non-verbal scale ('Paintrack') and second, a 100mm VAS, to a micro-computerised system ('Symtrack'). The system reminded the subject to carry out a score every hour by buzzing. The set-up encouraged good compliance and kept previous scores masked but even the most compliant subject found the system intrusive and they were less compliant after a few days. Nowadays, we are more used to personal portable devices and so I suspect acceptability would be improved. Whether compliance would be is questionable. Another interesting point is that ‘The 'Symtrack' was found to be limited by another of the down-fallings of visual analogue scales: clustering of scores around the extremes and centre (Wahlgren, 1995; Wahlgren et al., 1989). This flaw in visual analogue scales is not well acknowledged but has been detected in all aspects of their use, not just in dermatology (Torrance et al., 2001).

Scales do not provide a *qualitative* assessment of itch, just a measure of *intensity.* Questionnaires add the dimension of quality. Based upon the McGill Pain Questionnaire (Melzack, 1975), the Eppendorf itch questionnaire (Darsow, Mautner, Bromm, Scharein, & Ring, 1997) was developed to incorporate the scales just mentioned and adds that other dimension, qualitative assessment. It was the first questionnaire created to capture itch and was designed for uraemic itch. The questionnaire integrates a score of intensity with three qualitative dimensions: the sensory, for example, temporal and spatial; the affective, viz fear, tension; and the evaluative which assesses the overall pain experience. The itch questionnaire is behavioural. A zero score represents, 'I itch but never complain,' whilst a score of five is, 'I itch and itch interferes with rest plus/minus activity.' It provides an ordinal scale - rank orders without equidistance (Darsow et al., 1997). Yosipovitch created another questionnaire designed to assess uraemic itch in 2001(Yosipovitch et al., 2001). Whilst both of these enquired in the present tense about symptoms (did not require participants to report back over a time period) they were of limited value as they did not yield a summative score. An Itch Severity Scale was published in 2007. It added a score to Yosipovitch's questionnaire of 2001 (Majeski, Johnson, Davison, & Lauzon, 2007). Interestingly, it includes a diagram for the respondent to indicate where they experience the itch. A limitation to the usefulness of all these questionnaire is, however, that they are very detailed and take a long time to complete.

More recently developed questionnaires have been described specifically as 'Quality of Life' (QoL) questionnaires. The idea of QoL was derived from the perspective that, if you are operating in a medical system where resources are scarce, you should distribute the wealth fairly. Health Economists developed the concept of Quality Adjusted Life Years (QALY's) and these are presently used to justify health spending (Torrance, 1986; Tsuchiya & Dolan, 2005). QoL questionnaires and scores can be generic, for all diseases, and specific for system, for instance, dermatology QoL questionnaires and even specific to a condition.

Since this section is about quantifying itch, specifically, I will mention the first specific QoL questionnaire to measure the impact of itch, the ItchyQoL (Desai et al., 2008). This was based on the dermatology QoL questionnaires, Skindex-12 and Skindex-29. It enquires about the present state and requires answers to impact overall (not for a specific time). A further itch-specific QoL questionnaire is the 5-D itch scale (Elman, Hynan, Gabriel, & Mayo, 2010) measures five dimensions of living with itch: duration, degree, direction, disability and distribution. It yields a score and is quick to complete (5-10 minutes) It enquires about symptoms over the last two-weeks. Both of these questionnaires are quick to complete and yield a score.

b) ***Measuring the movements that occur as a consequence of itch: scratch***

The method of attempting to measure itch by using scratch as a proxy, is an obvious proposition based on its definition: scratch is 'coupled' to itch. Methodologically, it makes much more sense to quantify itch-related movement at night. Firstly, there is no issue regarding psychological overlay (consider how many people start to scratch at the mere mention of scabies) and secondly, organic conditions usually result in more itch at night (possibly a consequence of the warmth in bed). Thirdly, if motion monitors are to be used, a more specific result should be gained by a nocturnal study - other (rhythmical) motions, such as walking, should not complicate the picture. In this section I will report the chronological development of the field.

The 1970's

Dr. Savin, a dermatologist based here in Edinburgh, was the first to publish his observations of nocturnal scratch in itchy patients (Savin, Paterson, & Oswald, 1973). His group used facilities available in Edinburgh's Sleep Laboratory and in initial studies, four patients with

atopic dermatitis were directly observed overnight whilst recordings of EEG (electroencephalogram), EOG (electro-oculogram) and EMG's (electromyograms) of the forearm and sub mental area were made. It was noted in this study (and again demonstrated in further similar studies of more patients with atopic dermatitis as well as other pruritic conditions (Savin, Paterson, Oswald, & Adam, 1975)) that there was a higher frequency and duration of scratching bouts in stages I and II of Orthodox sleep whilst fewer occurred in stage III and even fewer in stage IV of Orthodox sleep. REM (unorthodox sleep) resulted in an intermediate, stage II level of scratching. The pattern of scratching, it was felt, was associated with the physiology of the sleep stage rather than condition. The studies, the group proposed, offered frequency and length of scratching bouts as quantifying features of itch-related movements.

Savin's desire to study scratch at the various stages of sleep had been stimulated by his observation that, as far as antipruritics were concerned - and especially antihistamines - it had not been demonstrated whether they worked by hypnotic effects or by acting as 'true' antipruritics. By this time - the early Seventies - suspicion was starting to arise that, although histamine was the classical experimental inducer of itch, it was unlikely to play a large role in chronic clinical itch where no wheal and flare was apparent. Therefore, having characterised the pattern of scratching in relation to stage of sleep, Savin's group went a step further and, in a trial whereby a sedating antihistamine and a non-sedating antihistamine were used in atopics, monitored in the sleep lab as before (Savin, Paterson, Adam, & Oswald, 1979), he showed that although the sedating antihistamine resulted in less scratching than the non-sedating drug, it's hypnotic actions placed the patient in stages III and IV of sleep for longer. This meant that less time was spent in stages I and II of sleep where more scratching occurs. When stages I and II were analysed with respect to the number of bouts of scratching, no fewer occurred than in a usual night. Savin, therefore,

surmised that the antihistamine reduced the amount of scratching because of its hypnotic effects, not because it was a true antipruritic.

One of the main limitations of Savin's studies are that the equipment required and all the wiring meant that similar studies were impractical for the Outpatient situation. This also meant that the observed characteristics may not have been 'natural' due to the lack of familiar surroundings and because of the awareness of electrical pads and wires. The specialised situation also meant that a long-term study of chronic disease patterns would be difficult - a subject is unlikely to consent to reside in a sleep lab for weeks on end. I also feel that a knowledge of the stage of sleep is only necessary if a therapeutic's action is being evaluated, i.e., whether it is sedative or not, and is not required outside this scenario.

One advantage to the EEG monitoring is that there can be no question about whether the patient is asleep or not. Scratch bouts were found to last longer if the subject was awake in this study. Due to the level of detail of observations made, useful information regarding the pathophysiology of itch, especially regarding the level of consciousness required to scratch was gained. Interestingly, the relationship between other repetitive movements, such as tooth grinding and scratch was made: both occur more often in stages I and II of sleep and, less often, in stages III and IV. Finally, and of great relevance to my studies into the frequency of action of human scratch, an EMG of the forearm is printed (Savin et al., 1973). Although the frequency of action is not stated in the paper I have calculated the EMG of the forearm as having a frequency of roughly 2Hz from the published 'read-out.' This frequency of action, 2Hz, is consistent with my own observations of disease/itch-related movement which will be reported in this thesis.

Another dermatologist, Professor Sam Shuster's interest in quantifying itch, it seems to me, came from a different aspiration. His determination was to be able to measure a subjective

experience in order to aid the understanding of itch's pathophysiology and also to allow the monitoring of chronic, itchy disease. In a paper published in 1975 (Felix & Shuster, 1975), two approaches made in order to quantify itch were described. In the first method, itchy patients were required to sleep overnight on the ward, in a hospital bed mounted on a metal frame with a proximity vibration transducer attached to one of the legs. Any body movement which moved the bed leg was detected by the transducer. The signal was amplified and continuously recorded on paper. The patients were directly observed (although not camera recorded) in order for the scratch movement read-out to be characterised/defined. The bed-leg recordings detailed the frequency of action of individual movements as well as the frequency and length of bouts of movement. The read-outs allowed the discrimination of scratch (rhythmical around a straight baseline) from restlessness (less rhythmical with a wandering baseline) and other (non-rhythmical) movements. Scratch movements were calculated, purely from the bed-leg read-outs (not from observation of limb movement) as having frequencies of 2Hz. Again, this is within the range of frequency of itch-related movements that I have observed. The bed-leg monitor also showed that less itchy people scratched in less frequent bouts but that the length of bouts and frequency of action of scratch was similar.

This group, having carried this study out and having noted limitations - especially regarding numbers of patients that could be studied in the one adapted bed on the ward - adapted a self-winding watch which logged a cumulative overnight score of itch and assessed this as an 'itch-monitor.' The main advantage, it was felt, was that if the watch's use proved successful, it could be used in the Outpatient scenario and the equipment would be unobtrusive. Using the bed-leg machine as the 'gold standard,' the self-winding watch scores were compared to this and to subjective itch scores. The overnight-movement score correlated well with the bed-leg measured scratch-movement time. The watch scores also correlated well with the

patients' subjective score of itch: mild, moderate or severe. Overall, itchy patients were shown to have higher watch scores than controls (Felix & Shuster, 1975).

Unfortunately, the self-winding watches were subject to criticism: they were inadequately sensitive to respond to all scratch movement; the mass of the pendulum meant the devices had positional sensitivity; the response was related to the velocity of the movement so that vigorous scratch registered a higher score than gentle scratch. Due to all of these facts, it was argued that the calibration technique employed (on a turntable revolving at a constant velocity) was invalid (Summerfield & Welch, 1980) The fact that the watch presented a cumulative score was another drawback as a pattern of movement could not be defined and scratch could not be separated from other movements.

The 1980's

Dermatologists are not the only specialists whose patients present with itch. Some devices designed to measure scratch have arisen from a desire to quantify cholestatic itch. In 1980, Summerfield described an electromagnetic movement detector which provided a cumulative duration of nocturnal limb movements (Summerfield & Welch, 1980). The detector was a little larger than a wristwatch and was tested when worn on the wrist/ankle. The group had studied Shuster's self-winding watches closely and attempted to improve on the design. Their sensor, which they claimed measured in all directions, registered a score when a movement exceeded a certain amplitude threshold (set after observing simulated scratch). The registered movement caused an increment in a mechanical or quartz watch. The devices were tested on Liver Ward in-patients - some with cholestasis and itch, some controls and some restless patients (they had non-itchy liver disease but due to encephalopathy did not sleep well). The device was sensitive enough to separate all three groups and there was good correlation with subjective scores. The group argued that, as the itchy patients moved their arms relatively more, then their movements represented scratch - the restless moved their

legs relatively more. Interestingly, a few patients with cholestasis were examined before and after biliary drainage interventions and the *subjective* score of itch improved *before* the *objective* score of movement, they *expected* to feel better, we assume. Unfortunately, the lack of longitudinal discrimination throughout the night (resulting from the cumulative score) and the motion sensor's not being specific for itch presumably accounts for these devices limited clinical use.

These early motion sensors, mounted on one limb, were too insensitive to register movements of other parts of the body and thus missed many of the non-stereotypical itch-related movements such as rubbing the face into the pillow or even legs rubbing together. Newer digital accelerometers (such as I have used) are so sensitive that even when the accelerometer is worn on the right wrist, scratching with the left hand is registered, as is truncal or leg movement (Benjamin et al., 2004; Bringhurst, Waterston, Schofield, Benjamin, & Rees, 2004).

Also in 1980, Aoki described the use of another new device (Aoki et al., 1980). It involved a paper strain gauge being attached to the back of the hand and relayed to an amplifier. In conceiving the design, it had been hoped that the finger movement associated with scratching could be measured and filtered according to the frequency of movement. The group described how they developed a computer analysis system for the strain-gauge read-out and decided to exclude movement outside 1-3Hz and only registered movement lasting more than 2 seconds – the basis of this was *simulated scratch* in controls. The group chose to study atopics, generalised eczema patients and controls who were studied overnight, on the ward and scratch was expressed as 'total scratching time.' Analysis of the read-outs showed that the device could separate those with itchy skin from controls and did not separate the atopics from those with generalised eczema. In a separate study (Aoki, Kushimoto,

Hishikawa, & Savin, 1991), the system was used alongside EEG monitoring and reflected Savin's previously discovered pattern of scratching activity in itchy conditions.

One problem with this study's design would appear to be that the filter frequency was decided after analysing a paper read-out of *simulated scratch in a normal volunteer who was awake* – could this be equivalent to an atopic's scratch overnight? It is known that decerebrate cats scratch at a lower frequency than cerebrate (Kuhta & Smith, 1990; Nojima & Carstens, 2003), and human studies have demonstrated that scratching whilst asleep is different from that when awake (Savin et al., 1975). It may have been more reliable to observe elicited experimental itch. However, the range of frequency for scratch-action which the group adopted is consistent with my detailed observations of subjects genuinely suffering from itch which I will present later. Despite the success of the studies using this system, Aoki's design has proved not to be robust enough for long-term use; mainly because the way the device had to be fitted made it impractical for home-use and resulted in the usual problems related to obtrusiveness of equipment on human behaviour.

The 1990's

An unusual method to measure scratch movement appears in the literature in 1991 (Mustakallio, 1991). Mustakallio describes his SCRADAR (Scratch Radar). The only information available on this device - which appears never to have been used in a clinical study - is an abstract. The device was described in more detail and demonstrated at a meeting. The SCRADAR required the subject to be covered in a fluorescent ointment whilst a radar was trained on the subject. The radar noted a disturbance in the fluorescence that occurred after scratch movements. No description on the frequency of radar sweeps or on how the data would be logged/analysed is given. It is difficult to ascertain how problems such as the cream rubbing off overnight with the usual nocturnal restlessness would have been avoided. The device appears not to have been used clinically at all.

There is a six-year gap in the literature before another device is described. In 1997 a 'Scratch-Monitor' was tested in atopics versus controls (Endo, Sumitsuji, Fukuzumi, Adachi, & Aoki, 1997). This device was a 25g box with a pressure sensor in the base which was taped to the back of the hand. After three or more pressure changes had occurred the number of scratch movements to occur was counted and the sensor was designed to filter out movements outside 1-2 Hz (again, this range was chosen after examining readouts from a sensor attached to a volunteer *simulating* scratch, the frequency stated results from a machine read-out, not direct observation). In practice, however, the design meant that there was considerable potential for variability in score: depending on how tightly the device was taped down (as it was a pressure sensor) also clenching the fist, for example, gave anomalous results. The authors talk about the 'frequency of scratch' being measured, this is not the case, however. The sensor, it seems was not sensitive enough to detect all movement and so a reliable frequency of cycles per second could not be made. They therefore rated the outputs by giving them an *arbitrary* number but then referred to the scores as 'frequency.' Use of the device was not extensive most probably due to its inconsistent sensitivity.

Two very similar systems involving piezoelectric sensors were inspired by the desire to quantify cholestatic itch (Molenaar, Oosting, & Jones, 1998; Talbot, Schmitt, Bergasa, Jones, & Walker, 1991). A piezo-electric pressure-sensor (acting as a contact-microphone) was glued to the nail of the middle finger then attached by a wire to a data processor and logger. The original design by Talbot's group was not portable and necessitated a hospital-based study whilst Molenaar's design utilised a chip, stored in a device worn on the waist to store data for later downloading. The pressure sensor was set to filter out movement outside 30 -100 Hz which was calculated, by Fourier transformation, as the frequency of nail vibration when scratching occurred. This appears to be a very high frequency until the length/flexibility of the nail as well as the ridging of skin is considered. The original and

improved piezo-electric devices have been used in clinical trial settings with some success (Bergasa et al., 1995; 1998; 1992). The device is, even so, rather impractical for day-to-day use and does not lend itself for chronic use - the subject cannot carry out their normal routines (including showering or bathing) with the device glued to the nail and probably do not behave naturally as they are aware of the system. Fitting the sensor requires trained hands. Apart from these standard criticisms, the sensors themselves are delicately hand-made items open to variation resulting in worries about score variability between devices.

In being very specific about the motion to be measured it is possible that some itch-related movements are excluded. The paper-strain gauge device, the 'Scratch-Monitor' and the nail-mounted monitor all aim to quantify *finger* movement and *specifically scratch*. It is evident, on watching itchy patients, that apart from the expected stereotypical scratch movements made with the hands, other movements also occur including rubbing with the hands, wrists and feet and rubbing the face into the pillow. A significant amount of time is spent rubbing the eyes with a balled fist. The paper-strain gauge, the 'Scratch Monitor' or the nail-mounted monitor would not detect these types of movement; therefore, a considerable amount of itch-related movement would go unnoticed.

Although time-intensive with regard to data processing, the present gold standard of scratch quantification is infrared video recording of itch-related movement. This was described in 1996 by Ebata's group (Ebata, Aizawa, & Kamide, 1996). Adult subjects (atopics and controls) were invited to the ward, videoed overnight and then the amount of time spent scratching compared between the two groups. The atopics were found to spend hundreds to tens of thousands of seconds scratching compared to the zero to one hundred seconds of controls. The same system was used, in hospital, to further characterise nocturnal scratch in atopics (Ebata, Aizawa, Kamide, & Niimura, 1999) and also to evaluate the effect of hypnotics on scratch (Ebata, Izumi, Aizawa, Kamide, & Niimura, 1998). Ebata's group

aimed to quantify the frequency of bouts only, not the frequency of action of scratch movements. Due to the fact that all itch-related movement can be logged and separated from restlessness, the system would appear to be sensitive and specific. Some difficulties occur when movement occurs under the cover, but since all the vast majority of itch-related movement is rhythmical, most of the movement under the covers can be classified and recorded (although rub versus scratch cannot be separated. A drawback to the system is that for some subjects it proves difficult to ignore a camera, video recorder and infrared light in the bedroom – even if it is their own.

The 2000's

Most recently, digital accelerometers have become the movement monitors of choice when assessing itch. Two groups have validated these machines against infrared video recordings although only Rees' group has observed subjects in the familiar surroundings of their homes (Bender, Leung, & Leung, 2003; Benjamin et al., 2004; Bringhurst et al., 2004).

Digital accelerometers were developed for the field of sleep research, for assessing tremors and to detect seizures. Available from two companies, commercially, they are similar in size, shape and weight to a wristwatch and, as such, are relatively unobtrusive. Their technology is described in detail in Chapter 2.

In validation studies the accelerometers scores showed a high correlation with the video observations and proved able to separate those with itch (atopics and others) from controls (Benjamin et al., 2004; Bringhurst et al., 2004; Ebata, Iwasaki, Kamide, & Niimura, 2001). It was observed in these studies, too, that the itch-related movements were not purely of the scratch-type, rubbing and writhing also occurred (Benjamin et al., 2004). Atopics were found to be asleep or motionless, on average, 46 minutes less than non-itchy controls - it must be assumed that atopics, therefore, expend much more energy overnight. The group showed that subjective scores (VAS of itch and sleeplessness) did not correlate well with the objective

accelerometer scores (Benjamin et al. 2004; Bringhurst et al. 2004). This result had been for the child subjects - where the parent is required to fill out the VAS score - but in fact, there was a (slightly) better correlation for children's scores than for the adults who had scored themselves.

The accelerometers are not, however, able to discriminate between itch-related movements and generalised movement. Even after one group set a 'threshold value', accelerometer scores were found not to be discriminatory (particularly against subjects getting up in the night, for example (Ebata et al., 2001)). The fact that the devices cannot discriminate between movements may *not* be considered a great downfall if it is accepted that, due to discomfort, itchy people do just *move more* that controls (and that therefore a cumulative score of movement is satisfactory). The lack of specificity may explain why the expected pattern of intensity of itch and related scratch according to stage of sleep throughout the night (as observed in Savin's studies) has not been demonstrated yet.

The present

When I first entered this field of research, the accelerometers were 'cutting edge technology.' At this point, I suspected that the (distant) future of the monitoring of itch-related movement was most likely to be a spin-off of CGI (Computer Generated Image) technology such as that used in the 'movie industry.' Sensors, which resemble ECG pads, are placed on specific sites of the body, limbs, even parts of the face and these provide reference points so that a computer, wirelessly tuned into these reference points, can generate an image which moves realistically. One example of the development of this technology used in medicine, is a vest which can monitor respiratory rate and ECG by using sensors woven into the cloth ('Smart shirt,' SensaTex, New York). This has been used to monitor firemen as they carry out their work. Another example is of sleeves with monitors woven inside, which have been developed to help assess the extent of flexion/extension after joint replacement. More ambitious designs include a wireless body area network of intelligent motion sensors which,

it is hoped, will assist rehabilitation (Jovanov, Milenkovic, Otto, & de Groen, 2005). The wearable networks can be integrated into clothing and use wireless technology - for ease of wearing and for transmission of data. The sensors transmit to a 'personal server,' for example, an enabled PDA, mobile phone or home computer carried on/nearby the subject and the 'personal server' downloads, remotely, to the health provider's computer. The health provider uses the information to direct/develop the rehabilitation exercise programme.

I did not predict the rapidity of development of technology for the fitness/personal training industry. Vast amounts of information can be gathered from technology which is not just portable but wearable (Jawbone, Applewatch, etcetera). As it stands, these devices have become ubiquitous and possess accelerometers, gyroscopes and are GPS enabled and so detailed movement data could be mined. What is disconcerting, is the report of variable sensitivity depending on the device and processing software. Indeed, worryingly, I detected processing errors in the software supplied with both of the accelerometers I used for my studies. Therefore, whilst the idea of an app to install on a patients' mobile device is attractive, there would be issues with interdevice variation which would make stansardised treatment protocols difficult to impose. Information from these modern, commonplace devices have a feature which would help solve a problem I found in my research: how to discern generalised locomotive movement from itch-related. GPS movement information would be able to discern this. No itch-app is available yet, but it not a huge leap to imagine that one could be designed. Not only could this capture objective movement data but offer the availability of daily, or more frequent diarising of subjective itch – one of the least biased way to monitor subjective disease.

VII: Disease-measurement tools used in this thesis

A number of tools have been developed with the goal of measuring the functional burden of disease as experienced by the patient. I introduced the idea in the 'Quantification of itch.' In a medical and economic climate where resource is scarce, the assessment of impact of disease and how it may be influenced by medical intervention, has become a major research programme. In England, the British Association of Dermatologists and the National Institute of Health and Clinical Excellence (NICE) has advocated the use of the Dermatology Life Quality Index (DLQI) and Psoriasis Area and Severity Index (PASI) as disease assessment tools for doctors to use on patients with psoriasis to determine whether they should receive certain expensive biological therapies (C. H. Smith et al., 2005) (http://www.nice.org.uk/Guidance/TA134).

The measures employed in this study have been chosen because they are used ubiquitously or are recommended in guidelines from bodies such as NICE (National Institute for Clinical Excellence). Descriptions have been limited to those directly pertinent or used in this thesis.

Disease-measurement tools, are divided into those assessing the *impact* of disease and those assessing *disease-extent*. As previously mentioned, the aim of developing these tools was to have a standardised measure, across the board. The problem with the disease-extent tools which require an estimation of skin involvement, is that intra and inter-observer variability occurs (Charman, Venn, & Williams, 1999; 2002), even with the aide of the 'rule of 9's' (Berth-Jones et al., 2006).

Disease impact measures used in this thesis

Disease-impact measures can be divided into generic, system specific and disease specific. The ones used in my studies follow.

1. Generic: Global health question (GHQ).

‘In general, would you say that your health is excellent, very good, good, fair or poor?’ (Ware, 1976). Poor scores ‘0’ and Excellent scores ‘4.’ The question takes seconds to complete. It is able to provide an overall assessment of patient-perceived health. Although deceptively simple, the question is powerful: decay of results has been demonstrated towards ‘end of life. (Bjorner, 2005)’ It is thought that it is as a function of its simplicity that respondents are able to combine information from all psychological and physical domains to give an unbiased, complete response.

2. Generic: SF-36 or RAND-36

This is the most extensively used health related quality of life questionnaire. It consists of 96 items and takes 5-10 minutes to complete (Stewart, Hays, & Ware, 1988; Ware & Sherbourne, 1992). The system was developed from the Rand Corporation's Health Insurance Experiment in the U.S.A. (Ware & Sherbourne, 1992). Questions pertain to the individual's typical day, the past 4 weeks, and experiences in general. The SF-36 measures health-related quality of life with 36 items along eight dimensions yielding one physical and one mental component summary score. The eight different dimensions are: (i) physical function: limitations in physical activities because of health problems; (ii) role plus or minus physical function: limitations in usual role activities because of physical health problems; (iii) bodily pain; (iv) general health perceptions; (v) vitality: energy and fatigue; (vi) social function: limitations in social activities because of physical or emotional problems; (vii) role plus or minus emotional function: limitations in usual role activities because of emotional health problems; and (viii) mental health. The eight dimensions range in score from 0 to 100, with higher scores indicating higher levels of function and/or better health.

3. System specific: Dermatology Life Quality Index (DLQI).

The DLQI was developed by Finlay and Khan in 1994 (Finlay & Khan, 1994) as a disease-specific measure intended to be simple, compact and applicable to patients with any skin disease. The DLQI is a questionnaire that measures how much a skin problem has affected the life of the patient over the previous seven days. It consists of ten items, six dimensions and one overall summary score. Each question has four answering alternatives: 'not at all', 'a little', 'a lot' and 'very much', with corresponding scores of 0, 1, 2 and 3, respectively. The overall summary score aggregates the score of each item and ranges between 0 (the best score) and 30 (the worst score). The six dimensions are: (i) symptoms and feelings; (ii) daily activities; (iii) leisure; (iv) work and school; (v) personal relationships; and (vi) treatment. This tool was developed after studying 120 consecutive dermatology patients and evaluating the areas of life most affected. The designers of the tool had practicality in mind - it was considered a priority that the tool should be short and quick to answer, in order to aide assessment during clinic visits –and they achieved this. In a ten year review of the DLQI's use, it was noted to have been used in 36 different skin diseases and to have been translated into 21 different languages (V. Lewis & Finlay, 2004).

4. Disease-specific: Visual Analogue Scale (VAS),

These can be used as a measure of a range of specific symptoms such as itch or pain (Hägermark & Wahlgren, 1992; Wahlgren, 1995). They have been described in an earlier section (IV, Psychophysics). In my studies and in other validated scores (Atherton et al., 1993) the visual analogue scales are standard 0-100mm scales or lines where 0 represents the most negative experience and 100 the most positive. The participant is required to make a mark on the line corresponding to the extent relevant to them at that time. Three horizontal visual analogue scales were used to assess the respondent's subjective measure of disease activity. The first question assessed the respondent's amount of itch on that day, with anchors

of ‘No itch’ (at 0) and Most terrible itch’ (at 100). The second VAS assessed sleep disturbance, with anchors of ‘No sleep’ (at 0) and ‘Best possible sleep’ (at 100). The third assessed disease extent, with anchors of ‘No disease’ (at) and ‘Worst possible disease’ (at 100).

5. Generic: The Day Reconstruction Method

This is a questionnaire which employs experience sampling methods to gain a report of emotions experienced whilst avoiding some the recall biases frequently encountered in questionnaires (Kahneman, 2004). Experiential Momentary Assessment (EMA) is a way to access experiential rather than stereotyped semantic memories (Robinson & Clore, 2002b; 2002a). The gold standard for accessing this memory is on-line reporting but the Day Reconstruction Method was developed as a more ’practical’ solution: it helps the respondent access the experiential memories by making them write a detailed diary of the previous day and then answer questions about each episode of the diary. The questionnaire consists of four packets:

Packet 1: Global satisfaction scores and demographics

Packet 2: Diary of previous day, required to break the day down into episodes

Packet 3: Questions about each of yesterday’s episodes

Packet 4: Questions about yesterday in general.

The questionnaire, according to the literature, takes 30-45 minutes to complete. This questionnaire has never been used in medical or dermatological research previously.

Disease-extent tools used in this thesis

There are a myriad of disease-extent tools available to dermatologists. An especially large range are available for scoring atopic dermatitis: EASI (Eczema Area Severity Index) (Tofte, 1998), SASSAD (Six area, Six sign, Atopic Dermatitis Score) (Charman et al., 2002), NESS

(Nottingham Eczema Severity Score) (Emerson, Charman, & Williams, 2000), to name a few but **SCOR**e for **A**topic **D**ermatitis extent (SCORAD) was most suitable for my studies as it integrated physician-assessed disease extent with subjective scores of itch and insomnia. I used the subjective itch score to compare with my objective accelerometer score.
SCORAD is the recognised extent tool for the European Working party for atopic dermatitis (Atherton et al., 1993; Kunz et al., 1997). SCORAD takes roughly five minutes to complete:-

First, SCORAD involves an estimation of area affected (A): a diagram using the 'rule of nines' is shaded to provide an interpretation of percentage of area affected. Second, severity of clinical signs is scored (B): a score of '0' for absence, through mild, moderate and severe, which scores '3', is applied for each of a) erythema, b) oedema/papulation, c) oozing/crust, d) excoriation, e) lichenification and f) dryness. Finally, VAS scores of itch and insomnia are completed (C): '0' indicates no itch or no sleeplessness and '100' indicates worse possible for itch or sleeplessness. These details are processed to give a score through the equation, $A/5 + 7B/2 + C$.

Psoriasis Area and Severity Index (PASI) (Fredriksson & Pettersson, 1978; Marks, 1989)was my obvious choice when the psoriasis extent studies were proposed. This is the disease extent tool advocated for psoriasis by NICE (National Institute of Clinical Excellence) (C. H. Smith et al., 2005). This tool requires no subjective input from patients and so it is a pure extent tool. PASI is assessed as follows:-

First, a representative area of psoriasis is selected for each body region: head and neck, arms, trunk and legs. The intensity of redness, thickness and scaling of the psoriasis is assessed as none (0), mild (1), moderate (2), severe (3) or very severe (4).
Second, the three intensity scores are added up for each of the four body regions to give

subtotals A1, A2, A3, A4. Each subtotal is multiplied by the body surface area represented by that region: A1 x 0.1 gives B1, A2 x 0.2 gives B2, A3 x 0.3 gives B3, A4 x 0.4 gives B4.
Third, the percentage area affected by psoriasis is evaluated in the four regions of the body. In each region, the area is expressed as nil (0), 1-9% (1), 13-29% (2), 30-49% (3), 50-69% (4), 70-89% (5) or 90-100% (6).
Fourthly, each of the body area scores is multiplied by the area affected: B1 x (0 to 6) = C1, B2 x (0 to 6) = C2, B3 x (0 to 6) = C3, B4 x (0 to 6) = C4.
Finally, the PASI score is calculated by: C1 + C2 + C3 + C4.

VIII: Reasoning for concern about biases in subjective symptom tools

Over the past twenty-five years, the field of cognitive behavioural psychology has developed. It has become important to measure well-being or happiness. This importance is borne out of continuing allegiance to Jeremy Bentham's ideas: maximising goodness for all, to clumsily paraphrase. If it has proven hard to capture itch, I am only relieved that I have not had to capture happiness: there is no reflex motor reaction for this. In any event, in trying to measure happiness, a number of problems surrounding the measurement of subjective states, quality of life and utility have been revealed (for reviews see 'Well-being: the Foundations of Hedonic Psychology' by Daniel Kahneman (Kahneman & Diener, n.d.))

Individuals appear to have trouble accessing their own feelings (Robinson & Clore, 2002a; 2002b; D. L. Thomas & Diener, 1990). The way information is gathered may alter, or influence, the patients own perception of their own feelings (McColl, 2005; Robinson & Clore, 2002a). Cognitive limitations may limit the value of subjective knowledge. This means that patients, due to adaptation or the hedonic treadmill, may not be able to remember

changes in their functional status, nor predict the effects of particular interventions or change in state (Kahneman, Krueger, Schkade, Schwarz, & Stone, 2006; Redelmeier, Rozin, & Kahneman, 1993).

In an effort to translate these findings to my research, I concentrated on the effect of two biases: framing and focusing on commonly used subjective disease tools.

'Framing bias' is the term widely used to explain the bias that stems from how the feeling, emotion or symptom is enquired about. Questions can be 'framed' in language or presented in a context that may elicit a stereotyped answer, for instance, by implying than an aspect of disease should be considered as a negative phenomenon, leading a respondent who may not have considered this aspect as negative before could, in this situation, consider it as such (McColl, 2005; Wright & Goodwin, 2002). As will be seen later, the enquiries in the DLQI questionnaire were rewritten in seemingly more neutral frames to assess whether it was vulnerable to this type of bias.

'Focusing' (bias) describes the way that the immediate context may alter how individuals perceive their own symptoms. For example, patients frequently anchor or skew their own assessment of disease by reference to others who they think are less or more fortunate (Del Missier, Ferrante, & Costantini, 2007; Kahneman et al., 2006; McColl, 2005; T. D. Wilson, Wheatley, Meyers, Gilbert, & Axsom, 2000). In order to test this I utilised sets of words used in psychology research which are proven to elicit emotions: negative, positive and neutral to see if response patterns changed in the face of this focalising. I also used a film about a person living with a distressing skin disorder in another study to see if the focus would change the way participants felt about their disease.

Ethics

Ethics committee approval for these studies was granted from Lothian Ethics Committee. The study reference numbers issued were:

1) LREC/2005/JLR/2/CM) **05/S1104/25**
2) Derm/06/JLR/1/CM **(6/S1104/56)**

So far as they can apply, I believe the protocols used in this work are in line with the Declaration of Helsinki (URL: http://www.wma.net/e/policy/b3.htm) and subsequent amendments on the World Medical Association website (URL: http://www.wma.net).

Statistical analysis

Significance:

- The level of significance was set to 0.05, throughout the studies.

Comparison of two groups:

- If data was parametrically distributed, Student's t test was undertaken.
- If the groups were within person, paired t tests were applied and if not unpaired t tests were applied.
- In some studies, the variance was difference between groups, specifally in accelerometer studies where subjects were compared to controls (heteroscedasticity), and so a t test to compare means of groups with different variance were used.
- If data was non-parametrically distributed, the Kruskal-Wallis test was applied.

Comparison of several groups:

- In this thesis, when comparison of variance across several groups was required, the data was found to be normally distributed and so one-way ANOVA was applied.

Correlation and correlation coefficients:

- Perfect correlation is represented by an 'r' level of 1.
- As a general 'rule of thumb' in this thesis, where mechanical components were being tested (for example, the accelerometers against each other), the

relationship was required to be as near to 1 as possible and preferably, over 0.95, as this is the accepted standard in engineering/mechanics.

- When biological systems were being compared (to each other or to mechanical components) an r level of over 0.5 was considered to represent a strong relationship.
- Pearson r was used to describe normally distributed, linear data.
- Spearman rank correlation was used for non-normally distributed data.

Regression analysis:

- Regression is used in Chapter 5. Following inspection using univariate analyses, non significant terms were removed from the regressions, and examination of factors were performed using Holm' s correction for pairwise t-tests
- Linear analysis was employed in Chapter 7. R and R square have been presented along with standard error and ANOVA's. The intercept, slope and standard error of the slope are presented too.

Autocorrelation:

- Longitudinal scoring patterns of objective accelerometer measured itch and subjective VAS score in Chapter 5 were analysed for autocorrelation.
- This type of analysis is one more commonly used to predict weather or financial fluctuations.

Nesting:

- Most of the experiments in Chapter 5 'nested' previous conclusions in each successive experiment, implicitly testing the conclusions of previous analyses.

Power

- Formal power calculations were not performed as the experiments and studies described in this these were exploratory. There were no previous studies upon which to base power calculations.

Software:

- All software was run on Mac OS X (Apple, California, USA).
- For general data management, Excel (Microsoft Ltd,, Seattle, USA). was used.
- Basic statistical analysis was carried out in Excel, using its Data Analysis plug-in.
- IBM SPSS, version 21, (IBM Corporation, New York, USA) was used for my analysis
- R, version 2.9 (R: The R Project for Statistical Computing, https://www.r-project.org) was used for analysis by Professor Rees.

The majority of statistical analysis was carried out by me. Professor Rees undertook the autocorrelation and regression analysis presented in Chapter 5 and published in the paper in Appendix I (Murray & Rees, 2011a).

Chapter 1

Direct observation of subjects in order to quantify and characterize itch-related movements.

I: Introduction to video study of itch

People have often asked me what my field of research is. I have explained that I am attempting to measure itch. The usual response is polite interest and then a change in conversation subject. A couple of people have asked me what 'I.T.C.H.' stands for. A few have asked to hear some more detail. If I have been invited to expand, I have explained that itch is coupled to scratch and that this is very fortunate as, unlike other sensations, there is a coupled physical response. I have explained that this scratch movement can be observed to give an 'itch-score.' I explain that direct-observation (and preferably a recording of this) is the most accurate way to measure the scratch movement that ensues from the itch sensation.

Direct observation of itch-related scratch movement is not a new phenomenon. The first studies started off in 'Sleep laboratories' (Felix & Shuster, 1975; Savin et al., 1973; 1975) and, as technology improved and became more portable, it was possible to actually undertake the study in the subject's own home (Benjamin et al., 2004; Bringhurst et al., 2004). Studying participants in their own homes is obviously preferable on many levels, but most particularly because the subject's behaviour does not have the potential for being altered by being in an 'unnatural' situation.

Direct observation of itch related scratch movement remains the 'gold-standard' to which all other methods of itch-related scratch movement quantification need to be measured against. It therefore made sense, since my purpose was to quantify itch, to start my research using this technology. In this chapter, I present details on the system I developed for my particular purpose and present some experiments carried out with the system.

II: Infrared videoing system.

The system comprised:

1) an infrared light source

2) an infrared camera on tripod

3) a time generator

4) a particular video cassette recorder (VCR): Panasonic NV-HS880 fitted with 'Extra long play.'

Figure 1.1 is a photograph detailing the equipment.

The infrared source was positioned to shed light across the subject's room. The infrared camera was positioned on a tripod to allow an unimpeded view across the subject's sleeping form. The camera was then connected to the video cassette recorder (VCR) and the recorder switched on. A specific type of VCR was required as this allowed 'Extra long play' which meant that a standard 4-hour VCR tape could record over 12 hours.

The mini-television was connected to the video recorder and the 'Time generator.' The 'Time generator' was set to accurate time. The mini-television allowed a way of ensuring that the 'Time generator' output was detected in the recording and that the positioning of the camera was seen to be satisfactory for capturing an unimpeded view of the subjects' sleeping form.

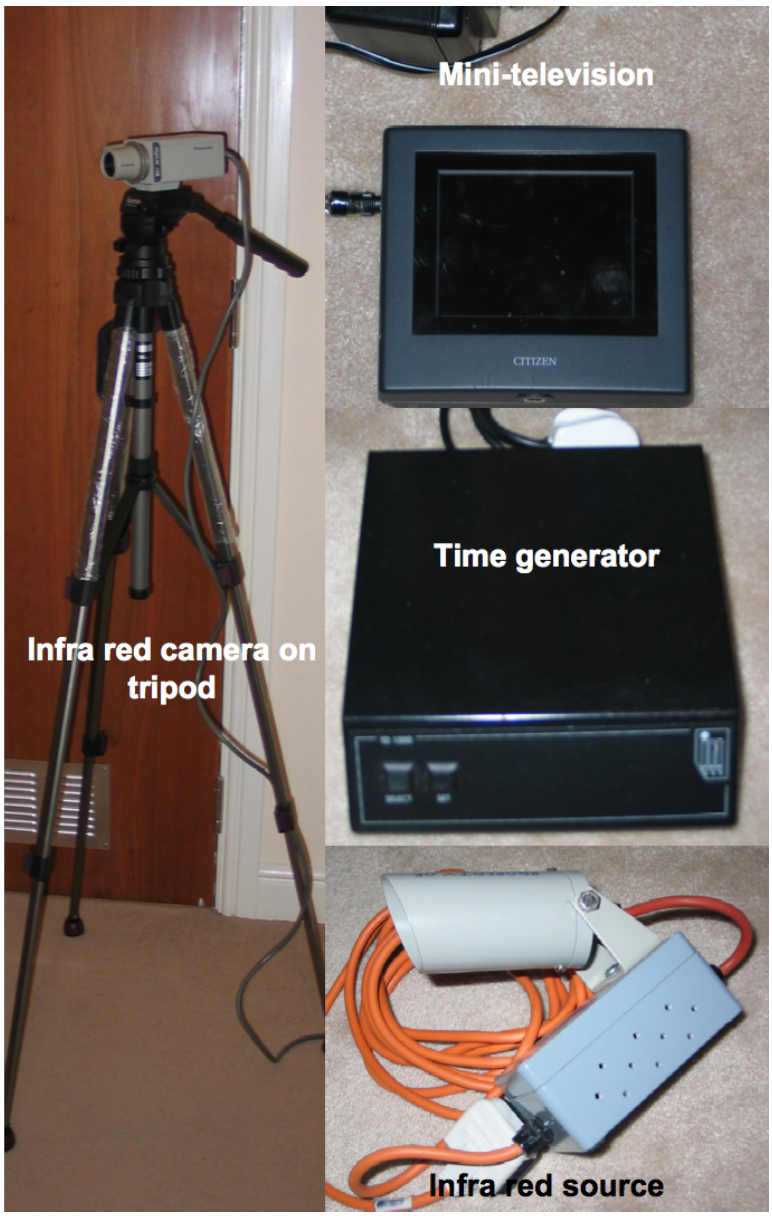
A 4-hour tape was inserted into the VCR. The VCR was set to record in 'Extra long play' mode and a test recording made for a few seconds and functionality checked. The recording was then stopped and the VCR left, turned on. When bedtime came, the parent was requested

to press the red, labelled, ‘Record’ button. At waking time, the parent was requested to turn the VCR off. Parents received full instructions in written, verbal and demonstrated forms.

This system improved upon previous systems which only allowed eight hours continuous recording: a child may have been sent to bed at 18:00hr and, if the recording was started immediately, recording would then cease at 02:00hrs, before the child awoke and out-with the Actiwatch’s original Macro’s output times, 22:00hr until 06:00hr. As I will explain, I created a new Macro which extended the available Actiwatch Plus-monitored time (see Chapter 2).

It should be noted that the system was placed as unobtrusively as possible within the room and merely effected an insignificant red-glow along with the quiet purr of the VCR and so it was minimally distracting to the study-subjects.

Figure 1.1. Photograph demonstrating all components of the infrared video system



III: Characterising the *'frequency of action'* of itch-related movement from direct observation of itchy subjects.

Aim

The objective of this study was to characterise the frequency, in Hertz (Hz), of human, nocturnal itch-related movement.

Background

When considering the design of a *specific* itch-monitor, there are two features of disease-related movement which can be exploited: the *frequency of action* of the movement and the *length of bout* of movement. Definitive data on the frequency of action of *human* itch-related movement was not available in the literature (whilst *length of bout* was (Savin et al., 1973; 1975)). Data on the frequency of action of other animal scratch are available: murine scratch, 10-30Hz (Brash et al., 2005); rat scratching, average frequency 8Hz (De Castro-Costa, Gybels, Kupers, & Van Hees, 1987) and cat scratch, average frequency 6.5Hz (Kuhta & Smith, 1990). Observations of primates in the literature state the *frequency of the bouts* of scratching, not the *frequency of action* of scratch. Without knowing the *frequency of action* of human scratch it was not possible to know what kind of range of frequencies as measured by new accelerometers should identify itch/disease-related movement.

Method

Study subjects

Eight children, all of whom had a diagnosis of atopic dermatitis, were studied. The children were recruited from a secondary care Paediatric Dermatology clinic and had been were diagnosed as having atopic dermatitis by a Consultant in the clinic (as described by Hanifin and Rajka (Hanifin et al., 2001)). Children with atopic dermatitis were chosen because their having itchy skin is a requirement for their diagnosis - so they would be expected to reliably

demonstrate itch/disease-related movement. The severity of disease was, as would be expected by the subject's being recruited from secondary care, moderate: SCORAD range 13.7 to 90.5, mean 33.3, median 32.2. The group's median age was 6 years (age range 2-13yrs, median age 6yrs, four male and four female). One night's video of each was examined.

Study method

Each study subject's nocturnal movements were observed using infrared video recordings. This involved my visiting the subject's house and setting up the infrared videoing system (see Section I.II).

The subject's parent was issued with verbal and written information including being asked to press the 'Record' button on the VCR when the child went to bed. The equipment and tape was recovered from the subject's house the next day. The infrared video recordings were played back in 'real-time' and the following observations were made:

1) For each hour of sleep the next
    a. generalised movement,
    b. scratch,
    c. rub,

   to occur after a specified, randomised time-point was observed and characterised.

2) For each of the episodes of movement,
    a. the length of the bout, measured *in seconds* with a stopwatch and
    b. the number of complete cycles of movement to occur within the episode

   were recorded. A 'bout,' or cycle of movement was counted when the part of the body being observed reached its starting point again.

A firm definition of each category of movement was made.

a) **Definition of scratch:** a ***rhythmical*** movement during which the *fingernails make contact* with the skin.
b) **Definition of rub:** a ***rhythmical*** movement where *fingernails are not the point of contact* with the skin, for example, rubbing the wrists together, rubbing eyes with a balled fist.
c) **Definition of generalised movement:** a ***non-rhythmical***, adjusting movement.

When the frequency of an action is described, in the unit Hertz (Hz) it is considered to represent the *number of complete cycles of movement completed in a second.* Therefore, the observations were converted into frequencies by dividing the number of completed cycles to occur in the observed episode by the number of seconds the episode lasted for.

The range of overnight recording was from eight to twelve hours, depending on how long the child slept for. Potentially, there should have been eight to twelve observed episodes for each category of movement, however, sometimes a subject did not scratch or rub in a one-hour period in which case no observation could be made. The definition about whether an observed action was scratch or rub (viz. contact with fingernails) was strictly adhered to and if there was any uncertainty then consecutive episode was studied.

Results

All of the subjects exhibited itch/disease-related actions. All demonstrated *stereotypical* scratch and rub movements. As noted in previous observatory studies, the children also demonstrated what may be considered *non-stereotypical* movements such as: rubbing the wrists together, rubbing the feet together, generalised body writhing-rubbing, eye-rubbing and kneeling up to rub the face into the pillow. Although non-stereotypical, these

movements were *all rhythmical*. All of the subjects also 'turned over,' adjusted the bedclothes and made other generalised, non-rhythmical movements.

The mean '*frequencies of action'* for each movement were:

- Generalised movement, 0.48Hz (standard deviation, ± 0.25).
- Rub, 0.98Hz (standard deviation, ± 0.36).
- Scratch, 1.85Hz (standard deviation, ± 0.55).

Individual's mean 'frequency of action' for each observed category of movement and the standard deviation is listed in Table 1.1. When analysed by one way ANOVA, the categories of movement are significantly different: $F_{2:21}$=56, p<0.001. There was some overlap of frequencies between the categories of movement, as illustrated in Figure 1.2 (red X = mean, red _ =median).

Individuals' 'frequency of action' rose consistently within person through 'Generalised movement', 'Rub' and Scratch. Figure 1.3 illustrates this finding.

Table 1.1. Mean (and standard deviation) directly-observed ‘frequency of action’ for each subject

	Adjusts position		Rub		Scratch	
	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
Subject 1	0.488	0.265	0.734	0.128	1.639	0.300
Subject 2	0.443	0.239	1.280	0.412	1.957	0.264
Subject 3	0.391	0.201	0.995	0.438	1.933	0.062
Subject 4	0.485	0.224	1.044	0.250	1.774	0.309
Subject 5	0.264	0.153	0.797	0.214	1.444	0.514
Subject 6	0.768	0.000	1.072	0.225	2.421	0.961
Subject 7	0.316	0.168	0.725	0.393	1.381	0.236
Subject 8	0.430	0.247	1.185	0.394	2.221	0.328

Figure 1.2. Hourly directly-observed ‘frequency of action’ for each action and for each subject, n=8

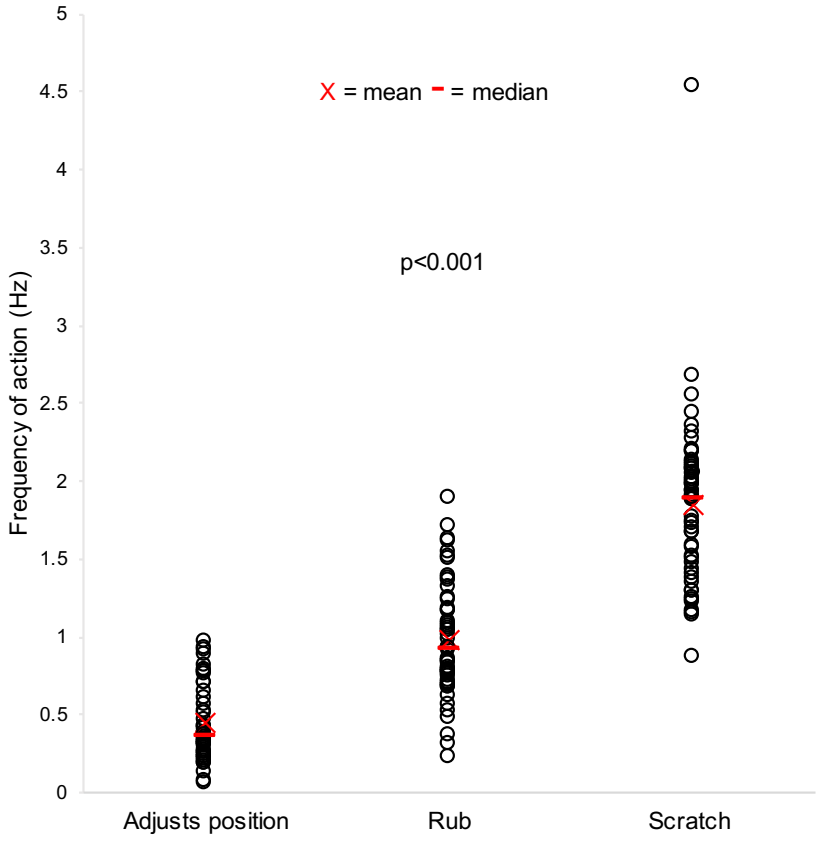
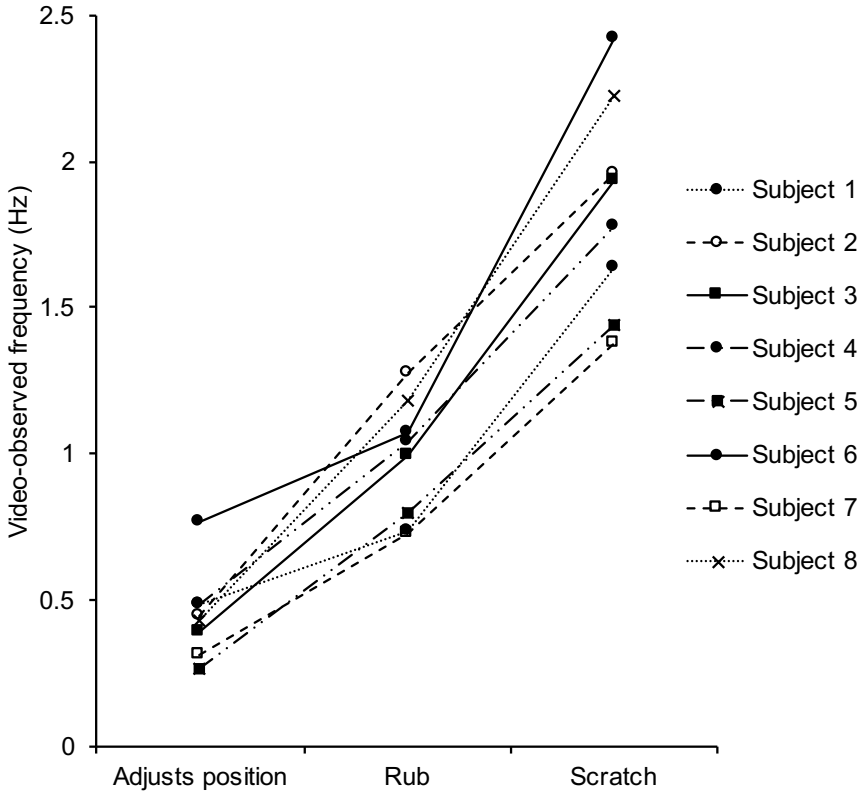


Figure 1.3. Mean directly-observed ‘frequency of action’ for each action and for each subject, n=8.



Conclusion

The mean values for '*frequency of action'* for each category of movement were significantly different. This means that it should be possible to distinguish scratch, rub or adjusting position based on the video-observed frequency of the action. There is, however, a lot of overlap in range of frequencies that each category of movement includes. There is less overlap if *all* itch-related movement, i.e. scratch *and* rub are considered as one group**.** Sometimes it is difficult to differentiate between scratch and rub so it could, therefore, be argued that it is valid to include these two categories of *rhythmical movement* together in opposition to the *non-rhythmical* 'generalised movement.' Grouping rub and scratch is also valid as both of these actions inflict damage upon the skin and, it is presumed, result in the clinical appearance of the rash of atopic dermatitis

The frequency of *human* scratch proved much lower than murine scratch: 10-25Hz. The observed range of frequency of action of itch-related movement is, however, consistent with Shuster's stated frequency of scratch, as measured by the bed-leg monitor, 2Hz (Felix & Shuster, 1975). It is interesting to note how the frequency of action of scratch would appear to be inversely proportional to the size of animal.

Overall, this study achieved its aim, namely, to characterise the frequency of human itch/disease-related movement. It has been shown that, by direct observation of subjects, it is possible to separate itch-related movement from generalised movement.

IV: Comparison of 'frequency of action' of human scratch compared to other animals.

Aim

To characterize itch-related movement for a range of mammals and compare their characteristics to human itch-related movement.

Background

A literature search for information on the frequency of action of scratch movement in humans and other animals revealed a dearth of information. There was clear literature to demonstrate that the frequency of action for murine scratch was 10-30Hz and 12.5Hz on average (Brash et al., 2005). This information was available from an objective measure of scratch. There was also published data on rat (8Hz) (De Castro-Costa et al., 1987) and cat (6.5Hz) (Kuhta & Smith, 1990) scratch-frequency of action. With respect to primates, however, there was no information on the 'frequency of action.' The focus of objectively measuring scratch in primates shifts away from the actual 'frequency of action' to how long 'bouts of scratch' are. Some human studies had mentioned putative 'frequency of action' of itch from analysis of electronic recordings (Aoki et al., 1980; Felix & Shuster, 1975; Savin et al., 1973; 1975) but none had measured the action through gold-standard direct observation.

From the published findings, it was hypothesised was that 'frequency of action' of scratch movement was inversely related to the size of the animal. The theory was that the speed of the movement (frequency of action: number of limb-cycles per second) should more precisely be related to limb length. In the absence of information exactly pertaining to limb length, published mean height or length and mass of standards of the species were used as a surrogate.

Method

A literature search was undertaken to glean any information on frequency of action of scratch in animals (*Felis catus* (cat)*, Rattus norvegicus* (rat) *and Mus musculus* (mouse). In order to glean more information than that available in the published literature, animals were directly observed: one day was spent at Edinburgh Zoo, if animals were spotted scratching, they were videoed opportunistically. YouTube was searched for video recordings of animals scratching. To qualify for the study, the scratch movement had to be undertaken by any limb, in clear view of the camera in such a way that digits could be seen making contact with another part of the body.

For each of the episodes of movement,

1) the length of the bout was measured *in seconds* with a stopwatch and
2) the number of complete cycles of movement to occur within the episode were recorded,
3) a 'bout,' or cycle of movement was counted when the part of the body being observed reached its starting point again.
4) the episode was then re-watched two more times and mean values taken.

The frequency of an action is described in the unit Hertz (Hz). It is considered to represent the *number of complete cycles of movement completed in a second.* Therefore, the observations were converted into frequencies by dividing the number of completed cycles to occur in the observed episode by the number of seconds the episode lasted for.

Results

During the visit to Edinburgh Zoo, mammals *Saimiri boliviensis* (squirrel monkey) *and Nomascus leucogenys* (white cheeked gibbon) were successfully video-recorded whilst

scratching. Due to the unpredictable nature of scratch bouts, further recordings were unsuccessful (due to difficulties capturing the event in time) despite other animals also being seen scratching on the day of the visit. The YouTube search yielded qualifying recordings of *Pan troglodytes* (chimpanzee) and *Gorilla graueri* (gorilla) scratching.

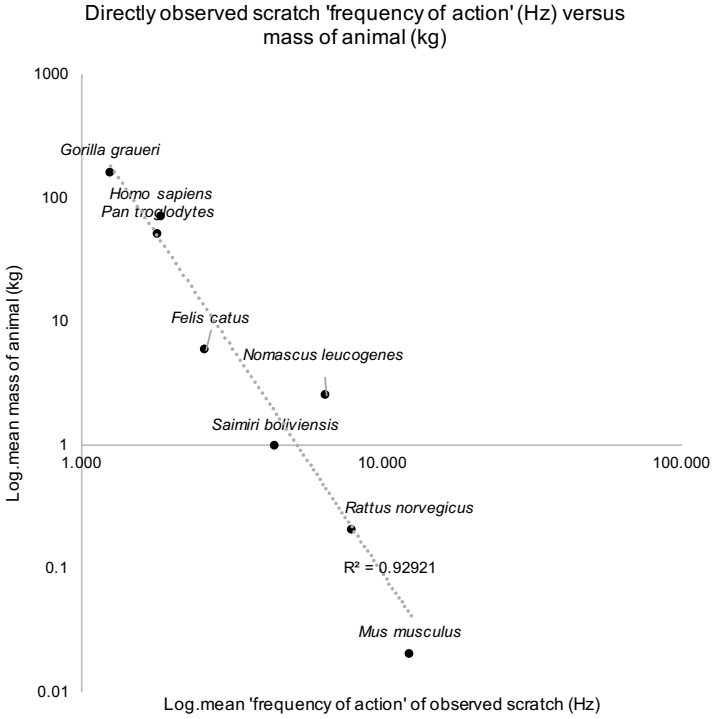
The 'frequency of action' observed from the video-recordings are summarized in Table 1.2 alongside standard lengths and masses of the animals (mean for male/female) sourced from the National Primate Research Center, University of Wisconsin – Madisons's website, 'Primate Info Net' and from University of Michigan Museum of Zoology 'Animal Diversity Web' website. Mean and standard deviation (for the animals I observed) are stated. It can be seen that 'frequency of action' does correlate with mass and length of the animal. The smallest animal, the mouse, exhibits highest frequency (or fastest) scratch with a mean frequency of 12Hz, and the largest animal, the gorilla, scratches with a frequency of 1.3Hz. Humans, which are the second biggest mammals observed, scratch at 1.8Hz on average.

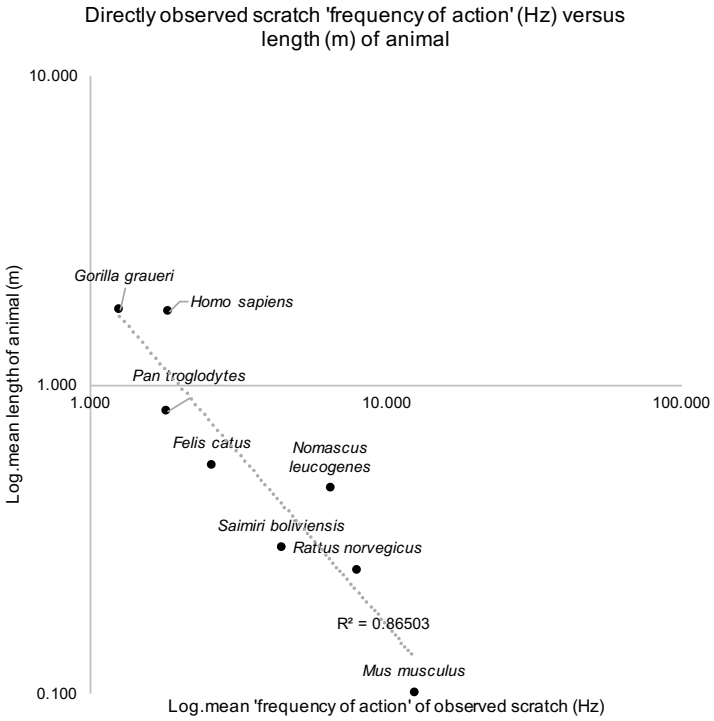
Figure 1.4 includes graphs of log. mean 'frequency of action' for each animal versus log. mass and log. length. There is a linear relationship indicating a strong relationship between both average log. mass and average log. length versus frequency of action. The relationship is best expressed as allometric scaling.

Table 1.2. Video-observed 'frequency of action' in different animals compared with length and mass. *N.B.: Standardised measurements of length and mass sourced from the National Primate Research Center, University of Wisconsin – Madisons's website, 'Primate Info Net' and from University of Michigan Museum of Zoology 'Animal Diversity Web' website.*

	Mean scratch frequency of action, Hz (standard deviation)	Length(m)	Mass (kg)
Gorilla graueri	1.248 ± 0.186	1.750	157.500
Homo sapiens	1.850 ± 0.361	1.730	70.000
Pan troglodytes	1.807 ± 0.145	0.816	50.000
Felis catus	6.500	0.460	2.500
Nomascus leucogenys	2.599 ± 0.033	0.546	5.700
Saimiri boliviensis	4.447 ± 0.186	0.294	0.950
Rattus norvegicus	8.000	0.250	0.200
Mus musculus	12.500	0.100	0.020

Figure 1.4. Graphs to show log. ‘frequency of action’ against log. mass (top) and log. length (bottom). Correlation and coefficient added.





Conclusion

This small experiment was an exploratory gesture to see how well the observed 'frequency of action' of human itch fitted with that of other species of mammal. The lack of published information required the video-observations to be undertaken.

It can be seen that there is a powerful relationship between mass and 'frequency of action' of scratch. There is a less powerful relationship between 'frequency of action' of scratch and length. It had been hypothesized that the greatest relationship would be between limb length and 'frequency of action' but the information on species standards for this data was not readily available. It had been hypothesized that length or height would correlate most closely with 'frequency of action' but the results do not concur with this. This could be explained by the fact that length is quoted as head to tail, a length of head to base of the spine would possibly be more relevant especially since humans do not have tails.

It is intriguing to see the demonstrated allometric relationship between the 'frequency of action' and mass or length. Allometric relationships are a feature of biological systems and well known examples include the relationship between Basal Metabolic Rate (BMR) and body mass and perhaps more significantly, for legged motion – indeed length of limb has been used to predict mode of action of locomotion for extinct creatures.

Overall, the study confirms that size is related to 'frequency of action' of scratch and that based on humans' standard sizes, the 'frequency of action' ascertained from our direct-observation experiments are in keeping with what would be predicted on this basis.

Chapter 2

Introduction and further validation of the use of accelerometers to measure itch-related movement

I: Introduction to accelerometers and the Actiwatch Plus

Actiwatches are wrist worn accelerometers. The 'Actiwatch Plus' were bought from Cambridge Neurotechnology, now known as CamNtech Ltd (Cambridge, U.K).

Accelerometers measure 'proper acceleration.' This means they measure acceleration relative to free fall. If an accelerometer is lying at rest on the floor, it measures gravity, 1G *upwards*, whereas if it were in freefall it would measure 0G.

The accelerometer works by containing a mass on a spring. When there is acceleration, the mass is displaced and the spring is able to accelerate the mass at the same rate as the casing. Measurement of the displacement gives the acceleration. In the Actiwatch accelerometer, a piezoelectric crystal is used to convert the mechanical movement into an electrical signal, see Figure 2.1.

The accelerometer integrates intensity, amount and duration of movement. The actiwatch senses movement 32 times per second and logs the highest counts every second. The epoch score is the sum of those highest scores per second. The Actiwatch is an analogue analyser, this means it uses a variable bandpass filter whose mid frequency is tuned to the range of frequencies to be measured. The unit Actiwatch has a bandpass filter of three to 11 Hz.

The Figure 2.2 demonstrates that the Actiwatch is the size and shape of a standard wristwatch. These accelerometers weigh approximately 20g. The measuring unit has slots at either end through which nylon (hypoallergenic) straps can be attached. Different straps were used for different patients

Figure 2.1. Diagram to explain how accelerometers work

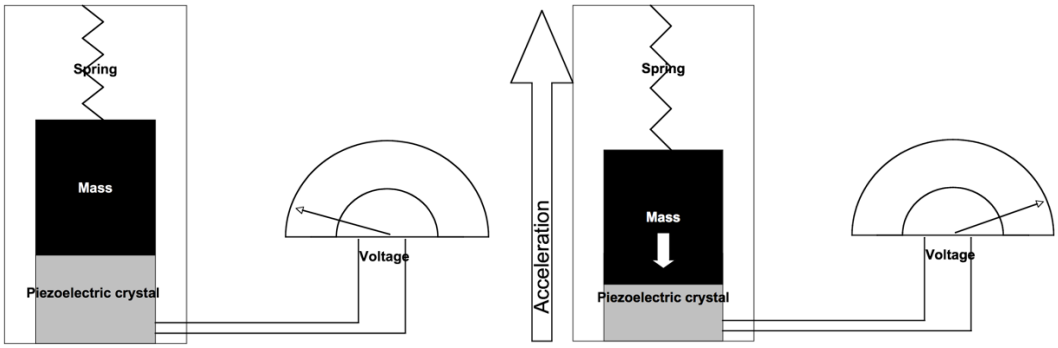


Figure 2.2. Photograph of an Actiwatch Plus accelerometer on a man's hand

Table 2.1. Measuring epochs available to program on Actiwatch Plus

Epoch length	Recording time
2 seconds	36h, 12 mins
5 seconds	3 days, 18h, 30mins
10 seconds	7 days, 18h, 24mins
15 seconds	11 days, 8h, 20mins
30 seconds	22 days, 16h, 40mins

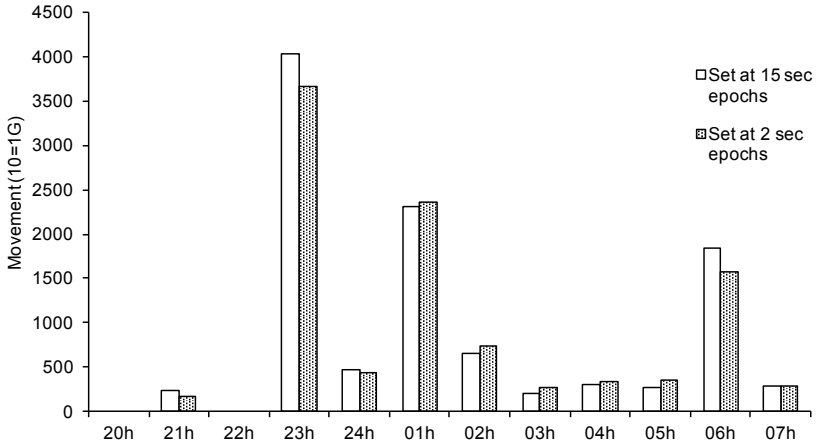
The Actiwatch Plus yields data in 'epochs.' A summary value for each epoch is recorded. The length of epoch can be set before a recording period. The shorter the epoch the shorter the recording time (as the memory is filled sooner). An epoch time of two seconds will allow recording for 36 hours, 12 minutes. No data is 'lost' with longer recording epochs as the output is a *summary* value, all that is lost is second-by-second discrimination of action. Table 2.1 illustrates the relationship of recording epoch to recording length.

Data is downloaded from the units by placing the measuring unit on a proprietary reader, also bought from Cambridge Neurotechnology. The reader can be connected to a PC. Cambridge Neurotechnology provided software for the analysis of movement and sleep. There are two problems with this software. The first problem is one specific to this study: the Cambridge Neurotechnology software yields summarized data, it sums the acceleration per epoch for chunks of time (for example, 30 minutes or one hour) and does not allow minute by minute scrutiny of acceleration. We really did need access to real time, second by second data in order to compare the accelerometer to video-recordings for validation purposes. The second problem has huge implications for all users of the Actiwatch Plus: when the raw data output acceleration is compared to the software output acceleration there is a difference. When I examined where the difference occurred and why, I found it to be to due to an error in the Cambridge Neurotechnology software coding: epochs were skipped by the software. The fault acts like a frame-shift mutation (to use DNA transcription as an analogy) so that summarized data yielded by the Cambridge Neurotechnology software after this point is out of synch with similarly analysed data passed through my own Macro. Due to these issues, raw data was extracted from .AWF and .AWD files and exported into Excel for the purpose of this study's analyses. The files of raw data are large: 20,000 Excel cells for a two second epoch ten-hour recording.

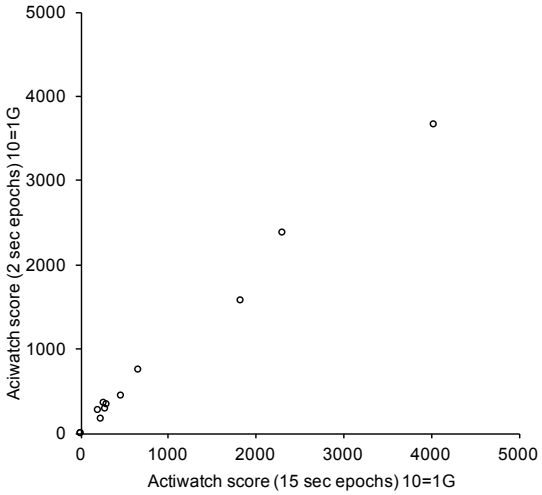
In order to prove that no data was lost by extending the epoch lengths, I carried out experiments where two watches were worn on the same wrist overnight. This experiment may sound superfluous based on my explanation of the physics of these devices, but the research team which had carried out pilot studies (Benjamin et al., 2004; Bringhurst et al., 2004) had been advised by a technical expert that this would be the case. Once I understood how the accelerometers worked I suspected that this was not the case and that data would not be 'lost' if longer recording epochs were used (the data would be summed, no acceleration data skipped or lost). The technical expert had also written a 'macro' which summarized the two second data into hourly chunks for the recording times of 10pm to 8am on the first night of recording only. This meant that even if the accelerometer had been worn and recorded for three nights, only one night's data was available.

To prove that no data was lost by increasing the recording epochs, a control subject wore two accelerometers on one wrist. One accelerometer was set to record two second epochs and the other 15 second epochs. I wrote my own macro which extracted hourly scores for two, 10, 15 and 30 second epochs. In my macro I summarized every recording hour from when the accelerometer was set to start at 18:00hr. The Actiwatch Plus which was set to capture 2 second epochs yielded 1800 Excel cells per hour and the Actiwatch Plus set to capture 15 second epochs which yielded 240 Excel cells per hour, the percentage difference is 13.3% and so you might expect this level of difference is data was being 'lost' in the capture process. The result is demonstrated in Figure 2.3. It can be seen that there is minimal difference in recording output when the 2 second epoch-recording Actiwatch Plus data is compared to the 15 second epoch-recording. The percentage difference between the devices is 4.1% which is most likely accounted for by variance between the devices and not by data being 'lost' in the detection process.

Figure 2.3. Graphs to summarise results from experiment whereby one subject (n=1) wore two Actiwatch Plus on right wrist overnight. Hourly Actiwatch Plus scores are compared. a) Bar chart to demonstrate hourly capture of movement: left bar (white), Actiwatch Plus set to capture 15 second epochs and right, (shaded bar), Actiwatch Plus set to capture 2 second epochs.



b) Scatter plot comparing Actiwatch Plus set to capture 15 second epochs versus 2 second epochs



The pilot study (Bringhurst et al., 2004) had also involved a few consecutive-night studies and these had demonstrated huge night-to-night variation in scores. One of my study aims was to increase the number of consecutive night studies to see if these variations were real and if there were any pattern or way of predicting the size of scores. I felt that compliance and participation would be increased and be more appealing if I could give the participants their accelerometers at the beginning of the study period (initially three nights but eventually six weeks) and not have to try to meet them each day to download data and reset the accelerometer. Therefore, as well as writing macros to compress various epoch lengths, I also wrote them to be able to extract data from many consecutive nights on one recording.

II: Validation of consecutive night Actiwatch Plus against infrared video-recordings

Aim

The Actiwatch Plus had previously been validated against the gold-standard of infra-red video recording(Benjamin et al., 2004). Significant night-to-night variation in score was detected in preliminary consecutive night studies (Bringhurst et al., 2004). This study sought to validate the Actiwatch Plus against infrared video recording in night-to-night recordings to see if the variation was a genuine phenomenon.

Background

The download from an accelerometer, or 'score,' can represent genuine scratch activity but is non-specific: it can equally represent other movements such as walking, playing etcetera. To try to skirt this problem, my group had always sought to quantify 'nocturnal activity.' Previously, these 'nocturnal hours' were determined by the limitations of a macro which only allowed data from 22:00h to 08:00h to be extracted and also by parent-recorded bed/sleep time of the child. Unfortunately, software errors were detected in the company supplied software and so it was necessary to deal with the raw data-downloads, see Chapter 2.I. My involvement in this research started by my writing a new macro which was not confined to 22:00-08:00hrs.

Previous subjects' consecutive night data was reanalysed primarily to address the question as to why there was such marked night-to-night variation in accelerometer score. It was observed that the hourly scores at the beginning of the recording night were markedly higher than those later in the recording. It was suspected that the early evening peak in score was most likely related to the subject not actually being asleep and therefore parent-recorded sleep-times not marrying with reality and not to be relied upon. In order to ascertain whether

the early evening surge in score was an itch-related activity, however, and thus an observation which could be exploited in capturing itch, further video studies were undertaken and are described here.

The study also afforded directly observed overnight activity to compare to Actiwatch Plus and further validate them in night-to-night studies (checking that the data was not corrupted in recordings that lasted several nights.

Method

Subjects

Six children, all of whom had a diagnosis of atopic dermatitis, were studied. The children were recruited from a secondary care Paediatric Dermatology clinic and had been were diagnosed as having atopic dermatitis by a Consultant in the clinic (as described by Hanifin and Rajka (Hanifin & Rajka, 1980)). The subjects had moderate to severe atopic dermatitis (assessed by SCORAD) as would be expected from their being recruited from a secondary care Dermatology clinic. The range of SCORAD was 16.2 to 33.6 and the median 25. The group's median age was 7 years (age range 3-10yrs, four male and two female).

Study method

Each study subject's nocturnal movements were observed using infrared video recordings. This involved my visiting the subject's house and setting up the infrared videoing system (see Chapter 1.II).

The subject's parent was issued with written and verbal instructions. They were asked to press the 'Record' button on the VCR when the child went to bed. The accelerometer was attached to the dominant hand’s wrist two hours before retiring. The videotape was

recovered from the subject's house the next day. The Actiwatch Plus data was downloaded and the infrared equipment set up to record the next night. The aim was to record data for three consecutive nights. The subjects were issued a questionnaire which asked for information such as the time of retiring to bed and VAS for symptoms.

Early evening analysis

The full video observations and parental noted bed-times were used to characterize movements throughout the recording period: specifically, early evening activities were scrutinized in order to explain the early evening peak in Actiwatch Plus recordings. The hourly total acceleration was extracted from the Actiwatch Plus data. The hours of analysis were dictated by the parent's recorded 'bedtime.' To illustrate, if a parent recorded bedtime as 20:00h, then the score from 20:00h to 21:00h was considered that for 'Hour 1' of sleep, and 21:00h to 22:00h 'Hour 2' etc. The videos were watched and the time of getting into bed was recorded, as was the time of 'presumed sleep' (defined as when the child was consistently lying still). The parent-recorded bedtime was compared to the actual sleep-time as reflected by the video recording. The waking times were noted and compared too, but this period of time was not scrutinized as no peak in score had been noted in this period.

Comparison of directly-observed movement

The infrared video recordings were played back in 'real-time' and every movement noted down. This viewing process took over one month to complete. For the purposes of analysis, although rhythmical and non-rhythmical movement was denoted, since the Actiwatch Plus is purely a movement monitor without any suggested or implicit sensitivity for a specific frequency of action, *total movement* recorded or timed per hour was compared to hourly Actiwatch Plus score. In order to investigate the night-to-night variation in Actiwatch Plus score, hour-by-hour patterns of scoring were plotted (were there specific or person-specific patterns?)

Results

One subject (subject 3) only managed to complete video-recording for one night and so this subject was not able to be included in the night-to-night analysis although the one night's video-recording was useful for the characterizing activity part of the experiment. The same subject's Actiwatch Plus was only worn on one night and not the same night as the video-recording. There were 14 night's recordings to examine overall. Not all hours of Actiwatch Plus recording yielded a similar video-tape hour's recording if the child were mobile during the night (and not on camera).

Early evening analysis and hour-by-hour consecutive night Actiwatch Plus scores

All of the following subjects' individual night-to-night hourly activity is summarized in Figure 2.4 and the sleep/waking times as recorded by the parent and observed on video displayed in Table 2.

Subject 1

This subject competed the three-night study. The early recording activity on Actiwatch Plus was shown to be due to the subject being awake. There was a mismatch in parent reported sleep time and the video-observed reality. There seems to be a similar scoring pattern each night longitudinally (note peaks at Hour 5 and Hour 8, see Figure 2.4).

The early peak in Actiwatch Plus score (hours one to three) is characterized on video as follows:-

Night 1: being in bed at 22:00h, watching TV, generally restless and scratching often.
Night 2: in bed at 19:38h, watching T.V. and frantically scratching until 19:50h when plays with dolls house. When playing, no scratching. At 21:15h puts dolls house away, watches TV and absent-mindedly scratches. 21:30h settles to sleep.

Night 3: in bed at 21:07h playing with dolls house, watching TV, no scratching. Puts dolls house away at 22:20h, watches TV and occasionally scratches. At 23:00h turns off TV and settle to sleep.

Subject 2

This subject also competed the three-night study. The early recording activity on Actiwatch Plus was shown to be due to the subject being awake. Although he was not asleep when the parent thought or recorded he was, he was in bed. This subject is very restless in bed and takes a long time to settle down to sleep. He scratches a lot but also generally moves a lot. There is hour-by-hour pattern for Actiwatch Plus score when nights are compared.

Subject 3

This subject only undertook the study for one night (as previously mentioned). The parent-noted sleep and wake times are only ten minutes apart. In the extra time before actually goes to sleep, the subject is in bed but is generally restless and scratches a lot.

Subject 4

This subject wore the Actiwatch Plus on each night, but due to some kind of malfunction, no movement was detected by the accelerometer until after the eighth hour of recording. The successfully completed two nights again illustrate the early peak in score but otherwise no pattern on hour-by-hour recording. This subject is in bed pretty much as the parent records and settles quickly without too much restlessness.

Subject 5

This subject successfully competed the three nights of the study. The early peak in score is noted again, but no hour-by-hour pattern when nights are compared. During the delay between recorded and presumed sleep-time, the subject is in bed and generally restless but

also scratches a lot. When the recorded time of getting up differs on night 3, the subject gets out of own bed at 00:05h and goes to her mother's bed. The parent recorded awake time is actually when the subject pulled the accelerometer off.

Subject 6

This subject completed two study nights. A small early recording hour peak is noted on night two. It is noted that the broad pattern of hour-to hour recording is similar one night to the next: peaks at Hour 4 and Hour 7. For both recording nights, during the mismatched time, the subject is in bed fidgeting. It is notable that there is a bigger mismatch in timings on night 2 and that this is when the peak in Actiwatch Plus score is larger too.

Figure 2.4. Each subject's nightly hour-by-hour Actiwatch Plus score

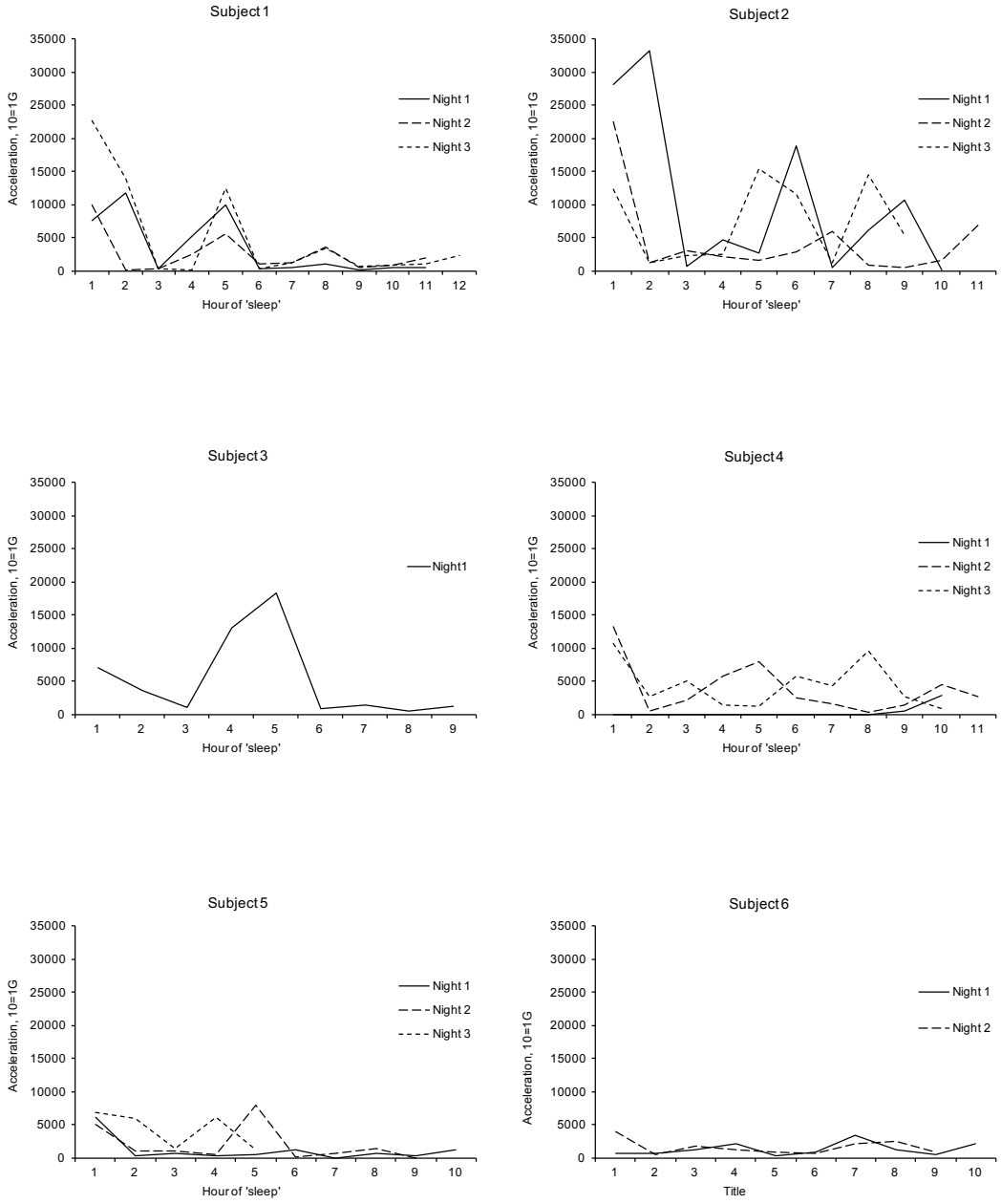


Table 2.2. Sleep/waking times as recorded by parent and as observed on video-recording

Subject	Study night	Parent says asleep	Observed sleep	Difference	Parent says awake	Observed awake	Difference
1	1	21:00h	22:30h	+90 mins	08:00h	07:30h	-30 mins
	2	21:00h	21:30h	+30 mins	08:00h	08:55h	+55 mins
	3	21:00h	23:00h	+120 mins	08:00h	08:55h	+55 mins
2	1	22:00h	22:05h	+5 mins	08:00h	06:50h	-70 mins
	2	21:00h	21:30h	+30 mins	08:00h	07:20h	-40 mins
	3	21:00h	21:30h	+30 mins	08:00h	05:00h	-180 mins
3	1	19:30h	19:40h	+10 mins	07:30h	08:00h+	+10 mins
4	1	22:00h	22:05h	+5 mins	08:00h	06:20h	-100 mins
	2	21:00h	21:20h	+20 mins	08:00h	07:20h	-40 mins
	3	21:00h	21:15h	+15 mins	08:00h	06:10h	-110 mins
5	1	20:10h	20:50h	+40 ins	07:05h	07:00h	-5 mins
	2	21:35h	21:40h	+5 mins	06:05h	06:05h	0 mins
	3	21:00h	22:20h	+80 mins	02:10h	00:05h	-115 mins
6	1	00:00h	00:20h	+20 mins	08:30h	09:50h	+80 mins
	2	21:00h	22:35h	+95 mins	08:00h	08:20h	+20 mins

Subjects' combined data

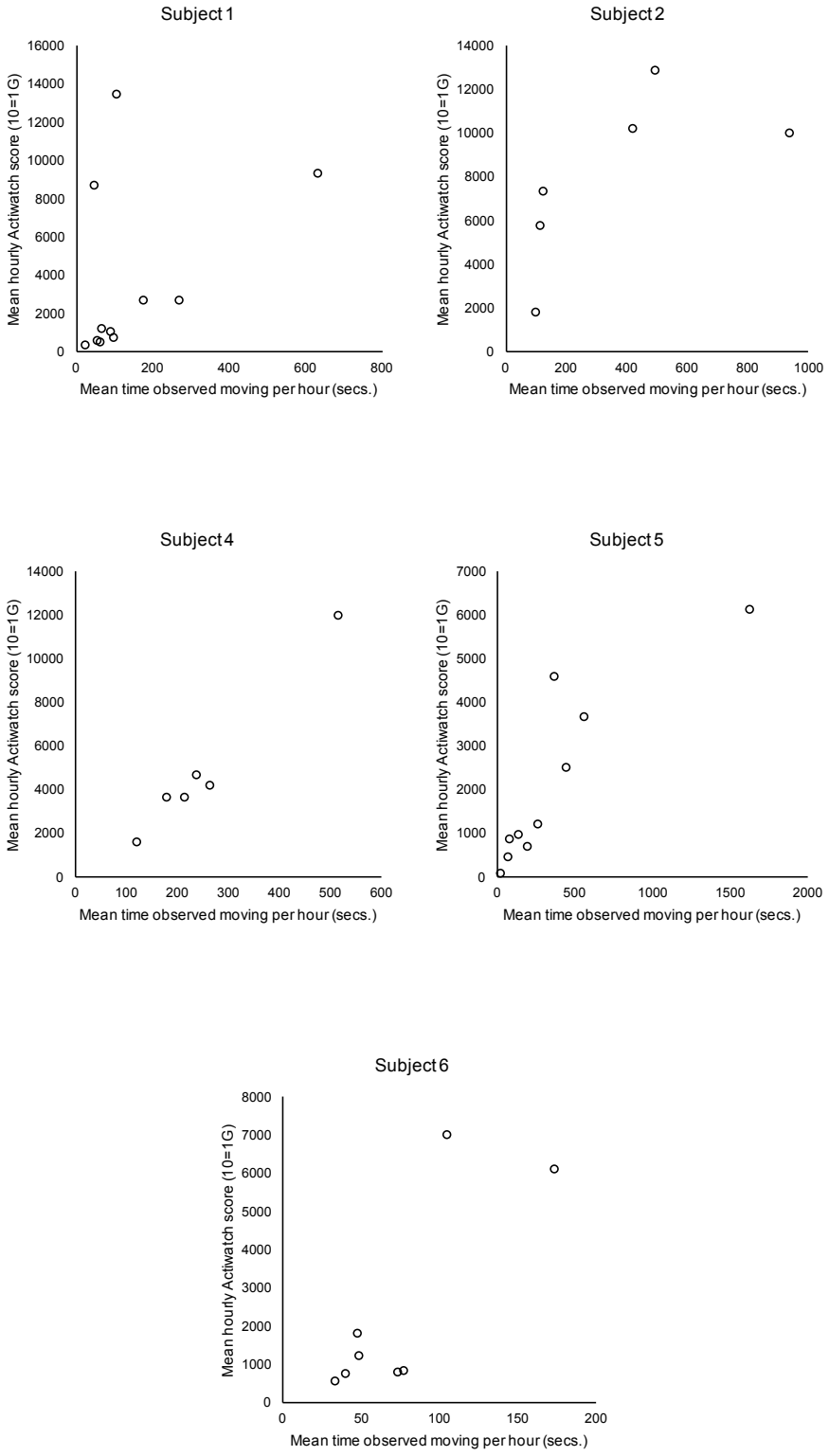
The mean difference in observed versus recorded sleep times is +39.7minutes (range +5 to +120 mins). This means that, on average, the parent recorded the child's sleep time approximately 40 minutes earlier than the actual time. When the difference between recorded sleep times and observed sleep times is analysed by paired t test, the difference is significant, p=0.005.

The mean difference in observed versus recorded wakening times was -30minutes (range -180 to +80 mins). This means that, on average the parent recorded the child's waking up time 30 minutes later than actually occurred. The paired t test analysis of the difference between recorded and observed waking times gives a p value of 0.047.

Validation of consecutive-night Actiwatch Plus scores against direct-observation of movement on video.

Each subject, except subject 3 (who had worn the accelerometer one night and completed the infrared videoing another night), had data available for analysis. The mean hourly Actiwatch Plus score per individual across nights was compared to the mean amount of time spent moving during the same hour, as observed from the video recording. The *mean* hourly data was analysed to allow for nesting of data: one subject having the same measurements each night, if all nights were correlated together would give a 'dishonest' picture. As expected, a close relationship is demonstrated between the gold-standard video recording and the Actiwatch Plus scores per hour. Each individual's Spearman rho demonstrates significance: Subject 1 $\rho=0.618$ $p=0.043$, Subject 2 $\rho=0.829$ $p=0.042$, Subject 4 $\rho=0.889$ $p=0.019$, Subject 5 $\rho=0.927$ $p<0.001$, Subject 6 $\rho=0.762$ $p=0.028$. Figure 2.5 summarises the results.

Figure 2.5. Graphs comparing mean hourly amount of movement for the study period (seconds) to mean hourly Actiwatch Plus score (10=1G,) all nights together.



Conclusion

This study sought to check that the night-to-night variation in Actiwatch Plus score was a 'real' phenomenon. It answered this question by affording the means of physically observing the subjects on consecutive nights on infrared video whilst the subject also wore the accelerometer. The study allowed any patterning of scoring to be observed too. The study allowed for further validation of the use of accelerometers to quantify nocturnal movement and confirms that there is no corruption in data, for instance or any other accelerometer recording error to account for the variation. The night-to-night variation in Actiwatch Plus score is reflected in directly-observed video movements and no generalized scoring patterns are reliably demonstrated (possibly minor peaks in score at about Hour 5 and Hour 8 for some individuals).

The early recording period Actiwatch Plus score was proven, in this experiment, to be due to two phenomena: inaccuracy in parent-recording sleep times and wakeful activity occurring in this period. Inaccuracy between actigraphic sleep-time and diarised wake/sleep times has been reported (Baker, Maloney, & Driver, 1999) (van den Berg et al., 2008) but in the published report by this group, the possibility of inaccuracy was raised as the actigraphy software was relied upon for sleep/wake times. At least in this study there can be no question of reliability: the times are available for direct observation on video.

The results demonstrate an expected close relationship between the amount of movement observed on camera and the Actiwatch Plus score. This is further validation for the use of the devices. It should be noted that time spent in *any* movement was compared to the accelerometer score. This was a deliberate choice as the accelerometer used in this experiment is a simple device which does not offer or attempt to specifically sense itch-related movement (as opposed to the DigiTrac).

Overall, therefore, it would appear that the night-to night variation in score is a real phenomenon. A larger multiple night, within subject Actiwatch Plus study was proposed on the basis of these results in order to characterize the night-to-night variation in score and attempt to explain it.

Chapter 3

Introduction and validation of a newer accelerometer, DigiTrac.

I: Introduction to DigiTrac accelerometers

During the course of the study, new accelerometers became available. The newer digital accelerometer sensed much more detailed information (detected three axes of movement) and furthermore which could transform data to detect acceleration at particular frequencies of action. It was hypothesized that, if we could determine the frequency of action of itch-related movement (scratch and rub), then these instruments should be able to specifically determine and measure this rhythmical movement. The recordings would, we felt be more accurate or discriminating than the Actiwatch Plus accelerometer. These newer instruments are known as DigiTrac accelerometers "DigiTrac - movement recorder," 2007 http://www.imsystems.net/DigiTrac/DigiTrac.htm IM Systems, Baltimore)

DigiTrac are wrist or leg-worn instruments that are capable of measuring acceleration movement in three axes (as opposed to the Actiwatch which measures in one axis, two planes). They contain three piezoceramic accelerometers (one for each measurement axis) and a gyroscope (to indicate the device's position).

The DigiTrac is larger than the Actiwatch and measures five centimeters in the long axis. When sensing in three axes, its recording time is ten hours. If it is sensing in just one axis the recording time is 30 hours. Figure 3.1 contains a photograph of the device on a woman's wrist (to allow size comparison).

The DigiTrac samples and records more frequently than the Actiwatch (the sampling rate is between ten and 40Hz). DigiTrac are sensitive to a range of frequencies from 0.8-14Hz by use of a band pass filter (as opposed to three to 11Hz in the Actiwatch).

The high frequency of sampling means that 'frequency of action' can be calculated. This means that the device is a 'spectrum analyzer.' A spectrum analyzer is a device used to

examine the spectral composite of electrical, acoustic or optical waveforms. A signal (here a movement) has a waveform that can be broken down into individual sine waves, and these are known as the 'spectral composite.'

Spectrum analysers can be:

a) Analogue analysers: use a variable bandpass filter whose mid frequency is tuned to the range of frequencies to be measured.

b) Digitral spectrum analysers: uses Fast Fourier Transform (FFT) to transform a waveform into the components of a frequency spectrum.

The DigiTrac is both an analogue and a digital spectrum analyzer. The output yields a detailed account of the acceleration at a range of frequencies of action.

The Fourier Transform decomposes a function of time or signal (here movement) into its component frequencies: any signal or waveform can be constructed by adding together a series of pure tones (sine waves) with appropriate amplitude and phase. The Fourier Transform is an equation to calculate the frequency, amplitude and phase of each sine wave within a signal.

The Discrete Fourier Transform (DFT) is a numerical equivalent of the Fourier Transform. The Fast Fourier Transform is just a computationally fast way to calculate DFT. A Fast Fourier Transform (FFT) is an algorithm to compute the discrete Fourier transform (DFT) and its inverse. FFT rapidly computes transformations by factorizing the DFT matrix into a product of sparse (mostly zero) factors. As a result, fast Fourier transforms are widely used for many applications in engineering, science, and mathematics.

The output from the digital accelerometer takes the form of 'bits' or 'points.' These are multiples as follows: 2, 4, 8, 16 etc. For fast Fourier Transform to be most accurate, the

epoch length should be as long as possible. An output of 256Pts equates to an epoch length of 6.4 seconds (it is possible to export 128Pts but this equates as a less preferable and shorter epoch length of 3.2 seconds). The accelerometer samples 40 times per second or once every 0.025seconds, an output of 256 (0.025x256) yields a 6.4 second epoch. The output also determines the discrimination wavelength or Hertz-wise: 256Pts equates to increments of 0.16 alternating with 0.15Hz (128Pts would equate to 0.32Hz increments).

The FFT read-out from the DigiTrac is a table. A sample is provided in Figure 3.2. The columns, running horizontally represent different frequencies, from 0.16-10Hz and the rows, running vertically represent consecutive epochs of the recording period. The FFT produces the frequency and amplitude composition of the DigiTrac signal, therefore, within the table is a value, in G-seconds (G-s), of the amplitude or acceleration, happening that specific category of frequency within a specific epoch of time. As can be seen in Figure 3.2, a frequency 'pure-tone' labels each column and then, vertically, epochs are labelled. Inside each box of the table is the acceleration for that frequency in that epoch of time.

A separate file is available and downloadable for each recording for each of the axes (x, y and z) as long as the device is set to record all three axes (as opposed to one).

I.M.Systems technical staff advise that a high pass filter in the unit removes the gravitational component so the orientation of the unit is irrelevant. This would negate the problem of the sensing axis for a particular movement being different with the subject being in a different position, for example, sitting up versus lying down.

Figure 3.1. Photograph of a DigiTrac accelerometer on a woman's wrist.

Figure 3.2. Sample of Excel spreadsheet containing DigiTrac FFT output table.

Frequency	0.16	0.31	0.47	0.63	0.78	0.94	1.09	1.25	1.41	1.56	1.72	1.88	2.03	2.19	2.34	2.5	2.66	2.81	2.97
22:00:00 PM	116	110	104	96	88	78	74	68	62	56	52	50	46	44	42	38	38	36	34
22:00:06 PM	0	0	2	0	0	0	0	2	2	0	0	0	0	0	2	0	0	0	0
22:00:13 PM	0	0	2	0	0	2	0	0	0	0	0	0	0	0	2	0	0	0	0
22:00:19 PM	0	0	0	0	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0
22:00:26 PM	0	0	2	0	2	2	0	2	2	0	0	0	0	0	0	2	0	0	0
22:00:32 PM	0	0	0	0	0	2	0	0	0	2	0	0	0	0	0	0	0	0	0
22:00:38 PM	0	0	2	0	2	2	0	2	0	2	0	2	0	0	0	0	0	2	0
22:00:45 PM	0	2	0	2	2	0	0	0	0	0	2	0	0	0	0	2	0	0	0
22:00:51 PM	0	0	0	0	2	0	2	2	0	2	0	0	2	0	0	0	2	2	0
22:00:58 PM	0	2	2	2	0	2	2	0	0	2	0	2	0	0	0	0	0	0	2
22:01:04 PM	0	0	0	0	2	0	0	0	0	0	2	2	0	0	0	0	0	0	0
22:01:10 PM	0	2	0	0	0	0	0	2	0	0	0	2	0	0	0	0	0	0	0
22:01:17 PM	0	0	0	0	2	0	2	0	0	2	0	0	0	0	0	0	0	0	0
22:01:23 PM	0	0	0	0	0	0	2	2	0	0	0	2	0	2	0	0	0	0	0
22:01:30 PM	0	0	2	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0

II: Experiment to validate inter-machine DigiTrac accuracy

It was necessary to assess the precision of the DigiTrac in the first instance. These experiments were not necessary for me to undertake on the Actiwatch, as they had been previously validated (Benjamin et al., 2004; Bringhurst et al., 2004).

Aim

To test how similar the DigiTrac read-outs are when moving at exactly the same speed, in the same direction: to test the 'precision' of DigiTrac.

Background

- Accuracy is the degree of conformity of a measured/calculated to quantity to its actual (true) value: how far the measured value is from the known reference level. Accuracy is a measure of veracity.

- Precision (also called reproducibility or repeatability) is the degree to which several measurements will show the same results.

- The relationship between accuracy and precision is illustrated in Figure 3.3 (over).

- Precision is characterised in terms of the standard deviation of the measurements. The coefficient of variation (CV) is a measure of dispersion of a probability distribution and is commonly applied in reliability theory.

- The coefficient of variation (CV) is defined as the ‘ratio of the standard deviation to the mean’ and is reported as a percentage. A CV of 20% or less is generally considered acceptable/realistic for mechanical equipment.

- CV=SD/Mean

Figure 3.3. Graph to illustrate relationship between accuracy and precision.

Coefficient of variation is ratio between standard deviation and mean.

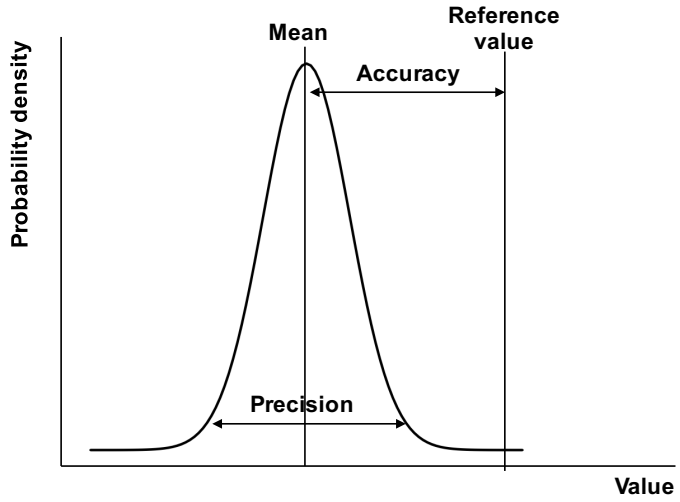
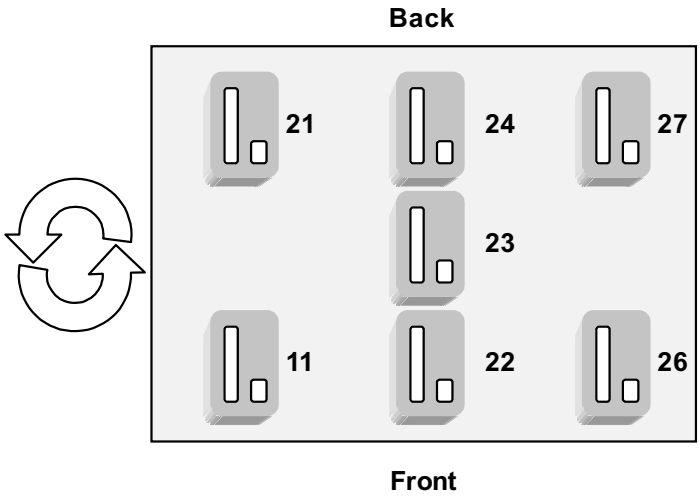


Figure 3.4. Diagram to illustrate placement of DigiTracs on the laboratory shaker



Method

All the DigiTrac (n=7) were placed face-down on a platform-style laboratory shaker (see Figure 3.4: the numbers indicate the serial number of the DigiTrac).

The DigiTrac were set to record three minutes of activity at each of eight increasing arbitrary shaker speed settings. This was repeated three times at each speed setting.

The average acceleration at each speed was then compared between devices and CV calculated.

Results

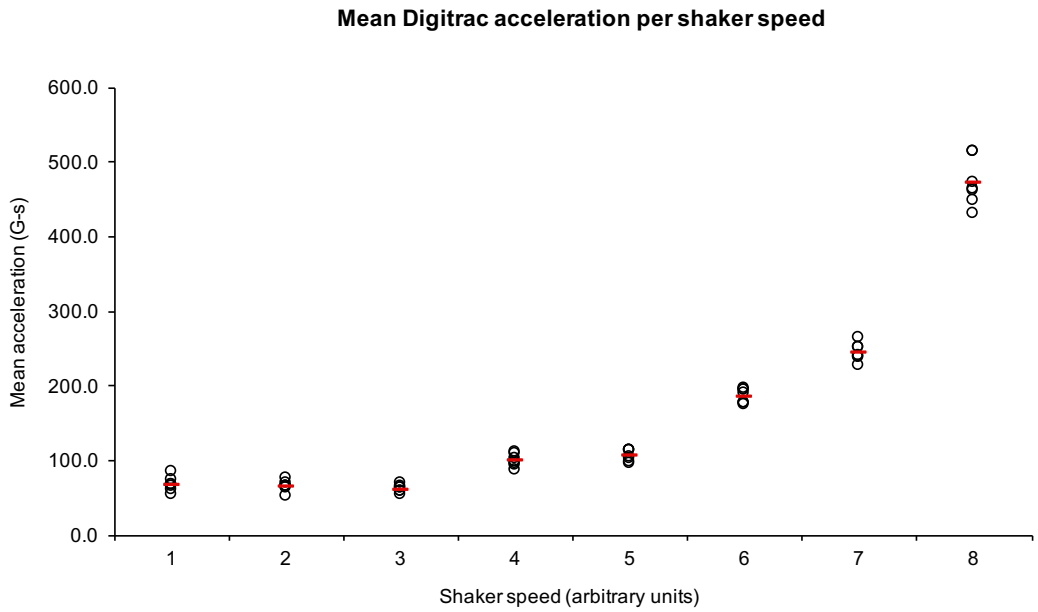
Table 3.1, lists the recorded mean acceleration, standard deviation and coefficient of variation for the seven devices (combined) at each shaker speed, from 1 to 8. It can be seen that the CV ranged from 4.9% for shaker speed setting number seven rising steadily to 14.3% for the lowest shaker speed setting, number one. This suggests better inter-device accuracy at higher frequency movement than at lower frequency movement. As previously mentioned, reassuringly, in the fields of electrical component testing, a CV of less than 20% is considered acceptable and so the devices detect acceptably at all frequencies of action.

A graph to illustrate the raw output for each device at each shaker frequency can be found in Figure 3.5.

Table 3.1. Table listing the recorded mean acceleration, standard deviation and coefficient of variation for the seven devices (combined) at each shaker speed, from 1 to 8.

	Mean accel'n per speed (G-s)	Standard deviation	Coefficient of variation (%)
1	67.2	9.6	14.3
2	64.5	7.2	11.2
3	61.2	5.5	9.0
4	99.0	8.8	8.9
5	105.6	7.6	7.2
6	186.2	9.6	5.2
7	243.9	12.1	4.9
8	471.9	31.9	6.8

Figure 3.5. Graph to demonstrate each device's mean output for each shaker frequency. The mean for each shaker frequency is illustrated as a red line.



Conclusion

The DigiTrac are acceptably precise (CV less than 20%). There is less inter-device accuracy at a lower frequency of action compared to a higher one but the experiment reassured us that they were fit for our purpose of assessing human itch-related scratch movement.

III: Experiment to assess sensitivity of DigiTrac accelerometer to itch-related movement.

Aim

To test the sensitivity of the DigiTrac to itch-related movement *in vivo*.

Method

One subject was studied. The subject wore the DigiTrac on her right wrist. The subject had a good understanding and ability to simulate stereotypical movement as she has previously been involved in scratch-measuring videotape experiments. The subject was simultaneously time-index videotaped. The experiment was in two parts:

1) First, the subject was required to simulate scratch by scratching her left forearm with her right hand for one minute. This experiment was designed to characterise the FFT of scratch movement and to see where the maximal activity was frequency-wise: was there favourable comparison with the video-observed mean frequency of scratch (1.85Hz).

2) Second, the subject was required to undertake a series of prescribed, stereotypical itch-related movements, scratch and rub only (and no non-rhythmical movements such as adjusting position) with both her right hand and her left hand. It was known that the Actiwatch Plus were sensitive enough to detect itch-related movement carried out on the non-Actiwatch PLus bearing arm (Benjamin et al., 2004; Bringhurst et al., 2004), but it was not known if the DigiTrac were similarly sensitive. The actions were carried out for ten second episodes starting at the beginning of the minute.

The prescribed movements were scripted as follows:

1) Scratching head with fingers left hand
2) Scratching head with fingers right hand
3) Scratching head with left arm
4) Scratching head with right arm
5) Scratching left elbow with right hand fingers
6) Scratching right elbow with left hand fingers
7) Scratching left elbow with right arm
8) Scratching right elbow with left arm
9) Scratching right loin with right fingers
10) Scratching left loin with left fingers
11) Scratching right loin with right arm
12) Scratching left loin with left arm
13) Rubbing left foot on right foot
14) Rubbing right foot on left foot
15) Rubbing eyes with left hand
16) Rubbing eyes with right hand
17) Rubbing face into pillow

The DigiTrac-derived FFT scores for all axes were extracted and uploaded to a PC. The sum of the FFT for each episode of movement was examined and compared to the time-index videoing of each movement. Since our video-study (Chapter1.III) had suggested that 0.5-2.5Hz should be the most discriminating output to detect itch-related movement, this range of outputs was also more closely examined.

Results

1) The FFT output was examined with a sum of activity at the pure tone frequencies and also with the mean at each frequency. As can be seen from the two graphs in Figure 3.6, the peaks and troughs at the frequency pure-tones is reflected similarly whether sum (Figure 3.6a) or mean (Figure 3.6b) of FFT is analysed. It can also be seen that the maximal activity is at 1.72Hz and at 1.88Hz and thus closely reflects the video-observed mean frequency of scratch.

2) The graph in Figure 3.7 shows the acceleration output of the DigiTrac for the sum of FFT (total score) and for the limited spectrum of 0.5-2.5Hz (all in G-s). Spikes in activity can be seen for all movement, most markedly if the movement was carried out by the arm bearing the DigiTrac. It can also be observed that movement carried out by the non-DigiTrac-bearing arm causes a rise above baseline and a small spike in activity. The limited spectrum of 0.5-2.5Hz also reflected the Total FFT pattern (including sensing itch-related movement by the non-accelerometer bearing arm).

Figure 3.6 a) Graph of DigiTrac FFT output for all stereotyped **scratch** movement: **sum of acceleration** at each pure tone (=frequency or cycle per second or Hertz).

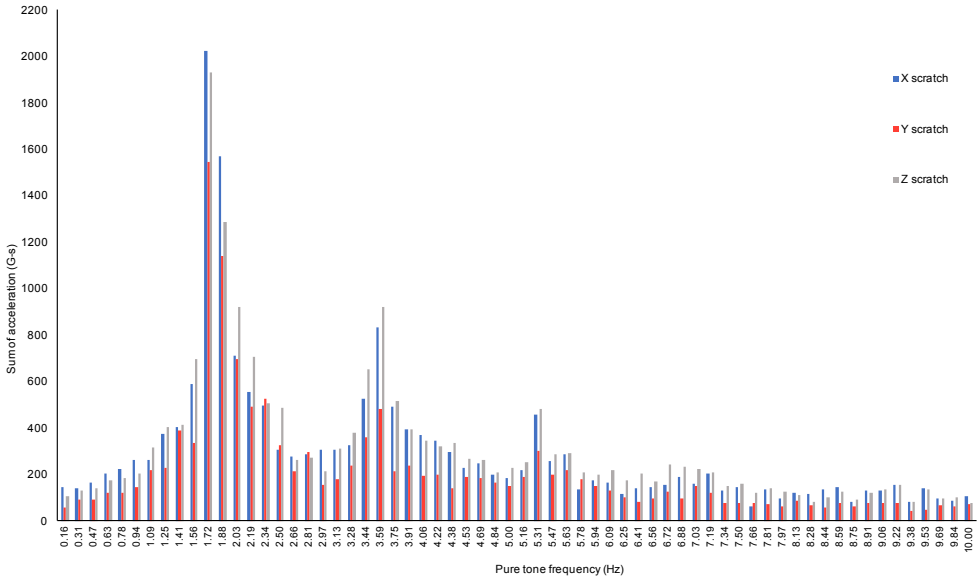


Figure 3.6 b) Graph of DigiTrac FFT output for all stereotyped **scratch** movement: **mean acceleration** at each pure tone (=frequency or cycle per second or Hertz).

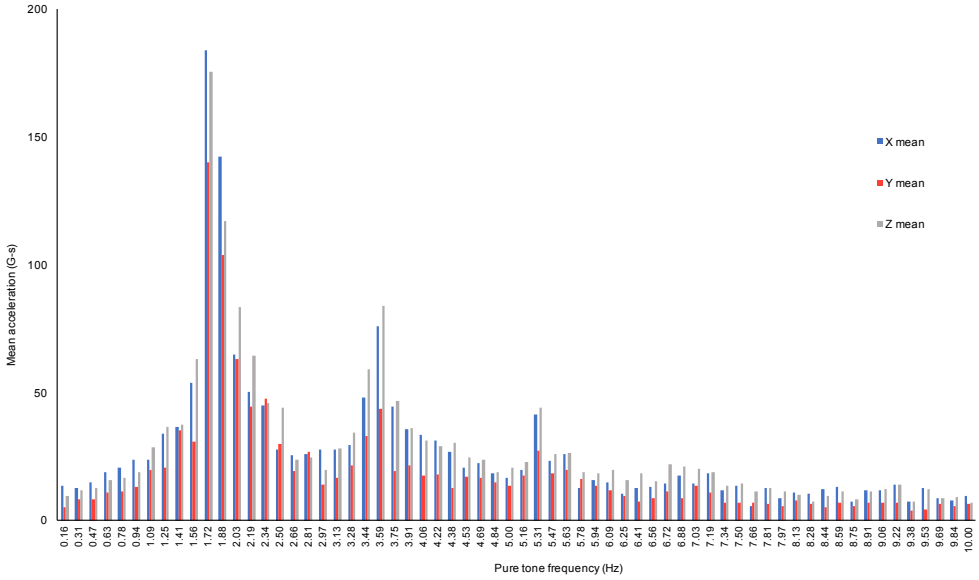
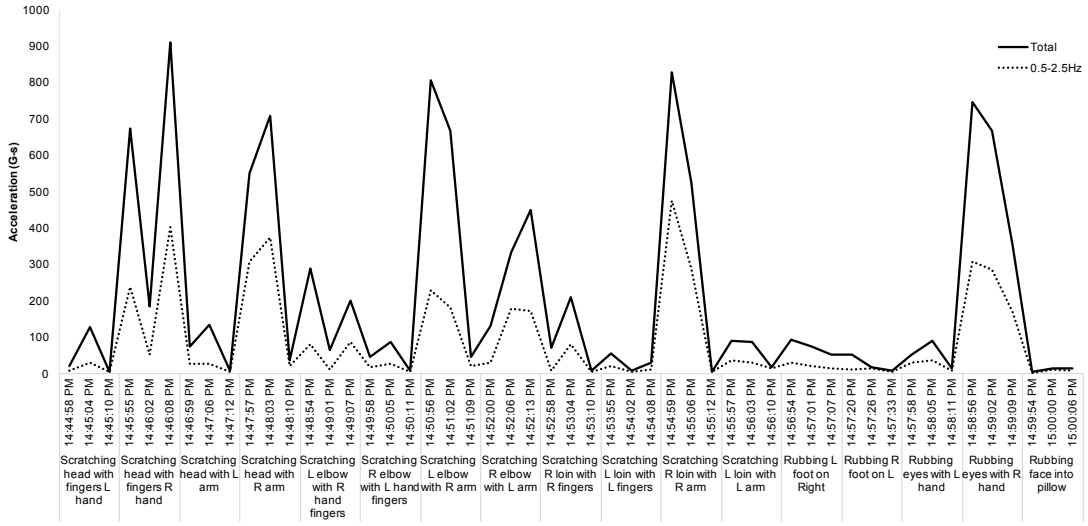


Figure 3.7. Graph of the acceleration output of DigiTrac for the sum of FFT (total score) and for the limited spectrum 0.5-2.5Hz (units all G-s) for stereotyped scratch movement. Solid line=total FFT and dashed line=0.5-2.5Hz.



Conclusion

The DigiTrac accelerometer has been demonstrated as capable of capturing itch-related movement in this experiment.

The fact that the maximal pure-tone frequency score was at exactly the mean video-observed score is encouraging for use of the DigiTrac as a movement-monitor able to discriminate and identify itch-related movement.

The sensitivity of the DigiTrac appeared to be at least as favourable as for the Actiwatch: it can detect scratch movement carried out by the non-accelerometer bearing arm.

The next step was to determine the best axis for capturing and discriminating itch-related movement.

IV: Experiment to assess which DigiTrac accelerometer axis output best to use for discriminating itch-related movement.

Aim

Decide which accelerometer axis output to use when measuring scratch movement. It was hypothesised, from observation of videos of scratch that the greatest sensitivity for scratch would be in the X axis as this would represent best detection of rhythmical movement at the wrist of the forearm through the elbow joint (see diagram of axes in Figure 3.8).

Method

One subject was studied. The subject wore the DigiTrac on her right wrist. She was required to undertake a series of prescribed, stereotypical itch-related movements (scratch and rub) and non-itch related movements (adjusting position) carried out for ten second episodes (starting at the beginning of the minute). The subject had a good understanding and ability to simulate stereotypical movement, as she has previously been involved in scratch-measuring videotape experiments. The subject was simultaneously time-index videotaped. The axes of movement in relation to the DigiTrac are illustrated in Figure 3.8.

The prescribed movements were scripted as previously but stereotyped 'generalised movements' were added in between the itch-related movements as follows:

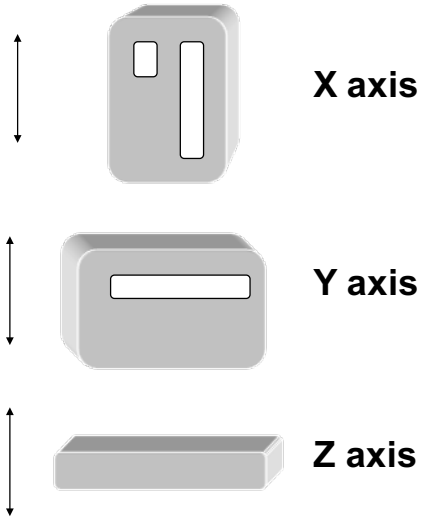
1. Stretching
2. Twitch
3. Lying on back, turn onto left side
4. Lying on back turn onto right side
5. Lying on left side turn onto right side
6. Lying on right side turn onto left side

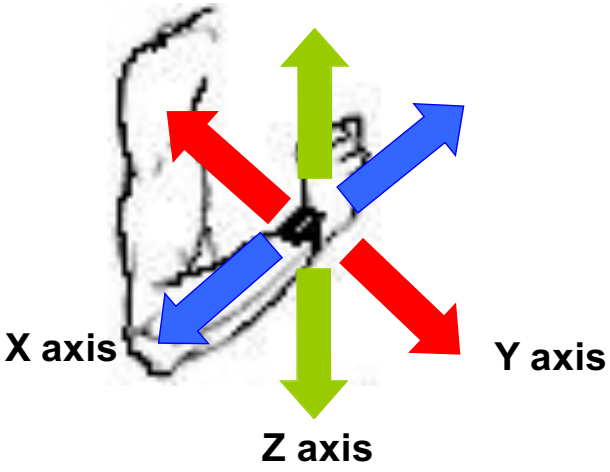
7. Lying on front turn onto back
8. Lying on front turn to right
9. Lying on front turn to left
10. Bring covers up.

The DigiTrac-derived FFT scores for each of the axes was extracted and uploaded to a PC. The sum of the FFT for each episode of movement in each axis was compared. The type of movement was identified and confirmed with the recorded timing of events and by observation of the time-indexed video.

For the purpose of analysis, anything itch-related (i.e. scratch and rub) was labelled ‘scratch’ (S), anything generalised was known as ‘adjusting position’ (A) and stillness was labelled ‘nil’ (N).

Figure 3.8. Diagrams to illustrate measuring axes of movement in relation to DigiTrac device.





Results

The different types of stereotyped movement as measured by each axis and the DigiTrac output is illustrated in Figure 3.9's graph. It demonstrates a consensus between the measuring axis. The graph also demonstrates that the maximal acceleration most often occurs during itch-related movement (S).

Table 3.2 demonstrates the comparison between X, Y and Z axes for the stereotyped movement. The overall mean sum of FFT for X axis was 261 G-s (SD 272.7), for Y was 396.5 G-s (SD 478.3) and for Z axis was 257.7 G-s (SD 258.6). Figure 3.10 summarised the data graphically. The amount of data overlap suggests no more sensitivity in any particular axis. An independent samples Kruskal-Wallis test demonstrates no difference in measuring axis: $p=0.476$.

Figure 3.9. DigiTrac output (sum of FFT acceleration, G-s) for different types of stereotyped movement: S=scratch, N=nil, A=adjusting position (n=1).

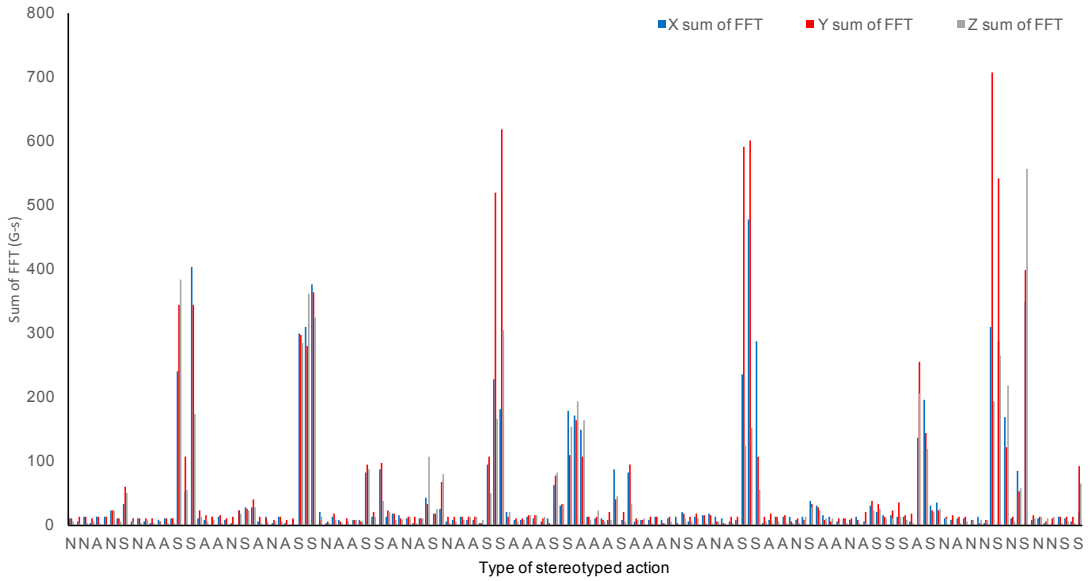
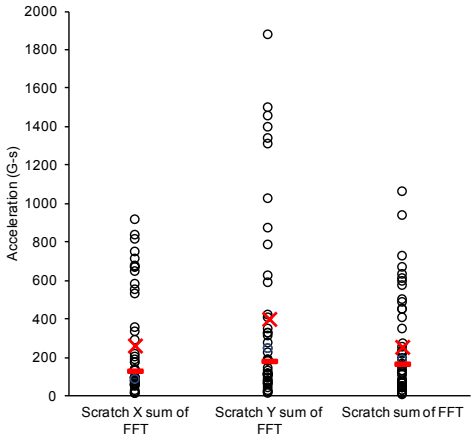


Table 3.2. Comparison between X, Y and Z axes for the stereotyped movement: sum of FFT and limited spectrum 0.5-2.5Hz listed.

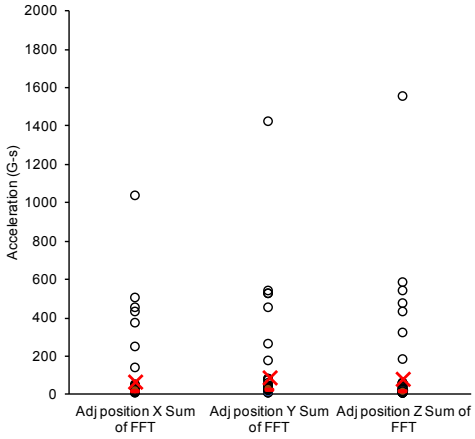
	X sum of FFT	Y sum of FFT	Z sum of FFT	X 0.5-2.5Hz	Y 0.5-2.5Hz	Z 0.5-2.5Hz
Scratch: sum	12266.0	18634.0	12110.0	5088.0	6920.0	4214.0
Scratch: mean	261.0	396.5	257.7	108.3	147.2	89.7
Scratch: median	130.0	184.0	164.0	36.0	40.0	38.0
Scratch: mode	88.0	106.0	38.0	30.0	12.0	6.0
Scratch: st dev	272.7	478.3	258.6	126.9	197.5	105.5
Adj posiition: sum	4008.0	5142.0	4868.0	1664.0	1884.0	1792.0
Adj position: mean	66.8	85.7	81.1	27.7	31.4	29.9
Adj position: median	15.0	23.0	12.0	10.0	14.0	8.0
Adj position: mode	10.0	24.0	6.0	8.0	10.0	4.0
Adj position: st dev	167.5	215.3	232.4	58.1	64.2	82.0
Nil: sum	426.0	532.0	284.0	310.0	360.0	212.0
Nil: mean	11.2	14.0	7.5	8.2	9.5	5.6
Nil: median	12.0	13.0	8.0	8.0	10.0	6.0
Nil: mode	12.0	12.0	8.0	10.0	10.0	6.0
Nil: stdev	5.2	7.0	4.8	3.9	3.4	3.7

Figure 3.10. Simulated stereotyped movements as measured by each sensing axis of the DigiTrac: a) scratch, b) adjusts position, c) nil or no movement. X =mean, - = median.

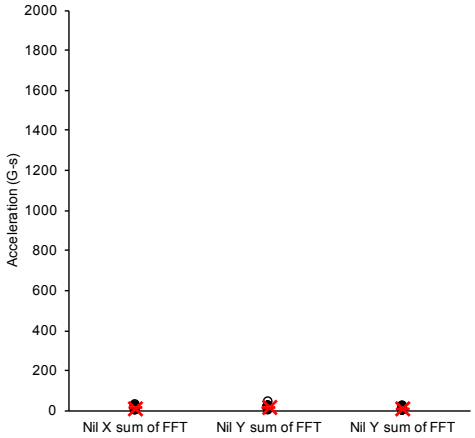
a.



b.



c.



Conclusion

The results suggest that any axis should be just as discriminating in order to pick up itch-related movement. For consistency's sake, the X axis was used in further experiments.

NB: Observed video studies demonstrated that frequencies of action for itch related movements were: scratch, 1.85Hz (SD±0.55) and rub, 0.98Hz (SD±0.36) and for non-itch generalised movement named 'Adjusting position' 0.48Hz (SD±0.25), see Chapter 1.III.

.

V: What is the frequency of action of human walking movement as measured by the DigiTrac?

Aim

To see if the DigiTrac read-out can distinguish between different movement: itch-related and other rhythmical non-itch related. If it were able to do so, it would be possible to study outside the usual night time restriction and also possible to recognise where there may be confounding: when a subject got up in the night to visit the toilet, for example.

Method

During a Paediatric clinic afternoon, ten subjects (patients and accompanying non-patients) were asked to wear a DigiTrac on each limb and walk around the Out-patient department for one minute.

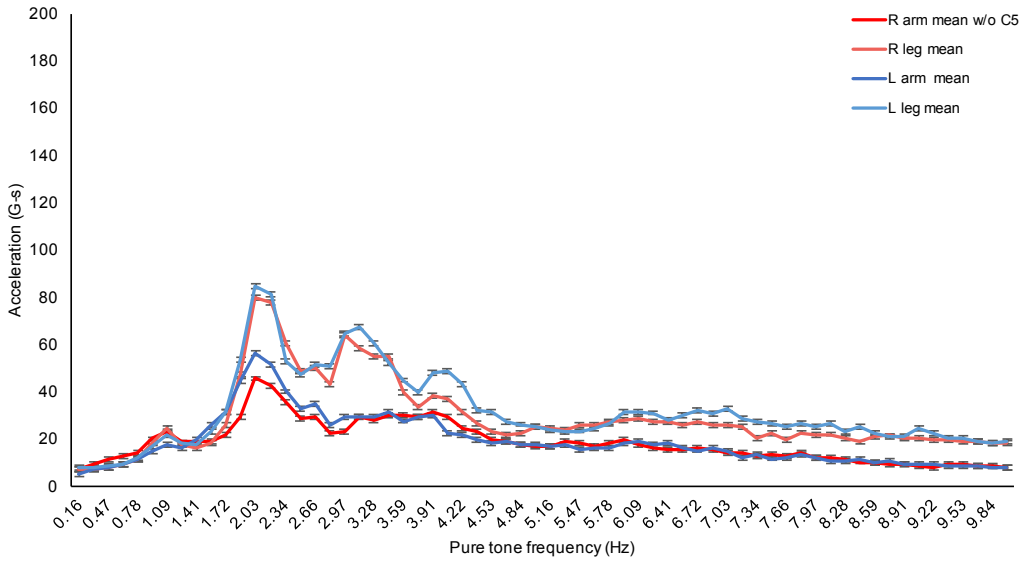
DigiTrac outputs for the X axis were downloaded and compared.

Results

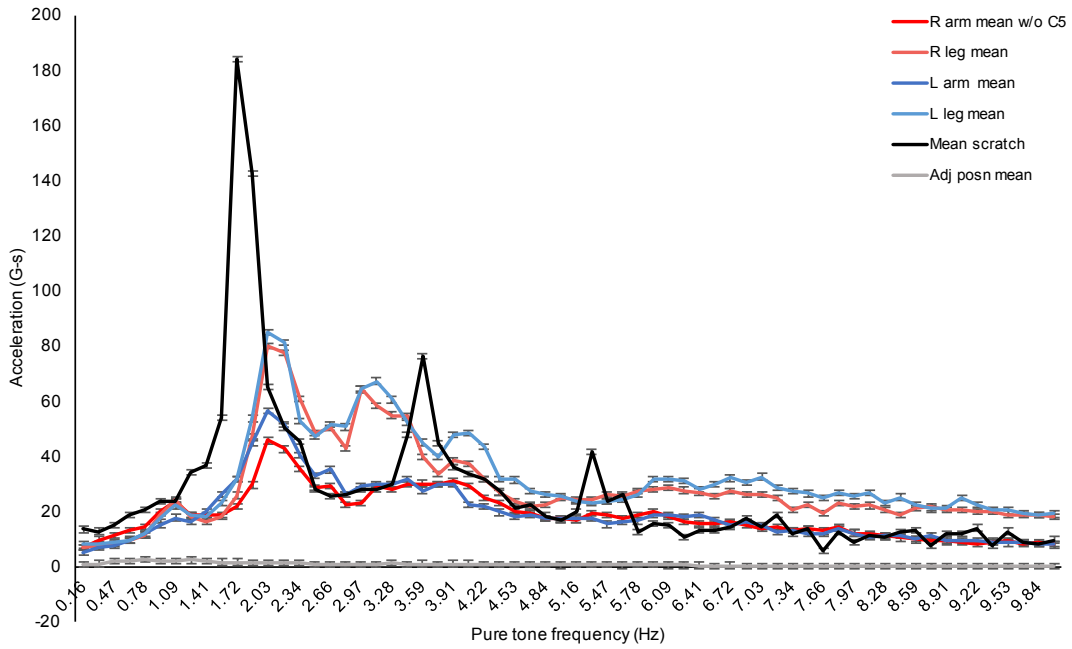
One subject's DigiTrac fell off the right arm. This readout was therefore removed for the purposes of analysis.

The graph in Figure 3.11a demonstrates the pattern of acceleration across the FFT in each pure tone. It can be seen that acceleration is higher in the legs than in the arms and that there is most acceleration at 2.03Hz for all four limbs and that there is a minor peak at 3Hz for leg movement. The mean and standard deviation error bars for each measuring limb is plotted.

Figure 3.11. a) Mean DigiTrac total FFT output of walking measured in each limb (error bars represent standard deviation).



b) Mean DigiTrac total FFT output of walking measured in each limb: stereotyped 'Scratch' and 'Adjust position' mean total FFT superimposed (error bars represent standard deviation).



The graph in Figure 3.11b superimposes stereotyped movement data onto the walking movement graph. This demonstrates that there is a slightly different pattern to the maximal accelerations on the pure tones for itch-related movement (black line) compared to walking: there is a peak at a lower frequency of action, 1.72 and 1.88 Hz with a minor peak at 3.59Hz for stereotyped scratch movement. It can be seen that generalized adjusting position movement (grey line) has minimal acceleration and no true peak.

Conclusion

It would appear from this data that walking has a different characteristic DigiTrac readout from itch-related movement. There are a couple of qualifications to note, however: firstly, walking is quite a natural movement and this has been compared to entirely manufactured scratch. Secondly, n=10 for the walking observation and the superimposed scratch data is from just one subject. When adjusting position is superimposed on the walking data, it can be seen that this is a low acceleration movement with no peak activity at any of the pure tones. This would fit with the observation that this movement is not rhythmical and not fast.

It should be noted that, unfortunately, the range of pure-tones for analysis which should discriminate itch-related movement (suggested from the video-observation study), 0.5-2.5Hz, does overlap with walking movement's (2-3Hz). However, the experiment's results do suggest that discrimination may be improved by taking *magnitude of movement* or acceleration into account as well as *frequency of action*, something which had not previously been considered.

The next step was to use the DigiTrac in a proper clinical observational setting.

Chapter 4

Clinical experiments using DigiTrac accelerometers

I: Translation of directly observed 'frequency of action' to DigiTrac FFT.

Aim

To use directly observed videos of subjects with itch in order to characterise the DigiTrac signal readout.

The study was also used, opportunistically, to validate the DigiTrac recording against Actiwatch Plus (which had already been validated against 'gold standard' infrared video recordings).

Background

The DigiTrac output yields data on acceleration in specific frequencies of action. In theory, therefore, since it is known that the spectrum of frequency of action of itch-related movement is 0.5-2.5Hz, it should be possible to gain specific itch-related movement measurement in this frequency range. In this experiment, DigiTrac output was compared to simultaneous overnight infra-red video recordings and the type of movement recorded at the highest magnitude activity in the 0.5-2.5Hz frequency range directly observed.

Method

Subjects

Six child-subjects were recruited. All had atopic dermatitis (as defined by Hanifin and Rajka criteria (Hanifin & Rajka, 1980). The subjects were recruited from the Paediatric Dermatology clinic, had been diagnosed by a Paediatric Dermatology Consultant and had, as would be expected from patients in a secondary care clinic, moderate eczema: SCORAD range was 18.8 to 37.0 and the median 27.7. The age range of the subjects was 3-15 years (median 9.5yrs). Four were male and two female.

Study method

All subjects wore a DigiTrac accelerometer overnight on the dominant wrist, an Actiwatch Plus on the non-dominant wrist and were simultaneously infra-red videoed at home in their own beds (as described in Chapter 1.II). Only one night was studied. Although, normally, accelerometers would be worn on the dominant wrist, the children were essentially too small to fit both devices on one forearm. Previous research had shown that there was a negligible difference in Actiwatch Plus score on dominant versus non-dominant wrist (Benjamin et al., 2004; Bringhurst et al., 2004) and so this device was placed on the non-dominant wrist.

Instructions were given verbally and in written format including when and how to start the video recording. Parents were also issued with a short questionnaire to complete which included general information such as when the child went to bed and got up and if they had taken medications.

The overnight DigiTrac recording started at 21:00hrs but, because of previous data suggesting a discrepancy between parent-recorded sleep-time and actuality, only the activity between 02:00hrs and 04:00hrs was analysed.

The DigiTrac and Actiwatch Plus outputs were uploaded into an Excel file.

a) ***Extracting a specific itch spectrum from DigiTrac output.***

   The DigiTrac analysis required the sum of acceleration, 0.5-2.5Hz to be extracted for each epoch. The sum and epoch was then ranked so that there was a table of Sum of FFT (G-s) starting with the largest magnitude in the first column, and the corresponding epoch was in the second column. In total there were 1125 epochs for the recording period. The video-recording was then examined and the action at the corresponding epoch was noted (in the third column). The movement was recorded as four categories: scratch, S, (itch-related rhythmical movement including scratch

and rub), adjusting position, A, nil, N, and off-camera, O. The process was repeated for each subject. The examination of the recordings in this intricate way took over one month to complete.

The data was organised into chunks of acceleration of 10G-s. The highest sum of FFT 0.5-2.5Hz was 680G-s and a large proportion were 0G-s. The proportion of each type of action for each these groups of 10G-s was plotted.

b) Validating DigiTrac against Actiwatch Plus

Total hourly scores for the DigiTrac entire sum of FFT and for 0.5-2.5Hz was compared to the hourly Actiwatch Plus score.

Results

a) Extracting a specific itch spectrum from DigiTrac output.

The individual's results are plotted in separate graphs in Figure 4.1. The graphs show the amplitude or size of acceleration on the x axis and the type of movement for that amplitude of movement on the y axis. Imagine a horizontal cut off at 50% (y axis-wise), the simplest way to appreciate the data is to consider that whatever colour line and whatever movement this represents, is the predominant movement at this point (running left to right across the x axis). It can be seen that at a low amplitude, most frequently, there is no movement and that above 400G-s the majority of movement is generalised or scratch.

Figure 4.2 combines the data for all the subjects. The graph shows that the majority of the movement in the amplitude 50-550G-s is scratch, but not exclusively scratch. Generalised movement happens at a similar amplitude but happens less often than scratch movement. If the amplitude is less than 50G-s the likelihood is that no movement at all is detected on the video. Over 550G-s, movement is just as likely to be itch-related as generalised (approximately 50% each).

Figure 4.1. Graphs to illustrate most intense DigiTrac scores related to directly observed movement on infrared video.

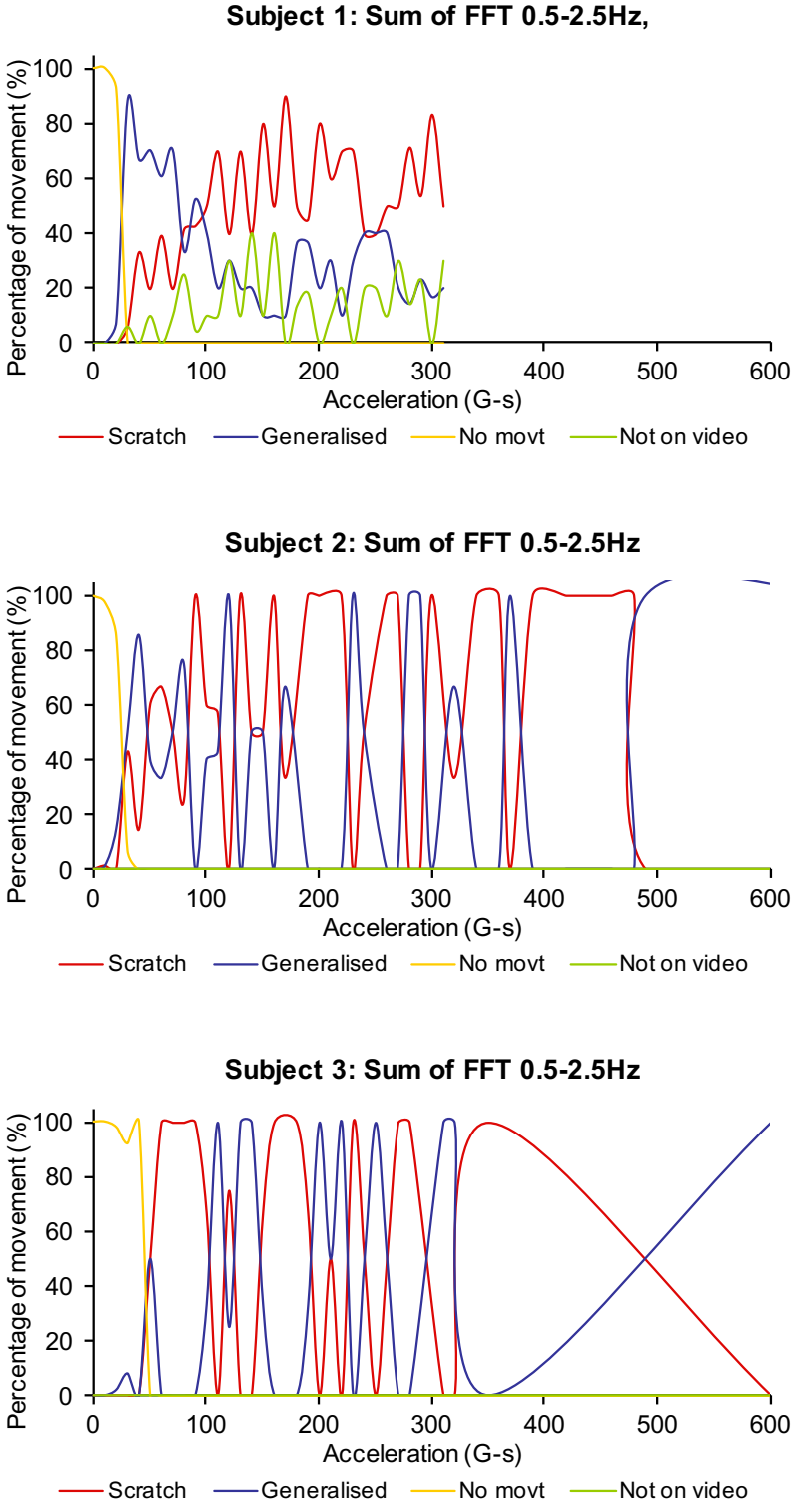
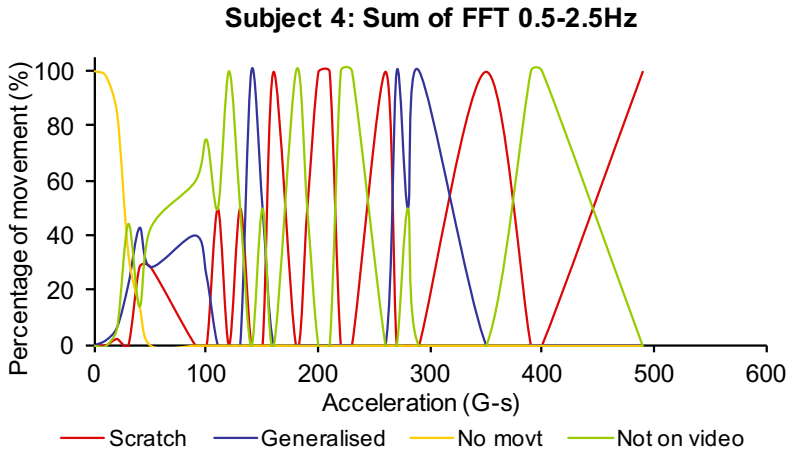
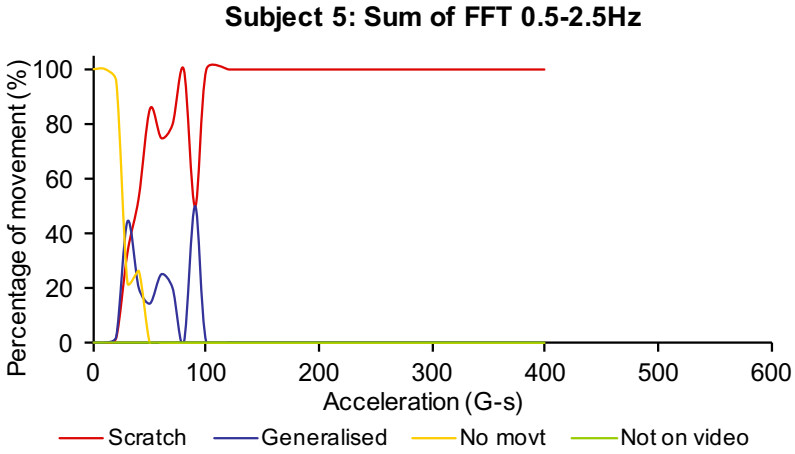


Figure 4.1. (cont'd)





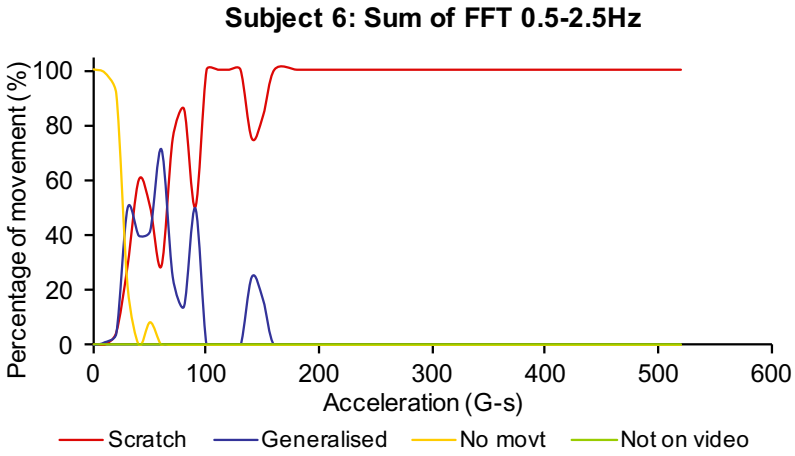
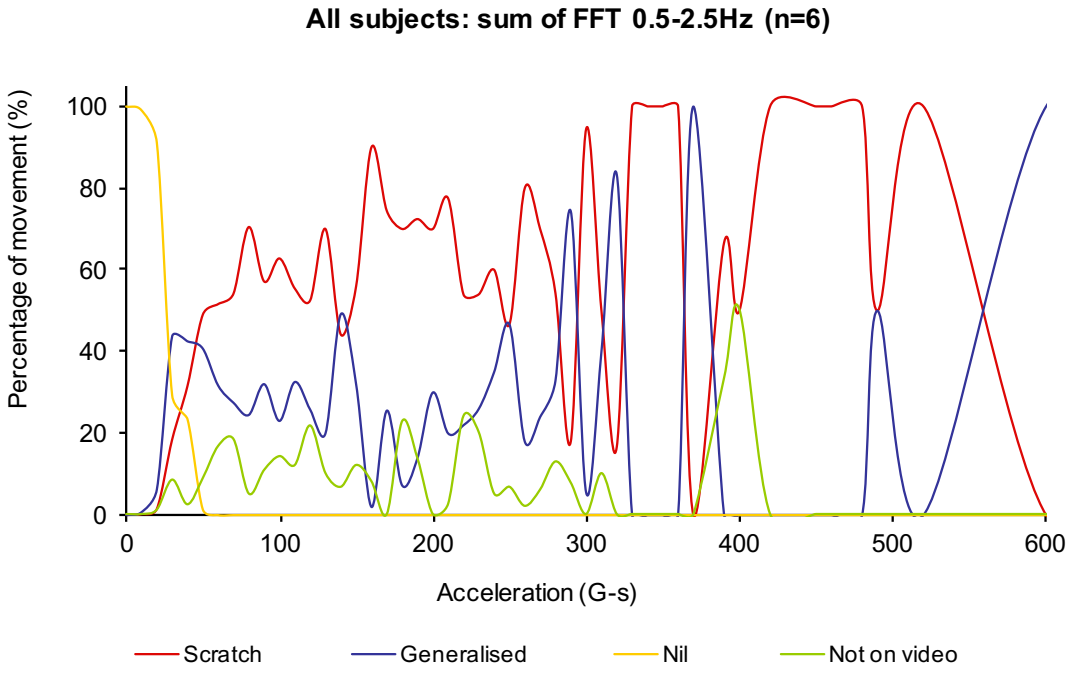


Figure 4.2. Graph to summarise all the subjects' activity. Note that the predominant type of movement is 'Scratch' at amplitudes of 50-550G-s.



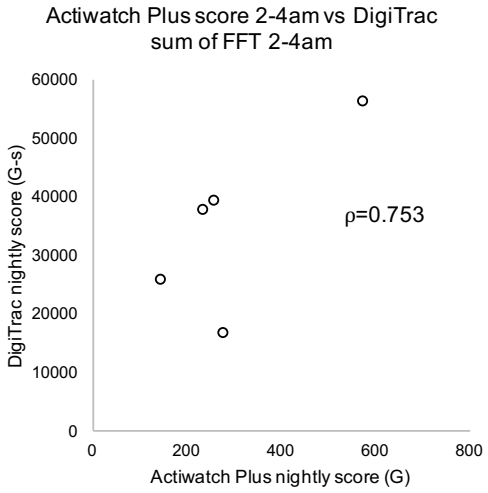
b) Validating DigiTrac against Actiwatch Plus

The total nightly DigiTrac score was compared to the same night's total nightly Actiwatch Plus score. It was apparent from the data download that Subject 1 forgot to put the Actiwatch Plus on.

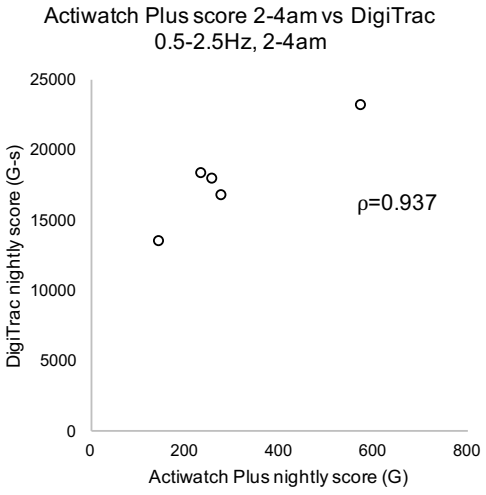
The total 2-4am Actiwatch Plus score was compared to the full spectrum 2-4am DigiTrac score. It was decided to compare the full spectrum FFT on the basis that the Actiwatch Plus was a general movement monitor. Correlation was noted for total Actiwatch Plus score versus total DigiTrac FFT score (Pearson ρ=0.753). The 2-4am Digitrac data for the restricted spectrum of 0.5-2.5Hz was also available and this was also compared to the Actiwatch Plus 2-4am score. The correlation between the limited DigiTrac spectrum and the Actiwatch Plus was higher (Pearson's ρ=0.937). The limits of the spectrum were determined by the directly observed frequency of action of itch-related movement, scratch and rub (see Chapter 1.III) Graphs to illustrate the data are in Figure 4.3.

Figure 4.3. Scatter plots to demonstrate Actiwatch Plus score compared to DigiTrac, a) Actiwatch Plus score versus total FFT DigiTrac output (sum of FFT), b) Actiwatch Plus score versus limited spectrum FFT DigiTrac output (0.5-2.5Hz).

a)



b)



Conclusion

This experiment demonstrates that the majority of movement in the frequency range 0.5-2.5Hz and between the amplitudes 50-550G-s, is itch-related 'scratch' movement. An amplitude of 0-100G-s suggests stillness. An amplitude of 400-500G-s upwards is a larger movement than scratch and suggests generalised or ambulatory movement. Unfortunately, not all movement of medium amplitude is itch-related, some can be low velocity, low amplitude movement related to changing position in bed.

A complicating factor is that it would appear that the majority of human movement occurs in the frequency range 0.5-2.5Hz too (as evidenced by my DigiTrac walking experiment, Chapter 3.V). This finding could be related to overall size and limb length of *Homo sapiens* as a species (a longer limb will take longer to complete a cycle of movement than a short one). One problem is that although psychological overlay and generalised movement is assumed to be kept to a minimum by undertaking studies overnight, it can not be assumed that a subject, especially a child will stay in bed. Children can quite often be on the move if they are not 'good sleepers.' It has been observed by personal anecdote and from discussion with contemporaries that children will get up out of their beds and bedrooms in the search for comfort, milk, etcetera. This could account for the 'Not on camera' movement. A bleary-eyed parent may not record such events: I have often woken up to find my little boy in bed with me in the morning never having been aware of his arrival in the night. Therefore, although parents were issued with questionnaires enquiring about the child's night's sleep which included asking what time they went to bed and if they had arisen overnight, it is entirely possible and likely that nocturnal excursions were not recorded as the parent may not have even consciously registered them.

The study offered the opportunity to validate the DigiTrac against the Actiwatch Plus. As expected, a relationship between the outputs of both types of device has been demonstrated

in this experiment. The relationship between the scores becomes stronger when the DigiTrac score is limited to the sum of 0.5-2.5Hz FFT. One explanation that can be offered to explain this would be differences in the band-pass filters for the devices and that introducing a further band-pass filter in the form of a restricted range of frequency from the FFT brings the results into closer alignment. A further explanation could be related to the fact that the devices were not worn on the same arms and so the filtering of DigiTrac score somehow again provided alignment for this.

On the basis of the observations and data from this experiment, it was hypothesised that the data should first be 'enriched' for the frequency range 0.5-2.5Hz. Since little or no movement was detected below 50-100G-s and larger amplitude, non-itch-related movement occurred at greater than 400-500G-s, a second 'enrichment' was proposed for an amplitude range: either 100-400G-s or 50-550G-s. Therefore, it was assumed that enriching for this frequency of action and amplitude of movement of DigiTrac should provide the most specific and discriminating measure of itch-related movement and this led to the next study.

II: Enriching DigiTrac data for itch-related movement

"Data enrichment is a general term that refers to processes used to enhance, refine or otherwise improve raw data." ***https://www.techopedia.com/definition/28037/data-enrichment***

Aim

To ascertain whether discrimination of itchy subjects from controls can be improved by enriching for a specific range of *'frequency of action'* and furthermore by a range of '*amplitude of movement*.'

Background

The DigiTrac had been specifically sought out as, a specific itch-measuring device. Through its ability to report data based on 'frequency of action,' and since itch-related movement, scratch and rub is rhythmical, it was postulated that its output should allow a more discriminating account of itch-related movement in nocturnal recordings. This study was an exploratory study to ascertain whether the enrichments of data (selecting or narrowing) suggested by previous studies would improve discrimination of itchy subjects from non-itchy controls.

The FFT of DigiTrac data allows break down of the signal for each epoch into a series of pure-tones 0-10Hz (see DigiTrac technical description, Chapter 3.I). If you add the acceleration across the pure tones you get the total acceleration per epoch (comparable to the Actiwatch Plus reading per epoch). Direct observation of subjects' itch demonstrated that itch-related movement occurs most often in a range of 0.5-2.5Hz. The absolute minimum 'frequency of action' of itch-related movement was 0.5Hz and the maximum 5Hz. The first

part of this experiment sought to decide if discrimination of itchy subjects from non-itchy controls would be improved by narrowing the DigiTrac output to a specific frequency range.

The study which compared DigiTrac output 0.5-2.5Hz to direct video recording (Chapter 4.I) suggested that a second enrichment of the DigiTrac readout should be applied for a specific a range of amplitude of movement or acceleration. For instance, it would appear, from video and accelerometer studies that most human movement occurs under 5Hz. Some of these movements will not be rhythmical, and so detecting a certain, specific '*frequency of action'* may prove helpful. However, rhythmical movement is not exclusively itch-related scratch or rub: for example, walking is a rhythmical action, but a large amplitude (entire limbs swing around the pendulums of shoulder and hip joints). It was hypothesised therefore, that excluding large amplitude rhythmical movements should improve the specificity of itch-related movement. A ceiling was therefore placed on amplitude of movement. Since no itch-related movement occurred below 50-100G-s a lower cut-off was proposed to improve specificity too.

Method

This was an exploratory study. Subjects who complained of itch associated with their conditions were compared to non-itchy controls.

Study subjects

Subjects were adults and children recruited from General and Paediatric clinics at the Department of Dermatology, Edinburgh. Controls were people who had no skin condition or itch who attended the department with patients. Nine subjects, four males and five females, age range 3-61 years and median, 12 years were recruited. Twelve controls, three males and nine females, age range 27-72 years and median 33 years were recruited.

Study Method

All participants were asked to wear a DigiTrac on the dominant wrist for one night. The DigiTrac were set to record movement from 18:00hrs. Participants were issued with a questionnaire to note the time of retiring to bed and if they had got up in the night. The recordings of between 02:00hrs to 04:00hrs were chosen for analysis as these were the hours which would mostly account for those spent in bed: adults presumed to have gone to bed, children not likely to have arisen yet.

a) Enriching for 'frequency of action.'

The DigiTrac data was uploaded to a P.C. To extract the data:

1) The entire FFT of the x axis acceleration data for 02:00hrs to 04:00hrs was exported into Excel.
2) Three different spectra were examined and where necessary, extracted from the entire FFT (0-10Hz) in order to enrich for 'frequency of action.'
    a. Total sum of FFT (0-10Hz)
    b. 0.5-2.5Hz: as suggested by mean direct video-observed frequency of itch-related movement.
    c. 0.5-5Hz: chosen as 5Hz was highest possible observed scratch frequency and 0.5Hz the slowest.

Subjects' outputs were compared to controls'.

b) Enriching for 'amplitude of movement.'

The full DigiTrac spectrum was uploaded to a P.C.

1) The entire FFT of the x axis acceleration data for 02:00hrs to 04:00hrs was exported into Excel.
2) The 0.5-2.5Hz spectrum was extracted.

3) Each epoch for each participant was then ranked for amplitude of movement (1125 for the 02:00hr to 04:00hr period).
4) The sum of acceleration for:
    a. all amplitudes
    b. 50-550G-s as broadest range dictated from previous study
    c. 100-400G-s as narrowest range dictated from previous study

Subjects' outputs were compared to controls'.

Since the data was normally distributed, parametric testing was applied (unpaired t test).

Results

Due to a software handling error, data was irretrievably lost on three control subjects such that only the data for FFT 0.5-2.5Hz was accessible. Data was only available on nine control subjects for all other analyses.

The itchy subjects' scores were generally higher than controls' even without enrichment of the DigiTrac output. The separation increases and the overlap decreases stepwise with enrichment for spectrum, or 'frequency of action' see Table 4.1a and Figure 4.4a. The maximal separation (marginally) is seen with enrichment for 0.5-5.0Hz. There is a lot of overlap between groups.

The results of the second enrichment for 'range of amplitude' on top of that for 'frequency of action' are summarised in Tables 4.1b and 4.1c and Figures 4.4b and 4.4c. The range becomes narrower and scatter is seen to be reduced by the enrichment. The effect is similar with either 'range of amplitude.'

Table 4.1a. DigiTrac total nightly scores: enrichment for 'frequency of action.' Last column on right, T test results expressed as p=/<.

	Subject (n=9)			**Control (n=9** except 0.5-2.5Hz n=12)			**Subject vs Control**
	Mean	Median	Standard deviation	Mean	Median	Standard deviation	**Unpaired T test**
Total FFT, all amplitudes	34320	35216	10801	23463	21868	7548	**0.020**
0.5-2.5Hz, all amplitudes	16749	16894	3955	12825	12910	3836	**0.034**
0.5-5Hz, all amplitudes	24131	24694	6538	17206	17220	6047	**0.033**

Table 4.1b. DigiTrac total nightly scores: enrichment for 'frequency of action' **0.5-2.5Hz** and 'range of amplitude.'

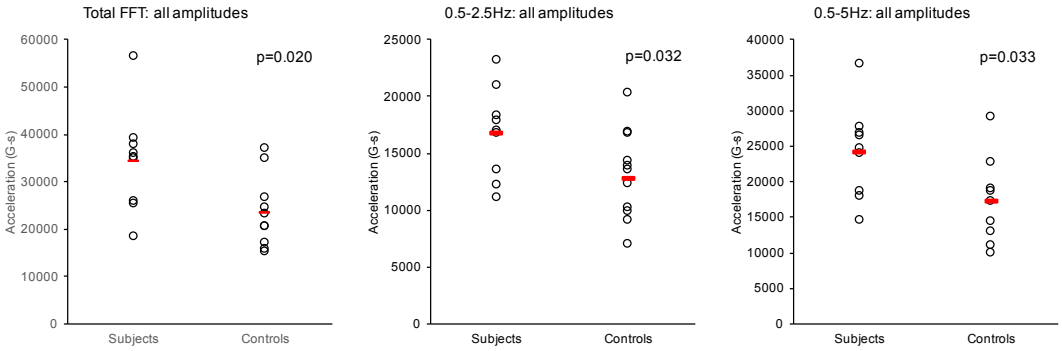
	Subject (n=9)			**Control (n=12)**			**Subject vs Control**
	Mean	Median	Standard deviation	Mean	Median	Standard deviation	**Unpaired T test**
0.5-2.5Hz, all amplitudes	16749	16894	3955	12825	12910	3836	**0.034**
0.5-2.5Hz, 50-550G-s	4086	3728	2784	898	721	596	**<0.001**
0.5-2.5Hz, 100-400G-s	5128	4386	2902	1395	1264	739	**<0.001**

Table 4.1c. DigiTrac total nightly scores: enrichment for both 'frequency of action' **0.5-5.0Hz** and 'range of amplitude.'

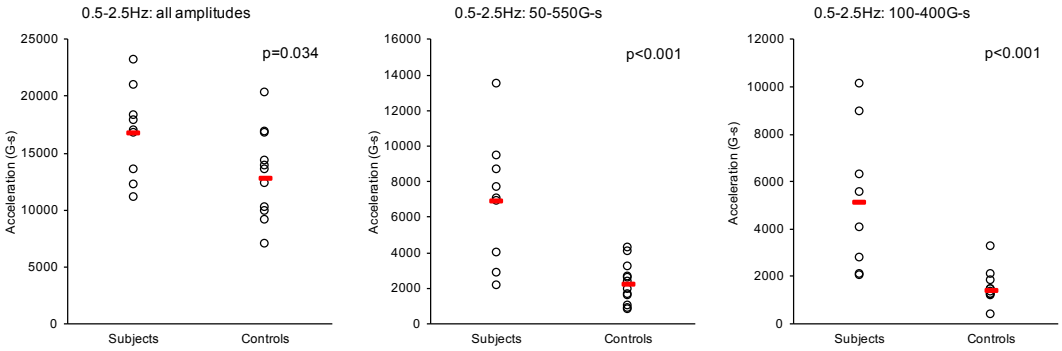
	Subject (n=9)			**Control (n=12)**			**Subject vs Control**
	Mean	Median	Standard deviation	Mean	Median	Standard deviation	**Unpaired T test**
0.5-5.0Hz, all amplitudes	24131	24694	6538	17206	17220	6047	**0.033**
0.5-5.0Hz, 50-550G-s	10469	11228	4208	3558	4110	1582	**<0.001**
0.5-5.0Hz, 100-400G-s	7596	7154	3338	2540	2684	1115	**0.001**

Figure 4.4. Graphs of nightly DigiTrac score, subjects versus controls: a) enriched for different 'frequency of action,' b) enriched for 'frequency of action' 0.5-2.5Hz and 'range of amplitude, c) enriched for 'frequency of action' 0.5-5.0Hz and 'range of amplitude.' Subjects: n=9, Controls: n=12, - = mean

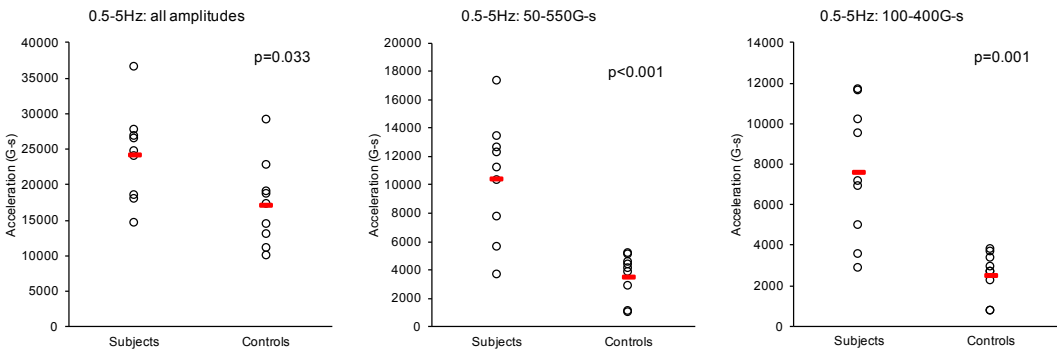
a)



b)



c)



Conclusions

This was a small exploratory experiment to see whether specificity of itch-related movement detection could be improved by using a more complex accelerometer, the DigiTrac, which was capable of providing information on frequency of action. There are some imperfections in the study but as a proof of concept study it fulfils its remit.

The imperfections implied above include that the study subjects are a mix of children and adults with different itchy conditions and that the controls are not matched in any way, a different range of ages and different distribution of sexes. The experiment offered a learning curve in handling the data: through the unfortunate incident of loss of data on three control subjects, safety-guards were put in place to prevent this occurrence in future studies.

There is evidently an advantage in enriching for 'frequency of action.' There is not much difference between the two proposed ranges. It would appear that, similarly, the further enrichment for amplitude improves specificity but that there is little difference between the two proposed ranges of 50-550G-s or 100-400G-s. The enrichment seems to work by improving closer clustering of the data (for controls). The experiment implies that the marginally superior separation is achieved by enriching first for a frequency range of 0.5-2.5Hz as both further enrichments for amplitude of 50-550G-s and 100-400G-s prove highly significant on unpaired t test.

On the basis of this pilot study, a more extensive study was then undertaken.

III: Use of DigiTrac accelerometer as a monitor of itch on consecutive nights in children with atopic dermatitis and comparing subjective and objective disease measures.

Aim

1) To use the DigiTrac accelerometer as an objective and specific measure of itch in children with atopic dermatitis for consecutive nights and compare these scores to controls subjects'.
2) To evaluate objective DigiTrac scores to subjective measure of itch (VAS) and disease-extent as measured by SCORAD.

Background

Use of a digital accelerometer had been validated to measure itch in children (Benjamin et al., 2004; Bringhurst et al., 2004).The study had raised interesting observations, firstly that there was considerable variation in night-to-night score and secondly that there was a disconnect between subjective and objective score. This study sought to discover whether the night-to-night variation and disconnect between subjective and objective score was because the Actiwatch Plus accelerometer was a non-specific movement monitor as opposed to the DigiTrac which, through enrichment of data could provide a more specific measure of itch-related movement.

Method

Study subjects

Subjects were consecutive children with atopic dermatitis (as defined by Hanifin and Rajka criteria (Hanifin & Rajka, 1980)) approached at the Paediatric clinic at the Royal Infirmary of Edinburgh and whose parents then agreed for them to participate. Controls were siblings of subjects or children of departmental members who had no present skin condition or itch.

Inclusion criteria:

Subjects

- Patients with atopic dermatitis defined by the Hanifin and Rajka criteria or patients with itchy skin due to other causes (skin or systemic either due to primary skin disease or systemic disease).
- Age >3 months
- Able to give meaningful consent or have a parent/guardian to give their consent if <16 years.

Controls

- Age >3 months
- Able to give meaningful consent or have a parent/guardian to give their consent if <16 years.

Exclusion criteria

Subjects

- Children under 3 months of age
- People not able to understand what is involved or, if ages <16 years do not have a parent or guardian who can understand and allow what is involved.
- People allergic to rubber.

Controls

- Children under 3 months of age
- People not able to understand what is involved or, if ages <16 years do not have a parent or guardian who can understand and allow what is involved.
- People allergic to rubber.
- People with itchy skin.

26 subjects and 26 controls were recruited. The subjects comprised 15 males, 11 females. The age range for subjects was 3 years to 15 years and the median was 6 years. The controls comprised 12 males and 14 females. The age range for controls was 2 years to 15 years and the median age was 11 years. As would be expected by the fact that subjects were recruited from a secondary care clinic, the range of SCORAD was 19.2 to 63.4 and the median 36.9.

Study Method

All participants were asked to wear a DigiTrac on the dominant wrist for three consecutive nights. The DigiTrac were set to record movement from 18:00hrs. The investigator met with the participant after each recording night in order to download the data and reset the DigiTrac for the next night. The opportunity was also taken to complete a disease extent score, SCORAD in subjects each time. Participants were issued with a questionnaire on which to note the time of retiring to bed and if they had got up in the night. Wherever possible, the questionnaire was completed by the actual participant but some of the children were too young to do this and so the forms were completed by proxy by the parent(s). The questionnaire also included a visual analogue scale (VAS) to be completed for itch and insomnia for the children with atopic dermatitis. Since there was only one investigator, no allowances needed to be made for inter-observer variability

The DigiTrac downloads for 02:00hrs to 04:00hrs were analysed for reasons previously described. Since best discrimination had been found in the previous study by combining the 0.5-2.5Hz spectrum with 100-400G-S and 50-550G-s amplitude range, these were the main enrichments applied. For comparison, full FFT spectra (or full DigiTrac readout) per epoch was also used for analysis.

Each individual's nightly DigiTrac data was uploaded to a P.C. Data was extracted as previously described:

1) The entire FFT of the x axis acceleration data for 02:00hrs to 04:00hrs was exported into Excel.
2) The limited spectra of 0.5-2.5Hz were extracted and each epoch output summed.
3) Each epoch sum for each participant was then ranked for amplitude of movement (1125 for the 02:00hr to 04:00hr period).
5) The sum of acceleration for:
    a. 50-550G-s as broadest range dictated from previous study
    b. 100-400G-s as narrowest range dictated from previous study.

Subjects' outputs were compared to controls'. Since mean and median were closely related, a parametric test was applied, an unpaired t test. The subject and control groups had different variance and thus two sample t tests for heteroscedastic (or samples of unequal variance) were run. When data was combined for all nights, mean data for all three nights was compared to allow for nesting of experimental results.

Subjects' subjective data (VAS) and SCORAD were compared to DigiTrac 0.5-2.5Hz 100-400G-s data. Unpaired t tests for heteroscedastic data were run.

Results

One subject's entire data had to be excluded as it became evident from the download of the DigiTrac that it had either malfunctioned or she had forgotten to put the DigiTrac on. Therefore, analyses were for 25 subjects (15 males, 10 females, the age range and median unchanged from previously stated). Regarding completeness of data otherwise: out of 75 nights' recordings for subjects, 11 (15%) were missed (no data on DigiTrac) For the controls, 4 nights were missed out of 78 (5%). There were more missing nights' data as the nights progressed for subjects due to their being less compliant as the nights went on.

The Table 4.2 summarises the subject characteristics, subjective score and disease extent score, SCORAD. Table 4.3 summarises the controls' characteristics.

Comparison of subjects' to controls' scores at varying levels of data enrichment

The DigiTrac downloads for each night were extracted and enriched as detailed in the Methods section, see Table 4.4 (subjects) and Table 4.5 (controls). The subject data was compared to controls' by unpaired t test. Firstly, it was analysed night-by-night (so that the statistics were not marred by overpopulation by the more reliable participants (Table 4.6).

Table 4.2. Subjects' phenotypic characteristics.

	Gender	Age	Other atopy?	Steroid	Food allergies	Family history	Who filled form?	SCORAD 1	SCORAD 2	SCORAD 3	VAS itch 1	VAS itch 2	VAS itch 3	VAS overall
Subject 1	Female	6	Hayfever	Potent	No	Yes	Adult	24	24	24	1.5	0.5	1.5	1.2
Subject 2	Male	3	No	Potent	Yes	No	Adult	24.7	22.7	23.7	2	2	5	3.0
Subject 3	Female	5	No	No	No	No	Adult	44.7	46.2	49.7	3	3	5	3.7
Subject 4	Female	3	No	Mild	No	No	Adult	44.5	39.5	46.5	3	2.5	3	2.8
Subject 5	Female	11	Asthma	Mild	No	No	Subject	34.2	34.2	33.7	6.5	6	6	6.2
Subject 6	Male	5	No	Mild	No	No	Adult	30.4	30.4	30.4	1.5	1	2.5	1.7
Subject 7	Male	7	Asthma	Mod	No	Yes	Adult	26.7	25.7	26.2	6	5.5		5.8
Subject 8	Female	11	No	Potent	No	Yes	Adult	25.4	25.9		1	4	7.5	4.2
Subject 9	Male	9	Asthma	Mod	No	Yes	Adult	34.2	41.2	45.2	5.5	4		4.8
Subject 10	Male	14	Asthma	Mod	No	Yes	Subject	38.5	37.5			7.5	7.5	7.5
Subject 11	Female	5	No	Mild	Yes	No	Adult		38.7	39.2	3	2.5	2	2.5
Subject 12	Female	6	Asthma	Potent	No	Yes	Adult	24.8	24.8	24.3	1.5	1.5	3	2.0
Subject 13	Male	15	Hayfever	Elidel	No	No	Subject	42.6	42.6	45.1	7.5	7.5		7.5
Subject 14	Female	15	No	Mild	Yes	Yes	Adult	37	36		3	3		3.0
Subject 15	Male	10	Hayfever	No	No	Yes	Adult	21.5	20.5		5	6	5	5.3
Subject 16	Male	6	No	Potent	Yes	No	Adult	61.6	61.6	63.6	1	1	5	2.3
Subject 17	Male	11	Asthma	Mild	Yes	Yes	Adult	32.6	32.6	41.1	3	3		3.0
Subject 18	Male	3	No	Mod	No	Yes	Adult	18.8	18.8		2.5	3	1.5	2.3
Subject 19	Female	4	Asthma	Mild	No	No	Adult	24.5	21.5	20	1	3	3.5	2.5
Subject 20	Female	12	No	Potent	No	No	Subject	30.4	33.9	34.4	3	3	2	2.7
Subject 21	Male	6	No	Potent	No	No	Adult	40.1	39.6	39.1	5	6	6.5	5.8
Subject 22	Male	10	Asthma	Mild	Yes	Yes	Adult	42.9	43.4	48.9	7	7.5	8	7.5
Subject 23	Male	5	Asthma	Mild	Yes	Yes	Adult	63.9	62.9	65.9	6	6.5		6.3
Subject 24	Male	8	No	Potent	Yes	Yes	Adult	30.9	30.4		2			2.0
Subject 25	Male	6	Asthma	Protopic	Yes	Yes	Adult	29.9						2.5

Table 4.3. Control participants' characteristics

	Gender	Age	Previous eczema	Atopy	Food allergy
Control 1	Male	12	Yes	Hayfever	No
Control 2	Male	9	No	No	No
Control 3	Female	15	No	No	No
Control 4	Female	6	No	No	No
Control 5	Female	11	No	No	No
Control 6	Female	13	Yes	No	No
Control 7	Male	15	No	No	No
Control 8	Male	15	Yes	Asthma	No
Control 9	Female	11	No	No	No
Control 10	Female	11	No	No	No
Control 11	Male	12	Yes	No	No
Control 12	Female	9	No	Asthma	No
Control 13	Female	4	No	No	No
Control 14	Male	15	No	Hayfever	No
Control 15	Male	2	No	No	No
Control 16	Female	7	No	No	No
Control 17	Female	4	No	No	No
Control 18	Male	15	Yes	Asthma	Dairy and fish
Control 19	Female	4	No	Asthma	No
Control 20	Female	8	No	No	No
Control 21	Female	12	No	No	No
Control 22	Male	15	No	No	No
Control 23	Male	7	No	No	No
Control 24	Male	12	No	No	No
Control 25	Female	4	No	No	No
Control 26	Male	3	No	No	No

Table 4.4. Pure DigiTrac scores and enriched data for each **subject** for each night (blank spaces are unrecorded nights).

	Night 1					Night 2					Night 3				
	Subjects: total FFT		Subjects: 0.5-2.5Hz			Subjects: total FFT		Subjects: 0.5-2.5Hz			Subjects: total FFT		Subjects: 0.5-2.5Hz		
	Total	100-400G-s	Total	50-550G-s	100-400G-s	Total	100-400G-s	Total	50-550G-s	100-400G-s	Total	100-400	Total	50-550	100-400
Subject 1	25018	2256	15872	1336	980	61426	6560	27300	9686	6820	38204	4524	18750	5332	4244
Subject 2	21860	2582	12222	2344	2200	20008	3340	11416	5844	2206	25884	4126	12834	3812	3060
Subject 3	12928	1758	7082	2378	1886	26752	5088	14614	4926	3638	80100	15442	28578	20576	17870
Subject 4	51572	12018	22180	11044	8900	20592	5154	11058	3340	2594	36250	10466	15948	7512	4046
Subject 5	22100	3054	13510	1766	1324										
Subject 6	47962	7838	19596	8240	6612	55180	6546	22076	10016	7552	120014	27378	36720	23878	17422
Subject 7	28832	3774	12416	4696	3286	40736	6386	15346	7482	4898	17476	4104	9570	1680	1170
Subject 8	39210	6606	17980	8140	7304	30474	5332	15622	6028	3802					
Subject 9	20612	3720	10228	2230	1554	20480	2548	10954	2958	2578					
Subject 10	81654	11190	30036	15604	11922	59046	16270	24758	16256	13620	28810	4868	15210	3014	2352
Subject 11						26778	7498	12252	5276	3920					
Subject 12	30302	2446	17248	1608	1186	37522	7546	19702	3700	2162	35742	4656	19790	2890	2260
Subject 13	60448	6962	18130	9004	7170	56444	6810	19160	11966	8656	75298	8604	23414	15664	9734
Subject 14	25762	6282	13524	3970	2746	100748	46290	30718	18346	10042					
Subject 15	37806	10980	18338	8634	6154	62470	13962	24584	15238	11514					
Subject 16	68506	1798	24120	15268	8964	35482	2938	15582	7346	4100	35332	4228	15448	7204	5794
Subject 17	7058	2490	4082	1480	720	16348	9150	6322	2848	1570	30110	8774	15712	6558	3808
Subject 18	35238	5684	16786	7028	5450	22332	2644	12362	3946	3488					
Subject 19	75378	24972	26002	14228	9486	33480	6928	14376	5270	3710	22582	6732	13114	3218	1972
Subject 20	21766	3214	12416	4666	2904	26800	2560	14124	4520	3708	21328	2998	11248	5368	3758
Subject 21	65302	14284	20960	14420	10226	33584	4616	12788	7812	4808	25672	4346	10708	4498	2664
Subject 22	24842	4804	11824	4310	3690	137264	15556	39144	28458	17002	99636	13810	28322	19640	14910
Subject 23	25304	6610	12158	2670	1250	91490	9322	32364	22554	17030	27998	5776	13402	5016	3842
Subject 24	56394	12688	23140	13470	10124	70766	18568	26978	17436	13294					
Subject 25	25296	4952	12168	2852	2052										

Table 4.5. Pure DigiTrac scores and enriched data for each **control** participant for each night (blank spaces are unrecorded nights).

	Night 1					Night 2					Night 3				
	Controls: total FFT		Controls: 0.5-2.5Hz			Controls: total FFT		Controls: 0.5-2.5Hz			Controls: total FFT		Controls: 0.5-2.5Hz		
	Total	100-400G-s	Total	50-550G-s	100-400G-s	Total	100-400	Total	50-550	100-400	Total	100-400	Total	50-550	100-400
Control 1	25196	4790	12450	4134	2952	35726	7894	17298	8214	5444	14330	3130	7774	2228	1786
Control 2	23692	4272	14390	2746	1886	33570	4946	16818	5382	4180	31846	3020	16588	4440	2288
Control 3	30922	3422	18604	4946	2466	24620	4330	14758	2822	2396					
Control 4	20360	5462	10130	4946	2096	16056	3306	8568	2822	1272	10074	2458	6486	17444	414
Control 5	31068	4070	17064	3828	2472	42420	8570	20318	1992	4828	47226	9166	22100	908	5464
Control 6	21026	3230	10244	3238	3340	22988	3774	11448	6468	2700	37746	7512	15898	7446	7228
Control 7	22420	3154	13720	4600	1862	20340	1384	12642	4208	628	20082	1976	12590	10618	648
Control 8	27224	3300	12302	2160	2630	15718	2726	7658	946	1270	24658	5290	11142	974	3080
Control 9	34910	5954	17670	4012	5954	46756	8144	19932	1868	5392	48312	10760	20934	4270	6998
Control 10	42042	6254	21602	4192	5168	21608	2702	13500	7508	1228	31624	2798	18016	9310	3600
Control 11	30260	9680	14832	8370	3938	20550	5722	11322	1978	2200	24326	5006	12996	4488	3320
Control 12	45618	4326	24534	6414	3536	37426	1764	20696	3032	3040	32116	2178	18358	4128	1266
Control 13	22240	2986	14306	5714	378	30258	6510	16682	2572	1286	28430	5110	16322	1656	1176
Control 14	34642	5348	18756	784	2030	40446	7786	21268	2458	4280	4874	1242	2766	2294	252
Control 15	23522	3808	11820	2706	2676						17922	2376	10550	5590	1090
Control 16	18850	8408	9142	546	2038	34192	15564	13836	3280	5402	14566	6800	7704	1828	1392
Control 17	43352	4234	21790	3428	4732	41938	2018	22746	8152	2898	41040	4170	22994	2268	4322
Control 18	16086	3026	9572	5604	1034	17378	4602	9730	4370	664	16690	3778	9556	5604	1246
Control 19	22616	5366	9180	1526	2378										
Control 20	18008	4348	8578	1500	1422	20624	3508	9518	1750	3174	15982	1468	8306	3242	2230
Control 21	27146	5728	14628	2284	1518	32190	5714	15506	3904	2788	30998	3828	15584	2560	2784
Control 22	28022	6562	14320	2388	1914	20506	2696	12184	3828	238	20604	1466	12666	3626	860
Control 23	30422	3634	15136	2650	2796	25820	4432	14362	812	2056	26054	3538	13824	1176	1266
Control 24	31620	6482	12906	3594	4484	30428	11008	12394	2978	2722	22842	4302	10748	2078	2486
Control 25	16146	3338	9070	5750	1414	39686	5714	15786		264	20148	3120	10340	5384	2070
Control 26	23360	8996	11698	3418	1836	17748	4608	9604	2378	390	49784	15966	18498	2516	6246

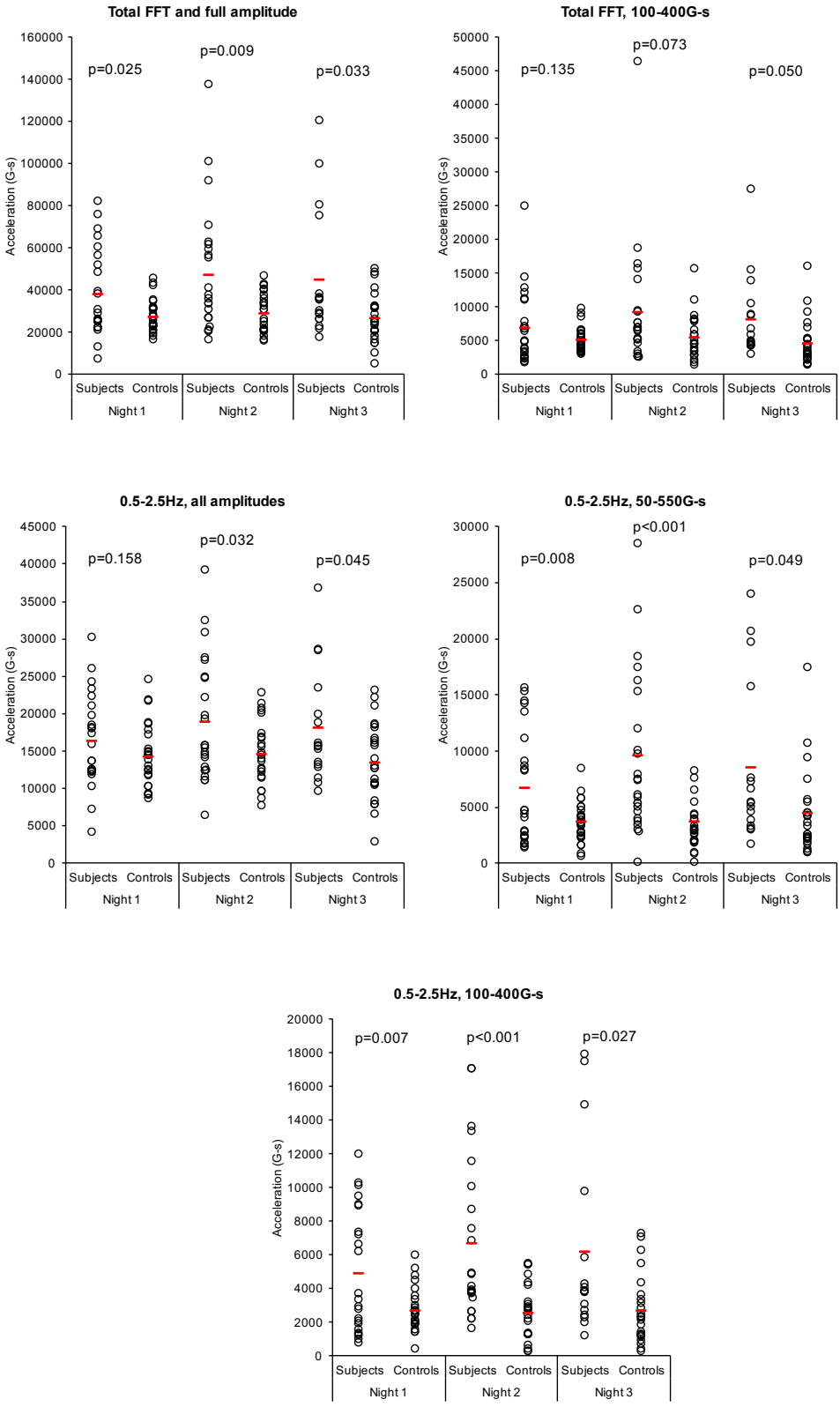
Table 4.6. Comparison of subjects' and controls' scores for each night.

	Night 1				Night 2				Night 3			
	Mean	Median	St deviation	T test	Mean	Median	St deviation	T test	Mean	Median	St deviation	T test
Subjects: total FFT, total acceleration	37965	29567	20574	**0.025**	47226	35482	30249	**0.009**	45027	32721	30980	**0.033**
Controls: total FFT, total acceleration	27337	26171	8003		28708	28039	9551		26345	24492	12053	
Subjects: total FFT 100-400G-s	6790	5318	5385	**0.135**	9201	6560	9266	**0.073**	8177	5322	6291	**0.050**
Controls: total FFT 100-400G-s	5007	4337	1859		5393	4605	3247		4602	3658	3415	
Subjects: 0.5-2.5Hz, total acceleration	16334	16329	6105	**0.158**	18852	15582	8266	**0.032**	18048	15580	7603	**0.045**
Controls: 0.5-2.5Hz, total acceleration	14171	14013	4308		14524	14099	4293		13448	12831	5204	
Subjects: 0.5-2.5Hz 50-550G-s	6724	4681	4950	**0.008**	9620	7346	6953	**<0.001**	8491	5350	7165	**0.049**
Controls: 0.5-2.5Hz 50-550G-s	3672	3511	1822		3640	2978	2154		4420	3434	3745	
Subjects: 0.5-2.5Hz 100-400G-s	4920	3488	3584	**0.007**	6640	4100	4821	**<0.001**	6182	3825	5611	**0.027**
Controls: 0.5-2.5Hz 100-400G-s	2652	2422	1322		2531	2548	1691		2646	2150	2048	

It can be seen that, if the nights are examined individually (Table 4.6), the separation between subjects and controls does not consistently reach significance. Using the entire FFT, for nights 1, 2 and 3, subject versus controls' means are 37965G-s and 27337G-s (p=0.025), 47226G-s and 28708G-s (p=0.009) and 45027G-s and 26345G-s (p=0.033) respectively. With enrichment for amplitude (100-400G-s) the means for subjects versus controls for nights 1, 2 and 3 are 6790G-s and 5007G-s (p=0.135), 9201G-s and 5393G-s (p=0.064) and 8177G-s and 4602G-s (p=0.050) respectively. It can be seen that the variance is large in Figure 4.5 impairing the ability to separate subjects from controls.

However, in the night-to-night analysis, enrichment for 'frequency of action' (whether 0.5-2.5Hz or 0.5-5Hz), with or without further enrichment for amplitude reliably separates the groups to the point of significance. Enrichment purely for 'frequency of action' 0.5-2.5Hz yields means for subjects versus controls for night 1, 2 and 3 of 16334G-s and 14171G-s (p=0.158), 18852G-s and 14524G-s (p=0.032) and 18048G-s and 13448G-s (p=0.045) respectively. Enrichment for 'frequency of action' 0.5-2.5Hz and amplitude 50-550G-s yields means for subject versus controls for night 1, 2 and 3 of 6724G-s and 3672G-s (p=0.008), 9620G-s and 3540G-s (p<0.001) and 8491G-s and 4420G-s (p=0.049) respectively. There is little difference in the improvement of separation by applying the filter of amplitude by 50-550G-s or 100-400G-s, but the later is slightly more effective: means for subject versus controls for nights 1, 2 and 3 are 4920G-s and 2652G-s (p=0.007), 6640G-s and 2531G-s (p<0.001) and 6182G-s and 2646G-s (p=0.027) respectively. The enrichment appears to reduce spread of the controls' scores whilst the subjects' remain very variable. See Figure 4.5.

Figure 4.5. Subjects' versus Controls' scores analysed separately for each night, subjects n=25, controls n=26, -=mean.



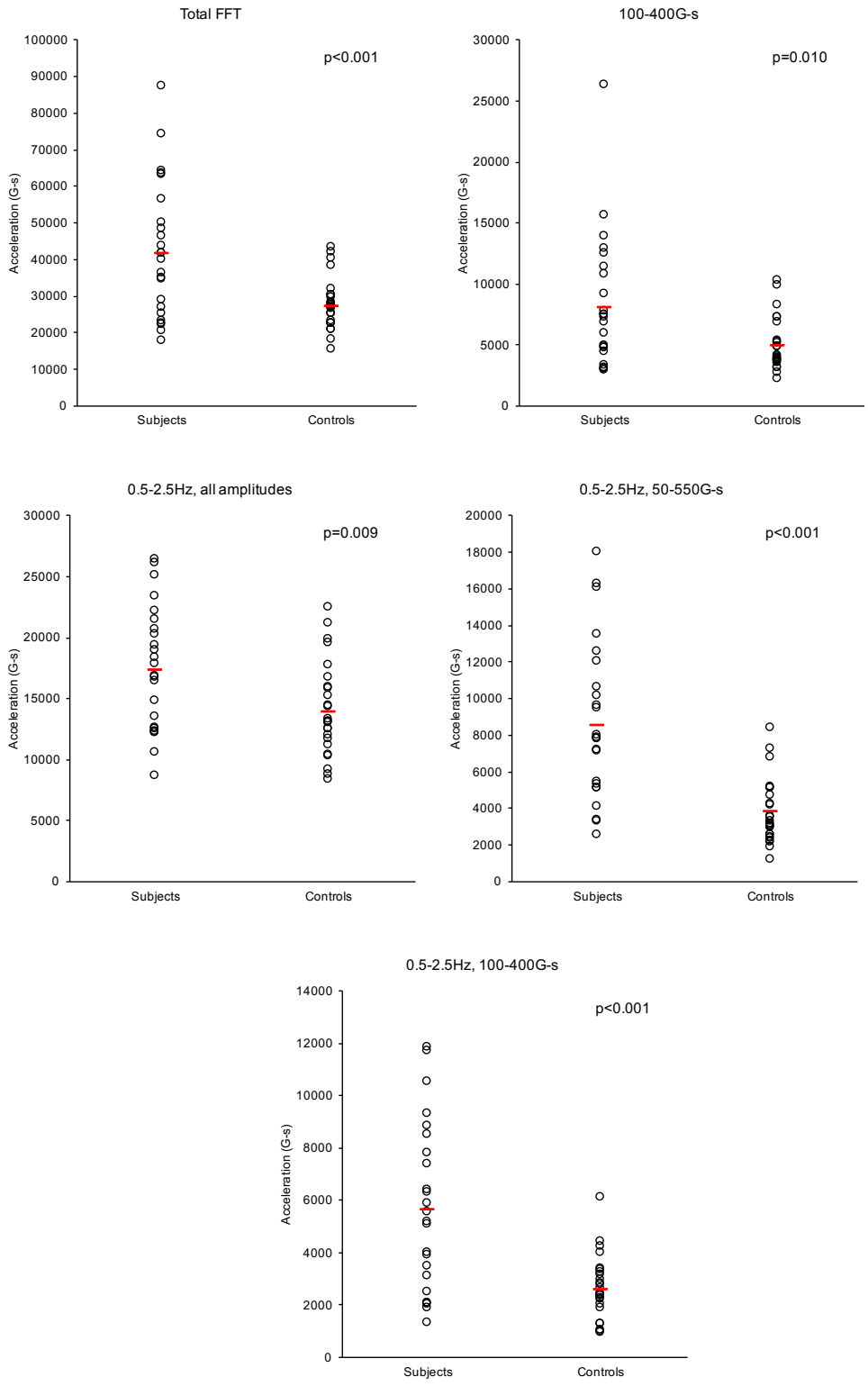
Secondly, all three nights' data was amalgamated to compare subjects' scores to controls.' Each night's data for each subject was meaned to allow for nesting in the study. The results are summarised in Table 4.7. It can be seen that whether the DigiTrac data is enriched or not, the groups are separated to the point of significance (Table 4.7). The means for subjects versus controls of the DigiTrac total FFT, unenriched are 41697G-s and 27253G-s respectively ($p<0.001$). The means for subjects versus controls of the DigiTrac total FFT enriched for amplitude 100-400G-s are 8054G-s and 4971G-s respectively ($p=0.010$). The means for subjects versus controls of the DigiTrac output enriched only by frequency 0.5-2.5Hz are 17345G-s and 13923G-s respectively ($p=0.009$). When the DigiTrac output is enriched for both frequency 0.5-2.5Hz and amplitude, the means of subjects versus controls are 8526G-s and 3855G-s ($p<0.001$) for 50-550G-s and 5644G-s and 2593G-s ($p<0.001$) respectively.

The overlap is reduced by clustering the data, or reducing variance, most especially of the control subjects through the enrichment. The combined data echoes the night by night analysis inasmuch as most separation is achieved by applying the enrichments of 0.5-2.5Hz and 100-400G-s: mean for subjects 8230G-s and mean for controls 3855G-s (t test $p<0.001$). The graphs in Figure 4.6 particularly illustrates the stepwise reduction in variance by data enrichment.

Table 4.7. Comparison of subjects' and controls' scores, all nights together. Last column on right: T test p=/<.

All three nights combined	Mean	Median	St deviation	T test
Subjects: total FFT, total acceleration	41697	39927	18262	**<0.001**
Controls: total FFT, total acceleration	27253	26815	7288	
Subjects: total FFT 100-400G-s	8054	7236	5310	**0.010**
Controls: total FFT 100-400G-s	4971	4068	2102	
Subjects: 0.5-2.5Hz, total acceleration	17345	16801	4994	**0.009**
Controls: 0.5-2.5Hz, total acceleration	13923	13053	3852	
Subjects: 0.5-2.5Hz 50-550G-s	8526	7813	4282	**<0.001**
Controls: 0.5-2.5Hz 50-550G-s	3855	3244	1806	
Subjects: 0.5-2.5Hz 100-400G-s	5644	5180	3164	**<0.001**
Controls: 0.5-2.5Hz 100-400G-s	2593	2405	1227	

Figure 4.6. Subjects' and controls,' all nights analysed together: subjects n=25, controls n=26, -=mean.



Night-to-night variation in score.

Night-to-night scores for each individual (and the mean, red X) are plotted to examine the amount of variation within individual and between individual, night by night. The data is summarised in Figure 4.7's graphs. The graphs demonstrate the large variance *within* and *between* person. The variance is greater for subjects rather than controls. The data also shows that there are not inherently 'high-scorers' or inherently 'low-scorers', by this, I mean there are not individuals who consistently give a high score night by night nor those who always yield low scores night-by-night. The graphs further illustrate how the enrichment steps, especially 0.5-2.5Hz and 100-400G-s improves the clustering, or reduces the variance, of controls' scores.

Subjective VAS score and SCORAD compared with objective (subjects only).

For the purposes of this part of the study, the DigiTrac data enriched by 0.5-2.5Hz and 100-400G-s was used as this seemed to provide the best separation of itchy subjects from controls, and therefore – it was assumed - was most likely to represent the objective measurement of itch. Tables 4.6 and 4.7 summarises the data. Figure 4.8 demonstrates the relationship between DigiTrac score (on x axis as deemed most likely to be a true, objective score) and subjective VAS score of itch. A poor relationship was noted between DigiTrac score and VAS: correlation coefficient, rho= 0.289. To answer whether a lack of relationship was related to whether the subject (rho=0.253) or a proxy (rho= 0.616) had completed the subjective data, these subjects were separated. Again, no significant relationship is demonstrated whether the patient completed their own score or not.

Figure 4.7. Subjects' DigiTrac scores on consecutive nights for different levels of enrichment: subjects n=25, controls n=26.

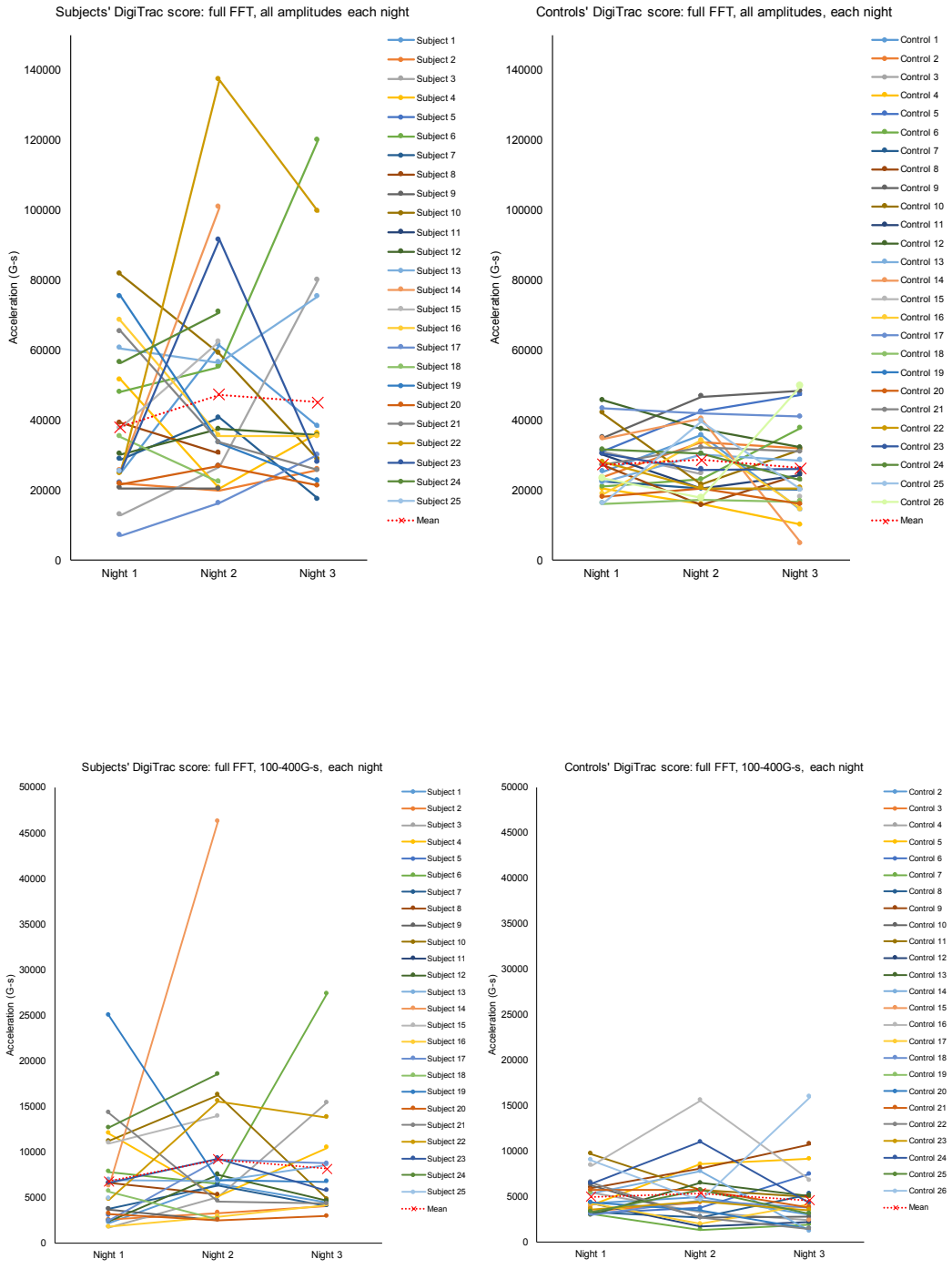


Figure 4.7. (cont'd).

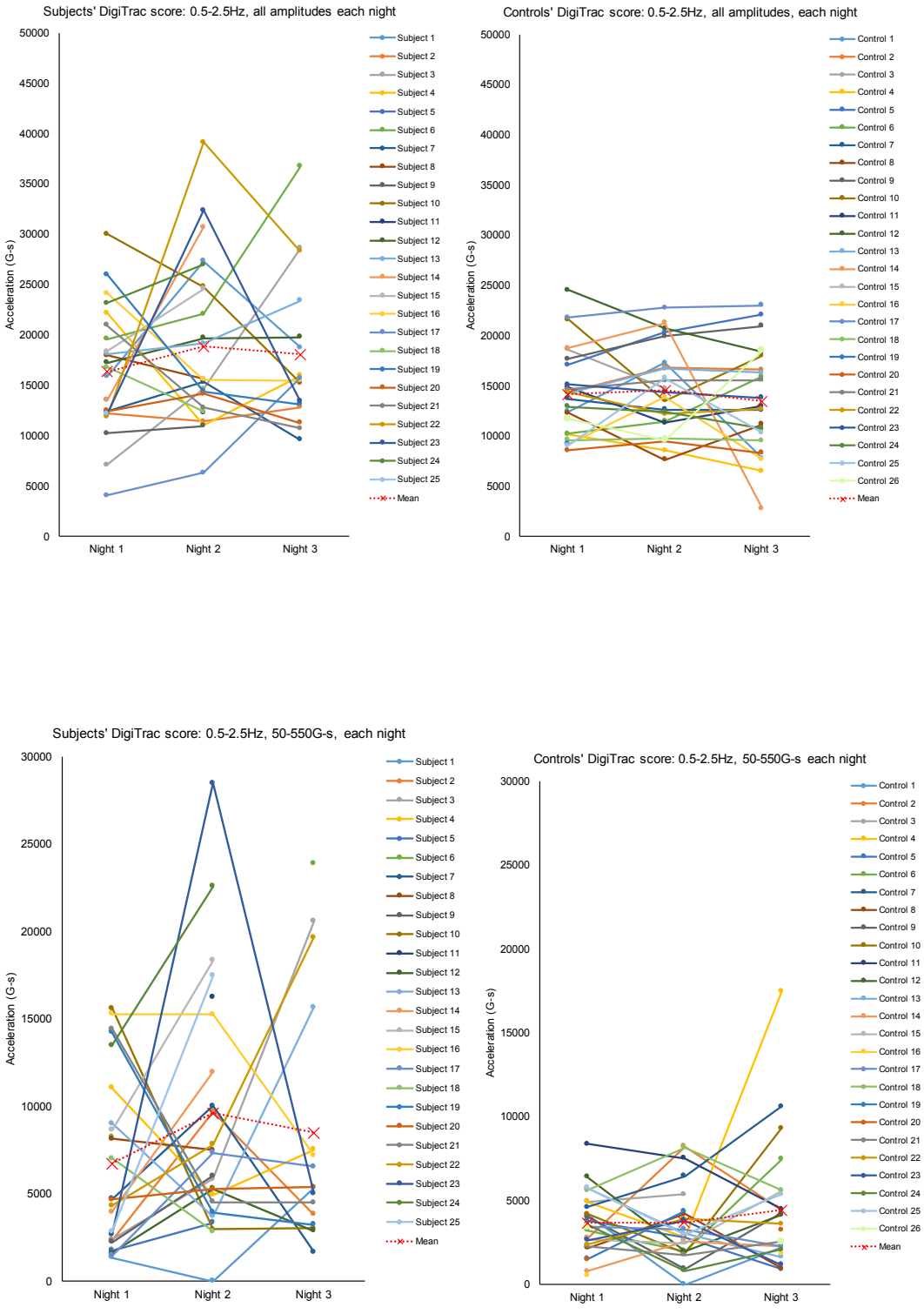


Figure 4.7. (cont'd).

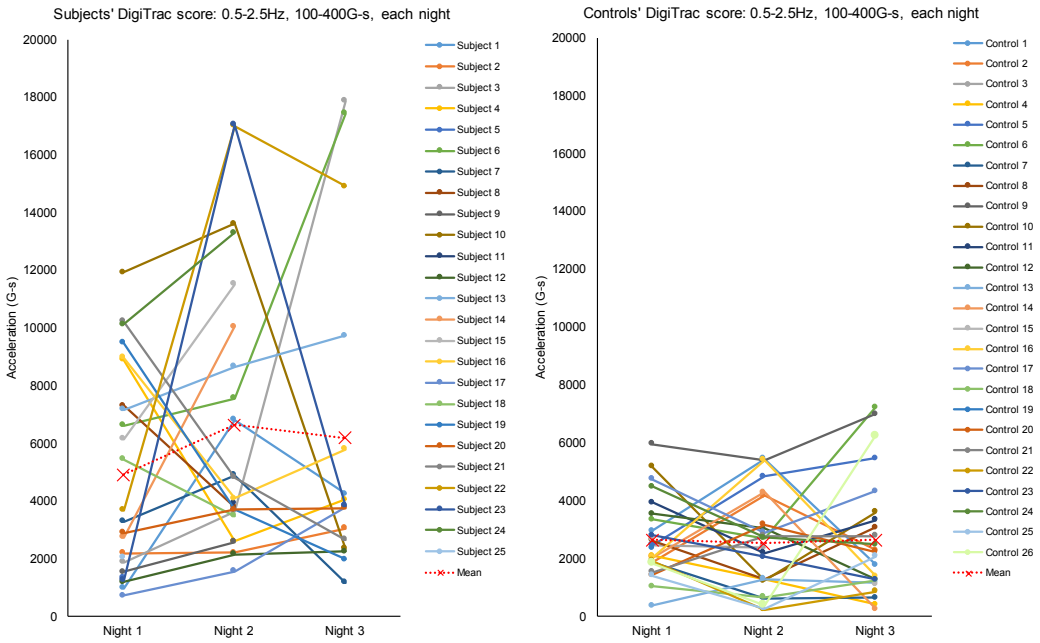
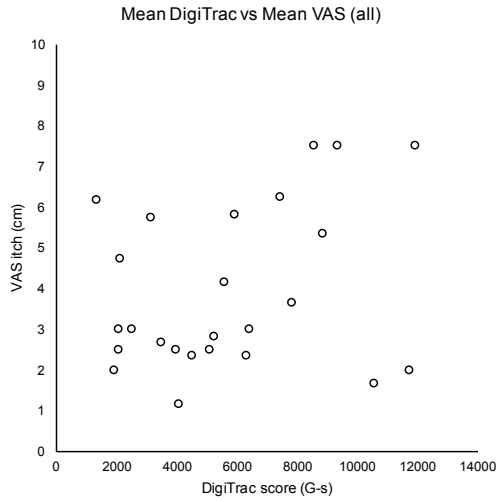
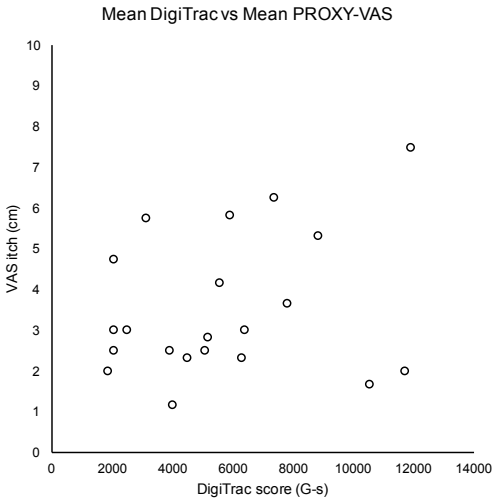


Figure 4.8. Scatterplot comparison of mean DigiTrac 0.5-2.5Hz 100-400G-s to mean subjective VAS scores, a) DigiTrac versus all VAS, correlation coefficient, rho= 0.289, b) DigiTrac versus proxy-VAS, correlation coefficient, rho= 0.253 and c) DigiTrac versus self-scored VAS, correlation coefficient, rho= 0.616.

a)



b)



c)

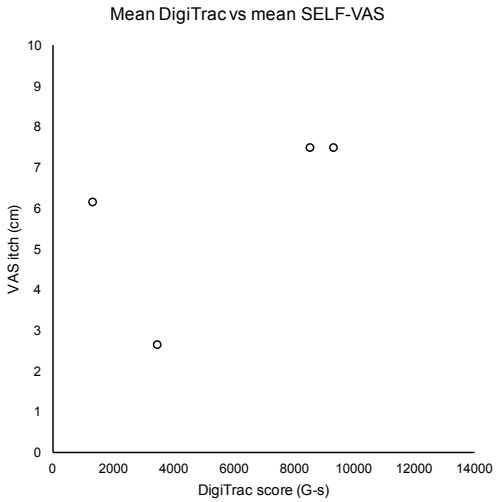
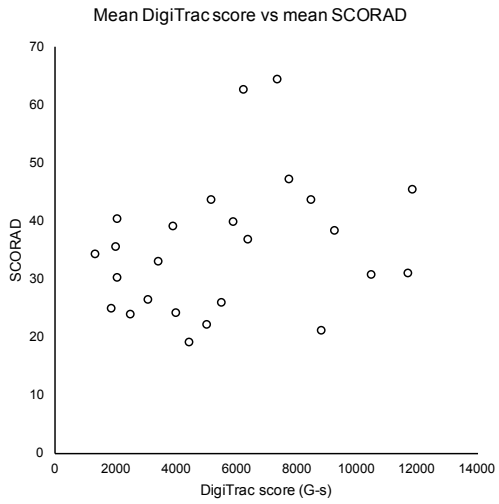


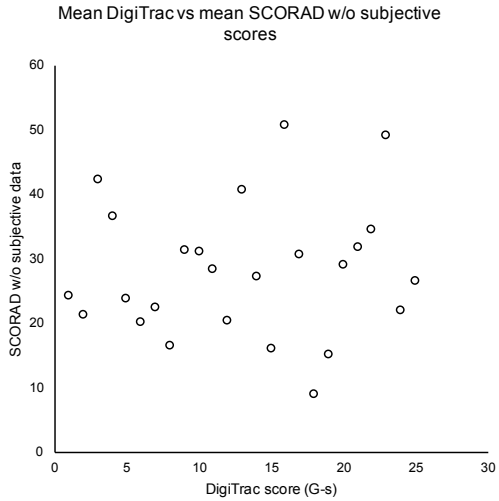
Figure 4.9 demonstrates the comparison of objective DigiTrac score and disease extent as measured by SCORAD: correlation coefficient, rho= 0.269. It can be seen that there is a poor relationship between the two. Since SCORAD involved combining VAS scores of itch and insomnia, and since it was thought that this may be a source of error (the majority of scores were made by proxy), the SCORAD without subjective data was also plotted and compared: correlation coefficient, rho=0.196. It can be seen that there is a lack of relationship between objective DigiTrac score and SCORAD whether this includes VAS scores or not.

Figure 4.9. Scatterplot comparison of mean DigiTrac 0.5-2.5Hz 100-400G-s score to mean disease extent measure, SCORAD a) including subjective scores, correlation coefficient, rho=0.269 and b) excluding subjective scores, correlation coefficient, rho=0.196.

a)



b)



Conclusion

This experiment had two objectives, firstly to extend the use of the DigiTrac, as a specific objective itch monitor into a larger study in order to answer some queries previously raised about variation in accelerometer score. The opportunity was also taken to see if the specific objective itch monitor scores correlated better than the non-specific movement monitors (Actiwatch Plus).

The results confirm postulations made on the basis of the previous, smaller study (Chapter 4.III), that DigiTrac could reliably separate itchy subjects from controls on the basis of the enriched data. Enrichment could be seen to improve separation stepwise through enrichment for 'frequency of action' and then for amplitude. The fact that the three nights' combined analysis demonstrates separation to the point of significance without so many levels of enrichment may allude to the fact that the separate night analysis is under-powered. The numbers of subjects and controls were made with the purpose of estimation of effect size rather than formal power calculations as this was an exploratory study.

There was large variation, within person, night-by-night in DigiTrac score whether the participant was an itchy subject or a control. It was suspected that some subjects may just score higher or lower than others, but the data does not suggest this: there was night to night variation in any direction. The range of scores reached much higher in the itchy subjects. The enrichment improves separation of itchy subjects from controls, and it would appear, as previously suggested in the pilot study, that this is in some ways because the enrichment allows better clustering of the controls whilst the massive range of itchy subjects' score remains untouched.

It is suspected that the age of subject may have implications on the accelerometer scores. As witnessed in infrared video studies and noticed by anecdote at home, small children generally move about more that teenagers and adults at night. Some condition could be added to the DigiTrac enrichment data to allow for this by a suitably qualified signal analyst. Unfortunately, this study did not afford the opportunity for this, but a later study described in Chapter 5.I did.

Overall, on the basis of this study, video-observation and previously published work by this unit, night-to-night variation in score would appear to be a true phenomenon.

Comparing objective to subjective measures was undertaken in order to see if the previously noted lack of correlation (Bringhurst et al., 2004) was to do with the fact that Actiwatch Plus accelerometers are non-specific movement monitors. DigiTrac-enriched data should provide a specific objective measure of itch to allow comparison with subjective and extent of disease data (SCORAD). However, no relationship was detected between objective DigiTrac score and VAS of itch. There was no correlation even when scores made by proxy were removed. This concurred with the Actiwatch Plus experiment findings. It would therefore appear that there is a true disconnect with subjective and objective itch scores.

No relationship was discovered, either, between the enriched DigiTrac score and SCORAD. SCORAD does involve and incorporate VAS scores of itch and insomnia and so, in an effort to correct for any bias afforded by this, the subjective scores were removed but correlation was still very poor. It is difficult to offer an explanation for why two seemingly objective scores should be divergent. It may be related to the previously mentioned variation in score: fleeting changes related to hotness, urticarial, for instance, may result in a very transient and abrupt itch response. This may account for the labile objective night-to-night DigiTrac score. DigiTrac responds to quick changes in condition. SCORAD is designed to assess the *chronic*

changes that skin makes in response to eczema such as lichenification. It may be because of this that the scores diverge.

It was at this point, that, having extended the use of accelerometers to my maximal ability (as a clinician and not a signal analyst) and frustrated by there seemingly being unanswered questions regarding subjective and objective disease measure divergence that the second part of this research was proposed.

Chapter 5

Measuring itch over time in chronic disease.

This chapter contains work published in a paper entitled “Are subjective accounts of itch to be relied on? The lack of relation between visual analogue itch scores and actigraphic measures of scratch.” It was written by Caroline S. Murray and Jonathan L. Rees. This was published in Acta Dermato-Venereologica in 2011, volume 91(1), pages 18 to 23 (Murray & Rees, 2011b). A copy of the paper is included in Appendix I. Other studies and data, not previously published are included too.

I: Longitudinal studies of objective and subjectively measured itch.

Aim

To use the Actiwatch Plus to characterise the features of night-to-night variation in objective itch score and to compare this to subjective measures of itch and disease extent.

Background

The experiments that led to this point had demonstrated:

a) Large night-to-night variation in objective accelerometer scores in small cluster studies.
b) That variation in objective score was a 'true' phenomenon, as verified by comparison to gold standard direct observation on infra-red video in consecutive night studies.
c) That accelerometers which had the potential to measure specific itch-related movement, (DigiTrac), as opposed to generalised movement (Actiwatch Plus) also demonstrated night-to-night variation.
d) A disconnect or lack of relationship between the verified objective accelerometer scores and subjective scores.

An objective score provides a logically attractive measure of disease: a 'true' measure which can be used to monitor disease or treatment, however, it misses a truly important dimension of disease, namely, how that individual actually experiences or perceives it. To provide an illustration, two individuals may have exactly the same disease extent as measured by PASI (Psoriasis Area Severity Index) but one, who enjoys swimming for instance, may regularly be placed in a situation where he/she is embarrassed by that disease and therefore notices more impact of the disease on their daily life than another individual who does not swim.

The experiments up to this point provided a source of confusion and concern: why were the objective and subjective not at all related? This experiment sought to further characterise the objective score night-to-night variation looking for patterns of scoring and, at the same time, to characterise night-to-night variation in subjective scores in the same way.

Methods

Participants were patients attending the Department of Dermatology in Edinburgh for treatment of itchy skin conditions. As such, as in previous experiments, these subjects represented a cohort with a more moderate to severe disease severity as they were attending secondary care clinics. After patients were approached and requested to consider taking part in the study, they were given written and oral information to consider. The patients were given time (at least 24 hours) to consider the invitation to participate in the study, and then contacted again by telephone. Those who agreed to participate then met the researcher at a mutually agreed appointment time and if they consented themselves or were consented by proxy the study went ahead

The studies have the use of Actiwatch Plus in common. Subjects were required to wear this for the prescribed number of nights of the study on the dominant wrist, or right wrist in a child who had not declared dominance, two hours before retiring to bed (reasons previously declared (Benjamin et al., 2004; Bringhurst et al., 2004)). The Actiwatch Plus has the facility of being able to alter recording epoch length. In order to be able to issue the accelerometer and for it to be set up to detect for the entirety of even the longest stint of the study (21 days), the epoch length was set to one minute. As explained in Chapter 2.I, lengthening the recording epoch does not reduce the sensitivity of the monitor as data is still sensed 32 times per second and the digital signal summed for whatever epoch is selected. For consistency's sake, the same epoch setting was used for all these studies.

Study 1: short cluster night-to-night study (up to 7 days).

Inclusion criteria

Subjects comprised 68 itchy adults and 50 itchy children. The control participants were 12 adults and 12 children. The subjects' diagnoses and characteristics are specified in Table 5.1. All the child-subjects had atopic eczema (as defined by Rajka and Hanifin criteria (Hanifin & Rajka, 1980)). The broad groups of diagnoses for adult-subjects were: eczema, psoriasis, cholestatic liver disease and pruritus of unknown cause (PUC). A total of 1654 nights were studied: 1573 subject-nights and 81 control-nights.

Subjects:

- Patients, children or adults, attending the department for treatment of itchy skin conditions.
- Age >3 months
- able to provide consent for self or parents able/willing to do so by proxy

Exclusion criteria

Subjects:

- Subjects under 3 months of age
- People not able to understand what is involved.
- People allergic to rubber (due to possible allergy to actiwatch).

Study method

Subjects were requested to wear the Actiwatch Plus (digital accelerometer) for three to seven nights. Participants were issued with written and oral instructions plus a questionnaire on which to note information such as any medications used, if they had got up overnight and subjective scores. Every day of the study, the participants were asked to complete a 10cm VAS of extent, itch and insomnia. In the case of children, adults (parent/guardian) completed the scores for the participants.

Study 2: longitudinal study for 42 nights

20 patients were recruited. Eleven females and nine men agreed to participate. The subjects' characteristics are listed in Table 5.1.

Inclusion criteria

Subjects:

- Adults, attending the department for treatment of atopic dermatitis (as defined by Hanifin and Rajka criteria (Hanifin & Rajka, 1980)
- Age 16 years
- able to provide consent.

Exclusion criteria

Subjects:

- Subjects under 16 years of age
- People not able to understand what is involved.
- People allergic to rubber (due to possible allergy to Actiwatch Plus).

Study method

The subjects were issued with a digital accelerometer and instructed to wear it every night for the next 42 nights. A total of 761 nights were studied (some participants forgot to apply the accelerometer every night). All subjects were issued with written and oral instructions plus a questionnaire to list information pertinent to the study (medications used, nocturnal ambulatory movement).

The participants were issued with dated sheets so that every day they should complete VAS scores of itch, disease extent and insomnia plus a 'ballot box' to post each day's scoring sheet into immediately after completing it. This was to enable, as far as possible, 'blinding'

to the previous day's VAS score and to reduce anchoring bias as much as possible. The sheet for Day 0 was completed and deposition demonstrated to the researcher by the participant.

On (or as near to) Day 21 of the study, the participant and researcher met so that the accelerometer could be downloaded and checked for full functionality. On (or as near to) Day 42 of the study the researcher met with the participant to collect and download the digital accelerometer and to collect the completed forms – all, it was expected, to be contained in the ballot box.

As well as the daily VAS, a measure of disease extent, SCORAD (Severity Scoring of Atopic Dermatitis) was completed by the researcher at the beginning, at the mid-point and at the end of the study. Subjective scores SF-36 (Ware & Sherbourne, 1992) and an adapted Day Reconstruction Method (DRM) (Kahneman, 2004), in which GHQ (Global Health Question) (Bjorner, 2005) was embedded and into which questions directly relating to skin disease and itch were added, were completed at the beginning, mid-point and the end of the study. See Figure 5.1 for a flowchart of the study. For descriptions of and the rationale behind choosing these subjective measures, please see Introduction, section VII.

Use of the DRM questionnaire was exploratory. The adapted DRM includes a 7-point Likert scale of itch and allows analysis of amount of time (in minutes) experiencing itch and other symptoms and emotions. A vast array of data was gleaned: information about every hour of the previous day on three occasions. Some of the data simply not relevant in this context, for instance, happiness whilst child-caring, cleaning, etcetera was not analysed. The adapted DRM is attached in Appendix III.

Subject characteristics are shown in Table 5.1 and overall completion data is shown in Table 5.2.

Figure 5.1. *Study 2:* flowchart. VAS=visual analogue scale, DLQI=Dermatology Life Quality Index, SCORAD=Severity Scoring of Atopic Dermatitis.

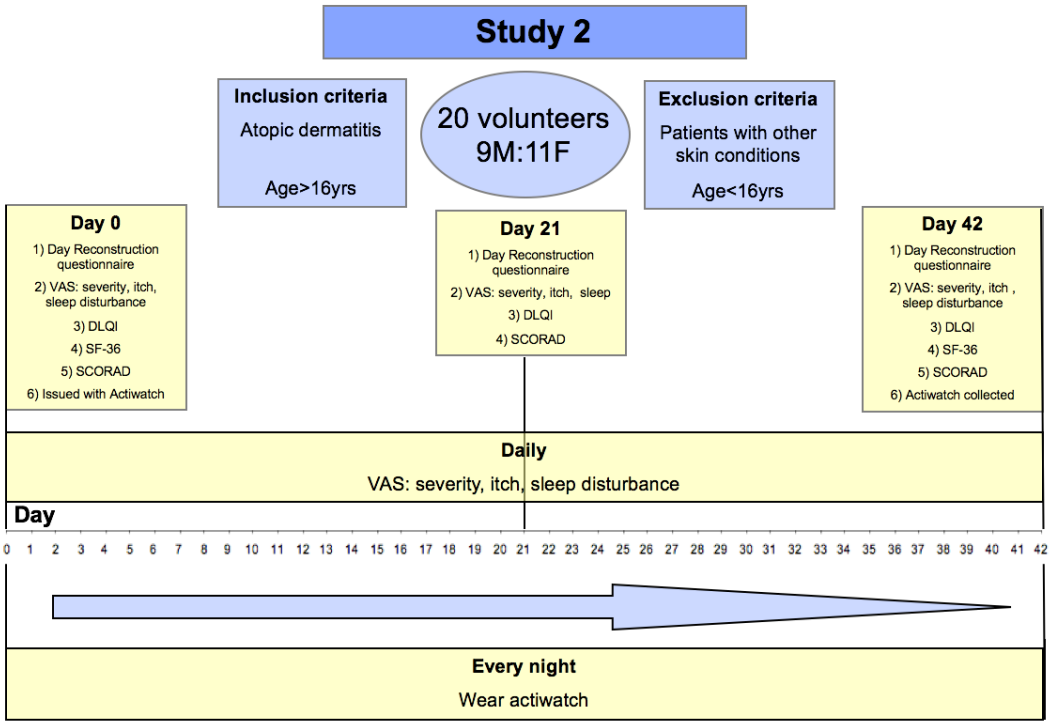


Table 5.1. Principal demographic characteristics of all study subjects (Eczema = atopic dermatitis, PUC=pruritus of unknown cause, Liver = hepatic disease associated itch. F=female, M=male. Children are <16 years; adults >16 years of age.

	Number	M:F ratio	Mean age (yrs)	Age range (yrs)	Nights studied	Eczema	Psoriasis	PUC	Liver
Child subjects	50	26:24	6.92	2-15	229	50	0	0	0
Child controls	12	2:10	10.33	6-14	42	n/a	n/a	n/a	n/a
Adult subjects	6	27:40	43.5	16-78	1344	30	25	4	8
Adult controls	13	5:8	38.23	24-71	39	n/a	n/a	n/a	n/a
Adult subjects	20	11:9	40.55	16-67	761	20	0	0	0

Table 5.2. *Study 2:* Completion data

Subject code	Completed VAS (days)	Total study days	Completion % VAS	Completed Actiwatch (nights)	Completion % Actiwatch
1	37	44	84.1	36	83.7
2	44	46	95.7	38	84.4
3	43	43	100	37	88.1
4	42	43	97.7	33	78.6
5	42	43	97.7	16	38.1
6	42	43	97.7	38	90.5
7	40	42	95.2	35	85.4
8	43	43	100	22	52.4
9	43	43	100	19	90.5
10	43	43	100	38	90.5
11	43	43	100	17	40.5
12	43	43	100	32	76.2
13	42	43	97.7	33	78.6
14	40	44	90.9	19	44.2
15	43	43	100	40	95.2
16	43	43	100	25	59.5
17	39	43	90.7	21	50
18	43	43	100	29	69
19	39	43	90.7	22	52.4
20	39	43	90.7	39	92.9

Analysis

This was an exploratory study and so formal power calculations were not made. The numbers of subjects recruited were decided with the purpose of estimation of effect size in mind and were opportunistic. The approach was to report p values or confidence limits on a pragmatic sample group based on previous experiments, availability and costs. Most of the experiments 'nested' previous conclusions in each successive experiment, implicitly testing the conclusions of previous analyses.

Actiwatch Plus data was downloaded via the proprietary reader into Excel (Microsoft Ltd, Seattle, USA) and managed in this software. The subjective and disease extent was also entered into and managed using Excel. Statistics were analysed in 'R' v2.9 or SPSS v21 (IBM) running on a Mac (OS 10).

In order to differentiate (in Study 1) between adults completing their own subjective symptom-scores and child-subjects, who had required an adult to complete them for them, analyses which involved examination of VAS itch scores were conducted separately for those under and over 16 years of age. Actiwatch Plus scores are non-normally distributed and were usefully log transformed. Following inspection using univariate analyses, non-significant terms were removed from the regressions, and examination of factors were performed using Holm's correction for pairwise t-tests. Holm's correction is required when it is necessary to counteract problems associated with multiple comparisons and issue also known as the Family-wise error rate (FWER). FWER is defined as, "the probability of making one or more false discoveries, or type I errors, among all the hypotheses when performing multiple hypotheses tests." It is similar to the Bonferroni method but more powerful and most appropriate in this particular circumstance. Within and between person variance was examined using analysis of variance.

Longitudinal scoring patterns of objective accelerometer measured itch and subjective VAS score in Study 2 were analysed for autocorrelation. This type of analysis is one more commonly used to predict weather or financial fluctuations (Kravchenko, Grechany, & Gadjiev, 2006; Lewellen, 2002). The subjective data was normally distributed and so parametric analyses were employed for comparisons within this type including Student t test and Pearson correlation coefficients.

Results

Study 1: Characterising determinants of actigraphy scores using short cluster night-to-night study (up to 7 days).

The majority of subjects completed three or more nights of the study. Just as in previous experiments, a large variation in Actiwatch Plus score was discovered. Slightly more variation in objective score, 56%, was found between person, and slightly less, 44%, existed within person: an individual's scores, overall varied by 44% night-to-night.

Subsequent analyses used mean Actiwatch Plus scores. Children (<16 years) and adults (>16 years) were examined separately, as mentioned previously to avoid the potential bias from subjective data being completed by proxy. The total data set results are shown in Figure 5.2. It can be seen that itchy subjects have higher Actiwatch Plus scores than controls and that those with liver disease associated itch appear to be scratching the most. There is much overlap, but second most itchy are those with atopic dermatitis (more subjects) and pruritus of unknown cause (very few subjects).

Children

Linear regression of log. actigraphy scores showed no effect of age (p=0.67) nor sex (0.365). Actiwatch Plus scores were, as expected, higher for atopic subjects than controls: mean (log)

actigraph score was 9.02 for atopics versus 8.29 for controls (p=0.021). VAS itch was not a significant predictor in either univariate (p=0.261) or multivariate (p=0.284) models.
The relation between Actiwatch Plus score and VAS itch score is demonstrated in Figure 5.3. A line of best-fit has been added but this does not significantly differ from zero. This finding concurs with previous experiments illustrating the disconnect between subjective and objective score.

Adults

Linear regression of log. actigraphy scores showed no effect of age nor sex. Mean log Actiwatch Plus score analysis showed a stepwise reduction in mean score through 10.4 for those with liver disease, 8.95 for those with atopic dermatitis, 8.99 for subjects with pruritus of unknown cause, 8.43 for psoriasis patients and 8.14 for controls. Pairwise t test are shown in Table 5.3, and it is clear that many of the comparisons were significant, notably for eczema patients compared to all other groups except pruritus of unknown cause.

Mean VAS itch scores were 6.64, 4.83, 4.68 and 2.53 in patients with liver disease, psoriasis, atopic dermatitis and patients with itch of unknown cause. The only significant difference between these groups was between those with liver disease and those with itch of unknown cause (p=0.04, pairwise t test): the spread of values within each group was large.

Univariate analyses demonstrated that both diagnosis (p<0.001) and VAS itch (p=0.042) were significant predictors of actigraphy scores but the R square value for VAS itch was small at 0.06. The R square for diagnostic category was more convincing at 0.51. Once diagnosis was entered into the regression, VAS itch was no longer significant (p=0.27). The relation by diagnostic group between VAS itch score and Actiwatch Plus score is shown in Figure 5.4. Lines of best-fit have been added but do not differ from zero. Thus, the disconnect between subjective score and objective is captured for adults too in this study.

Figure 5.2. *Study 1:* Mean overnight actigraphy score for all participants (children and adults) separated by diagnosis: control (n=32), liver a.k.a hepatic itch (n=8), unknown a.k.a. pruritus of unknown cause (n=4) , psoriasis (n=25) and atopic derm a.k.a eczema (n=80). Actiwatch score units: 10=1G.

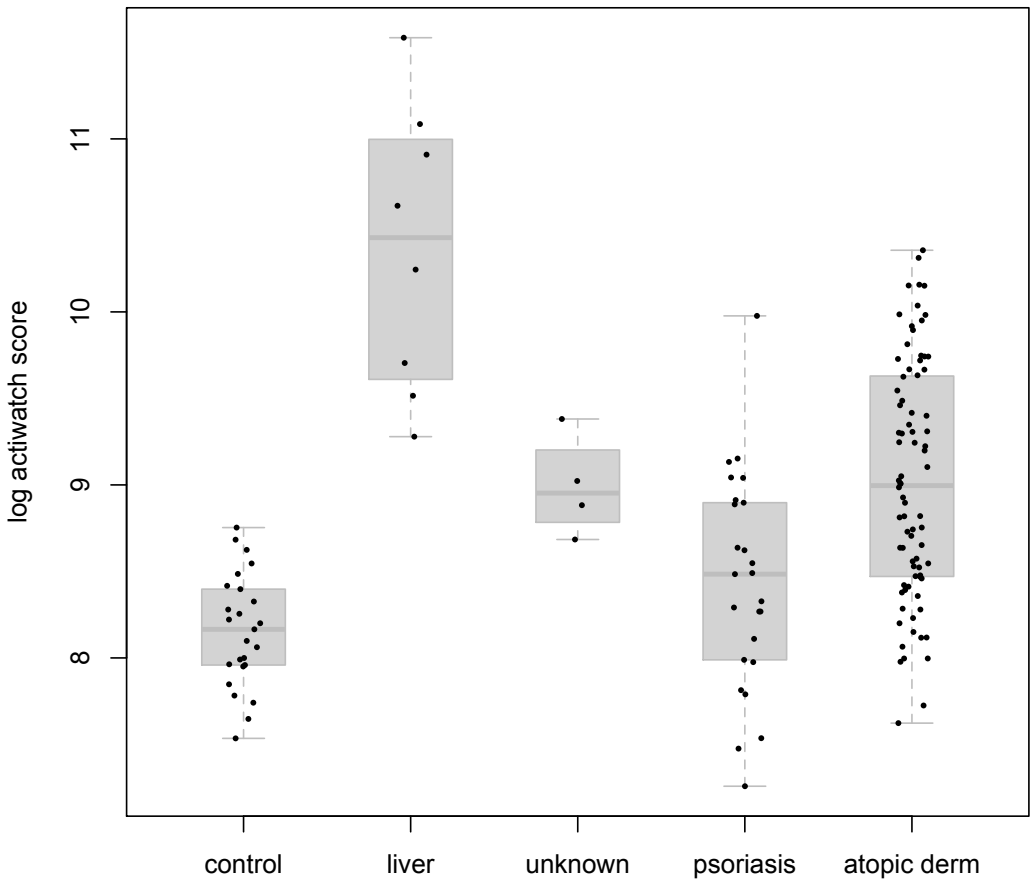


Figure 5.3. *Study 1*: Regression analysis of children's actigraph (mean log overnight) score on VAS score of itch. Line= 'best fit' regression line (not significantly different from 0). Actiwatch score units: 10=1G.

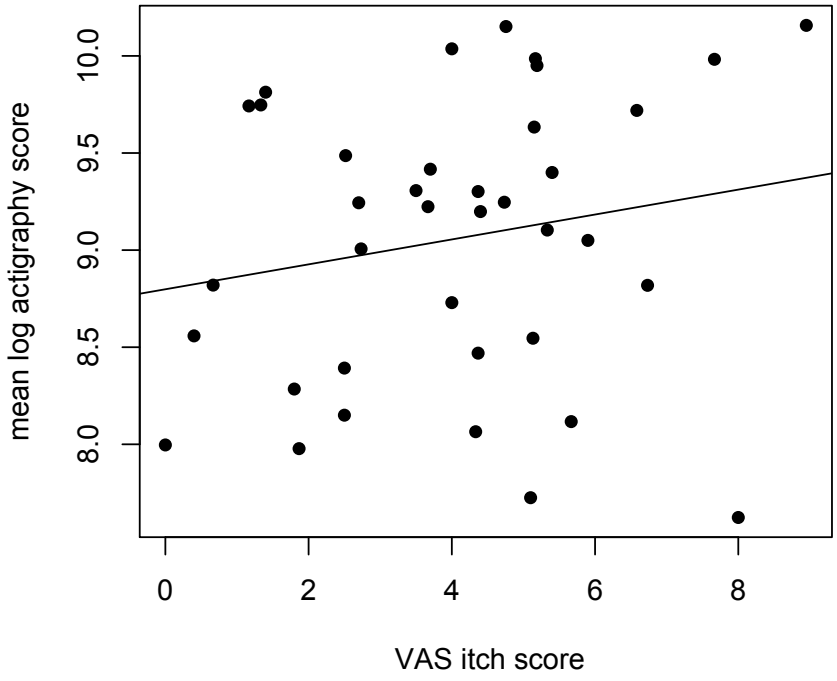
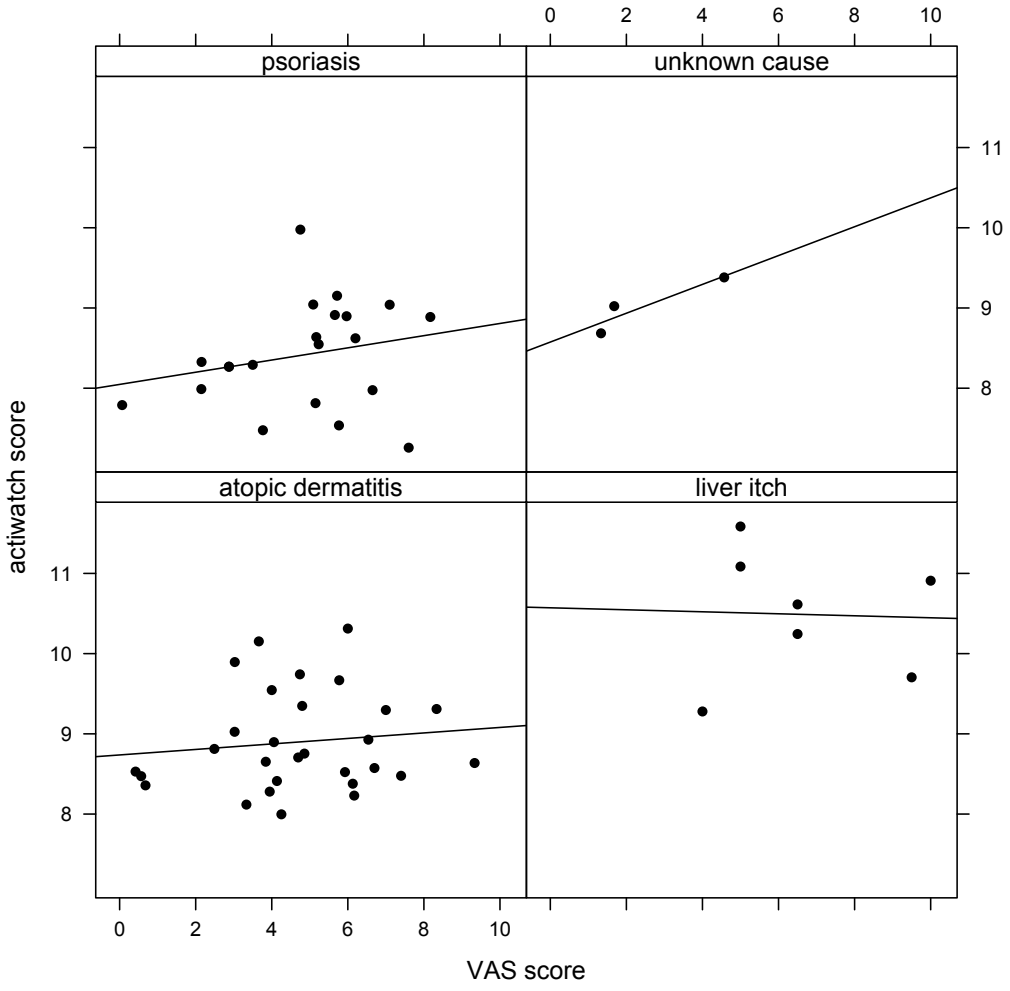


Table 5.3. *Study 1:* Summary of principal differences between Actiwatch Plus scores for adults by diagnostic group. Pairwise comparisons using t-tests with pooled error. Eczema = atopic dermatitis, Liver hepatic disease associated itch, PUC=pruritus of unknown cause.

	Eczema	Control	Liver	Psoriasis
Control	0.002	-	-	-
Liver	<0.001	<0.001	-	-
Psoriasis	0.018	0.356	<0.001	-
PUC	0.890	0.079	0.003	0.300

Figure 5.4. *Study 1:* Regression analysis of adults' Actiwatch Plus (mean log overnight) score on VAS itch separated by diagnostic group, see: top left for psoriasis, top right for pruritus of unknown cause (=PUC), bottom left for atopic dermatitis and bottom right for liver itch (=hepatic itch). 'Best fit' regression lines shown (none significantly different from 0). Actiwatch Plus score units: 10=1G.



Study 2: longitudinal study for 42 nights

This experiment involved a different cohort of subjects from those recruited for Study 1: 20 adults with eczema.

a) Characterising longitudinal objective and subjective scoring patterns using prospective 42 night study: comparison of daily objective Actiwatch Plus score with daily VAS itch score

Nocturnal Actiwatch Plus score and VAS itch scores were recorded over a 42-day study period. Little relation was found between the Actiwatch Plus score and VAS itch. This finding is illustrated in Figure 5.5. Lines of best fit have not been added but the distribution of the data points clearly demonstrates no relationship between these two measures. Spearman's rho correlation coefficient for each subject is listed in Table 5.4.

The data in Table 5.4 needs to be considered cautiously as night-to-night for the same person cannot be assumed to be independent and so it is not meaningful to report correlations for each person. A different approach was therefore taken to the data: could forecasting analysis detect a pattern of scoring in either the objective or subjective data? The empirical autocorrelation structure of the VAS itch and actigraphy scores was examined (Figures 5.6 and 5.7). A line representing p=0.001 significance is plotted. It can be seen that the pattern for actigraphy and for VAS itch is very different. For VAS itch, there is a clear pattern of autocorrelation most strong at a lag of one night, but visible up to five nights. For actigraphy there is no good evidence of any lag effect. Time series modelling using moving averages improved the fit of the model for the VAS itch data.

Figure 5.5. *Study 2*: each subject's night-to-night objective log. Actiwatch Plus score (10=1G) compared with corresponding day's VAS score (cm)

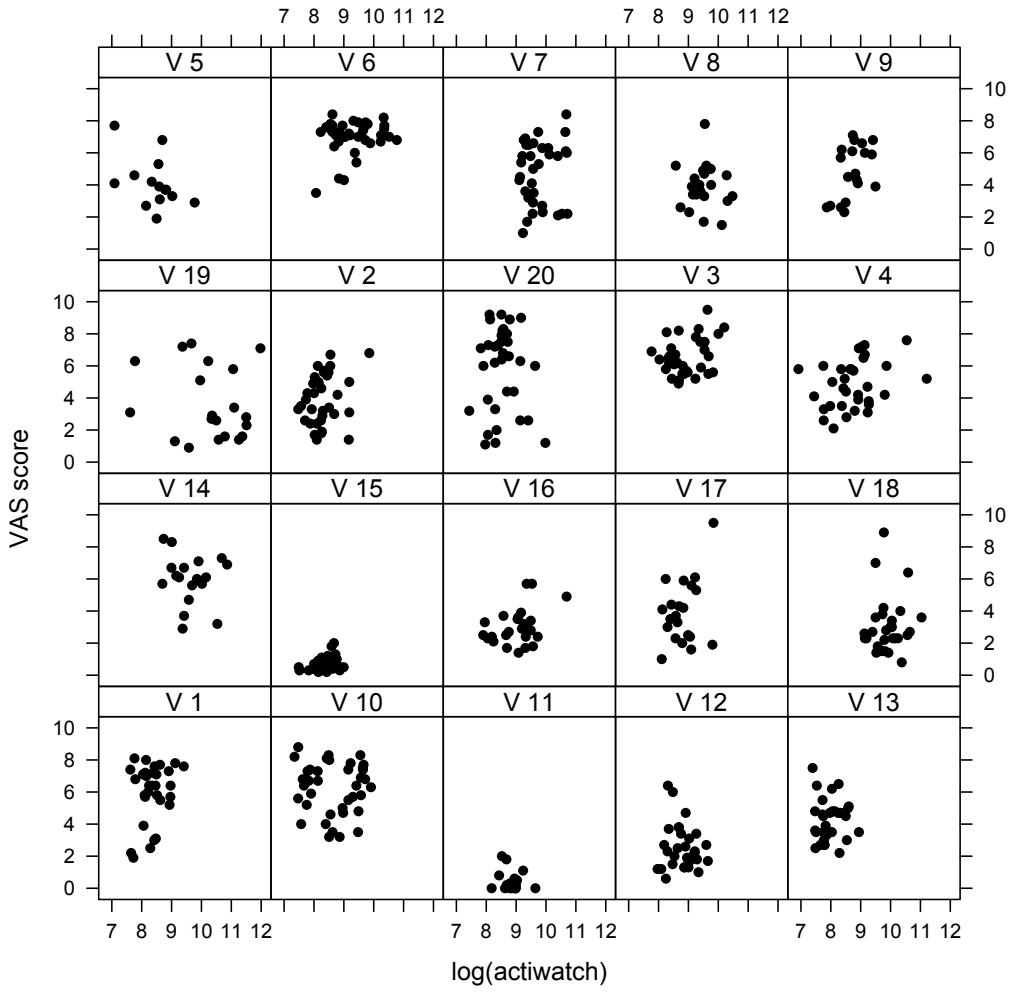


Table 5.4. *Study 2:* Spearman's rho correlation coefficients for Actiwatch Plus score versus VAS of itch.

Subject	Spearman's rho	Significance	Number of pairs
1	0.119	0.523	36
2	0.264	0.125	37
3	0.146	0.040	35
4	0.287	0.105	33
5	-0.377	0.184	14
6	0.009	0.959	39
7	0.091	0.605	35
8	0.500	0.860	22
9	0.444	0.057	19
10	-0.010	0.053	38
11	-0.164	0.528	17
12	-0.063	0.734	32
13	0.081	0.657	33
14	-0.114	0.653	19
15	0.263	0.101	40
16	0.241	0.245	25
17	0.214	0.351	21
18	0.111	0.565	29
19	-0.155	0.514	20
20	0.096	0.576	36

Figure 5.6. *Study 2:* Graph demonstrating autocorrelation structure of massed individuals' mean log overnight Actiwatch Plus scores. Dashed line represents p=0.001 significance (autocorrelation above or below line suggests significance).

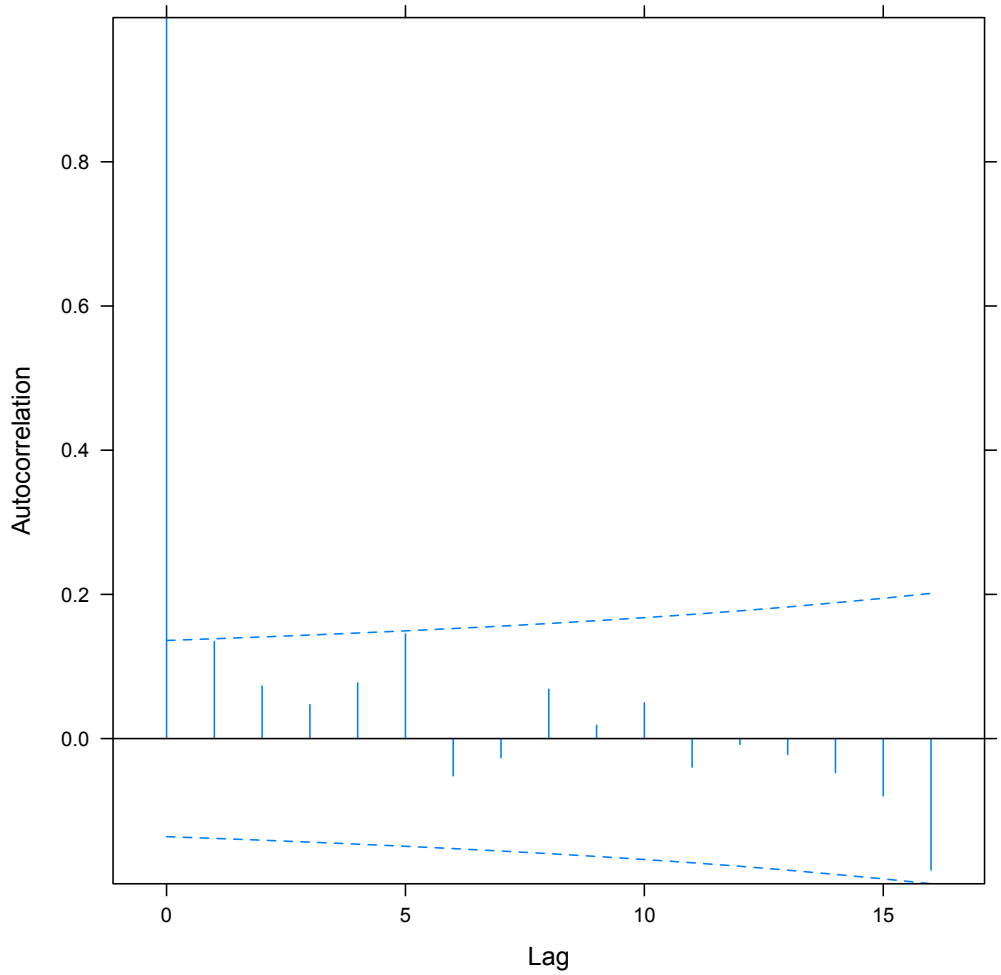
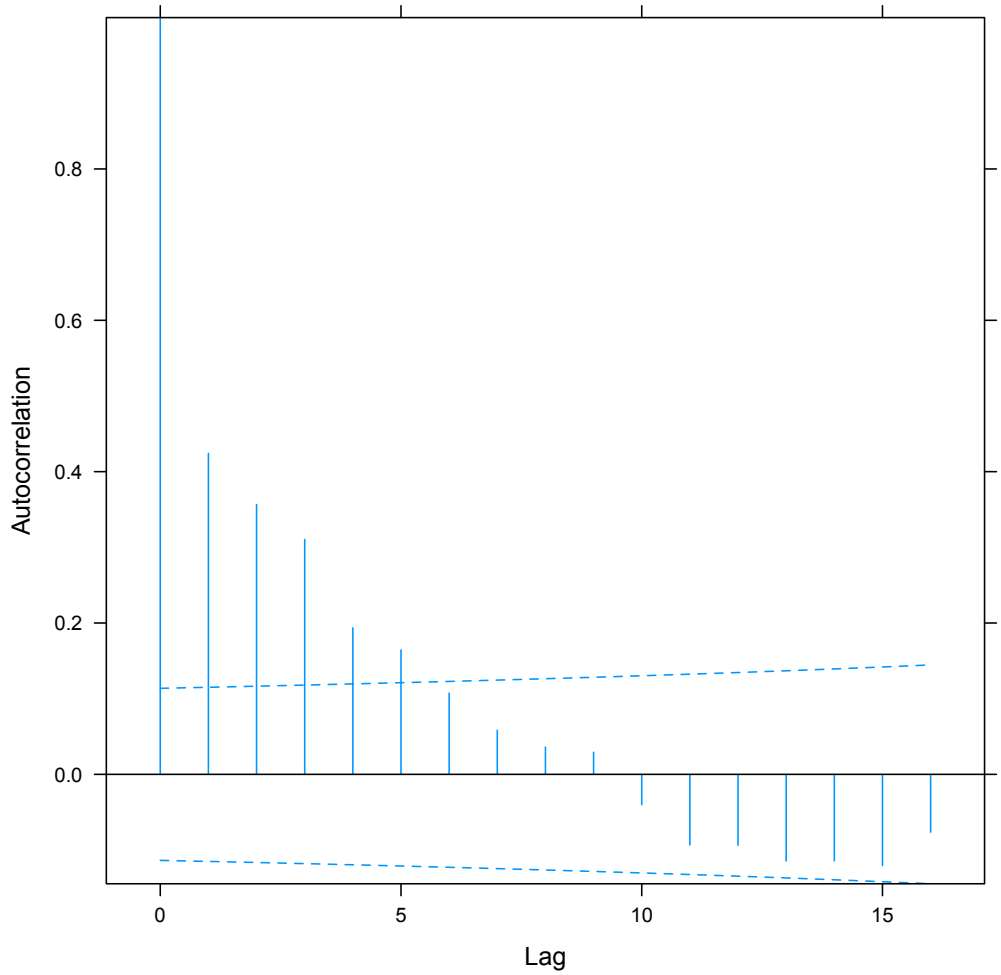


Figure 5.7. *Study 2*: Graph demonstrating autocorrelation structure of massed individuals' VAS itch scores. Dashed line represents p=0.001 significance (autocorrelation above or below line suggests significance).



b) Comparison of objective and subjective measures of disease.

All subjects bar one (who missed the middle assessment) completed the validated subjective symptom questionnaires DLQI (Dermatology Life Quality Index) and SF-36 (otherwise known as RAND-36) at the beginning, middle and end of the 42-day study period (as well as undertaking daily VAS scores). The researcher also examined the subjects to complete the disease extent measure, SCORAD, at the same time points. As an exploratory venture, the Day Reconstruction Method (DRM) was completed by the subject at the beginning, middle and end of the study too. The subjective and extent measures were compared to each other as well as to the objective Actiwatch Plus score.

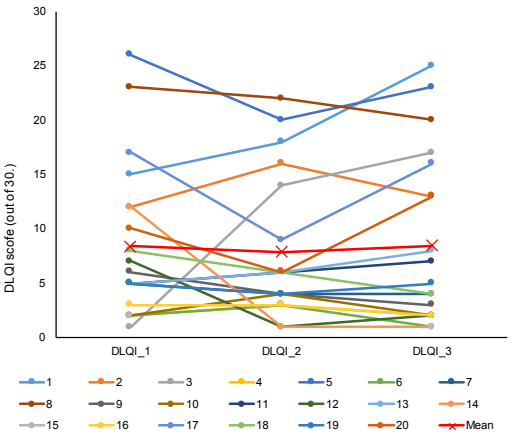
There was no significant difference in DLQI, SF-36 or SCORAD over the 42-night period (as analysed by ANOVA). It had been predicted that, since the DRM questionnaire took up to two and a half hours to complete, the number of episodes would decrease each time the participant had to repeat the process, but this was not the case (ANOVA confirms lack of difference). Table 5.5 summarises the DLQI, SF-36 and SCORAD score characteristics. Figure 5.8 summarises the DLQI, SF-36 and SCORAD scores plus the number of DRM episodes.

Table 5.5. *Study 2:* Day 1, Day 21 and Day 42 DLQI, SF-36 and SCORAD mean raw scores (n=20).

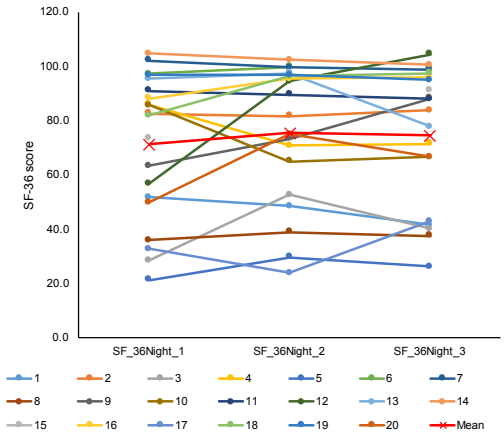
	Mean	Median	St dev	Overall Mean	Overall range
SF-36 Day 1	71.203	82.229	26.614	73.578	21 to 104
SF-36 Day 21	75.294	81.646	25.488		
SF-36 Day 42	74.363	83.750	25.147		
DLQI Day 1	8.400	5.500	7.148	8.254	1 to 26
DLQI Day 2	7.895	6.000	6.641		
DLQI Day 3	8.450	4.500	7.970		
SCORAD Day 1	22.075	14.500	15.240	19.517	1.5 to 54
SCORAD Day 21	17.895	14.500	13.915		
SCORAD Day 42	18.500	13.500	15.628		

Figure 5.8. *Study 2:* Graphs to show how scores change over the length of the study: assessments on Day 1 (_1), Day 21 (_2) and Day 21 (_3). Raw scores for a) DLQI, b) SF-36, c) SCORAD raw scores and d) number of DRM episodes

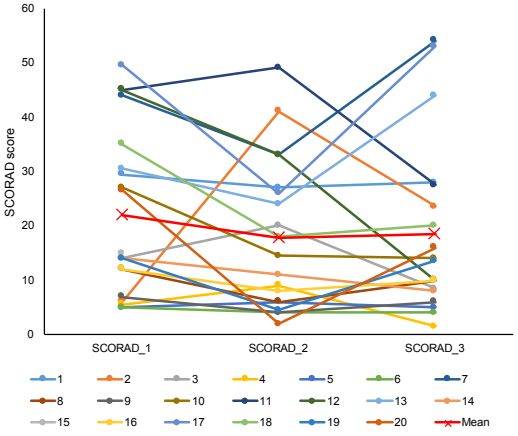
a)



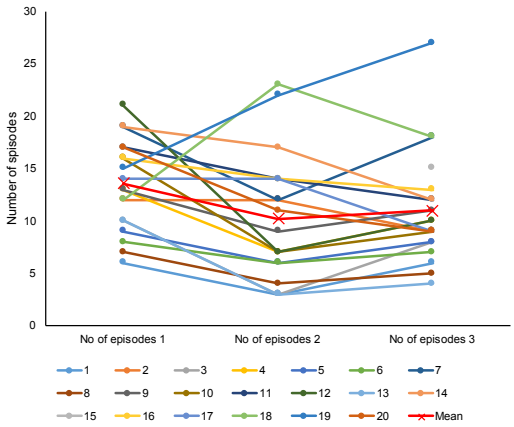
b)



c)



d)



i) Correlation data

An expected *negative* correlation was found between the extensively validated DLQI and SF-36 raw scores (Pearson's correlation coefficient=-0.814, $p<0.001$). The relationship was expected to be negative since a higher DLQI score indicates poor health whilst a lower SF-36 indicates poor health (Ware & Sherbourne, 1992). There was a strong *positive* relationship between DLQI and DRM itch raw scores (Pearson's correlation coefficient=0.786, $p<0.001$). These scores run in the same direction with regard to severity and so a positive correlation was expected. DRM itch was, strongly and negatively related to SF-36 (Pearson's correlation coefficient=-0.677, p=0.001), again, correlation ran in the expected direction. There relationship between SCORAD and all other measures, subjective and objective is poor - none are significant, see Table 5.6.

The DRM itch data shows a poor relationship with DRM raw scores for positivity (Pearson's correlation coefficient=-0.004, p=0.986) and happiness (Pearson's correlation coefficient=-0.306, p=0.189). There is a significant positive correlation with DRM itch and raw scores for negativity (Pearson's correlation coefficient=0.505, p=0.023), sadness (Pearson's correlation coefficient=0.659, p=0.002) and strongest and perhaps most understandably with being uncomfortable (Pearson's correlation coefficient=0.952, $p<0.001$).

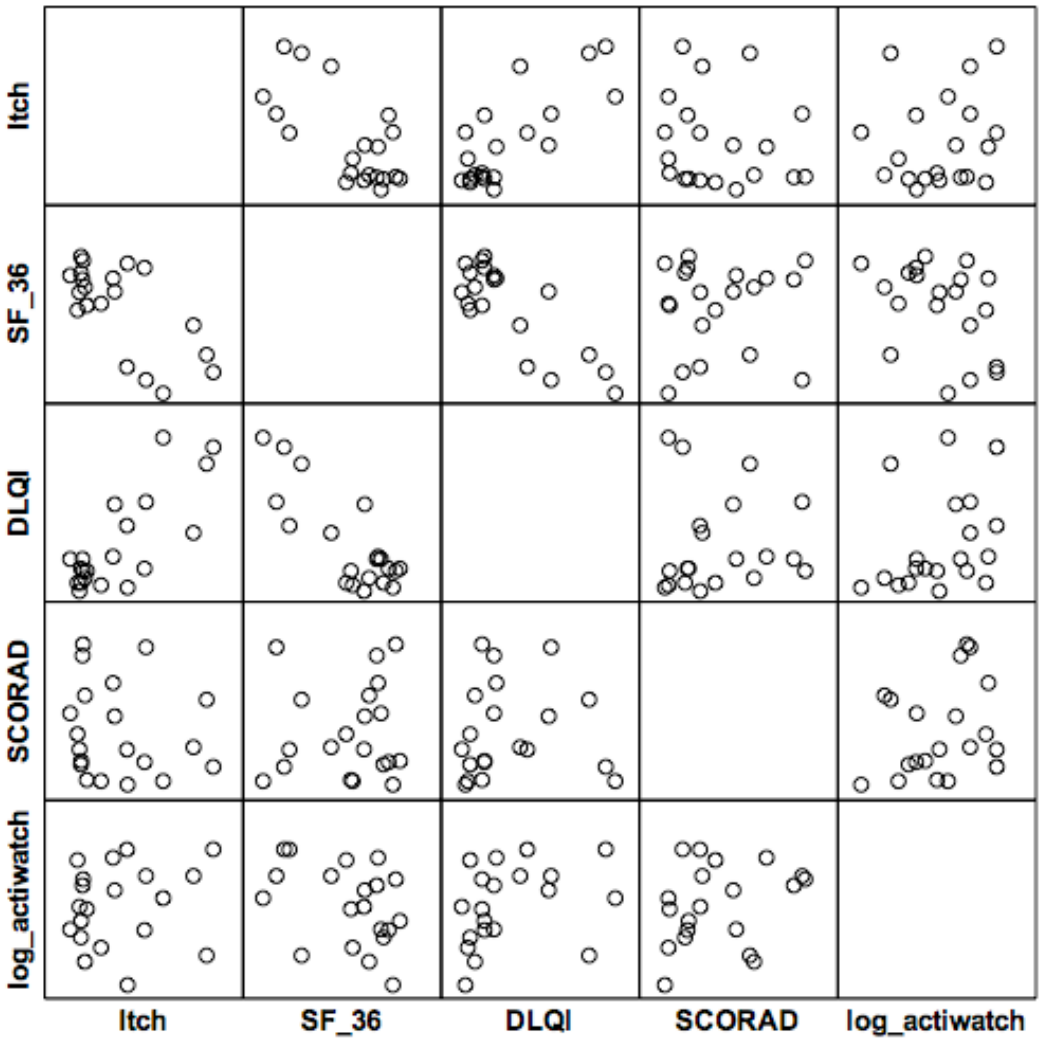
There is an expected negative directional correlation between Log.actiwatch score and SF-36 (Pearson's correlation coefficient = -0.382, p=0.082), but the relationship is weak. The relationship between Log.actiwatch and SCORAD and DLQI is also poor: Pearson's correlation coefficient=0.275, p=0.240 and 0.327, p=0.160 respectively. Log.actiwatch score is no better related to DRM itch raw scores (Pearson's correlation coefficient=0.153, p=0.519).

Data is summarised in Table 5.6 and in the scatterplot in Figure 5.9.

Table 5.6. *Study 2:* Comparison of subjective, disease extent and Actiwatch Plus score on three occasions over 42 days. Top number in cell is Pearson's correlation coefficient and the lower one the p-value (t test). Subjects, n=20.

	Positive	Negative	Happy	Sad	**Itch**	Uncomfortable	SF36	DLQI	SCORAD	log.actiwatch
Positive	n/a	0.235 p=0.319	0.503 p=0.024	0.243 p=0.302	**-0.004** **p=0.966**	0.091 p=0.703	n/a	n/a	n/a	n/a
Negative	0.235 p=0.319	n/a	-0.329 p=0.157	0.889 p<0.001	**0.505** **p=0.023**	0.652 p=0.002	-0.518 p=0.019	0.535 p=0.015	-0.18 p=0.448	n/a
Happy	0.503 p=0.024	-0.329 p=0.157	n/a	-0.324 p<0.001	**-0.306** **p=0.189**	-0.292 p=0.212	0.363 p=0.116	n/a	n/a	n/a
Sad	0.243 p=0.302	0.889 p<0.001	-0.324 p=0.163	n/a	**0.559** **p=0.002**	0.719 p<0.001	-0.713 p<0.001	0.629 p=0.003	0.07 0.769	n/a
Itch	**-0.004** **p=0.966**	**0.505** **p=0.023**	**-0.306** **p=0.189**	**0.559** **p=0.002**	**n/a**	**0.952** **p<0.001**	**-0.677** **p=0.001**	**0.786** **p<0.001**	**-0.143** **p=0.546**	**0.153** **p=0.519**
Uncomfortable	0.091 p=0.703	0.652 p=0.002	-0.292 p=0.212	0.719 p<0.001	**0.952** **p<0.001**	n/a	-0.745 p<0.001	0.872 p<0.001	-0.174 p=0.463	0.221 p=0.329
SF36	n/a	-0.518 p=0.019	0.363 p=0.116	-0.713 p<0.001	**-0.677** **p=0.001**	-0.745 p<0.001	n/a	-0.814 p<0.001	0.071 p=0.767	-0.393 p=0.082
DLQI	n/a	0.535 p=0.015	n/a	0.629 p=0.003	**0.786** **p<0.001**	0.872 p<0.001	-0.814 p<0.001	n/a	0.028 p=0.902	0.327 p=0.160
SCORAD	n/a	-0.18 p=0.448	n/a	0.070 p=0.769	**-0.143** **p=0.546**	-0.174 p=0.463	0.071 p=0.767	0.029 p=0.902	n/a	0.275 p=0.240
log.actiwatch	n/a	n/a	n/a	n/a	**0.153** **p=0.519**	0.221 p=0.329	-0.393 p=0.082	0.327 p=0.160	0.275 p=0.240	n/a

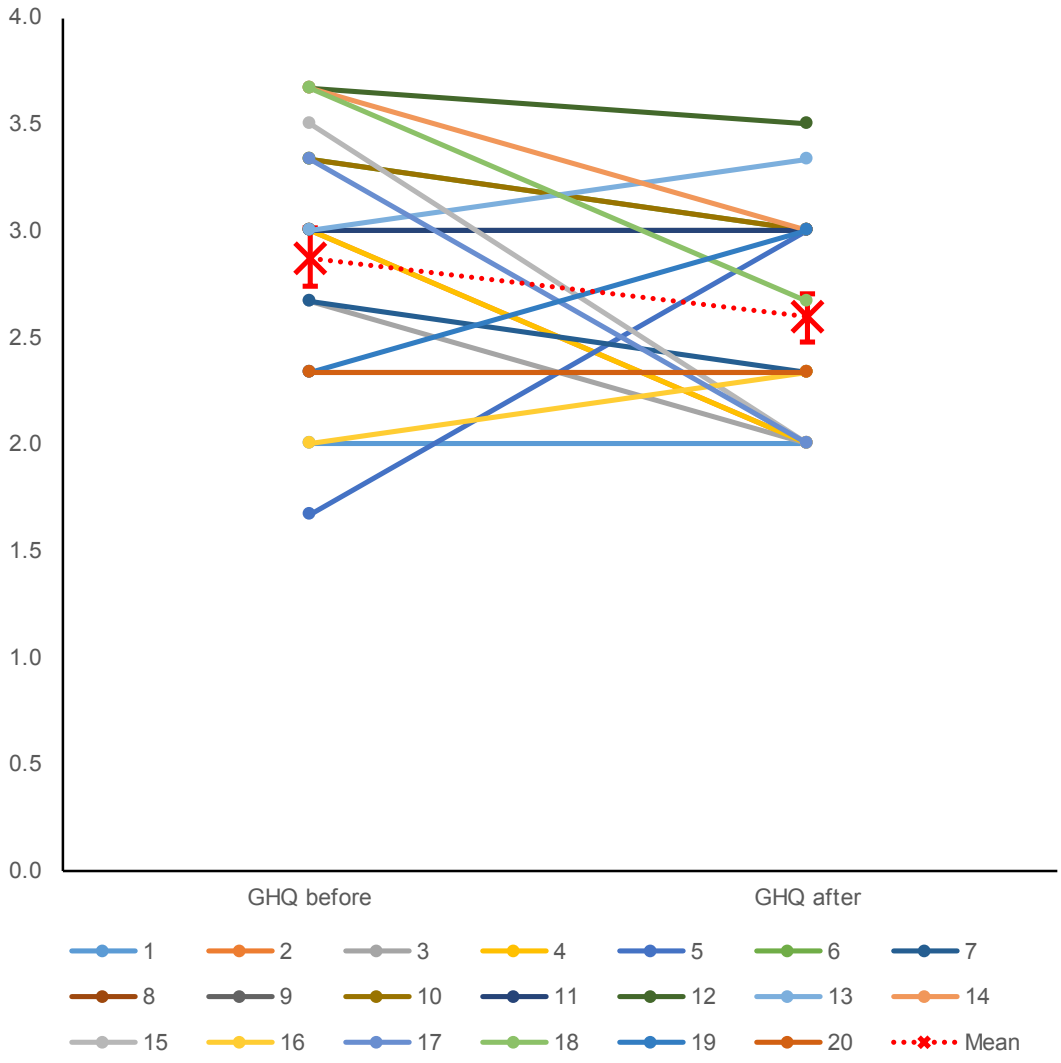
Figure 5.9. *Study 2*: Comparison of subjective, disease extent and Actiwatch Plus score on three occasions over 42 days (n=20).



ii) Global Health Question.

This question was embedded in DRM twice: first at the beginning and then at the end. The reason for the two placements was to see if focussing the mind acutely by undertaking the DRM and minimising recall bias made any difference to the score. The mean score (each subject over three occasions) before and after DRM does not appear markedly different graphically, see Figure 5.10, but the variable is ordinal. The mean before is 2.88 (SEM ±0.135) and after 2.59 (SEM ±0.113) and p=0.083 (paired t test). A lower score means a poorer perception of health.

Figure 5.10. *Study 2:* Line chart to demonstrate Global Health Question response before and after Day Reconstruction Method questionnaire. Mean GHQ before and after are compared within person. Red cross and dashed line represents mean and SEM error bars.



iii) Day Reconstruction Method measured itch: in and out of work.

It is stated that being awake or being distracted reduces the perception of itch. The DRM questionnaire afforded the possibility of testing this theory. The questionnaire was adapted to include questions about how itchy the subject felt whilst they were inside or out of work. The results confirmed the effect: the mean percentage time (minutes) subjects felt incredibly itchy outside and inside work respectively were 12.967mins(SE±5.188) and 6.333mins (SE±1.992), p=0.487, the mean time subjects felt quite itchy outside or in work respectively were 29.608mins (SE±4.290) and 27.578mins (SE±5.913), p=0.541, the mean time respondents were hardly itchy outside or in work were 29.267mins (SE±3.991) and 26.722mins(SE±3.735), p=0.574 respectively and finally, the means for not feeling itchy at all when outside and in work were 26.508mins (SE±5.769) and 38.944mins (SE±7.154), p=0.027. The trends suggests that distraction at work reduces the perception of itch but the only significant difference in means is in those not feeling itchy at all: as expected more people felt least itchy in work rather than out of work. The results are summarised in Figure 5.11 and Table 5.7

Figure 5.11. *Study 2:* Bar chart to show reported DRM itch outside (plain white bars) and inside work hours (patterned bars). Units of itch: mean percentage time (minutes) felt itchy.

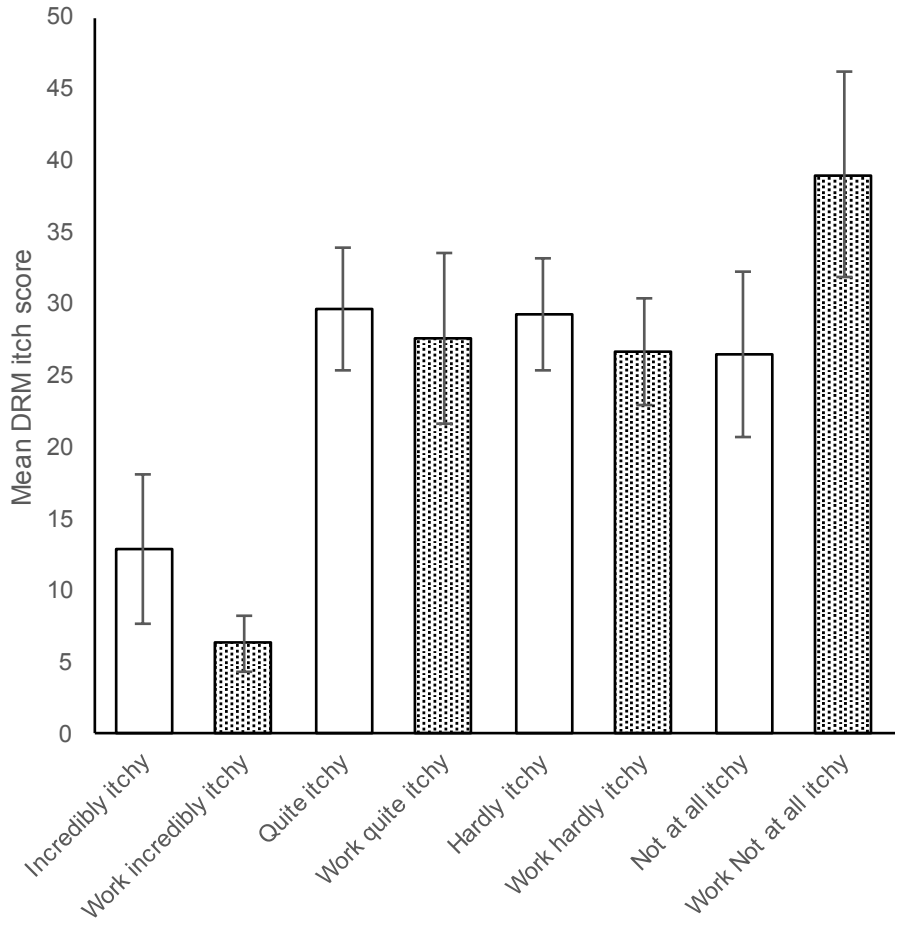


Table 5.7*. Study 2:* Table to summarise paired t test results for itch outside and inside work. Units: mean percentage time (minutes) felt itchy. .

	Mean	Standard deviation	Standard error	Paired t test
Incredibly itchy	12.967	23.201	5.188	0.487
Work Incredibly Itchy	6.333	8.908	1.992	
Quite itchy	29.608	19.187	4.290	0.541
Work Quite itchy	27.578	26.443	5.913	
Hardly noticeably itchy	29.267	17.847	3.991	0.574
Work Hardly noticeably itchy	26.722	16.704	3.735	
Not at all itchy	26.508	25.801	5.769	0.027
Work Not at all itchy	38.944	25.801	7.154	

Conclusion

The aim of this study was to use the Actiwatch Plus to characterise the features of night-to-night variation in objective itch score and to compare this to subjective measures of itch and disease extent. The work builds upon previous studies and confirms that the relation between objective measures of scratch and subjective itch is poor.

Study 1 demonstrated that there is no obvious relation between itch assessed using VAS scores and objective Actiwatch Plus scores for either children with eczema or adults with a range of diagnoses. Previous work has validated actigraphy against direct observation of children scratching at night(Benjamin et al., 2004; Bringhurst et al., 2004), and both scratching and restlessness are correlated with each other, and higher in those with atopic dermatitis than in controls. Itch related behaviour at night is not as stereotyped as when awake, and actigraphy is, we believe, a useful practical assay for scratch when compared with the time consuming nature of direct observation of video recordings. Although we analysed children and adults separately we saw no influence of sex or age on Actiwatch Plus score. The main determinant of differences in Actiwatch Plus score was diagnostic group: those with liver disease were the most severely affected. We have assumed that restlessness due to other factors differing between diagnostic groups is not confounding the scores. Although the subjective VAS itch scores differed between the different diagnostic groups, the differences were largely non-significant, again emphasising a disconnect between VAS scores and objective accelerometer score

The large variation in objective accelerometer score was confirmed again in Study 1. It was shown that about 60% of variation was between subject and roughly 40% within subject. This fact emphasises that nocturnal movements vary from night to night within any one subject, and of course limits the power of actigraphy scores for experimental studies unless repeated sampling is carried out.

Study 2 afforded the opportunity to look in more detail at the relation between VAS itch and Actiwatch Plus score over time. Little convincing relation between subjective VAS and objective Actiwatch Plus was found even in the long-lasting 42-day study. Examination of the autocorrelation showed a very different pattern for actigraphy and VAS itch. The VAS itch scores clearly showed an effect of lag, and it is not clear why this should be. Subjects were asked to post their scores into ballot boxes in an attempt to minimise filling out of results in batches, and to attempt to minimise knowledge of the score on one day influencing the score on the next day. This strategy would not be entirely robust against 'gaming' (posting a batch of results at the end, for instance) and electronic diaries or other methods of continuous sampling would perhaps provide a better approach (Kahneman, 2004; Langan, Bourke, Silcocks, & Williams, 2006; Robinson & Clore, 2002b). We suspect that the presence of the temporal pattern for VAS itch scores (and not actigraphy) is more easily explained by subjects anchoring their scores based on recollection of previous scoring (Bjorner, 2005; Kahneman & Tversky, 1996). Overall – and whether we can explain the data or not – the results suggest that the use of VAS scores for itch is limited as a subjective measure of disease activity in the context of the chronic diseases we studied.

In order to determine whether the problem was an inherent one with VAS scores or more generally to do with subjective assessment of disease, Study 2 drew in other validated subjective score of itch and of disease extent to compare to each other and to the objective Actiwatch Plus. These experiments proved that the disconnect is not because the VAS in itself is a flawed measure (it's limitations are part of published literature, (Bjorner, 2005; Torrance et al., 2001)). The fact that there is a lack of relationship between DLQI, SF-36 and Actiwatch Plus score confirms the disconnect between subjective and objective. Since part of SCORAD is VAS itch, although it is a disease extent measure, it is not surprising that there was a lack of relationship between this and Actiwatch Plus score. Reassuringly, the subjective scores do correlate appropriately with each other.

This study also offered an opportunity to explore the use of a different questionnaire, the Day Reconstruction Method, designed by Daniel Kahneman who works in the field of well-being research and cognitive behavioural psychology (Kahneman, 2004). It was designed to capture what must be one of the most difficult feelings to define, happiness. The publication which described the questionnaire (Kahneman, 2004) threw up some quite controversial findings: people happiest socialising with friends and not relatives, mothers finding childcare only one step better than their least satisfying activity, cleaning the house. The questionnaire captures 'true' emotion rather than stereotyped answers, the ones we feel we should give: 'Of course I'm happiest spending time with my husband/children.' The questionnaire 'works' by requiring the participant to build a diary of the day, hour by hour and relive each episode whilst asking questions about it. This reduces recall bias to a minimum – it is known that for most perception, how participants feel in the most recently past five minutes dictates the response even if the enquiry is about a longer period, for instance, the past week for DLQI, the past month for SF-36(Kahneman et al., 1993; Redelmeier & Kahneman, 1996; Redelmeier, Katz, & Kahneman, 2003).

There was no better relationship between DRM measured itch and objective Actiwatch Plus score. It was reassuring, again, to see that DRM aspects correlated appropriately with the other subjective measures and that again SCORAD was not related to the score. It is intriguing to see that, even when recall bias is reduced to a minimum, the disconnect between subjective and objective still exists, there is obviously some elusive facet of disease that we can not directly measure with presently available tools.

One of the most powerful tools for assessing overall health is the Global Health Question (GHQ), a deceptively simple question which just requires a simple choice of four responses. Decline in GHQ raw score can be tracked to a respondent's death (Bjorner, 2005). The fact that the response changed for the worse after DRM in this study is interesting: it would

appear to suggest that after really thinking about the hourly impact of disease participants do feel worse about their disease and health.

The impact that distraction can have on disease perception was an unexpected serendipitous aspect of itch perception which was captured by the DRM. The fact that participants did report being less troubled by itch whilst at work and more when not distracted at home is what had been expected. It demonstrates the sensitivity of the tool.

One limitation of this study is that the work was based on those attending secondary care, and therefore was probably skewed towards patients with more severe disease, as reflected by the SCORAD scores. The subjects were volunteers and, since the study involved a significant amount of time and contact with the investigators, the study group may not be typical of the reference population, another limitation. It was not possible to prospectively control the severity of disease during the study period—if patient groups at the extreme of the severity of the symptom scale had been used it is possible different conclusions might be made, although the SCORAD scores for the eczema group were as high or higher than in many therapeutic trials.

Overall, this experiment demonstrated that objective scoring using actigraphy (Actiwatch Plus or DigiTrac) may not be perfect but does confer a measure on a certain aspect of itchy disease. Some form of measurement of perception of disease or impact of disease on life should be utilised too but it should be borne in mind that the tools available to us now are not perfect either. To this researcher's mind, although it is appealing to use subjective tools because, for instance, they are used often, they are used by prominent research teams or because they are convenient, they should not be entirely relied upon.

Chapter 6

The effect of bias on subjective disease-scores.

This chapter contains work published in a paper entitled “How Robust are the Dermatology Life Quality Index and Other Self-reported Subjective Symptom Scores when Exposed to a Range of Experimental Biases?” written by Caroline S. Murray and Jonathan L. Rees. This was published in Acta Dermato-Venereologica in 2010, volume 90(1), pages 34 to 38 (Murray & Rees, 2010). A copy of the paper is included in the Appendix II.

I: Robustness of the Dermatology Life Quality Index and Other Self-reported Subjective Symptom Scores to a Range of Biases

Aim

To test widely used subjective-symptom measures for vulnerability to bias: are they measuring semantic beliefs about disease rather than actual experience?

Background

This work in this thesis has demonstrated a disconnect between objective and subjective symptom measures. Through the experiments previously described, I had validated and refined objective disease measurement in the form of actigraphy. The disconnect between objective and subjective score was still present and so the next logical step was to ascertain whether the problem or cause for the inconsistency was actually due to a 'fault' in the subjective scoring method.

Subjective-symptom measures are widely used and most commonly validated by checking for reliability, responsiveness and validity. Reliability is checked by testing for internal consistency (Cronbach's alpha), see Introduction, which ascertains that tools are measuring the same psychometric variable. Validity is checked by test-retest consistency. The subjective symptom tools had not been tested for vulnerability to bias. In these short experiments I tested the effects of focusing (Del Missier et al., 2007; T. D. Wilson et al., 2000) and framing (Del Missier et al., 2007; Wright & Goodwin, 2002) biases on Dermatology Life Quality Index (DLQI), Global Health Question (GHQ) and visual analogue scores (VAS). Please see section Introduction, section VII for descriptions of these tools.

Method

Participants:

A total of 215 patients were recruited. For each study, consecutive patients who agreed to take part were enrolled from the Royal Infirmary's Department of Phototherapy in Edinburgh. Details of specific diagnoses were not sought, the usual throughput of the Phototherapy department would suggest that the majority (70%) of the patients had psoriasis, a minority (roughly 10%) had eczema and 20% other conditions (for instance generalised pruritus). Most participants scored the DLQI in the region of 6–10, which equates to a "moderate" effect of the skin condition on the quality of life, this level of severity was expected inasmuch as recruitment was from secondary care.

All study procedures and patient interactions were conducted using a consistent, written script. The interaction script and subject information sheet explained that the studies were to determine which sort of questionnaire or score was most accurate in assessing symptoms. The interventions were described in general terms ("You will be given a list of words to memorise" or "If you are randomised into a certain group you may see a film broadcast on terrestrial television") in order to minimize unintentional "unblinding".

All participants completed the GHQ ("In general, for someone of your age, would you say that your health is excellent, very good, good or poor?"), DLQI and VAS of disease extent, itch and insomnia, always in the same order.

Inclusion criteria

- Patients attending the Phototherapy department for treatment.
- Age >14 years
- Able to give meaningful consent.

Exclusion criteria

- Subjects under 14 years of age
- People not able to understand what is involved.

Experiment 1

My hypothesis was that, if subjects were exposed to certain mood-eliciting words, they would affect the subjective score accordingly, for instance, if they had read negative words, their subjective scores would suggest worse disease.

Forty patients were randomised into two groups. Block randomisation was undertaken and all the paperwork for the study including instructions was placed in a sealed envelope so that the study was double-blinded.

Group 1 were asked to read 10 negative words, had 1 minute to memorise them and were then asked to write them out. After this, they completed the GHQ, DLQI and the VAS's of disease extent for itch and insomnia. The words are listed in Table 6.1.

Group 2 went through an identical process but the participants were given a list of 10 neutral words.

One field of Psychology research has identified and used words which elicit certain affective states. Groups 1 and 2's words, for this part of the experiment, are listed in Table 6.1 and were taken from the 'Balanced Affective Word List' (http://www.sci.sdsu.edu/CAL/wordlist/origwordlist.html). The words were matched for total character and syllable length.

A further 41 patients were randomised into 2 groups:

Group 3 were asked to read 10 negative words and then write them out (without the necessity of memorising them). The participants then completed GHQ, DLQI and VAS's or disease extent, itching and insomnia. The words are in Table 6.2. Group 4 went through an identical process but participants were given a list of 10 positive words.

Groups 3 and 4's words were taken from the University of Florida's NIMH Centre for the Study of Emotion and Attention (http://csea.phhp.ufl.edu/Media.html#bottommedia) and are listed in Table 6.2. Again, the words were matched for total character and syllable length.

Experiment 2

My hypothesis was that if a subject saw a film highlighting the negative aspects of having a skin disease, then this would make them focus on the negative aspects of living with their skin disease and so their subjective scores would imply that they had worse disease.

Fifty-four patients were block randomised into two groups:

Group 1 completed GHQ then VAS's for disease extent, itch and insomnia, after having watched a 10 minute clip from a terrestrial television broadcast ('Real Families: My Skin Could Kill Me' broadcast before the 'watershed' on ITV1 in October 2005) about living with the severe skin condition, Harlequin ichthyosis.

Group 2 just completed the subjective tools without having watched the television clip.

Table 6.1. *Experiment 1:* Lists of negative and neutral words. Words selected from Balanced Affective Word list, http://www.sci.sdsu.edu/CAL/wordlist/origwordlist.html

Group 1	Group 2
Negative words	**Neutral words**
worry	wagon
ashamed	aluminium
gloom	green
bad	bus
sick	can
suffering	submarine
unhappy	vitamin
itch	iron
misery	margin
rejected	resident

Table 6.2. *Experiment 1:* Lists of negative and positive words, University of Florida's NIMH Centre, http://csea.phhp.ufl.edu/Media.html#bottommedia

Group 3	Group 4
Negative words	**Positive words**
abuse	angel
bankrupt	birthday
betray	beauty
cancer	caress
cruel	cheer
funeral	freedom
gloom	glory
hatred	humour
hurt	home
jail	joke
misery	mother
poison	pretty
pollute	passion
rabies	reward
rejected	romantic
sad	sun
sick	sexy
suicide	snuggle
terrible	treasure
tragedy	triumph

All subjects were questioned in the same way and in the same experimental room, whether they had watched the film or not. The randomisation result (to watch the film or not) was included in the questionnaire envelope and was opened, with the interviewer being present in the study room, the interviewer then adopted the appropriate script (for whether the participant was to watch the film or not) from that point.

Experiment 3

In this study, the hypothesis was that if the DLQI focused on negative aspects of disease, then re-framing it into "neutral" frames should result in scores implying a better quality of life. We also hypothesized that if the DLQI focused on the negative, then this may negatively affect the responses to other subjective symptom scores.

Eighty patients were randomised into two groups and each of these two groups further split into 2 sub-groups, giving a total of 4 sub-groups into which subjects were block-randomised. Half the subjects answered the GHQ (Global Health Questionnaire) and standard DLQI (Dermatology Life Quality Index,)(Finlay & Khan, 1994) whilst the other half answered an altered Dermatology Life Quality Index (ADLQI) and the standard Global Health Questionnaire. The ADLQI mirrored the standard DLQI but an attempt was made for each question to be re-written in a neutral frame, thereby, minimising the possibility of a positive or negative framing and potentially reducing the possibility of a stereotyped answer. The standard DLQI is shown in Figure 6.1a and the altered version in Figure 6.1b.

Division of the two groups allowed the ordering of the examination to be manipulated, with half the subjects receiving the GHQ first and then either the DLQI or the ALDQI, with the other half receiving the GHQ second.

Demographic variables including age and sex, together with the results, were de-identified and recorded in Excel (Microsoft, Seattle). Statistical analyses were undertaken using R-software (http://www.R-project.org) (Team RDC. A language in environment of statistical computing Vienna, Austria. 2006.)

Figure 6.1a. *Experiment 3:* Dermatology Life Quality Index

DERMATOLOGY LIFE QUALITY INDEX

DLQI

Hospital No: Date: Score:

Name: Diagnosis:

Address:

The aim of this questionnaire is to measure how much your skin problem has affected your life OVER THE LAST WEEK. Please tick one box for each question.

1.	Over the last week, how **itchy**, **sore**, **painful** or **stinging** has your skin been?	Very much A lot A little Not at all	☐ ☐ ☐ ☐	
2.	Over the last week, how **embarrassed** or **self conscious** have you been because of your skin?	Very much A lot A little Not at all	☐ ☐ ☐ ☐	
3.	Over the last week, how much has your skin interfered with you going **shopping** or looking after your **home** or **garden?**	Very much A lot A little Not at all	☐ ☐ ☐ ☐	Not relevant ☐
4.	Over the last week, how much has your skin influenced the **clothes** you wear?	Very much A lot A little Not at all	☐ ☐ ☐ ☐	Not relevant ☐
5.	Over the last week, how much has your skin affected any **social** or **leisure** activities?	Very much A lot A little Not at all	☐ ☐ ☐ ☐	Not relevant ☐
6.	Over the last week, how much has your skin made it difficult for you to do any **sport**?	Very much A lot A little Not at all	☐ ☐ ☐ ☐	Not relevant ☐
7.	Over the last week, has your skin prevented you from **working** or **studying**?	yes no	☐ ☐	Not relevant ☐
	If "No", over the last week how much has your skin been a problem at **work** or **studying**?	A lot A little Not at all	☐ ☐ ☐	
8.	Over the last week, how much has your skin created problems with your **partner** or any of your **close friends** or **relatives**?	Very much A lot A little Not at all	☐ ☐ ☐ ☐	Not relevant ☐
9.	Over the last week, how much has your skin caused any **sexual difficulties**?	Very much A lot A little Not at all	☐ ☐ ☐ ☐	Not relevant ☐
10.	Over the last week, how much of a problem has the **treatment** for your skin been, for example by making your home messy, or by taking up time?	Very much A lot A little Not at all	☐ ☐ ☐ ☐	Not relevant ☐

Please check you have answered EVERY question. Thank you.

© Dermatology Life Quality Index. A Y Finlay, G K Khan, April 1992
www.dermatology.org.uk. This must not be copied without the permission of the authors.

Figure 6.1b. *Experiment 3:* Altered Dermatology Life Quality Index.

AMENDED DERMATOLOGY LIFE QUALITY INDEX

DLQI

Hospital No: Date: Score:

Name: Diagnosis:

Address:

The aim of this questionnaire is to measure how much your skin problem has affected your life OVER THE LAST WEEK. Please tick one box for each question.

1.	Over the last week, did you notice any **symptoms** or **sensations** from your skin?	Very much A lot A little Not at all	☐ ☐ ☐ ☐	
2.	Over the last week, has how **happy or confident** you've been had anything to do with your skin?	Very much A lot A little Not at all	☐ ☐ ☐ ☐	
3.	If you think about **everyday things** such as **going shopping** or **doing the garden,** has your skin altered what you've done**?**	Very much A lot A little Not at all	☐ ☐ ☐ ☐	Not relevant ☐
4.	When you think about what you **chose to wear** last week, did your choices have anything to do with your skin?	Very much A lot A little Not at all	☐ ☐ ☐ ☐	Not relevant ☐
5.	Did your choices regarding **socialising/taking part in** activities have anything to do with your skin last week?	Very much A lot A little Not at all	☐ ☐ ☐ ☐	Not relevant ☐
6.	Did your choices regarding taking part in **sporting activities** have anything to do with your skin last week?	Very much A lot A little Not at all	☐ ☐ ☐ ☐	Not relevant ☐
7.	If you were **not at work or studying** this week was it because of your skin ?	yes no	☐ ☐	Not relevant ☐
	If you **were at work or studying** last week, did **what you achieve** have anything to do with your skin?	A lot A little Not at all	☐ ☐ ☐	
8.	When you think about how **well you have got on with your partner/ any close friends** over the past week, has your skin got anything to do with it ?	Very much A lot A little Not at all	☐ ☐ ☐ ☐	Not relevant ☐
9.	Considering your **sex-life** over the last week, is your skin relevant to it?	Very much A lot A little Not at all	☐ ☐ ☐ ☐	Not relevant ☐
10.	Compared to other daily activities, over the last week, how much time/effort have you spent on **treating your skin**?	Very much A lot A little Not at all	☐ ☐ ☐ ☐	Not relevant ☐

Please check you have answered EVERY question. Thank you.

Results

The majority of variables were non-normally distributed, medians were therefore compared using the Kruskal-Wallis (KW) analysis of variance (ANOVA) and using Fisher's exact test for count data. The significance level was set at $p<0.05$. Due to the limited range of the GHQ questionnaire, these results were also examined using Fisher's exact test but this did not alter any of the conclusions and is not presented.

Experiment 1: the impact of affect-eliciting words.

A total of 81 subjects were studied and their characteristics are shown in Table 6.3. There were four intervention groups, numbered 1 to 4 as mentioned previously. There were no significant differences in the sex allocation (Fishers test p=0.81) nor median ages (Kruskal Wallis, p=0.70) between the four groups. Median scores for the four groups and p values using the Kruskal Wallis ANOVA are shown in Table 6.3. As can be seen, there were no significant differences evident.

Experiment 2: the impact of watching a film about living with a severe skin condition.

54 subjects were studied. Their characteristics, the median group-scores and Kruskal-Wallis ANOVA results are shown in Table 6.4.

There was no significant sex difference in the two groups, those who were shown the video and those who were not (Fisher, p=0.27). The median age of those shown the video was 52 as compared with those who were not shown the video of 33, a difference that is highly significant (KW, p=0.001). However, scatter plots did not show any obvious relation between age and the outcome measures so this difference was ignored. Median scores and p

values for the Kruskal Wallis ANOVA are shown in Table 6.4. As can be seen, there are no significant differences evident.

Experiment 3: re-wording DLQI into neutral frames.

80 subjects completed this experiment and their characteristics are summarised in Table 6.5. GHQ and quality of life scores (QI) were examined following four 'treatments.' The first 'treatment,' the DLQI, was compared with the second 'treatment', Altered-DLQI (ADLQI), and following this, the ordering of GHQ and DLQI/ADLQI were studied (hence treatment groups were numbered as follows: 1, 2 (DLQI) and 3, 4 (ADLQI).

There were no significant differences in sex (Fisher, p=0.17) or age (KW, p=0.92) between the four groups. Median scores and Kruskal Wallis ANOVA across the four groups for QI (DLQI and ADLQI) and GHQ are listed in Table 6.5. These differences are not significant (KW for QI p=0.47 and GHQ p=0.76). The ordering had no effect on GHQ (p=0.60) or QI (p=0.5) scores and therefore groups 1 & 2, and 3 & 4 were combined. Medians for the ADLQI and the DLQI for these combined groups were 8.5 and 7, respectively, a difference that was close to statistical significance with a *p* value of 0.07 (Kruskal Wallis test).

Table 6.3. *Experiment 1:* Study subjects' characteristics and comparison of medians with Kruskal Wallis ANOVA p values.

	Age in years:	**Sex**	**Median DLQI**	**Median GHQ**	**Median VAS extent**	**Median VAS itch**	**Median VAS insomnia**
	mean (range)	M:F ratio	(interquartile range)	(interquartile range)	(interquartile range)	(interquartile range)	(interquartile range)
Group 1 Negative, n=20	39.8 (17-71)	12:8	8.50 (5.00-11.75)	2.00 (1.00-2.00)	3.60 (2.80-5.60)	2.15 (0.90-4.33)	3.60 (1.50-7.20)
Group 2 Neutral, n=20	41.4 (18-68)	9:11	9.50 (5.75-14.25)	2.00 (2.00-2.00)	3.70 (1.80-5.30)	2.15 (1.08-4.33)	4.80 (2.85-5.73)
Group 3 Negative, n=19	39.0 (20-72)	10:9	6.00 (2.00-10.50)	2.00 (2.00-3.00)	2.50 (0.60-5.00)	1.50 (0.55-3.95)	1.90 (0.70-5.75)
Group 4 Positive, n=22	37.4 (16-74)	11:11	9.00 (2.25-12.50)	2.00 (2.00-3.00)	4.80 (2.50-7.50)	1.30 (0.40-3.00)	3.40 (1.30-6.00)
p=Kruskal-Wallis			**0.41**	**0.44**	**0.35**	**0.46**	**0.27**

Table 6.4. *Experiment 2:* Study subjects' characteristics and comparison of medians with Kruskal Wallis ANOVA p values.

	Age in years:	**Sex**	**Median DLQI**	**Median GHQ**	**Median VAS extent**	**Median VAS itch**	**Median VAS insomnia**
	mean (range)	M:F ratio	(interquartile range)	(interquartile range)	(interquartile range)	(interquartile range)	(interquartile range)
Group 1 Video, n=27	50.3 (17-77)	14:13	8.00 (4.50-9.00)	3.00 (2.00-3.00)	4.50 (2.95-6.15)	2.00 (0.85-3.90)	3.90 (2.10-6.35)
Group 2 No video, n=27	35.8 (16-74)	9:18	9.00 (4.50-15.00)	2.00 (1.50-3.00)	3.20 (1.65-6.40)	2.70 (0.95-5.15)	2.20 (1.45-5.05)
p=Kruskal-Wallis			0.34	0.20	0.50	0.11	0.21

Table 6.5. *Experiment 3:* Study subjects' characteristics and comparison of medians with Kruskal Wallis ANOVA p values.

	Age in years:	**Sex**	**Median QI**	**Median GHQ**
	mean (range)	M:F:Unknown	(interquartile range)	(interquartile range)
Group 1 DLQI then GHQ, n=21	45.1 (28-68)	14:6:1	6.50 (2.75-10.75)	3.00 (2.00-3.00)
Group 2 GHQ then DLQI, n=21	42.8 (16-79)	9:11:1	7.00 (4.50-11.00)	2.00 (2.00-3.00)
Group 3 ADLQI then GHQ, n=19	42.3 (19-81)	9:9:1	8.00 (6.75-10.50)	2.00 (1.00-3.00)
Group 4 GHQ then ADLQI, n=19	44.0 (17-76)	8:11:0	8.50 (6.00-13.00)	2.00 (1.75-3.00)
p=Kruskal-Wallis			**0.47**	**0.76**

N.B: DLQI: Dermatology Life Quality Index; ADLQI: Altered Dermatology Life Quality Index; GHQ: Global Health Question; QI: Quality of Life Index score.

Conclusions

The results presented are, essentially, negative and, in that sense, they can be viewed as reassuring: the commonly used subjective symptom-measures we employ in dermatology are resistant to the forces of bias. Using the criteria of statistical significance, we were unable to significantly alter the scores with the various attempts at manipulation of the context or wording of the questionnaires or VAS.

There are, however, a number of limitations to the work. Although we studied several hundred subjects, it is always possible that, with larger groups, differences that we observed may have been formally significant. Second, even within the experimental paradigm we adopted, there were limitations to the way the experiments were carried out. For instance, although I used a video of a child affected with skin disease, I could not find a suitable video that I thought was meaningful to use as a control. It was extraordinarily difficult to alter the wording of the DLQI without producing a caricature of it. The differences seen between the altered DLQI and the genuine DLQI approach significance but, of course, interpretation of these differences is not straightforward. The fact that a different questionnaire produces a different median score should not be viewed as too surprising and nor does it invalidate in any way, even if the differences had been significant, the absolute score on the DLQI. It is just perhaps that the median value would differ rather than any relation between the DLQI and a host of other disease features.

Although the study did not demonstrate any effects of framing or contextual factors in our study, the study itself was experimental and may not reproduce the sorts of real life factors that will influence the way people respond to questionnaires. The results do not mean that the tools are invulnerable to the effects of say, the stress of a clinic running late, liking or not

liking the clinician or 'gaming' scores to gain a particular goal, for instance, eligibility for a biologic treatment for psoriasis. It would be difficult to capture such influences.

This experiment did not set out to compare different questionnaires, or measures of aspects of diseases - there is already a large literature on this and on the advantages and disadvantages of speciality or disease specific scoring systems versus more generic questionnaires, such as EuroQOL or SF-36, for instance. (Both, Essink-Bot, Busschbach, & Nijsten, 2007; Chren, Lasek, Quinn, & Covinsky, 1997; Hongbo, Thomas, Harrison, Salek, & Finlay, 2005). Much as the universe continues to expand, so, it seems, does the number of scoring systems being proposed and advocated in the literature. This study's results were published and its final proposal was that if nothing else a 'weather eye' should be kept on the entire field of subjective-symptom measures and that tools should continue to be validated in the face of new or improved understanding of the human psyche.

With respect to attempting to explain the disconnect between objective actigraph score and subjective scores of itch, the explanation was still outstanding. My final conclusion in this, the context of the thesis, therefore is this: inasmuch as the the tools tested seem robust in the face of pressure applied by biases, the disconnect between objective and subjective can not be explained by vulnerability of the subjective symptom tools to focus and framing bias. As mentioned above, they may be vulnerable to other pressures, but the study rules out the effect of these two commonly occurring biases specifically.

Chapter 7

Effect of computer-generated 3-D models and photographs on self and third person measurement of disease-extent in psoriasis

Psoriasis extent assessed by 3-D models and the effect of own photographs on assessing self extent

Aims

1) To assess whether anonymised 3-D models of psoriasis change or improve perceived psoriasis extent.
2) To assess whether subjects' own photographs would improve self-assessed extent in a clinical treatment setting.

Background

The evidence presented in this thesis, so far, demonstrates a poor relationship between objective and subjective disease scores. This made me consider another disconnect: between doctors' and patients' assessment of disease. If this exists, surely there is potential for dissatisfaction in the care interaction especially if there are different perspectives on what is considered an acceptable treatment outcome.

When searching the literature, I just found one study which had investigated this in the field of quality of life (Janse et al., 2004). This meta-analysis found that there was more of a discrepancy between doctors' and patients' assessed subjective domains (cognition, emotion and pain) which was deemed 'moderate to good,' as opposed to objective domains (sensation, self care and mobility), which were 'good to excellent.' This is not unexpected as the objective items are more tangible. What was telling, however, was that they only found 12 studies (from an original search of over 10,000 papers) where patient and doctor reports were included and, in fact, none of these included studies had actually reported the discrepancy in results(Janse et al., 2004).

It is not unexpected that these subjective-disease questionnaire results may diverge but there is also evidence of variance in measuring disease extent. Several studies have checked for intra- and inter-operator variability in scoring this in dermatology. Variability has been found in scoring eczema with the available tools (Charman et al., 1999) and psoriasis (Marks, 1989; Ramsay & Lawrence, 1991) (Berth-Jones et al., 2006). It appears that the crux of the matter may be that neither patients nor practitioners are very accurate or reliable at estimating affected surface area coverage. In a study using one-dimensional cut-out models of psoriasis of known surface area coverage, it was shown that there was considerable difference in assessment: dermatologists under-estimated coverage whilst nurses over-estimated and, the most accurate at assessing area were 'rule of 9'-naïve and dermatology-naïve medical students (Tiling-Grosse & Rees, 1993).

It occurred to me that, perhaps, standardised photographs of known surface area coverage, which could be used by patients and doctors for comparison, could be helpful, however, problems with data protection and personal features preclude unbiased patient-to-photograph comparisons. Also, it is very difficult to achieve accurate calculation of the actual surface area covered on photographs due to their being 2-D representations of 3-D objects. Despite evolving technology, there are not yet good systems of determining psoriasis extent by a computer analysis of a photograph (Ashcroft, Wan Po, Williams, & Griffiths, 1999; Park et al., 2004; Savolainen, Kontinen, Alatalo, Röning, & Oikarinen, 1998).

I therefore proposed a series of computer-generated images of increasing severity (or disease coverage). I wrote to the Dean of the Edinburgh College of Art to introduce myself and the potential study. I received a very helpful response. He introduced me to the artist, Beverley Hood, who worked in the field of animation, who had designed 'avatars' and who, therefore, was ideally suited and open to a collaboration. A series of 36 computer generated images comprising 36 disease severities or psoriasis coverages ranging from 1.95% to 62% were

designed. The models were anonymous, featureless figures, which, therefore, allowed an unbiased relationship between assessor and model: a blank canvas for the respondent to map himself or herself onto.

The images were used, firstly, to investigate perception of disease extent across different medical and non-medical groups and, secondly, for patients to assess their own disease extent and to compare this with assessment with PASI and DLQI. The overall aim was to unify patients' and doctors' perspectives of extent and thus hopefully encourage more satisfying clinical interactions on both sides.

The reasoning for introducing the other facet of own-photograph prompts to the second part of this study was to ascertain if the 'hedonic treadmill' (Kahneman, Diener, & Schwarz, 1999; Riis et al., 2005) was affecting self-perception of disease-extent in a chronic disease, psoriasis. It was speculated that, if this were the case, adaption to the situation could tend to make patients assess their disease as less severe or extensive than it really was. The hedonic treadmill can, of course, work in the other direction too: if a patient is so used to their usual condition, they may not perceive the improvements that treatments effect. This was why the effect was tested on a group about to embark upon treatment. It had been noted in Phototherapy clinics that patients appeared pleasantly surprised (curiously) when seeing their own photographs at mid-way review. They mostly commented that they had improved more than they had thought. It was this anecdote and the cognitive behavioural theory which informed the second study.

Method

Development of the computer models.

I selected photographs of psoriasis patients. A range of patterns and severities were selected from patients attending secondary care and included chronic plaque psoriasis and guttate

disease. The full range of available photographs did not include true erythroderma. The photographs were standardised (standard lighting and positional conditions), did not include the face and the subject had consented to their use for research and teaching. They were sorted into nine folders of increasing extent. It was expected that the severities would run from one to ten where one was no psoriasis and ten maximal, hence nine severities, starting at two, minimal psoriasis, on this scale. No formal measurements of affected surface area were made on the photographs.

Meanwhile, the artist, Beverley Hood had created a humanoid 3-D figure or 'avatar.' When designing the body-shape of the model, a realistic, naturally proportioned, figure was purposefully produced – not an idealised, fantasy body commonly associated with computer generated models (viz. Lara Croft). Since it was recognised that the creation of each model would be so intricate, it was decided that, certainly initially (and for the purpose of this study) a model with female form only would be designed and therefore we would only recruit female patients for parts of the studies which involved the participants comparing themselves to the models. Figure 7.1 illustrates the 'avatar' without any psoriasis.

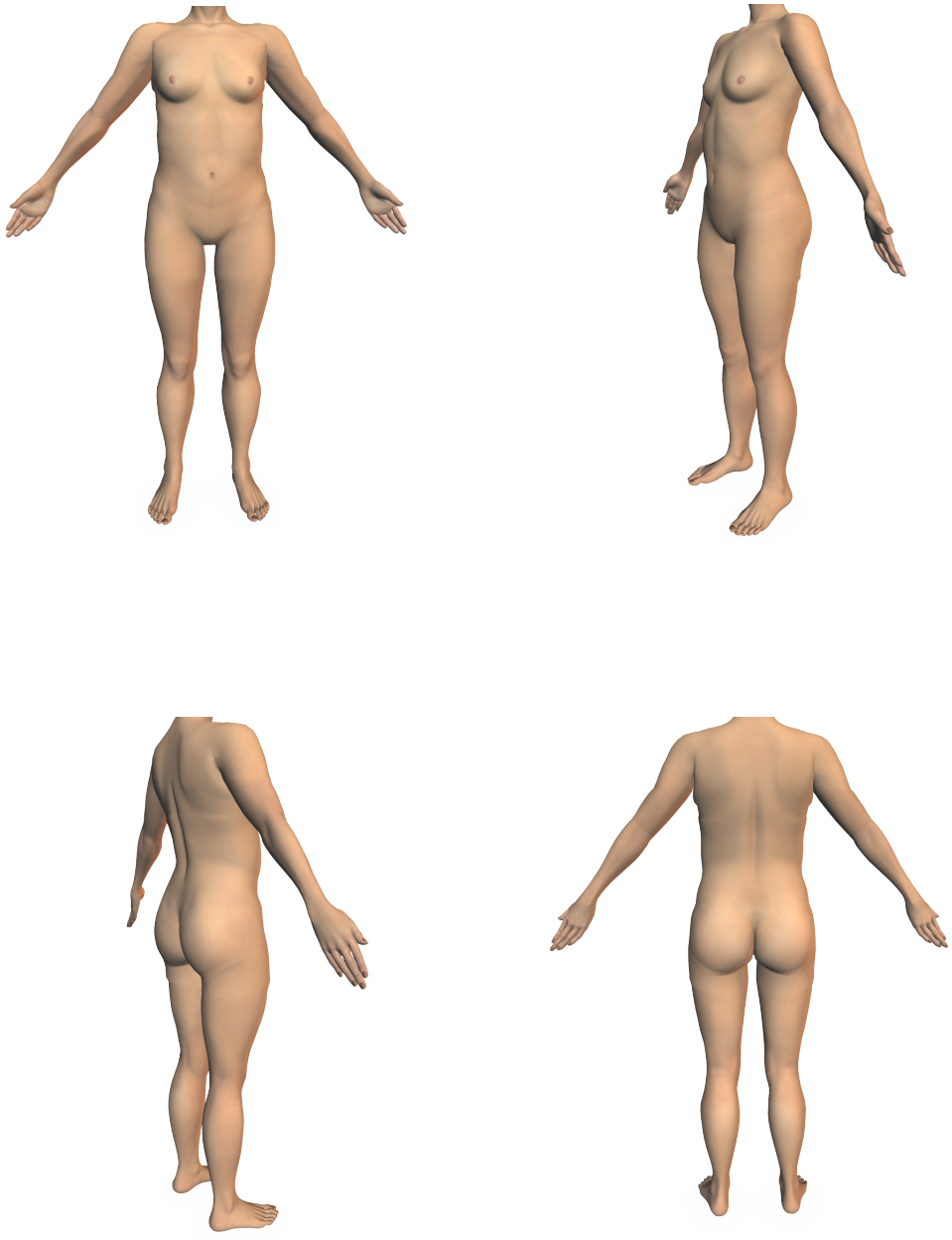
The folders of psoriasis photographs were passed to the artist and she then created the 3-D models as representative composites of different disease coverage, they were not, therefore, mere illustrations of one patient photographed. A series of computer images comprising nine psoriasis severities (each with views of front, back, half front and half back) were designed using the 3-D character-modelling package, 'Poser' (Smith Micro Software Inc., Santa Cruz, CA) and the imaging software, 'Adobe Photoshop' (Adobe Systems Inc., San Jose, CA).

Although only nine models of increasing severity were designed, the fact that the whole figure could be viewed from four different angles (front, back, half front and half back) provided 36 different percentage-coverage views available for evaluation. 'Weighting' of

severity relating to certain body sites affected: areas that cannot easily be covered, the face and the hands, has been demonstrated (Finlay & Coles, 1995). This is why our models were designed with deliberate avoidance of coverage in these sites. The first 3-D first models had unaffected faces in situ but sample respondents found these models harder to relate to – even the bland facial features were distracting – and so headless models were created.
Figure 7.2 demonstrates the psoriasis images created.

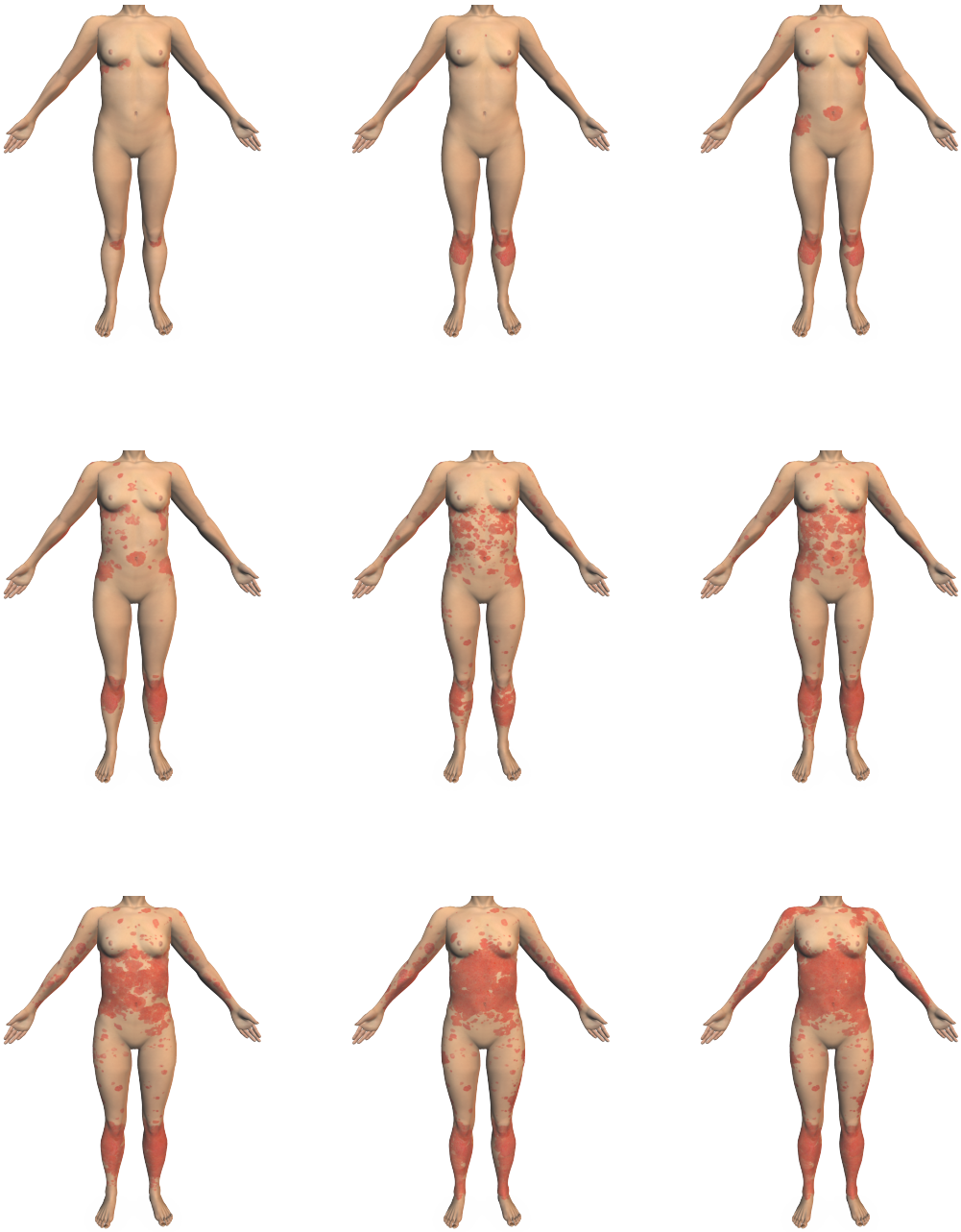
The images of psoriasis-affected models had been entirely artistically created: it was not decided that they should represent certain percentage disease coverages 'up front.' Therefore, the percentage surface area covered by disease had to be discerned. The 3-D image was flattened in 'Adobe Photoshop.' The resulting image is best compared to an animal skin rug (see Figure 7.3a). The 'Histogram' feature of 'Adobe Photoshop' yields the number pixels of each colour. I therefore carefully coloured each plaque of psoriasis the opposite spectral colour from skin-colour (blue) having removed the top rendering layer of the image. The 'Histogram' feature could therefore reveal the number of pixels of psoriasis compared to unaffected skin. Figure 7.3a shows an example of the flattened image and an example of the same image coloured in order to assess percentage disease coverage (Figure 7.3b). This flattened image allowed surface area coverage for the entire 3-D model, but not for the posed views, so these had to be estimated from the 3-D form, colouring the plaques blues and using the 'Histogram' feature as previously described. The surface area coverage for each image was calculated by dividing the number of pixels of plaque coverage by the entire image pixel number. Psoriasis coverage ranged from 1.95% to 62.00%, see Table 7.1. Figures 7.4 and 7.5 are graphs demonstrating the BSA (body surface area) % coverage of psoriasis in each 3-D pose (Figure 7.4) and combined and in order of BSA % extent in Figure 7.5. It can be seen that the rise in extent is linear (R square=0.976).

Figure 7.1. Computer generated 3-D ‘avatar.’



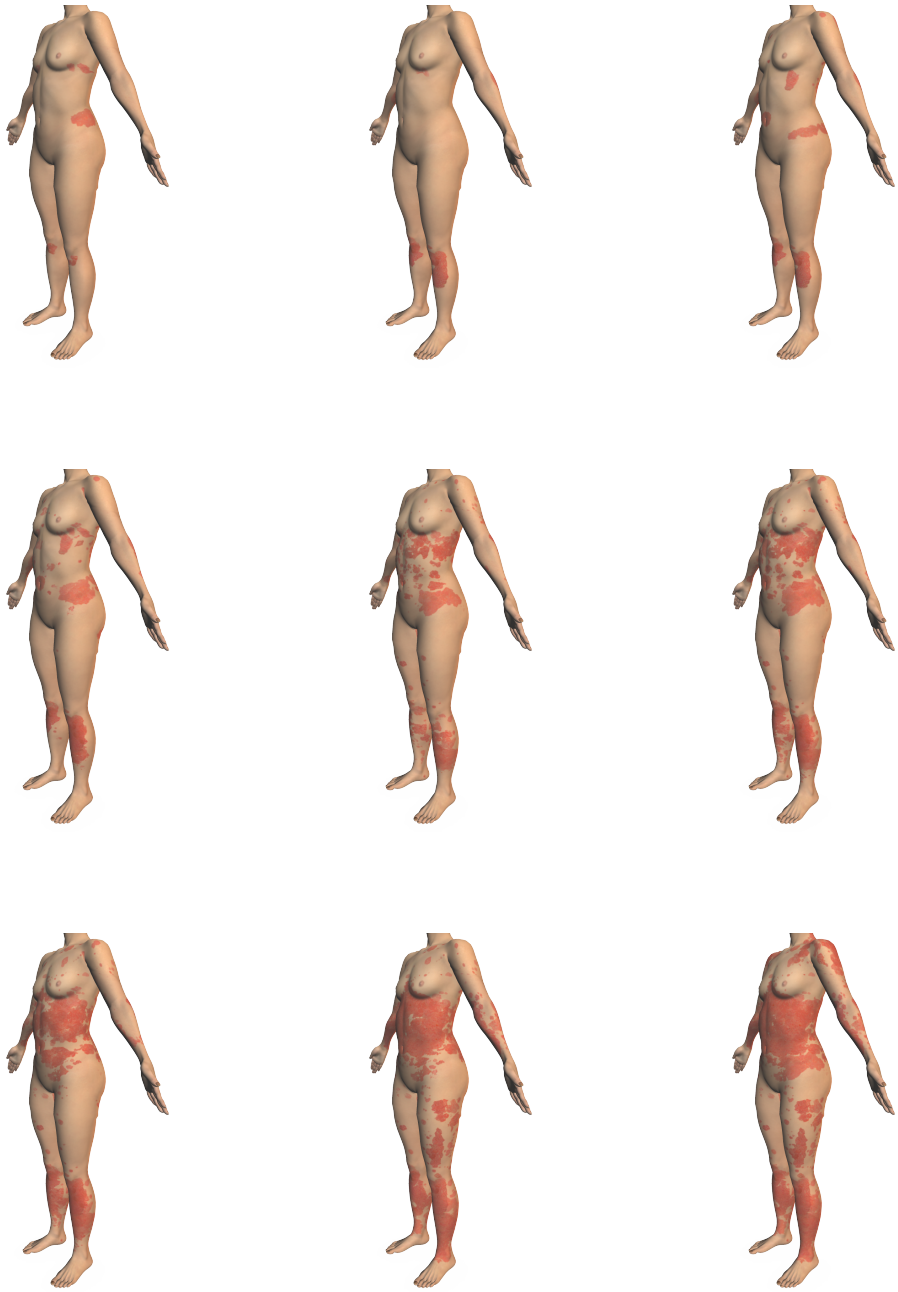
These images are reproduced with the kind permission of Beverley Hood and are her property and thus not for reproduction.

Figure 7.2a. Computer generated 3-D images of psoriasis, increasing severity from left to right, top to bottom: front view.



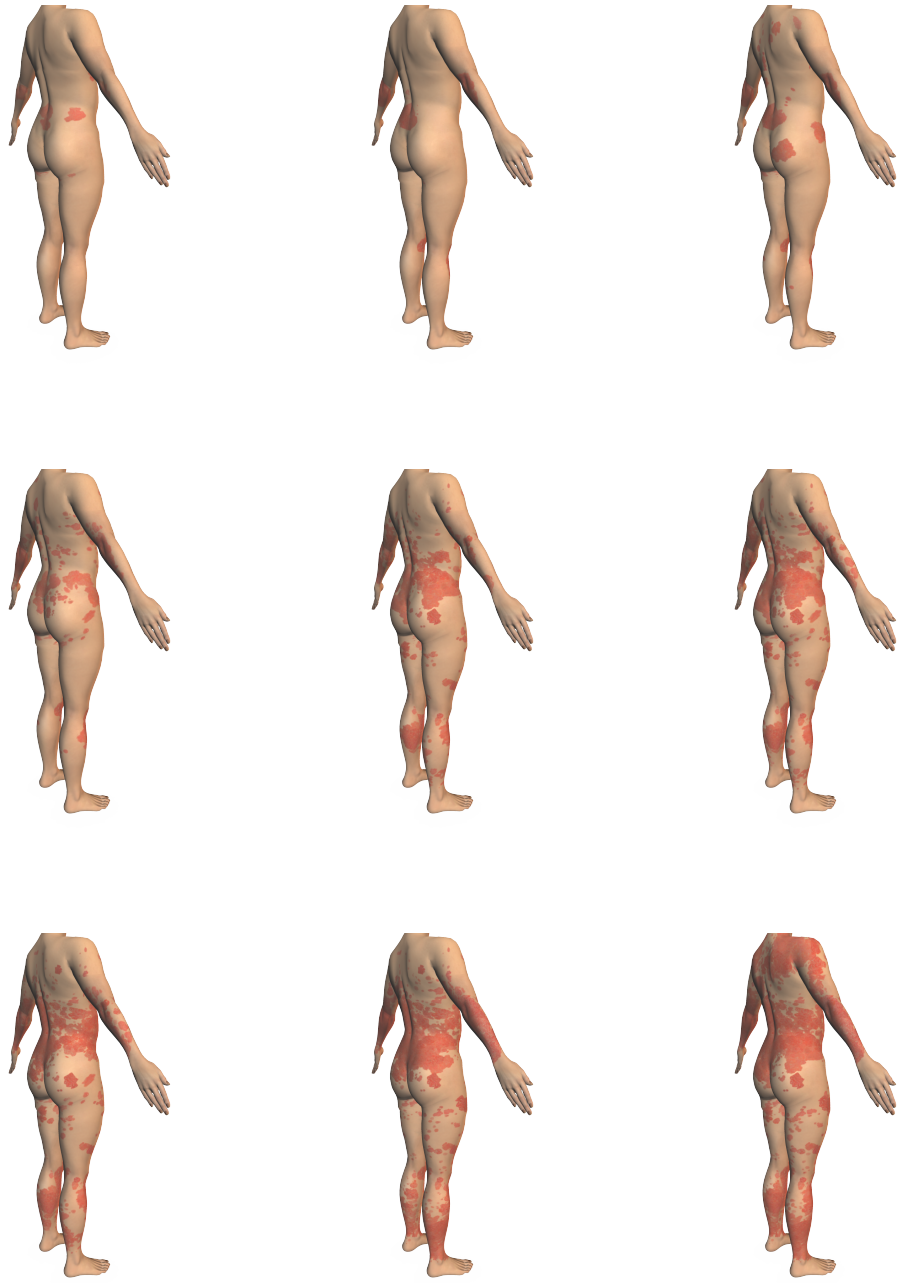
These images are reproduced with the kind permission of Beverley Hood and are her property and thus not for reproduction.

Figure 7.2b. Computer generated 3-D images of psoriasis, increasing severity from left to right, top to bottom: half front view.



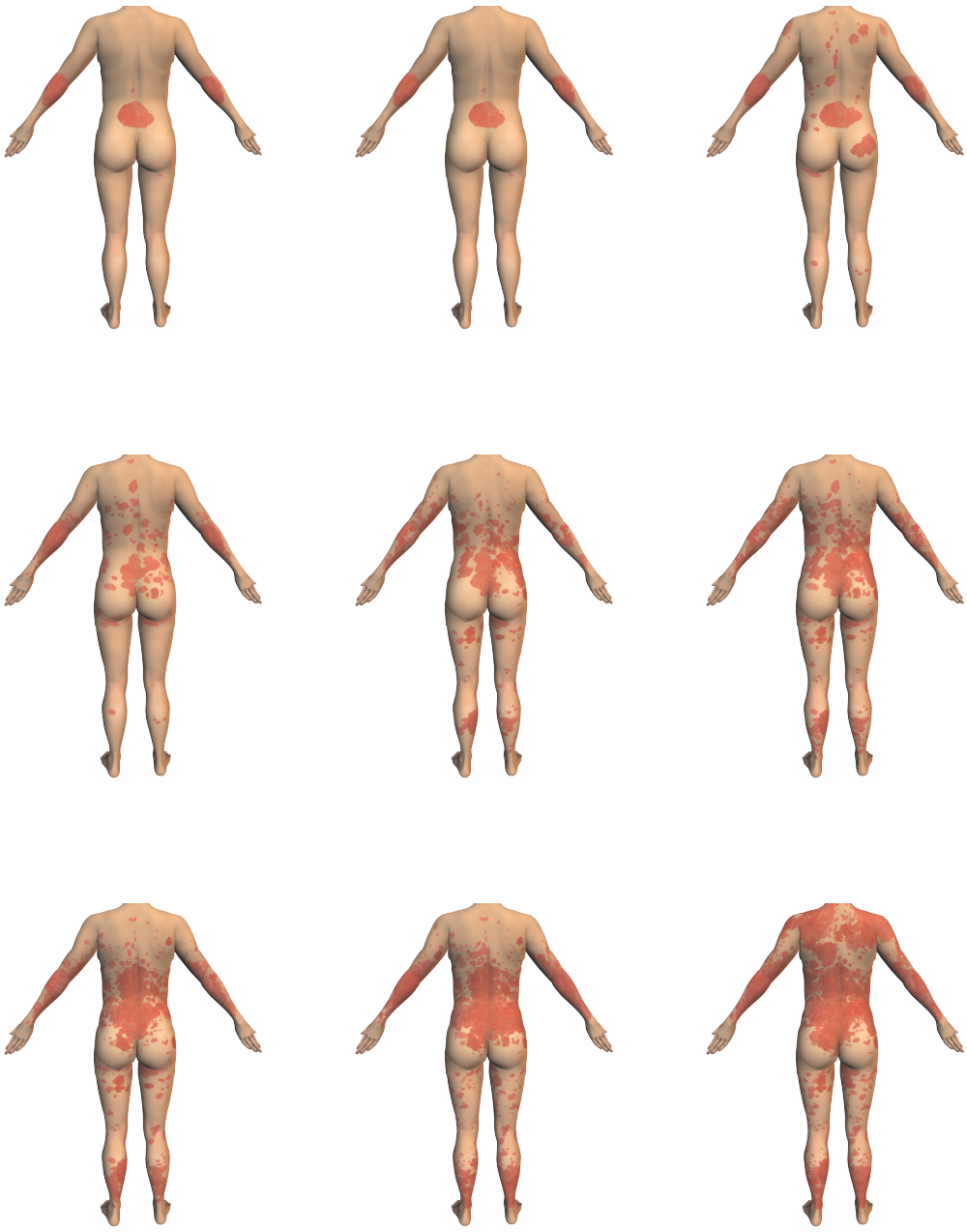
These images are reproduced with the kind permission of Beverley Hood and are her property and thus not for reproduction.

Figure 7.2c. Computer generated 3-D images of psoriasis, increasing severity from left to right, top to bottom: half back view.



These images are reproduced with the kind permission of Beverley Hood and are her property and thus not for reproduction.

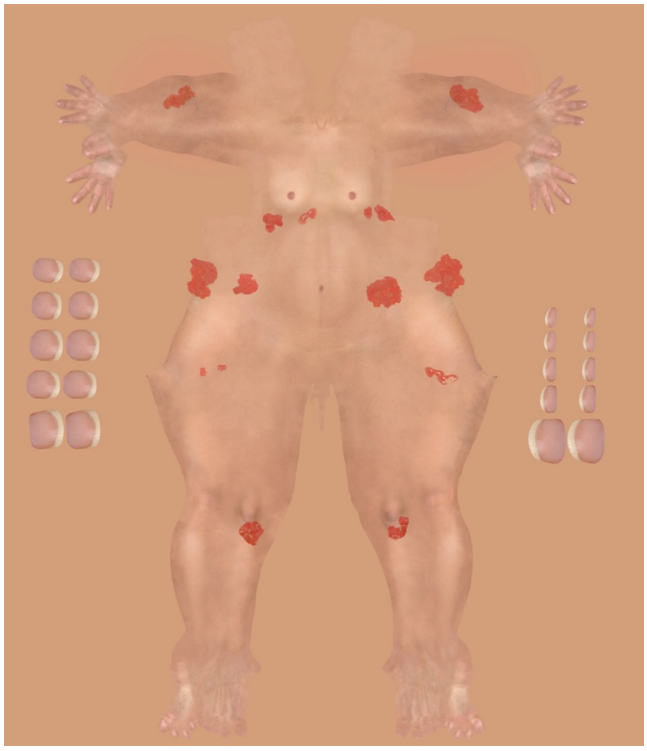
Figure 7.2d. Computer generated 3-D images of psoriasis, increasing severity from left to right, top to bottom: back view.



These images are reproduced with the kind permission of Beverley Hood and are her property and thus not for reproduction.

Figure 7.3. Computer generated 2D images of psoriasis, a) original, b) psoriasis plaques coloured blue for surface-area analysis.

a)



b)

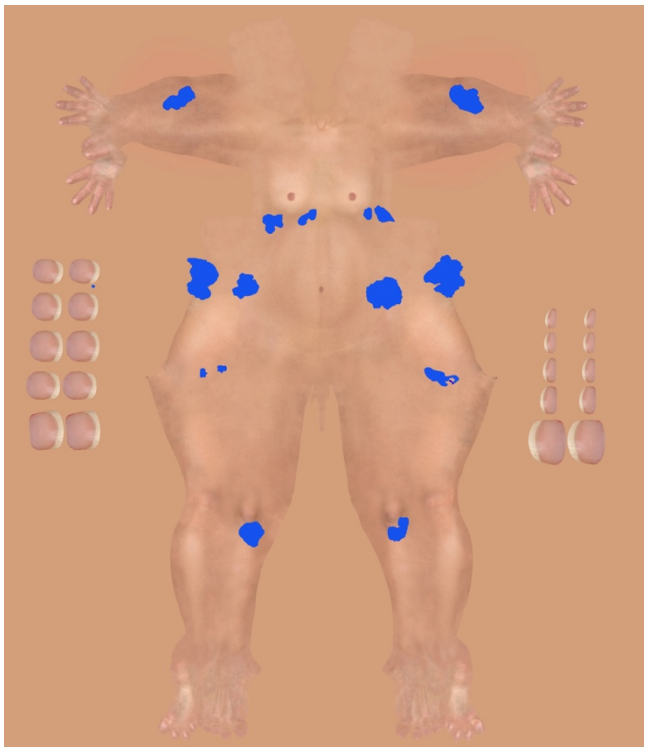


Table 7.1. Percentage Body Surface Area (BSA) coverage for the 3-D model posed views (see Figure 2).

	Pose and Body Surface Area coverage			
	Front	**Half front**	**Half back**	**Back**
1	2.0	2.3	3.7	4.6
2	5.5	4.7	5.1	7.6
3	9.5	7.8	10.7	14.7
4	15.4	14.8	17.1	20.7
5	21.2	19.9	26.1	28.5
6	28.2	27.4	33.3	35.9
7	36.3	31.8	35.4	37.2
8	43.6	44.1	44.3	47.0
9	50.7	51.5	57.6	62.0

Figure 7.4. Line graph to show BSA (Body Surface Area) coverage (%) for each view in rising severity order.

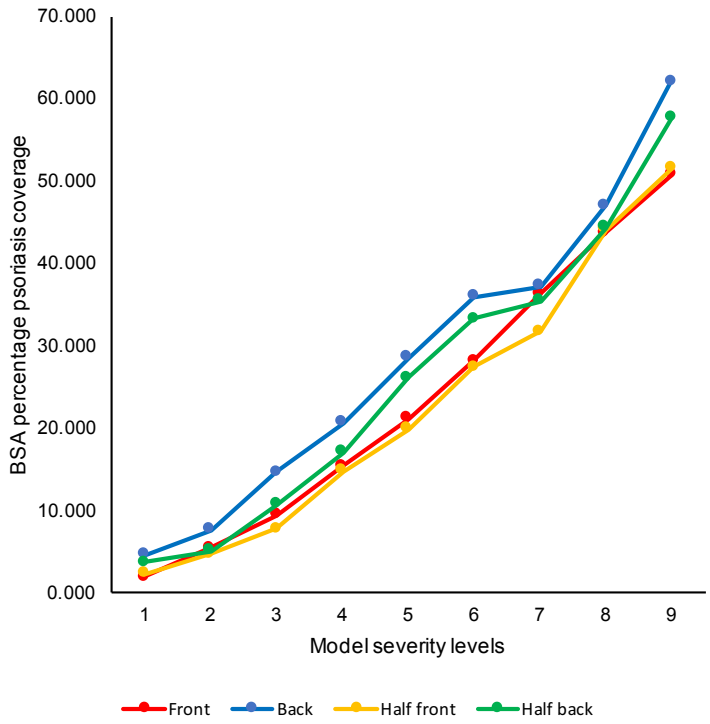
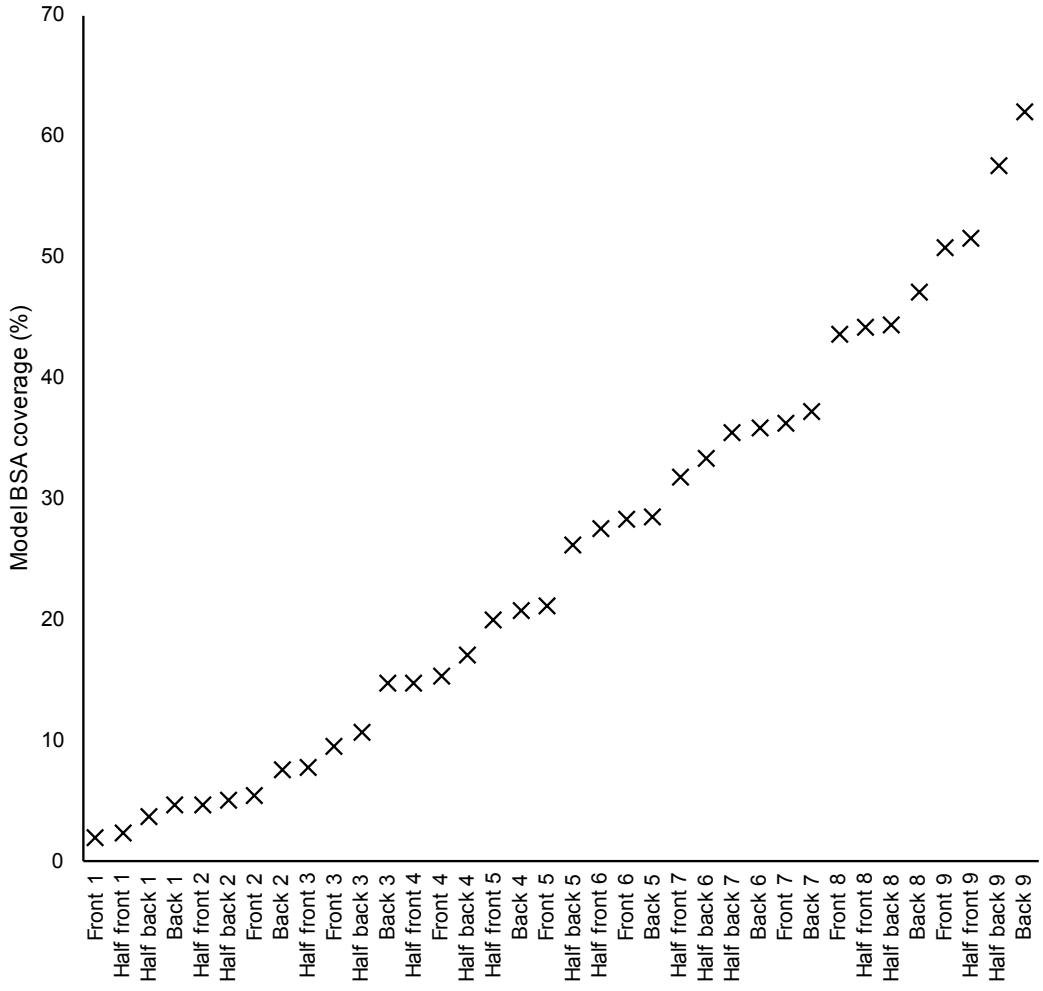


Figure 7.5. Scatter plot to show all 3-D models in order of increasing BSA (Body Surface Area) coverage (%).



Study 1: ***Do standardised 3-D images of psoriasis improve inter-assessor consistency for disease extent?***

Study 1 participants

60 participants were recruited: 12 dermatology doctors, 8 dermatology nurses, 5 general nurses, 10 medical students and 5 administrative staff as well as 20 female psoriasis patients The psoriasis patients were females approached consecutively who were attending the Royal Infirmary of Edinburgh's Psoriasis and Phototherapy clinics. All study procedures and subject interactions were conducted with a consistent written script and answer-sheet.

Study 1 protocol

The 3-D models were presented in random order of severity in four separate folders. Each folder contained a different view (front, back, half front and half back) of the models. The folders were presented in the image viewing system of Mac OS X, Leopard (Apple, California). Table 7.2 shows the *actual* order of severities presented to the subject. Each participant was asked to: 'Go through each image in order and give the model a score out of 10 where 10 is the most severe and 0 the least severe. If you think two models are the same you can give the same score.' They were also asked to pick one model that they thought best-represented 'mild disease,' one which represented 'moderate' disease and, finally, one model to represent 'severe' disease. This process was repeated for each folder and therefore, each view, and respondents had no access to their responses or to figures from the other 'views.' Participants were also asked to score their Global Health Question (GHQ) response.

The participants' characteristics are summarised in Table 7.3. Disease extent was not formally assessed: moderate to severe level of disease was expected from the source of recruitment, a tertiary referral clinic and estimated with the GHQ (Fayers, 2005).

Table 7.2. *Study 1:* Random order of severities (number and %BSA coverage) presented to participants

	Position view and model severity			
Order of presentation	**Front**	**Half front**	**Half back**	**Back**
First folder	6	1	4	1
Second folder	1	3	7	7
Third folder	3	9	5	2
Fourth folder	2	6	3	9
Fifth folder	9	4	9	8
Sixth folder	4	7	1	5
Seventh folder	8	2	2	4
Eighth folder	7	5	6	3
Ninth folder	5	8	8	6

Table 7.3. *Study 1:* Participants characteristics and GHQ (Global Health Question response), n=60

Type of respondent	Number	Median age, years (and range)	Male:Female	Median GHQ (and range)
Administrative assistant	5	54 (23-56)	1:4	3.0 (3.0-4.0)
Dermatology doctor	12	33.5 (27-51)	5:7	4.0 (3.0-4.0)
Dermatology nurse	8	47 (39-53)	2:6	2.5 (1.0-3.0)
General nurse	5	38 (20-51)	0:5	3.0 (2.0-4.0)
Medical student	10	22 (19-25)	6:4	3.5 (3.0-4.0)
Patient	20	46.5 (19-77)	0:20	1.88 (0-4.0)

The scoring was found to be normally distributed, therefore parametric tests were undertaken: means were compared by standard ANOVA and linear regression and ANOVA performed when comparing respondent scores of extent to the models' extent (Body Surface Area, BSA, %).

The respondents were clustered as follows for the purposes of analysis:

1) Dermatology doctors, n=12: included Consultants and Registrars trained and practising in dermatology.
2) Dermatology nurses, n=8: nurses specifically trained in and practising in dermatology
3) General nurses, n=5: nurses from other specialities, Health Care Assistants (none of whom undertook specific dermatology care).
4) Administrative, n=5: staff who work in administrative roles such as receptionists, Medical Records staff. Also included 'Research Technicians' as these had no clinical dermatology experience.
5) Medical students, n=10: students on the first day of their dermatology attachment who had no previous dermatology experience.
6) Patients, n=20: the volunteers with psoriasis

Study 2: ***Do standardised 3-D images of psoriasis and photograph prompts improve self assessed disease extent?***

Study 2 participants

21 patients were recruited. All were female adult patients attending the Royal Infirmary of Edinburgh. Consecutive female patients were approached who were attending the Phototherapy or Psoriasis tertiary referral clinics. It was decided only to pick female patients as the computer models were female forms (as stated above).

Patients who were *about to commence* phototherapy or systemic treatment were given information about the study. The patients were given time to consider the invitation to participate in the study, and then contacted again by telephone. Those who agreed to participate then had an appointment to meet the researcher arranged.

The subjects' median age was 40 (range 21 to 62 years). The mean baseline Global Health Question response was 2 (range 0 to 3). The baseline mean DLQI was 15.9 with a range of 3 to 27 and mean baseline PASI was 8.9 with a range of 3.2 to 14.6.

Inclusion criteria

- Psoriasis (as diagnosed by Consultant).
- Female
- Starting phototherapy/Ciclosporine/Biologics
- Age >16yrs

Exclusion criteria

- Patients with other skin conditions
- Males

- Not about to commence Phototherapy or systemic treatment
- Age <16yrs

Study 2 protocol

On the day that the treatment (Phototherapy or systemic treatment) was due to commence (Day 0), the patient met the researcher and:

1) Clinical, standardized photographs were taken.
2) PASI (Psoriasis Area Severity Index) was scored by the researcher.
3) The computer-generated models were presented in random order of severity in Leopard, Mac OS X (as in Study 1). Only 'front' and 'back' views were shown (as these approximated to the clinical photograph views). The participants had full control of which models to bring into central view and were shown how to move around amongst the images by using the computer keyboard. The participants were then asked to select the model which, "most represents how you feel about your psoriasis today." This selection was called 'Selection 1.'
4) Subjects then completed a VAS score of disease severity and DLQI.
5) The participants were then asked to look at that day's (Day 0) photographs and then presented with a new viewing folder of images (presented in a different order) and asked to select a model which "most represents how you feel about your psoriasis today." This selection was called 'Selection 2.'

On Day 42, the patient and researcher met again.

1) Clinical photographs were taken
2) The PASI scored by the (same) doctor.
3) The computer-generated models were presented in random order of severity in Leopard, Mac OS X (as on Day 0 and Study 1). The participants were asked to select

the model which, “most represents how you feel about your psoriasis today.” The chosen model was called ‘Selection 3.’

4) Subjects then completed a VAS score of disease severity and DLQI.
5) The participants were asked to look at that day’s (Day 42) photographs and then presented with a new viewing folder of images (presented in a different order) and asked to select a model which “most represents how you feel about your psoriasis today.” This model was called ‘Selection 4.’
6) Finally, the subjects were shown their own Day 0 photographs. Again, after studying these, they were presented with a further folder of images presented in a different order and asked to select a model “most represents how you feel about your psoriasis today.” This final chosen model was called ‘Selection 5.’

Please see Figure 7.6 for a flowchart of the study. Table 7.4 shows the actual order of severities presented to the subjects.

Statistical analysis

Data was analysed using software (Excel, Microsoft Ltd, USA and StatsDirect Ltd, United Kingdom). Pair-comparisons were made using Student’s t test as the data was normally distributed data and by ANOVA for comparison across several groups. Correlations were evaluated by using Spearman’s rho and linear regression of assessed score on actual pixel score was calculated, as was ANOVA of the regression coefficients and intercepts.

Figure 7.6. *Study 2* flowchart

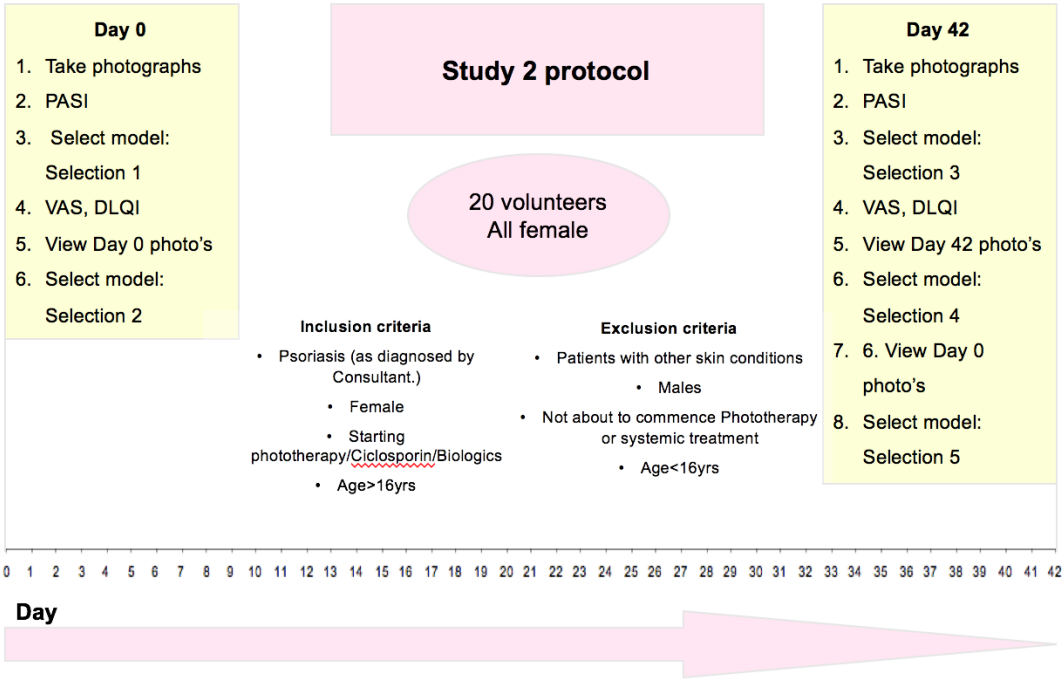


Table 7.4. *Study 2:* Order images presented to patients

Set 1	**Set 2**	**Set 3**
Selections 1 and 3	Selections 2 and 4	Selection 5
Front 6	Front 4	Front 8
Back 1	Back 5	Back 2
Front 1	Front 8	Front 6
Back 7	Back 2	Back 5
Front 3	Front 7	Front 9
Back 2	Back 8	Back 4
Front 2	Front 2	Front 2
Back 9	Back 3	Back 1
Front 9	Front 9	Front 5
Back 8	Back 4	Back 3
Front 3	Front 6	Front 1
Back 5	Back 7	Back 8
Front 8	Front 5	Front 4
Back 4	Back 1	Back 6
Front 7	Front 3	Front 7
Back 3	Back 6	Back 7
Front 6	Front 1	Front 3
Back 6	Back 9	Back 9

Results

Study 1: ***Do standardised 3-D images of psoriasis improve inter-assessor consistency for disease extent?***

Participants were not selected or controlled for as this was an exploratory study, therefore the groups displayed differences beyond their being medical, non-medical or patients. Table 7.3 summarises these features but importantly it should be noted that the age and Global Health Question (GHQ) results are significantly different across the groups. Notable differences are that the patients score the lowest median GHQ response, 1.88 (poorer general well-being) followed by dermatology nurses (median 2.5) the general nurses and administrative staff (median 3), medical students (median 3.5) and finally, dermatology doctors (median 4). The ages of the groups from oldest to youngest are: administrative staff (median 54 years), dermatology nurses (median 47 years), patients (median 46.5 years), general nurses (median 38 years), dermatology doctors (median 33.5 years and finally medical students, (median 22 years). The groups were compared and found to be significantly different for age and GHQ score response (ANOVA, $p<0.001$). Since baseline differences in the groups had been demonstrated, scoring differences were expected too across the different respondent groups.

Comparing how the different groups score the models for extent 'out of ten.'

Each group of participants' assessed severity correlated positively and linearly with the actual severity (actual pixel coverage), even when all 36 severity-extents were viewed in random order. These mean scores and standard deviation for each pose and each respondent group are summarized in Table 7.5. There is a lot of data to in the table and the easier way to appreciate the results is via the graphs. Figure 7.7 includes graphs showing the mean scores for all respondents' severity score out of 10 (different labels for each group) compared to actual BSA % coverage of poses. It can be seen that dermatology doctors tend to score less

generously with respect to severity (lower part of the scatter) compared to non-medical and non-dermatologically trained respondents (higher part of the scatter).

ANOVA analysis of respondent group for severity (1 through to 9) is summarised in Table 7.6. When scoring the 'front' pose, dermatology doctors score lowest (mean 5.11) followed by dermatology nurses (mean 5.21), then medical students (mean 5.23), then patients (mean 5.39), administrative staff (mean 5.64) and finally the most generous were general nurses (mean 6.38). The order of scoring for 'half front' pose, from lowest to highest is: dermatology doctors (mean 4.74), dermatology nurses (mean 4.75), medical students (4.99), patients (mean 5.54), administrative staff (mean 5.56) and the highest scorers were general nurses (mean 6.07). The order of scoring for 'half back' pose, from lowest to highest is: dermatology doctors (mean 4.89), dermatology nurses (mean 5.06), medical students (5.18), administrative staff (mean 5.56), patients (mean 5.62), and the highest scorers were general nurses (mean 6.07). The order of scoring for 'back' pose, from lowest to highest is: dermatology doctors (mean 4.95), medical students (5.20), dermatology nurses (mean 5.29), administrative staff (mean 5.60), patients (mean 6.01), and the highest scorers were general nurses (mean 6.16). The group differences were not significantly different when analysed by ANOVA for the 'front' pose ($p=0.183$. They were just significant for 'half front' ($p=0.019$) and not for 'half back' ($p=0.087$). The group differences were significantly different for the 'back' pose ($p=0.005$).

Table 7.5. *Study 1:* Mean severity scores for each 3-D model.

	Mean severity score (standard deviation)								
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Dermatology doctors: FRONT	1.500 (±0.688)	1.833 (±0.603)	2.750 (±0.467)	4.667 (±0.647)	5.667 (±0.924)	6.250 (±0.874)	7.250 (±0.751)	7.750 (±0.647)	8.333 (±1.191)
Dermatology nurses: FRONT	1.875 (±0.991)	2.500 (±0.756)	3.625 (±0.744)	4.375 (±0.518)	5.750 (±0.886)	6.375 (±0.744)	6.875 (±0.641)	7.500 (±0.756)	8.000 (±0.926)
General nurses: FRONT	1.400 (±0.548)	2.600 (±0.894)	2.800 (±0.837)	6.800 (±1.095)	7.200 (±1.304)	7.800 (±1.304)	9.200 (±0.837)	9.600 (±0.548)	10.000 (±0.000)
Medical students: FRONT	1.200 (±0.422)	1.900 (±0.568)	2.700 (±0.483)	4.700 (±0.675)	6.000 (±0.943)	6.400 (±0.966)	7.600 (±0.843)	8.300 (±0.823)	9.000 (±0.667)
Administrative staff: FRONT	1.600 (±1.517)	2.200 (±1.095)	3.400 (±1.140)	4.800 (±1.483)	5.800 (±1.304)	6.400 (±0.894)	8.200 (±0.837)	8.800 (±0.447)	9.600 (±0.894)
Patients: FRONT	1.850 (±1.089)	2.550 (±1.572)	4.550 (1.146)	5.400 (±1.095)	6.450 (±1.234)	7.450 (±0.759)	7.600 (±1.273)	8.500 (±1.000)	9.700 (0.571)
Dermatology doctors: HALF FRONT	1.583 (±0.674)	1.917 (±0.539)	2.667 (±0.786)	4.333 (±0.934)	4.750 (±1.104)	6.000 (±1.044)	6.333 (±0.647)	7.000 (±0.944)	8.083 (±1.265)
Dermatology nurses: HALF FRONT	1.625 (±0.744)	1.875 (±0.835)	2.750 (±0.707)	4.375 (±0.916)	4.500 (±1.195)	6.000 (±1.069)	6.000 (±1.195)	7.250 (±1.165)	8.375 (±1.188)
General nurses: HALF FRONT	1.400 (±0.548)	2.200 (±0.447)	2.800 (±0.837)	5.800 (±1.304)	7.000 (±1.225)	8.000 (±0.707)	8.200 (±0.837)	9.400 (±0.894)	9.800 (±0.447)
Medical students: HALF FRONT	1.300 (±0.483)	1.600 (±0.516)	2.600 (±0.516)	4.200 (±0.632)	5.100 (±1.287)	6.500 (±1.080)	6.900 (±0.994)	7.800 (±0.789)	8.900 (±0.738)
Administrative staff: HALF FRONT	2.000 (±1.225)	2.200 (±1.643)	3.200 (±1.304)	4.800 (±1.304)	5.600 (±1.517)	7.200 (±0.837)	7.000 (±1.225)	8.400 (±0.548)	9.600 (±0.548)
Patients: HALF FRONT	1.700 (±1.031)	2.100 (±0.968)	2.750 (0.639)	5.150 (±1.531)	5.750 (±1.372)	7.100 (±1.373)	7.450 (±1.468)	8.800 (±0.894)	9.100 (±1.071)
Dermatology doctors: HALF BACK	1.750 (±0.622)	2.000 (±0.739)	3.000 (±0.739)	3.667 (±0.985)	5.333 (±1.303)	6.583 (±1.240)	5.917 (±1.084)	7.417 (±1.311)	8.417 (±1.379)
Dermatology nurses: HALF BACK	2.250 (±0.707)	2.625 (±0.518)	3.375 (±0.744)	4.000 (±0.756)	5.500 (±0.756)	6.125 (±1.356)	6.250 (±1.282)	7.125 (±1.458)	8.250 (±1.035)
General nurses: HALF BACK	2.400 (±1.140)	2.600 (±1.140)	4.000 (±1.871)	4.400 (±1.673)	6.600 (±2.302)	8.200 (±1.304)	7.600 (±2.074)	8.800 (±1.643)	10.000 (±0.000)
Medical students: HALF BACK	2.300 (±1.767)	1.800 (±0.632)	3.200 (±1.619)	3.800 (±1.033)	5.800 (±1.317)	6.900 (±0.876)	6.200 (±1.476)	7.800 (±0.789)	8.800 (±1.549)
Administrative staff: HALF BACK	2.000 (±1.225)	2.400 (±1.140)	3.400 (±1.140)	4.600 (±1.517)	5.800 (±1.304)	7.200 (±1.304)	7.200 (±0.837)	8.000 (±1.000)	9.400 (±0.894)
Patients: HALF BACK	1.900 (±1.210)	2.250 (±0.910)	4.050 (±1.356)	3.950 (±0.945)	6.050 (±1.468)	7.550 (±1.099)	6.650 (±1.387)	8.600 (±1.046)	9.550 (±0.605)
Dermatology doctors: BACK	1.917 (±0.289)	1.750 (±0.866)	3.667 (±0.651)	4.333 (±0.651)	5.417 (±0.900)	6.083 (±0.669)	5.833 (±1.193)	7.083 (±0.996)	8.500 (±1.168)
Dermatology nurses: BACK	2.250 (±0.707)	3.000 (±0.535)	4.000 (±0.535)	4.750 (±0.707)	5.750 (±0.886)	6.125 (±1.126)	6.000 (±1.309)	7.250 (±0.886)	8.500 (±1.069)
General nurses: BACK	1.800 (±0.447)	2.400 (±0.894)	5.000 (1.871)	6.000 (±1.414)	7.400 (±1.342)	7.400 (±2.074)	7.000 (±1.871)	8.400 (±0.894)	10.000 (±0.000)
Medical students: BACK	1.600 (±0.699)	2.400 (±0.699	3.700 (±0.675)	4.400 (±0.516)	5.300 (±0.823)	6.700 (±1.252)	6.400 (±0.966)	7.000 (±0.943)	9.300 (±0.675)
Administrative staff: BACK	2.400 (±0.894)	2.600 (±0.894)	4.200 (±1.095)	4.800 (±1.095)	5.800 (±1.095)	6.400 (±1.517)	7.000 (±0.707)	7.400 (±1.140)	9.800 (±0.447)
Patients: BACK	1.850 (±1.089)	2.550 (±1.572)	4.550 (1.146)	5.400 (±1.095)	6.450 (±1.234)	7.450 (±0.759)	7.600 (±1.273)	8.500 (±1.000)	9.700 (0.571)

Figure 7.7. *Study 1:* Respondents' mean scores compared to actual BSA % coverage.

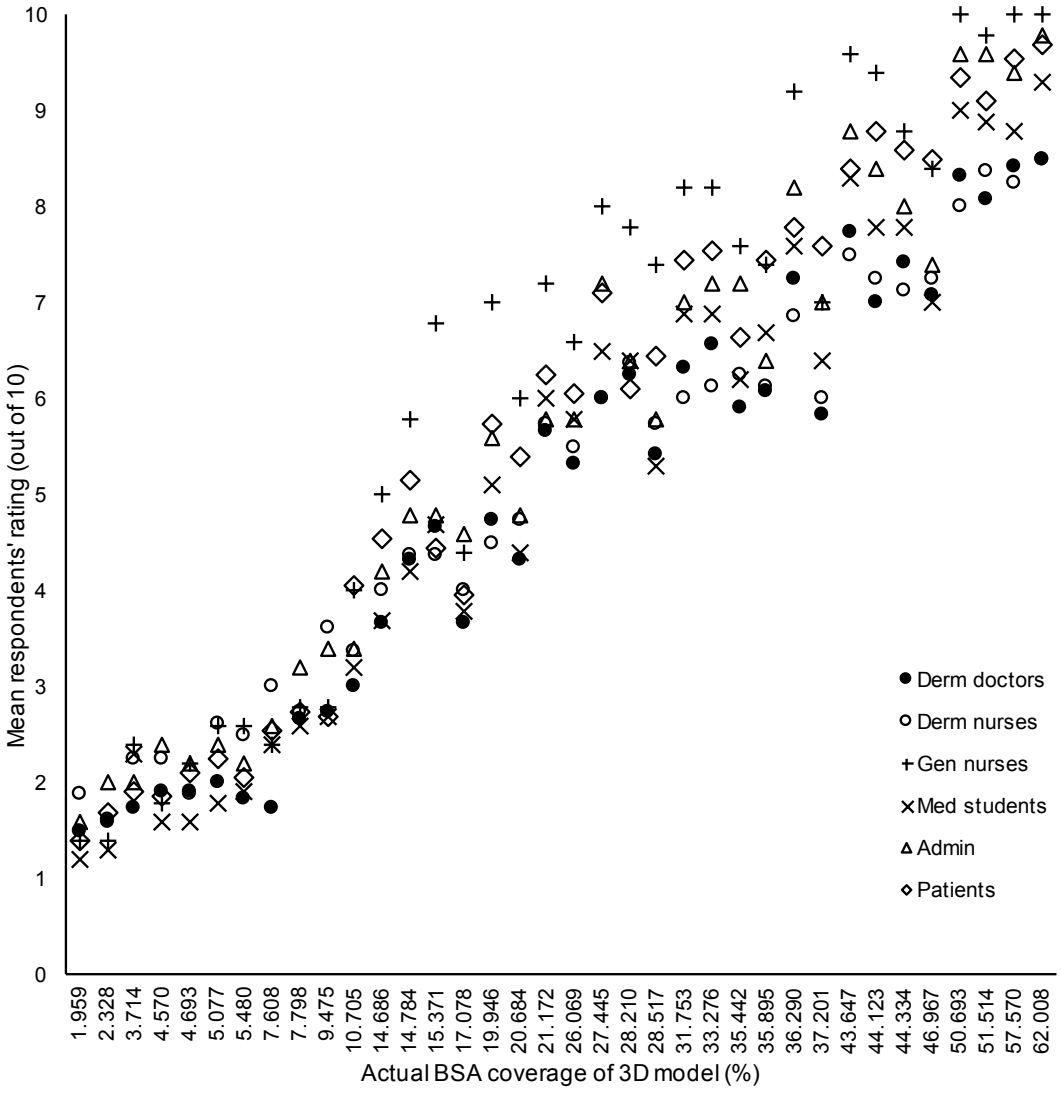


Table 7.6. *Study 1:* Mean group scores (out of 10) for each pose and ANOVA).

Groups	Count	Sum	Mean	Variance	F	P-value
Front Admin	45	254	5.644	8.598	1.516	0.183
Front Derm doctor	108	552	5.111	6.492		
Front Derm nurse	72	375	5.208	4.956		
Front Gen nurse	45	287	6.378	10.468		
Front Med student	90	471	5.233	8.383		
Front Patient	180	970	5.389	8.831		
Half front Admin	45	250	5.556	7.889	2.732	0.0189
Half front Derm doctor	108	512	4.741	5.558		
Half front Derm nurse	72	342	4.750	5.937		
Half front Gen nurse	45	273	6.067	9.882		
Half front Med student	90	449	4.989	7.427		
Half front Patient	180	998	5.544	8.484		
Half back Admin	45	250	5.556	7.298	1.937	0.0867
Half back Derm doctor	108	529	4.898	6.261		
Half back Derm nurse	72	364	5.056	4.842		
Half back Gen nurse	45	273	6.067	9.200		
Half back Med student	90	466	5.178	7.114		
Half back Patient	180	1011	5.617	7.925		
Back Admin	45	252	5.600	6.018	3.433	0.005
Back Derm doctor	108	535	4.954	5.278		
Back Derm nurse	72	381	5.292	4.322		
Back Gen nurse	45	277	6.156	8.134		
Back Med student	90	468	5.200	5.960		
Back Patient	180	1081	6.006	7.492		

Figure 7.8 shows each respondent groups' mean severity score compared to the actual severity (BSA % coverage) and the regression of assessed severity on actual severity. All groups demonstrated a correlation coefficient of greater than 0.94 and the group showing the strongest relationship with the least spread around the mean was 'Dermatology Doctors.' The group with most variance was 'General nurses' which included nurses from other specialities and non-specialist trained Health Care Assistants.

Linear regression of subject's severity assessment on actual disease coverage demonstrated a median gradient 0.142 (range 0.117-0.154, SE slope ±0.006; median intercept 1.94 (range 1.621-2.266); analysis of variance for regression $p<0.001$; R square 0.875-0.951. Table 7.7 demonstrates the correlation coefficients and linear regression analysis of assessed severity compared to actual severity (BSA%).

Figure 7.8. *Study 1:* Respondent groups' mean severity score (standard deviation error bars) compared to the actual severity (BSA % coverage).

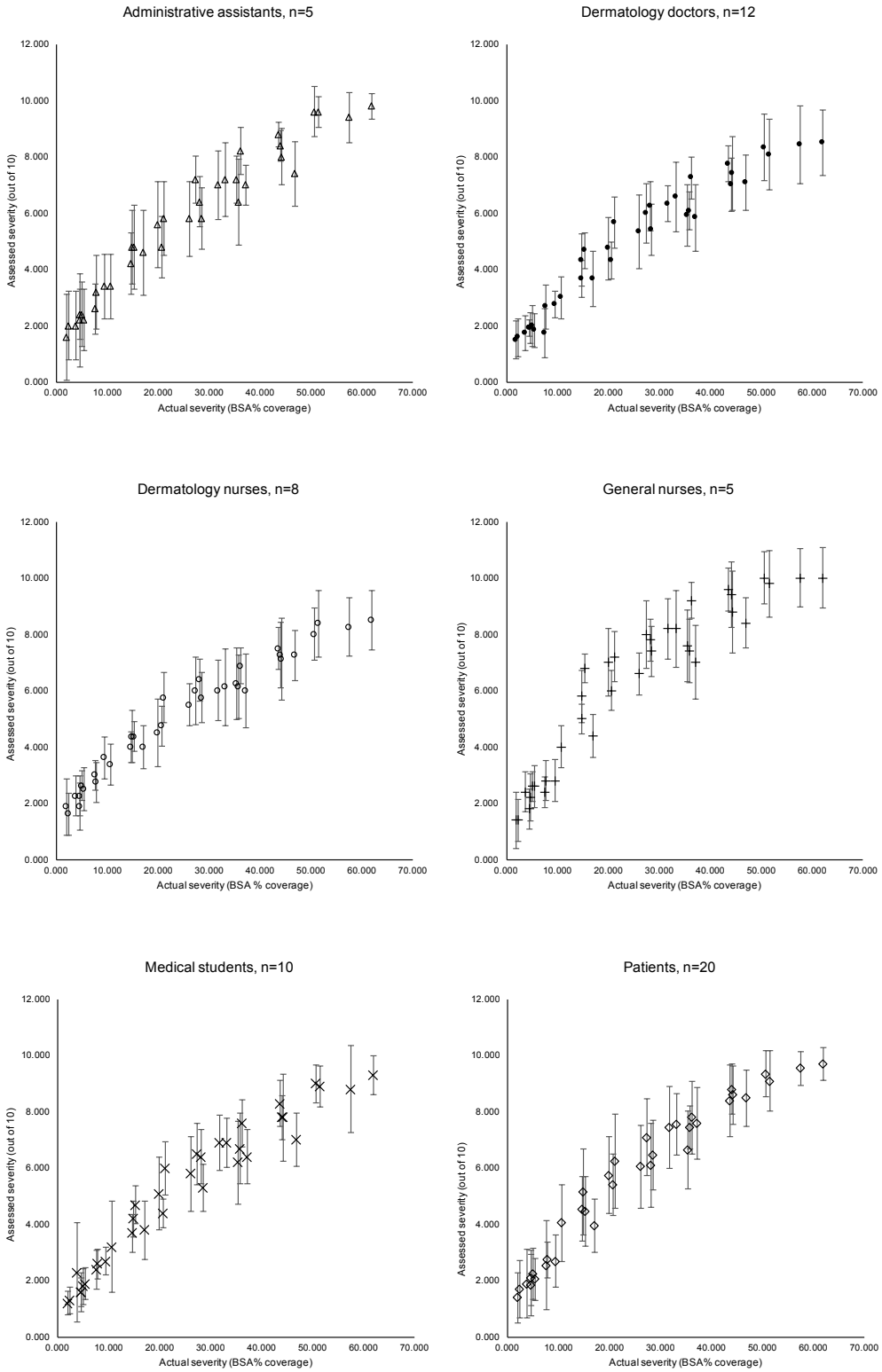


Table 7.7. *Study 1:* Correlation coefficients and linear regression analysis of assessed severity compared to actual severity (BSA%).

	R	R square	Standard error	Intercept	Slope	Standard error of slope	p-value
Administrative staff	0.973	0.946	0.594	2.004	0.142	0.006	<0.001
Dermatology doctors	0.967	0.936	0.589	1.712	0.127	0.006	<0.001
Dermatology nurses	0.975	0.951	0.468	2.125	0.117	0.005	<0.001
General nurses	0.936	0.875	1.03	2.266	0.154	0.010	<0.001
Medical students	0.968	0.937	0.645	1.621	0.141	0.006	<0.001
Patients	0.971	0.942	0.646	1.882	0.149	0.006	<0.001

Comparing how the different groups assign the models categories: mild, moderate and severe

The comparison of selection of models for 'mild,' 'moderate' and 'severe' disease was limited to dermatology practitioners (doctors and nurses) in comparison to the psoriasis patients. The results are summarized in Figure 7.9 and Table 7.8.

Figure 7.9 contains a clustered bar chart of all groups' mean responses and the standard deviation. It can be seen that the results/responses are broadly the same across the respondent type. The graph illustrates that the variance is more restricted in the dermatology-trained (doctors and nurses) than in the untrained.

Table 7.8 summarises the means and ANOVA of the responses across the groups. The difference in response depending on the type of person is significantly different for mild, moderate and severe disease. As mentioned earlier, the groups were not matched for many features including age and GHQ response and so this may not be considered surprising.

Figure 7.9. *Study 1:* Clustered bar chart to show responses for mild, moderate and severe disease for 'front', 'half front', 'half back' and 'back' poses. Error bars represent standard deviation.

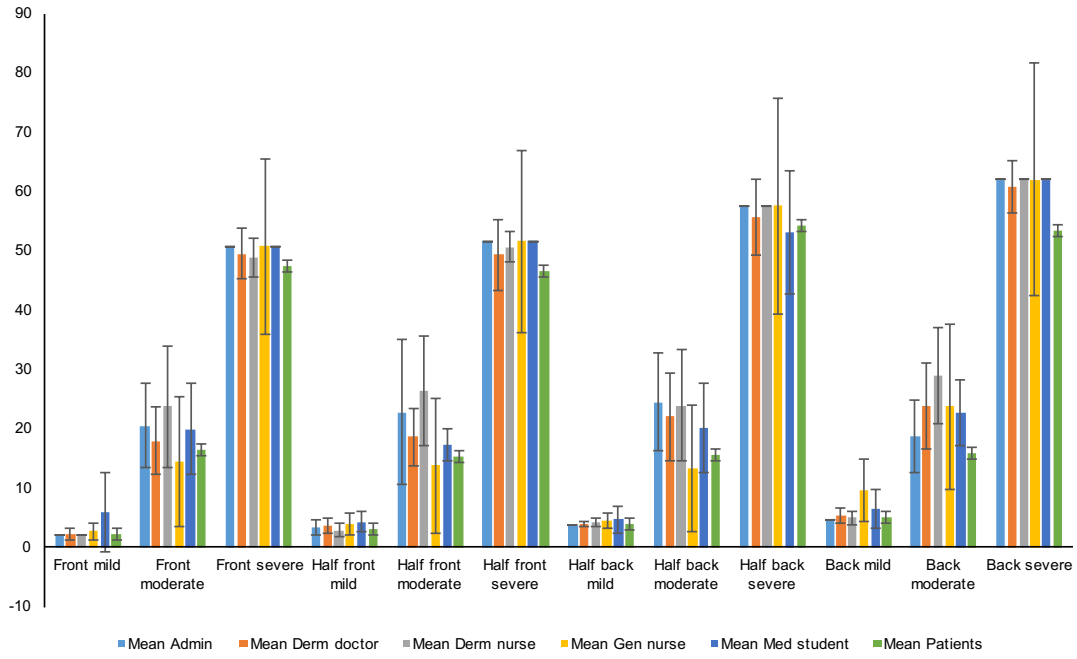


Table 7.8 *Study 1:* Mean responses for each group compared for severity (ANOVA).

Groups	Count	Sum	Mean	Variance	F	P-value
Mild Admin	20	67.588	3.379	1.289	5.582	<0.001
Mild Derm doctor	48	163.388	3.404	1.115		
Mild Derm nurse	32	112.432	3.514	2.021		
Mild Gen nurse	20	103.532	5.177	17.675		
Mild Med student	40	213.988	5.350	15.626		
Mild Patient	80	284.507	3.556	1.949		
Moderate Admin	20	431.565	21.578	69.290	8.804	<0.001
Moderate Derm doctor	48	989.612	20.617	43.761		
Moderate Derm nurse	32	822.456	25.702	82.197		
Moderate Gen nurse	20	326.490	16.325	83.188		
Moderate Med student	40	800.007	20.000	38.760		
Moderate Patient	80	1265.632	15.820	65.151		
Severe Admin	20	1108.930	55.447	22.530	2.900	<0.001
Severe Derm doctor	48	2582.708	53.806	49.485		
Severe Derm nurse	32	1752.803	54.775	32.791		
Severe Gen nurse	20	1108.930	55.447	22.530		
Severe Med student	40	2173.125	54.328	45.732		
Severe Patient	80	4032.394	50.405	123.612		

The data in Table 7.9 is the comparison of assessment for a model to represent mild, moderate and severe disease for the dermatology-trained (doctors and nurses) versus patients. There is no significant difference when each group assesses mild disease: means 3.45% and 3.56% respectively ($p=0.137$) with the same range of response, 1.96% to 7.61%. There is a large and significant difference ($p<0.001$) when the groups assess moderate disease: the mean for dermatology-trained is 22.65% as opposed to 15.82% for patients. The ranges of selected disease for moderate are also quite different and significantly so: the dermatology-trained chose 9.48% to 46.97% and patients chose a range of 4.69% to 51.51%. The assessment of severe disease is just significantly different ($p=0.040$) for the dermatology trained (mean is 54.19%) versus patients' (mean is 50.40%). The lower threshold for severe disease for the dermatology-trained was 31.75% and for patients, much lower at 14.686%. The maximal disease choices were the same for the dermatology-trained and patients (62.008%).

Overall, the models and extent selected for 'mild' disease were the same for dermatology practitioners compared to patients. The range of models considered representative for 'moderate' and 'severe' disease was larger for both dermatology practitioners and patients, although the range was larger for the patients (not to a significant level, however). With respect to the selection of the models representative of 'severe' disease, patients had a significantly lower threshold compared to dermatology practitioners when patients selected compared to dermatology staff selecting (t test: $p=0.012$).

Table 7.9. *Study 1:* Comparison of dermatology staff to patients' assessment of %BSA: mild, moderate and severe disease (t test).

Group	Mean	Range	T test, p=
Mild dermatology staff	3.44775	1.959 to 7.608	0.137
Mild patients	3.556335	1.959 to 7.608	
Moderate dermatology staff	22.65085	9.475 to 46.968	<0.001
Moderate patients	15.820405	4.693 to 51.514	
Severe dermatology staff	54.1938875	31.753 to 62.008	0.040
Severe patients	50.404925	14.686 to 62.008	

Study 2: ***Do standardised 3-D images of psoriasis and photograph prompts improve self assessed disease extent?***

Table 7.10 details the raw scores from this part of the study. As would be expected, since the subjects were undergoing treatments (phototherapy or starting systemic treatment), the validated disease extent and subjective score- measures show improved disease, or reduced extent over time. The raw scores and means are detailed in Figure 7.10.

The subjective disease scores improved over time. The DLQI was significantly reduced over time: the start of the study mean score was 15.9 and the end of study mean 8.1 ($p=0.001$), suggesting improved quality of life. SF-36 scores were significantly increased, indicating an improvement in quality of life: start mean was 54.7 and end mean was 69.9 ($p=0.012$). The GHQ mean was not significantly different but rose, suggesting improved well-being: mean at the start 1.9 and at the end 2.3 ($p=0.111$). The mean VAS of extent fell significantly from 6.3 to 2.8 ($p=0.001$).

Measure of disease extent fell over the treatment period. The PASI decreased significantly, the start mean was 8.9 and the end mean 4.9 ($p<0.001$). Extent as measured by self-comparison of patients to the 3-D models also fell significantly overall from the start of the study (mean 21.8) to the end (mean 9.0), $p<0.001$.

This seems to be a pertinent point at which to present information about the 3-D models' psoriasis extent scores (not PASI *per se,* as there was no data to score for close-up assessment of the plaques such as scaling, etcetera from the models but extent measured as if PASI). My (blinded) extent scores of the models correlate positively (as expected) and linearly with actual extent. Regression analysis yields an R of 0.974, R square of 0.948 with an intercept of 4.42 and a slope of 0.424, $p<0.001$. The data is summarised in Figure 7.11.

Table 7.10. *Study 2:* Respondents' raw scores. '1' represents Day 0's assessments and '2' represents Day 42's assessments. Comparison between two groups, t test (paired) and between selections 1-5 by ANOVA.

Subject	Age	DLQI 1	DLQI 2	SF-36 1	SF-36 2	GHQ 1	GHQ 2	VAS extent 1	VAS extent 2	PASI 1	PASI 2	Part 1		Part 2		
												Selection 1	Selection 2	Selection 3	Selection 4	Selection 5
1	40.0	20.0		68.7		3.0		7.5		11.1		62.0	28.5			
2	45.0	26.0	25.0	27.7	70.0	2.0	3.0	7.8	4.6	11.1	9.0	47.0	28.5	28.5	50.7	36.3
3	28.0	22.0	2.0	31.5	42.4	0.0	2.0	9.4	2.0	7.9	3.2	7.6	4.6	4.6	4.6	4.6
4	34.0	27.0	5.0	50.3	87.7	2.0	3.0	5.3	0.8	10.2	1.8	28.2	21.2	2.0	2.0	2.0
5	55.0	14.0	10.0	64.3	59.9	2.0	2.0	1.9	1.1	6.2	0.7	4.6	4.6	2.0	2.0	2.0
6	55.0	20.0	22.0	47.3	43.4	3.0	2.0	6.3	8.4	11.3	7.7	28.5	35.9	9.5	9.5	28.5
7	62.0	3.0	1.0	99.0	96.2	2.0	2.0	9.0		7.2		7.6	35.9			
8	30.0	10.0	5.0	59.6	54.9	1.0	1.0	6.3	4.3	14.2	13.8	28.2	21.2	5.5	5.5	5.5
9	21.0	18.0	1.0	68.9	91.6	3.0	3.0	8.6	0.3	4.2	2.2	9.5	37.2	2.0	5.5	21.2
10	27.0	12.0	16.0	95.9	75.0	3.0		3.8	5.0	10.7	6.8	20.7	35.9	9.5	9.5	21.2
11	61.0	17.0	15.0	10.9	31.4	1.0	1.0	9.3	8.9	13.9	13.5	14.7	37.2	14.7	14.7	14.7
12	32.0	8.0		53.5		2.0		7.6		3.2		9.5	15.4			
13	54.0	26.0		27.4		1.0		5.0		5.0		9.5	9.5			
14	59.0	15.0	5.0	55.3	102.9	3.0	3.0	7.2	1.9	14.2	4.2	47.0	28.5	7.6	7.6	4.6
15	31.0	23.0	5.0	44.9	86.8	3.0	2.0	9.6	0.8	7.5	1.8	20.7	21.2	2.0	2.0	2.0
16	44.0	18.0	13.0	39.7	56.3	1.0		3.6	3.0	14.6	7.2	9.5	5.5	5.5	5.5	14.7
17	39.0	8.0	2.0	49.8	87.2	1.0	3.0	1.8	0.3	3.6	0.7	2.0	2.0	2.0	2.0	2.0
18	7.0	5.0	5.0	70.6	90.0	3.0	3.0	3.7	2.0	6.8	2.6	5.5	5.5	5.5	5.5	2.0
19	41.0	15.0	5.0	73.8	49.8	1.0	1.0	8.5	2.5	12.7	5.0	28.5	37.2	2.0	21.2	28.2
20	34.0	13.0	3.0	78.9	91.6	2.0	3.0	8.9	0.8	8.2	0.7	50.7	43.6	5.5	2.0	5.5
21	53.0	13.0	5.0	31.3	39.9	1.0	2.0	1.4	0.2	3.2	2.0	9.5	5.5	2.0	2.0	2.0
Mean	40.571	15.857	8.056	54.722	69.826	1.905	2.250	6.310	2.759	8.905	4.876	21.476	22.124	6.494	8.929	11.588
p=		0.001		0.012		0.111		0.001		<0.001				<0.001		

Figure 7.10. *Study 2:* Day 0 (labelled as '1') and Day 42 (labelled as '2') raw scores and mean for all disease extent and subjective disease measures.

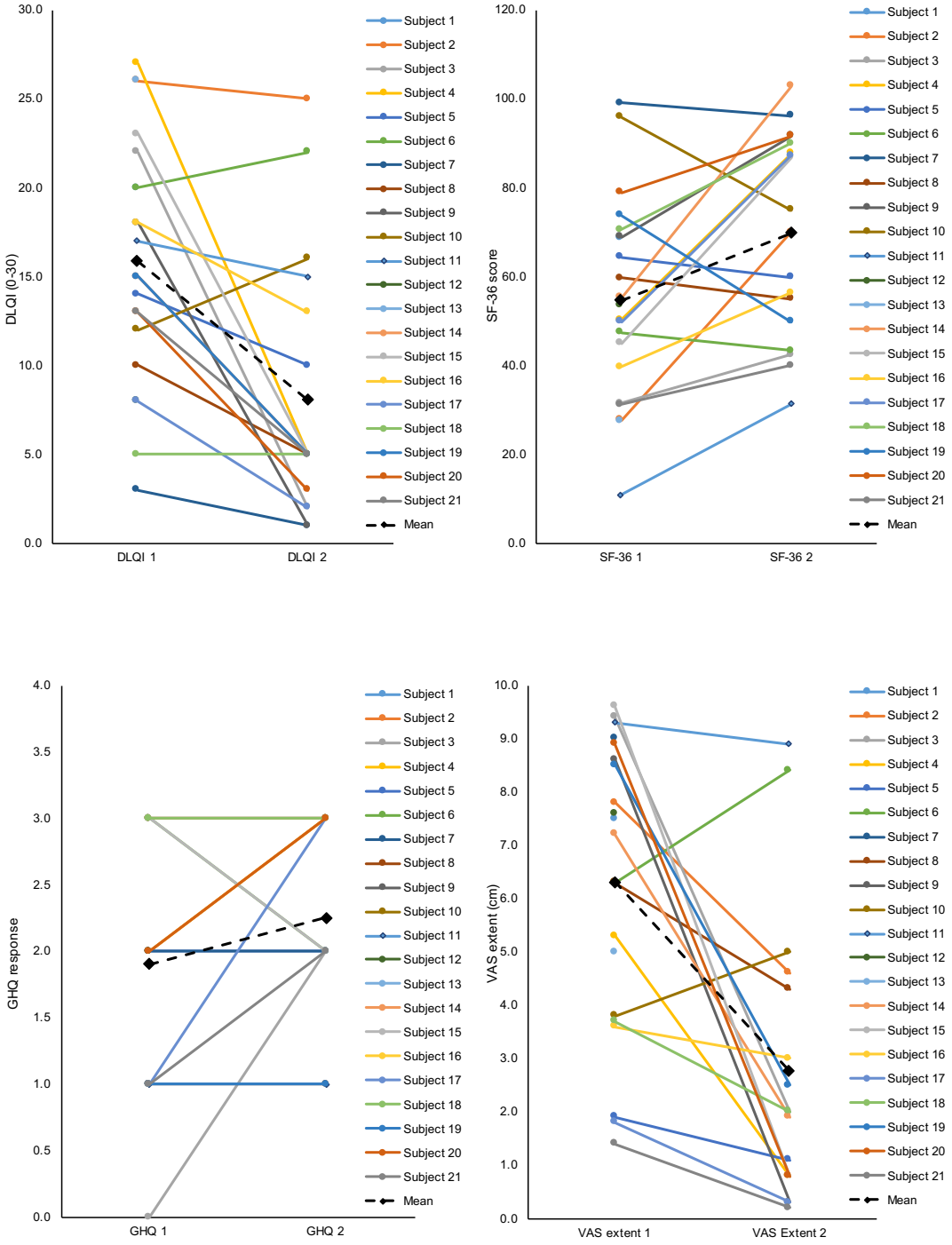


Figure 7.10 *(cont'd)*

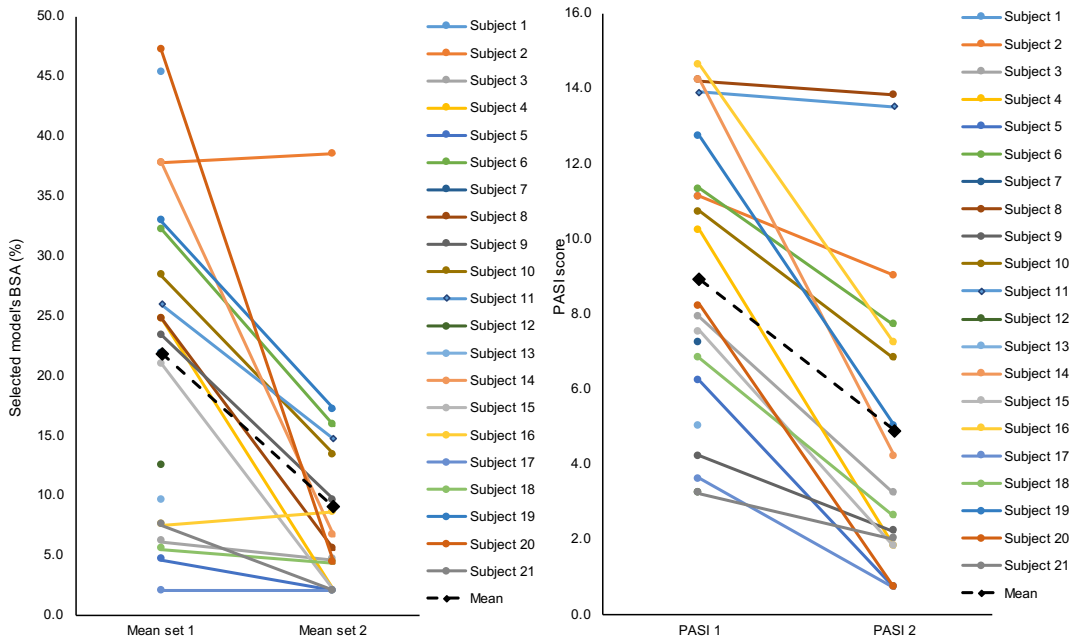
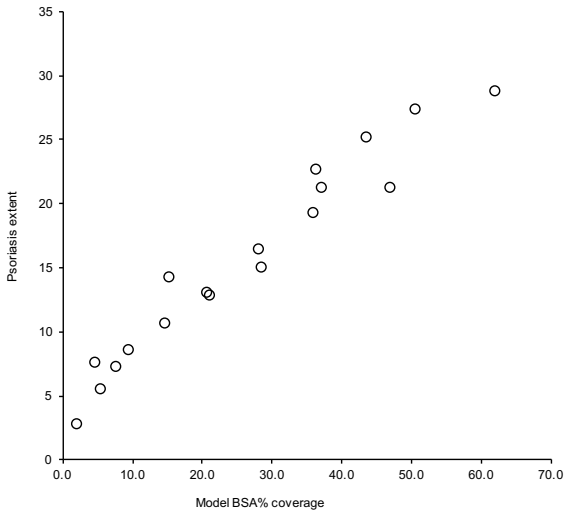


Figure 7.11. *Study 2:* 3-D model score (actual BSA%) compared to investigator blinded extent-score of model.



Selection of models and the effect of clinical photographs

One of the aims of this study was to see if self-expressed disease extent could be improved by showing the respondents photographs of themselves. In part 1, they scored their extent as expressed by comparison to the 3-D models (selection 1) then saw their photographs from that day then scored themselves (against 3-D model) again (selection 2). The theory was that, since perceived disease may be different from reality, the photographs (of reality) should give the patients a prompt and improve scoring.

In part 2, they scored their extent by comparing themselves to the 3-D models (selection 3) then re-scored themselves having seen *that days* photographs (selection 4) and, finally, they were shown their *start of study* photographs and scored themselves again (selection 5). The theory here was that by reminding respondents of how bad they had been at the start and reminding them of their present state may provide a necessary reality-check to improve their perception of their own disease and perception of their improvement with treatment.

Figure 7.12 summarises the results. The graph show that generally the extent scores rise after prompting with photographs. At the start of the study, the mean score prior to seeing the photographs was 21.5 and this rose to 22.0, a non-significant rise, after seeing the photographs ($p=0.110$).

As previously mentioned, the mean photograph-determined extent score significantly fell from Day 0 compared to Day 42 (means overall 21.8 at the start and 9.0 at Day 42).

At the end of the study, the first selected model of extent mean was 6.5. After seeing that day's pictures, the selected model BSA% coverage rose to 8.9 ($p=0.17$). When re-scoring after having seen the start of study photographs, the self-assessed extent rose to 11.6,

p=0.184 when selection 4 was compared to selection 5. The difference in scoring was significantly different (p=0.032) when selection 3 (no photograph prompt) was compared to selection 5 (baseline photograph prompt). See Figure 7.12 for graphical summary.

The results suggest that the photographs manipulate the self-scoring in some way, but whether this was in a helpful way or not was not clear. In an attempt to address this, the relationship of self-assessed disease extent (3-D model BSA% coverage) was examined, The results are summarised in Figure 7.13 and the regression analysis is detailed in Table 7.11. There is a lot of variance and this is only improved when the mean Day 0 and Day 42 scores are plotted against PASI. The photographs appear to cause deterioration in the relationship.

Figure 7.12. *Study 2:* Raw scores (and mean) for model-selected (expressed as BSA%). Selections 1 and 2 for Day 0 and selections 3,4 and 5 for Day 42. Selections 1 and 3 unprompted by photographs from that day. Selections 2 and 4 after seeing that day's photographs and Selection 5 after seeing Day 0 photograph. Grey dotted line separates first part from second part of the study and represents 42 days having elapsed.

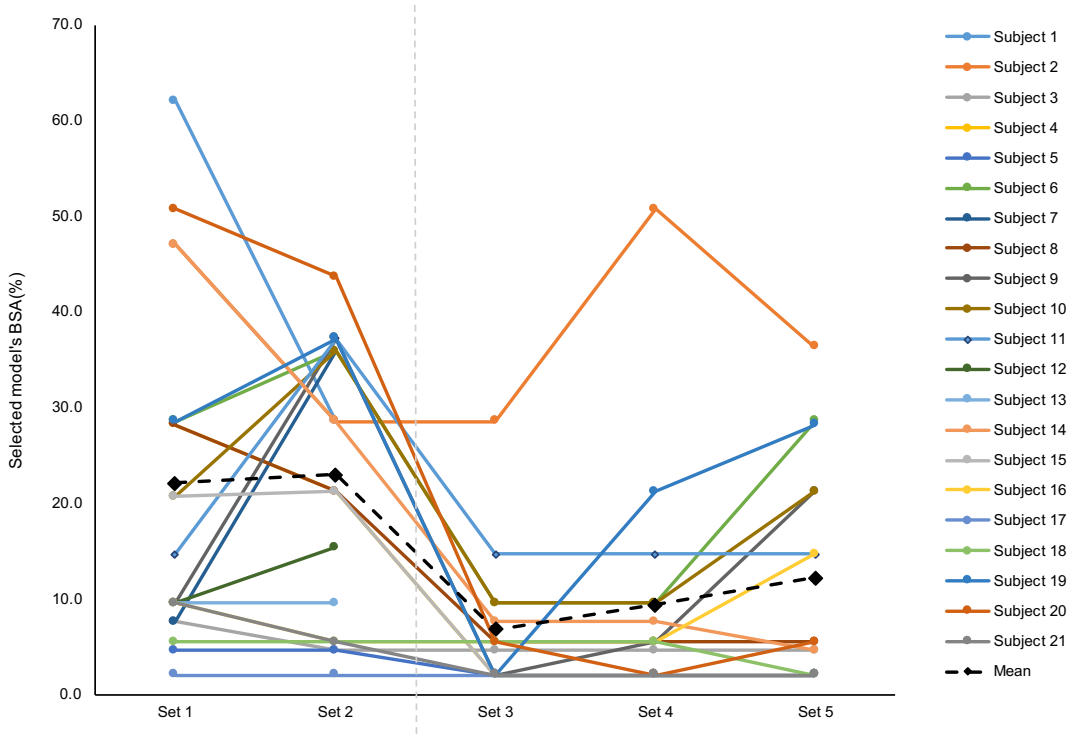


Figure 7.13: *Study 2:* Scatterplots of self-assessed extent by 3-D model (BSA%) on PASI (investigator scored). Selections 1 and 3 unprompted by photographs from that day. Selections 2 and 4 after seeing that day's photographs and Selection 5 after seeing Day 0 photograph.

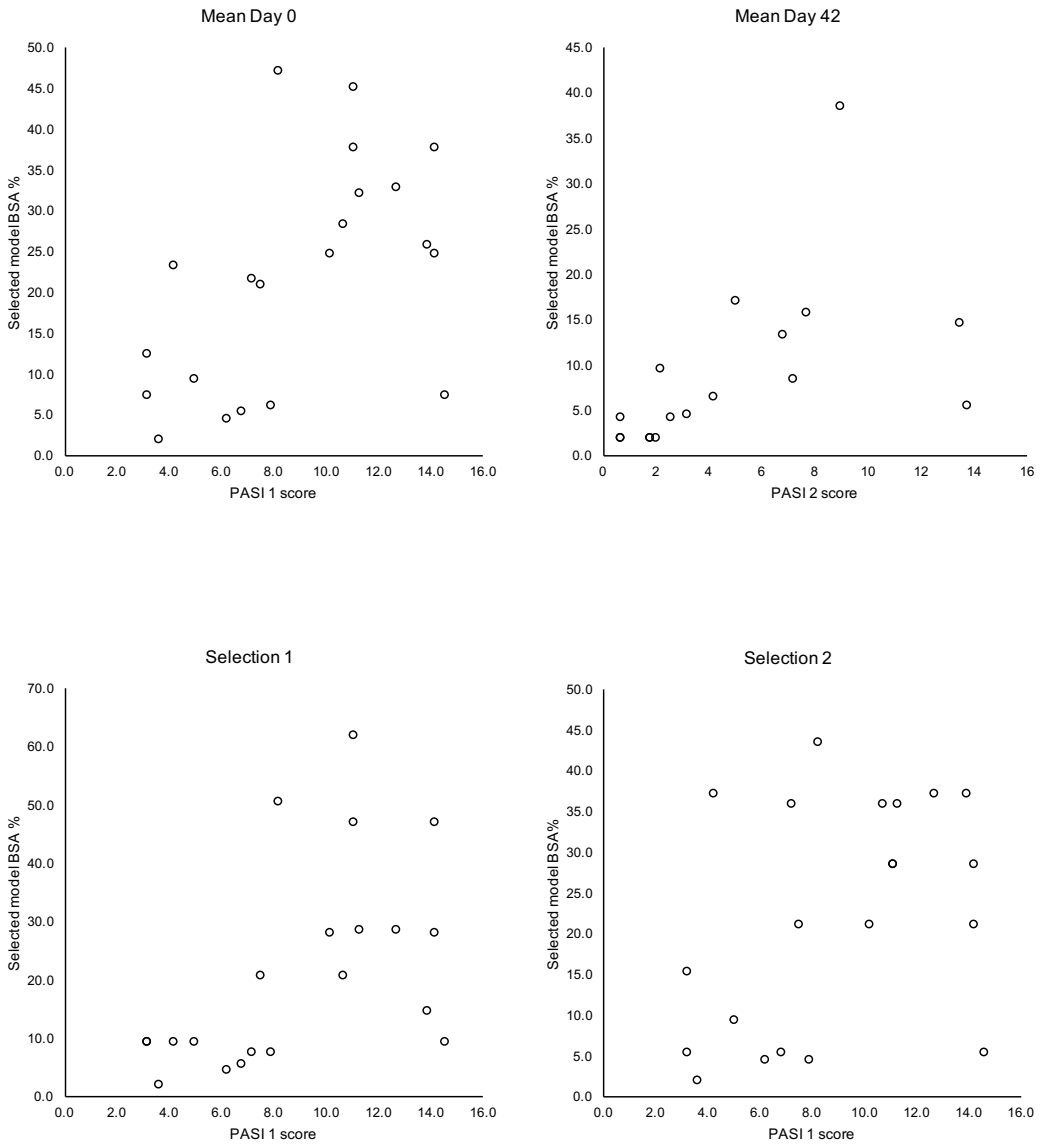


Figure 7.13. *Study 2: cont'd*

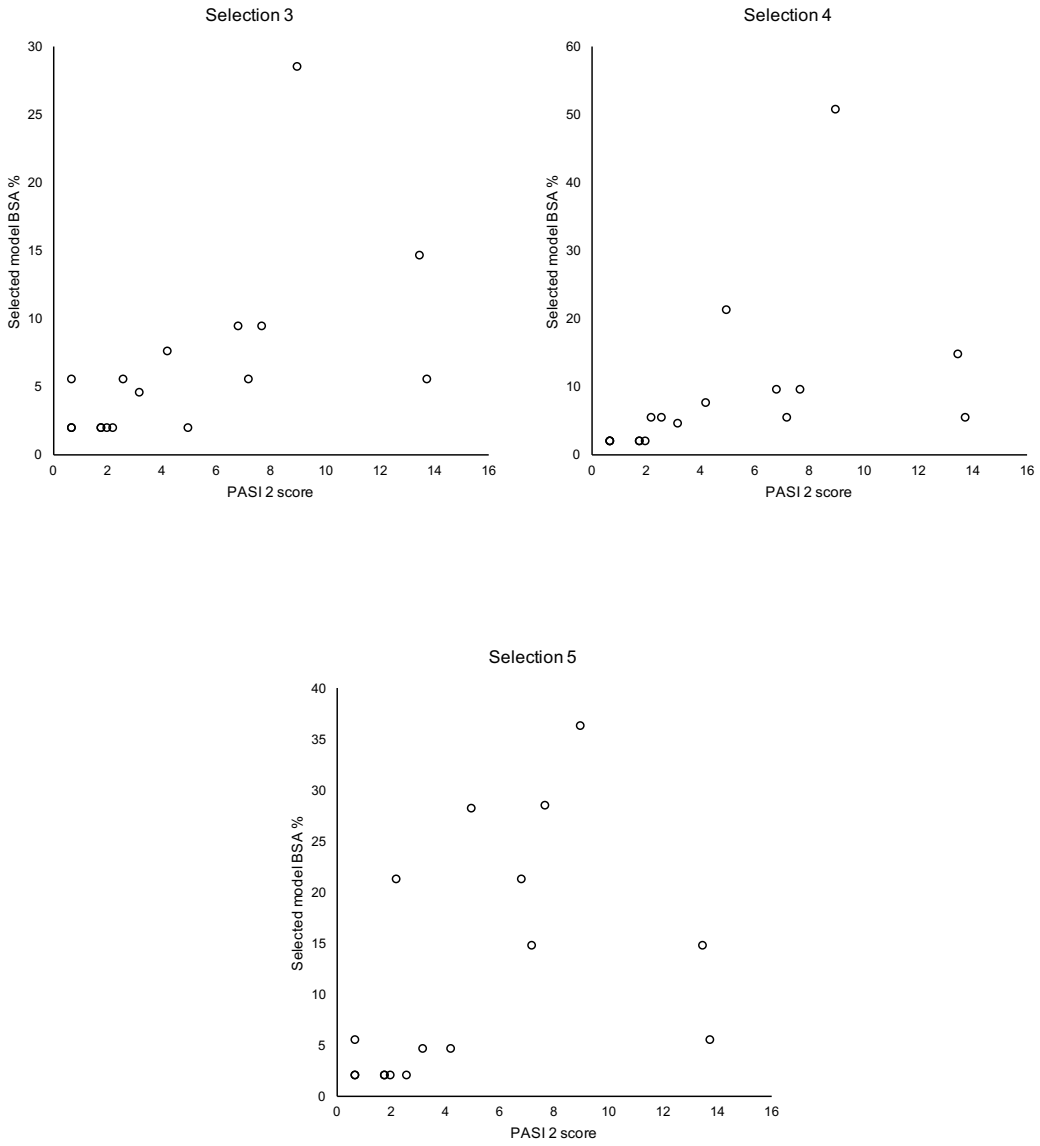


Table 7.11. *Study 2:* Regression analysis of self-assessed extent by 3-D model (BSA%) on PASI (investigator scored). Selections 1 and 3 unprompted by photographs from that day. Selections 2 and 4 after seeing that day's photographs and Selection 5 after seeing Day 0 photograph.

	R	R square	Standard error	Intercept	Slope	Standard error of slope	ANOVA p=
PASI 1 vs Selection 1	0.526	0.276	15.189	0.333	2.374	0.881	0.014
PASI 1 vs Selection 2	0.402	0.162	13.100	9.160	1.456	0.760	0.071
PASI 1 vs Mean Day 0	0.531	0.282	12.074	4.746	1.915	0.701	0.013
PASI 2 vs Selection 3	0.583	0.339	5.631	1.965	0.929	0.335	0.014
PASI 2 vs Selection 4	0.450	0.203	11.021	2.693	1.279	0.655	0.070
PASI 2 vs Selection 5	0.460	0.211	10.508	5.484	1.252	0.624	0.063
PASI 2 vs Mean Day 42	0.526	0.277	8.103	3.381	1.153	0.481	0.030

Conclusions

In this experiment I have presented realistic, but anonymous 3-D models of psoriasis with, thanks to their basis on graphical display, a known and accurate percentage surface area coverage of disease. This pilot study demonstrates that these figures were easy for different groups to relate to and that the presentation system using Mac OS X, Leopard, user-friendly (even for elderly people who had never used a computer before). The consistency and accuracy of all groups' scores of severity when using the 10-point scale is encouraging.

It was not expected or designed that the 3-D models should provide a linear scale but this seems to have been the case, as I have demonstrated. PASI is not a linear scale (Carlin, Feldman, Krueger, Menter, & Krueger, 2004). The lower scores cluster at the bottom of the range but the spectrum of disease extent that the models represented seems to have been at the linear area of the relationship of extent when measured against body surface area as seen in Figure 7.11. A linear scale should provide a more sensitive measure of extent.

Study 1 demonstrated, that with the aide of the models, all types of people, the dermatology naïve, dermatology practitioners and patients could, almost entirely consistently and linearly score disease extent. Dermatology doctors and nurses tended to score lower BSA, perhaps because they were always comparing the model to the most severe disease in their mind's eye and this level of severity was not a possibility to the non-dermatology respondents. There was a significant difference when groups score were compared across respondent type but there was scope for confounding here as the groups were different for other reasons (age, gender, GHQ). The main variability in scoring was for the 'back' view, which was significantly different across groups, but since the other views were more consistently scored, overall, consistency was the rule.

The beneficial effect of the models, in drawing together different groups' perspectives of disease extent when using the 10-point scoring system should be compared and contrasted to the effect that the models, along with the three-point scale of mild, moderate and severe disease elicits. These comments from respondents allude to different groups' perspectives on disease: from a patient, "They're all severe compared to me"; from a few doctors, "None of them are really severe."

This study demonstrates an increase in the variance in the groups assessed severity once assessors are required to choose from the three categories of mild, moderate and severe disease. Patients and the dermatology-trained were in alignment regarding mild disease. Patients considered disease severe at a lower threshold than the dermatology trained – not surprising as the dermatology practitioners are working in a secondary and tertiary care setting. The concept of moderate disease is, I think, quite obscure and this is reflected by the large variance in scoring for this severity. Again, patients had a lower threshold for this than the dermatology trained individuals. It could be argued that the concept of 'moderate disease' is an abstract phenomenon even to dermatology practitioners (and therefore more prone variation) and irrelevant to patient groups. Any marked Likert scale may not be an ideal scoring method as has been demonstrated in experimental itch studies (Wahlgren, 1995). In view of this study's findings, I would suggest that the use of categories such as mild, moderate and severe disease is avoided.

Study 2 put the 3-D models into the clinical situation. Patient-assessed disease extent (using the models) was compared to doctor-assessed extent (using PASI). It has previously been shown that there is a poor correlation between doctor and patient assessed disease extent (Jacobson & Kimball, 2004). Another study suggested that the SAPASI, or self-assessed PASI bore a close resemblance to doctor assessed PASI but, in the study which mooted this

(Sampogna, Sera, Abeni, IDI Multipurpose Psoriasis Research on Vital Experiences (IMPROVE) Investigators, 2004), the extent of disease was based on the doctor's assigning a percentage to the patients' shaded area on a line-drawing and so any inaccuracy in assessing percentage surface area would not be evident.

The clinical part of our study demonstrated that a significant decrease improvement in disease after six week's worth of treatment. This was evidenced by the usual validated subjective scores (DLQI and SF-36) improving significantly and disease extent scores, doctor's, i.e. PASI or patients', i.e. VAS extent improving significantly too. The self-scoring as expressed by BSA% from the 3-D model selection improved significantly too.

It was suspected that the patients' perception of disease extent would alter if they could see their clinical photographs. This suspicion grew from a clinical observation that, in Phototherapy mid-way review clinics, patients commonly expressed surprise at how much they had improved after they saw their 'Start' photographs and from appreciating the potential effects of the 'hedonic treadmill' (Kahneman et al., 1999). In fact, our study demonstrated that our subjects thought that they had significantly more disease at week 6 after seeing their 'Start' photographs (the opposite of what we expected). The correlation coefficients between patient and doctor assessed disease weaken and become less significant after patients see their photographs, this would suggest that it is better, therefore, to rely on a patient's 'gut instinct' of just how extensive their disease is.

One major limitation of this study is that the subjects in the clinical part of the study all had to be female (in order for them to be able to relate reliably to the female-form models). It has been shown that there is a gender difference in the perception of disease extent (R. K. Roenigk & Roenigk, 1978; Schmid-Ott, Jaeger, Kuensebeck, Ott, & Lamprecht, 1996) and so it would be valuable to develop male models and repeat the study with male subjects.

Some may also consider the fact that the system is not 'dynamic' to be a limitation. This may or may not be the case. It would be of great interest to compare a patient's, or doctor's perception of disease if a computer model existed which allowed the disease extent to be 'drawn on' and the pixel coverage calculated. However, the model may not be useful clinically if the perspective prejudice was reflected in the 'drawing on process.' Apart from this, there would be problems relating to accuracy as the model would not have the same surface area as the subject and accuracy of representation on the model may not be realistic or achievable (in the clinic scenario) as it is time consuming to accurately transcribe each actual plaque onto a model.

The most severe model actually only had a disease surface-area coverage of 62%. It would appear that overestimation of severe disease is commonplace (see previous descriptions(Ramsay & Lawrence, 1991)) and is possibly unavoidable if we continue to estimate it in the same way. The textbook definition of erythroderma requires greater than 80% surface area coverage. Having developed these models, my perspective of body surface area has changed too: I genuinely doubt that the definition of erythroderma, with 80% BSA coverage can be realistic now that I appreciate what 80% coverage really means. My experience is that patients are routinely described as erythrodermic who are not 80% covered in disease. Part of the issue, as illustrated by the Tiling-Grosse (Tiling-Grosse & Rees, 1993) and Lawrence (Ramsay & Lawrence, 1991) papers, is that many small plaques increase the error in judging surface-area. I would propose that small-plaque models were developed to use accordingly.

With the ready availability of photographic services in hospitals, it may appear sensible to employ computerised image analysis to quantify disease extent, however this analysis is still unreliable, despite advances in technology. The main problem is that this system involves 3-D analysis of a 2-D image photograph (Ashcroft et al., 1999; Savolainen et al., 1998)

Secondly, there is lack of distinction between diseased and non-diseased skin and this is complicated by the effects of shadow, the optical effect of scaling etc. so a straightforward computer algorithm to 'tell psoriasis from 'normal skin' is not yet achievable (Ahmad Fadzil, Ihtatho, Mohd Affandi, & Hussein, 2009). Photographic systems exist where there are cameras mounted 'in the round' which capture an image and then software 'morphs' them into a digitised 3-D representation but the systems are not perfect and, commonly, 'gaps' appear in the image and their availability is limited. It may seem remarkable that, in the days where 3-D printing is quite readily available we can not access a similar technology for psoriasis but a certain amount of extrapolation is required for these algorithms to work and this would not be appropriate in a clinical or, most particularly, a trial situation.

It is important to 'have a handle' on patients' perspective on disease extent, since, whilst a study may demonstrate the chances of reducing the percentage PASI with a certain treatment, if the difference is not perceptible or worthwhile to the patient, can or should it be considered useful? If the assessment and response to treatment of a disease is not viewed from the same perspective, patient and doctor groups may be communicating at crossed-purposes meaning dissatisfaction for both parties. Such accurate assessment would also be sensible in making decisions regarding the distribution of wealth in the health services.

In summary, in a collaboration with the Edinburgh College of Art, I have developed a novel system of psoriasis severity models, using graphical display. I have shown that respondents are able to relate to the models and that, the models' use, along side a 10-point scale, results in consistent and accurate scoring of disease-extent. Importantly, the combination of the models and scale seems to unite respondents' perspective on the disease. I have demonstrated that patients can, quite effectively, score their self-disease extent using the 3-D models I developed and that their own photographs were not necessarily a benefit to them, if compared to doctor assessed PASI.

Summary Conclusion

Since conclusions for separate experiments are placed throughout the thesis, for easier accessibility, I will summarise them here.

Chapter 1: Direct observation of subjects in order to quantify and characterise itch-related movement.

Direct observation and recording of movement remains the 'gold-standard' for quantifying itch-related movements. I developed an improved infra-red videoing system which allowed recording of individuals, in their own homes, for up to twelve hours. Using this system, I made recordings of children with atopic dermatitis and characterised the itch-related movements.

Nocturnal movements were categorised into: generalised movement, which is not specifically itch-related; rub, rhythmical movements where fingernails are not the point of contact with the skin and scratch, rhythmical movements where fingernails make contact with the skin.

All subjects demonstrated the itch-related movements but were also noted to be generally restless. The mean frequencies of action for each movement were: generalised movement, 0.48Hz (SD±0.25); rub, 0.98Hz (SD±0.36) and scratch, 1.85Hz.

These results concurred with previous suggested frequency of action of scratch in humans derived from electronic data in subjects (Felix & Shuster, 1975) and in scratch simulated by

adults (Aoki et al., 1980). The study achieved its aim of characterising human itch-related movement and demonstrating that it is possible to separate itch-related movement from generalised movement on the basis of frequency of action.

In order to put the results in context, and because detailed information about frequency of action of scratch was available for mice (Brash et al., 2005), in a small experiment I observed different mammals scratching behaviour. Frequency of action was higher, or the movement faster, in the smallest animals (mouse 12.5Hz) and lower or slower in larger animals (gorilla largest, 1.3Hz followed by humans, 1.85Hz). I postulated that the frequency of action related most closely to limb length, but standardised measurements for this were not available. The frequency of action did correlate closely with standard mass and standard length of the animal, the relationship was not linear, but was best expressed by a power relationship. The closest correlation was between the mass of the animal and frequency of action.

Chapter 2: Introduction and further validation of the use of accelerometers to measure itch-related movement.

Having improved the data extraction from these devices, by writing new macros and having checked functionality with regard to using the devices to record for longer than one night, I performed validation studies of these devices against infrared videoing in short consecutive night studies. These studies set about checking if the night-to-night variation in score previously noted was a 'real' phenomenon.

It was found that the variation in score was 'real' when validated against videoing but that, in the early evening and morning this was likely to be due to inaccuracy in the subject-recorded sleep and wake-times. This phenomenon had previously been reported in accelerometer sleep-studies (van den Berg et al., 2008) but not validated against video-

recordings. There was good correlation in Actiwatch Plus detected movement and video recorded, directly-observed movement during sleep. No particular patterns in activity were found – these had been examined as it had previously been mooted that sleep stage may determine scratching activity (Savin et al., 1973; 1975). As a result of this chapter's findings, much longer longitudinal studies were proposed, but also, it was decided to see if newer technologies could improve upon the Actiwatch Plus's motion capture.

Chapter 3: Validation of a newer accelerometer: DigiTrac

A newer accelerometer became available. Due to improvements in technology, it was able to store more detailed data and sense in more axes such that 'frequency of action' data would be extractable. Since itch-related movement had been proven to be rhythmical and have a certain range of 'frequency of action' (Chapter 1), this device was attractive.

Once I acquired some of these DigiTrac accelerometers, I had to perform some basic validation tests. The machines were found to be acceptably precise when compared against each other: the coefficient of variance (CV) was highest at low acceleration (14.3%) and lowest at high acceleration (6.8%). Since the industry standard for CV is 20%, these findings assured precision.

I then moved on to assess the DigiTrac's ability to capture simulated itch-related movement. FFT analysis of the DigiTrac output clearly demonstrated scratch and rub movements. Encouragingly, the peak activity on FFT corresponded with the 'frequency of action' of itch-related movement directly observed in actual subjects (so not simulated scratch) in Chapter 1.

Before the DigiTrac could really be put into clinical experimental use, I had to determine which measuring axis of the three available, I should use to best detect itch-related

movement. In actual fact, there was not much difference between the ability to capture itch-related movement in any axis, so just in the interests of consistency, I proposed that the X axis be used from then on.

Finally, since I felt that walking was the most likely rhythmical movement to reduce specificity of itch-related movement overnight, this movement was characterised on DigiTrac recordings and compared to itch-related movement. It was found that walking had a different FFT profile to itch-related movement, but that most FFT activity was at 2-3Hz and since itch-related movement has a range of 0.5-2.5Hz there was potential for confusion (or confounding). The study highlighted that discrimination of movement could also be improved by considering the 'frequency of action' and also the *magnitude* of activity, something which had not previously been contemplated.

Chapter 4: Clinical experiments using the DigiTrac accelerometers

The DigiTrac were used in itchy subjects to validate and characterise the readout/capture in patients (as opposed to healthy subjects simulating scratch). Six children with atopic dermatitis were infrared videoed at home and wore the DigiTrac. First, an 'itch spectrum' was determined by plotting the amount of time undertaking scratch versus rub versus generalised movement on video. The majority of movement that occurred between 50-550G-s was scratch (but not exclusively scratch). Subjects were motionless at less than 50G-s and movement was just as likely to be generalised or itch-related over 550G-s. Secondly, the DigiTrac were validated against Actiwatch Plus successfully. It was suspected that walking movement may have clouded clarity of the DigiTrac recording: as discovered in Chapter 3, the spectra did overlap.

On the basis of the observations, it was hypothesised that the data should first be 'enriched' for the frequency range 0.5-2.5Hz. Since little or no movement was detected below 50-

100G-s and larger amplitude, non-itch-related movement occurred at greater than 400-500G-s, a second ‘enrichment’ was proposed for an amplitude range: either 100-400G-s or 50-550G-s. Therefore, it was assumed that enriching for this frequency of action and amplitude of movement of DigiTrac should provide the most specific and discriminating measure of itch-related movement and this led to the next study.

This small exploratory study involved itchy adults and children (n=9) with different itchy disorders and compared them to controls (n=12). The DigiTrac FFT outputs were ‘enriched’ for different ranges of filter for ‘frequency of action’ (0.5-2.5Hz and 0.5-5Hz) and ‘amplitude’ (50-550G-s and 100-400G-s), suggested by previous studies’ observations so that the maximal separation of subjects from controls could be determined. Overall, enrichment for both ‘frequency of action’ and ‘amplitude’ improved separation of itchy subjects from controls. There was a marginally superior separation for enrichment by ‘frequency of action’ 0.5-2.5Hz and ‘amplitude’ 100-400G-s.

A larger study was then undertaken. It compared 26 subjects with atopic dermatitis to 26 control subjects on short runs of consecutive nights and one aim was to see if the large night-to-night variation in accelerometer score was also a feature of DigiTrac recordings, i.e., whether it was detected by a specific itch-movement detector (as opposed to a generalised movement detector, the Actiwatch Plus). The study, through scoring disease-extent with SCORAD, also determined the relationship of objective to subjective scores and physician assessed disease-extent.

The results demonstrated that the DigiTrac FFT output could separate itchy subjects from controls. A large variance in score was found for subjects and a smaller one for controls. Layers of enrichment, stepwise first for ‘frequency of action’ and then for ‘amplitude’ improved the separation of itchy subjects from controls. The enrichment seemed to work by

reducing variance in the control group. The greatest separation from the itchy compared to controls was for the combined three-night data. This suggests that lack of power may have been an issue. Power had not been calculated formally as this had been a pilot study.

The study confirmed that there was large night-to-night variation in score, this was found to be a 'real' phenomenon and also confirmed a disconnect between subjective disease-score and objective itch measured as DigiTrac score. These had both been findings of work with the Actiwatch Plus. Furthermore, the objective accelerometer DigiTrac score did not have a good relationship with physician-assessed extent, SCORAD, even when subtracting the subjective score-component.

I took a moment, at this point, to reflect on what I had discovered. I had proven that Actiwatch Plus and DigiTrac accelerometers were both suitable and validated for the purpose of capturing itch-related movement in studies. I had discovered that there was true large variation in objective accelerometer score night-by-night within and between person, but I had only assessed this in small clusters of nights: would this 'even out' or demonstrate a pattern over time? The results also kept telling me that there was a disconnect between the objective and subjective disease scores, again, only in small cluster studies: would these even out or demonstrate a pattern over time. I next had to try to explain and investigate these issues.

Chapter 5: Measuring itch over time in chronic disease

This chapter reports longitudinal studies which addressed a few issues specifically including the night-to-night variation in accelerometer score and subjective/objective score disconnect in some instances over a period of six weeks. For shorter clusters, 68 adults and 50 children with any itchy condition were recruited. The six-week study recruited a separate 20 adult patients with any itchy skin condition and involved subjects wearing the Actiwatch Plus

accelerometers every night for 42 nights whilst scoring VAS of itch and insomnia daily. Other subjective measures of disease, SF-36, DLQI and GHQ were assessed and disease extent was calculated using SCORAD.

The study confirmed the great variation in accelerometer scores: about 60% was between subject and 40% within subject. This finding limits the usefulness of actigraphy in experimental studies, unfortunately. The accelerometer scores were found not to be influenced by age or sex. The main determinant of differences in score was disease type: liver disease most severely affected, followed by other causes of itch including atopic dermatitis and psoriasis. Controls had lower scores.

The study confirmed the lack of relationship between subjective and objective itch scores. This was true for children with eczema and adults with a variety of itchy disorders. The second part of the study afforded the capability to compare the subjective and objective over a much longer time than any other study had allowed, six weeks. Still little convincing evidence of a relationship between the subjective VAS of itch and actigraphy was found. The fact that a strong autocorrelation pattern of scoring was found in the VAS data is interesting: it appears that subjects revolve scores around an individual anchor-point, even without access to the previous day's scores (viz ballot box). This is a recognised 'anchoring' phenomenon(Ariely, Loewenstein, & Prelec, 2003; Kahneman & Sugden, 2005). An effect of lag was found in the autocorrelation of VAS scores. This is difficult to explain. No such autocorrelation effect was found in the actigraphy scores. Overall, these results are not very encouraging for the accuracy or validity of VAS scores.

The use and comparison of other subjective symptom tools answered the question as to whether the disconnect was due to flaws in the VAS, or due to systematic differences in objective and subjective disease measurements. The fact that none of the subjective disease

tools had a convincing relationship with actigraphy, but did with each other, suggests that the disconnect is basically subjective versus objective. SCORAD (±subjective components) did not correlate well with subjective or objective tools.

Trialling the new DRM questionnaire was an interesting exercise. Never having used it before, it was instructive to see that the form took between 30 minutes to 2hrs 30 minutes to complete. This renders it impractical for regular clinic use. Again, DRM responses correlated well with the other validated subjective tools, SF-36, DLQI, VAS, but not with actigraphy or SCORAD (±subjective components). Two intriguing findings surfaced from using the DRM. The first was that the GHQ result worsened at the end of the process compared to the start: this suggests that, when giving a true rather than gut reaction, subjects do feel worse about their health. The difference was not significant but the number of participants was small. The second was that distraction from itch when at work and vice versa was significantly displayed in the scores.

Whilst there were limitations (discussed at the end of this chapter), overall these studies demonstrated that objective scoring using actigraphy may not be perfect but does confer a measure on a certain aspect of itchy disease. Some form of measurement of perception of disease or impact of disease on life should be utilised too but it should be borne in mind that the tools available to us now are not perfect nor tell the whole picture either.

Chapter 6: The effect of bias on subjective disease scores.

Subjective-symptom measures are widely used and most commonly validated by checking for reliability, responsiveness and validity. The subjective symptom tools commonly used in dermatology had not been tested for vulnerability to bias. In these short experiments I tested the effects of focussing and framing biases on Dermatology Life Quality Index (DLQI), Global Health Question (GHQ) and visual analogue scores (VAS).

215 patients with different conditions and moderate disease (as measured by DLQI) were recruited. Three different studies were undertaken: the first required people to read and/or write mood eliciting words prior to completing the subjective disease scores, the second required a group to see a film about a person living with severe disease then complete the tools and, finally, the last study compared the usual DLQI with an amended one rewritten to be more neutrally framed.

None of the interventions made any significant differences to any of the subjective scores. Although a 'negative' result, the findings were reassuring as they suggested that the subjective symptom scores are robust to the effects of focus and framing bias. The results do not mean that the tools are invulnerable to the effects of say, the stress of a clinic running late, liking or not liking the clinician or 'gaming' scores to gain a particular goal, for instance, eligibility for a biologic treatment for psoriasis. The poor relationship between subjective and objective score remained unexplained and whilst the subjective disease tools were impervious to these rather basic bias-manipulations, it should not be assumed that the subjective tools are perfect or sufficient to measure a disease unilaterally.

Chapter 7: Effect of computer-generated 3-D models and photographs on self and third person measurement of disease-extent in psoriasis

A poor relationship between disease-extent scores (SCORAD and PASI) and objective accelerometer score had been determined in Chapters 4, 5 and 6. This encouraged me to scrutinise these disease-extent measurement tools. I discovered that intra and inter-operator variability had been detected despite the use of these standardised tools (Charman et al., 1999; Ramsay & Lawrence, 1991). I also discovered that work had been carried out on estimation of surface area in the trained and untrained (Tiling-Grosse & Rees, 1993). Estimation of BSA seemed likely to be the source of inter-operator variance. The fact that

doctors underestimated disease whilst dermatology naïve medical students were more accurate at gauging BSA made me wonder where patients would fit in, especially in the context of having read that patients were potentially not always aware of their own severity of disease as they had adapted to it over time, 'the hedonic treadmill' (Brickman, Coates, & Janoff-Bulman, 1978; Frederick & Loewenstein, 1999; Kahneman et al., 1999). I decided to try to unify perspective of disease between doctors and patient in an effort to improve communication between the parties.

In a collaboration with Edinburgh College of Art, anonymous 3-D computer models were created which simulated psoriasis. Models of nine different known severities were designed. These models were based on photographs of a range of psoriasis severities and were, quite inadvertently found to increase in severity, expressed as percentage Body Surface Area (%BSA), linearly. PASI is not a linear scale, thus reducing its sensitivity at the lower severity of disease (Carlin et al., 2004).

In a first study, it was demonstrated that all groups of respondents (n=60), dermatology trained and not, and even patients, scored randomly presented models accurately. Doctors tended to underestimate disease and most variance in scoring was found in general nurses. It was found that there was more variation in scoring if, rather a score out of 10 was to be chosen, the respondents were required to assign the models as having mild, moderate or severe disease. This could represent a problem previously reported in Likert scales previously (Wahlgren, 1995). It could also be due to the fact that concepts of mild, moderate and severe disease are more abstract, especially moderate disease which had most variance in scoring.

In the second part of the study, female psoriasis patients, about to commence treatment were required to score their disease-extent with the aide of the 3-D models, with and without the

aide of photographs. The study involved 21 patients with mild to moderate psoriasis and monitoring over six weeks – including PASI, DLQI and VAS. As expected the subjective scores and disease-extent scores improved over time, including the 3-D picture scores. The subjective scores correlated well with each other.

When it came to the effect of the photographs on patient assessed disease, it appeared that showing them reduced corroboration between doctor and patient-score extent. The 3-D models alone, however, appeared to unify perspective on disease (doctor versus patient).

Limitations

Subjects for all studies were recruited from secondary and tertiary care clinics. The subjects were, therefore, skewed towards those with more severe disease, especially those with eczema as evidenced by the SCORAD scores (higher than those typically reported in trials).

The subjects 'volunteered', this selects patients of a different type from the average clinic attender. Some of the studies allowed a significant amount of time with the investigator (Chapter 5), another step away from the general norm. It was not possible to control the severity of disease throughout the study period: patients at the extremes of severity may have responded differently.

In the psoriasis studies reported in Chapter 7, subjects were about to embark upon therapy. There was no way of 'blinding' this event from patient or investigator. As a result, an expectation of improvement may have resulted in bias.

Compliance: assumptions had to be made for accelerometer studies regarding compliance as,

whilst it was sometimes reported by subjects that they had forgotten to wear the accelerometer, subjects did not always mention this. It was obvious from 'eye-balling' the summary scores if the watch had been worn or not and so, if this were the case, since the impetus was on separating disease from controls, these night-scores were excluded. Compliance tables are detailed in with the appropriate study results.

As well as the accelerometers, compliance was required in the subjective form filling. This was easy to confirm if questionnaires were completed in the investigator's presence (SF-36, DRM, GHQ) but not so straightforward for daily VAS scores. Sometimes there were 'honest omissions' forms left unfilled, but I had no way of proving that scores were made daily and not done in batches or without a prompt of the previous day's score – although I tried to encourage daily scores and blinding by providing a ballot box in the Chapter 5 study.

An obvious limitation of the 3-D psoriasis models is that they are female form. This may have had implications for the first part of the study when males and females respondents were required to 'score' the models but based on the fact that this was a third person process, it was expected that it should not be. The female 3-D models also made it imperative that only female patients were recruited to the part of the second study where respondents had to score themselves against the models. The results of this study should not be extrapolated to males without validation, especially since sex differences in the effect of psoriasis exist (R. K. Roenigk & Roenigk, 1978). Finally, the most severe 3-D model only had a disease surface-area coverage of 62% whilst true erythroderma is 80% coverage. This should not have been a problem as long as none of the study participants were erythrodermic (and they were not) but would be a problem if studies/use of the models was to be extended.

All of these studies were, essentially, exploratory. As a result, 'power' could not be calculated. There is a strong possibility, therefore, that with larger groups some differences

would have become formally significant. This is particularly pertinent to the separation of the itchy subjects' from controls' actigraphy scores throughout the thesis and to the 'Effect of bias' experiments in Chapter 6.

Discussion

My research started with the aim of solving the problem of measuring itch. It evolved into aiming to improve measuring disease. Whilst I cannot say that I have entirely solved these problems, I can say that I have made significant inroads into understanding itch, its associated behaviour and its measurement. I have also scratched the surface (genuinely, no pun intended) of understanding the other disease-measurement tools and how they measure another aspect of living with disease. I hope the work will inform further research and help, eventually to success in the goal of accurate measurement.

The studies in this thesis approached measuring disease from the objective angle first. I have shown that a simple accelerometer, the Actiwatch Plus, is a convenient and validated way to capture objective itch-related movement, even over a long period of time - up to six weeks. I have developed and validated the use of another accelerometer, DigiTrac. This device was expected to be superior, as it could detect 'frequency of action' of movement and I had proven that itch-related movement was rhythmical (Chapter 1). In video-studies I characterised the different 'frequencies of action' of itch-related movement in humans: scratch and rub. I showed that the itch-related movements, demonstrated in patients with itchy disease, were separate and different from the activity of generalised movement (Chapter 1). I put humans' frequency of itch-related movement in context by comparing it to other mammals' (Chapter 1). I proved human itch-related 'frequency of action' by using gold standard video recording and this was consistent with previous proposals by Shuster (Felix & Shuster, 1975) and Aoki (Aoki et al., 1980), of action at about 2Hz. However, a finding from my characterizing the 'frequency of action' of walking study in Chapter 3, should have alerted me to a problem: walking peaked at about 2 to 3 Hz. Itch-related movement peaked at about 0.5 to 2.5 Hz. There was an overlap.

The simple fact of the matter is that a lot of human movement is rhythmical – limbs swing about the pendulum of joints – consider your arm swinging as you walk. Factoring in my animal-scratch experiment's finding that mass predicts 'frequency of action', along with the observation that a lot of animal characteristics are related to their size overall, I suspect that all human movement occurs at a frequency of between 0.5 to 3 Hz. Therefore, perhaps my hope of discriminating itch-related movement from other movements was overly optimistic. I tried to negotiate this problem by filtering of the amplitude of movement. One limitation to this is that it also had the potential to filter out movement in the non-DigiTrac-wearing limb. This would reduce the sensitivity of detection of itch-related movement overall.

Nevertheless, despite my not being a trained signal analyst, I believe that I developed the output such that DigiTrac could detect specific itch–related movement to a discriminating degree (to separate itchy subjects' scores from controls').

My studies do not suggest that either accelerometer is superior to the other. The DigiTrac can offer very detailed information but this may not be necessary. Actiwatch Plus scores can separate itchy subjects from control to a similar degree to DigiTrac. Both accelerometers demonstrate great variance in score especially in itchy subjects. I demonstrated that accelerometer score could predict disease group (itchy versus controls and specifically liver disease from other itchy disease). Actiwatch Plus is possibly superior to DigiTrac as it is a simple device, therefore, data analysis is quicker and more straightforward. Actiwatch Plus is also smaller and is generally more convenient as it can be set to measure and issued to subjects for several weeks. Another reason to explain why, despite the more detailed movement monitoring, the DigiTrac was found not to be highly superior to Actiwatch Plus in detecting itch-related movement, may be to do with the factory-fitted bandpass filters incorporated in the machines. It could well be that the filter incorporated in Actiwatch Plus, may have, in fact made them precisely fit for the purpose of detecting and itch-related and

human movement (see Chapter 2.I).

One final comment on accelerometer technology which I would not like to go unnoticed is that both devices, Actiwatch Plus and DigiTrac were supplied with their own user-friendly interface software. Tempting though it was to use this, I am glad that I validated it as I discovered 'bugs' in the software for both devices. The raw accelerometer data can be accessed in .csv files and exported into Excel. This is safer than relying on the company-supplied software for accuracy's sake. Software errors happen, some have devastating consequences, viz destruction of NASA 'Mariner 1' rocket in 1962 due to an incorrect formula in FORTRAN, fives deaths in the 1980's due to a bug in the controlling code for Therac-25 radiation machines. Obviously, bugs in this software would not have the potential for such harm but is worth knowing to avoid it before using the devices in an experimental situation where truly valid results are a necessity. The error found was a jump in code such that time and activity was then disconnected. This effect would have caused increasing error with time, so would be a real problem with long study periods.

There are two major findings in this work: proof of large night-to-night variation in objective accelerometer scores and poor correlation between objective-accelerometer measured itch and subjective scores or physician scores of disease-extent. I believe that I have thoroughly proved the veracity of the first finding: night-to-night variation exists within and between people, especially the itchy. The finding was repeatedly shown: in Actiwatch Plus and DigiTrac studies, in small clusters of nights and even in long, six-week studies. I suppose, having proved the phenomenon, it would be remiss not to try to explain it to some extent. I suspect that there are many fleeting factors which may affect itch in clinical disease and account for the night-to-night variation in objective itch score. It is now clear that there are, at least, two pathways for itch signaling (see Introduction). I suspect that in chronic itchy conditions, inflammatory factors (skin disease) and/or central factors (for instance opioids in

cholestasis) are the main driver(s) behind the sensation. Consider atopic dermatitis: I suspect the cytokine system and even proteases (viz. house dust associated serine proteases activating itch via PAR-2 (Kauffman, Tamm, Timmerman, & Borger, 2006)) 'prime the system.' This would set a background level of itch. Then other factors could be superimposed, more transiently, to allow for night-to-night change. We know that heat exacerbates itch sensation, therefore a hot day, a hot bath, or a hot curry could make for an itchier night. An argument just prior to bed and the associated vasodilation might quite easily be seen to affect itchiness. We also know that there are central, cortical factors that can affect itch, therefore factors which make sufferers more aware of their condition could transiently the increase its awareness and therefore activity: I have been driven mad by itch whilst reading books and papers about it. The central activation of itch may be why patients so often tell us that their skin is worse when they are stressed (although there are other pathways potentially culpable for this). Finally, although I believe that the literature quite clearly suggests that itch in atopic dermatitis is most likely not histamine-mediated, atopics' predisposition to Type I allergy obviously paves the way for transient histamine–mediated exacerbation of it. In clinic, I have observed atopic dermatitis patients with periorbital eczema, not responding to topical steroids and emollients. If Type I allergy to mould is demonstrated (a common problem in student flats, it would appear) and antihistamines are prescribed in addition to the usual steroid and emollient, on review, they are transformed. Perhaps some of the night-to-night variation is therefore explained by superimposed Type I disease to foods, or other mediators.

Some variance in accelerometer score was observed within person and between subjects in the control group. This was not at all to the same same degree as it was in the subjects. I suspect this is due to transient phenomena detailed above but without the inflammatory or central 'priming' of the itch-pathway.

The second major findings of this study, was the disconnect between objective and subjective measures of itch. It is accepted that an objective measure of disease, no matter how logical it is, cannot hope to express the impact a disease or symptom has on a person or on their emotional state. The formalized quantification of this effect was borne out of the most well-meaning of principles: really discerning how a person feels about their disease or health state and then aiming to improve in "goodness" for all apropos Jeremy Bentham and John Stuart Mill's aim of utilitarianism. However, on examining the literature it appears to me that the basic principle has been lost along the way, probably, I hate to say, in an effort to rationalize health spending in a situation of limited resources (Nord, 1999). There is a proliferation of quality of life assessment tools but it is not clear what exactly they are measuring and most importantly whether they represent what an actual patient feels or thinks. I had no idea that quality of life schools and questionnaires were developed by asking a healthy person what they imagined it might be like to live with the condition (Nord, 1999). This methodology seems essentially flawed. It also appears, however, that people actually living in certain situations or with certain conditions find it hard to clearly express, via quality-of-life surveys, how they feel.

One issue is thought to be the ‘hedonic treadmill’ (Brickman et al., 1978; Kahneman et al., 1999). It is this which makes patients with chronic disease on haemodialysis score a better quality of life than someone with an acute disease: just because they have got used to the "awfulness" of the situation (Riis et al., 2005). Adaptation or the ‘hedonic treadmill’ also accounts for paraplegics’ life-satisfaction score settling to near pre-accident levels, one year later (Brickman et al., 1978). It can happen with happiness too, with lottery winners being shown to be no happier than controls (Brickman et al., 1978). It surprised me to read that patients with terminal disease score a better quality of life than those without (Fayers, 2005). Psychologists argue that this to do with not having to deal with uncertainty: at least they know what is certainly going to happen, soon. It also raises questions about the validity of

quality-of-life scores as far as I am concerned.

Over the past 25 years, researchers working in the fields of Economics and Cognitive Behavioural Psychology, have sought to provide clarity to the situation, as I have alluded to above. They have also sought to raise awareness of the systematic problems with measuring subjective symptoms. In-roads have been made in understanding how episodic, experiential memories differ from semantic stereotyped ones and trying to improve tools in order to tap into these and into how biases can affect answers elicited from questions (Kahneman, 2004; Robinson & Clore, 2002a; 2002b). One anecdote which can help to explain the difference in episodic and semantic memories is by recalling a roller-coaster ride. When you remember the ride, you remember a collective memory of the event: it was exciting, it was scary, I felt sick. You do not actually relive each second. If you did have the kind of memory which enabled you to remember, in second-by-second detail, you would have no desire or requirement to repeat the experience. Our brains simply can not store such detailed information in a readily accessible manner as there is not enough room.

Acknowledgement of the superior quality of episodic experiential memory has made its way into the field of medicine in the form of diarizing. Paper journals or forms can be completed. Subjects can defeat researchers by lack of compliance. One (rather sneaky) group placed light sensors into diaries to check if they had been completed daily or if many forms were filled in retrospect and confirmed their suspicions of so call 'parking lot compliance' (Smyth & Stone, 2003). As technology and our connection with it improves, apps or online reporting may improve the experiential monitoring and patients' compliance.

In my research, I compared my version of experiential reporting, daily VAS scores, to the objective accelerometer scores over weeks. No correlation was found between the scores. Interestingly, however, significant autocorrelation was detected for the VAS score, despite

the respondents seemingly having been blinded to the previous day's score (by posting it in a ballot box). This suggests VAS are open to the effects of another bias, 'anchoring effect' (Ariely et al., 2003; Kahneman & Sugden, 2005). This is a situation where an individual scores around their own individual reference point.

My studies showed was that there was, at least, a corroboration between the subjective tools tested. It appears that whatever they are testing, they are testing it reliably. The fact that the symptom-tools were robust in the face of my imposed biases is reassuring regarding reliability too. There still remains, however, an unexplained disconnect between the subjective and objective scores: some other facet of disease which is, as yet, intangible and which is, as a consequence, immeasurable.

Since undertaking this research, I find it depressing to see how many studies are published which make statements purely on the basis of subjective disease scores, especially in the field of itch. For instance, a recently published study has stated that the itch in lichen planus is nothing to do with IL-31, although it is expressed in this condition, because the VAS scores of itch were not affected by different levels of the cytokine (Welz-Kubiak, Kobuszewska, & Reich, 2015). I accept that subjective tools have a value in the clinical situation, surely objective ones do in a trial or experimental one. An argument in times gone by would have been that accelerometers are difficult and expensive to acquire, but I really do think that small cheap devices are possible with improvements in present technology and it may even be possible to install an app on a device the patient already owns, an Applewatch, for instance.

An editorial comment about the paper I published from these studies, which tested subjective measures to bias, made a very interesting point: despite measuring patient reports in clinical trials, the data is most often not reported (Williams, 2010). My considering other disconnects

and ‘perspective on disease’ led me to the last tranche of my studies. The views of patients and doctors are not necessarily united even though they aspire to the same goal, successful treatment. If doctors are honest, they really want to see an improvement in a patient's condition if they have prescribed a treatment. They have also seen the whole spectrum of disease, unlike a patient, and so their weighing up of severity may well be different to their patient’s. Meanwhile, patients only have themselves and their own experienced severity for reference.

When investigating the two perspectives, I found it amazing that this has been so neglected by the majority of quality of life literature. In one meta-analysis, of over 10,000 papers, only twelve contained data from both groups and none clearly reported it (Janse et al., 2004). It seems obvious that if patient and doctor have different views on what is an acceptable outcome, there is a huge chance of disenchantment on both sides. Work summarized in this thesis has shown how faceless, featureless, 3-D models of psoriasis are able to unite perspective on extent of disease, from dermatologists through general nurses and even to patients. Taking inspiration from published work I had read about in the field of Well-being research (Kahneman et al., 1999), I also decided to see if the models and own-patient photographs would sway personally assessed extent of disease. Curiously, seeing their own photographs reduced corroboration between patient and doctor-assessed extent. I am not sure why this should be. Perhaps the photographs were an unpleasant ‘reality check’ to the patients and influenced the score accordingly.

As it stands, the 3-D pictures are only in the female form. I would love to extend the study and to acquire equivalent 3-D male models. I really do think that the models could prove a helpful aide in the day-to-day clinical and trial settings. One particularly helpful feature of them is that their extent increases linearly. A linear scoring pattern is more sensitive to change than a clustered model. PASI clusters scores at the lower end of the severity

spectrum and so is less sensitive to change here.

Overall, I have shown that you can measure itch objectively. The objective scores do not marry with subjective scores and since they measure different aspects of disease, both should be registered. Clinical doctors and researchers need to be aware of biases and perspectives on disease which may affect assessment of disease and keep an open, questioning mind.

References

Ahmad Fadzil, M. H., Ihtatho, D., Mohd Affandi, A., & Hussein, S. H. (2009). Objective assessment of psoriasis erythema for PASI scoring. *Journal of Medical Engineering & Technology*, *33*(7), 516–524. http://doi.org/10.1080/07434610902744074

Andoh, T., Yageta, Y., Konno, M., Yamaguchi-Miyamoto, T., Takahata, H., Nojima, H., et al. (2008). Evidence for Separate Involvement of Different μ-Opioid Receptor Subtypes in Itch and Analgesia Induced by Supraspinal Action of Opioids. *Journal of Pharmacological Sciences*, *106*(4), 667–670. http://doi.org/10.1254/jphs.08004SC

Andrew, D., & Craig, A. D. (2001). Spinothalamic lamina I neurons selectively sensitive to histamine: a central neural pathway for itch. *Nature Neuroscience*, *4*(1), 72–77. http://doi.org/10.1038/82924

Aoki, T., Kushimoto, H., Hishikawa, Y., & Savin, J. A. (1991). Nocturnal scratching and its relationship to the disturbed sleep of itchy subjects. *Clinical and Experimental Dermatology*, *16*(4), 268–272.

Aoki, T., Kushimoto, H., Kobayashi, E., Ogushi, Y., Akimoto, T., Horiko, T., & Fujita, M. (1980). [Estimation of scratch movement of hands in night by use of computer system (author's transl)]. *Nihon Hifuka Gakkai Zasshi. the Japanese Journal of Dermatology*, *90*(8), 737–739.

Ariely, D., Loewenstein, G., & Prelec, D. (2003). "Coherent Arbitrariness": Stable Demand Curves Without Stable Preferences. *The Quarterly Journal of Economics*, *118*(1), 73–106. http://doi.org/10.1162/00335530360535153

Arthur, R. P., & Shelley, W. B. (1955). Experimental evidence for an enzymatic basis for itching in man. *Nature*, *175*(4464), 901–902.

Ashcroft, D. M., Wan Po, A. L., Williams, H. C., & Griffiths, C. E. (1999). Clinical measures of disease severity and outcome in psoriasis: a critical appraisal of their quality. *British Journal of Dermatology*, *141*(2), 185–191.

Atanassoff, P. G., Brull, S. J., Zhang, J., Greenquist, K., Silverman, D. G., & LaMotte, R. H. (1999). Enhancement of experimental pruritus and mechanically evoked dysesthesiae with local anesthesia. *Somatosensory & Motor Research*, *16*(4), 291–298.

Atherton, D., Bieber, T., Bonifazi, E., Broberg, A., Calza, A., Coleman, R., et al. (1993). Severity scoring of atopic dermatitis: The scorad index. *Dermatology*, *6*(1), 95. http://doi.org/10.1016/0923-1811(93)91263-T

Baker, F. C., Maloney, S., & Driver, H. S. (1999). A comparison of subjective estimates of sleep with objective polysomnographic data in healthy men and women. *Journal of Psychosomatic Research*, *47*(4), 335–341.

Ballantyne, J. C., Loach, A. B., & Carr, D. B. (1988). Itching after epidural and spinal opiates. *Pain*, *33*(2), 149–160.

Bell, J. K., McQueen, D. S., & Rees, J. L. (2004). Involvement of histamine H4 and H1 receptors in scratching induced by histamine receptor agonists in Balb C mice. *British Journal of Pharmacology*, *142*(2), 374–380. http://doi.org/10.1038/sj.bjp.0705754

Bender, B. G., Leung, S. B., & Leung, D. Y. M. (2003). Actigraphy assessment of sleep disturbance in patients with atopic dermatitis: An objective life quality measure. *Journal of Allergy and Clinical Immunology*, *111*(3), 598–602. http://doi.org/10.1067/mai.2003.174

Benjamin, K., Waterston, K., Russell, M., Schofield, O., Diffey, B., & Rees, J. L. (2004). The development of an objective method for measuring scratch in children with atopic dermatitis suitable for clinical use. *Journal of the American Academy of Dermatology*, *50*(1), 33–40. http://doi.org/10.1016/s0190-9622(03)02480-0

Bergasa, N. V., Alling, D. W., Talbot, T. L., Swain, M. G., Yurdaydin, C., Turner, M. L., et

al. (1995). Effects of naloxone infusions in patients with the pruritus of cholestasis. A double-blind, randomized, controlled trial. *Annals of Internal Medicine*, *123*(3), 161–167.

Bergasa, N. V., Schmitt, J. M., Talbot, T. L., Alling, D. W., Swain, M. G., Turner, M. L., et al. (1998). Open-label trial of oral nalmefene therapy for the pruritus of cholestasis. *Hepatology (Baltimore, Md.)*, *27*(3), 679–684. http://doi.org/10.1002/hep.510270307

Bergasa, N. V., Talbot, T. L., Alling, D. W., Schmitt, J. M., Walker, E. C., Baker, B. L., et al. (1992). A controlled trial of naloxone infusions for the pruritus of chronic cholestasis. *Gastroenterology*, *102*(2), 544–549.

Berth-Jones, J., Grotzinger, K., Rainville, C., Pham, B., Huang, J., Daly, S., et al. (2006). A study examining inter- and intrarater reliability of three scales for measuring severity of psoriasis: Psoriasis Area and Severity Index, Physician"s Global Assessment and Lattice System Physician"s Global Assessment. *British Journal of Dermatology*, *155*(4), 707–713. http://doi.org/10.1111/j.1365-2133.2006.07389.x

Bickford, R. (1938). Experiments relating to the itch sensation. *Clinical Science*, *3*, 377–386.

Bjorner, J. (2005). Self-rated health. In P. Fayers & R. Hays (Eds.), *Assessing quality of life in clinical trials* (Second). Oxford.

Blunk, J. A., Schmelz, M., Zeck, S., Skov, P., Likar, R., & Koppert, W. (2004). Opioid-induced mast cell activation and vascular responses is not mediated by mu-opioid receptors: an in vivo microdialysis study in human skin. *Anesthesia and Analgesia*, *98*(2), 364–70– table of contents.

Both, H., Essink-Bot, M.-L., Busschbach, J., & Nijsten, T. (2007). Critical review of generic and dermatology-specific health-related quality of life instruments. *The Journal of Investigative Dermatology*, *127*(12), 2726–2739. http://doi.org/10.1038/sj.jid.5701142

Boulay, J.-L., O'Shea, J. J., & Paul, W. E. (2003). Molecular phylogeny within type I cytokines and their cognate receptors. *Immunity*, *19*(2), 159–163.

Brash, H. M., McQueen, D. S., Christie, D., Bell, J. K., Bond, S. M., & Rees, J. L. (2005). A repetitive movement detector used for automatic monitoring and quantification of scratching in mice. *Journal of Neuroscience Methods*, *142*(1), 107–114. http://doi.org/10.1016/j.jneumeth.2004.08.001

Brickman, P., Coates, D., & Janoff-Bulman, R. (1978). Lottery winners and accident victims: Is happiness relative? *Journal of Personality and Social Psychology*, *36*(8), 917–927. http://doi.org/10.1037//0022-3514.36.8.917

Bringhurst, C., Waterston, K., Schofield, O., Benjamin, K., & Rees, J. L. (2004). Measurement of itch using actigraphy in pediatric and adult populations. *Journal of the American Academy of Dermatology*, *51*(6), 893–898. http://doi.org/10.1016/j.jaad.2004.05.039

Bromm, B., Scharein, E., Darsow, U., & Ring, J. (1995). Effects of menthol and cold on histamine-induced itch and skin reactions in man. *Neuroscience Letters*, *187*(3), 157–160.

Carlin, C. S., Feldman, S. R., Krueger, J. G., Menter, A., & Krueger, G. G. (2004). A 50% reduction in the Psoriasis Area and Severity Index (PASI 50) is a clinically significant endpoint in the assessment of psoriasis. *Journal of the American Academy of Dermatology*, *50*(6), 859–866. http://doi.org/10.1016/j.jaad.2003.09.014

Carstens, E., & Akiyama, T. (2014). Spinal Coding of Itch and Pain. Boca Raton (FL): CRC Press/Taylor & Francis.

Cevikbas, F., Wang, X., Akiyama, T., Kempkes, C., Savinko, T., Antal, A., et al. (2014). A sensory neuron–expressed IL-31 receptor mediates T helper cell–dependent itch: Involvement of TRPV1 and TRPA1. *Journal of Allergy and Clinical Immunology*, *133*(2), 448–460.e7. http://doi.org/10.1016/j.jaci.2013.10.048

Charman, C. R., Venn, A. J., & Williams, H. C. (1999). Measurement of body surface area involvement in atopic eczema: an impossible task? *British Journal of Dermatology*, *140*(1), 109–111.

Charman, C. R., Venn, A. J., & Williams, H. C. (2002). Reliability testing of the Six Area, Six Sign Atopic Dermatitis severity score. *British Journal of Dermatology*, *146*(6), 1057–1060.

Chren, M. M., Lasek, R. J., Quinn, L. M., & Covinsky, K. E. (1997). Convergent and discriminant validity of a generic and a disease-specific instrument to measure quality of life in patients with skin disease. *Journal of Investigative Dermatology*, *108*(1), 103–107.

Costa, R., Marotta, D. M., Manjavachi, M. N., Fernandes, E. S., Lima-Garcia, J. F., Paszcuk, A. F., et al. (2008). Evidence for the role of neurogenic inflammation components in trypsin-elicited scratching behaviour in mice. *British Journal of Pharmacology*, *154*(5), 1094–1103. http://doi.org/10.1038/bjp.2008.172

Craig, A. D., Reiman, E. M., Evans, A., & Bushnell, M. C. (1996). Functional imaging of an illusion of pain. *Nature*, *384*(6606), 258–260. http://doi.org/10.1038/384258a0

Darsow, U., Drzezga, A., Frisch, M., Munz, F., Weilke, F., Bartenstein, P., et al. (2000). Processing of histamine-induced itch in the human cerebral cortex: a correlation analysis with dermal reactions. *Journal of Investigative Dermatology*, *115*(6), 1029–1033. http://doi.org/10.1046/j.1523-1747.2000.00193.x

Darsow, U., Mautner, V. F., Bromm, B., Scharein, E., & Ring, J. (1997). [The Eppendorf Pruritus Questionnaire]. *Der Hautarzt; Zeitschrift Für Dermatologie, Venerologie, Und Verwandte Gebiete*, *48*(10), 730–733.

Davidson, S., Zhang, X., Yoon, C. H., Khasabov, S. G., Simone, D. A., & Giesler, G. J. (2007). The Itch-Producing Agents Histamine and Cowhage Activate Separate Populations of Primate Spinothalamic Tract Neurons. *Journal of Neuroscience*, *27*(37), 10007–10014. http://doi.org/10.1523/JNEUROSCI.2862-07.2007

De Castro-Costa, M., Gybels, J., Kupers, R., & Van Hees, J. (1987). Scratching behaviour in arthritic rats: a sign of chronic pain or itch? *Pain*, *29*(1), 123–131.

Del Missier, F., Ferrante, D., & Costantini, E. (2007). Focusing effects in predecisional information acquisition. *Acta Psychologica*, *125*(2), 155–174. http://doi.org/10.1016/j.actpsy.2006.07.005

Desai, N. S., Poindexter, G. B., Monthrope, Y. M., Bendeck, S. E., Swerlick, R. A., & Chen, S. C. (2008). A pilot quality-of-life instrument for pruritus. *Journal of the American Academy of Dermatology*, *59*(2), 234–244. http://doi.org/10.1016/j.jaad.2008.04.006

Drzezga, A., Darsow, U., Treede, R. D., Siebner, H., Frisch, M., Munz, F., et al. (2001). Central activation by histamine-induced itch: analogies to pain processing: a correlational analysis of O-15 H2O positron emission tomography studies. *Pain*, *92*(1-2), 295–305.

Ebata, T., Aizawa, H., & Kamide, R. (1996). An infrared video camera system to observe nocturnal scratching in atopic dermatitis patients. *The Journal of Dermatology*, *23*(3), 153–155.

Ebata, T., Aizawa, H., Kamide, R., & Niimura, M. (1999). The characteristics of nocturnal scratching in adults with atopic dermatitis. *British Journal of Dermatology*, *141*(1), 82–86.

Ebata, T., Iwasaki, S., Kamide, R., & Niimura, M. (2001). Use of a wrist activity monitor for the measurement of nocturnal scratching in patients with atopic dermatitis. *British Journal of Dermatology*, *144*(2), 305–309.

Ebata, T., Izumi, H., Aizawa, H., Kamide, R., & Niimura, M. (1998). Effects of nitrazepam on nocturnal scratching in adults with atopic dermatitis: a double-blind placebo-controlled crossover study. *British Journal of Dermatology*, *138*(4), 631–634.

Elman, S., Hynan, L. S., Gabriel, V., & Mayo, M. J. (2010). The 5-D itch scale: a new measure of pruritus. *British Journal of Dermatology*, *162*(3), 587–593. http://doi.org/10.1111/j.1365-2133.2009.09586.x

Emerson, R. M., Charman, C. R., & Williams, H. C. (2000). The Nottingham Eczema Severity Score: preliminary refinement of the Rajka and Langeland grading. *British*

Journal of Dermatology, *142*(2), 288–297.
Endo, K., Sumitsuji, H., Fukuzumi, T., Adachi, J., & Aoki, T. (1997). Evaluation of scratch movements by a new scratch-monitor to analyze nocturnal itching in atopic dermatitis. *Acta Dermato-Venereologica*, *77*(6), 432–435.
Fayers, P. M. (2005). Assessing Quality of Life in Clinical Trials. Oxford University Press, USA.
Felix, R., & Shuster, S. (1975). A new method for the measurement of itch and the response to treatment. *British Journal of Dermatology*, *93*(3), 303–312.
Finlay, A. Y., & Coles, E. C. (1995). The effect of severe psoriasis on the quality of life of 369 patients. *British Journal of Dermatology*, *132*(2), 236–244.
Finlay, A. Y., & Khan, G. K. (1994). Dermatology Life Quality Index (DLQI)—a simple practical measure for routine clinical use. *Clinical and Experimental Dermatology*, *19*(3), 210–216. http://doi.org/10.1111/j.1365-2230.1994.tb01167.x
Frederick, S., & Loewenstein, G. (1999). Well-being: The foundations of hedonic psychology. (D. Kahneman, E. Diener, & N. Schwartz, Eds.) (pp. 302–309). New York.
Fredrickson, B. L., & Kahneman, D. (1993). Duration neglect in retrospective evaluations of affective episodes. *Journal of Personality and Social Psychology*, *65*(1), 45–55.
Fredriksson, T., & Pettersson, U. (1978). Severe Psoriasis – Oral Therapy with a New Retinoid. *Dermatology*, *157*(4), 238–244. http://doi.org/10.1159/000250839
Goswami, S. C., Thierry-Mieg, D., Thierry-Mieg, J., Mishra, S., Hoon, M. A., Mannes, A. J., & Iadarola, M. J. (2014). Itch-associated peptides: RNA-Seq and bioinformatic analysis of natriuretic precursor peptide B and gastrin releasing peptide in dorsal root and trigeminal ganglia, and the spinal cord. *Molecular Pain*, *10*(1), 44. http://doi.org/10.1186/1744-8069-10-44
Han, L., Ma, C., Liu, Q., Weng, H.-J., Cui, Y., Tang, Z., et al. (2012). A subpopulation of nociceptors specifically linked to itch. *Nature Neuroscience*, *16*(2), 174–182. http://doi.org/10.1038/nn.3289
Hanifin, J. M., Thurston, M., Omoto, M., Cherill, R., Tofte, S. J., & Graeber, M. (2001). The eczema area and severity index (EASI): assessment of reliability in atopic dermatitis. EASI Evaluator Group. *Experimental Dermatology*, *10*(1), 11–18.
Hanifin, J., & Rajka, G. (1980). Diagnostic features of atopic dermatitis. *Acta Derm Venereol Stockh*, *Suppl. 92*, 44–47.
Hägermark, O. (1992). Peripheral and central mediators of itch. *Skin Pharmacology : the Official Journal of the Skin Pharmacology Society*, *5*(1), 1–8.
Hägermark, O., & Strandberg, K. (1977). Pruritogenic activity of prostaglandin E2. *Acta Dermato-Venereologica*, *57*(1), 37–43.
Hägermark, O., & Wahlgren, C. F. (1992). Some methods for evaluating clinical itch and their application for studying pathophysiological mechanisms. *Journal of Dermatological Science*, *4*(2), 55–62.
Hägermark, O., Rajka, G., & Bergvist, U. (1972). Experimental itch in human skin elicited by rat mast cell chymase. *Acta Dermato-Venereologica*, *52*(2), 125–128.
Hägermark, O., Strandberg, K., & Hamberg, M. (1977). Potentiation of itch and flare responses in human skin by prostaglandins E2 and H2 and a prostaglandin endoperoxide analog. *Journal of Investigative Dermatology*, *69*(6), 527–530.
Hongbo, Y., Thomas, C. L., Harrison, M. A., Salek, M. S., & Finlay, A. Y. (2005). Translating the science of quality of life into practice: What do dermatology life quality index scores mean? *Journal of Investigative Dermatology*, *125*(4), 659–664. http://doi.org/10.1111/j.0022-202X.2005.23621.x
Hoon, M. A. (2015). Molecular dissection of itch. *Current Opinion in Neurobiology*, *34*, 61–66. http://doi.org/10.1016/j.conb.2015.01.017
Hsieh, J. C., Hägermark, O., Ståhle-Bäckdahl, M., Ericson, K., Eriksson, L., Stone-Elander, S., & Ingvar, M. (1994). Urge to scratch represented in the human cerebral cortex during itch. *Journal of Neurophysiology*, *72*(6), 3004–3008.

Imamachi, N., Park, G. H., Lee, H., Anderson, D. J., Simon, M. I., Basbaum, A. I., & Han, S. K. (2009). TRPV1-expressing primary afferents generate behavioral responses to pruritogens via multiple mechanisms. *Proceedings of the National Academy of Sciences*, *106*(27), 11330–11335. http://doi.org/10.1073/pnas.0905605106

Inagaki, N., Nagao, M., Igeta, K., Kawasaki, H., Kim, J. F., & Nagai, H. (2001). Scratching behavior in various strains of mice. *Skin Pharmacology and Applied Skin Physiology*, *14*(2), 87–96.

Jacobson, C. C., & Kimball, A. B. (2004). Rethinking the Psoriasis Area and Severity Index: the impact of area should be increased. *British Journal of Dermatology*, *151*(2), 381–387. http://doi.org/10.1111/j.1365-2133.2004.06035.x

Janse, A. J., Gemke, R. J. B. J., Uiterwaal, C. S. P. M., van der Tweel, I., Kimpen, J. L. L., & Sinnema, G. (2004). Quality of life: patients and doctors don't always agree: a meta-analysis. *Journal of Clinical Epidemiology*, *57*(7), 653–661. http://doi.org/10.1016/j.jclinepi.2003.11.013

Jinks, S. L., & Carstens, E. (1998). Skin cooling attenuates rat dorsal horn neuronal responses to intracutaneous histamine. *NeuroReport*, *9*(18), 4145–4149. http://doi.org/10.1097/00001756-199812210-00027

Jinks, S. L., & Carstens, E. (2000). Superficial dorsal horn neurons identified by intracutaneous histamine: chemonociceptive responses and modulation by morphine. *Journal of Neurophysiology*, *84*(2), 616–627.

Jovanov, E., Milenkovic, A., Otto, C., & de Groen, P. C. (2005). A wireless body area network of intelligent motion sensors for computer assisted physical rehabilitation. *Journal of Neuroengineering and Rehabilitation*, *2*(1), 6. http://doi.org/10.1186/1743-0003-2-6

Kahneman, D. (2004). A Survey Method for Characterizing Daily Life Experience: The Day Reconstruction Method. *Science*, *306*(5702), 1776–1780. http://doi.org/10.1126/science.1103572

Kahneman, D., & Diener, E. (Eds.). (n.d.). Well-being: the foundations of hedonic psychology (First).

Kahneman, D., & Sugden, R. (2005). Experienced Utility as a Standard of Policy Evaluation. *Environmental and Resource Economics*, *32*(1), 161–181. http://doi.org/10.1007/s10640-005-6032-4

Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review*, *103*(3), 582–91– discusion 592–6.

Kahneman, D., Diener, E., & Schwarz, N. (1999). Well-Being. Russell Sage Foundation.

Kahneman, D., Fredrickson, B. L., Schreiber, C. A., & Redelmeier, D. A. (1993). When More Pain Is Preferred to Less: Adding a Better End. *Psychological Science*, *4*(6), 401–405. http://doi.org/10.1111/j.1467-9280.1993.tb00589.x

Kahneman, D., Krueger, A. B., Schkade, D., Schwarz, N., & Stone, A. A. (2006). Would you be happier if you were richer? A focusing illusion. *Science*, *312*(5782), 1908–1910. http://doi.org/10.1126/science.1129688

Kale, O. O. (1998). Onchocerciasis: the burden of disease. *Annals of Tropical Medicine and Parasitology*, *92*(3), 101–115. http://doi.org/10.1080/00034989859636

Kardon, A. P., Polgár, E., Hachisuka, J., Snyder, L. M., Cameron, D., Savage, S., et al. (2014). Dynorphin acts as a neuromodulator to inhibit itch in the dorsal horn of the spinal cord. *Neuron*, *82*(3), 573–586. http://doi.org/10.1016/j.neuron.2014.02.046

Kauffman, H. F., Tamm, M., Timmerman, J. A. B., & Borger, P. (2006). House dust mite major allergens Der p 1 and Der p 5 activate human airway-derived epithelial cells by protease-dependent and protease-independent mechanisms. *Clinical and Molecular Allergy : CMA*, *4*(1), 5. http://doi.org/10.1186/1476-7961-4-5

Kim, D.-K., Kim, H. J., Kim, H., Koh, J.-Y., Kim, K.-M., Noh, M.-S., et al. (2008). Involvement of serotonin receptors 5-HT1 and 5-HT2 in 12(S)-HPETE-induced scratching in mice. *European Journal of Pharmacology*, *579*(1-3), 390–394.

http://doi.org/10.1016/j.ejphar.2007.11.005
Koga, K., Chen, T., Li, X.-Y., Descalzi, G., Ling, J., Gu, J., & Zhuo, M. (2011). Glutamate acts as a neurotransmitter for gastrin releasing peptide-sensitive and insensitive itch-related synaptic transmission in mammalian spinal cord. *Molecular Pain*, *7*(1), 1. http://doi.org/10.1186/1744-8069-7-47
Krause, L., & Shuster, S. (1983). Mechanism of action of antipruritic drugs. *Bmj*, *287*(6400), 1199–1200. http://doi.org/10.1136/bmj.287.6400.1199
Kravchenko, K. L., Grechany, G. V., & Gadjiev, G. D. (2006). Correlation between Drosophila population sizes and solar activity parameters. *Biophysics*, *51*(3), 466–470. http://doi.org/10.1134/S0006350906030201
Krueger, G., Koo, J., Lebwohl, M., Menter, A., Stern, R. S., & Rolstad, T. (2001). The Impact of Psoriasis on Quality of Life: Results of a 1998 National Psoriasis Foundation Patient-Membership Survey. *Archives of Dermatology*, *137*(3), 280–284. http://doi.org/10.1016/0190-9622(93)70062-X
Kuhta, P. C., & Smith, J. L. (1990). Scratch Responses in Normal Cats: Hindlimb Kinematics and Muscle Synergies. *Journal of Neurophysiology*, *64*(6), 1653–1667.
Kunz, B., Oranje, A. P., Labrèze, L., Stalder, J. F., Ring, J., & Taïeb, A. (1997). Clinical validation and guidelines for the SCORAD index: consensus report of the European Task Force on Atopic Dermatitis. *Dermatology (Basel, Switzerland)*, *195*(1), 10–19.
Lagerström, M. C., Rogoz, K., Abrahamsen, B., Persson, E., Reinius, B., Nordenankar, K., et al. (2010). VGLUT2-Dependent Sensory Neurons in the TRPV1 Population Regulate Pain and Itch. *Neuron*, *68*(3), 529–542. http://doi.org/10.1016/j.neuron.2010.09.016
LaMotte, R. H., Shain, C. N., Simone, D. A., & Tsai, E. F. (1991). Neurogenic hyperalgesia: psychophysical studies of underlying mechanisms. *Journal of Neurophysiology*, *66*(1), 190–211.
Langan, S. M., Bourke, J. F., Silcocks, P., & Williams, H. C. (2006). An exploratory prospective observational study of environmental factors exacerbating atopic eczema in children. *British Journal of Dermatology*, *154*(5), 979–980. http://doi.org/10.1111/j.1365-2133.2006.07153.x
Lewellen, J. (2002). Momentum and Autocorrelation in Stock Returns. *Review of Financial Studies*, *15*(2), 533–564. http://doi.org/10.1093/rfs/15.2.533
Lewis, T. (1926). The Blood Vessels of the Human Skin. *British Medical Journal*, *2*(3418), 61–62.
Lewis, V., & Finlay, A. Y. (2004). 10 Years Experience of the Dermatology Life Quality Index (DLQI). *Journal of Investigative Dermatology Symposium Proceedings*, *9*(2), 169–180. http://doi.org/10.1111/j.1087-0024.2004.09113.x
Liu, Q., Tang, Z., Surdenikova, L., Kim, S., Patel, K. N., Kim, A., et al. (2009). Sensory Neuron-Specific GPCR Mrgprs Are Itch Receptors Mediating Chloroquine-Induced Pruritus. *Cell*, *139*(7), 1353–1365. http://doi.org/10.1016/j.cell.2009.11.034
Liu, Y., Samad, O. A., Zhang, L., Duan, B., Tong, Q., Lopes, C., et al. (2010). VGLUT2-Dependent Glutamate Release from Nociceptors Is Required to Sense Pain and Suppress Itch. *Neuron*, *68*(3), 543–556. http://doi.org/10.1016/j.neuron.2010.09.008
Lynn, B. (1991). "Silent" nociceptors in the skin. *Trends in Neurosciences*, *14*(3), 95.
Majeski, C. J., Johnson, J. A., Davison, S. N., & Lauzon, G. J. (2007). Itch Severity Scale: a self-report instrument for the measurement of pruritus severity. *British Journal of Dermatology*, *156*(4), 667–673. http://doi.org/10.1111/j.1365-2133.2006.07736.x
Marks, R. (1989). Assessment of Disease Progress in Psoriasis. *Archives of Dermatology*, *125*(2), 235. http://doi.org/10.1001/archderm.1989.01670140087017
McColl, E. (2005). Context effects and proxy assessments. In P. Fayers & R. Hays (Eds.), *Assessing Quality of Life in Clinical trials* (Second, pp. 131–146). Oxford: Oxford University Press.
Melzack, R. (1975). The McGill Pain Questionnaire: major properties and scoring methods.

Pain, *1*(3), 277–299.
Mishra, S. K., & Hoon, M. A. (2013). The cells and circuitry for itch responses in mice. *Science*, *340*(6135), 968–971. http://doi.org/10.1126/science.1233765
Molenaar, H. A., Oosting, J., & Jones, E. A. (1998). Improved device for measuring scratching activity in patients with pruritus. *Medical & Biological Engineering & Computing*, *36*(2), 220–224.
Murray, C. S., & Rees, J. L. (2010). How robust are the Dermatology Life Quality Index and other self-reported subjective symptom scores when exposed to a range of experimental biases? *Acta Dermato-Venereologica*, *90*(1), 34–38. http://doi.org/10.2340/00015555-0768
Murray, C. S., & Rees, J. L. (2011a). Are subjective accounts of itch to be relied on? The lack of relation between visual analogue itch scores and actigraphic measures of scratch. *Acta Dermato-Venereologica*, *91*(1), 18–23. http://doi.org/10.2340/00015555-1002
Murray, C. S., & Rees, J. L. (2011b). Are subjective accounts of itch to be relied on? The lack of relation between visual analogue itch scores and actigraphic measures of scratch. *Acta Dermato-Venereologica*, *91*(1), 18–23. http://doi.org/10.2340/00015555-1002
Mustakallio, K. K. (1991). Scratch radar for the measurement of pruritus. *Acta Dermato-Venereologica. Supplementum*, *156*, 44.
Namer, B., Carr, R., Johanek, L. M., Schmelz, M., Handwerker, H. O., & Ringkamp, M. (2008). Separate peripheral pathways for pruritus in man. *Journal of Neurophysiology*, *100*(4), 2062–2069. http://doi.org/10.1152/jn.90482.2008
Nojima, H., & Carstens, E. (2003). Quantitative assessment of directed hind limb scratching behavior as a rodent itch model. *Journal of Neuroscience Methods*, *126*(2), 137–143. http://doi.org/10.1016/S0165-0270(03)00074-8
Nord, E. (1999). Cost-Value Analysis in Health Care. Cambridge University Press. http://doi.org/10.1002/1099-1050(200012)9:8<740::AID-HEC569>3.0.CO;2-N
Park, S. Y., Ha, S.-H., Yu, D.-S., Son, S.-W., Kim, D.-J., Kim, I.-H., et al. (2004). Quantitative evaluation of severity in psoriatic lesions using three-dimensional morphometry. *Experimental Dermatology*, *13*(4), 223–228. http://doi.org/10.1111/j.0906-6705.2004.00122.x
Ramsay, B., & Lawrence, C. M. (1991). Measurement of involved surface area in patients with psoriasis. *British Journal of Dermatology*, *124*(6), 565–570.
Reddy, V. B., Shimada, S. G., Sikand, P., LaMotte, R. H., & Lerner, E. A. (2010). Cathepsin S elicits itch and signals via protease-activated receptors. *The Journal of Investigative Dermatology*, *130*(5), 1468–1470. http://doi.org/10.1038/jid.2009.430
Redelmeier, D. A., & Kahneman, D. (1996). Patients' memories of painful medical treatments: real-time and retrospective evaluations of two minimally invasive procedures. *Pain*, *66*(1), 3–8.
Redelmeier, D. A., Katz, J., & Kahneman, D. (2003). Memories of colonoscopy: a randomized trial. *Pain*, *104*(1), 187–194. http://doi.org/10.1016/S0304-3959(03)00003-4
Redelmeier, D. A., Rozin, P., & Kahneman, D. (1993). Understanding patients' decisions. Cognitive and emotional perspectives. *Jama*, *270*(1), 72–76.
Rees, J., & Murray, C. S. (2005). Itching for progress. *Clinical and Experimental Dermatology*, *30*(5), 471–473. http://doi.org/10.1111/j.1365-2230.2005.01852.x
Riis, J., Loewenstein, G., Baron, J., Jepson, C., Fagerlin, A., & Ubel, P. A. (2005). Ignorance of Hedonic Adaptation to Hemodialysis: A Study Using Ecological Momentary Assessment. *Journal of Experimental Psychology General*, *134*(1), 3–9. http://doi.org/10.1037/0096-3445.134.1.3
Roberson, D. P., Gudes, S., Sprague, J. M., Patoski, H. A. W., Robson, V. K., Blasl, F., et al. (2013). Activity-dependent silencing reveals functionally distinct itch-generating sensory neurons. *Nature Neuroscience*, *16*(7), 910–918. http://doi.org/10.1038/nn.3404
Robinson, M. D., & Clore, G. L. (2002a). Belief and feeling: Evidence for an accessibility

model of emotional self-report. *Psychological Bulletin*, *128*(6), 934–960. http://doi.org/10.1037//0033-2909.128.6.934
Robinson, M. D., & Clore, G. L. (2002b). Episodic and semantic knowledge in emotional self-report: Evidence for two judgment processes. *Journal of Personality and Social Psychology*, *83*(1), 198–215. http://doi.org/10.1037//0022-3514.83.1.198
Roenigk, R. K., & Roenigk, H. H. (1978). Sex differences in the psychological effects of psoriasis. *Cutis*, *21*(4), 529–533.
Ross, S. E., Mardinly, A. R., McCord, A. E., Zurawski, J., Cohen, S., Jung, C., et al. (2010). Loss of Inhibitory Interneurons in the Dorsal Spinal Cord and Elevated Itch in Bhlhb5 Mutant Mice. *Neuron*, *65*(6), 886–898. http://doi.org/10.1016/j.neuron.2010.02.025
Rossbach, K., Nassenstein, C., Gschwandtner, M., Schnell, D., Sander, K., Seifert, R., et al. (2011). Histamine H1, H3 and H4 receptors are involved in pruritus. *Neuroscience*, *190*(C), 89–102. http://doi.org/10.1016/j.neuroscience.2011.06.002
Rukwied, R., Lischetzki, G., Mcglone, F., Heyer, G., & Schmelz, M. (2000). Mast cell mediators other than histamine induce pruritus in atopic dermatitis patients: a dermal microdialysis study. *British Journal of Dermatology*, *142*(6), 1114–1120. http://doi.org/10.1046/j.1365-2133.2000.03535.x
Sampogna, F., Sera, F., Abeni, D., IDI Multipurpose Psoriasis Research on Vital Experiences (IMPROVE) Investigators. (2004). Measures of clinical severity, quality of life, and psychological distress in patients with psoriasis: a cluster analysis. *Journal of Investigative Dermatology*, *122*(3), 602–607. http://doi.org/10.1046/j.0022-202X.2003.09101.x
Savin, J. A., Paterson, W. D., & Oswald, I. (1973). Scratching during sleep. *The Lancet*, 296–297.
Savin, J. A., Paterson, W. D., Adam, K., & Oswald, I. (1979). Effects of trimeprazine and trimipramine on nocturnal scratching in patients with atopic eczema. *Archives of Dermatology*, *115*(3), 313–315.
Savin, J. A., Paterson, W. D., Oswald, I., & Adam, K. (1975). Further studies of scratching during sleep. *British Journal of Dermatology*, *93*(3), 297–302.
Savolainen, L., Kontinen, J., Alatalo, E., Röning, J., & Oikarinen, A. (1998). Comparison of actual psoriasis surface area and the psoriasis area and severity index by the human eye and machine vision methods in following the treatment of psoriasis. *Acta Dermato-Venereologica*, *78*(6), 466–467.
Schmelz, M. (2001). A neural pathway for itch. *Nature Neuroscience*, *4*(1), 9–10. http://doi.org/10.1038/82956
Schmelz, M. (2003). Chemical Response Pattern of Different Classes of C-Nociceptors to Pruritogens and Algogens. *Journal of Neurophysiology*, *89*(5), 2441–2448. http://doi.org/10.1152/jn.01139.2002
Schmelz, M., Schmidt, R., Bickel, A., Handwerker, H. O., & Torebjörk, H. E. (1997). Specific C-receptors for itch in human skin. *Journal of Neuroscience*, *17*(20), 8003–8008.
Schmid-Ott, G., Jaeger, B., Kuensebeck, H. W., Ott, R., & Lamprecht, F. (1996). Dimensions of stigmatization in patients with psoriasis in a "Questionnaire on Experience with Skin Complaints'. *Dermatology (Basel, Switzerland)*, *193*(4), 304–310.
Shelley, W. B., & Arthur, R. P. (1955a). Mucunain, the active pruritogenic proteinase of cowhage. *Science*, *122*(3167), 469–470.
Shelley, W. B., & Arthur, R. P. (1955b). Studies on cowhage (Mucuna pruriens) and its pruritogenic proteinase, mucunain. *A.M.a. Archives of Dermatology*, *72*(5), 399–406.
Shuster, S. (1966). Trends in dermatology. *Transactions of the St. John's Hospital Dermatological Society*, *52*(2), 163–175.
Simone, D. A., Alreja, M., & LaMotte, R. H. (1991a). Psychophysical studies of the itch sensation and itchy skin (“alloknesis”) produced by intracutaneous injection of histamine. *Somatosensory & Motor Research*, *8*(3), 271–279.

Simone, D. A., Sorkin, L. S., Oh, U., Chung, J. M., Owens, C., LaMotte, R. H., & Willis, W. D. (1991b). Neurogenic hyperalgesia: central neural correlates in responses of spinothalamic tract neurons. *Journal of Neurophysiology*, *66*(1), 228–246.

Singer, E. M., Shin, D. B., Nattkemper, L. A., Benoit, B. M., Klein, R. S., Didigu, C. A., et al. (2013). IL-31 Is Produced by the Malignant T-Cell Population in Cutaneous T-Cell Lymphoma and Correlates with CTCL Pruritus. *Journal of Investigative Dermatology*, *133*(12), 2783–2785. http://doi.org/10.1038/jid.2013.227

Smith, C. H., Anstey, A. V., Barker, J. N. W. N., Burden, A. D., Chalmers, R. J. G., Chandler, D., et al. (2005). British Association of Dermatologists guidelines for use of biological interventions in psoriasis 2005. *British Journal of Dermatology*, *153*(3), 486–497. http://doi.org/10.1111/j.1365-2133.2005.06893.x

Smyth, J. M., & Stone, A. A. (2003). Ecological Momentary Assessment Research in Behavioural Medicine. *Journal of Happiness Studies*, *4*, 35–52.

Steinhoff, M., Neisius, U., Ikoma, A., Fartasch, M., Heyer, G., Skov, P. S., et al. (2003). Proteinase-activated receptor-2 mediates itch: a novel pathway for pruritus in human skin. *The Journal of Neuroscience : the Official Journal of the Society for Neuroscience*, *23*(15), 6176–6180.

Stewart, A. L., Hays, R. D., & Ware, J. E. (1988). The MOS short-form general health survey. Reliability and validity in a patient population. *Medical Care*, *26*(7), 724–735.

Summerfield, J. A., & Welch, M. E. (1980). The measurement of itch with sensitive limb movement meters. *British Journal of Dermatology*, *103*(3), 275–281.

Sun, Y.-G., & Chen, Z.-F. (2007). A gastrin-releasing peptide receptor mediates the itch sensation in the spinal cord. *Nature*, *448*(7154), 700–703. http://doi.org/10.1038/nature06029

Takaishi, M., Uchida, K., Suzuki, Y., Matsui, H., Shimada, T., Fujita, F., & Tominaga, M. (2015). Reciprocal effects of capsaicin and menthol on thermosensation through regulated activities of TRPV1 and TRPM8. *The Journal of Physiological Sciences*, *66*(2), 1–13. http://doi.org/10.1007/s12576-015-0427-y

Talbot, T. L., Schmitt, J. M., Bergasa, N. V., Jones, E. A., & Walker, E. C. (1991). Application of piezo film technology for the quantitative assessment of pruritus. *Biomedical Instrumentation & Technology / Association for the Advancement of Medical Instrumentation*, *25*(5), 400–403.

Thomas, D. L., & Diener, E. (1990). Memory accuracy in the recall of emotions. *Journal of Personality and Social Psychology*, *59*(2), 291–297. http://doi.org/10.1037//0022-3514.59.2.291

Tiling-Grosse, S., & Rees, J. (1993). Assessment of area of involvement in skin disease: a study using schematic figure outlines. *British Journal of Dermatology*, *128*(1), 69–74.

Tofte, S. (1998). Eczema area and severity index (EASI): A new tool to evaluate atopic dermatitis. *Journal of the European Academy of Dermatology and Venereology*, *11*, S197. http://doi.org/10.1016/S0926-9959(98)95291-6

Torebjörk, H. E. (1974). Afferent C units responding to mechanical, thermal and chemical stimuli in human non-glabrous skin. *Acta Physiologica Scandinavica*, *92*(3), 374–390. http://doi.org/10.1111/j.1748-1716.1974.tb05755.x

Torebjörk, H. E., & Ochoa, J. (1981). Pain and itch from C fiber stimulation. *Soc. Neurosci. Abstr.*, *7*, 228.

Torrance, G. W. (1986). Measurement of health state utilities for economic appraisal. *Journal of Health Economics*, *5*(1), 1–30.

Torrance, G. W., Feeny, D., & Furlong, W. (2001). Visual Analog Scales: Do They Have a Role in the Measurement of Preferences for Health States? *Medical Decision Making*, *21*(4), 329–334. http://doi.org/10.1177/02729890122062622

Tsuchiya, A., & Dolan, P. (2005). The QALY model and individual preferences for health states and health profiles over time: a systematic review of the literature. *Medical Decision Making*, *25*(4), 460–467. http://doi.org/10.1177/0272989X05276854

Tuckett, R. P. (1982). Itch evoked by electrical stimulation of the skin. *Journal of Investigative Dermatology*, *79*(6), 368–373.
Twycross, R., Greaves, M. W., Handwerker, H., Jones, E. A., Libretto, S. E., Szepietowski, J. C., & Zylicz, Z. (2003). Itch: scratching more than the surface. *QJM : Monthly Journal of the Association of Physicians*, *96*(1), 7–26.
Ui, H., Andoh, T., Lee, J.-B., Nojima, H., & Kuraishi, Y. (2006). Potent pruritogenic action of tryptase mediated by PAR-2 receptor and its involvement in anti-pruritic effect of nafamostat mesilate in mice. *European Journal of Pharmacology*, *530*(1-2), 172–178. http://doi.org/10.1016/j.ejphar.2005.11.021
van den Berg, J. F., van Rooij, F. J. A., Vos, H., Tulen, J. H. M., Hofman, A., Miedema, H. M. E., et al. (2008). Disagreement between subjective and actigraphic measures of sleep duration in a population-based study of elderly persons*. *Journal of Sleep Research*, *17*(3), 295–302. http://doi.org/10.1111/j.1365-2869.2008.00638.x
Vogelsang, M., Heyer, G., & Hornstein, O. P. (1995). Acetylcholine induces different cutaneous sensations in atopic and non-atopic subjects. *Acta Dermato-Venereologica*, *75*(6), 434–436.
Wahlgren, C. F. (1995). Measurement of itch. *Seminars in Dermatology*, *14*(4), 277–284.
Wahlgren, C. F., & Ekblom, A. (1996). Two-point discrimination of itch in patients with atopic dermatitis and healthy subjects. *Acta Dermato-Venereologica*, *76*(1), 48–51.
Wahlgren, C. F., Ekblom, A., & Hägermark, O. (1989). Some aspects of the experimental induction and measurement of itch. *Acta Dermato-Venereologica*, *69*(3), 185–189.
Walter, B., Sadlo, M. N., Kupfer, J., Niemeier, V., Brosig, B., Stark, R., et al. (2005). Brain activation by histamine prick test-induced itch. *Journal of Investigative Dermatology*, *125*(2), 380–382. http://doi.org/10.1111/j.0022-202X.2005.23817.x
Ware, J. E. (1976). Scales for measuring general health perceptions. *Health Services Research*, *11*(4), 396–415.
Ware, J. E., Jr., & Sherbourne, C. D. (1992). The MOS 36-ltem Short-Form Health Survey (SF-36). *Medical Care*, *30*(6), 473–483. http://doi.org/10.1097/00005650-199206000-00002
Welz-Kubiak, K., Kobuszewska, A., & Reich, A. (2015). IL-31 Is Overexpressed in Lichen Planus but Its Level Does Not Correlate with Pruritus Severity. *Journal of Immunology Research*, *2015*(5), 1–6. http://doi.org/10.1155/2015/854747
Williams, H. (2010). Subjective Dermatology Outcomes: How You Frame the Question May Not Be That Important After All. *Acta Dermato-Venereologica*, *90*(1), 3–4. http://doi.org/10.2340/00015555-0765
Wilson, S. R., Gerhold, K. A., Bifolck-Fisher, A., Liu, Q., Patel, K. N., Dong, X., & Bautista, D. M. (2011). TRPA1 is required for histamine-independent, Mas-related G protein–coupled receptor–mediated itch. *Nature Publishing Group*, *14*(5), 595–602. http://doi.org/10.1038/nn.2789
Wilson, S. R., Thé, L., Batia, L. M., Beattie, K., Katibah, G. E., McClain, S. P., et al. (2013). The Epithelial Cell-Derived Atopic Dermatitis Cytokine TSLP Activates Neurons to Induce Itch. *Cell*, *155*(2), 285–295. http://doi.org/10.1016/j.cell.2013.08.057
Wilson, T. D., Wheatley, T., Meyers, J. M., Gilbert, D. T., & Axsom, D. (2000). Focalism: a source of durability bias in affective forecasting. *Journal of Personality and Social Psychology*, *78*(5), 821–836.
Woodward, D. F., Nieves, A. L., Hawley, S. B., Joseph, R., Merlino, G. F., & Spada, C. S. (1995). The pruritogenic and inflammatory effects of prostanoids in the conjunctiva. *Journal of Ocular Pharmacology and Therapeutics : the Official Journal of the Association for Ocular Pharmacology and Therapeutics*, *11*(3), 339–347.
Wright, G., & Goodwin, P. (2002). Eliminating a framing bias by using simple instructions to 'think harder' and respondents with managerial experience: comment on “breaking the frame.” *Strategic Management Journal*, *23*(11), 1059–1067. http://doi.org/10.1002/smj.265

Yosipovitch, G., Greaves, M. W., & Schmelz, M. (2003). Itch. *Lancet (London, England)*, *361*(9358), 690–694. http://doi.org/10.1016/S0140-6736(03)12570-6

Yosipovitch, G., Zucker, I., Boner, G., Gafter, U., Shapira, Y., & David, M. (2001). A questionnaire for the assessment of pruritus: validation in uremic patients. *Acta Dermato-Venereologica*, *81*(2), 108–111.

Appendix: I

Are Subjective Accounts of Itch to be relied on? The lack of relation between Visual Analogue Itch Scores and Actigraphic Measures of Scratch. Caroline Siân Murray and Jonathan L Rees. Acta Dermato-venerologica, 2011, vol. 91(1), pp. 18-23.

Published with the kind permission of Acta Dermato-venerologica

Acta Derm Venereol 2011; 91: 18–23

INVESTIGATIVE REPORT

Are Subjective Accounts of Itch to be Relied on? The Lack of Relation Between Visual Analogue Itch Scores and Actigraphic Measures of Scratch

Caroline Siân MURRAY and Jonathan L. REES
Department of Dermatology, The University of Edinburgh, Edinburgh, United Kingdom

There is a widespread belief that subjective accounts of disease are key components of measures of disease severity and quality of life. In the present study we have set out to test this hypothesis using visual analogue scales (VAS) for itch, as a subjective measure, and actigraphy as an objective measure. One-hundred and seventeen itchy children and adults (and 25 controls) were studied for clusters of nights (total number 1,654) and actigraphy scores and VAS itch taken daily. Fifty-six percent of the night-to-night variation in actigraphy scores occurred between different individuals, while 44% was intra-subject. Neither age nor sex (children's or adults') predicted actigraphy scores, and the only significant predictor of actigraphy score was disease type ($p = 0.001$, $r^2 = 0.51$). In a multivariate model VAS itch score was not a significant determinant of actigraphy scores for either children or adults ($p = 0.26$). In order to see if there was a relation between VAS itch and actigraphy within the same patients (rather than between patients), 20 eczema patients wore the actigraph and scored VAS itch nightly for 42 nights. Little relationship was found between the actigraphy score and the VAS itch. Empirical autocorrelation analysis of VAS itch and actigraphy score reveal a clear autocorrelation for subjective VAS scores that was not found for the objective actigraphy score. Our data suggest a dissociation between scratch and perceived or recalled itch. One explanation is that VAS itch scores suffer from considerable anchoring, and context bias, and that their use in measures of disease severity is problematic. ***Key words: actigraphy; pruritus; visual analogue scale; bias.***

(Accepted August 16, 2010.)

Acta Derm Venereol 2011; 91: 18–23

Professor Jonathan Rees, Grant Chair of Dermatology, Department of Dermatology, Room 4.018, First Floor, The Lauriston Building, Lauriston Place, EH3 9HA Edinburgh, United Kingdom. E-mail: jonathan.rees@ed.ac.uk

In recent years there has been a considerable growth in interest in subjective measures of disease severity in dermatology (1–3). Much of the focus of this effort has been on attempts to measure the impact of disease on an individual sufferer, accepting that objective accounts may be inadequate to express disease impact. To take an example, the impact of psoriasis affecting, say, 50% of the body area may be greater in a person who likes to regularly swim compared with somebody who does not swim (4). An objective score such as the PASI (5) will not reflect these aspects of disease that are important (4). These arguments are widely acknowledged and accepted (6–8).

Some measures of disease severity or disease impact include individual subjective measures such as itch. An example would be the SCORAD, which includes subjective accounts of itch using a visual analogue scale (VAS) (9). Inclusion of itch in such scores is predicated on the idea that distress due to itch may provide information on suffering that is not fully accounted for by other measures (1, 10).

Measurement of itch presents its own problems (10). Itch is arguably the principal symptom of skin disease (10–12). Like pain, itch is subjective and it is difficult to see how it can be measured except by subjective account. However, much as pain has consequences – acute pain, for instance, may be accompanied by vocalisation, or an increase in heart rate – itch provokes the desire to scratch (10), a behaviour that may be independently verified. A number of strategies have been suggested for how to measure scratch and building on pioneering work by Felix & Shuster (13) we have used wristworn accelerometers or actigraphs (Actiwatch Plus, Cambridge Neurotechnology), worn at night to record scratch as a proxy measure of itch, having validated this technique against infrared videoing of children and manual scoring of scratch movements (14, 15). As compared with earlier work, our focus has been to develop strategies that can be used by subjects in their own homes and that take advantage of digital recording and analytical methods. Others have followed analogous approaches (16–18). Our goal can be considered analogous to the use of peak flow meters to monitor disease severity in asthma, in that we wish to develop a symptom tool that can be used for longitudinal monitoring of disease activity.

One unexpected finding from our previous work on children was that there was little relation between subjective accounts of itch and objective measurements recorded using an actigraph – despite the latter being

© 2011 The Authors. doi: 10.2340/00015555-1002
Journal Compilation © 2011 Acta Dermato-Venereologica. ISSN 0001-5555

validated by direct observation of scratch behaviour through infrared videoing of children in bed at night (15). This raises an important issue: assessment, recall and formulation of subjective accounts may be prone to error or artefact (19–22). If this is accepted, then whatever the virtues of subjectively recalled impacts of disease, error remains.

Our aim in the present work was to explore the relation between subjective accounts of itch using visual analogue scales, and scratch assessed objectively using actigraphs. In the process we also wished to address some limitations of our previous work (14,15): rather than examining correlations between subjects, we examined scores for each subject at multiple time points such that within-person correlations between itch and scratch could be examined. Our results are subject to more than one interpretation, but suggest that subjective measures of itch in this particular context may indeed be prone to artefact, perhaps due to anchoring bias (7, 23, 24), and that their uncritical use of them may be misleading.

METHODS

Ethical permission was received from Lothian NHS Ethics Committee for all experiments.

Study 1

Sixty-eight itchy adults and 50 itchy children were recruited. Twelve adult controls and 12 child controls also took part. The subjects were recruited from clinics at the Royal Infirmary of Edinburgh. In so far as was practical, consecutive eczema patients were approached and given information about the study. For inclusion in the study, all participants had to be older than 3 months of age and be able to provide consent or have a parent/guardian willing to do so. Subjects had to have been diagnosed by a Consultant Dermatologist as having a characteristically itchy skin condition and to have complained of an associated symptom of itch. Controls had no evidence of skin disease and reported no symptom of itch. The participants were given time to consider the invitation to take part in the study and then contacted again by telephone. Those who agreed to participate then had an appointment to meet the researcher arranged. Written, informed consent was acquired from all participants and guardians of participants.

Sample sizes were chosen opportunistically, based on availability and what was considered practical in the light of prior experiments (14, 15)

Table I details the subjects and specifies their diagnosis and characteristics. All the child subjects had atopic eczema. The broad groups of diagnoses for adult subjects were: eczema, psoriasis, cholestatic liver disease and pruritus of unknown cause (PUC). A total of 1,654 nights were studied: 1,573 subject nights and 81 control nights.

Subjects were requested to wear the actigraph on the wrist of their dominant hand (for consistency's sake, but also please see reference for further explanation (14)) for three to seven nights. Every day of the study, the participants also completed 10-cm VAS for extent, itch and insomnia. In the case of children, adults (parent/guardian) completed the scores for the participants.

Study 2

Twenty patients were recruited. All were adult eczema patients attending the Royal Infirmary of Edinburgh. For inclusion into the study, the participant had to be 16 years old or over and have atopic dermatitis, as diagnosed by a Consultant Dermatologist. They were approached as described for Study 1 and similarly gave written, informed consent. Eleven females and nine men agreed to participate. Table I gives the subjects' characteristics. (Although not a designated part of this study, SCORAD scores were available for the volunteers: the median was 34.2 with a range of 9.7 to 67.7).

All subjects were issued with an actigraph and instructed to wear it on the wrist of their dominant hand every night for the next 42 nights. A total of 761 nights were studied (patients forgot to take part on some nights).

All subjects were issued with dated sheets showing the VAS for itch, disease extent and insomnia. All of the subjects had a 'ballot' box and they were instructed to post each day's scoring sheet into the ballot box immediately after completing it. The sheet for day 0 was completed and deposition demonstrated to the researcher by the participant. On (or as near to) day 21 of the study, the participant and researcher met so that the actigraph could be checked for full functionality. On (or as near to) day 42 of the study, the researcher met with the participant to collect the actigraph and download its data, and to collect the completed forms – all, it was expected, to be contained in the ballot box.

Actigraphy

The digital accelerometer used was the Actiwatch Plus (Cambridge Neurotechnology, Cambridge, United Kingdom). This instrument detects and measures movement by utilising a piezoelectric accelerometer that logs the integration of intensity, amount and duration of movement in three axes. The Actiwatch senses and logs 32 times/s and a summation value is available for the measurement epoch. Epoch length was set at one minute. The Actiwatch is supplied with a proprietary reader that enables downloading to a computer. The digital accelerometer output was exported to software (Excel; Microsoft Ltd, Seattle, USA) for analysis. The nocturnal score was a simple sum of all the scores logged between 1 am and 5 am (a time range shown in previous unpublished studies to exclude most extraneous ambulatory movement).

Table I. *Principal demographic characteristics of the study subjects*

		Total		Age, years		Nights studied	AD	Psoriasis	PUC	Hepatic itch
		n	F:M ratio	Mean	Range	*n*	*n*	*n*	*n*	*n*
Study 1	Child subjects	50	24:26	6.92	2–15	229	50	0	0	0
	Child controls	12	10:2	10.33	6–14	42	N/A	N/A	N/A	N/A
	Adult subjects	67	40:27	43.5	16–78	1344	30	25	4	8
	Adult controls	13	8:5	38.23	24–71	39	N/A	N/A	N/A	N/A
Study 2	Adult subjects	20	9:11	40.55	16–67	761	20	0	0	0

AD: atopic dermatitis; PUC: pruritus of unknown cause; F: female; M: male. Subjects were classified according to age: child, <16 years; adult, >16 years.

Statistical analyses

Data was managed using Excel (Microsoft Ltd, Seattle, USA.) and exported to 'R' v2.9 running on a Mac OS 10.6 for all statistical analyses (25). In order to differentiate between adults completing their own subjective symptom scores and child subjects – who required an adult to complete them for them – analyses which involved examination of VAS itch scores were conducted separately for those under and over 16 years old. Actiwatch scores are non-normally distributed and were log-transformed (14, 15). For studies comparing disease groups, where data were available for more than one night, mean scores were used.

Within- and between-night analyses were performed by ANOVA using person as a factor (26). Following univariate analyses, accelerometer scores were modelled using linear regression with sex and disease treated as factors and age and VAS as continuous variables. Non-significant terms were removed from the model and levels of each factor compared using pairwise Student's *t*-tests with Holm's correction (25,26).

For Study 2, VAS itch and log actigraph scores were modelled using a mixed-effect model with person as the unit using the nmle package in 'R'. Since this does not take full account of the correlation structure, the autocorrelation function in 'R' ('acf') was used and moving model averages compared using ANOVA. Detailed exploration of different lag periods or different classes of time series models was not performed. Graphical representation of the autocorrelation structure shows to what extent a score on day n, is correlated with scores on subsequent successive days (n+1, n+2, n+3...).

RESULTS

Study 1

Determinants of actigraphy scores. For the majority of subjects, readings from three or more nights were available. Fifty-six percent of the variation was between-person and 44% within-person (i.e. due to variation from night to night for the same person) (ANOVA). Means for each subject were calculated for subsequent analyses. Children (subjects <16 years old) and adults (>16 years) were examined separately. The total data set is shown in Fig. 1.

Children. Linear regression of log actigraphy scores showed no effect of age ($p=0.67$) or sex ($p=0.365$). As expected, scores were higher for subjects with atopic dermatitis than controls: mean log actigraphy score 9.02 vs. 8.29 ($p=0.021$). VAS itch was not a significant predictor in either univariate ($p=0.261$) or multivariate ($p=0.284$) models. Fig. 2 shows the relation between actigraphy and VAS itch scores (the slope of the regression line shown is not significantly different from 0).

Adults. Sex and age had no effect on actigraphy scores. Mean log actigraphy scores were 10.4, 8.95, 8.43, 8.99 and 8.14 in patients with liver disease, atopic dermatitis, psoriasis, pruritus of unknown cause and controls respectively. Measures of uncorrected effect can be seen in Fig. 1, and a matrix of Holm-corrected pairwise

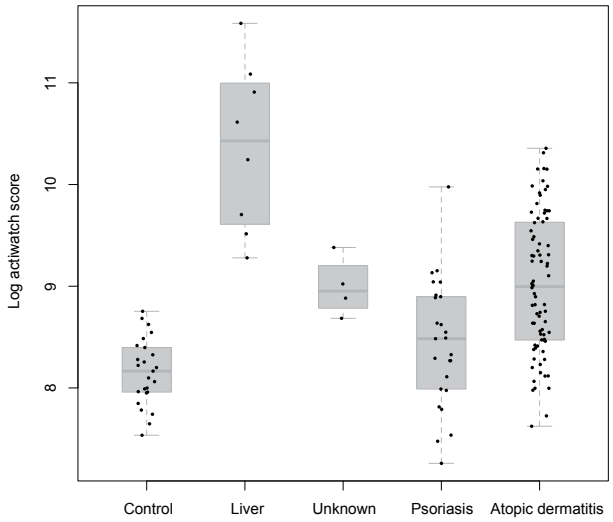


Fig. 1. Study 1. Mean overnight actigraphy score for all participants (children and adults) separated by diagnosis: control, liver (=hepatic itch), unknown (=pruritus of unknown cause), psoriasis and atopic dermatitis.

Student's *t*-tests performed following linear regression using pooled error is shown in Table II. As can be seen, scores were significantly higher for subjects with liver disease compared with other disease groups and controls. Scores for subjects with atopic dermatitis, but not those with psoriasis, were significantly higher than in controls.

Mean VAS itch scores were 6.64, 4.83, 4.68 and 2.53 in patients with liver disease, psoriasis, atopic dermatitis and patients with itch of unknown cause, respectively. The only significant difference between these groups was between those with liver disease and those with itch of unknown cause ($p=0.04$; Student's *t*-test). The spread of the values within each group was large.

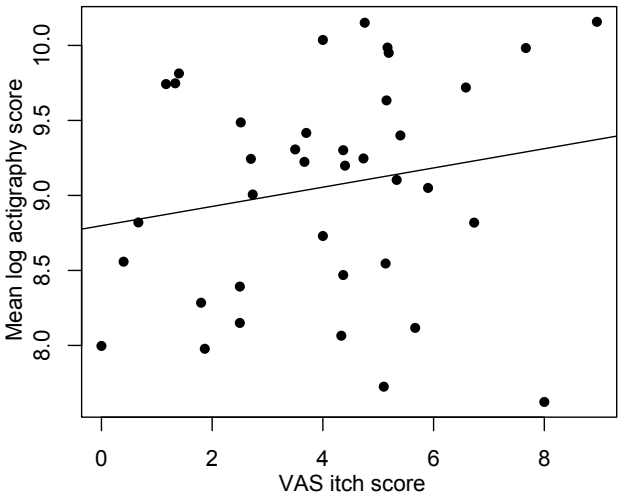


Fig. 2. Study 1. Regression analysis of children's mean log overnight actigraphy score and visual analogue scale (VAS) itch score. A 'best-fit' regression line (slope not significantly different from 0) is shown.

Table II. *Summary of principal differences between actigraphy scores for adults by diagnostic group. Pairwise comparisons are made using Student's* t*-tests with pooled error following linear regression. Mean log scores are: control, 8.14; liver, 10.4; PUC, 8.99; psoriasis, 8.43 and AD, 8.95.*

	Control	Liver	PUC	Psoriasis
Liver	<0.0001	–	–	–
PUC	0.0782	0.0034	–	–
Psoriasis	0.3564	0.0001	0.3033	–
AD	0.0015	0.0001	0.8884	0.0184

PUC: pruritus of unknown cause; AD: atopic dermatitis.

In univariate analyses both diagnosis ($p<0.001$) and VAS itch ($p=0.042$) were significant predictors of actigraphy scores, but the r2 value for VAS itch was only 0.06 (r^2 for diagnostic category = 0.51). Once diagnosis was entered into the regression equation, VAS itch was no longer significant ($p=0.27$). The relation, by diagnostic group, between VAS itch score and actigraphy is shown in Fig. 3 (best-fit lines are shown, but their slopes do not differ significantly from 0).

Study 2

In a different cohort of 20 adults with eczema, actigraphy and VAS itch scores were recorded over 42 nights. Scatter plots of VAS itch against actigraphy are shown for each of the twenty persons in Fig. 4. It can be seen that there appears to be little relation between actigraphy and VAS itch scores. Because scores from night to night for the same person cannot be assumed to be independent, we initially modelled the data using a mixed-effect model with person as the grouping factor. In order to take account of the correlation structure, we then compared this model (using ANOVA) with a simple moving average model using the corARMA function (25, 26) with a delay of 2 or 3 days. This model provided a better fit to the data ($p<0.001$). The empirical autocorrelation structure for VAS itch is shown in Fig. 5, with a dashed line representing statistical significance ($p=0.001$) plotted. The autocorrelation bar crossing the $p=0.001$ dashed line indicates that the data on successive days are highly correlated. This is most evident for lags of 1 to 3 days (the correlation on day 0 is of course perfect as the same figure is being compared), but the stepwise diminution of the height of the autocorrelation bar reflects a diminution of the correlation from any one day (n) over subsequent successive days (n+1, n+2, n+3...). No pattern for autocorrelation was seen for actigraphy scores (data not shown).

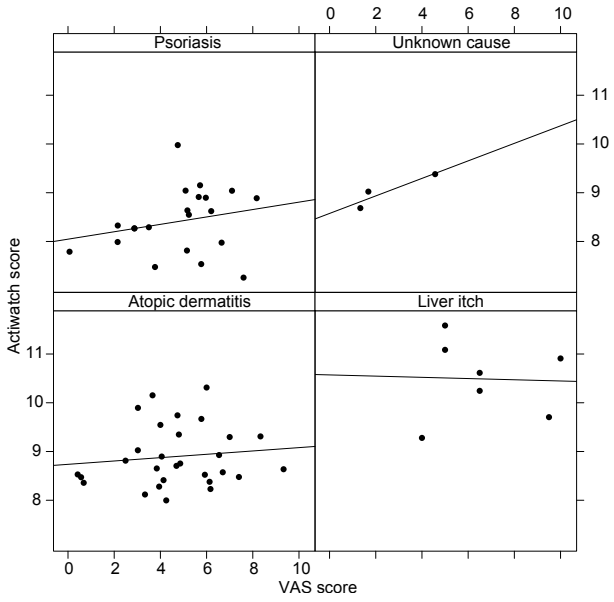


Fig. 3. Study 1. Regression analysis of adults' Actiwatch mean log overnight score and visual analogue scale (VAS) itch score separated by diagnostic group: psoriasis, unknown cause (=pruritus of unknown cause), atopic dermatitis and liver itch (=hepatic itch.) 'Best-fit' regression lines are shown (none of whose slopes is significantly different from 0).

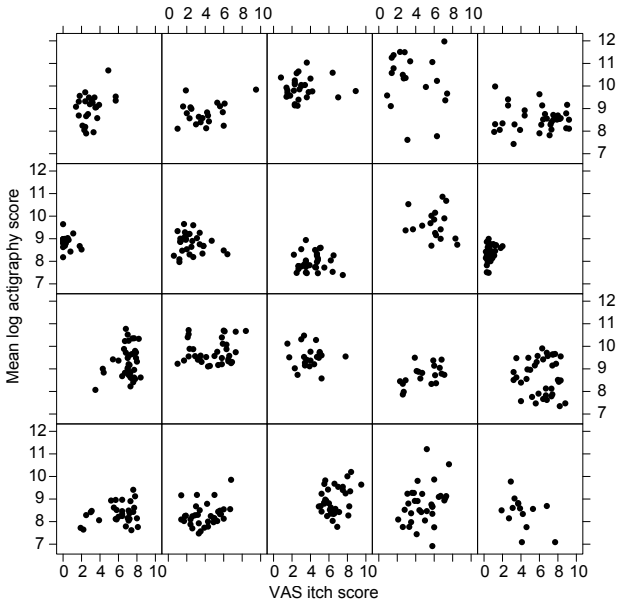


Fig. 4. Study 2. Individual scatter plots of visual analogue scale (VAS) itch against mean log overnight actigraphy score for every participant.

DISCUSSION

The current work builds upon previous studies and in large part confirms that the relation between objective measures of scratch and VAS itch is poor. Before discussion, some limitations of the work are highlighted.

The work was based on those attending secondary care, and therefore was probably skewed towards patients with more severe disease, as reflected by the SCORAD scores. Patients were volunteers and, since the study involved a significant amount of time and contact with the investigators, the study group may not

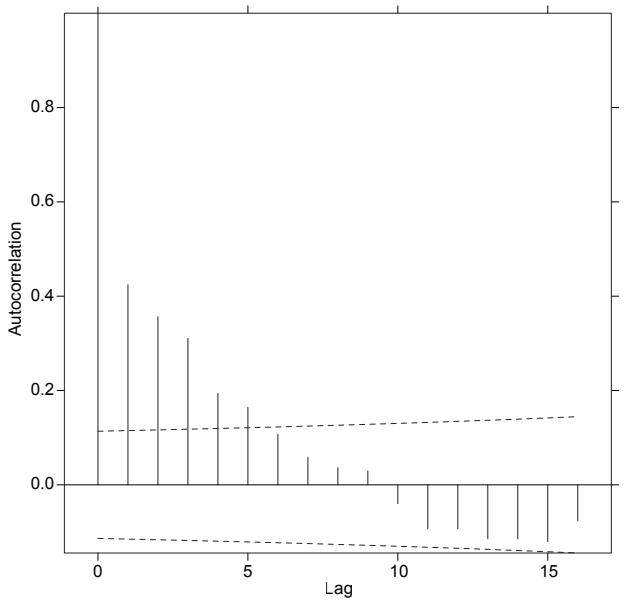


Fig. 5. Study 2. Graph demonstrating the autocorrelation structure of grouped individuals' visual analogue scale itch scores. Dashed line: statistical significance ($p=0.001$) (autocorrelation lines crossing this bar indicate statistical significance).

be typical of the reference population. We obviously were not able to control prospectively the severity of disease during the study period – if patient groups at the extreme end of the scale for symptom severity had been used, it is possible different conclusions might have been made, although the SCORAD scores for the eczema group were as high or higher than in many therapeutic trials.

In study 1 we have shown that there is no obvious relation between itch assessed using VAS scores and actigraphy scores for either children with atopic dermatitis or adults with a range of diagnoses (see Fig 2 and 3. Previous work has validated actigraphy against direct observation of children scratching at night, and both scratching and restlessness are correlated with each other, and higher in those with atopic dermatitis than in controls (14, 15). Itch-related behaviour at night is not as stereotyped as when awake, and actigraphy is, we believe, a useful practical assay for scratch when compared with the time-consuming nature of direct observation of video recordings. Although we analysed children and adults separately, we saw no influence of sex or age on actigraphy scores. The key determinant of differences in actigraphy scores was diagnostic group, with those with liver disease the most severely affected. This assumes, however, that restlessness due to other factors differing between diagnostic groups is not confounding our scores – an assumption that needs to be tested. Although VAS itch scores differed between the different diagnostic groups, the differences were largely non-significant, again emphasising a disconnect between VAS scores and actigraphy.

In study 1 we showed that approximately 60% of the recorded variation was between-subject and approximately 40% within-subject. This fact emphasises that nocturnal movements vary from night to night within any one subject, and of course limits the power of actigraphy scores for experimental studies unless repeated sampling is carried out. It is quite possible that other factors – night of the week and work patterns – contribute to this within-person variation, as well as variation in the disease that is causing the scratch. We have not specifically examined these factors.

In study 2 we had a chance to look in more detail at the relation between VAS itch and actigraphy over time. Again, we saw little convincing relation, and examination of the autocorrelation shows a very different pattern for actigraphy and VAS itch. The VAS itch scores clearly show an effect of lag, and the question is: why? We asked our subject to post their scores in an attempt to minimise the filling out of results in batches, and to limit knowledge of the score on one day influencing the score on the next day. This strategy is not foolproof and electronic diaries or other methods of continuous sampling would perhaps provide a better approach (27, 28). We suspect that the presence of the temporal pattern for VAS itch scores (and not actigraphy) is more easily explained by subjects anchoring their scores based on recollection of previous scoring (7, 23). Whether or not this is indeed the correct interpretation, coupled with the results of study 1, our data suggest that the use of VAS scores for itch may be limited in their utility as a subjective measure of disease activity in the context of the chronic diseases we studied. We suspect that objective scores such as actigraphy – however imperfect – deserve more attention for monitoring disease activity.

ACKNOWLEDGEMENT

This work was supported by the Wellcome Trust (grant number 076754).

REFERENCES

1. Finlay AY, Khan GK. Dermatology Life Quality Index (DLQI) – a simple practical measure for routine clinical use. Clin Exp Dermatol 1994; 19: 210–216.
2. Both H, Essink-Bot ML, Busschbach J, Nijsten T. Critical review of generic and dermatology-specific health-related quality of life instruments. J Invest Dermatol 2007; 127: 2726–2739.
3. Chen SC. Dermatology quality of life instruments: sorting out the quagmire. J Invest Dermatol 2007; 127: 2695–2696.
4. Rees JL. Trials, evidence, and the management of patients with psoriasis. J Am Acad Dermatol 2003; 48: 135–143.
5. Fredriksson T, Pettersson U. Severe psoriasis – oral therapy with a new retinoid. Dermatologica 1978; 157: 238–244.
6. Nord E. Cost-value analysis in health care: Making sense out of QALYs. Cambridge: Cambridge University Press, 1999.

7. Bjorner JB, Fayers P. Self-rated health. In: Fayers P, Hays R, Idler E, eds. Assessing quality of life in clinical trials: methods and practice. Oxford: Oxford University Press, 2005: 309–323.
8. Basra MK, Fenech R, Gatt RM, Salek MS, Finlay AY. The Dermatology Life Quality Index 1994-2007: a comprehensive review of validation data and clinical results. Br J Dermatol 2008; 159: 997–1035.
9. Stalder JF, Taieb A, Atherton DJ, Bieber T, Bonifazi E, Broberg A, et al. Severity scoring of atopic dermatitis: the SCORAD index. Consensus Report of the European Task Force on Atopic Dermatitis. Dermatology 1993; 186: 23–31.
10. Wahlgren CF. Measurement of itch. Semin Dermatol 1995; 14: 277–284.
11. Shuster S, Fisher GH, Harris E, Binnell D. The effect of skin disease on self image [proceedings]. Br J Dermatol 1978; 99 (Suppl 16): 18–19.
12. Rees JL, Laidlaw A. Pruritus: more scratch than itch. Clin Exp Dermatol 1999; 24: 490–493.
13. Felix R, Shuster S. A new method for the measurement of itch and the response to treatment. Br J Dermatol 1975; 93: 303–312.
14. Benjamin K, Waterston K, Russell M, Schofield O, Diffey B, Rees JL. The development of an objective method for measuring scratch in children with atopic dermatitis suitable for clinical use. J Am Acad Dermatol 2004; 50: 33–40.
15. Bringhurst C, Waterston K, Schofield O, Benjamin K, Rees JL. Measurement of itch using actigraphy in pediatric and adult populations. J Am Acad Dermatol 2004; 51: 893–898.
16. Ebata T, Iwasaki S, Kamide R, Niimura M. Use of a wrist activity monitor for the measurement of nocturnal scratching in patients with atopic dermatitis. Br J Dermatol 2001; 144: 305–309.
17. Bender B, Leung S, Leung D. Actigraphy assessment of sleep disturbance in patients with atopic dermatitis: An objective life quality measure. J Allergy Clin Immunol 2003; 111: 598–602.
18. Hon KL, Lam MC, Leung TF, Kam WY, Lee KC, Li MC, et al. Nocturnal wrist movements are correlated with objective clinical scores and plasma chemokine levels in children with atopic dermatitis. Br J Dermatol 2006; 154: 629–635.
19. Redelmeier DA, Kahneman D. Patients' memories of painful medical treatments: real-time and retrospective evaluations of two minimally invasive procedures. Pain 1996; 66: 3–8.
20. Redelmeier DA, Katz J, Kahneman D. Memories of colonoscopy: a randomized trial. Pain 2003; 104: 187–194.
21. Broderick JE, Schwartz JE, Vikingstad G, Pribbernow M, Grossman S, Stone AA. The accuracy of pain and fatigue items across different reporting periods. Pain 2008; 139: 146–157.
22. Stone AA, Broderick JE, Kaell AT, DelesPaul PA, Porter LE. Does the peak-end phenomenon observed in laboratory pain studies apply to real-world pain in rheumatoid arthritics? J Pain 2000; 1: 212–217.
23. Kahneman D, Tversky A editors. On the reality of cognitive illusions. Psychol Review 1996; 103: 582–586.
24. Kahneman D, Tversky A. Evaluation by Moments: Past and Future. In: Choices, Values and Frames. New York: Cambridge University Press and the Russell Sage Foundation, 2000.
25. R Development Core Team. R: A language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing, 2009. Available from: http://www.R-project.org
26. Crawley MJ. The R Book.Chichester, England: John Wiley and Sons Ltd, 2007.
27. Kahneman D, Krueger AB, Schkade DA, Schwarz N, Stone AA. A survey method for characterizing daily life experience: The day reconstruction method. Science 2004; 306: 1776–1780.
28. Langan SM, Bourke JF, Silcocks P, Williams HC. An exploratory prospective observational study of environmental factors exacerbating atopic eczema in children. Br J Dermatol 2006; 154: 979–980.

Appendix: II

How Robust are the Dermatology Life Quality Index and Other self-reported Subjective Symptom Scores when Exposed to a Range of Experimental Biases? Caroline Siân Murray and Jonathan L Rees. Acta Dermato-venerologica, 2010, vol. 90 (1), pp. 34-38.

Published with the kind permission of Acta Dermato-venerologica

Acta Derm Venereol 2010; 90: 34–38

INVESTIGATIVE REPORT

How Robust are the Dermatology Life Quality Index and Other Self-reported Subjective Symptom Scores when Exposed to a Range of Experimental Biases?

Caroline S. MURRAY and Jonathan L. REES
Department of Dermatology, University of Edinburgh, Edinburgh, United Kingdom

Subjective-symptom tools used in dermatology have rarely been experimentally tested for cognitive "focus" and "framing" biases. We investigated the effects of affective biases on the Dermatology Life Quality Index (DLQI), the Global Health Question and visual analogue scores. Two experiments tested the response to affect-eliciting words and film. We demonstrated no significant difference in median DLQI scores for subjects exposed to negative vs. neutral words (medians 8.5 and 9.5, respectively), or negative vs. positive words (medians 6.0 and 9.0, respectively, overall $p=0.41$.) Median DLQI scores were similar for groups who had (8.0), or had not (9.0), seen a video clip about a severe skin condition ($p=0.34$). Finally, we compared an Amended DLQI (ADLQI), the DLQI re-worded into neutral "frames", with the standard DLQI. ADLQI median scores were higher (ADLQI 8.25, DLQI 6.75), but not significantly so ($p=0.47$). We have been unable to demonstrate any effects of the biases studied, but the statistical power of our study is modest.
Key words: dermatology; bias; quality of life.

(Accepted September 7, 2009.)

Acta Derm Venereol 2010; 90: 34–38.

Caroline Siân Murray, Department of Dermatology, Room 4.018, First Floor, The Lauriston Building, Lauriston Place, Edinburgh, EH3 9HA, UK. E-mail: carolinesianmurray@ed.ac.uk

It is widely accepted that objective measures of disease based on patho-biological variables are insufficient to measure the personal impact of disease. There are at least two reasons for this. First, we do not have objective correlates of many states, such as pain or itch (1, 2), and secondly, how the disease process affects an individual person will depend on a range of individual and contextual factors. For instance, the visibility of extensive psoriasis may be a greater burden to an individual who likes to go swimming, than to an individual who never goes swimming. Another example is that a person who previously had very severe disease, for instance bad childhood eczema, will use that as a comparator for their present state: their current disease state might be viewed differently if there was no previous history of skin disease (3).

In recent years, a number of tools have been developed with the goal of measuring the functional burden of disease as experienced by the patient. Examples include the Global Health Questionnaire (GHQ), which has been used to provide an overall assessment of patient-perceived health (4), whilst another generic tool, the visual analogue scale (VAS), has been used as a measure of a range of symptoms such as itch or pain (5–7). For skin disease, one of the most widely used tools is the Dermatology Life Quality Index (DLQI) (8, 9). The designers of the tool had practicality in mind – it was considered a priority that the tool should be short and quick to answer, in order to aid assessment during clinic visits, and thus this 10-question tool was designed. In a medical and economic climate where resources are scarce, the assessment of quality of life and how it may be influenced by medical intervention has become a major research programme. In England, the British Association of Dermatologists and the National Institute of Health and Clinical Excellence (NICE) has advocated the use of the DLQI as a disease assessment tool for patients with psoriasis, to determine whether they should receive certain expensive biological therapies (10) (http://www.nice.org.uk/Guidance/TA134).

Work over the last 25 years, in cognitive psychology and especially in the field of happiness research, has revealed a number of problems surrounding the measurement of subjective states, quality of life and utility (for reviews, see Kahneman et al. (11)). First, and strange as it may initially appear, individuals may not be able to access their own feelings (12, 13), and the way in which information is gathered may alter, or influence, the patient's own perception of their own feelings (3, 14–16). Secondly, a number of cognitive limitations may limit the value of subjective knowledge: patients may not be able to remember changes in their functional status, nor predict the effects of particular interventions or change in state (12, 13, 17–19).

In the present paper, we set out to explore the effects of contextual or "framing" biases on commonly used subjective measures including the DLQI in dermatology. We use the term "framing bias" widely to include bias that stems from how the feeling, emotion or symptom is enquired about. Questions can be "framed" in language or presented in a context that may elicit a stereotyped

© 2010 The Authors. doi: 10.2340/00015555-0768
Journal Compilation © 2010 Acta Dermato-Venereologica. ISSN 0001-5555

answer; for instance, by implying that an aspect of disease should be considered as a negative phenomenon, leading a respondent to consider this aspect as negative where they did not before (3). Secondly, the immediate context may alter how individuals perceive their own symptoms. For example, patients frequently anchor or skew their own assessment of disease by reference to others who they think are less or more fortunate (3, 5).

We therefore designed three experiments in which either the wording of the DLQI was altered, or the immediate context in which individuals completed the DLQI, GHQ or VAS for common symptoms was manipulated. The manipulation was performed using video, listing of negative or neutral words, and alterations in the actual wording of the DLQI.

METHODS

Participants

An opportunistic sample of 215 patients was recruited. Because of the absence of similar prior work, formal power calculations were not performed. For each study, consecutive patients who agreed to take part were enrolled from the Royal Infirmary's Department of Phototherapy in Edinburgh. Details of specific diagnoses were not sought, the usual throughput of the phototherapy department would suggest that the majority (70%) of the patients had psoriasis, a minority (approximately 10%) had eczema and 20% other conditions (for instance generalized pruritus). Ethics committee approval was granted by the Lothian Ethics Committee (LREC reference: 06/S1104/56).

All study procedures and patient interactions were conducted using a consistent, written script. The interaction script and subject information sheet explained that the studies were to determine which sort of questionnaire or score was most accurate in assessing symptoms. The interventions were described in general terms ("You will be given a list of words to memorise" or "If you are randomised into a certain group you may see a film broadcast on terrestrial television") in order to minimize unintentional "unblinding".

All participants completed the GHQ ("In general, for someone of your age, would you say that your health is excellent, very good, good or poor?"), DLQI and VAS of disease extent, itch and insomnia, always in the same order. The DLQI is a 10-question tool, the score of which is acquired by summing the score for each question. The higher the DLQI score, the more severely their quality of life is affected (maximum score 30.) Most participants scored in the region of 6–10, which equates to a "moderate" effect of the skin condition on the quality of life.

Experiment 1

Our hypothesis was that, if subjects were exposed to certain mood-eliciting words, they would affect the subjective score accordingly, for instance, if they had read negative words, their subjective scores would suggest worse disease.

Forty patients were randomized into two groups. Group 1 were asked to read 10 negative words, had one minute to memorize them and were then asked to write them out. After this, they completed the GHQ, DLQI and the VAS of disease extent for both itch and insomnia. Group 2 went through an identical process, but the participants were given a list of 10 neutral words (certain fields of psychology research have identified and use words that elicit certain affective states.) The words for this part of the experiment are listed in Table I and were taken from the "Balanced Affective Word List" (http://www.sci.sdsu.edu/CAL/wordlist/origwordlist.html). The words were matched with respect to total character and syllable length. A further 41 patients were randomized into two groups. Group 3 were asked to read 10 negative words and then write them out (without the necessity of memorizing them). The participants then completed GHQ, DLQI and VAS or disease extent, itching and insomnia. Group 4 went through an identical process, except participants were given a list of 10 positive words. These words were taken from the University of Florida's NIMH Centre for the Study of Emotion and Attention (http://csea.phhp.ufl.edu/Media.html#bottommedia) and are listed in Table I. The source of affect-eliciting words was altered as this afforded a larger scope of words with more recent and more extensive validation. Again, the words were matched for total character and syllable length.

Experiment 2

Our hypothesis was that if a subject saw a film highlighting the negative aspects of having a skin disease, then this would make them focus on the negative aspects of living with their skin disease and so their subjective scores would imply that they had worse disease.

Fifty-four patients were randomized into two groups, with Group 1 completing GHQ then VAS for disease extent, itch and insomnia, after having watched a 10-min clip from a terrestrial television broadcast ("Real Families: My Skin Could Kill Me", which was broadcast before the "watershed" on ITV1 in October 2005) about living with the severe skin condition, Harlequin ichthyosis. Group 2 just completed the subjective tools without having watched the television clip. All subjects were questioned in the same way and in the same experimental room, whether or not they had watched the film. The randomization result (to watch the film or not) was included in the questionnaire envelope and was opened, with the interviewer present in the study room, the interviewer then adopted the appropriate script (for whether or not the participant was to watch the film) from that point.

Table I. *Experiment 1: words presented to each intervention group*

Negative words (Group 1)	Neutral words (Group 2)	Positive words (Group 3)	Negative words (Group 4)
worry	wagon	angel	abuse
ashamed	aluminium	birthday	bankrupt
gloom	green	beauty	betray
bad	bus	caress	cancer
sick	scan	cheer	cruel
suffering	submarine	freedom	funeral
unhappy	vitamin	glory	gloom
itch	iron	humour	hatred
misery	margin	home	hurt
rejected	resident	joke	jail
		mother	misery
		pretty	poison
		passion	pollute
		reward	rabies
		romantic	rejected
		sun	sad
		sexy	sick
		snuggle	suicide
		treasure	terrible
		triumph	tragedy

Table II. *Experiment 1: subject characteristics, median scores and p-values of Kruskal-Wallis analysis of variance (ANOVA)*

	Age, years mean (range)	M/F	Median DLQI (interquartile range)	Median GHQ (interquartile range)	Median VAS (interquartile range)		
					Extent	Itch	Insomnia
Group 1 (Neg) $n=20$	39.8 (17–71)	12/8	8.50 (5.00–11.75)	2.00 (1.00–2.00)	3.60 (2.80–5.60)	2.15 (0.90–4.33)	3.60 (1.50–7.20)
Group 2 (Neut) $n=20$	41.4 (18–68)	9/11	9.50 (5.75–14.25)	2.00 (2.00–2.00)	3.70 (1.80–5.30)	2.15 (1.08–4.40)	4.80 (2.85–5.73)
Group 3 (Neg) $n=19$	39.0 (20–72)	10/9	6.00 (2.00–10.50)	2.00 (2.00–3.00)	2.50 (0.60–5.00)	1.50 (0.55–3.95)	1.90 (0.70–5.75)
Group 4 (Pos) $n=22$	37.4 (16–74)	11/11	9.00 (2.25–12.50)	2.00 (2.00–3.00)	4.80 (2.50–7.50)	1.30 (0.40–3.00)	3.40 (1.30–6.00)
p=Kruskal-Wallis			0.41	0.44	0.35	0.46	0.27

DLQI: Dermatology Life Quality Index; GHQ: Global Health Question; VAS: visual analogue scale.

Experiment 3

In this study, our hypothesis was that if the DLQI focused on negative aspects of disease, then re-framing it into "neutral" frames should result in scores implying a better quality of life. We also hypothesized that if the DLQI focused on the negative, then this may negatively affect the responses to other subjective symptom scores.

Eighty patients were randomized into two groups and each of these two groups further split into two sub-groups, giving a total of four sub-groups. Half the subjects answered the GHQ and standard DLQI, whilst the other half answered an altered DLQI (ADLQI) and the standard GHQ. The ADLQI mirrored the standard DLQI, but an attempt was made for each question to be re-written in a neutral frame, thereby, minimizing the possibility of a positive or negative framing and potentially reducing the possibility of a stereotyped answer. The ADLQI is shown in the electronic appendix (http://adv.medicaljournals.se/article/abstract/10.2340.00015555-0768/app1). Division of the two groups allowed the ordering of the examination to be manipulated, with half the subjects receiving the GHQ first, and then either the DLQI or the ADLQI, with the other half receiving the GHQ second.

Demographic variables including age and sex, together with the results, were de-identified and recorded in Excel. Statistical analyses were undertaken using R-software (http://www.R-project.org (20)).

RESULTS

Examination of raw data, not surprisingly, showed that the majority of variables were non-normally distributed. Medians were therefore compared using the Kruskal-Wallis (KW) analysis of variance (ANOVA), or for count data, Fisher's exact test for r × c contingency tables. Formal significance was taken at $p<0.05$. Because of the limited range of the GHQ questionnaire, results were also examined using Fisher's exact test, but this did not alter any of the conclusions and is not presented.

Experiment 1

The impact of affect-eliciting words. A total of 81 subjects were studied and their characteristics are shown in Table II. There were four intervention groups, numbered 1–4, as mentioned above. There were no significant differences in the sex allocation (Fishers test, $p=0.81$) nor median ages (Kruskal-Wallis, $p=0.70$) between the four groups. Median scores for the four groups and p-values using the Kruskal-Wallis ANOVA are shown in Table II. As can be seen there are no significant differences evident.

Experiment 2

The impact of watching a film about living with a severe skin condition. A total of 54 subjects were studied. Their characteristics and the median group-scores and Kruskal-Wallis ANOVA are shown in Table III.

There was no significant sex difference between the two groups, those who were shown the video and those who were not (Fisher, $p=0.27$). The median age of those shown the video was 52 compared with those who were not shown the video of 33; a difference that is highly significant (KW, $p=0.001$). However, scatter plots did not show any obvious correlation between

Table III. *Experiment 2: subject characteristics, median scores and p-values for Kruskal-Wallis analysis of variance (ANOVA)*

	Age, years mean (range)	Sex M/F	Median DLQI (interquartile range)	Median GHQ (interquartile range)	Median VAS (interquartile range)		
					Extent	Itch	Insomnia
Group 1 (video) $n=27$	50.3 (17–77)	14/13	8.00 (4.50–9.00	3.00 (2.00–3.00)	4.50 (2.95–6.15)	2.00 (0.85–3.90)	3.90 (2.10–6.35)
Group 2 (no video) $n=27$	35.8 (16–74)	9/18	9.00 (4.50–15.00)	2.00 (1.50–3.00)	3.20 (1.65–6.40)	2.70 (0.95–5.15)	2.20 (1.45–5.05)
p=Kruskal-Wallis			0.34	0.20	0.50	0.11	0.21

DLQI: Dermatology Life Quality Index; GHQ: Global Health Question; VAS: visual analogue scale.

Table IV. *Experiment 3: subject characteristics, median scores and p-values of Kruskal-Wallis analysis of variance (ANOVA)*

	Age, years mean (range)	Sex M/F/Un known	Median QI (interquartile range)	Median GHQ (interquartile range)
Group 1 (DLQI then GHQ) $n=21$	45.1 (28–68)	14/6/1	6.50 (2.75–10.75)	3.00 (2.00–3.00)
Group 2 (GHQ then DLQI) $n=21$	42.8 (16–79)	9/11/1	7.00 (4.50–11.00)	2.00 (2.00–3.00)
Group 3 (ADLQI then GHQ) $n=19$	42.3 (19–81)	9/9/1	8.00 (6.75–10.50)	2.00 (1.00–3.00)
Group 4 (GHQ then ADLQI) $n=19$	44.0 (17–76)	8/11/	8.50 (6.00–13.50)	2.00 (1.75–3.00)
$p=$Kruskal-Wallis			0.47	0.76

DLQI: Dermatology Life Quality Index; ADLQI: Amended Dermatology Life Quality Index; GHQ: Global Health Question; QI: Quality of Life Index score.

age and the outcome measures, so this difference was ignored. Median scores and *p*-values for the Kruskal-Wallis ANOVA are shown in Table IV. As can be seen, there are no significant differences evident.

Experiment 3

Re-wording DLQI into neutral frames. A total of 80 subjects were studied and their characteristics are summarized in Table IV. GHQ and quality of life scores (QI) were examined following four "treatments". The first "treatment", the DLQI, was compared with the second "treatment", ADLQI and, following this, the ordering of GHQ and DLQI/ADLQI were studied (hence treatment groups were numbered as follows: 1, 2 (DLQI) and 3, 4 (ADLQI.)

There were no significant differences in sex (Fisher, $p=0.17$) or age (KW, $p=0.92$) between the four groups. Median scores and Kruskal-Wallis ANOVA across the four groups for QI (DLQI and ADLQI) and GHQ are listed in Table IV. These differences are not significant (KW for QI $p=0.47$ and GHQ $p=0.76$). The ordering had no effect on GHQ ($p=0.60$) or QI ($p=0.5$) scores and therefore groups 1 and 2, and 3 and 4 were combined. Medians for the ADLQI and the DLQI for these combined groups were 8.5 and 7, respectively, a difference that was close to statistical significance with a *p*-value of 0.07 (Kruskal-Wallis test).

DISCUSSION

The results presented are, essentially, negative and, in that sense, they can be viewed as reassuring. Using the criteria of statistical significance we were unable to significantly alter the scores with the various attempts at manipulation of the context or wording of the questionnaires of VAS. There are a number of limitations to the work we present.

Although we studied 215 subjects, we did so in the absence of formal power calculations and a type II error is always possible. Whereas, if the effect of any biases had been major, then we may have detected it, more modest effects will probably have gone undetected. We cannot rule out clinically relevant effects, although our data provide the effect estimates for future studies. Secondly, even within the experimental paradigm we adopted, there were limitations to the way the experiments were carried out. For instance, although we used a video of a child affected by skin disease, we did not find a suitable video that we thought was meaningful to use as a control. We also found it extremely difficult to alter the wording of the DLQI without producing a caricature of it. The differences seen between the altered DLQI and the genuine DLQI approach significance, but of course, interpretation of these differences is not straightforward. The fact that a different questionnaire produces a different median score is not unexpected and, even if the difference had been significant, it does not invalidate, in any way, the use of the DLQI. Another facet of this experiment is that it demonstrates that the DLQI itself would not appear to bias answering of other scores: the GHQ scores were similar whether the participants had been exposed to DLQI or to the supposedly neutral-framed ADLQI.

Although we have not demonstrated any effects of framing or contextual factors in our study, the study itself was experimental and may not reproduce the sorts of real life factors that will influence the way people respond to questionnaires. For instance, and rather mundanely, a patient whose appointment has been delayed excessively, one can imagine, might be considered more likely to weight his or her own disease more heavily. It would be difficult to capture such influences. Furthermore the use of measures such as DLQI as justification for therapy (or denial of therapy) as in the UK is also much more complex than some appreciate (21). Clinical anecdote suggests that patients are quite capable of "gaming" the system to achieve what they feel are appropriate, and one should remember that quality of life, health status and patients' perception of these measures are distinct (21). It is difficult not to imagine that if patients are meaningfully consented, and the purpose of the DLQI as a justification of clinical need is explained, that patients will not moderate their answers accordingly.

Finally, it was not our purpose to compare different questionnaires, or measures of aspects of diseases. There is already a large literature on this and on the advantages and disadvantages of speciality or disease-specific scoring systems vs. more generic questionnaires, such as

EuroQOL or SF-36, for instance (22–24). We do feel that it is important, however, that in view of the fact that there is an increasing literature on the design, use, limitations of various disease-scoring systems and on cognitive psychology as a whole, that this information is acknowledged and used to continue to validate the subjective tools that we commonly use.

ACKNOWLEDGEMENTS

Professor A. Y. Finlay for his allowing us to use DLQI.
This study was supported by Wellcome Project Grant – 076754

The authors declare no conflict of interest.

REFERENCES

1. Rees J, Murray CS. Itching for progress. Clin Exp Dermatol 2005; 30: 471–473
2. Bringhurst C, Waterston K, Schofield O, Benjamin K, Rees JL. Measurement of itch using actigraphy in pediatric and adult populations. J Am Acad Dermatol 2004; 51: 893–898.
3. McColl E, Fayers P. Context effects and proxy assessments. In: Fayers P and Hays R, editors. Assessing quality of life in clinical trials. Second edn. Oxford: Oxford University Press; 2005, p. 131–146.
4. Bjorner JB. Self-rated health. In: Fayers P and Hays R, editors. Assessing quality of life in clinical trials. Second edn. Oxford: Oxford University Press; 2005, p. 309–323.
5. Wahlgren CF. Measurement of itch. Semin Dermatol 1995; 14: 277–284.
6. Wahlgren CF. Itch and atopic dermatitis: clinical and experimental studies. Acta Derm Venereol Suppl 1991; 165: 1–53.
7. Hägermark O, Wahlgren CF. Some methods for evaluating clinical itch and their application for studying pathophysiological mechanisms. J Dermatol Sci 1992; 4: 55–62.
8. Finlay AY, Khan GK. Dermatology Life Quality Index (DLQI) – a simple practical measure for routine clinical use. Clin Exp Dermatol 1994; 19: 210–216.
9. Lewis V, Finlay AY. 10 years experience of the Dermatology Life Quality Index (DLQI). J Invest Dermatol Symp Proc 2004; 9: 169–180.
10. Smith CH, Anstey AV, Barker JN, Burden AD, Chalmers RJ, Chandler D, et al. British Association of Dermatologists guidelines for use of biological interventions in psoriasis 2005. Br J Dermatol 2005; 153: 486–497.
11. Kahneman D, Diener E, Schwarz R. Well-being: the foundations of hedonic psychology. 1st edn. Schwartz R, editor. New York: Russell Sage Foundation Publications; 2003.
12. Redelmeier DA, Kahneman D. Patients' memories of painful medical treatments: real-time and retrospective evaluations of two minimally invasive procedures. Pain 1996; 66: 3–8.
13. Redelmeier DA, Katz J, Kahneman D. Memories of colonoscopy: a randomized trial. Pain 2003; 104: 187–194.
14. Robinson MD, Clore GL. Episodic and semantic knowledge in emotional self-report: evidence for two judgment processes. J Pers Soc Psychol 2002; 83: 198–215.
15. Robinson MD, Clore GL. Belief and feeling: evidence for an accessibility model of emotional self-report. Psychol Bull 2002; 128: 934–960.
16. Kahneman D, Krueger AB, Schkade DA, Schwarz N, Stone AA. A survey method for characterizing daily life experience: the day reconstruction method. Science 2004; 306: 1776–1780.
17. Fredrickson BL, Kahneman D. Duration neglect in retrospective evaluations of affective episodes. J Pers Soc Psychol 1993; 65: 45–55.
18. Kahneman D, Krueger AB, Schkade D, Schwarz N, Stone AA. Would you be happier if you were richer? A focusing illusion. Science 2006; 312: 1908–1910.
19. Redelmeier DA, Rozin P, Kahneman D. Understanding patients' decisions. Cognitive and emotional perspectives. JAMA 1993; 270: 72–76.
20. Team RDC. A language in environment of statistical computing. Vienna, Austria 2006.
21. Bradley C. Importance of differentiating health status from quality of life. Lancet 2001; 357: 7–8.
22. Both H, Essink-Bot ML, Busschbach J, Nijsten T. Critical review of generic and dermatology-specific health-related quality of life instruments. J Invest Dermatol 2007; 127: 2726–2739.
23. Chren MM, Lasek RJ, Quinn LM, Covinsky KE. Convergent and discriminant validity of a generic and a disease-specific instrument to measure quality of life in patients with skin disease. J Invest Dermatol 1997; 108: 103–107.
24. Hongbo Y, Thomas CL, Harrison MA, Salek MS, Finlay AY. Translating the science of quality of life into practice: what do dermatology life quality index scores mean? J Invest Dermatol 2005; 125: 659–664.

Electronic appendix

No.	Title	Available at:
1	Amended Dermatology Life Quality Index	http://adv.medicaljournals.se/article/abstract/10.2340.00015555-0768/appendix

Appendix: III

Day Reconstruction Method Questionnaire: adapted for Dermatology

Packet 1

Research suggests that the best way to get the most accurate account of your feelings and experiences is to ask you to make a diary of a day and then to ask you questions about it rather than just asking you questions without a diary to refer to.

This questionnaire booklet feels large but you will see that this is just because we've tried to provide you with enough paper for you to make your diary on – you don't have to use all the paper if you don't need to – it's not thick because there are lots of questions.

Just so you know, you can make personal notes or use shorthand in the diary section as we don't need to see it. The only rule is that you shouldn't look at Packet 3 before having completed Packet 2.

Now, first we have some general questions about your life. Please answer these questions by ticking the answer that best describes your opinion.

1. Taking all things together, how satisfied are you with your life as a whole these days? Are you

__ very satisfied, __ satisfied, __ not very satisfied, __ not at all satisfied?

2. Next, let's turn to your life at home. Overall, how satisfied are you with your life at home? Are you

__ very satisfied, __ satisfied, __ not very satisfied, __ not at all satisfied?

3. And how about your job? Overall, how satisfied are you with your present job? Are you

__ very satisfied, __ satisfied, __ not very satisfied, __ not at all satisfied?

4. Now we would like to know how you feel and what mood you are in when you are at home. When you are at home, what percentage of the time are you

in a bad mood	____%
a little low or irritable	____%
in a mildly pleasant mood	____%
in a very good mood	____%
	Sum = 100%

5. We would also like to know how you feel and what mood you are in when you are at work. When you are at work, what percentage of the time are you

in a bad mood	____%
a little low or irritable	____%
in a mildly pleasant mood	____%
in a very good mood	____%
	Sum = 100%

Next, we would like to ask for some background information about you.

1. What year were you born? _____

2. What is your sex? __ Male __ Female

3. Are you :-

__ single (never married) __ married __living with a partner

__ divorced/separated __ widowed

4. What skin diagnosis do you have?

__ Eczema or Dermatitis
__ Psoriasis

5. How old were you when you started to have this skin problem?
_____

6. What creams and/or ointments do you use?

_____

_____

7. Do you have any other medical problems? If so please list them:

_____

_____

8. If you take any tablets/medications/inhalers/over-the-counter medicaments please list them:

_____

_____

Thank you!

You may now start on Packet 2

Packet 2

Yesterday

We would like to learn what you did and how you felt yesterday. Not all days are the same – some are better, some are worse and others are pretty typical. **Here we are only asking you about yesterday.**

Because many people find it difficult to remember what exactly they did and experienced, we will do this in three steps:

1. On the next page, we will ask you when you woke up and when you went to sleep yesterday.

2. We'd like you to reconstruct what your day was like, as if you were writing in your diary. Where were you? What did you do and experience? How did you feel? Answering the questions on the next page will help you to reconstruct your day. This diary packet is only for you, to help you remember and describe what happened during the first half of yesterday. It is yours to keep, so your notes are strictly personal and confidential. You do not need to turn it in. Nobody will read what you jot down about your day.

3. After you have finished reconstructing your day in your diary, we will ask you specific questions about this time (these questions are in Packet 3). In answering these questions, we'd like you to consult your diary page and the notes you made to remind you of what you did and how you felt.

To begin, please circle the day of the week that YESTERDAY was:

Monday Tuesday Wednesday Thursday Friday Saturday Sunday

Diary Pages

About what time did you wake up yesterday? _____

And when did you go to sleep? _____

- On the next three pages, please describe your day. Think of your day as a continuous series of scenes or episodes in a film.
- Give each episode a brief name that will help you remember it (for example, “commuting to work”, or “at lunch with B”, where B is a person or a group of people).
- Write down the approximate times at which each episode began and ended. The episodes people identify usually last between 15 minutes and 2 hours.
- Indications of the end of an episode might be going to a different location, ending one activity and starting another, or a change in the people you are interacting with.
- There is one page for each part of the day – Morning (from waking up until noon), Afternoon (from noon to 6:00 pm) and Evening (from 6:00 pm until you went to bed).
- There is room to list 10 episodes for each part of the day, although you may not need that many, depending on your day.
- It is not necessary to fill up all of the spaces – use the breakdown of your day that makes the most sense to you and best captures what you did and how you felt.
- Try to remember each episode in detail, and write a few words that will remind you of exactly what was going on. Also, try to remember how you felt, and what your mood was like during each episode.
- What you write only has to make sense to you, and to help you remember what happened when you are answering the questions in Packet 3.
- Remember, what you write in your diary will not be seen by anybody else.
- **Packet 2 is yours to keep if you wish – you don’t have to turn it in with the rest of your questionnaire.**

Morning

To include from waking up until just before lunch.

Episode name	**Time began**	**Time ended**	**Notes to self:** What happened? How you felt?
1M_____	_____	_____	_____
2M _____	_____	_____	_____
3M _____	_____	_____	_____
4M _____	_____	_____	_____
5M _____	_____	_____	_____
6M _____	_____	_____	_____
7M _____	_____	_____	_____
8M _____	_____	_____	_____
9M_____	_____	_____	_____
10M_____	_____	_____	_____

Afternoon

To include from lunch ‘til just before dinner.

Episode name	**Time began**	**Time ended**	**Notes to self:** What happened? How you felt?
1A Lunchtime	_____	_____	_____
2A _____	_____	_____	_____
3A _____	_____	_____	_____
4A _____	_____	_____	_____
5A _____	_____	_____	_____
6A _____	_____	_____	_____
7A _____	_____	_____	_____
8A _____	_____	_____	_____
9A_____	_____	_____	_____
10A_____	_____	_____	_____

Evening

To include from dinnertime ‘til just before went to bed.

Episode name	**Time began**	**Time ended**	**Notes to self:** What happened? How you felt?
1E Dinnertime	_____	_____	_____
2E _____	_____	_____	_____
3E _____	_____	_____	_____
4E _____	_____	_____	_____
5E _____	_____	_____	_____
6E _____	_____	_____	_____
7E _____	_____	_____	_____
8E _____	_____	_____	_____
9E_____	_____	_____	_____
10E_____	_____	_____	_____

Please look over your diary once more. Are there any other episodes that you’d like to revise or add more notes to? Is there an episode that you would want to break up into two parts? If so, please go back and make the necessary adjustments on your diary pages. If not, you may go on to Packet 3.

Thank You

You may now start on Packet 3.

Packet 3

How Did You Feel Yesterday?

Before we proceed, please look back at your diary pages...

How many episodes did you record for the Morning? _____

How many episodes did you record for the Afternoon? _____

How many episodes did you record for the Evening? _____

Now, we would like to learn in more detail about how you felt during those episodes. For each episode, there are several questions about what happened and how you felt. Please use the notes on your diary pages as often as you need to.

Please answer the questions for every episode you recorded, beginning with the first episode in the Morning. To make it easier to keep track, we will ask you to write down the number of the episode that is at the end of the line where you wrote about it in your diary. For example, the first episode of the Morning was number 1M, the third episode of the Afternoon was number 3A, the second episode of the Evening was number 2E, and so forth.

It is very important that we get to hear about all of the episodes you experienced yesterday, so please be sure to answer the questions for each episode you recorded. After you have answered the questions for all of your episodes - including the last episode of the day (just before you went to bed) - you can go on to Packet 4.

First Morning Episode

Please look at your Diary and select the **earliest** episode you noted in the Morning.

When did this first episode begin and end (e.g., 7:30am)? *Please try to remember the times as precisely as you can.*

This is episode number _____, which began at _____ and ended at _____.

What were you doing? (please check all that apply)

__ commuting
__ shopping
__ doing housework
__ eating
__ socializing
__ nap/resting
__ relaxing
__ intimate relations
__ other: please specify_____

__ working
__ preparing food
__ taking care of your children
__praying/worshipping/meditating
__ watching TV
__ computer/internet/email
__ on the phone
__ exercising

Where were you?

__ At home __ At work __ Somewhere else

Were you interacting with anyone (including on the phone, in a teleconference, etc)?

__ no one → *skip next question.*

If you were interacting with someone (please check all that apply)

__ spouse/significant other
__ friends
__ co-workers
__ others: listed_____

__ my children
__ parents/relatives
__ boss
__clients/customers/students/patients

How did you feel during this episode?

Please rate each feeling on the scale given. A rating of 0 means that you did not experience that feeling at all. A rating of 6 means that this feeling was a very important part of the experience. Please circle the number between 0 and 6 that best describes how you felt.

	Not at all						Very much
Impatient for it to end	0	1	2	3	4	5	6
Happy	0	1	2	3	4	5	6
Frustrated/annoyed	0	1	2	3	4	5	6
Depressed/sad	0	1	2	3	4	5	6
Competent/capable	0	1	2	3	4	5	6
Hassled/pushed around	0	1	2	3	4	5	6
Warm/friendly	0	1	2	3	4	5	6
Angry/hostile	0	1	2	3	4	5	6
Worried/anxious	0	1	2	3	4	5	6
Enjoying myself	0	1	2	3	4	5	6
Criticized/put down	0	1	2	3	4	5	6
Tired	0	1	2	3	4	5	6
Itchy	0	1	2	3	4	5	6
Uncomfortable	0	1	2	3	4	5	6

Next Episode

Now look at your Diary and select the episode that immediately followed the one you just rated.

This is episode number _____, which began at _____ and ended at _____.

What were you doing? (please check all that apply)

__ commuting
__ shopping
__ doing housework
__ eating
__ socializing
__ nap/resting
__ relaxing
__ intimate relations
__ other: please specify_____

__ working
__ preparing food
__ taking care of your children
__praying/worshipping/meditating
__ watching TV
__ computer/internet/email
__ on the phone
__ exercising

Where were you?

__ At home __ At work __ Somewhere else

Were you interacting with anyone (including on the phone, in a teleconference, etc)?

__ no one → *skip next question.*

If you were interacting with someone (please check all that apply)

__ spouse/significant other
__ friends
__ co-workers
__ others: listed_____

__ my children
__ parents/relatives
__ boss
__clients/customers/students/patients

How did you feel during this episode?

Please rate each feeling on the scale given. A rating of 0 means that you did not experience that feeling at all. A rating of 6 means that this feeling was a very important part of the experience. Please circle the number between 0 and 6 that best describes how you felt.

	Not at all						Very much
Impatient for it to end	0	1	2	3	4	5	6
Happy	0	1	2	3	4	5	6
Frustrated/annoyed	0	1	2	3	4	5	6
Depressed/sad	0	1	2	3	4	5	6
Competent/capable	0	1	2	3	4	5	6
Hassled/pushed around	0	1	2	3	4	5	6
Warm/friendly	0	1	2	3	4	5	6
Angry/hostile	0	1	2	3	4	5	6
Worried/anxious	0	1	2	3	4	5	6
Enjoying myself	0	1	2	3	4	5	6
Criticized/put down	0	1	2	3	4	5	6
Tired	0	1	2	3	4	5	6
Itchy	0	1	2	3	4	5	6
Uncomfortable	0	1	2	3	4	5	6

Next Episode

Now look at your Diary and select the episode that immediately followed the one you just rated.

This is episode number _____, which began at _____ and ended at _____.

What were you doing? (please check all that apply)

__ commuting
__ shopping
__ doing housework
__ eating
__ socializing
__ nap/resting
__ relaxing
__ intimate relations
__ other: please specify_____

__ working
__ preparing food
__ taking care of your children
__praying/worshipping/meditating
__ watching TV
__ computer/internet/email
__ on the phone
__ exercising

Where were you?

__ At home __ At work __ Somewhere else

Were you interacting with anyone (including on the phone, in a teleconference, etc)?

__ no one → *skip next question.*

If you were interacting with someone (please check all that apply)

__ spouse/significant other
__ friends
__ co-workers
__ others: listed_____

__ my children
__ parents/relatives
__ boss
__clients/customers/students/patients

How did you feel during this episode?

Please rate each feeling on the scale given. A rating of 0 means that you did not experience that feeling at all. A rating of 6 means that this feeling was a very important part of the experience. Please circle the number between 0 and 6 that best describes how you felt.

	Not at all						Very much
Impatient for it to end	0	1	2	3	4	5	6
Happy	0	1	2	3	4	5	6
Frustrated/annoyed	0	1	2	3	4	5	6
Depressed/sad	0	1	2	3	4	5	6
Competent/capable	0	1	2	3	4	5	6
Hassled/pushed around	0	1	2	3	4	5	6
Warm/friendly	0	1	2	3	4	5	6
Angry/hostile	0	1	2	3	4	5	6
Worried/anxious	0	1	2	3	4	5	6
Enjoying myself	0	1	2	3	4	5	6
Criticized/put down	0	1	2	3	4	5	6
Tired	0	1	2	3	4	5	6
Itchy	0	1	2	3	4	5	6
Uncomfortable	0	1	2	3	4	5	6

Next Episode

Now look at your Diary and select the episode that immediately followed the one you just rated.

This is episode number _____, which began at _____ and ended at _____.

What were you doing? (please check all that apply)

__ commuting
__ shopping
__ doing housework
__ eating
__ socializing
__ nap/resting
__ relaxing
__ intimate relations
__ other: please specify_____

__ working
__ preparing food
__ taking care of your children
__praying/worshipping/meditating
__ watching TV
__ computer/internet/email
__ on the phone
__ exercising

Where were you?

__ At home __ At work __ Somewhere else

Were you interacting with anyone (including on the phone, in a teleconference, etc)?

__ no one → *skip next question.*

If you were interacting with someone (please check all that apply)

__ spouse/significant other
__ friends
__ co-workers
__ others: listed_____

__ my children
__ parents/relatives
__ boss
__clients/customers/students/patients

How did you feel during this episode?

Please rate each feeling on the scale given. A rating of 0 means that you did not experience that feeling at all. A rating of 6 means that this feeling was a very important part of the experience. Please circle the number between 0 and 6 that best describes how you felt.

	Not at all						Very much
Impatient for it to end	0	1	2	3	4	5	6
Happy	0	1	2	3	4	5	6
Frustrated/annoyed	0	1	2	3	4	5	6
Depressed/sad	0	1	2	3	4	5	6
Competent/capable	0	1	2	3	4	5	6
Hassled/pushed around	0	1	2	3	4	5	6
Warm/friendly	0	1	2	3	4	5	6
Angry/hostile	0	1	2	3	4	5	6
Worried/anxious	0	1	2	3	4	5	6
Enjoying myself	0	1	2	3	4	5	6
Criticized/put down	0	1	2	3	4	5	6
Tired	0	1	2	3	4	5	6
Itchy	0	1	2	3	4	5	6
Uncomfortable	0	1	2	3	4	5	6

Next Episode

Now look at your Diary and select the episode that immediately followed the one you just rated.

This is episode number _____, which began at _____ and ended at _____.

What were you doing? (please check all that apply)

__ commuting
__ shopping
__ doing housework
__ eating
__ socializing
__ nap/resting
__ relaxing
__ intimate relations
__ other: please specify_____

__ working
__ preparing food
__ taking care of your children
__praying/worshipping/meditating
__ watching TV
__ computer/internet/email
__ on the phone
__ exercising

Where were you?

__ At home __ At work __ Somewhere else

Were you interacting with anyone (including on the phone, in a teleconference, etc)?

__ no one → *skip next question.*

If you were interacting with someone (please check all that apply)

__ spouse/significant other
__ friends
__ co-workers
__ others: listed_____

__ my children
__ parents/relatives
__ boss
__clients/customers/students/patients

How did you feel during this episode?

Please rate each feeling on the scale given. A rating of 0 means that you did not experience that feeling at all. A rating of 6 means that this feeling was a very important part of the experience. Please circle the number between 0 and 6 that best describes how you felt.

	Not at all						Very much
Impatient for it to end	0	1	2	3	4	5	6
Happy	0	1	2	3	4	5	6
Frustrated/annoyed	0	1	2	3	4	5	6
Depressed/sad	0	1	2	3	4	5	6
Competent/capable	0	1	2	3	4	5	6
Hassled/pushed around	0	1	2	3	4	5	6
Warm/friendly	0	1	2	3	4	5	6
Angry/hostile	0	1	2	3	4	5	6
Worried/anxious	0	1	2	3	4	5	6
Enjoying myself	0	1	2	3	4	5	6
Criticized/put down	0	1	2	3	4	5	6
Tired	0	1	2	3	4	5	6
Itchy	0	1	2	3	4	5	6
Uncomfortable	0	1	2	3	4	5	6

Next Episode

Now look at your Diary and select the episode that immediately followed the one you just rated.

This is episode number _____, which began at _____ and ended at _____.

What were you doing? (please check all that apply)

__ commuting
__ shopping
__ doing housework
__ eating
__ socializing
__ nap/resting
__ relaxing
__ intimate relations
__ other: please specify_____

__ working
__ preparing food
__ taking care of your children
__praying/worshipping/meditating
__ watching TV
__ computer/internet/email
__ on the phone
__ exercising

Where were you?

__ At home __ At work __ Somewhere else

Were you interacting with anyone (including on the phone, in a teleconference, etc)?

__ no one → *skip next question.*

If you were interacting with someone (please check all that apply)

__ spouse/significant other
__ friends
__ co-workers
__ others: listed_____

__ my children
__ parents/relatives
__ boss
__clients/customers/students/patients

How did you feel during this episode?

Please rate each feeling on the scale given. A rating of 0 means that you did not experience that feeling at all. A rating of 6 means that this feeling was a very important part of the experience. Please circle the number between 0 and 6 that best describes how you felt.

	Not at all						Very much
Impatient for it to end	0	1	2	3	4	5	6
Happy	0	1	2	3	4	5	6
Frustrated/annoyed	0	1	2	3	4	5	6
Depressed/sad	0	1	2	3	4	5	6
Competent/capable	0	1	2	3	4	5	6
Hassled/pushed around	0	1	2	3	4	5	6
Warm/friendly	0	1	2	3	4	5	6
Angry/hostile	0	1	2	3	4	5	6
Worried/anxious	0	1	2	3	4	5	6
Enjoying myself	0	1	2	3	4	5	6
Criticized/put down	0	1	2	3	4	5	6
Tired	0	1	2	3	4	5	6
Itchy	0	1	2	3	4	5	6
Uncomfortable	0	1	2	3	4	5	6

Next Episode

Now look at your Diary and select the episode that immediately followed the one you just rated.

This is episode number _____, which began at _____ and ended at _____.

What were you doing? (please check all that apply)

__ commuting
__ shopping
__ doing housework
__ eating
__ socializing
__ nap/resting
__ relaxing
__ intimate relations
__ other: please specify_____

__ working
__ preparing food
__ taking care of your children
__praying/worshipping/meditating
__ watching TV
__ computer/internet/email
__ on the phone
__ exercising

Where were you?

__ At home __ At work __ Somewhere else

Were you interacting with anyone (including on the phone, in a teleconference, etc)?

__ no one → *skip next question.*

If you were interacting with someone (please check all that apply)

__ spouse/significant other
__ friends
__ co-workers
__ others: listed_____

__ my children
__ parents/relatives
__ boss
__clients/customers/students/patients

How did you feel during this episode?

Please rate each feeling on the scale given. A rating of 0 means that you did not experience that feeling at all. A rating of 6 means that this feeling was a very important part of the experience. Please circle the number between 0 and 6 that best describes how you felt.

	Not at all						Very much
Impatient for it to end	0	1	2	3	4	5	6
Happy	0	1	2	3	4	5	6
Frustrated/annoyed	0	1	2	3	4	5	6
Depressed/sad	0	1	2	3	4	5	6
Competent/capable	0	1	2	3	4	5	6
Hassled/pushed around	0	1	2	3	4	5	6
Warm/friendly	0	1	2	3	4	5	6
Angry/hostile	0	1	2	3	4	5	6
Worried/anxious	0	1	2	3	4	5	6
Enjoying myself	0	1	2	3	4	5	6
Criticized/put down	0	1	2	3	4	5	6
Tired	0	1	2	3	4	5	6
Itchy	0	1	2	3	4	5	6
Uncomfortable	0	1	2	3	4	5	6

Next Episode

Now look at your Diary and select the episode that immediately followed the one you just rated.

This is episode number _____, which began at _____ and ended at _____.

What were you doing? (please check all that apply)

__ commuting
__ shopping
__ doing housework
__ eating
__ socializing
__ nap/resting
__ relaxing
__ intimate relations
__ other: please specify_____

__ working
__ preparing food
__ taking care of your children
__praying/worshipping/meditating
__ watching TV
__ computer/internet/email
__ on the phone
__ exercising

Where were you?

__ At home __ At work __ Somewhere else

Were you interacting with anyone (including on the phone, in a teleconference, etc)?

__ no one → *skip next question.*

If you were interacting with someone (please check all that apply)

__ spouse/significant other
__ friends
__ co-workers
__ others: listed_____

__ my children
__ parents/relatives
__ boss
__clients/customers/students/patients

How did you feel during this episode?

Please rate each feeling on the scale given. A rating of 0 means that you did not experience that feeling at all. A rating of 6 means that this feeling was a very important part of the experience. Please circle the number between 0 and 6 that best describes how you felt.

	Not at all						Very much
Impatient for it to end	0	1	2	3	4	5	6
Happy	0	1	2	3	4	5	6
Frustrated/annoyed	0	1	2	3	4	5	6
Depressed/sad	0	1	2	3	4	5	6
Competent/capable	0	1	2	3	4	5	6
Hassled/pushed around	0	1	2	3	4	5	6
Warm/friendly	0	1	2	3	4	5	6
Angry/hostile	0	1	2	3	4	5	6
Worried/anxious	0	1	2	3	4	5	6
Enjoying myself	0	1	2	3	4	5	6
Criticized/put down	0	1	2	3	4	5	6
Tired	0	1	2	3	4	5	6
Itchy	0	1	2	3	4	5	6
Uncomfortable	0	1	2	3	4	5	6

Next Episode

Now look at your Diary and select the episode that immediately followed the one you just rated.

This is episode number _____, which began at _____ and ended at _____.

What were you doing? (please check all that apply)

__ commuting
__ shopping
__ doing housework
__ eating
__ socializing
__ nap/resting
__ relaxing
__ intimate relations
__ other: please specify_____

__ working
__ preparing food
__ taking care of your children
__praying/worshipping/meditating
__ watching TV
__ computer/internet/email
__ on the phone
__ exercising

Where were you?

__ At home __ At work __ Somewhere else

Were you interacting with anyone (including on the phone, in a teleconference, etc)?

__ no one → *skip next question.*

If you were interacting with someone (please check all that apply)

__ spouse/significant other
__ friends
__ co-workers
__ others: listed_____

__ my children
__ parents/relatives
__ boss
__clients/customers/students/patients

How did you feel during this episode?

Please rate each feeling on the scale given. A rating of 0 means that you did not experience that feeling at all. A rating of 6 means that this feeling was a very important part of the experience. Please circle the number between 0 and 6 that best describes how you felt.

	Not at all						Very much
Impatient for it to end	0	1	2	3	4	5	6
Happy	0	1	2	3	4	5	6
Frustrated/annoyed	0	1	2	3	4	5	6
Depressed/sad	0	1	2	3	4	5	6
Competent/capable	0	1	2	3	4	5	6
Hassled/pushed around	0	1	2	3	4	5	6
Warm/friendly	0	1	2	3	4	5	6
Angry/hostile	0	1	2	3	4	5	6
Worried/anxious	0	1	2	3	4	5	6
Enjoying myself	0	1	2	3	4	5	6
Criticized/put down	0	1	2	3	4	5	6
Tired	0	1	2	3	4	5	6
Itchy	0	1	2	3	4	5	6
Uncomfortable	0	1	2	3	4	5	6

Next Episode

Now look at your Diary and select the episode that immediately followed the one you just rated.

This is episode number _____, which began at _____ and ended at _____.

What were you doing? (please check all that apply)

__ commuting
__ shopping
__ doing housework
__ eating
__ socializing
__ nap/resting
__ relaxing
__ intimate relations
__ other: please specify_____

__ working
__ preparing food
__ taking care of your children
__praying/worshipping/meditating
__ watching TV
__ computer/internet/email
__ on the phone
__ exercising

Where were you?

__ At home __ At work __ Somewhere else

Were you interacting with anyone (including on the phone, in a teleconference, etc)?

__ no one → *skip next question.*

If you were interacting with someone (please check all that apply)

__ spouse/significant other
__ friends
__ co-workers
__ others: listed_____

__ my children
__ parents/relatives
__ boss
__clients/customers/students/patients

How did you feel during this episode?

Please rate each feeling on the scale given. A rating of 0 means that you did not experience that feeling at all. A rating of 6 means that this feeling was a very important part of the experience. Please circle the number between 0 and 6 that best describes how you felt.

	Not at all						Very much
Impatient for it to end	0	1	2	3	4	5	6
Happy	0	1	2	3	4	5	6
Frustrated/annoyed	0	1	2	3	4	5	6
Depressed/sad	0	1	2	3	4	5	6
Competent/capable	0	1	2	3	4	5	6
Hassled/pushed around	0	1	2	3	4	5	6
Warm/friendly	0	1	2	3	4	5	6
Angry/hostile	0	1	2	3	4	5	6
Worried/anxious	0	1	2	3	4	5	6
Enjoying myself	0	1	2	3	4	5	6
Criticized/put down	0	1	2	3	4	5	6
Tired	0	1	2	3	4	5	6
Itchy	0	1	2	3	4	5	6
Uncomfortable	0	1	2	3	4	5	6

Next Episode

Now look at your Diary and select the episode that immediately followed the one you just rated.

This is episode number _____, which began at _____ and ended at _____.

What were you doing? (please check all that apply)

__ commuting
__ shopping
__ doing housework
__ eating
__ socializing
__ nap/resting
__ relaxing
__ intimate relations
__ other: please specify_____

__ working
__ preparing food
__ taking care of your children
__praying/worshipping/meditating
__ watching TV
__ computer/internet/email
__ on the phone
__ exercising

Where were you?

__ At home __ At work __ Somewhere else

Were you interacting with anyone (including on the phone, in a teleconference, etc)?

__ no one → *skip next question.*

If you were interacting with someone (please check all that apply)

__ spouse/significant other
__ friends
__ co-workers
__ others: listed_____

__ my children
__ parents/relatives
__ boss
__clients/customers/students/patients

How did you feel during this episode?

Please rate each feeling on the scale given. A rating of 0 means that you did not experience that feeling at all. A rating of 6 means that this feeling was a very important part of the experience. Please circle the number between 0 and 6 that best describes how you felt.

	Not at all						Very much
Impatient for it to end	0	1	2	3	4	5	6
Happy	0	1	2	3	4	5	6
Frustrated/annoyed	0	1	2	3	4	5	6
Depressed/sad	0	1	2	3	4	5	6
Competent/capable	0	1	2	3	4	5	6
Hassled/pushed around	0	1	2	3	4	5	6
Warm/friendly	0	1	2	3	4	5	6
Angry/hostile	0	1	2	3	4	5	6
Worried/anxious	0	1	2	3	4	5	6
Enjoying myself	0	1	2	3	4	5	6
Criticized/put down	0	1	2	3	4	5	6
Tired	0	1	2	3	4	5	6
Itchy	0	1	2	3	4	5	6
Uncomfortable	0	1	2	3	4	5	6

Next Episode

Now look at your Diary and select the episode that immediately followed the one you just rated.

This is episode number _____, which began at _____ and ended at _____.

What were you doing? (please check all that apply)

__ commuting
__ shopping
__ doing housework
__ eating
__ socializing
__ nap/resting
__ relaxing
__ intimate relations
__ other: please specify_____

__ working
__ preparing food
__ taking care of your children
__praying/worshipping/meditating
__ watching TV
__ computer/internet/email
__ on the phone
__ exercising

Where were you?

__ At home __ At work __ Somewhere else

Were you interacting with anyone (including on the phone, in a teleconference, etc)?

__ no one → *skip next question.*

If you were interacting with someone (please check all that apply)

__ spouse/significant other
__ friends
__ co-workers
__ others: listed_____

__ my children
__ parents/relatives
__ boss
__clients/customers/students/patients

How did you feel during this episode?

Please rate each feeling on the scale given. A rating of 0 means that you did not experience that feeling at all. A rating of 6 means that this feeling was a very important part of the experience. Please circle the number between 0 and 6 that best describes how you felt.

	Not at all						Very much
Impatient for it to end	0	1	2	3	4	5	6
Happy	0	1	2	3	4	5	6
Frustrated/annoyed	0	1	2	3	4	5	6
Depressed/sad	0	1	2	3	4	5	6
Competent/capable	0	1	2	3	4	5	6
Hassled/pushed around	0	1	2	3	4	5	6
Warm/friendly	0	1	2	3	4	5	6
Angry/hostile	0	1	2	3	4	5	6
Worried/anxious	0	1	2	3	4	5	6
Enjoying myself	0	1	2	3	4	5	6
Criticized/put down	0	1	2	3	4	5	6
Tired	0	1	2	3	4	5	6
Itchy	0	1	2	3	4	5	6
Uncomfortable	0	1	2	3	4	5	6

If you have more episodes to rate, please ask the attendant for additional forms.

Have you rated all of your episodes, including the last episode of the day, just before you went to bed? **If so, you may go on to Packet 4.**

Packet 4

A Few More Questions about Yesterday

Now that you have told us about your day in detail, we have a few more general questions.

First, we would like to know **overall** how you felt and what your mood was like yesterday. **Thinking only about yesterday,** what percentage of the time were you

in a bad mood	____%
a little low or irritable	____%
in a mildly pleasant mood	____%
in a very good mood	____%

Sum = 100%

Now we would like to know **overall how you itchy** you were yesterday. **Thinking only about yesterday,** what percentage of the time were you

Incredibly itchy	____%
Quite itchy	____%
Hardly noticeably itchy	____%
Not at all itchy	____%

Sum = 100%

Now we'd like to know **how typical yesterday was** for that day of the week (i.e., for a Monday, for a Tuesday, or so on). Compared to what that day of the week usually is like, yesterday was (please circle one)

Much worse	**Somewhat worse**	**Pretty typical**	**Somewhat better**	**Much better**
1	2	3	4	5

Now we would like to know overall how you felt and what your mood was like **at work** yesterday. Thinking only about the time you spent **at work yesterday**, what percentage of the time were you

in a bad mood	____%
a little low or irritable	____%
in a mildly pleasant mood	____%
in a very good mood	____%

Sum = 100%

Thinking **only about the time you spent at work yesterday**, what percentage of the time were you

Incredibly itchy	____%
Quite itchy	____%
Hardly noticeably itchy	____%
Not at all itchy	____%

Sum = 100%

Now we'd like to know how yesterday compares to a typical day at work. Compared to a typical day at work, my time spent at work yesterday was (please circle one)

Much worse	**Somewhat worse**	**Pretty typical**	**Somewhat better**	**Much better**
1	2	3	4	5

How Do Others See You?

In this section, we would like to learn how others see you.

What would the people who know you say about you? For each of the following, please indicate where they would place you on the scale below.

On this scale, a -3 means that this is much less a characteristic of you than of other people. A 0 means that others would see you as about average. A +3 means it is much more characteristic of you than of others. Please circle the number between -3 and +3 that best describes what others would say about you.

	Much less than others			**About average**			**Much more than others**
enthusiastic	-3	-2	-1	0	+1	+2	+3
optimistic	-3	-2	-1	0	+1	+2	+3
laughs easily	-3	-2	-1	0	+1	+2	+3
very healthy	-3	-2	-1	0	+1	+2	+3
always sees the bright side	-3	-2	-1	0	+1	+2	+3
comfortable everywhere	-3	-2	-1	0	+1	+2	+3
enjoys good food	-3	-2	-1	0	+1	+2	+3
enjoys being in company	-3	-2	-1	0	+1	+2	+3
pessimistic	-3	-2	-1	0	+1	+2	+3
often worries for nothing	-3	-2	-1	0	+1	+2	+3
a bit depressed	-3	-2	-1	0	+1	+2	+3
often angry	-3	-2	-1	0	+1	+2	+3
tense and uncomfortable	-3	-2	-1	0	+1	+2	+3
seems quite ill	-3	-2	-1	0	+1	+2	+3

1. During the past month, how would you rate your overall sleep quality?

__very good __ fairly good __ fairly bad __very bad

2. During the past month, on average how many hours of actual sleep did you get at night?

Average hours of sleep per night _____

3. Last night, how many hours of actual sleep did you get?

Hours of sleep last night _____

4. During the past month, how often have you had trouble staying awake while driving, eating meals, or engaging in social activity?

__ not at all during the past month
__ less than once a week/
__ once or twice a week
__ three or more times a week

5. During the past month, how much of a problem has it been for you to keep up enough enthusiasm to get things done?

__ no problem at all
__ only a very slight problem
__ somewhat of a problem
__ a very big problem

6. How satisfied are you with your health these days? Are you

__ very satisfied __ satisfied __ not very satisfied __not at all satisfied

You have now completed the survey. Please review each packet to be sure you have answered all the questions.

After you have checked your answers, put all of the numbered packets (except the diary if you wish to keep it) in the large envelope.

Thank you very much for participating