Measuring subjective health perceptions:

Insights from psychological theory and social research

Steven Charles Hope

Thesis submitted for the degree of

Doctor of Philosophy

City University

Department of Sociology

Social Research Methodology Centre

April 2007

Volume 1

Table of contents

Volume 2

List of Tables

List of Figures

Acknowledgements

I would like to thank my principal supervisor, Ingrid Schoon, for her guidance and unstinting support during the course of the PhD project. I would also like to thank my co-supervisor, Dick Wiggins, for his insightful observations and suggestions at critical stages of the work, along with the other members of the PhD steering group, Bob Erens and Amanda Sacker, for their invaluable advice on research design and analysis. I must also mention a number of people and organisations who helped me plan and organise the research: Dr Simon Barton and the staff and volunteers of the HIV/GUM Directorate, Chelsea & Westminster Hospital; Lighthouse, West London; London East Aids Network; Debbie Collins, of the National Centre for Social Research; and the City University Sociology Department. Thanks also to the Economic and Social Research Council, who provided the funding for this work to be undertaken. For their encouragement throughout, I would like to thank my family, friends and colleagues. Finally, I must express my gratitude to all those participants who took part in the qualitative and quantitative studies conducted for this thesis.

Abstract

The central premise of this thesis is that the generation of a response to a questionnaire item on subjective health perceptions is highly complex, dynamic, and contextualised within an individual's unique frame of reference. Consequently, an integrative framework for the investigation of response processes and subjective health was adopted, informed by psychological theory, and techniques drawn from qualitative, cognitive, survey and psychometric traditions.

The focus of research involved the detailed examination of a single health status measure, the SF-12v2 (a brief, multi-dimensional health status measure), completed by comparative samples of university and HIV participants (representing nominally healthy and health problem groups). The use of the SF-12v2 and the two samples allowed comparisons to be made between methods and samples throughout the research.

Qualitative (cognitive) techniques provided rich and useful information on response processes used when completing the health status measure, with considerable variation in item interpretation and clear contextual influences on response strategies employed. Nevertheless, differences in response processes between samples generally related to substantive health problems, which could be summarised and investigated quantitatively.

The quantitative research indicated that the SF-12v2 possessed generally good psychometric properties, although not identical for the two samples, and that the relationship between contextual factors, response process and response could be meaningfully examined. Finally, path analyses demonstrated that a unified model of response could be developed and tested, linking contextual factors, response strategy and SF-12v2 scores. The results showed that personality and objective health factors influenced physical and mental health scale scores, although direct and mediated pathways differed by outcome and sample.

In conclusion, this research framework has offered important insights into the response processes involved in the completion of a health status measure. The use of multiple qualitative and quantitative techniques has provided a more detailed understanding of response from different methodological perspectives. Nonetheless, further work is required to more fully develop this contextual model of response.

Glossary

AIDS	Acquired Immune Deficiency Syndrome
ART	Anti-retroviral Therapy
BP	Bodily Pain
CASM	Cognitive Aspects of Survey Methodology
CD4	CD4 T-cell count
CFA	Confirmatory Factor Analysis
DIF	Differential Item Functioning
EFA	Exploratory Factor Analysis
GH	General Health
GP	General Practitioner
GUM	Genitourinary Medicine
HAART	Highly Active Antiretroviral Therapy
HIE	Health Insurance Experiment
HIV	Human Immunodeficiency Virus
HRQOL	Health-related Quality of Life
IRT	Item Response Theory
MCS	Mental Component Scale
MH	Mental Health
MIMIC	Multiple-Indicator-Multiple-Cause
MOS	Medical Outcome Study
NBS	Norm-Based Scoring
NHS	National Health Service
NICE	National Institute for Health and Clinical Excellence
PCA	Principal Component Analysis

PCS	Physical Component Scale
PF	Physical Functioning
PIQ	Process Interrogation Questionnaire
QOL	Quality of Life
RE	Role Emotional
RP	Role Physical
RV	Relative Validity
SEM	Structural Equation Modelling
SF	Social Functioning
SF-12 SF-12v2	Short Form 12, versions 1 and 2
SF-36 SF-36v2	Short Form 36, versions 1 and 2
STI	Sexually Transmitted Infection
VL	Viral Load
VT	Vitality
WHO	World Health Organization
ICD-10	International Classification of Diseases (International Statistical Classification of Diseases and Related Health Problems) 10th Revision

Introduction

Background

This thesis investigates processes involved in responding to a health status measure among people with diverse health experiences, and, in seeking to develop a detailed understanding of these processes, adopts an integrative approach, linking psychological and social research theory and research. The research is comparative, including a health problem group (people with HIV) and a nominally healthy group (people recruited from a university), and focuses on a single health status measure, the Short Form 12-item questionnaire, version two (SF-12v2), providing a template for the extension of health status questionnaire validation beyond current standard techniques.

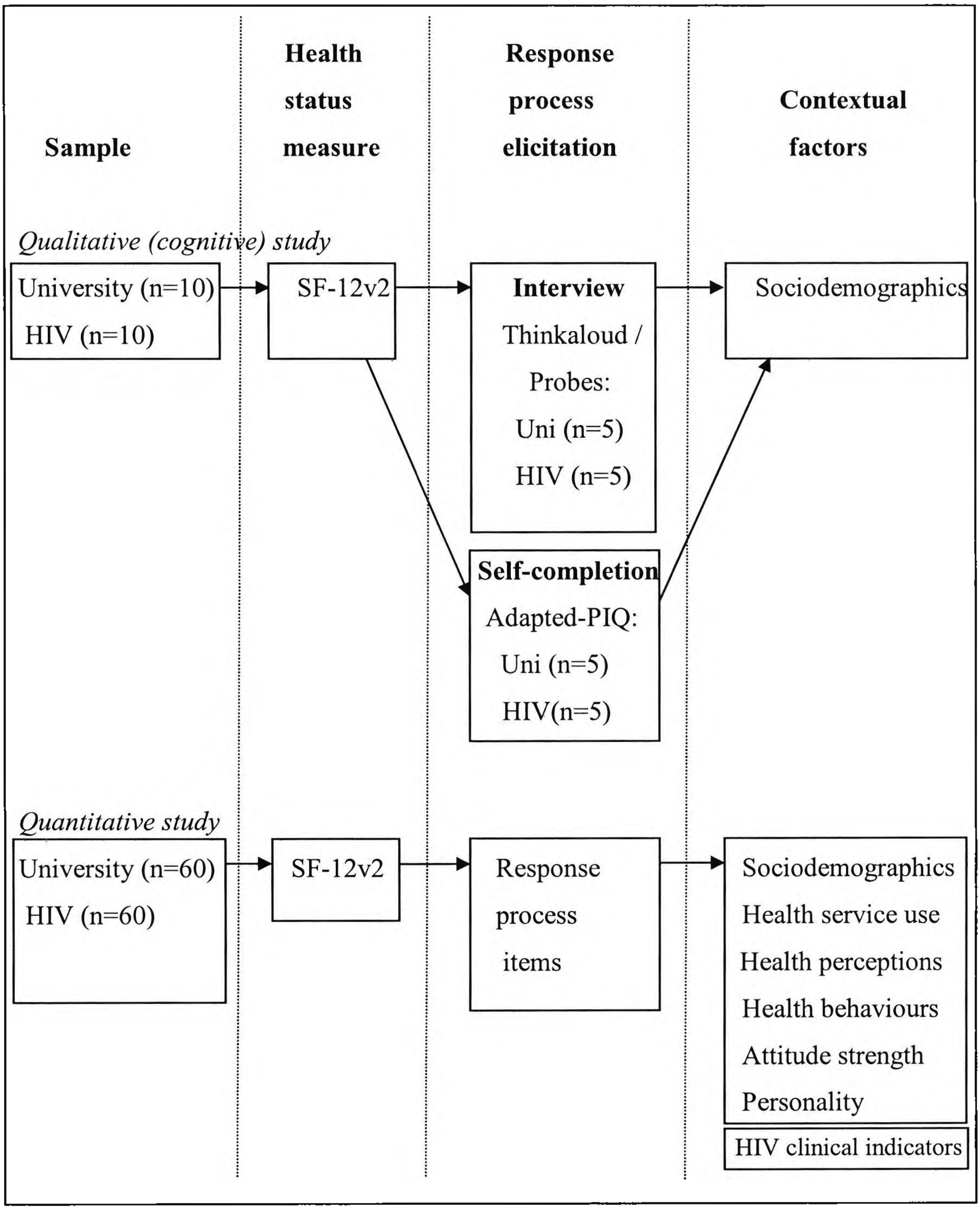
While the psychometric testing of health status measures is widespread, demonstrating good reliability and validity, less attention has been paid to the investigation of underlying response processes (McColl et al, 2003). In relation to the SF-12v2, this is the first project to validate it for people with HIV, to statistically model differential item functioning, and to investigate processes underlying responses to the measure. The study will therefore have relevance for psychology, psychometric testing, and HIV research, bringing new impetus for research.

The thesis is structured in three distinct parts, and is theoretically informed by both cognitive and psychometric models for the evaluation of health status measurement. Part One includes five sections: first, an outline of the research questions; second, an overview of the history and development of the structured measurement of health perceptions; third, a description of the origins and development of the health status measure used in this research, the SF-12v2; fourth, health perceptions and chronic illness, focusing on HIV/AIDS and health status; and fifth, theory and research on survey response processes, particularly in relation to health.

Parts Two and Three of the thesis describe qualitative (cognitive) and quantitative research studies undertaken to address specific questions about health status measurement and response processes in HIV positive and university comparative samples. The research uniquely combines methodologies in order to investigate health status from multiple perspectives: individual contextual factors, response processes, and final response. A range of methods and techniques were employed: in-depth qualitative research into subjective aspects of the response process; classical and modern psychometrics techniques for analyses of item and scale functioning; quantitative analysis of contextual influences on response process and response. Finally, Part Four provides an overall conclusion to the thesis, integrating the qualitative and quantitative findings in relation to the literature, and discussing future research possibilities.

Methodological overview

Figure 1.1: Overview of study design



Small and large-scale studies, using very different approaches, were employed to investigate the complexities of the response process. Figure 1.1 summarises the research design, the upper half of the figure showing the qualitative (cognitive) study and the lower half the survey study. Although different, both studies share several characteristics: comparative HIV and university samples, the SF-12v2 health status measure, and an investigation of response processes and individual contextual factors.

There were three main reasons for including people with HIV as a research sample. First, while a number of research projects have investigated HIV and health status, including the SF-12v1, no published paper has yet included the psychometrically improved SF-12v2. Second, and more fundamental to the current thesis, since health status is so widely accepted in HIV research, it is crucial to find out more about the meaning of health status measures for this health group. Although research has revealed interesting relationships between aspects of HIV and health status, none has systematically investigated the processes by which people with HIV/AIDS go about answering health status measures. Third, any study of health status response processes for chronic illnesses has to consider both the nature of the disease and the sociodemographic composition of the population from which the sample is drawn. There are some important characteristics of HIV that make it particularly suitable for a study of response process and health status. Unlike many other chronic diseases, HIV is generally more prevalent in a younger population (most people diagnosed with HIV in the UK are aged between 20-44 years: UK Collaborative Group for HIV and STI Surveillance, 2005), which reduces some of the issues resulting from both multiple coexisting chronic diseases and the effects of aging. In addition, clear ordinal measures of disease progression are available for HIV, which are seldom found in other chronic diseases (the virological markers CD4 and viral load) (Taylor, 2000).

The core of the design of this research, both qualitative and quantitative, involves a comparison between people with HIV and a contrasting (nominally healthy) population, in order to identify findings particular to the health problem group. It was decided to recruit samples from City University students and staff. Reasons for this were ease of recruitment, age profile and likely healthiness of the sample. To elaborate on these reasons, since the research was being carried out at the University, it was possible to recruit from both student and staff populations. Second, students and staff would be expected to have a younger age profile than the general population, a finding replicated among people with HIV. Third, students and staff are studying or working and therefore would be expected to be among the healthier groups in the population (Mathers and Schofield, 1998).

The methodological approaches adopted in the two studies differ markedly, one being a small qualitative (cognitive) study and the other a quantitative survey. The adoption of these diverse approaches relates to the comparative strengths and weaknesses of each technique and the very different data they can provide. In addition, it provides an opportunity to compare findings and methods, in a manner akin to triangulation or convergent validity (Campbell and Fiske, 1959).

Qualitative (cognitive) techniques aim to elicit conscious cognitive processes, investigating "meaning" assigned to the question and response by respondents. Although qualitative and cognitive are not synonymous (Campanelli, 1997), there are shared characteristics. Most do not recruit large, representative samples, instead using small purposive samples, representing different subpopulations of interest. Cognitive studies have been used to investigate question interpretation and discover problems of ambiguity and illogicality (Converse and Presser, 1986; Jenkinson, 1995). Similarly, cognitive techniques have also provided important insights into the meaning of answers obtained (McColl, 2003). A range of cognitive techniques are used in the present study in order to develop a deeper understanding of SF-12v2 response processes in the university and HIV samples, including think-aloud, retrospective probes, and an open-ended self-completion questionnaire.

However, small-scale cognitive research has been criticised. The artificiality of the cognitive interview situation and the tasks respondents are requested to perform have been questioned (Campanelli, 1997; Drennan, 2003; Willis, 1994). Cognitive researchers have been criticised for not being explicit about assumptions underlying the methods employed or the analysis undertaken. When a problem with a survey question is highlighted through cognitive methods, it has been argued that there is no clear indication of the "frequency, magnitude and impact of the phenomenon discovered within the cognitive model" (Bjorner et al, 2003). It has also been claimed that in a large sample, deviations in the response process do not have a significant impact of the results obtained, being "averaged out" in group analyses or being identified as error variance (Norman, 2003). Psychometric analysis has shown that most respondents are consistent in their responses to health surveys, casting doubt on the need to use smaller-scale cognitive research except in a few cases where discrepancies are highlighted (Bjorner et al, 2003). Finally, the ability of respondents to access their internal cognitive processes has been questioned (Collins, 2003). There is a range of evidence that people may not have insights into many of their cognitive processes (Bem, 1972; Nisbett and Wilson, 1977).

Some of these criticisms have themselves been challenged. First, a number of problems with survey items have been identified using cognitive methods. Second, the claim that response process problems may be overcome using large samples has been criticised both because validation depends on shared meaning, and, if meaning varies, large samples would not overcome problems comparing groups or change over time in individuals (Mallison, 2002). Third, it has been argued that those higher-level cognitive processes which are important to the respondent and central to their concept of "self" are accessible (Markus, 1983).

Nevertheless, quantitative statistical techniques using larger samples can be used to analyse aspects of response that are impossible to investigate using a qualitative approach, and therefore survey data were incorporated in the research design.

The survey data were used to assess the psychometric properties of the SF-12v2 scales in the HIV and university samples. This included classical psychometric concepts, such as reliability and validity (Rust and Golombok, 1999). In addition, further analyses were undertaken using modern psychometric methods that allow a measurement model to be specified and tested, in order to investigate the magnitude of any deviations from a hypothesised scale structure. For questionnaires to be valid, all respondents should understand and respond to items in the same way. For example, all those who are at the same level of an underlying construct should respond identically to the items measuring that construct. Variation may be problematic if discovered between individuals, but if there are systematic relative differences in the way questions are answered between different population groups, referred to as differential item functioning (DIF), this could undermine measurement (Fleishman and Lawrence, 2003). In this thesis, DIF analyses were undertaken using MIMIC (Multiple-Indicator-Multiple-Cause) modelling (Muthén, 2002), an extension to confirmatory factor analysis (CFA), which allows a hypothetical model of the relationship between items and underlying latent factors to be tested. As a first stage, a CFA was used to identify whether the items satisfactorily related to the two latent physical and mental health factors the SF-12v2 is hypothesised to measure. This model was then tested to see whether an exogenous variable (in this case, HIV sample membership), influenced constituent scale items without mediation by the hypothesised latent variables. That is to say, whether there were any direct effects of HIV sample membership on SF-12v2 item responses after controlling for the influence of the underlying latent factors. If no DIF is detected, any scoring differences observed between the sample groups are considered to reflect true differences in perceived health, rather than an artefact of the measure used (Fleishman and Lawrence, 2003).

There are three notable problems facing those who seek to use statistical techniques to identify DIF. First, a large sample is necessary in order to have confidence in the statistical modelling; second, a generally constant bias affecting all items in a scale equally, such as might be due to personality characteristics, cannot be detected; and third, the identification of DIF does not of itself provide information about why this has occurred. A greater depth of understanding of DIF, and the processes giving rise to it, can only be achieved using other techniques, such as cognitive interviewing.

Statistical analyses were also used to investigate factors that might influence SF-12v2 response. This included a comparison of the distributions of contextual and response process variables in the two samples, and the relationships between these factors and SF-12v2 response. Finally, a conceptual model was developed, unifying context, response process and response. This *contextual model* specified a set of direct and mediated pathways linking life events, dispositional characteristics, and response strategy to SF-12v2 physical and mental health summary scores. Path modelling techniques provided an opportunity to test the hypothesised model in relation to the observed data (Byrne, 2001; Wright, 1934). However, although these final analyses combined important influences on SF-12v2 scores in a theoretically meaningful sequence (Brief et al, 1993), a quantitative approach cannot capture the subtlety of individual influences on response.

No single approach is likely to provide a complete understanding of the nature of response, or the relationship between psychometric properties of a measure and the response process. The division between qualitative (cognitive) and quantitative (psychometric) approaches to the investigation of response processes, mirroring that seen in other areas of the social sciences, has been questioned. It has been suggested that the two can be complementary, particularly since modern psychometric approaches are able to model variations in response (Bjorner et al, 2003). The shortcomings of both qualitative (cognitive) and quantitative approaches highlight the need for an approach based on multiple methods, which has been adopted in this research.

Having outlined some of the methodological background of the research included in the thesis, the next section details the specific research questions addressed, relates them to the methodology employed and briefly outlines some of the key findings.

How do people understand and interpret the questions they are completing? That is, what are the conscious processes involved in answering health status items?

This question employed qualitative methods in an attempt to understand the cognitive processes involved in completion of the SF-12v2 items for university and HIV samples, each comprising ten participants.

Method: Two different approaches were used: half of the participants in each group were randomly assigned to answer a semi-structured self-completion questionnaire on response process issues, and the remainder took part in a cognitive interview, including both a prospective think-aloud component and retrospective probing about specific aspects of response (see Figure 1.1). Comments were analysed and grouped according to themes relating to stages of the response process (comprehension, retrieval/judgement, and response) (Tourangeau et al, 2000). Similarities and differences in themes were identified between the two participant groups.

Key findings: Participants were able to provide meaningful information about response processes, verbally or in writing. Despite some differences by sample and method, important aspects of the response process could be identified. There was variation in the interpretation of key terms, particularly for items that asked about general health, role performance and mental health. Differences in response processes between samples tended to relate to the manifold effects of HIV permeating participants' lives, influencing health and lifestyle. The university participants were more likely to refer to general perceptions in generating their answers, while HIV participants were more likely to refer to specific situations. Although there was some evidence of adaptation relating to chronic health limitations, responses generally reflected health deficits. Participants who found the SF-12v2 response options to be unclear indicated that they used the relative position of categories to guide their responses. Overall, participants actively engaged in the response process, using information gained from previous items to guide their progress through the questionnaire.

The remaining questions were investigated using quantitative data collected using a questionnaire survey administered to two samples. Although the target samples were both of 60 respondents (see Figure 1.1), the achieved samples comprised 64 university respondents and 72 HIV positive respondents.

Can the reliability and validity of a standard health status measure, the SF-12v2, be demonstrated for both healthy and health problem groups, using a classical psychometric approach?

Reliability and validity of the SF-12v2 have been demonstrated in general population and health problem groups (Ware et al, 2004). However, much of this work was carried out with SF-12v2 items included among a battery of other items. In addition, validation work has not yet been conducted with HIV positive respondents. Therefore, the psychometric properties of the SF-12v2 were tested here for both university and HIV samples.

Method: The survey data were analysed in terms of standard psychometric reliability and validity criteria in the two samples, including item facility, internal consistency, discriminant and convergent validity and scale principal component analysis.

Main findings: The SF-12v2 was found to perform well in both samples, according to standard classical psychometric criteria, with evidence that the scales were reliable and valid. However, multiple limitations in the HIV sample were reflected in numerous unpredicted correlations between physical and mental health items, and consequently, a principal component structure that was less well defined according to physical and mental health components than for the university sample. Nevertheless, significantly lower mean scores in the HIV sample compared to the university sample suggest that SF-12v2 physical and mental scales are sensitive to HIV-related problems.

Are responses to individual items in a health status measure invariant, regardless of health experiences?

The previous classical psychometric analyses demonstrated SF-12v2 reliability and validity. Following this, a modern psychometric approach, involving statistical modelling techniques, was used to model item response, to determine whether there were systematic differences in the way certain items were answered by respondents from the two sample groups, an effect known as Differential Item Functioning (DIF). If discovered, DIF could undermine the results obtained, as the comparability of scores would be compromised.

Method: DIF was estimated using Multiple-Indicator-Multiple-Cause (MIMIC), an extension of structural equation modelling (SEM). A model representing underlying physical and mental health dimensions of the SF-12v2 was assessed, with sample membership included as an exogenous variable.

Results: The first stage of the MIMIC analysis indicated that the two-factor model, based on latent physical and mental health factors, adequately reflected the combined university and HIV sample SF-12v2 data. In the second stage, the influence of sample on item response was tested. This demonstrated that patterns of scoring differed between HIV and university samples, particularly for the Bodily Pain scale, which inflated the overall physical score attained by the HIV sample in comparison to the university sample. This effect, however, was small. These findings suggest that, although there was evidence of DIF in SF-12v2 response, the effect of DIF on the scale scores attained was negligible.

What is the relationship between contextual factors (namely, individual factors, such as health experiences, perceptions, and behaviours, personality and sociodemographics), response processes and the answers given to questions about health status?

Relationships between a range of contextual and response process variables and the SF-12v2 were investigated in the two samples. This analysis had two functions. First, as a form of construct validation of the SF-12v2, investigating whether predicted associations were apparent. Second, the results informed the development of a response process model, which was subsequently tested with a path model.

Method: Statistical analyses were used to compare sample composition in terms of contextual and response process variables, and also to investigate relationships between contextual factors, response processes, and actual SF-12v2 response in each sample, with results compared between samples. Subsequently, a simple path model of response processes was tested in each sample with the SF-12v2 mental and physical health summary scale scores as the outcome.

Main findings: The two samples differed on a range of contextual variables, including sociodemographics, general health, attitude strength, personality traits, health behaviours and health service use. The HIV sample was more diverse than the university sample on most health-related measures, generally reported poorer health, more experience of health problems and greater use of services. In addition, when compared to the university sample, they made less use of a strategy of basing responses on general perceptions (they rather made a greater use of recalled specific experiences); rated the SF-12v2 easier to answer; and indicated that they felt the SF-12v2 was more useful for measuring their health. In both samples, SF-12v2 scores varied according to the contextual influences and response processes reported, generally as predicted, providing evidence for the construct validity of the SF-12v2. Effects were usually larger in the HIV sample, reflecting the wider range of health experiences in this sample.

Path modelling identified direct pathways between personality and health service use and SF-12v2 scores, differing according to physical or mental health. There were strong relationships between Neuroticism and mental health scores, and health service use and physical health scores in both samples. Additionally, increased Neuroticism and health service use were both directly associated with greater attitude strength (health concerns and thoughts). However, pathways involving response strategy (based on a count of the number of times a general perception strategy was reported) varied between samples and SF-12v2 physical and mental health summary scales. These initial results suggest that a contextual model provides a useful foundation for future work to develop more sophisticated models of response processes, taking into consideration both the outcome employed and the sample.

To conclude, different qualitative (cognitive) and quantitative techniques were used to conduct a detailed comparative analysis of the SF-12v2 in HIV and university sample, including detailed cognitive work on the different stages leading to a response, the psychometric properties of the measure in the samples, and quantitative analyses investigating relationships between contextual factors, response processes, and response, concluding with a path model of the individual contextual influences on judgement and response.

Part One: Overview of the theoretical and research background to the thesis

This section details the theoretical and research background which has informed the research questions and techniques employed, including: a historical overview of health perception measurement, generally and in relation to the SF-12v2 measure; an introduction to the complexities of health status measurement and chronic illness, including HIV/AIDS; and finally, an outline of a cognitive framework for survey response processes, which is a significant influence on the approach adopted in this thesis.

1.1 Issues in the measurement of health perceptions

The first part will describe the historical development of the measurement of health perceptions to the present day, when multidimensional questionnaires are commonly used to measure subjective health perceptions. Furthermore, this account covers the uses to which questionnaires on health perception are put, the types of questionnaires available, and the range of content of these questionnaires. Having provided a general introduction to these issues, the next section will include an account of the development of the SF-12v2, the measure included in the research conducted for this thesis.

Concepts and definitions

Crude mortality data have been collected in western nations for over one hundred years, mostly aggregated to regional or national levels (McDowell and Newell, 1987). These figures are derived from routinely collected data and are used to highlight the health of a population. Such indicators play an important role in informing health policy. However, in developed countries, simple mortality data are no longer seen as adequate measures of health. There has been a shift over time from viewing health as survival, to seeing it in terms of successful performance of daily activities and positive themes, such as happiness, well-being and quality of life, each concept capable of being defined in many different ways (Bowling, 2004; McDowell and Newell, 1987).

In 1948, the World Health Organization (WHO) published the following definition of health:

"Health is a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity" (WHO, 1948).

This definition continues to be used, and acknowledges that health is a complex multidimensional concept, incorporating components that are not simply associated with survival. However, this definition is broad and can only act as a starting point for those researchers wishing to study health in more detail. In addition, it raises one issue that needs to be considered by all those investigating health: how to reconcile the wish for a standard definition with the subjective nature of perceived health.

In the years since the WHO definition was proposed, there has been growing interest in the measurement of subjective health perceptions. Subjective health status is now used in many ways, such as health service, policy and clinical research (Bowling, 2001), where they have been claimed to provide an insight into the patient's perspective on their health (Ware, 1993). However, there is no single explicit theory underlying the measurement of subjective health perceptions and, consequently, there has been a proliferation of methods and measures (Bullinger, 2002; Carr et al, 2001; Gill and Feinstein, 1994). The most commonly adopted approach for the measurement of subjective health perceptions uses self-report questions, generally combined to form questionnaires (Hyland, 1992), although classification of questionnaires is also confused and varied. Between the years 1966 and 1986, for example, more than 3000 individual measures of "quality of life" were identified in a review by Prutkin and Feinstein (2002).

The concept of health status has been used to describe the measurement of subjective perceptions of health, being first cited in the Index Medicus in 1974 and used increasingly since (Bergner, 1989). However, measures of subjective health perceptions are sometimes claimed to assess health-related quality of life (HRQoL), referring to a separate research tradition drawn from the social sciences, investigating "quality of life" measured in terms of a broad range of dimensions (Prutkin and Feinstein, 2002). For most health questionnaires, the term health status is probably more appropriate, since they generally contain only a limited range of dimensions focusing on health-related functional status (Bowling, 2004). For this reason, health status will be the predominant term used in this work, referring more narrowly to a measure containing items on mental and physical functioning and well-being.

Use of health status measures

Stewart and Ware (1992) identified four important areas for the use of health status measures, namely health care policy studies, clinical trials, monitoring health of general population, and clinical decision making. In particular:

1. *Health care policy studies*, which can be further subdivided between health care systems and medical outcomes. Interest in health status measures has coincided with a period of change in health care systems throughout the world (Bergner, 1985; Donovan et al, 1993; Garratt et al, 1993; Hunt et al, 1986; Katz et al, 1992; Stewart and Ware,

1992). In the UK, the National Health Service (NHS) has undergone substantial reorganisations in recent years. This has led to the re-examination of service needs of local populations, and reviews of health care provision, including those conducted by the National Institute for Health and Clinical Excellence (NICE) (National Institute for Health and Clinical Excellence, 2005). A lot of research is being carried out into medical outcomes that provide more information than simple clinical endpoints such as death or discharge from hospital (Bergner, 1985; Garratt et al, 1993; Hunt et al, 1986; Jenkinson et al, 1993; Katz et al, 1992; Ware, 1991). The subjective element of outcome has been defined as "the extent to which a change in the patient's functioning or well-being meets the patient's needs of expectations" (Stewart and Ware, 1992). These outcomes have a clinical element in assessing the effectiveness of treatments, but they also relate to health policy and health care provision, mentioned above.

2. *Clinical trials*. These are used to evaluate the effectiveness of new treatments and procedures. Traditional measures used to assess clinical trials have included clinical symptoms and cost. However, these are now commonly supplemented with health status measures in order to assess patient functioning or well-being (Bergner and Rothman, 1987; Hunt et al, 1986; Patrick and Deyo, 1989; Stewart and Ware, 1992; Ware, 1991).

3. *Monitoring health of general population*. Standardised mortality rates and treatment survival rates do not provide detailed information about population health, first because they do not relate to most of the population outside the health care system (Stewart and Ware, 1992) and, second, they are epidemiological indicators rather than measures of the health of individuals in the population. Stewart and Ware (1992) remark that standard measures of health status could be used to monitor health in the general population and provide norms for groups within the population.

4. *Clinical decision-making*. The use of health status measures has been suggested for everyday clinical practice (Patrick and Deyo, 1989; Stewart and Ware, 1992). Patient well-being is often discussed informally during consultations between patients and health care professionals. Stewart and Ware (1992) write that the routine use of health status measures in practice would aid clinical decision-making and allow patients and professionals to make more informed decisions about individual patients. However, health status measures are most widely used to provide aggregate data about groups of individuals, such as patients with a particular health problem, or a general population sample.

Types of measures

It has been suggested that health status measures can be broadly divided between *generic* and *disease-specific* measures (Patrick and Deyo, 1989). *Generic measures* focus on "...such basic human values as emotional well-being and the ability to function in everyday life" (Ware, 1991) and should be applicable in any study of health, regardless of the population. They have been advocated for a number of tasks, such as comparisons between disease types, disease severity, treatment regimens, change over time, and across demographic and cultural subgroups (Patrick and Deyo, 1989; Ware, 1991). In contrast, *Disease-specific measures* focus on issues that are pertinent to a particular disease or condition. However, a third type of measure can be added, *individualised quality of life measures* (Joyce et al, 2003; Ruta et al, 1994). Unlike the generic and specific measures, these are not limited by a series of closed questions. Instead, respondents are requested to select and rate areas of their health, or life more broadly, that are important to them. These measures relate to a perspective where individual context and preferences are considered to be a crucial aspect of the assessment of health status (May and Warren, 2001).

The research conducted for this thesis deals with issues surrounding a generic measure, the SF-12v2. It has been suggested that generic and disease-specific measures are complementary and should be used alongside one another (Patrick and Deyo, 1989). However, the development and conceptualisation of generic measures is different from disease specific measures, trying to encapsulate a much broader sense of health, capable of being relevant to respondents with any health problem or none at all. The research included in the thesis does, however, reflect on issues relating to person-specific measures, incorporating a rating for the perceived usefulness of each item for assessing respondents' health.

Content of measures

Although some health status developers argue that they at last incorporate the lay perspective into the health care system (Hunt et al, 1986; Stewart and Ware, 1992), most structured health status measures are based on a model which focuses on the impact of ill health on physical and mental functioning in society, rather than positive aspects of health asserted by the WHO definition (Bowling, 2004). Researchers and clinicians are generally familiar with the methods of natural science, and are willing to accept psychometrically derived, structured health status measures, even if little detail is given on any theoretical foundations (Carver et al, 1999; Donovan et al, 1993).

The development of measures based on expert opinion or adaptations of previous scales and items rather than the concerns of laypeople continues this pattern (Bowling, 2001; Rogerson et al, 1989). Stainton Rogers (1991) criticises the psychometric techniques used to design questionnaires to assess health by making three points. First, the researcher predetermines the results they are likely to receive; second, researchers interpret what respondents mean when they answer without reference to the respondent's own understanding; third, variability, ambiguity and inconsistency are constrained by the psychometric method whereas they are likely to play a real part in people's everyday experiences of health. Moreover, little is known about whether the items included in measures are relevant to the respondents who are asked to complete them. Studies investigating content from alternative perspectives have found that items selected as important by lay people do differ both from items selected using standard psychometric techniques emphasising homogeneity, and items rated important by health care professional (Carver et al, 1999; Juniper et al, 1997; Kane et al, 1998; May and Warren, 2001; Rothwell et al, 1997).

Thus, despite the widespread use of standardised health status measures, there is also criticism surrounding their development and application. Although this section has provided a brief account of some of the general issues in health status measurement, the next section details the origins and development of a particular health status measure, the SF-12v2, which is the focus of the research carried out in this thesis.

1.2 SF health status measures

The Short Form "SF" series of measures have a history typical of many generic health status measures, involving development and refinement, validation studies and the collection of normative data. The theoretical basis and psychometric qualities of the current SF-12v2 cannot be fully understood without reference to the long-term development process from which it has evolved; in particularly, the relation between the SF-12 and the SF-36, which are described below.

Background to the SF health status measurement programme

"SF" measures came about as the result of two large-scale research projects, the Health Insurance Experiment (HIE) and the Medical Outcome Study (MOS). The HIE was carried out by the RAND corporation with the goal to investigate options for funding health services in the United States, requiring the development of scales to measure '…a broad array of functioning and well-being concepts' in adults and children, measuring both physical and mental health (Stewart and Ware, 1992). The HIE contained 108 items and took 45 minutes to complete. There was a clear conception of health as multidimensional, for which scoring profiles could be produced rather than a single score (McHorney et al, 1993). Items related to the WHO definition of health of physical, mental, and social health (WHO, 1948), along with general health perceptions, which was considered to reflect unique information about health (Ware et al, 1980).

Although the measure was reported to be reliable and valid for use with a general population, many people refused to complete it because of its large number of items. The researchers recognised the need for a measure that was not as lengthy, yet more comprehensive than single item health measures, eliciting a rating of perceived general health, used in many clinical investigations (Ware and Sherbourne, 1992). To enable some information to be collected from participants, a shorter questionnaire was developed based on the HIE questionnaires, designed to be both comprehensive and brief, which was used successfully (Ware et al, 1993).

Subsequently, the HIE team developed a 'comprehensive' 18-item short-form health survey measuring physical functioning, role limitations due to poor health, and current health perceptions, which was used in a general survey carried out in 1984 (Ware et al, 1993). Two additional questions measuring social functioning and bodily pain were added to the 18-item questionnaire in 1986, to create the "Short-Form" (SF)-20, which was used with the elderly and adults with chronic medical conditions in the Medical Outcomes Study, a project to investigate '... specific characteristics of providers, patients, and health systems on outcomes of care' (Hays et al, 1993). A major aim of the MOS was the advancement of state of the art methods designed to routinely monitor outcome of patient care. The study consisted of both cross-sectional and longitudinal components. The cross-sectional component served to screen patients with target tracer conditions (hypertension, congestive heart failure, recent myocardial infarction, diabetes, and severe depression), who were followed in detail in the longitudinal component. The cross-sectional sample consisted of 22462 adult patients. The size of the cross-sectional sample necessitated the use of a short-form survey, the SF-20. Analysis of the MOS data sets showed the SF-20 to be reliable and valid (Stewart and Ware, 1992). However, additional research showed that individuals who had scored at the floor (lowest possible score) on some SF-20 scales had later health decrements undetected by the measure (Hays et al, 1993).

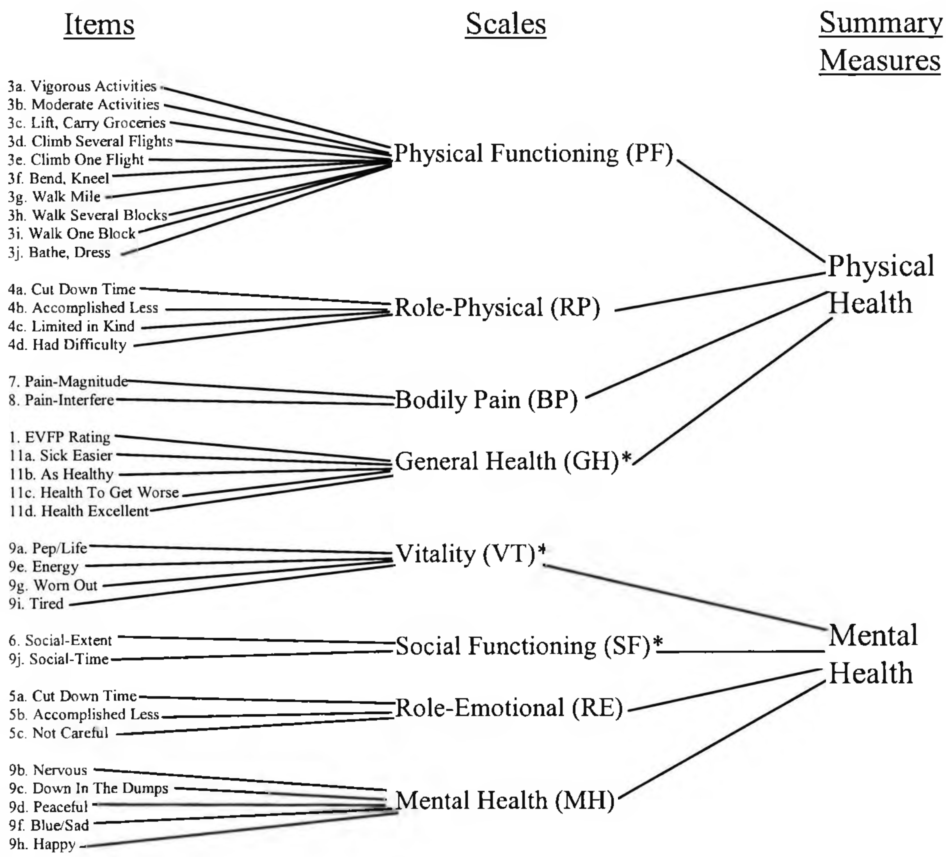
Initial development and validation of measures

The next stage in the process of the development of a "Short Form" measure was the SF-36, dating from 1988-1990 (Ware, 1993; Ware & Sherbourne, 1992). The SF-36 was planned to be a short, yet comprehensive, multidimensional measure, capable of identifying differences between groups and individuals over time and across populations. The dimensions that constitute health status as measured by SF measures were derived from the WHO definition of physical, mental and social well-being, concepts found in the literature, and constructs developed for the HIE and MOS projects (Ware and Donald, 1992).

The development of the SF-36 scales typically involved defining a scale content area, drawing an item pool from earlier instruments and expert opinion, carrying out a pilot study to obtain basic response data, and selecting items to include in the scale through the use of multi-trait methods and factor analysis. Of the original forty health concepts included in the MOS, eight were included in the SF-36 (Stewart & Ware, 1992), and the SF-36 was validated against the original MOS measures from which it was developed. The SF-36 was designed to have low respondent burden, taking approximately ten to fifteen minutes to complete, either by self-administration, or by interview over the telephone or face-to-face. Thirty-five items were included in the scales with an additional item on perceived changes in health over the course of one year.

The SF-36 has been validated in numerous studies from many countries since it was first published, a recent survey of medical publications revealing that it is the most widely evaluated patient-assessed measure (the SF-12 was number twenty-two) (Garratt et al, 2002). There is evidence that the scales differ in their ability to detect different conditions as hypothesised, with the Physical Functioning (PF) scale more valid for chronic medical conditions and the Mental Health (MH) scale more valid for psychiatric conditions (McHorney et al, 1993), that scale scores are associated with health service use (Brazier et al, 1992), and that the SF-36 is sensitive to health changes over time (Garratt et al, 1994). As alluded to, the conceptual model for the SF-36 represents eight scales underpinning two overarching health dimensions, physical health and mental health (Figure 1.2).

Figure 1.2: SF-36 measurement model



* Significant correlation with other summary measure.

Source: Ware (2000)

Physical health consists of the scales Physical Functioning (PF), Role-Physical (RP) and Bodily Pain (BP) and General Health (GH), while mental health includes the scales Role Emotional (RE), Mental Health (MH), Vitality (VT) and Social Functioning (SF). Subsequent factor analytical studies have confirmed the existence of physical and mental health components (Ware et al, 1993). Three of the physical health scales have been found to correlate highly with only the physical component (PF, RP and BP) and two of the mental health scales to correlate highly with only the mental component (RE and MH). The General Health (GH), Vitality (VT) and Social Functioning (SF) are associated with both components, although all more strongly to their hypothesised components. The confirmation of the physical and mental components provided an opportunity to carry out research using higher-level concepts, a Physical Component Summary (PCS-36) and a Mental Component Summary (MCS-36). These weighted linear combinations of all eight scales capture between 80-85% of the scoring variance (Ware et al, 2004). The finding that two summary scales are capable of explaining most of the variance previously measured using eight scales led the development team to investigate whether a shorter measure could be produced. Regression analysis showed that a twelve-item subset of the items in the SF-36 reproduced at least 90% of the scoring variance in PCS-36 and MCS-36 in general population and patient samples while remaining comprehensive, with one or two items drawn from each of the eight scales (Ware et al, 1994). These items comprise the SF-12, a shortened measure which provides summary scores for physical and mental health, but which takes on average only 2 to 3 minutes to complete. Validation for the shorter measure was claimed, with SF-12 and SF-36 summary scales strongly correlated, scoring differences between health groups as hypothesised, and evidence that, for group comparisons, the same patterns of significant results were obtained for both measures (Ware et al, 1994).

The development of the SF-12v2 measure

Ongoing development of the SF measures has resulted in revisions to the SF-36 and, consequently, the SF-12, based on the use of psychometric analysis and qualitative techniques, including focus groups and cognitive interviewing (Ware et al, 2000). The first, version 2.0 of the SF-36 (SF-36v2), was introduced in 1996. Changes to the original included alterations to instructions and question wording, to improve sense; and simplification of question layout. Additionally, response options were changed to improve item and scale characteristics. The number of response categories for physical and emotional role functioning items increased from two to five. This has increased the range and improved measurement precision, together with reducing floor and ceiling effects for the Role Physical (RP) and Role Emotional (RE) scales. In contrast, the number of response categories for the Mental Health (MH) and Vitality (VT) items was reduced from six to five, due to the ambiguity of one of the response options, found using the Thurstone Method of Equal-Appearing Intervals (Keller et al, 1998). These changes have also simplified response, reducing the number of response category formats from six in the SF-36v1 to three in the SF-36v2.

In the original SF measures, scale scores ranged from 0 – 100, where 0 indicated poor health and 100 no health problems for that scale. Although the original scoring is also available for version two of the SF measures, norm-based scoring was implemented as the standard. Norm-based scoring (NBS) involves linear transformations of original scores to produce T-scores, normed in relation to the 1998 United States general population figures, which have a mean of 50 and a standard deviation of 10. The advantage of norm-based scoring is that it allows for straightforward comparisons: between scales within a measure, different SF measures that have been scored using NBS, different health groups, and across countries in relation to this standard. A one-point difference in scores on any Short Form NBS scale can be interpreted as one tenth of a standard deviation unit.

Table 1.1: Items included in the SF-12v2

Scale	Item number	Abbreviated content (item name)
Physical Functioning (PF)	2A 2B	Moderate activity (PF02) Climb several flights of stairs (PF04)
Role Physical (RP)	3A 3B	Accomplished less (RP02) Limited in kind of activities (RP03)
Bodily Pain (BP)	5	Pain impact (BP02)
General Health GH)	1	Health in general (GH01)
Vitality (VT)	6B	Lot of energy (VT02)
Social Functioning (SF)	7	Social impact of health/well-being (SF02)
Role Emotional (RE)	4A 4B	Accomplished less (RE02) Did work less carefully (RE03)
Mental Health (MH)	6A 6C	Felt calm (MH03) Felt downhearted (MH04)

The SF-12 was revised in accordance with the SF-36v2, including the wording changes, the increase in response options for Role Physical (RP) and Role Emotional (RE) items, and the reduction of Mental Health (MH) and Vitality (VT) options. Table 1.1 shows the twelve items included in the SF-12v2, together with a summary of the content of each item, item number of the questionnaire and the item name. The changes to response options have been shown to reduce floor and ceiling effects observed for the original SF-12 and to provide more finely graded scales. Research into item functioning carried out following improvements to items and scoring algorithms, have led the developers to suggest that scores can be provided reliably for all the eight scales as well as the physical and mental summary scores (Ware et al, 2004). The meaning of scales is shown in Table 1.2.

Table 1.2: Meaning of floor and ceiling scores on SF-12v2 scales

	Definition	
Scales	**Lowest Possible Score (Floor)**	**Highest Possible Score (Ceiling)**
Physical Functioning (PF)	Very limited in performing physical activities	Performs physical activities without limitations due to health
Role-Physical (RP)	Problems with work or other daily activities as a result of physical health	No problems with work or other daily activities
Bodily Pain (BP)	Very severe and extremely limiting pain	No pain or limitations due to pain
General Health (GH)	Evaluates personal health as poor	Evaluates personal health as excellent
Vitality (VT)	Feels full of energy none of the time	Feels full energy all of the time
Social Functioning (SF)	Frequent interference with normal social activities due to physical or emotional problems	Performs normal social activities without interference due to physical or emotional problems
Role-Emotional (RE)	Problems with work or other daily activities as a result of emotional problems	No problems with work or other daily activities
Mental Health (MH)	Feelings of psychological distress all of the time	Feels peaceful, happy, and calm all of the time
Physical Component Scale (PCS)	Limitations in physical, social, and role activities, severe bodily pain, lack of energy, health rated "poor"	No physical limitations, disabilities, or decrements in well-being, high energy level, health rated "excellent"
Mental Component Scale (MCS)	Frequent psychological distress, social and role disability due to emotional problems, health rated "poor"	Frequent positive affect, absence of psychological distress and limitations in usual social/role activities due to emotional problems, health rated "excellent"

Source: Adapted from Ware et al (1993).

Normative data for the SF-12v2 were gathered from representative samples of the 1998 non-institutionalised general US population (n=7069 for the standard SF-12v2). Data were collected by postal survey, with a response rate of 68%. The mean age of the sample was 50.7 years (age range 18-96) and 23% were sixty-five years or older. Most respondents were women (59.6%), of white ethnicity (84.2%), and with least 12 years of education (79.8%). Sampling weights were used in order to match the 1998 age and gender distribution of the US population, according to US Census Bureau data (Ware et al, 2004).

The developers claim that:

"The SF-12v2 is replacing the SF-36 as the instrument of choice in many population surveys that require a shorter instrument." (Ware et al, 2004).

Although the SF-36v2 has better coverage of health domains, the same eight scales and the summary profiles are included in the SF-12v2 with a third of the items. Brevity is an advantage in studies of response processes, making it suitable for intensive investigation of individual items, without overburdening the respondent. There is evidence that respondents find the shortness of the SF-12 a favourable characteristic in comparison to longer measures, such as SF-36 (Ware, 2004). All "SF" measures include items that cover different manifestations of multidimensional health status: functional status (observable, tangible standards, external to the individual), perceived well-being (subjective judgements about an individual's affective state which cannot be inferred by external behaviour), social and role disability, and general health perceptions (a personal evaluation of current health, susceptibility to illness and health outlook) (Ware et al, 1993).

Reliability

Reliability can be measured using a number of standard procedures to produce estimates of the stability and consistency of a measure (Rust and Golombok, 1999). Forms of reliability reported in relation to the SF-12 are: internal consistency, alternate forms / equivalence, and test-retest reliability, details of which follow. The coefficient produced by tests of reliability provides an indication of the size of the error variance in relation to the true score variance. A reliability coefficient of 0.70 (indicating that 70% of total measured variance is true score) or more is considered adequate for group-level analyses, such as those conducted in surveys (Ware et al, 2004).

Internal consistency

This technique assesses the extent to which items in a scale are equivalent, taken as an indication that they measure the same concept. A conventional approach to testing internal consistency involves calculating Cronbach's alpha, the average inter-item correlation (Cronbach, 1951). However, the appropriateness of this approach has been called into question when a scale has few items (Ware et al, 2004). The SF-12v2 was designed to be relatively heterogeneous, with items selected because they contain unique variance in the estimation of physical and mental health. In addition, none of the eight scales contains more than two items and a number of scales are represented only by a single item. However, the SF-12v2 developers report that an internal consistency approach was adopted to test the reliability of the SF-12v2 summary scales, PCS-12 and MCS-12, taking into account the reliability of the SF-12v2 scales as well as covariance among them. US general population data showed overall internal consistency to be good for both summary profiles (PCS-12 =0.89 and MCS-12=0.86). Comparative estimates according to health, age and gender also revealed high estimates for internal consistency: PCS-12 (range: 0.80-0.90) and MCS-12 (range: 0.82-0.88) (Ware et al, 2004).

Alternative form / equivalence

An alternate forms approach was also used by the developers to test the reliability of the eight SF-12v2 scales. This involves administering to the same subjects both the scale being tested and a different, equivalent set of items. The correlation between the scores obtained on the two forms represents the equivalence of the alternate forms of the scale. Since the SF-12v2 items were drawn from a much larger items pool measuring the same eight scales, it was possible for the test developers to compare the SF-12v2 items with other, comparable items contained within the original item pool. Alternate forms reliability was calculated by correlating SF-12v2 scale scores with a total score for each health concept from the item pool, excluding the SF-12v2 items. In the US general population, overall alternate form reliability estimates were all above the 0.70 recommended level, ranging from 0.73 (BP) to 0.86 (RP). Comparative estimates, according to health, age and gender, were lower among those who did not report any chronic conditions ("healthy") compared to those with one or more health conditions, falling below the recommended level for three scales (BP, SF (note: also for males), and RE). The test developers argue that these results were due to there being less variability in the scores estimated from the total item pool and the SF-12v2 scale among the "healthy" respondents compared to the others (Ware et al, 2004). This would indicate that "healthy" respondents achieved similar scores for both the SF-12v2 scales and equivalent scales from the item pool. In contrast, among the health problem respondents, perceived distinctions between items influenced response, leading to greater variability in the scores achieved for both the SF-12v2 and the item pool. Such findings suggest that the SF-12v2 measurement properties may differ according to the health experiences of a particular sample.

Test-Retest reliability

Retest reliability shows the extent to which scores on a measure can be generalised over different occasions, shown by correlating the results obtained from each administration. However, it is difficult in practice to measure test-retest reliability in health because real changes could have taken place over time. No coefficients were reported by the SF-12v2 developers, although it has been demonstrated to be adequate for the SF-12: PCS=0.89; MCS=0.76 (Ware et al, 1996).

Validity

In psychometrics, the operationalisation of validity as a concept takes different forms, the overall aim of which is to assess whether a test is meaningfully measuring what it purports to measure. The validity of a test is problematic, since it relates to underlying theory and assumptions rather than simply issues of statistical procedure. Different forms of validity have been reported for the SF-12v2, including content validity, and construct validity, including convergent and discriminant validity, detailed below. However, health status measures, such as the SF-12v2, have been criticised for a lack of face validity, relating to the acceptability of measure by respondents, and this is also discussed.

Content validity

This involves the systematic examination of the test content to determine whether it covers a representative sample of the area under investigation. As explained earlier, the content of the areas included in the SF-12v2 are based on two theoretical dimensions, physical and mental health which are further divided into the eight scales, included in all SF measures. The developers argue that the SF measures contain "... health concepts that represent basic human values and are relevant to everyone's health status and well-being" (Ware, 1987). Many items were selected from instruments and scales "...that have been in use for more than 20 years" (Ware et al, 1993). Although it may be argued that a widely used or long-standing item has demonstrated robustness, this is an assumption that may not necessarily be borne out empirically. Older items may not retain their initial meanings and may be difficult for a contemporary respondent to answer (Schuessler, 1982). Items should be reviewed and tested in the context in which they will be used. In addition, the selection of items based on those used in previous measures may also limit the content considered appropriate, with underlying concepts restricted to those previously developed. As a consequence, critics have argued that the empirical, psychometric approach taken during the development of the SF measures insufficiently involves lay people in defining the content of the measures, and instead relates to policy-makers concerns (Hunt and McKenna, 1993), and health professionals' assumptions about issues relevant to health status (Hunt and McKenna, 1992).

Construct validity

This refers to the extent to which the test may be said to measure intended theoretical constructs. Construct validity is difficult to establish because it involves evidence from many sources and a coherent theoretical structure for the concepts being measured. It requires the gradual accumulation of information about the construct from various sources and, therefore, support for the validity of the SF-12v2 has included evidence from earlier research including the SF-12v1 and the SF-36v1.

The principal approach to demonstrating the validity of the SF-12v2 has involved testing for differences in scores between groups known to differ according to physical or mental health problems, both, or none at all. This work has identified that, as hypothesised, scores obtained on the physical health scales are better able to discriminate between those with physical conditions compared to others, while scores on mental health scales can best discriminate those with mental health conditions. Comparisons with the corresponding SF-36v2 scales has indicated that, although the longer measure is more valid, being better able to discriminate between health groups, similar patterns of scoring are shown for the SF-12v2 (Ware et al, 2004). The analyses of the relationship between contextual and response process variables included in this thesis may also be considered to relate to construct validity, since the contextual factors were hypothesised to have specific patterns of association with health status, which could be tested.

Additional support for validity relates to studies produced by the developers and others using the original SF-12, which has been in the public domain since 1995. These provide considerable evidence that the summary scales vary in theoretically consistent ways (Ware et al, 2004). For example, showing predicted differences according to specific physical or mental health problems, severity of medical condition, the presence of comorbid conditions, the existence of acute symptoms, and in comparison with other health status measures. Age differences have also been shown, with lower physical health scores, and, to a lesser extent, higher mental health scores among older people. A follow-up investigation using self-rated health transitions demonstrated that those who rated their health to have declined over the year were more likely report worsening physical and mental health SF-12 scores compared to others. Along with published research into a range of other medical and psychiatric conditions (see Ware et al, 2004), the original SF-12 scale scores have been associated with clinical signs in studies of HIV patients (Delate and Coons, 2000; Han et al, 2002), and to increase as predicted following the initiation of anti-HIV treatment (Mannheimer et al, 2005). More information on the HIV studies will be provided in the next section.

Finally, tests of the validity of the item structure refer to original work conducted on the SF-36v1. Convergent and discriminant forms of construct validity were demonstrated, with scales correlating more strongly with those scales they were hypothesised to relate to, and weakly with others (Ware et al, 1993). In addition, the hypothesised factor structure, involving overarching physical and mental health dimensions, has been demonstrated, with items or scales loading highly on the factors as predicted (Ware et al, 2000). Analyses confirming the hypothesised two-factor structure for SF measures may be considered evidence for the construct validity of the measure, showing that the test measures underlying theoretical constructs.

Face Validity

It has been suggested that meaning is central to understanding subjective views and essential for the validity of subjective health measures (Hunt, 1997; Mallinson, 2002). In addition, on a more practical level, if the questions seem irrelevant or pointless, respondents may become bored and disengage from the task, or start to satisfice, providing answers without considering them properly, even if the measure is otherwise empirically valid (Anastasi, 1968; Krosnick, 1999). A problem with validating ratings of subjective experiences, such as health status, is how to assess the accuracy of these responses, as there are seldom any objective criteria on which to make such judgements (Beyer and Bowden, 1997) without resorting to external indicators, for example health service use. Indeed, conventional validation techniques often involve examining the relationships between a measure and external criteria, or analysing items to test whether a hypothesised factor structure is achieved. Such approaches have been adopted when validating the SF measures, which were developed with little input from laypeople or patients. However, a number of researchers have expressed the wish for a broader definition of validity, linking the psychometric properties of a measure with the use to which it is to be put. This may take the form of an explicit consideration of the appropriateness, meaningfulness and usefulness of the measure (Messick, 1998); for example, in terms of social policy implications of the results (Bornstein, 1996), or individuals' views of the questions they are being asked (May and Warren, 2001). It has been suggested that an extension to the standard psychometric validation should be carried out by questionnaire developers, systematically investigating the meaning and interpretation of questions, including the processes engaged by respondents when answering questions (Messick, 1995).

Testing the meaning of questions is considered to be one of the primary aims of questionnaire development process (Converse and Presser, 1986). A pragmatic approach has often been adopted by those who develop health status questionnaires, aimed at gaining the acceptance of the respondents in the questionnaire by attempting to make instruments that are relevant, in order to reduce the risk of error due to misinterpretation and/or lack of motivation. If items have a high response rate, this is taken to indicate acceptance and comprehension of the item. Dixon et al (1994) argue that this is not a good indicator of acceptance since respondents may co-operate without necessarily accepting the content of the items they are completing. The results from research investigating the influence on questionnaire score of respondent understanding and acceptance of items, or face validity, has been equivocal. Holden and Jackson (1979) found that greater acceptance was empirically associated with criterion validity for personality scales. Turner and Fiske (1968) found that items that had been interpreted by the respondents as the developers intended correlated most highly with overall test score. However, Adams (1950) found that the relationship between acceptability and empirically assessed criterion validity was small and undependable. Nevo (1985) indicated that face validity could be measured by asking respondents directly, although whether the respondent can make judgements about the validity of a test has been questioned (Secolsky, 1987). It has been suggested that patients should be asked to indicate which items in health questionnaires were "personally important" to them, although this suggestion does not seem to have been widely implemented (Gill and Feinstein, 1994).

Overall evaluation of the SF-12v2

In summary, the SF-12v2 is a twelve item multidimensional instrument that is generally completed by the respondent in approximately 2-3 minutes. It is part of a family of measures that have been developed over many years, during which time statistical techniques have been used to improve measurement quality while reducing the number of items required for accurate measurement of health status. The SF-12v2 thus represents a trade-off between comprehensiveness and brevity. Psychometric analyses have shown that the SF-12v2 can measure eight mental and physical health scales reliably and validly. The SF-12v2 therefore represents a standard, multiple item measure of health status, similar to many others that are currently in use world-wide. The research in this thesis takes the opportunity to study in more detail the properties of the SF-12v2, including the use of reliability (internal consistency), and validity (construct, including convergent and discriminant validity and factor structure, and perceived face validity of the SF-12v2 as a measure of subjective health), plus modern psychometric analysis, modelling of item response.

Having discussed the historical and methodological issues involved in the development of health status in general, and in relation to the SF-12v2, the next section focuses on the complexities of the relationship between health status and health. Although some of the issues surrounding respondent perceptions have been mentioned earlier, including the question of who determines what health status should be, the next section provides more detail on some of the problems relating to the measurement of health among those with a chronic health problem, such as HIV, together with an overview of research on the relationship between health status and HIV.

1.3 HIV/AIDS and health

Both the qualitative (cognitive) and quantitative studies included in the thesis compare people with a health problem (HIV) and a nominally healthy group (university students and staff). There is considerable evidence that health status scores vary according to illness. However, less is known about response processes and health status. This section describes research on health experiences and response process, before introducing specific findings from research on HIV and health status.

Health experiences and health status response processes

The scant evidence available suggests that personal experience of a serious, chronic health problem may be an important personal contextual influence on response processes to questions about health (May and Warren, 2001; Warnecke et al, 1996). It has been suggested that people with no health problems or with longstanding illnesses may have fewer problems thinking about long-term health than people with intermittent health problems (Jenkinson et al, 1996). Qualitative research has shown that respondents with health problems can have difficulty meaningfully relating their own perceptions to the fixed statements and response categories provided in a health status questionnaire (Donovan et al, 1993), and it is often difficult for them to disentangle considerations of general health from those of a particular illness (Allison et al, 1997; Dijkers, 1999; Jenkinson et al, 1996). In these cases, the respondent can be viewed as an active participant in the response process (Tourangeau and Rasinski, 1988), interpreting and reinterpreting questions in order to make sense of them (Manderbacka, 1998), with interpretation functioning as a mediating mechanism between contextual factors (including personality and objective health measures) and response (Brief et al, 1993). Other research has looked at retrospective assessments of quality of life ("then" ratings) in order to investigate change in perceptions (Albrecht and Devlieger, 1999; Allison et al, 1997; Dijkers, 1999). This is dependent on memory and the assumption that evaluations of importance do not change over time. However, it is difficult to justify such an assumption. There is evidence that people adapt to illness over time, adjusting their perceptions of health and illness, giving rise to the common finding that respondents with illnesses do not provide a health status rating that would be expected of a "sick" person (Allison et al, 1997; Manderbacka, 1998; Wright et al, 1992). Research on "response shift" has shown that people may reorder or change of goals and values during the course of a chronic illness (Sprangers and Schwartz, 1999).

The nature of change in health perceptions may be complex and varied, influenced by personality factors, and related to evolving patterns of expectations and comparisons (Allison et al, 1997). Different models have been outlined, ranging from alpha change (the conventional understanding of change, where the fundamental construct being measured remains the same), to gamma change (where the respondent's terms of reference change completely over time). Several psychological mechanisms underlying the dynamism of health status have been suggested, including adaptation, coping, self-control, uncertainty, self-concept, and expectations (Allison et al, 1997). Schechter (1993) proposes that analyses of health question response processes should be carried out according to varying groups, such as those with a health problem compared to others without such problems, in order to learn whether processes differ. This is the approach adopted in this work, comparing response processes reported in HIV (health problem) and university (nominally healthy) samples.

HIV and health status

HIV is an infection that progressively damages the immune system, causing a range of opportunistic infections or tumours, many of which would be negligible in a healthy individual. As the immune system weakens, illnesses become increasingly more severe, leading eventually to an AIDS diagnosis. Health status has been demonstrated to be generally lower among people with symptomatic HIV than in the general population, including research using the SF-12 in a sample of HIV positive patients recruited in a clinical setting, which showed lower physical and mental health among people with HIV compared to the US population norms (Delate and Coons, 2000).

However, health status varies between people with HIV/AIDS, and this variation may provide insights into HIV progression and treatment efficacy. HIV/AIDS is now generally considered a chronic condition, so that quality of life is a main concern, rather than mere survival. As with many other chronic illnesses, health among HIV positive people is multifaceted. In addition to the negative consequences of HIV/AIDS and concomitant illnesses, drugs taken to combat HIV can often have adverse effects in themselves. The complexity of disease and symptom management have led to the widespread use of health status measures as a research and outcome tool among people with HIV/AIDS. New treatments are being developed and routinely trialled, for which subjective perceptions of health are considered an important component, alongside traditional clinical measures (Han et al, 2002).

A number of studies have shown associations between health status or quality of life and objective markers of HIV progression, including disease stage, symptoms, hospitalisation, weight loss, but, most commonly, virological markers of disease progression (viral load and CD4) (Call et al, 2000; Campsmith et al, 2003; Carierri et al, 2003; Delate and Coons, 2000; Gill et al, 2002; Mannheimer 2005; Murri et al, 2003).

There is also evidence that health status research can provide unique information about treatment issues, disease progression and mortality. The efficacy of many anti-HIV regimes is unknown, and they may have severe side effects (Delate and Coons, 2000). However, good adherence to a treatment regimen has been shown to relate to better health status (Carierri, 2003; Mannheimer et al, 2005). Longitudinal research has shown that health status varies along with disease and treatment status and may even predict clinical outcomes. For example, clinical trial evidence, using the SF-12, has shown that that poorer health status is linked to subsequent disease progression and death (Han et al, 2002). The initiation of HAART (Highly Active Anti-Retroviral Therapy) has been associated with an improvement in subjective health status among people with HIV (Carierri, 2003; Mannheimer et al, 2005), although adverse effects on health following a change in therapy has been related to worsening mental health status (Eriksson et al, 2005). It has been suggested that the introduction of HAART as the standard treatment of HIV/AIDS has resulted in less dramatic changes in health status over time compared to the pre-HAART era. Stability in health status ratings has been observed in people on HAART, despite a concurrent increase in reported symptoms (Saunders and Burgoyne, 2002).

The next section provides a detailed introduction to the survey response process, which has been alluded to throughout this section. The particular perspective adopted in this research is based on a cognitive model of response, which is the framework for methods and analysis described in the remainder of the thesis.

1.4 Survey response processes

Before any survey question is answered, whether about health or another topic, a potentially complex cognitive process has been undertaken. This section outlines developments in the conceptualisation of the nature of the response process, a model of the response process, specific issues about response processes and health status, the relationship between contextual factors and response, and, finally, approaches to survey response process research.

Conceptualisation of the response process

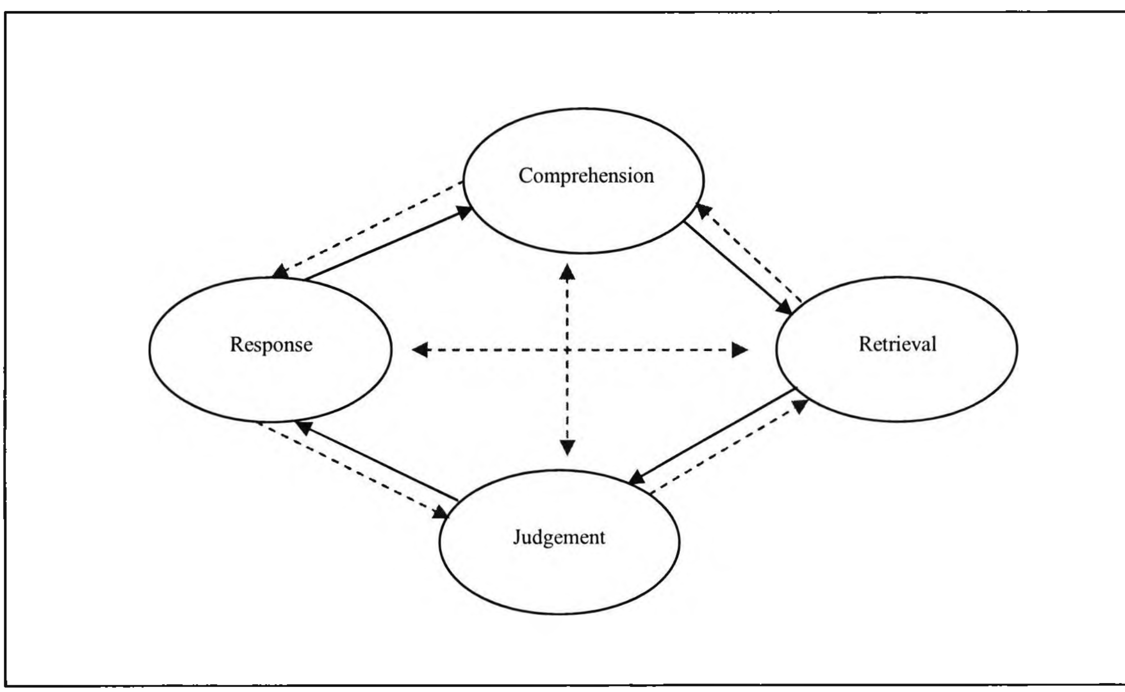
Survey researchers have traditionally paid little attention to the processes leading to the responses that are the focus of their work. There has been a tendency to assume that respondents are willing to put effort into understanding questions, accurately recalling information and then accurately formulating answers. A strict embodiment of this belief posits that respondents have values capable of expression through meaningful responses, provided they are asked appropriately (Fischoff, 1991). In this view, a standardised instrument and instructions will lead to a sampling of beliefs, followed by an appropriate response, representing a "true" attitude (Hyman, 1954; Lazarsfeld, 1944).

This perspective has been challenged in recent years, notably in the Cognitive Aspects of Survey Methodology (CASM) programme (Jabine et al, 1984; Sirken et al, 1999). Much of this research has focused on the cognitive processes employed by respondents when interpreting and responding to questions. A number of experimental and non-experimental techniques have been used to investigate the cognitive processes of response that are accessible to the respondent (Jobe, 2003). Findings from this research have shown that questionnaire response is typically complex, with many influences, some of which may provide important information about the way people deal with the world around them, and with the task of responding to a questionnaire (Collins, 2003).

A cognitive model of response

There is increasing evidence that questionnaire response processes comprise a range of tasks, leading to answers which not only relate to the substantive survey questions, but many other factors. The tasks involved in this process have been operationalised in terms of a four-stage information-processing model of response (Figure 1.3), comprising tasks required to understand and respond to a survey question (Tourangeau et al, 2000):

Figure 1.3: Four-stage model of response process



(Source: Collins et al, 2003)

Although outlined as a series of distinct stages, from item comprehension, through to retrieval of information, judgement and final response, in reality the response process is likely to be dynamic rather than linear, with interactions between stages and external influences on response (Collins, 2003; Tourangeau and Rasinski, 1988). In this four-stage model, respondents are active participants in the survey process, attempting to manage their image, trying to make sense of questions and providing answers if possible, and striving for consistency (Clarke and Schober, 1992). However, answering a question requires cognitive effort, and therefore processing goals may influence the way people give an answer. Potential problems that can undermine the validity of response have been shown to occur at different stages of the process (Warnecke et al, 1996).

The four-stage model of response and health status

The four-stage model of response formed the basis of the research carried out in the thesis. In the qualitative (cognitive) research, probes and analysis were structured according to the four stages, and in the quantitative research, the two central stages, retrieval and judgement were focused on, particularly in relation to any contextual influences on them. As with all cognitive models, the stages are, of course, only a simplified framework in which to investigate the complexity of the reality underlying the processes leading to a response. However, they provide a useful structure for research, as will be demonstrated below.

Although an item about health may appear straightforward to understand, *Comprehension* is potentially complex and involves the interpretation of the meaning of the particular words and phrases in the context of the item and the survey, and the respondents' own health experiences. Over a number of years, studies have shown that the meaning of questions intended by the researcher and the respondent can vary enormously, firstly because of variation in the interpretation of meaning, and second, if questions are difficult or ambiguous, respondents may modify the question so that they can answer it more easily (Belson, 1981; Donovan et al, 1993; Turner and Fiske, 1968). Without a common understanding, the validity of measurement is called into question.

The next two stages, *retrieval* and *judgement*, are central to the research conducted in the thesis, as they are related to the generation of an answer to the health status item. At these stages, it is hypothesised that information is retrieved or reconstituted, and answers formed, based on the cognitive strategies employed. Although all stages relate to the individual perspectives of respondents, retrieval and judgement are particularly interesting because they concern a range of individual contextual factors that relate to the strategy employed to generate an answer.

Retrieval requires information to be extracted from long-term memory. However, memory is not simply a reflection of events; for example, information may never have been stored in memory, rare experiences may not be recalled and common ones may no longer be distinct (Tourangeau et al, 2000). There is evidence that people may have only 'generic memories' for common events, such as multiple visits to a health centre, which may not be stored individually (Jobe and Mingay, 1989). If the information required is not stored in memory, is not accessible, or if the respondent is not motivated to seek the information, this stage cannot be carried out satisfactorily, leading to inference (Tourangeau et al, 2000).

Judgement calls for the formulation of an answer to the item, either involving recall of specific events or information, or the use of cognitive strategies, or heuristics, to construct an answer. Heuristic strategies may include estimates based on a general impression, or the recollection of summary information about the rate of occurrence of an event (Tourangeau, et al, 2000). Alternatively, the information available to the respondent may lead them to consider a representativeness heuristic (i.e. recollection of a salient medical condition), an availability heuristic (i.e. a recent health experience, easily recalled from memory), or an anchor-and-adjust heuristic (i.e. adjusting the answer in relation to the question and response categories provided) (Barofsky, 2003; Sudman et al, 1996).

Many health status items deal with issues which are not well defined, and for which the answer has to be constructed by respondents from their own perspective. For example, when asked about their "overall health", respondents' frames of reference may differ, depending on whether they include or exclude issues, such as particular health problems, when answering a question (Mallinson, 2002). A more subtle variation of this may result from temporal adaptation processes. It has been suggested that people with chronic illnesses may both adapt to their condition and recalibrate their judgements of good health and limitations, with a consequent effect on item interpretation and response (Allison et al, 1997; Heyink, 1993; Schwartz and Sprangers, 1999).

When considering a subjective health item for which there is no objective answer, respondents may employ comparison processes in order to generate a judgement. These could be comparisons with other people, such as those of the same age or with the same condition (Baron-Epel and Kaplan, 2001; Michalos, 1986; Garratt and Ruta, 1999). There may also be internal comparisons (such as current compared to past health), comparisons with an ideal, and comparisons involving expectations, hopes and aspirations (Calman, 1984; Cella and Tulsky, 1990; Krupinski, 1980).

However, there is also evidence that a heuristic strategy may be employed that is based on general self-perceptions rather than specific experiences or people. This has been found to be more common if the item is considered to be vague (Schwartz et al, 1987) and among those who have limited experience of the issue they are being asked about, such as the healthy compared to those with health problems (Rothman and Schwartz, 1998). Finally, the difficulty or easiness of the task has also been considered to vary according to the cognitive complexity of the judgement strategy, so that, in relation to health, those with serious problems or no problems at all might have a simpler task making a judgment than others, who would have to carry out a more complex cognitive assessment of the degree of limitation they experience (Jenkinson et al, 1996).

Finally, the *Response* itself involves the respondent mapping their answer to the available options, a task influenced by format, including problems resulting from categories that are unclear or inappropriate, and the respondent's willingness to provide an answer, which can be affected by the factors such as self-presentation and social desirability (Tourangeau et al, 2000).

In this account of the response process as a series of cognitive stages, there exists a continuum of "degrees of thoroughness" of question interpretation and response, where at one extreme lies *optimizing* and at the other *satisficing* (Krosnick et al, 1996). Optimizing, or optimally processing, relates to the "ideal" response process and involves careful processing at each stage, from comprehension through to retrieval, judgement and response. In contrast, satisficing involves less thorough processing, ranging from weak to strong satisficing. Weak satisficing involves executing all the stages of the response process, but less carefully and with possible biases. In strong satisficing, the respondent provides an apparently acceptable answer without going through the stages of retrieval or judgement. Motivation to optimize relates to the personal importance of the question for the respondent, beliefs about the usefulness of the consequence of responding to the questionnaire, and direct experience of the subject. Satisficing is thought to occur with more difficult tasks, lower respondent ability or motivation, and as a result of fatigue (Krosnick, 1999).

This section has outlined response process in terms of a *cognitive model*, comprising a series of interrelated stages. As illustrated in Figure 1.3, the generation of a response is likely to be a dynamic and iterative process, rather than a clear progression from one stage to the next. The descriptions of the retrieval and judgement stages, in particular, have demonstrated how individual influences may affect the strategy adopted to provide an answer. The next section expands on this, providing more detail on a range of potential contextual factors which have been shown to be associated with health status and which may therefore influence response process.

Contextual influences

Response processes are not isolated events; they exist within a context, including questionnaire design, the situation in which the measure is being administered, and individual factors. In terms of questionnaire design, subtle influences have been identified, with, for example, earlier health questions sensitising respondents to later questions (Bowling, 1995; Pavot and Diener 1993). Situational factors, such as room layout and approach taken by the researcher, as well as mode effects, in which the method of administration, such as interview versus self-completion questionnaire, have also been shown to influence response (Groves, 1989; Rust and Golombok, 1998). The current research, however, focuses primarily on individual factors that may influence the response process.

Individual characteristics

Several such individual factors have been identified, including health experiences (introduced earlier), culture, life events, personality, emotions and motivation (Barofsky, 2003; Furnham, 1983; Krosnick, 1999; Sadana et al, 2000; Totman, 1987; Warnecke et al, 1996). It has been argued that even with aggregated survey data, contextual information is required in order to understand the meaning of health question responses, for example indicating possible adaptation to health problems and explaining apparently paradoxical results (Mallison, 2002). In terms of a more complex understanding of response, contextual factors may co-occur, and operate by affecting objective health or aspects of the response process. The range of potential factors that may influence response, from discrete life experiences to enduring personal characteristics, have been summarised in two contrasting concepts: *bottom-up* processing, corresponding to the direct effects of events, versus *top-down* processing, describing the influence of personality and higher-level cognitive processes (Diener, 1984), and there is evidence for simultaneous and interrelated top-down and bottom-up influences on response (Brief et al, 1993; Headey, et al, 1991).

Attitude structure

An understanding of self-reported health perceptions may be informed by more general attitude theory and research (Sehulster, 1994), reflecting the individual context of the respondent, in terms of the particular cognitive framework they possess which will influence response to a question on health. *Self-health attitudes*, referred to in order to answer health status questions, may share many of the features of strong attitudes (Petty and Krosnick, 1995) or self-schemata (Markus, 1977, 1983): they are the product of direct experience, they may vary in intensity, and personal importance, knowledge, and the accessibility of information required in order to answer questions; and those with experience of health problems may analyse information more systematically than those with less experience (Rothman and Schwartz, 1998). This suggests that people with significant health experiences may have a more complex cognitive framework about health with which to answer questions, assign specific meanings to terms, and may also use different decision heuristics when answering.

Another important explanatory factor considered to relate to cognitive structure involves experience of health problems in other people, both in terms of the understanding of the effects of health problems this may bring, but also in relation to judgement processes. An important explanatory factor for health judgement is considered to be social comparisons with others (Calman, 1984; Michalos, 1986; Garratt and Ruta, 1999). Social comparisons may be of benefit in allowing people to evaluate their own health, and facilitate adaptation to illness (Heidrich and Ryff, 1993 and 1995). Studies on the relationship between self-rated health and mortality have found that self-assessments involve comparison with reference groups (Idler and Angel, 1990).

Many of the attitude strength dimensions that are thought to be relevant to the cognitive structure of health perceptions were included in the quantitative research conducted for this thesis. These include direct experience (general health, long-standing illness and recent contact with health services), intensity (health concerns), accessibility (frequency of thinking about health), identification (experience of health problems among others), and importance (perceived relevance of items for measuring health).

Personality factors

Considerable evidence exists in support of a taxonomy of personality that includes five broad trait dimensions: Extraversion, Neuroticism, Openness to experience, Agreeableness and Conscientiousness. These traits have been reported in a number of studies across diverse cultures, have been shown to relate to beliefs and predict behaviours and have been shown to be stable over a number of years (Gustavsson et al, 1997; Langston and Sykes, 1997; McRae and John, 1992). Two of the traits, in particular, have been associated with health outcomes and quality of life, in patient and community samples: Neuroticism and, to a lesser extent, Extraversion (Goodwin and Engstrom, 2002; Korotkov and Hannah, 2004), with poorer health among those with higher ratings of Neuroticism and better health reported among those with higher ratings of Extraversion.

In research among people with HIV, Extraversion and Neuroticism have been associated with self-reported health status, adherence to drug medication and disease progression (Penedo et al, 2003). However, personality may also be associated with reporting styles, such as social desirability and acquiescence (Diener et al, 1991). These reporting styles may be seen as strategies adopted by the respondent when faced with the task of completing the items contained within a questionnaire. Such strategies may or may not represent a deliberate attempt at impression management by the respondent (Anastasi, 1968; Krosnick, 1999). Loevinger (1966) describes these strategies as manifestations of a source of valid variance rather than error variance, reflecting test-taking attitudes and demonstrating the level of self-conceptualisation that the individual has achieved, while other research has shown that response may sometimes be mediated by a personality disposition, such as self-monitoring (that is, "self-observation and self-control guided by situational cues to social appropriateness") (Furnham, 1981). In addition, coping mechanisms and subsequent adjustment may relate to personality factors such as optimism and self-mastery (Andrews and Withey, 1976; Pearlin and Schooler, 1978), suggesting that personality may be related to adaptation to illness. As

Neuroticism and Extraversion are the traits with the most evidence linking them to health status, they were investigated in the quantitative study.

Age

A number of sociodemographic factors have been associated with reported health status. There is consistent evidence that self-reported physical functioning declines with age while mental health increases in older age groups (Franks et al, 2003). This also applies to people with HIV (Campsmith et al, 2003; Mannheimer et al, 2005; Murri et al, 2003). However, a study of differential item functioning (DIF) in the SF-12, which statistically adjusted for differences in scale response patterns between age groups, reported that mental health perceptions actually declined with age (Fleishman and Lawrence, 2003), supporting a suggestion that older people have lower expectations and are more likely to give a positive assessment of their health than would be justified by any objective judgement (Blaxter, 1990).

Gender

Women generally report poorer health status than men in the general population (Franks et al, 2003), despite the fact that women have longer life expectancies (Barford et al, 2006). The gender effect also applies among people with HIV (Campsmith et al, 2003; Mannheimer et al, 2005). This finding has been discussed in terms of response style, different perceptions, and more non life-threatening chronic illness suffered by women (Franks et al, 2003). However, in DIF analyses, with adjustment for differences in patterns of response, the gender effect on self-rated physical and mental health in the SF-12 did not substantially alter, suggesting either constant bias affecting all scale items equally, or a meaningful difference in subjective health ratings between males and females (Fleishman and Lawrence, 2003).

Education

Generally, those with lower levels of education report poorer subjective health perceptions (general population: Franks et al, 2003; people with HIV: Campsmith et al, 2003; Murri et al, 2003). In addition, there is evidence that morbidity and mortality are also higher among those who have less education (Franks et al, 2003). However, it could be that self-reports are influenced by cognitive ability, so that, for example, those with lower cognitive ability are more likely to satisfice when answering (Krosnick, 1999), or subcultural factors, including lower expectations about their health among those from lower social classes (Blaxter, 1990; Calnan, 1987).

Ethnicity

In some studies, being black has been associated with lower health status and increased mortality (Franks et al, 2003). Among people with HIV, initial results showing worse health status among black and Hispanic respondents were attenuated in multivariate analyses (Campsmith et al, 2003). However, it has also been reported that black respondents have higher mental health scores than others, although these differences were reduced when results were adjusted for DIF between ethnic groups, with a suggestion that the findings may reflect cultural norms (Fleishman and Lawrence, 2003).

Health behaviours

Perceptions of poorer health have been reported for cigarette smokers (Schmitz et al, 2003; Wilson et al, 1999) and problem drinkers (Friedman et al, 1999). This is likely to be related to the wide variety of illnesses caused by these behaviours (Department of Health, 2003 and 2004), although there appears to be no evidence showing whether health behaviours influence the response process, either independently or in relation to other contextual influences.

Work roles

Being out of the labour force has been consistently associated with poor health status, with some longitudinal evidence of a decline in health following unemployment (Bartley, 1994; Mathers and Schofield, 1998). This relationship has been replicated among people with HIV, with poorest subjective health reported among unemployed respondents (Murri et al, 2003).

In conclusion, Part One has identified the theoretical and methodology framework in which the research for the thesis was conducted. Health status measures are now widely used in many areas of research and practice. However, there are concerns about the use of standardised measures for the assessment of an issue as complex and personal as perceptions about health. The long-term development of a particular generic instrument, the SF-12v2, the health status measure at the centre of the research included in this thesis, was documented, including a discussion of reliability and validity. The remaining sections referred either directly or indirectly to the potential influences on health status response. Chronic illnesses may influence both health and cognitive processes leading to a health status judgement. The evidence in terms of HIV indicates that poorer health status is associated with being HIV positive and disease progression, although there is no published research on the influence of HIV on response processes. A cognitive model of health status response was outlined, demonstrating the potential complexity of the cognitive tasks involved in generating a health status response. The framework included four stages of response, comprehension, retrieval, judgement and response and, in addition, various strategies for response formulation. Finally, a range of potential contextual influences on the response process and final response were described.

Parts Two and Three that follow contain the substantive research, opening with a qualitative (cognitive) assessment of SF-12v2 response processes for university and HIV participants, relating to the four stage cognitive model of response detailed earlier.

Part Two: A qualitative (cognitive) study on response processes for the SF-12v2 health status measure among university and HIV participants

Overview

Response processes were investigated using cognitive interviewing (think-aloud) and self-completion techniques (see Figure 2.1 for a diagrammatic representation of the study design). The findings revealed that many SF-12v2 items were perceived to be vague, and item interpretations differed between participants. Most participants employed a range of general and specific strategies in order to produce their answers. Those with health problems were more likely to have difficulty answering the SF-12v2, and were more likely to relate their response to specific experiences. Despite some evidence for the influence of adaptation and satisficing on results, those with health problems generally gave answers consistent with reported limitations. A comparison of the results indicated that, while retrospective interviewing with probes provided the more detailed information, other techniques could identify similar, and non-contradictory findings, contributing to a debate about method effects in cognitive research.

2.1 Introduction

As a first stage in exploring the SF-12v2 response processes reported by university and HIV participants, a small-scale qualitative (cognitive) study was carried out. The principal aim of the study was to address the question:

How do people understand and interpret the questions they are completing? That is, what are the conscious processes involved in answering health status items?

The study employed two distinct qualitative (cognitive) approaches to gain a better understanding of response processes (See Figure 2.1). One involved a face-to-face interview, during which participants verbalised their thoughts while completing SF-12v2, followed by answering retrospective probes about aspects of the response process, from comprehension to response. The other involved administering a semi-structured self-completion questionnaire, including questions on item comprehension and the formation of an answer. This section introduces these approaches and the techniques used.

Cognitive techniques for assessing response processes: an overview

A range of cognitive techniques exists, including one-to-one cognitive interviews (including undirected prospective think-aloud and retrospective probing), focus group discussions, card sorts, vignettes, measurement of response latency, and confidence ratings (Campanelli, 1997; Collins, 2003). Most cognitive testing is carried out during the development of survey instruments, although they have also been used to study fundamental cognitive processes (Collins, 2003; McColl, 2003; Willis, 1994). Cognitive techniques are rooted in the assumption that, using the appropriate methods, it is possible to gain an insight into the cognitive functions employed by respondents during the survey response process.

Cognitive interviews

It has been claimed that cognitive interviewing can provide insights into what respondents are thinking when selecting a response (Barofsky, 2003). Cognitive interviews cover a range of approaches, from the verbalisation of thoughts by the participant during a survey, with little or no intervention from the researcher (prospective think-aloud), through to a structured interview, in which the researcher asks predefined probe questions, either before or after the completion of the survey (Campanelli, 1997).

Think-aloud

For the think-aloud technique, participants are asked to give a verbal account of their thoughts, either concurrently, as they are answering a question, or retrospectively, immediately after they have answered. Think-aloud has an open format that allows the respondent to articulate their thoughts without the use of probes, therefore also removing interviewer bias (Willis, 1994). However, the dependence on the participant verbalising their thought processes means that this technique is particularly demanding, and training is required before think-aloud can be carried out. The ability of respondents to articulate their thought processes as they complete a survey has been questioned (Drennan, 2003). Although it has been claimed that thinking-aloud merely slows down cognitive processes rather than altering them, the task of thinking aloud may sometimes interact with the processes being reported (Conrad et al, 1999). As attitudes may be influenced by the context in which they are expressed, it has been suggested that they may be changed by the act of thinking aloud (Wilson et al, 1995). More specifically, the response process may be complicated by the additional cognitive effort required during the interview to the point where the results obtained are distorted (Barofsky, 2003; Drennan, 2003; Willis, 1994).

Furthermore, questions may not be interpreted identically in both conventional completion and think-aloud conditions. Being told to think about the task may change the way it is undertaken, so that, for example, the respondent may take more care in answering the question (optimizing), since the nature of the cognitive interview has given the question greater significance than it would otherwise have (Drennan, 2003;

Willis, 1994). This attitude change effect has been found to apply more to weak attitudes than to strong ones because these attitudes are less central to the respondent and are held less securely than is the case with attitudes based on knowledge and experience (Wilson et al, 1989). Verbalising tasks require training and may nevertheless favour the articulate (Collins, 2003; Willis, 1994). Concerns have also been raised that cognitive interviews do not represent a real survey situation (Campanelli, 1997), although artificiality is a criticism that could be levelled at any survey (Drennan, 2003).

Cognitive probes

Probes are questions designed to elicit information from participants about particular parts of the response process. They can be asked *prospectively*, during survey completion, or *retrospectively*, once the survey has been finished. Probe questions are either written beforehand or developed during the interview, as required. They can relate to specific stages of the response process or be more general questions about the overall response (Collins, 2003). Probe questions may be used to enquire about any stage of the response process. For example, asking about the meaning of particular words or phrases (comprehension stage) to the respondent, about the time period the respondent was thinking when answering a question (retrieval and judgement stages), or how one response category differs from others (response stage). More general questions may not relate to any single stage of processing, such as asking how easy a question was to answer, or how the respondent went about answering the question. Interviewer control can be seen as a positive aspect of probing, allowing the interviewer to probe areas of interest or ambiguity (Conrad, 1999). In addition, unlike the think-aloud technique, a participant answering probe questions does not need to be trained in order to take part in an interview (Willis, 1994).

However, concerns have been raised about the use of probes. Asking questions about the response process may introduce bias by asking for information to which the respondent does not have access, leading them to think about issues that they would not otherwise have considered, or, when carried out *prospectively*, influencing later responses (Campanelli, 1997; Conrad et al, 1999; Willis, 1994). In addition, memory limitations may affect responses when *retrospective* probing is used (Campanelli, 1997).

Self-completion

The self-completion approach adopted was based on work carried out on the response processes used in completion of personality measures (Gordon and Holden, 1996; Turner and Fiske, 1968). The first instrument using this approach was the Process Interrogation Questionnaire (PIQ). The purpose of the PIQ was reported to be, '...to elicit information regarding the processes experienced by subjects after reading and before answering the personality questionnaire items presented to them.' (Turner and Fiske, 1968). Data collection was by a paper-based self-completion questionnaire designed to identify response process used by respondents, focusing mainly on keyword comprehension and overall response process. Since the original PIQ study, other research has used similar process-elicitation tools. Some studies have used the information provided on the open process questions to identify themes, and develop closed lists of response strategies for personality measures. These closed lists have then been included in quantitative studies of the response process, with respondents asked to indicate the most important response strategies adopted for particular questions from a fixed series of options (Gordon and Holden, 1996). However, it was decided to adopt an open-ended approach in the current research in order to gain a better understanding of the response process as it applies to health status questions.

Although there is some comparative evidence that administering cognitive testing in either a face-to-face interview or self-completion questionnaire can provide useful information, interviews appear to provide more complete and detailed information, possibly because interviews allow non-verbal prompting, restatements and clarification of probes, while greater effort is required to write a response to a question in a self-completion questionnaire (Davis et al, 1995).

Given the potential problems outlined with any cognitive method, the study was not limited to a single approach. Instead, some participants undertook both think-aloud and probing, while others completed the self-completion questionnaire. The application of these techniques in the qualitative (cognitive) study is outlined in more detail in the Methods section that follows.

2.2 Methods

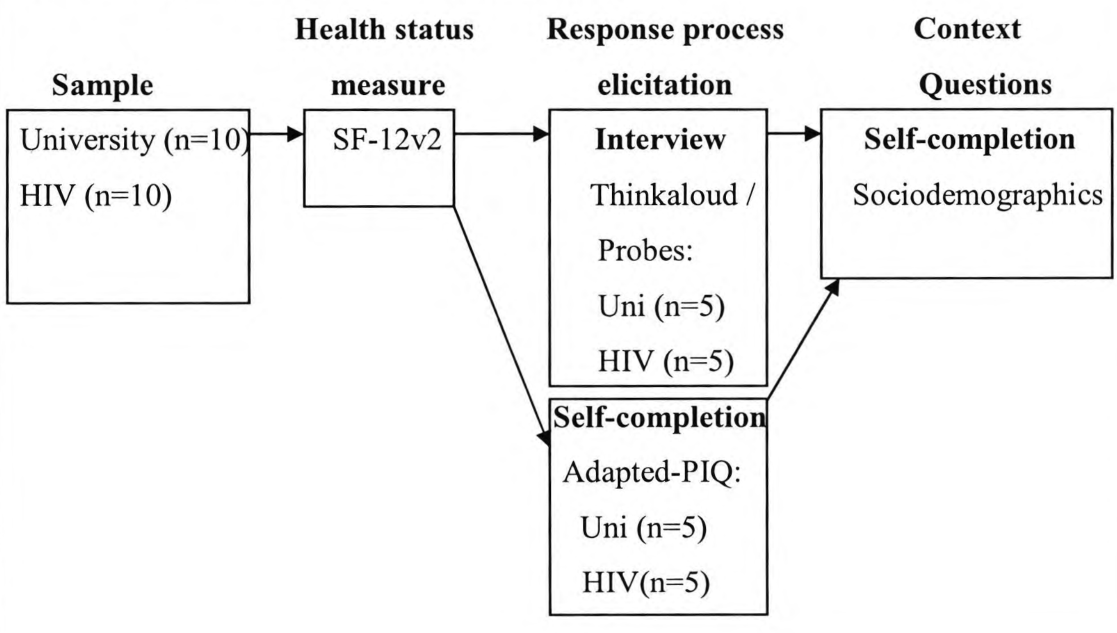
This section outlines the methods used in the qualitative (cognitive) study, including an introduction to the methodology employed, the development of the materials, piloting, ethical considerations, recruitment data collection and qualitative analyses.

Design

The aim of this study was to elicit information about response processes involved in completion of the SF-12v2 health status measure from both a nominally healthy sample (university students and staff) and a health problem group (people with HIV), using different cognitive techniques in order both to understand the influence of method on results and to overcome possible biases inherent in any single method, highlighted in the previous section. Figure 2.1 demonstrates the design of the study.

Generally, there is wide variation in the numbers of participants recruited to take part in any qualitative research, and sample size typically depends on the aims of the study and the number of participants required in order to achieve an understanding of the research area (Kvale, 1996). Similarly, in relation to cognitive interviews, sample sizes vary, but as few as four can be sufficient (Willis, 1994), while other studies have been reported with samples of six (Forsyth et al, 1992). In this study, there were ten participants in each sample group. All participants completed both the SF-12v2 health status measure and a series of sociodemographic contextual items. A target of five participants from each sample for a particular method seemed to be an acceptable number for a small-scale study. By random allocation, one half of each sample took part in a cognitive interview while the other half completed the self-completion instrument.

Figure 2.1: Qualitative (cognitive) study design



Development of materials and pilot work

Before the study was undertaken, materials and procedures were subjected to expert review and small-scale pilots. Both interview and self-completion techniques were based on existing materials and testing procedures (Collins, 2003; Schechter, 1993; Turner and Fiske, 1968), and therefore piloting was limited.

Cognitive interview schedule

The researcher, members of the PhD steering group (Appendix 1), comprising experienced health and social researchers, and an expert in cognitive interviewing from the National Centre for Social Research, London, reviewed the topics and wording of the probes for each SF-12v2 question. The probes were based on existing questions designed to elicit information about cognitive processes (Collins, 2003; Schechter, 1993), adopting the four-stage model of response outlined in the previous section, including comprehension, retrieval, judgement and response (Tourangeau et al, 2000), along with more general probes. The retrospective probing interview was semi-structured, with questions and topics adapted, supplemented or removed during the course of the interview in relation to the answers received during the think-aloud component of the session. Following the review of the probes, the procedure and probes were tested for sense with friends, colleagues, and people with HIV working in the Chelsea & Westminster Hospital. In total four people took part in this pilot. Both procedure and probes functioned well and were used in the main cognitive interview. The final version of the probes is included in Table 2.1. The probes were available for use with each SF-12v2 item, and the italicised words and phrases included in the square brackets refer to specific terms included in each SF-12v2 item. The full version of the interview schedule, including a script to guide the researcher through the task instructions, SF-12v2 questions, and potential probe questions are included in Appendix 2.

Table 2.1: Think-aloud probe questions

COMPREHENSION
• What do you think this question is getting at?
• What does [*word or phrase*] mean to you?
RETRIEVAL/JUDGEMENT
• What time period were you thinking about when you answered this question? (from when to when?)
• How did you work out [*time period*]?
• When answering the question, did you think about your future health in any way? (general health question)
RESPONSE
• You gave the answer [*response*]. How is that different from [lo*wer response category*]. And [*higher response category*]?
• Can you tell me the difference between [*one response cate*gory] and [*second response category*]?
GENERAL
• How did you go about answering this question? What sorts of things were you thinking of while you were answering?
• Was it easy or difficult to answer? Why?
• Was there a time in your life when you would have described your health as [*lower response categories*]. And [*higher response categories*]? What happened that has changed how you feel about your health? (general health question)
• Do you have any current medical condition that is affecting your health? Have you had it recently? (general health question)
• After talking about it, would you still describe your health as [*response*]? (general health question)

Self-completion instrument

In the original Process Interrogation Questionnaire (PIQ) (Turner and Fiske, 1968), upon which the current instrument is based, each page of the questionnaire included an individual personality item followed by six questions aimed to assess: (a) how the respondent decided to answer the item in the way they did; (b) whether they focused on any word or phrase in the item; (c) how easy or difficult they found it to answer the item; (d) whether they based their answer on a particular situation/instance, several situations/instances, or a general picture of themselves; (e) the amount of experience they had of the situation suggested in the item; (f) and, finally, they were asked to write one sentence, in their words, stating what the item meant to them. Half of the probe questions required only a written response (a, b, f), one had closed and open-response elements (c), and two were closed response only (d, e).

The researcher and the researcher's PhD Steering group reviewed these questions, in order to make sure they were suitable for use with health questions and simple to understand and complete. Five of the six question areas were retained. PIQ question 'e' (experience with the situation suggested by the statement) was excluded from the revised PIQ, since it was judged not to relate straightforwardly to the particular items used in the SF-12v2. Wording changes are included in Table 2.2. The revised questions were piloted for sense with five people who had not worked on the cognitive interview pilot. As in the earlier pilot, these were friends, colleagues and people with HIV working in the Chelsea & Westminster Hospital. The pilot indicated that the questions and format worked well. The final adapted-PIQ was presented in a self-completion questionnaire booklet which was thirteen pages long, including a front page containing the title, instructions and the identification number for the participant, and a single page for each of the twelve SF-12v2 items, followed by the five adapted-PIQ questions (Appendix 3).

Table 2.2.1: Adaptations to the Process Interrogation Questionnaire

PIQ	ADAPTED PIQ
a. State briefly how you decided to answer the question in the way you did. That is, as best you can, say what process went on during the period after reading the statement and before circling T or F [the response format of the personality items]	a. State briefly how you decided to answer the question in the way you did. That is, as best you can, say what you thought during the period between first seeing the question and selecting your answer. ➢ This was amended, with 'process went on' replaced by the less technical, 'you thought'. In addition, 'after reading the statement and before circling T or F' was replaced by, 'between first seeing the question and selecting your answer'. The revision was suggested as a more natural and understandable description of the overall response process. Throughout the adapted PIQ, 'question' replaced 'statement' in accordance with the format of the SF-12v2 health status measure.
b. Did you focus on any word or phrase in the statement? If so, which one? Did this word or phrase have some special meaning for you or present you with some problem? If you could, would you: Eliminate the word or phrase? Change it? If so, *how* would you change it? If you would do either (1) or (2), would your answer to the statement change?	b. Did you focus on any word(s) or phrase(s) in the question? If so, which one(s)? And why did you focus on the word(s) or phrase(s)? ➢ This was changed to permit more than one word or phrase to be a focus. The wording was simplified so as to concentrate on reasons for focusing on words or phrases.
c. Did you find it *very easy, fairly easy, fairly hard,* or *very hard* to make a decision about your answer? Can you say why this was?	c. Did you find it *very easy, fairly easy, fairly hard,* or *very hard* to make the decision about your answer? Can you say why this was? ➢ 'A decision' was changed to 'the decision' to signify that the question is concerned with the answer selected.

Table 2.2.2: Adaptations to the Process Interrogation Questionnaire

PIQ	ADAPTED PIQ
d. Did you feel that, in answering, you used a *particular situation or instance, several situations or instances,* or did you think of the answer as reflecting a *general picture* of yourself?	d. When answering the question, did you feel that you had a *particular situation or instance* in mind, *several situations or instances,* or did you think of the answer as reflecting a *general picture* of yourself? ➢ The word order of the question was modified, with 'When answering the question' moved to the start of the question. The PIQ asked whether, 'you used… a situation', whereas the adapted version asked whether, 'you had…a situation… in mind'. Both changes were made to simplify the wording and clarify the question task.
e. Would you say that you have had *much, some, or little or no* experience with the situation suggested by the statement.	➢ Excluded (see explanation)
f. Write one sentence in your own (and different) words stating what the sentence meant to you	e. Write one sentence in your own words explaining what the question meant to you. ➢ The question was adapted, removing '(and different) words' since this was considered to unnecessarily complicate the question. 'Explaining' replaced 'stating' because it was felt to be a more likely to elicit a detailed response.

Sociodemographic questionnaire

A one-page questionnaire, 'About Yourself', containing three explanatory sociodemographic questions: age, sex and employment status was devised (Appendix 4). The questionnaire was tested for sense with both sets of pilot participants.

Ethical considerations

Study participants were expected to provide personal information about themselves and their health. Therefore, all research materials and methods were submitted to the appropriate ethics committees for scrutiny before any research was undertaken.

City University sample

An application was made to the City University ethics committee in the summer of 2003 (see Appendix 5 for application headings). The committee reviewed the application and, given the focus of the research on participants' health, requested more evidence that participants would have access to information and support, should it be required. The research was discussed with representatives of the National Centre for Social Research and the City University Health and Counselling Services. As a result, provision was made for potential participants' distress, and both the subject matter and the availability of the counselling service were stated in the written explanatory statement provided to participants and reiterated to each participant verbally by the researcher during the study preliminaries (Appendix 6). A separate 'Helplines' sheet was composed, including contact details for the university health and counselling services, along with other external support services (Appendix 7). All participants completed a consent form before the study began (Appendix 8). The revised ethics application was resubmitted to the University and accepted in the autumn of 2003.

HIV sample

The research proposal was also submitted to the ethics committees of the Riverside Health Authority Research Ethics Committee (responsible for Chelsea & Westminster Hospital) in December 2003, under consideration by Chairman's Action (see Appendix 9 for key application headings). The committee reviewed the application and requested additions to the participant information sheet: an invitation paragraph, explaining that the participant should think carefully about the research before agreeing to take part; and the inclusion of more information on the storage, use, access and destruction of the tapes and transcripts (Appendix 10). As with the university sample, participants were also given a Helplines information sheet (Appendix 11). Finally, the use of a new, standard consent form was requested (Appendix 12). These changes were made and the application resubmitted. This was accepted in the spring of 2004.

Confidentiality and anonymity

Given the sensitive nature of the research on health, and the identification of people with HIV, it was very important that anonymity and confidentiality was to be preserved during all aspects of the research process. For organisational purposes, names were used to identify participants. Each participant was allocated an identification number that is used on all written materials, tape recordings, and computer files. The researcher was the only person to know the code for identification numbers and this list was stored in a locked filing cabinet. Names of participants were written on the consent forms, but the forms did not include participants' allocated identification numbers and were stored securely separately from the identification code. Data from paper questionnaires were input by the researcher on password-protected computer equipment. No names or addresses of participants were stored on computer. Original questionnaires were stored in a locked filing cabinet. Think-aloud interview tapes were transcribed onto password-protected computer equipment by the researcher. Tapes were stored in a locked filing cabinet. The computer data, tapes and questionnaires will be destroyed once the results of the study have been fully documented.

Recruitment

Both samples were purposive and selected not to be representative of the populations from which they were drawn, but, rather, to investigate in detail aspects of response process among participants from the two samples.

University sample

A poster advertisement was drawn up including basic details about study aims and methodology, inclusion criteria and contact details (Appendix 13). This was posted on noticeboards within Departments of the School of Social Sciences. In addition, after a brief explanation of the study by the researcher, the poster was handed out to students in lectures during the graduate programme in Social Research Methods. Finally, the contents of the poster were included in an email sent to three email lists: students enrolled on the City University MSc in Advanced Social Research Methods, members of the City University research student society, and the list of City University research students held by the university research student administrator. Data collection took place between December 2003 and January 2004.

HIV sample

HIV/AIDS patients were recruited from the HIV/GUM Directorate at Chelsea and Westminster Hospital. Recruitment was by word-of-mouth among the patient volunteer service, a group of HIV/AIDS patients who carry out voluntary work in the Directorate. Data collection took place in May 2004.

Data collection

Participants were randomly assigned to undertake either the adapted-PIQ or the cognitive interview approach. Half of the participants received the adapted-PIQ, while the other half undertook the interview procedures. Before recruitment began, the identification numbers used in each sample were randomly divided between the two approaches, so that five numbers were assigned to the adapted-PIQ approach and five to the think-aloud approach. Identification numbers were sequentially allocated to participants in order of recruitment. That is, when a participant contacted the researcher, they were given the next available identification number in the particular sample to which they were recruited, which had already been assigned to one approach.

Participants from both samples took part in a single one-to-one session with the researcher, lasting approximately 45 minutes.

University sample

Participants from City University were interviewed in the researcher's office at the university. Each participant was greeted by the researcher at the main entrance to the building and taken to the office, where they were invited to sit at a desk, followed by the researcher. Participants were presented with a written explanatory statement and a sheet containing Helplines available should the participant wish to ask for advice or support following the study session. The documentation was discussed and any questions raised by the participant were answered by the researcher. Afterwards, the participant and the researcher signed an informed consent form.

HIV sample

HIV participants were interviewed in a private office in the HIV/GUM Directorate, at Chelsea & Westminster Hospital. The participant was collected from the entrance to the Directorate of HIV/GUM building and taken to the office. An identical procedure to that carried out with the university participants involved sitting at a table, going through the information sheet and Helplines and finally signing the consent form after any participant questions and comments had been addressed.

In both samples, either the adapted PIQ or cognitive methods procedures followed.

Adapted PIQ / SF-12v2 procedure

The researcher explained the task to the participant and highlighted that the instructions were also contained on the front of the questionnaire booklet. The booklet was handed to the participant, along with a pen and some spare sheets of paper. The participant was left to complete the questionnaire, although the researcher remained in the same room in order to answer any questions or problems that might arise. Once the adapted-PIQ was completed, the researcher asked the participant to complete a separate sheet containing sociodemographic questions ('About you'). That concluded the study session. Afterwards, the researcher debriefed the participant, providing additional details about the study and fielding any comments or questions.

Cognitive interview / SF-12v2 procedure

The researcher read from a script, explaining the task to the participant. Training is required in the think-aloud technique as it requires some practice before it can be performed fluently. The researcher went through three examples, unrelated to health, that are regularly used in think-aloud research (included in Appendices 2 and 14). When the participant was familiar with the think-aloud procedure, the researcher explained that the study task would follow and confirmed that the participant was still comfortable having the remainder of the session tape recorded. The participant was provided with a pen and asked to tick a response box for each of the 12 questions. The tape recorder was started and the participant asked to begin. The researcher placed the printed SF-12v2 questions in front of the participant, one item at a time (Appendix 14). The general prompt, "what are you thinking?" was used if the participant was silent for more than a few seconds. On completion of the think-aloud procedure, the researcher asked whether the participant would be willing to answer some additional questions about the answers they had given. The researcher presented each of the twelve questions to the participant again. The researcher asked the probe questions, with additional follow-up questions for clarification. After all twelve questions had been reviewed, the tape recorder was switched off. The researcher asked the respondent to complete the separate questionnaire containing sociodemographic questions ('About you'). That concluded the study session. Afterwards, the researcher debriefed the participant, providing additional details about the study and fielding any comments or questions.

Analysis

All tape-recorded data were transcribed verbatim into Microsoft Word wordprocessor documents. Written information from the adapted-PIQ was also typed into a word-processing package for ease of manipulation. Responses to the sociodemographic questionnaire were entered into an MS Excel spreadsheet. All data were stored by the participant identifier.

The "Framework" approach, developed in the National Centre for Social Research (Ritchie and Lewis, 2003) was used for the analysis of data collected using both cognitive interview and adapted-PIQ. "Framework" involves the development of a rigorous "thematic framework", under which data are sorted, synthesised and compared. A key requirement of this approach is a transparent connection between data and higher-level themes, so that themes can be related to the original data, and comparisons made across themes and participants. This is possible because "Framework" is based on organising data into a grid, with a column representing a theme and each row a respondent, with a cell in the grid relating to answers provided by a respondent, referring to a specific theme. For the purposes of structuring the material collected, all data were grouped according to the four-stage response process model described earlier, including comprehension, retrieval, judgement and response (Tourangeau et al, 2000).

Major themes relating to the response process were described and illustrated with examples from the data. Although the model adopted is based on the four categories, for the purposes of analysis and reporting, retrieval and judgement stages were combined since the nature of the response process makes it unlikely that participants will be able to distinguish between these two stages, or to clearly differentiate them in the cognitive data (Conrad et al, 1999).

Initially, typed questionnaires and interview transcripts were examined with the intention of generating a coded index of the themes for each stage of the response process. After this, the questionnaires and transcripts were reviewed and coded according to the original index categories. At this stage, the indexing categories were refined. The list of codes used for classification is included in Appendix 15. Once the indexing categories were stable and comprehensive, data from each case were synthesised within the appropriate cell of an MS Excel spreadsheet, so that no data about the response process remained unclassified. The grid was then interrogated so that the main themes describing aspects of the response process were identified, described and reported, with appropriate illustrations from the data.

2.3 Results

Results were first analysed separately according to sample (that is university or HIV) and method (cognitive interview, in terms of think-aloud and probes, and also self-completion questionnaire responses). However, as the findings were found to be similar, the combined results are reported. Before presenting findings about the response process, a simple sociodemographic description of both samples is shown for each method. This is followed by the results of the "Framework" analyses, divided according to SF-12v2 scale and response process components, with quotes included to illustrate themes that were identified. Finally, the results are discussed in terms of methodology and sample.

Sample descriptives

Since the study was qualitative in nature, university and HIV samples were very small, with only ten participants in each, further divided into four groups according to method. Despite this, some descriptive sociodemographic data are given to provide a context for the substantive findings (Tables 2.3 and 2.4). The university sample comprised more females than males, in contrast to the exclusively male HIV sample, and was also generally younger. Most participants in both samples combined two or more work or domestic roles. Nine out of the ten university participants were students. Four HIV participants were classed as long-term sick (always combined with another activity), and another unemployed.

Table 2.3: Sample sociodemographics (Cognitive interview)

ID	Age	Gender	Roles
University sample			
2	22	Female	Studies full-time
5	27	Female	Employed part-time, Studies full-time
6	41	Female	Employed part-time, Studies full-time
7	24	Female	Studies full-time
8	35	Male	Employed full-time, Studies part-time
HIV sample			
11	49	Male	Longterm sick, Volunteer work
12	49	Male	Longterm sick, Volunteer work
15	36	Male	Employed full-time, Studies part-time
18	36	Male	Volunteer work
19	51	Male	Volunteer work

Table 2.4: Sociodemographics and SF-12v2 scores (adapted-PIQ)

ID	Age	Gender	Roles
University sample			
1	31	Female	Studies full-time, Carer
3	30	Male	Employed full-time, Studies part-time
4	31	Male	Employed full-time
9	25	Female	Employed full-time, Studies part-time
10	37	Female	Employed part-time, Studies full-time, Carer
HIV sample			
13	49	Male	Long-term sick, Voluntary work
14	56	Male	Long-term sick, Early retirement
16	34	Male	Unemployed
17	59	Male	Employed full-time
20	35	Male	Employed full-time, Studies part-time

Analysis of themes by scales

General Health Scale

Single item scale:

GH01: In general, would you say your health is: Excellent / Very Good / Good / Fair / Poor

Summary

- The general health scale was imprecise in terms of health definition and time scale, allowing participants to consider a range of issues when answering.
- The main response strategies reported were: a general perception, recall of specific experiences and comparisons with other people. The general perception strategy included undefined self-reflection and internal comparisons with the past or an ideal.
- Disease issues dominated many of the HIV participants' comments.

Comprehension

The question wording was vague, leading to diverse interpretations of meaning, and the definition of the *key term* 'health' varied. Without any examples, it could encompass many aspects of well being and functioning, physical fitness, health history, or health problems.

In both samples, health was commonly regarded as a multidimensional concept:

"... physical health, mental health, ability to do things, you know... Everything, sleeping" [University sample: 2: Cognitive interview: Retrospective]

HIV participants similarly referred to a broad and varied definition of health, including medical opinion, energy, and non-health issues, such as fitness and humour. Once again an overarching definition of health was provided:

"I don't just think about health in terms of physical, erm, I'm not just thinking in terms of physical health, but I'm thinking about mental health" [HIV sample: 19: Cognitive Interview: Prospective]

"How do I feel overall, not focusing on specifics" [HIV sample: 14: PIQ]

However, a more limited definition could also be adopted, such as physical health and presence or absence of serious physical illnesses:

"I'd take it to mean something like... any existing conditions, illnesses, or anything like that... I generally just thought of physical health rather than anything else" [University sample: 5: Cognitive Interview: Retrospective]

And, more narrowly still, physical fitness and functioning:

"I was defining it as physical – as: am I fit, can I run up and down stairs, will I get puffed out, would I... ache, I was kind of thinking about those things" [University sample: 8: Cognitive Interview: Retrospective]

The openness of the item allowed those participants with HIV to include a consideration of their condition within a definition of health:

"My overall health, really, erm, in particular I was thinking about my HIV" [HIV sample: 15: Cognitive Interview: Retrospective]

The General Health item does not indicate a *timeframe*. Therefore, it was possible to interpret the item over any duration, or none at all, and considerable variation resulted. Most participants indicated that they had some sense of time in mind when answering, based on a series of events, or a less clearly-defined duration.

This period varied from 'now':

"When I look at that, I'm just saying, 'my health is now good', not how it was 6 months ago, or last year when I had 'flu, or, erm, when I started taking this combination, erm, I felt absolutely dreadful, so – this was a year ago – the er... it's now I'm thinking of, and I'm saying that's good now. That's the way that strikes me." [HIV sample: 19: Cognitive Interview: Retrospective]

To recent events:

"I'd say, yeah, ok, it is based on recent experiences" [University sample: 7: Cognitive Interview: Retrospective]

To events over a longer duration:

"… I think it was, kind of, a span [of events] of an overall amalgamation of everything together" [University sample: 8: Cognitive Interview: Retrospective]

Or a period of time:

"… I would say I was thinking about [health] in the longer-term period rather than in specifically a recent period" [University sample: 6: Cognitive Interview: Retrospective]

However, a general perception, not obviously related to time or events, was also mentioned:

"I was thinking sort of like health in general" [University sample: 5: Cognitive Interview: Retrospective]

Among people with HIV, time could include the period before and since HIV diagnosis:

"Since being diagnosed I haven't, I've been hospitalised once, y'know… in general, my health has been very good." [HIV sample: 15: Cognitive Interview: Retrospective]

"I remember what my health was like and I know what it feels like now." [HIV sample: 13: PIQ]

Retrieval / Judgement

Simple and complex comparison processes were evident in the responses provided, including internal and external references.

The main strategy adopted was based on *general perceptions*, indicating the use of strategies focused on the self. Under this broad heading, were an undefined *self-reflection*, and *comparisons* with the past or an ideal.

For example, *self-reflection* was here combined with a comparison with an ideal, providing a *ceiling* on the value that the participant could place on their own health:

"I'm fairly fit I think, quite healthy... In general, I'd say my health was "very good"; I wouldn't say it was "excellent" because I think one can improve, I think, we can exercise a bit more and things..." (rated 'very good') *[University sample: 8: Cognitive Interview: Retrospective]*

A *comparison* process could involve *discounting* a discrete health problem against a *general perception* of good health. In this example, a positive rating of health was followed by the revelation of a longstanding medical condition, suggesting a process of *adaptation* to limitations:

"I ended up going for "very good" because, overall, I didn't think I had too many complicated issues in my life... [medical condition] affects my ability to do some things... [it] is a muscular weakness in the legs which affects my ability to walk and balance. So, that affects some things, like stairs and... erm... balance, but other than that those things you can adjust around, so... er... other than that, no, I haven't got anything affecting me" (rated 'very good') *[University sample: 5: Cognitive Interview: Retrospective]*

In the HIV sample, judgement was dominated by disease considerations. In the following examples, the participants reported *a general perception* but indicated *adaptation* to disease limitations:

"At the moment I feel fine physically and mentally, in relation to HIV, and I think that's the thing as well, that has to be always remembered." (rated 'good') *[HIV sample:19: Cognitive Interview: Prospective].*

"I considered my feelings within the health parameters that I am in" (rated 'fair') *[HIV sample: 14: PIQ]*

Specific situations were only referred to by HIV participants, particularly in relation to experiences with HIV, including medical issues:

"I feel good in myself, but judging by the medical results, I would say my health was 'fair', mainly because I had a CD4 trial and my CD4's dropped and my viral load has shot up and so I've got to go on new drugs, so that's the reason why I'm saying 'fair', but generally I feel good in myself, no problems." (rated 'fair') *[HIV sample: 12: Cognitive Interview: Prospective].*

Finally, some participants revealed that they had based their judgement on a *comparison with others*. However, the particular comparison group differed, thereby influencing the judgement made.

For example, *comparison with others* of the same age, who were healthier (an upward comparison), which led to a *ceiling* on the response:

"I straight-away think, compared to other people... I thought, overall, people I know, my age, you know, doing the same thing... I didn't put excellent because I think there are healthier people, and I can think of examples of them, of people who probably drink less than me and sleep 9 hours a day, you know, exercise regularly" (rated 'very good') *[University sample: 2 : Cognitive Interview: Retrospective]*

There was also evidence of *comparison with others* among HIV participants, although, in the following example, the strategy was used in combination with a general perception of life with HIV, indicating a focus on disease issues whether the strategy was internal or external:

"In relation to other people I know (mostly positive) and in relation to my previous health and illness" (rated 'very good') *[HIV sample: 20: PIQ]*

Response

In the general health scale item, responses are a continuation of the main part of the question. Issues of undefined terms (vague quantifiers, leading sometimes to a consideration of the relative position of the options), social desirability, and emotional reactivity, were raised.

The role of *relative position* was explicitly acknowledged, with participants appearing to rule out options according to their location on the continuum:

"I... had to, kind of, think of all these things in my head and... put that in a category... ...And I suppose I knew I certainly wasn't 'excellent' but I was there between 'very good' and 'good' ... and because I, kind of, looked on the page and recognised the, kind of, five categories, knew which ones I wasn't, I then, kind of, had to find answers in my head around how would I justify either of those two things" (rated 'very good') *[University sample: 8: Cognitive Interview: Retrospective]*

Some participants appeared to have ruled out all other response options to arrive at a final one:

"I knew it wasn't good, and poor was too harsh, so took the one between" (rated 'fair') *[HIV sample: 14: PIQ]*

There was also *vacillation* between adjacent categories (suggesting *vague quantifiers*):

"there wasn't much difference between 'very good' and 'good', I ended up going for 'very good' because, overall, I didn't think I had too many complicated issues in my life, so that's why I ended up putting 'very good'. I guess if I hadn't, if I thought I'd more than, let's say an average rate of illnesses or something I might have put 'good', or if it had more of a permanent effect on my life or something". (rated 'very good') *[University sample: 5: Cognitive Interview: Retrospective]*

Vacillation led to a criticism of the response, with a suggestion of an additional option:

"'Good', which I ticked, did not seem quite adequate to describe my health; so, I considered 'fair'. Probably something in the middle ('okay') would have been adequate." (rated 'good') *[University sample: 4: PIQ]*

And there was evidence for *end-aversion*, with the avoidance of the most positive response category, 'excellent' in both the university sample:

"... I think 'good' is in the middle, so it's quite neutral, you know. If I'm... I didn't put 'excellent' because I think there are healthier people...so I wouldn't regard myself as in excellent health, but I think it's 'very good' – it's better than 'good'" (rated 'very good') *[University sample: 2: Cognitive Interview: Retrospective]*

And the HIV sample:

"Saw 'excellent' but did not want to use that would tempt fate" (rated 'very good') [HIV sample: 17: PIQ]

In addition, the presence of HIV could act to put a *ceiling* on response:

"I'm doing good. I couldn't really put 'very good' until the day that it's gone, or it's, sort of, repressed to the point where it has almost no impact on your health" (rated 'good') *[HIV sample: 19: Cognitive Interview: Retrospective]*

Issues of *social desirability* occur at the response stage, leading to a more positive response than would be appropriate. One participant revealed the contradictory impulses behind their final response:

"I thought about answering it honestly. I also wanted to reflect a positive answer" (rated 'good') *[University sample: 3: PIQ]*

Finally, the general nature of the question could result in *emotional reactivity*:

"... frustrating, it tells nothing about me, or my health" (rated 'very good') *[HIV sample: 20:PIQ]*

Physical Functioning scale

Two- item scale:

The following questions are about activities you might do during a typical day. Does your health now limit you in these activities? If so, how much?

PF02: Moderate activities, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf... Yes, limited a lot / Yes, limited a little / No, not limited at all

PF04: Climbing several flights of stairs... Yes, limited a lot / Yes, limited a little / No, not limited at all

Summary

- Item PF02 was considered vague and open to subjective interpretation.
- Item PF04 provided a concrete example which was meaningful to most, although still interpretation varied.
- General perception and specific experiences response strategies were recalled; in addition, comparisons with others were also mentioned.
- Those without problems answered the items more easily; others, particularly HIV participants, dealt with issues of disease adaptation and restrictions.

Comprehension

In order to answer a health item, a sense of a *timeframe* is required. Across samples and methods, participants did not appear to have a problem considering that the PF items referred to an average day. In addition, the immediacy of the item could be understood by the inclusion of the word 'now':

"... the word 'now' – sets the question in a very specific time frame" [University sample: 1: PIQ]

The time reference for both questions was 'during a typical day'. In order to comprehend the items, this *key term* required clarification:

"'During a typical day'... I need to think about what 'a typical day' means to me before I can answer the question." [University sample: 10: PIQ]

Participants in the cognitive interview were asked to define a 'typical day' included in the question. Definitions were based on general experience, or a routine or average day, including a weekday, "a normal day" and "everyday". For example:

"A typical day is a kind of day you have most of the time... and a non-typical day would be maybe one of the weekend days you know, some celebration or something ... and you would be feeling a bit less able to do things... But normally, if everything is fine, and I wasn't up until 6 in the morning, that's a typical day" [University sample: 2: Cognitive Interview: Retrospective]

Or simply a general level of physical health:

"I was thinking generally, y'know." [HIV sample: 11: Cognitive Interview: Retrospective]

PF02 (Moderate activity)

The first of the scale items included the *key term* 'moderate activities'. Although supported by examples, the definitions provided differed due to the item being perceived as *vague*.

One participant remarked that the definition of moderate activities would be subjective:

"Probably, what you determine as a moderate activity depends on your overall level of health or fitness, anyway" [University sample: 7: Cognitive Interview: Retrospective]

And a participant with HIV reflected on moderate activities in relation to illness considerations:

"My idea of moderate activities might be some other people's idea of-of a lot of activity, erm, because I've always – I've got a way, I know that, erm, I think, I expect to be able to do a lot and, erm, I-I sometimes push myself too far." [HIV sample: 19: Cognitive Interview: Retrospective]

The term was generally interpreted in relation to physical effort, although the degree of exertion considered necessary differed. Most suggested that moderate activities are unchallenging for healthy people: activities requiring movement but not much effort, strength or speed:

"The ability to do day-to-day things, really. Things you'd normally do and maybe not really think about if you thought your health was fine." [University sample: 5: Cognitive Interview: Retrospective]

Although others indicated that moderate activities might be slightly strenuous, requiring effort and fitness to perform them:

"I'd say for a young person in London… nothing, I mean, I was talking, kind of, carrying shopping bags, I'm talking about, like, kind of, 12 bottles of mineral water, 54 bottles of wine – I'm talking about heavy things from the car, I was thinking "moderate" as, kind of, heavy bags rather than pushing a vacuum cleaner." [University sample: 8: Cognitive Interview: Retrospective]

HIV participants who considered they had health limitations related this item to physical exertion. For example:

"It all boils down to strength, y'know, it's the only way I can judge it… I don't have the stamina" [HIV sample: 12: Cognitive Interview: Retrospective]

There was *inconsistent use of examples.* Some participants felt the range and level of effort required varied between the examples provided:

"I think that these two – pushing a vacuum cleaner and bowling – are definitely more moderate than the others" [University sample: 7: Cognitive Interview: Retrospective]

Most participants *reinterpreted* the item, considering and disregarding the examples provided and instead generated examples of moderate activities they felt were more relevant to their own lives.

"... 'activities', interpreted this as related to work, being a mother – not necessarily as the 'moderate activities' you describe..." [University sample: 10: PIQ].

However, the examples contained in the question were also considered useful in relation to the apparent vagueness of the item:

"Confused at first about how to define limitations – on seeing examples and 'moderate' it was easier to come to an answer" [University sample: 14: PIQ]

PF04 (Climb several flights of stairs)

By the third item in the SF-12v2, there was already evidence of *context* effects, with participants starting to look for, and finding, similarities between items.

Both PF items share the same format. The structural and formatting similarities between the two questions guided comprehension of the question:

PF04: "I recognised it as similar to the previous question and followed a similar pattern of thought" [University sample: 3: PIQ]

Context was explicitly mentioned as a way to identify that the item was measuring level of physical exertion beyond that required for the routine tasks included in PF02:

"... I think, yeah, it was referring to more sort of physical exertion, not just routine daily tasks" [University sample: 6: Cognitive Interview: Retrospective]

However, the item was also considered to refer to general physical ability rather than strength and endurance:

"I was thinking three or four floors... Because I think if it was more it would be more physical activity and this is more a general question rather than testing your physical strength and endurance." [University sample: 2: Cognitive Interview: Retrospective]

HIV participants considered that the question tapped issues of energy, stamina, fitness and physical health limitations. For example, *"Fitness... How I cope with making physical effort" [HIV sample: 17: PIQ]*

Retrieval / Judgement

General perceptions, specific experiences and comparison with others were identified in both samples.

General perceptions were more commonly reported by those who were physically unlimited by health problems when carrying out routine activities or restricted by health problems.

In the following examples, there was the suggestion that a lack of problems, with consequently fewer indications of detailed cognitive work, made the process of responding seem *easier*:

PF02: "I straight away think that... er... that this question is more appropriate for disabled or older people, people who'd have difficulties with moderate activities and I think mostly young people wouldn't... so I didn't have any hesitation with this... I don't actually have any problems with mobility" (rated 'no, not at all') *[University sample: 2: Cognitive Interview: Retrospective]*

PF02: "I instantly thought, 'no, not limited at all', and did not think anything else." (rated 'no, not limited at all') *[University sample: 10: PIQ]*

However, general perceptions of health limitations were also reported. Some participants with HIV indicated that they had based their decision on a *general perception*, with evidence of *adaptation* to illness:

PF02: "I'm not really limited – I feel that I'm not limited, although I do limit... put a limit on myself, so that's why I put that, that I am limited a bit... I obviously can climb, I can, I can do a few, well I can do loads, actually, but then I would, again, restrict myself, because I don't want to overdo it' (rated 'yes, limited a little') *[HIV sample: 11: Cognitive Interview: Retrospective].*

And the effects of HIV and treatment were mentioned in relation to *general perceptions*:

PF04: 'I can actually climb stairs, but I'd still have to say that it was limited a little, because of the, I think it's the physiological factors – HIV itself and the after effects of drugs and whatever, that some days are worse than others' (rated 'yes, limited a little') *[HIV sample: 19: Cognitive Interview: Retrospective]*

Even among those who considered themselves healthy, a *comparison* with an ideal could result in a *general perception* of lack of fitness and a rating of limitation:

PF02: "I was thinking about how much effort you'd have to put in to doing it.... I was thinking about fitness and probably because, for me, I don't really have any health problems ... And relative to my physical abilities, I suppose" (rated 'yes, limited a little') *[University sample: 7: Cognitive Interview: Retrospective]*

In contrast, *comparison* with a conceptualisation of other people as limited may result in a rating that does not seem to relate to other information provided.

For example, knowing that one generally has difficulty climbing stairs can be *discounted* by a downward *comparison with others*:

PF04: "Well, I sometimes get a little bit short of breath but it doesn't actually stop me doing it. Erm... I think it's no. I think most people get out of breath climbing stairs" (rated 'no, not at all') *[University sample: 6: Cognitive Interview: Prospective].*

Among people with HIV, a *comparison* with past health was made between current state (living with HIV), indicating either that they had limitations earlier in the disease course, or a process of *adaptation*:

PF02: "My activities are not at all limited living with HIV now – so I answered in that way" (rated 'no, not limited at all') *[HIV sample: 20: PIQ].*

Reference to *specific situations* informed the judgement made to item PF04 in particular, relating perhaps to the concrete nature of the *key term* 'several flights of stairs', about which participants could recall particular instances:

PF04: "That's quite a good example, because, I think, quite often we are in situations where we do climb stairs ... So, that's a nice, practical example. I can, kind of, really *imagine doing it..."* (rated 'no, not at all') *[University sample: 8: Cognitive Interview: Prospective]*

A number of participants who indicated limitation, based their judgement on specific health problems and situations:

PF04: "That would be, as I said, several flights of stairs because I have that sort of balance, that kind of issue... Yes, if I had to climb several flights of stairs, that would be a problem... erm... And the other thing I thought was, 'yeah, I just climbed some stairs!', that's something to do with it..." (rated 'yes, limited a little') *[University sample: 5: Cognitive Interview: Retrospective]*

Participants also acknowledged other limitations, including smoking in relation to this item and recalled specific experiences where they had been short of breath had affected them:

PF04: "I am a smoker, so I thought of that as limiting my ability to climb several flights of stairs" (rated 'yes, limited a little') *[University sample: 10: PIQ]*

And some participants stated that they could straightforwardly relate the question to specific experiences where physical health problems had limited their ability to climb stairs:

PF04: "Reminded me of discomfort doing it" (rated 'yes, limited a lot') *[HIV sample: 16: PIQ].*

Response

Both Physical Functioning items had three response categories, two options for health limitation and a third for no limitation. The three options, rather than five used in other response scales, were felt by some to be an advantage in forcing a response:

"If you think about it, you can start going "mmm... is it this or that?", and there are only three choices here, so that probably stops you being less indecisive, because quite often you've got the big scales with multiple places that you can place your tick on the line" [University sample: 7: Cognitive Interview: Retrospective]

In addition to being simpler to understand and use than larger scales:

"Because there there's only three, which I - three's easy, 'cause three you just look and they're there in front on you..." [HIV sample: 11: Cognitive Interview: Retrospective]

However, participants could also find three options a restriction:

"I went for option 'Yes, limited a little', but I sometimes find myself between 'Yes, limited a little' and 'Yes, limited a lot' " [HIV sample: 13: PIQ]

Role Limitations

Four questions asked participants to estimate the amount of time in the previous four weeks that the performance of roles (work or other regular daily activities) had been limited, first by physical and then emotional health. The four questions were considered similar and also vague by participants, as commented on later.

Role Physical scale

Two- item scale:

During the past 4 weeks, how much of the time have you had any of the following problems with your work or other regular daily activities as a result of your physical health?

RP02: Accomplished less than you would like... All of the time / Most of the time / Some of the time / A little of the time / None of the time

RP03: Were limited in the kind of work or other activities... All of the time / Most of the time / Some of the time / A little of the time / None of the time

Summary

- The items were considered vague and interchangeable by many. Definitions of the key terms varied.
- General perception and specific experiences response strategies were recalled, along with a suggestion of satisficing.

- Those without problems answered the items more easily; others, particularly HIV participants dealt with issues of disease adaptation and restrictions.

Comprehension

The *key term* 'Regular daily activities' included at the start of the scale was generally considered to include a comprehensive range of activities ("everything I do", "what you do in your day", "things I wanted to do normally"), although it was also defined more narrowly, and more strenuously:

"Probably things like going to the gym – not that I'd do that daily – but... so I suppose I thought more about more exerting sort of activities... yeah, I think the word 'activities' makes me think of sport and things that require a bit more exertion than perhaps routine, just sort of cooking, whatever" [University sample: 6: Cognitive Interview: Retrospective]

There was evidence that participants were able to identify patterns of structure and layout in the questionnaire, indicating a *context* effect. In this example, perceived similarities between items RP02 and RP03 were felt to simply the task:

RP03: "That question's easier to understand following the first question, the first part of that question...erm... because the structure of the sentences is the same... Erm... and it was quite clear what I was answering at that point" [University sample: 8: Cognitive Interview: Retrospective]

However, some participants found the items *vague* and could not distinguish between questions RP02 and RP03, seeing both in relation to physical health limitations:

"[RP03] is very similar to the previous one – find [RP02] and [RP03] a bit vague." [University sample: 10: PIQ]

"This I found very similar to the one before... generally I just put the same answer...". [HIV sample: 11: Cognitive Interview: Retrospective]

RP02 (Accomplished less)

The *key term* '<u>Accomplished less</u>' was taken to include a sense of limitation due to physical health, the phrase being interpreted in terms of restrictions to the accomplishment of goals in work or other activities:

"I think it is asking if your physical health stopped you from doing things you'd have done if it wasn't there... If you didn't have the problem" [University sample: 2: Cognitive Interview: Retrospective]

Participants could also consider "scenarios", or situations that could be covered by the question:

"'Accomplished less than I would like'... That would have meant, basically, if I'd have been in bed with 'flu, or something like that, and then I'd not been able to get all the work that I'm supposed to be doing done... I had to come up with these little scenarios in my head ..." [University sample: 7: Cognitive Interview: Retrospective]

However, a problem was reported interpreting the negative wording of the question:

"I had to think about that bit of the question because, although you've underlined it, accomplished - because accomplished is quite a positive word, isn't it? – but then you've 'accomplished less' and that, kind of, that kind of, threw me a bit... I think, I'm almost, I'm almost expecting the question to be 'accomplished more than I'd like" [University sample: 8: Cognitive Interview: Prospective]

RP03 (Limited in kind of activities)

In the second item the *key term* 'kind' in the phrase, 'Were limited in the <u>kind</u> of work or other activities...' was generally defined as particular types of work or other activities:

"[kind means] you're limited to a certain type of work or activities" [University sample: 5: Cognitive Interview: Retrospective]

Although the 'kind of work or activities' could be defined as requiring a considerable amount of effort:

"Maybe lifting things, moving filing cabinets, things like that..." [University sample: 7: Cognitive Interview: Prospective]

The formatting of the question, underlining 'kind' also made it clearly a focus of the question:

"I did, kind of, look at the underlined word of '<u>kind</u>'. "Were limited in the <u>kind</u> of work or other activities" [University sample: 8: Cognitive Interview: Retrospective]

However, the apparent *vagueness* of the items led to some *reinterpretation* of their meaning, in one case in relation to working life:

RP02 and RP03: "Have I managed to cope with my day's work?" [HIV sample: 17: PIQ]

Retrieval / Judgement

A *general perception* strategy was commonly identified, particularly when the participant was largely unaffected by health problems. In a number of cases, a *comparison* between expectations and achievements was identified:

RP02: "I always accomplish what I set out to accomplish in a day" (rated 'None of the time') *[HIV sample: 15: Cognitive Interview: Retrospective].*

RP02: "… I think I do have…erm… realistic goals and I think I do achieve them most of the time… Yeah, most of the time I get done what I want to do… Certainly not restricted by my health, physical health" (rated 'a little of the time') *[University sample: 6: Cognitive Interview: Prospective]*

Similarly, among those with illnesses, there was evidence for a *comparison* between expectations and achievements. In the following example, there was the recognition of a discrepancy between expectations and achievements, caused by a health problem:

RP02: "I'm considering, like I have this… erm… this… medical condition, I suppose, which does affect my ability to do some things, but that's sort of part of my life, so I don't know, it's nothing extra and, accomplished less than I'd like, I don't know, yes it does affect you and I can't do some of the things I might have liked but… I don't know… I suppose 'some of the time'" (rated 'some of the time') *[University sample: 5: Cognitive Interview: Prospective]*

The task was *easy* for those who had a *general perception* of themselves as having good health:

RP03: "My health is good, so I didn't have to think about them too hard when I looked at them at first" (rated 'none of the time') *[University sample: 2: Cognitive Interview: Retrospective]*

However, the level of fitness and endurance required by the activities considered may vary, and if the threshold of performance is high enough, creating a *ceiling* on response, even apparently healthy people may consider that they are limited:

RP03: "That would be 'some of the time'... because... but, again, there are reasons... for that... I don't know... maybe lifting things, moving filing cabinets, things like that... which is not a regular part of my work" (rated 'some of the time') *[University sample: 7: Cognitive Interview: Prospective]*

Both items included a four-week recall period. When answering, some recalled *specific situations* over four weeks:

RP02: "'cause the last four weeks have been hectic for me, erm, because I got back from holiday, and sorting out staff, and rushing around..." (rated 'some of the time') *[HIV sample: 12: Cognitive Interview: Retrospective]*

However, *timeframe* was not necessarily adhered to. When asked retrospectively about the four weeks, some participants speculated that they may have either considered shorter periods:

RP02: "Maybe I've used a week and sort of just used that as a way to judge it ..." (rated 'a little of the time') *[University sample: 6: Cognitive Interview: Retrospective]*

Or recent events over a less well-defined timeframe:

RP02: "Just thinking back on it... the activities over the past few weeks. Erm... I probably didn't go back a full... as far back as four weeks, just what I remember mainly over the last couple of weeks" (rated 'a little of the time') *[University sample: 5: Cognitive Interview: Retrospective]*

For both Role Physical items, the health burden of HIV was referred to by most HIV positive participants, in relation to reviewing *specific situations* and limitations:

RP02: "I thought about restrictions that have been placed on me by my health situation in the last month" (rated 'some of the time') *[HIV sample: 14: PIQ]*

However, strong *satisficing* was also indicated, with a participant indicating that they gave a response without making a judgement about the answer to the item:

RP03: "...you sort of scan up and down a few times and then think, 'well, I'll pick the middle one'" (rated 'some of the time') *[HIV sample: 11: Cognitive Interview: Retrospective].*

Response

The apparent *easiness of* the task for those with no limiting health problems (who could simply select, 'None of the time', in contrast to other participants) was highlighted:

"I think it is easier to answer these kinds of questions if you don't have any problems or if you do have a lot of problems, yes. If once or twice there was something wrong, I'd probably read more carefully the differences between the other categories" (rated 'none of the time') *[University sample: 2: Cognitive Interview: Retrospective]*

However, some participants with health problems indicated that they had problems with *vague quantifiers*:

"I have to put 'some of the time', because almost everyday it has that effect in the morning, so, but then, so, but then, every so often I have a day – I don't know why this is, it's as if some strange reaction occurs with the drugs, or, maybe the HIV itself does it – erm, then I feel just off for the whole day, and I can't, don't feel like doing anything, I just need a rest – and it's above and beyond the fact that people need a rest, anyway – it's a sort of extra feeling, so, when I put all that together, I have to put 'some of the time'... but... that might be 20% of the time or 15% - it's not 'all of the time', and it's manageable" (rated 'some of the time') *[HIV sample: 19: Cognitive Interview: Retrospective]*

In addition, it appeared that the middle response, 'some of the time', was used to indicate an average, due to its *relative position* rather than the label:

"Because of the fluctuation day to day the... answer might seem a bit vague" (rated 'some of the time') *[HIV sample: 13: PIQ]*

Role Emotional scale

Two- item scale:

During the past 4 weeks, how much of the time have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious)?

RE02: Accomplished less than you would like... All of the time / Most of the time / Some of the time / A little of the time / None of the time

RE03: Did work or other activities less carefully than usual … All of the time / Most of the time / Some of the time / A little of the time / None of the time

Summary

- Most of the points raised for the RP also apply to the RE scale.
- The items were considered vague and interchangeable by many. Definitions of the key terms varied.
- The RE scale was considered similar to RP: some found this helpful while others were confused.
- General perception and specific experiences response strategies were recalled, and there was evidence of satisficing.
- HIV participants raised disease issues which related to these items.

Comprehension

The subjective quality of emotional states resulted in the items being considered *vague* and variation in interpretation of the *key terms* employed.

'Emotional problems', as used in these questions, was generally considered to relate to mild affective problems, including day-to-day anxiety and sadness, feelings and mood, anxiety and tearfulness, and frustration.

However, some took issue with the use of the term:

"In the context of this question, ['emotional problems'] is a bit stronger that I would've put it, because I think a problem would be more sort of… er… more permanent or… er… bigger, as opposed to a temporary issue, but it does ask about the past four weeks… it just seems a bit negative, I guess" [University sample: 5: Cognitive Interview: Retrospective]

And the juxtaposition of depression and anxiety was questioned, indicating that this item could have a *double meaning*:

"Depression is serious, a serious condition, but anxious is less serious, and can be dealt with easily. You can get over it without any help" [University sample: 2: Cognitive Interview: Retrospective]

Sometimes participants *reinterpreted* the item accordingly:

"Feeling depressed, I'd be very careful about that one, because that's a very serious condition, so, I'd say it was mainly, I mainly related it to forms of anxiety" [HIV sample: 19: Cognitive Interview: Retrospective]

However, the inclusion of examples was found to be useful in defining what was perceived to be a *vague* term:

"'Emotional problems' ... in my mind I quickly need to define them before I can answer the question... establishing meaning of question needs some time... Found the explanation of 'emotional problems' (depressed or anxious) helpful" [University sample: 10: PIQ]

As with the Role-Physical scale, there was evidence of *context* effects: the two questions were considered *vague*, leading them to be understood to be asking the same question:

RE03: "Basically the same as the one before [RE02], very similar" [HIV sample: 18: Cognitive Interview: Prospective]

However, other participants indicated that they used the *context*, the similarities and discrepancies between questions, to support their reading and interpretation of the items. The layout and wording of the Role-Emotional questions were similar to those used for the Role Physical questions and there was evidence that participants focused on elements of the questions and formatting that were similar, and identified those aspects that were novel:

RE02: "'Accomplished less' had already appeared in [RP02]. It was easier because I'd already got over the hurdle of getting it" [University sample: 8: Retrospective]

RE02: "I more-or-less expected this kind of question ([Role Physical] and [Role Emotional] are analogous) and had almost decided on my answer before I had even *read this... 'Emotional problems', 'depressed', 'anxious'. They were what distinguished this question from [Role Physical]" [University sample: 4: PIQ]*

Familiarity with the format of the questions had also simplified the task for some:

RE03: "Read question quicker because of last question [RE02]... I was already familiar with the general gist of the question... [focused on] 'less carefully than usual' [RE03] – because it was underlined" [University sample: 3: PIQ]

Retrieval / Judgement

Judgements based on *general perceptions* and *specific experiences* were reported, and there was also evidence for *satisficing*.

In the university sample, there was an example of a *general perception* of positive mental health:

RE03: "I have not experienced any emotional problems in the last 4 weeks." (rated 'none of the time') *[University sample: 1: PIQ]*

While the subject matter had changed to emotional problems, some HIV participants still referred to the effects of their disease. This could be represented as a *general perception* of a negative psychological impact of HIV-related problems, without reference to the nature of the limitation on performance:

RE02: "There's a lot of feelings connected with HIV ...it can be quite crushing, just to think this, kind of, virus is just there and won't go away and you have to always be taking medication" (rated 'some of the time') *[HIV sample: 19: Cognitive Interview: Prospective].*

Specific experiences also related to the effects of HIV were also recalled:

RE02: "Anxious – I have had some anxiety recently over my physical health" (rated 'a little of the time') *[HIV sample: 14: PIQ]*

In both samples, recent stressful life events were recalled which had affected performance:

RE02: "I have been a bit anxious in the last four weeks because I have come to London. So… I chose 'A little of the time'" (rated 'a little of the time') *[University sample: 2: Cognitive Interview: Retrospective]*

RE02: "these were playing on my mind most of the time …not focusing on my work or other jobs I would … normally do" (rated 'most of the time') *[HIV sample: 12: Cognitive Interview: Retrospective].*

However, there was some evidence for the use of *satisficing*. In the following example, a participant admitted that they could not recall specific incidents when they were limited by emotional problems, although they assumed that they would have been affected by anxiety at some time during the four weeks:

RE02: "I'm sure 'anxious' has somehow come into it… Erm… 'a little of the time'" (rated 'a little of the time') *[University sample: 7: Cognitive Interview: Prospective]*

In the next example, strong satisficing was indicated, in which response was unrelated to the contents of the items:

RE02: *"…after I've got to the third page, I 've just started ticking willy-nilly because I'm just bored, I've had enough, I just want to get to the end…"* (rated 'some of the time') *[HIV sample: 11: Cognitive Interview: Prospective].*

Response

As was the case with previous items, because of *vague quantifiers*, participants used the *relative position* of categories to help them decide on their meaning in relation to one another:

"It is easy with four weeks, I looked at the answers like four weeks. 'All of the time' is all of the four weeks. 'None of the time' is none of the four weeks. So, one week is the four weeks is 'A little of the time', and I chose that…. Yes, I was thinking of four weeks… I looked at the position… I chose the answer that was more than none of the four weeks" (rated 'a little of the time') *[University sample: 2: Cognitive Interview: Retrospective]*

And one participant attempted to estimate the proportion of time 'some of the time' (the middle option) would relate to:

"I had to put down that, erm, I sometimes would accomplish less 'some of the time' in the last four weeks because of having my brain cluttered up with all these things. Well, what I mean by that is probably, maybe, out of a whole week, erm, 10% of the time" (rated 'some of the time') *[HIV sample: 19: Cognitive Interview: Retrospective]*

However, there was evidence that *social desirability* considerations arose in relation to mental health problems:

"I thought about whether I wanted to tell the researcher about my [negative affect]." (rated 'none of the time') *[University sample: 4: PIQ]*

Bodily Pain scale

Single item scale:

BP02: During the past 4 weeks, how much did pain interfere with your normal work (including both work outside the home and housework)? Not at all /A little bit / Moderately / Quite a bit / Extremely

Summary

- The item was considered straightforward to understand and answer by most: reasons given included that pain was a more understandable concept, and questions on physical issues are easier to answer than those on mental health or role performance.
- General perception and specific experiences response strategies were recalled. The general perception strategy was considered easy when there was no pain present.
- HIV participants raised disease issues which related to this item.

Comprehension

In contrast to most other items, Bodily Pain was generally considered a simple concept:

"I thought this was quite straightforward. Does anything cause you pain? Does that affect what you do? And... erm... that's pretty much a straightforward answer to it" [University sample: 5: Cognitive Interview: Retrospective]

A definition provided referred to pain existing in a variety of forms:

"A huge spectrum… everything from extreme pain through to… finger throbbing" [HIV sample: 19: Cognitive Interview: Retrospective]

Once again, participants brought to the question their experience of the previous questions, indicating *context* effects:

"Thought about the gist of the question, having previously read the others, looked for specifics – 'pain' and answered accordingly… [focused on] 'past 4 weeks' and 'pain', because underlined" [University sample: 4: PIQ]

Retrieval / Judgement

Most participants based their judgements on a *general perception* that they had not experienced pain. With an indication that the absence of a problem meant that the response process was considered to be *easy*:

"It is easy for me to answer, because I have no pain" (rated 'not at all') *[University sample: 2: Cognitive Interview: Prospective]*

And physical symptoms were also considered more straightforward, and therefore *easier* to consider than emotional ones:

"Much easier to judge physical symptoms, for me, anyway, than perhaps deal with the quantity of them, certainly the emotional ones, it's very difficult to quantify emotional symptoms… for me" (rated 'not at all') *[University sample: 6: Cognitive Interview: Retrospective]*

Those in pain could also indicate that they based their judgement on a *general perception*, should they be in pain most of the time:

"it's fact; it's how I feel… Because I'm in pain – discomfort a lot of the time" (rated 'quite a bit') *[HIV sample: 13: PIQ]*

However, those with pain were more likely to consider *specific experiences*. In the following example, the participant mentioned the effects of pain at work, although they added that they had adapted their working life to minimise any interference:

"Period/menstrual pain affects my work… Menstrual pain is worse if I am sat at a desk for long periods so I try to rearrange my work so I can be more active on those days." (rated 'a little bit') *[University sample: 9: PIQ]*

Among HIV patients, pain was mentioned in relation to the disease, with the suggestion of *adaptation* to a chronic problem:

"it's probably HIV itself, but also the medications, cause a, sort of, aching in the body … it's a sort of dull pain…it doesn't really stop you doing things, but it sort of slows down everything, and you get so used to it that you don't notice it, but sometimes I feel that I'm just, y'know, I'm really having to force myself, so, I would say that, in the last four weeks, this has probably happened… erm… 'a little bit'" (rated 'a little bit') *[HIV sample: 19: Cognitive Interview: Prospective]*

Mental Health scale

Two item scale:

This introduction relates to a group of three items also including the Vitality item: These questions are about how you feel and how things have been with you during the past 4 weeks. For each question, please give one answer that comes closest to the way you have been feeling. How much of the time during the past 4 weeks…

MH03: Have you felt calm and peaceful?… All of the time / Most of the time / Some of the time / A little of the time / None of the time

MH04: Have you felt downhearted and depressed?… All of the time / Most of the time / Some of the time / A little of the time / None of the time

Summary

- Unlike preceding items, complex interpretation and judgement processes were necessary: these items required all participants to consider the degree to which they had particular feelings during a time period.
- The items were considered vague, including double meanings, and normative assumptions; individual definitions varied.
- Complex response processes were reported.

- The items were answered in relation to general perceptions and specific experiences, and satisficing was also suggested.
- Context effects were identified.

Comprehension

The two MH scale items were considered to be *vague*. In this situation, participants shown that they undertook considerable cognitive effort in order to define and interpret the items.

MH03 (Felt calm)

Definitions of the *key term* 'calm and peaceful' varied. Some considered a positive interpretation, either adding nothing to this key phrase, or mentioning words such as "relaxed".

The term was sometimes defined in relation to an opposite: "angry and restless", "stress, feeling anxious", "stressed out":

"To be able to get on with daily life without a great deal of stress and, erm, er, not feeling at war with the world, feeling at peace with the world in terms of, erm, if someone is rude to you, you just think, 'oh well, most people haven't been rude today', and that person was and it doesn't matter, erm, and, erm, y'know, just feeling – you know you feel light as you walk around and do things and, erm, not-not-not worried, well, not very worried about anything… you can deal with things in a smooth, effective way" [HIV sample: 19: Cognitive Interview: Retrospective]

The juxtaposition of 'calm' and 'peaceful' was questioned; individually they were considered *vague* terms and combined they gave the item a *double meaning*, leaving the question open to *reinterpretation* by the participant:

"I've never used the word 'peaceful', because peaceful for me sounds a bit, erm, sounds a bit religious - and I'm not, and it sounds, kind of… somebody dies peacefully, don't I I don't know if calm is a word I would… I would say 'relaxed', I suppose, more than calm so I suppose it's a terminology thing – I really had to think about those words before I could, kind of, gauge what the question was asking me. But 'peaceful' makes me think of… am I sitting down with, kind of, a tape in the background with waves

going on... and that's something I'd never do, anyway, so... I found that, kind of, difficult to gauge what it was asking me. But again, I've just interp-I've just changed those words for words that I would, kind of... for me would, kind of, gauge that's what, that's what the question's asking me, kind of, then answer the question... I was, kind of, substituting those words for something else" [University sample: 8: Cognitive Interview: Retrospective]

The question was also problematic because of a perceived *normative assumption.* Some questioned whether being 'Calm and peaceful' was necessarily positive:

"I quite like questions that have like the two extremes with the one statement on one side and one statement on the other side, so it means that you know exactly what people want you to distinguish between... erm... and... from 'calm and peaceful'... I mean, is being excited and having fun the opposite of 'calm and peaceful'? Or is the opposite of being 'calm and peaceful' being stressed out and pissed off most of the time..." [University sample: 7: Cognitive Interview: Retrospective]

MH04 (Felt downhearted)

The second item included the *key term*, 'downhearted and depressed'. Definitions of the included generally milder affective problems, including mood, feeling miserable, and a negative approach to life. This is another example of a question capable of having a *double meaning* as a result of the use of the two terms. As with the Role Emotional items, the use of 'depressed' was queried for being too serious to include alongside another term:

"Depressed is something serious... And I think that downhearted must not be as bad... it is difficult to think about them together... depression is very serious" [University sample: 2: Cognitive Interview: Retrospective]

However, there were differing views about this. There was also evidence that participants made use of *context* and were familiar with the format and contents of the questionnaire.

"...just reminds me of the previous questions about emotional problems, erm, feeling depressed and that kind of thing ... you'd associate one with the other. Wouldn't think of each problem separately, I wouldn't look at it thinking, 'oh, I might have only been downhearted' and 'have I been depressed?'. I'd just treat it as one" [University sample: 5: Cognitive Interview: Retrospective]

Retrieval / Judgement

When answering both questions, *general perceptions* and emotions relating to *specific situations* were reported.

A number of participants based their perception on a *general perception* of themselves and their personality, not necessarily related to events in the previous four weeks:

MH04: "I'd say never depressed or downhearted, always look on the positive side generally, and, kind of, always... It's the old, kind of, can-do approach, isn't it – if someone says to me, 'now what do you think? How will I do it differently?' Then I sort of get them to say, 'yes'. So, I don't think I'd be downhearted" (rated 'none of the time') *[University sample: 8: Cognitive Interview: Retrospective]*

MH03: "I'm that sort of person... I am calm and cool and people can throw things at me and I just have to deal with it" (rated 'most of the time') *[HIV sample: 12: Cognitive Interview: Prospective]*

However the possible *double meaning* of the item could influence the judgement process. In the following example, the participant experienced recent sadness but based their response on a general picture of themselves as not being depressed:

MH04: "I can't think about depression and something that is not as serious. I had to answer 'None of the time' because I am not depressed" (rated 'none of the time') *[University sample: 2: Cognitive Interview: Retrospective]*

In addition, providing a summary of emotions over four weeks was considered a *difficult* task because of the *variability* of emotions (in comparison to physical symptoms):

MH03: "I thought about my feelings over the past month in general... A little less easy as feelings can vary and change very frequently" (rated 'some of the time') *[University sample: 1: PIQ]*

Among HIV participants, *general perceptions* were often reported in relation to the disease.

For example, *general perceptions* of the negative emotions associated with HIV problems, unrelated to specific occasions or events:

MH03: "A lot of the time I'm stressed and frustrated at the limitations of my health" (rated 'some of the time') *[HIV sample: 13: PIQ]*

However, there was evidence for the effects of HIV being mentioned and *discounted*, suggesting *adaptation*:

MH04: 'I mean in my life I have been a little downhearted and depressed, but, considering my treatment, not really, no I don't get down about it' (rated 'none of the time') *[HIV sample: 18: Cognitive Interview: Prospective].*

Along with an indication that long-standing *general perceptions* of mood were involved in the response process, suggesting that the answer related to personality factors:

MH03: "Never was very calm pre-diagnosis and never have been calm, so what relevance is this question?... Looked at how I felt – reacted" (rated 'most of the time') *[HIV sample: 20: PIQ]*

Participants who selected a response reflecting a lack of calm and peacefulness or presence of downheartedness and depression could consider *specific experiences*, either over a period of time or according to recent life events.

The judgement process might include a *comparison* between usual feelings and feelings evoked by recent specific experiences:

MH03: "How much of the time during the past 4 weeks... Have you felt calm and peaceful...erm... 'Some of the time'... It's... it's... I'm not quite sure, er, because it's been, you know, quite a rough four weeks because a lot of new things happened, but, but usually I feel peaceful most of the time, so I'd say some of the time now, just a little bit less than that. It's quite... it's quite strange that you are asking about the past four weeks, 'cause it's a very distinct answer. If you'd have said, in the past 6 weeks, I'd have problems with answering, because 6 weeks ago and four weeks ago, it's just two completely different things totally. So, it's quite good, it's very relevant to me" (rated 'some of the time') *[University sample: 2: Cognitive Interview: Prospective]*

In addition, the "stress" of recent life events could be recalled:

MH03: "Christmas and New Year were a bit frantic at times and [mentioned currently undergoing a specific life event] so there are the odd moments of stress" (rated 'some of the time') *[University sample: 9: PIQ]*

Ambivalence with the perceived *normative assumption of* the question, whereby feeling 'calm and peaceful' 'all of the time' was considered to be the positive response, was also mentioned.

In the following example, the participant's judgement was based on comparing perception of themselves as a generally relaxed person with recent experiences. Of note, the question was reversed and considered as a negative during initially consideration:

MH03: "I've had lots of pressures at work and I'm probably not as, as calm as I'd like to be... Cause I think... erm... 'Cause I think I-I'm relaxed most of the time, but sometimes I get a wee energy, kind of, hype and I need that sometimes to get me up and going, so I wouldn't want to feel calm and peaceful all of the time, I'd want to feel that I had lots of things going on to get me, kind of adrenaline bursts, kind of, to keep me alert" (rated 'most of the time') *[University sample: 8*: Cognitive Interview: *Retrospective]*

And it was considered that not feeling 'calm and peaceful' can be positive:

MH03: "It isn't necessarily a bad thing, because, when I think of it in terms of the work I-I do, it was constructive not feeling calm and peaceful, actually, it was, was part of the process of thinking my way through various things in relation to that, so it wasn't a bad thing, there" (rated 'A little of the time') *[HIV sample: 19: Cognitive Interview: Retrospective]*

Finally, confusion caused by the *normative assumptions* of the question provoked difficulties in reconciling the response to the item with experience:

MH03: "I answered it as being calm and peaceful 'most of the time' but I was also having quite a lot of fun" (rated 'most of the time') *[University: 7: Cognitive Interview: Retrospective]*

As with previous items in this sequence, there was evidence for *satisficing*, with the judgement based on an assumption that, during the course of the four weeks, the participant must have experienced emotional problems:

MH04: "...you kind of think, 'aw, well they've mentioned it, so I must have done a bit 'Some of the time", and if I think about it a bit more, then, maybe, yeah, 'A little of the *time' is right, but if I answered it straight off then I probably would have gone, 'aw, 'None of the time', even though it does have 'past 4 weeks' underlined twice..."* (rated 'a little of the time') *[University sample: 7: Cognitive Interview: Retrospective]*

Response

Participants provided some explanation of the categories, indicating that they had ruled out options based on their *relative position*. In the following examples, extreme categories were ruled out and the middle option chosen, seemingly based on an *average* for *variable* emotions:

MH03 "... the past four weeks have been quite stressful, mainly workwise, and generally I see myself as a calm and peaceful person who can cope fairly well, but I suppose, because of the recent sort of experiences, I couldn't say 'none' and I couldn't say 'all' either, so I went for the middle... I think I hesitated over it... I think what was going on was I was thinking that I am, generally speaking, a calm and peaceful person, and then it was, sort of, trying to balance my normal sort of state with the last four weeks, and, sort of, comparing those two I found that quite difficult, I think. And to remember, actually, that sort of length of period, exactly what had happened in those four weeks" (rated 'some of the time') *[University sample: 6: Cognitive Interview: Retrospective]*

MH04: "'Downhearted and depressed'. 'Some of the time', again relating to what I was saying earlier, it's just been a bit up-and-down recently" (rated 'some of the time') *[University sample: 5: Cognitive Interview: Prospective]*

However, the terms in the response represented *vague quantifiers* for some participants:

MH03: "It was not clear to me what 'some of the time' and 'a little of the time' mean... Few people who have a 9-5 job feel 'calm and peaceful' 'all of the time'" (rated 'a little of the time') *[University sample: 4: PIQ]*

Comments from one participant both criticised the options available and indicated *social desirability* considerations, and *emotional reactivity*:

MH03: "I thought about answering the question honestly in relation to my [mental health condition] (and did), however, I felt frustrated ... the question is very limited in categorising a complex emotion in this way... it's very hard to realistically measure a mood by the use of five categories." (rated 'some of the time') *[University sample: 3: PIQ]*

Vitality scale

Single item scale:

This statement refers to the group of three items including the two MH scale items:

These questions are about how you feel and how things have been with you during the past 4 weeks. For each question, please give one answer that comes closest to the way you have been feeling. How much of the time during the past 4 weeks…

VT02: Did you have a lot of energy? All of the time / Most of the time / Some of the time / A little of the time / None of the time

Summary

- The key phrase was ambiguous to a number of participants and a number of interpretations were given.
- General perceptions and specific experiences were reported.
- Energy was a difficult sensation to report because of its variability.
- People with HIV mentioned their condition in relation to response.

Comprehension

Most participants focused on the *key term* energy. However, the phrase 'a lot of energy' varied and caused some confusion. Some participants considered the question to be about feeling energetic while others referred to negative feelings, such as feeling tired or stress and exhaustion.

One participant explained that they considered the question to have a *double meaning*:

"To me, question means both "having energy" and "feeling tired/not having energy"" [University sample: 10: PIQ]

The inclusion of 'a lot' in relation to energy also added to the difficulties with comprehension expressed:

"Had to decide just how much energy is a lot of energy" [University sample: 1: PIQ]

A participant speculated that they defined energy as the opposite of calm and peaceful (the previous item in the SF-12v2), reflecting the effect of *context* on the interpretation of ambiguous terms:

"Being asked about being "calm and peaceful" probably made me say, 'aw, well, I was quite calm and peaceful', so... saying that I had a lot of energy is... the opposite of that" [University sample: 7: Cognitive Interview: Retrospective]

Retrieval / Judgement

Participants based their responses on *general perceptions* or *specific experiences.*

It was reported that the task was *easy* when the participant had no problems and could report a *general perception*:

"It is easy for me to answer because I always have enough energy to do things" (rated 'all of the time') *[University sample: 2: Cognitive Interview: Retrospective]*

Although a clear self-perception also made the task seem *easier*:

"Some people thrive on nervous energy, and I think I may be one of those people" (rated 'most of the time') *[HIV sample: 15: Cognitive Interview: Prospective]*

However, it was conversely considered a *difficult* task when energy was considered *variable* and problematic to summarise:

"I never really have a lot of energy; but just enough most of the time...Energy levels fluctuate a lot. You have to think a bit more about it." (rated 'some of the time') *[University sample:1: PIQ]*

Some HIV participants reported that they considered a *general perception* of their energy in relation to HIV-related problems:

'I have lots of energy. It does come in waves. A lot of people [location] think I've got loads of energy, because I rush around doing things, but I know I could have a lot more... if my body didn't ache so much, erm, the outside appearance can be very deceiving, y'know, he goes to the gym, he works out, he's got muscles, but underneath, sometimes, I don't have the stamina, I just don't have the energy, so 'some of the time' is right. I wouldn't put 'most of the time', for sure.' (rated 'some of the time') *[HIV sample: 12: Cognitive Interview: Retrospective]*

Participants could make a *general perception* involving a *comparison* between perceived levels of energy currently and at a younger age:

"I think, I think I do have lots of energy, and I think have it most of the time. I think I'd probably... If you, kind of, look at, kind of, your life, if you like, your lifespan, I'd probably say I've got less energy now than I had, like, 5 years ago, but I'm sure that's the same for everyone that's getting older" (rated 'most of the time') *[University sample: 8: Cognitive Interview: Prospective]*

Some participants indicated that the question had a *normative assumption*, which they might not subscribe to. In the following example, a participant provided an answer based on a *general perception*, although they interpreted the item to indicate that a lot of energy was positive, which they did not have to accept:

"I did think in relation to this question, 'A little of the time', but it's not me... as in, the person I am; it's just that I'm not a naturally exuberant, just jumping-around the place full of energy kind of person, anyway, which is why I ended up choosing 'a little of the time'. But if you were doing person-specific, I wouldn't say that, I mean I'd just think I was just normal, but lot of energy? No... Slightly tricky, because I didn't really think that it relates very well to me, but... I had to answer it as it's written" (rated 'a little of the time') *[University sample: 5: Cognitive Interview: Retrospective]*

In relation to *specific experiences*, judgement could take the form of a comparison between a general perception of usual levels of energy and specific experiences:

"I haven't been as physically active as I normally am... That's been more about other pressures... Erm... [reads question quietly]... 'some of the time', I suppose" (rated 'some of the time') *[University: 6 sample: Cognitive Interview: Prospective]*

Specific experiences could occur as a result of HIV-related medical problems:

"I feel a little bit tired in the afternoons… I'm fine all morning and suddenly my energy levels take a dip" (rated 'some of the time') *[HIV sample: 18: Cognitive Interview: Retrospective].*

For some participants, the process of making the judgement was *difficult* as it involved *variability*. A widespread perception guiding many judgements was that 'energy' is a finite resource possessed by the individual that is drawn on by events in everyday life.

In the following example, the participant referred to a comparison of resources and utilisation of energy in *specific situations*:

"I had quite a stressful/exhausting time in the last four weeks and have gone through a lot of changes as well as worked a lot. Although I was often exhausted, I thought the fact that I had managed to do what I needed to do indicated that I had had a lot of energy – just I had used it up sometimes." (rated 'most of the time') *[University sample: 4: PIQ]*

Response

The relative position of response categories again led the middle option, 'some of the time', to be seemingly considered indicative of an *average* for a *variable* sensation:

"It comes in waves" (rated 'some of the time') *[HIV sample: 12: Cognitive Interview: Prospective]*

"Some days are better than others' (rated 'some of the time') *[HIV sample: 18: Cognitive Interview: Retrospective]*

As with the questions on mental health, the response options were criticised and *social desirability* considerations mentioned:

"…again there seems a lack of choice in answers … I wanted to write 'most of the time' because it seems a more positive answer but answered 'some of the time' as the more truthful" (rated 'some of the time') *[University sample: 3: PIQ]*

Social Functioning scale

Single item scale:

SF02: During the past 4 weeks, how much of the time has your physical health or emotional problems interfered with your social activities (like visiting friends, relatives, etc.)?... All of the time / Most of the time / Some of the time / A little of the time / None of the time

Summary

- Most people felt that they could understand the social functioning scale.
- General perceptions and specific experiences were reported, along with satisficing.
- People with HIV mentioned their condition in relation to response.

Comprehension

'Social activities' was identified as a *key term* for this single item assessment. This was generally defined as being with other people in non-routine or work settings: socialising, going out, social events, entertaining, going to the pub, drinking and having dinner, were all given as examples. None required the use of the examples 'like visiting friends, relatives, etc)':

"Social activities... are when you socialise with other people... when you go out. I didn't need them... It would have been easy to answer without examples, but I checked them, anyway" [University sample: 2: Cognitive Interview: Retrospective]

Most participants focused on elements of the phrase 'interfered with your social activities (like visiting friends, relatives, etc.)'. One participant found the term 'interfered' *vague*:

"...I thought of 'interfere' to mean that I had not visited a friend because of a [negative affect]. But then I thought 'interfered' could also mean 'influenced'" [University sample: 4: PIQ]

Retrieval / Judgement

A *general perception* of the self as sociable was revealed by a number of participants, leading them to consider that the judgement was *easy*:

"That was very easy. I thought, I thought… erm… I think if I broke my leg I'd still go out in the pub, kind of thing – I'd still manage to get there somehow" (rated 'none of the time') *[University sample: 8 : Cognitive Interview: Retrospective]*

HIV participants also mentioned debilitating effects of their illness which would affect them:

"In terms of just general fatigue, I suppose, occasionally, I suppose… often [socialise] when I feel under the weather, because it usually makes me feel better" (rated 'a little of the time') *[HIV sample: 19: Cognitive Interview: Prospective]*

The effects of HIV were also mentioned in relation to the effects treatment changes had had on emotional health. In the following example, the *general perception* is based on a *comparison* of the participant's current situation with a time when they experienced treatment side effects:

"It is how things are at the moment… when I started taking [anti-HIV medication] my emotional state went through the floor. I'm not on [anti-HIV medication] now, as a result things have got a bit better" (rated 'most of the time') *[HIV sample: PIQ: 13]*

Another participant referred to a *general perception* that was particularly combative, indicating possible *adaptation* to HIV or personality effects:

"I'm not going to let HIV get in way of my life" (rated 'none of the time') *[HIV sample: PIQ: 20]*

Specific situations were also recalled in which social events were missed due to physical health or emotional well-being. A *comparison* process was evidence, involving a general perception, in which social events are not or only seldom missed, with discrete events:

"I didn't go to a social event because I felt under the weather, shall we say, emotionally, not physically… erm… so again, 'A little of the time'; it's not that often" (rated 'a little of the time') *[University sample: 6 : Cognitive Interview: Prospective]*

Specific experiences were also recalled in relation to a few occasions in which health or well being had limited participants' social lives:

"It hasn't really stopped me going out... erm... maybe 'a little of the time'... I didn't really fancy going out the other night, because I was waiting for [stressful life event], so, maybe 'a little bit of the time" (rated 'a little of the time') *[HIV sample: 15: Cognitive Interview: Prospective]*

One participant could not recall any incidents where this had happened but indicated that they adopted a *satisficing* approach, estimating that they must have been affected 'a little of the time' on the basis that they knew that a mental health problem would have made it likely:

"I thought I knew that my [mental health problem] does at times interfere with my social activities, but I couldn't remember any instances in the past four weeks, so I answered just what seemed to be the most honest/likely answer (like 'probably a little'.... I couldn't truly remember any instances, whilst recognising it's quite likely – I guess if I put 'none of the time' I don't feel I would have truly represented my health" (rated 'a little of the time') *[University sample: PIQ: 3]*

Response

Participants who considered the response options to be *vague quantifiers* had to consider the *relative positions* of the categories. In the following example, the participant was ruling out options in relation to a specific life event:

"There are five possible answers, erm – 'none of the time' isn't right, erm, it did interfere with my social activities, but 'a little of the time' because I only missed going to see a couple of friends. If I hadn't gone out last night, then I may have put 'some of the time' – that's very middle of the road, erm, it has certain aspects of 'most of the time' and certain aspects of 'all of the time' and that's how I distinguish it; if there's five possible answers, the middle one must be middle of the road, if you like; it wasn't middle of the road, it was slightly more of the time, so, if that makes sense." (rated 'a little of the time') *[15: HIV sample: Cognitive Interview: Retrospective]*

2.4 Comparison and summary

The results obtained bear out other qualitative research that has demonstrated the complexity of the concept of health, open to a range of understandings (Calnan, 1987). In practice, when answering any health question, but particularly a vague item, such as GH01 (general health), respondents have to implicitly define and weight various health dimensions (Pavot and Diener, 1993). As was shown in this study, numerous issues were raised, including physical and mental health, fitness, chronic disease medication side effects, and even non-health factors, such as humour. Similarly, in the Health and Lifestyles Survey (HALS) (1984), respondents were asked what they understood by the concept of health: physical and psychological fitness, functional ability; a moral dimension, including will power, self-discipline, self-control; health behaviours, including smoking, drinking and exercise; plus the perception as health as a tangible entity that could be used up through neglect or increased through nurturance (Blaxter, 1990).

Although there were differences between samples, they generally related to the substantive health differences that influenced the types of response processes reported. *University* and *HIV* participants tended to discuss the same issues about comprehension, retrieval/judgement and response. However, samples varied by age, gender and roles occupied as well as HIV status. It is possible that the results provided related to these factors, although there was no clear evidence for this from the analysis. The major difference was the fundamental influence of HIV on all aspects of lifestyle, including the effects of medication, lethargy, pain, and related illnesses: the disease clearly affected all aspects of the lives of many of the HIV sample participants. There was evidence of adaptation but there were still clear deficits indicated in responses and comments provided.

All results were grouped according to response stages from comprehension to response. In terms of *comprehension*, many items were considered vague, and there was considerable variation in interpretation of terms. However, this was often not problematic, because participants generally indicated their healthiness or otherwise over the course of all twelve items. Participants generally found physical health questions easier to answer because the terms were considered more concrete. In contrast, mental health and role items appeared particularly problematic, because of the difficulty of the terms included in the items, as well as issues of mood and temporal variability. Items with two or more meanings were present, despite it being generally argued that items should only contain a single meaning (Gerber and Wellens, 1997). For example, some participants queried the juxtaposition of terms (RE items: "…feeling depressed or anxious"; MH04: "downhearted and depressed"; MH03: "calm and peaceful"), and examples which were considered not to be comparable (PF02: "Moderate activities, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf"). As a consequence, items were reinterpreted, as participants could choose how terms were defined and related to one another. Participants sometimes reported emotional reactivity, especially if an item was felt to underestimate the complexity of perceived physical or mental health and lifestyles (such as GH01 (general health): "it tells nothing about me, or my health", or MH04 (downhearted and depressed): "the question is very limited in categorising a complex emotion in this way"). Question ordering and layout were found to be important; question interpretation and response were influenced by earlier questions, and this occurred from very early in the questionnaire completion, for example, in relation to the Role-Physical and Role-Emotional items, which were perceived by a number of participants to be very similar.

When *retrieving and judging* their answers, participants found it easier to use question timescales if they related to specific events in their lives or their recent experience. Otherwise, they either ignored them or framed the response generally. Many participants did not use the examples contained in items, instead preferring to think of their own. Two main strategies were reported: a general perception (including general self-reflection, and comparisons with past health or an ideal) and specific experiences (recollection of previous situations or events). A general perception was reported more commonly by those without health problems, although it was also reported by participants with a chronic illness, such as HIV, both findings suggesting the presence of *self-health attitudes* or *self-schemata*, summary cognitions related to health (Markus, 1977; Sehulster, 1994). Those who reported specific situations were more likely to recall particular health problems and experiences of health limitations. There was also evidence for another comparison strategy, this time with other people. The strategy differed according to whether the comparison was upward (to someone healthier) or downward (someone less well) (Singer, 1994). In addition, some participants reported non-optimal processing of items (satisficing), whereby question detail was skipped and answers generated in relation to the available response options or conjecture, rather than experience or general perceptions (Krosnick et al, 1996).

Investigation of the *response stage* revealed that most items were considered to have vague quantifiers, which could generally only be understood according to the relative position of the categories. In this respect, there was evidence that the middle category was perceived to be an average (for example, described in relation to "fluctuation", "comes in waves", and "generally speaking"), and also that categories were ruled out systematically to leave a final selection. Some participants admitted that social desirability concerns affected thoughts about response, particularly for mental health items, although it did not seem to have influenced the final response.

In terms of the approaches to data collection, it was found that there was much more detail obtained from cognitive interviewing, particularly with retrospective probing, which included the ability to clarify issues. However, the different techniques have strengths and weaknesses. On the downside, it is not always straightforward to conduct a cognitive interview; participants must first learn to use the think-aloud technique and then to continue to use it while answering a questionnaire, which may take considerable effort for some participants. In addition, a large amount of information may be collected that is unrelated to the response process, due to the undirected nature of the technique. The adapted-PIQ is a much simpler approach, and it provides structured information about the response process. However, it focuses particularly on comprehension and broad response information, and any comments provided on the PIQ cannot be clarified with participants.

In conclusion, this study has shown that different cognitive methods and samples can provide rich and useful information on the response processes used when completing a health status measure. Of course, these data are only suggestive of the response processes and may not have reflected actual cognitive processes ongoing during the administration of the SF-12v2. A first noteworthy finding, however, is that both interview and self-completion provided useful, comparable data, leading to the decision to combine results when presenting the findings. Although the interview was the more detailed and thorough approach, many of the findings from the interviews were also suggested by the self-report data, and in terms of resources, self-completion measures are much less cost and time intensive. Another important finding is the limited number of judgement strategies reported, and the similarity in types of strategy between samples.

The success of the PIQ self-completion questionnaire, and the finding that only a limited number of retrieval / judgement strategies were reported in the qualitative (cognitive) study, led to the decision that it would be possible to construct a fixed-category response strategy item. Thus, the self-completion questionnaire used in the quantitative study that will be described in Part Three included a response strategy item alongside every SF-12v2 item.

Part Three: A quantitative study to investigate contextual influences, response process and the SF-12v2

Overview

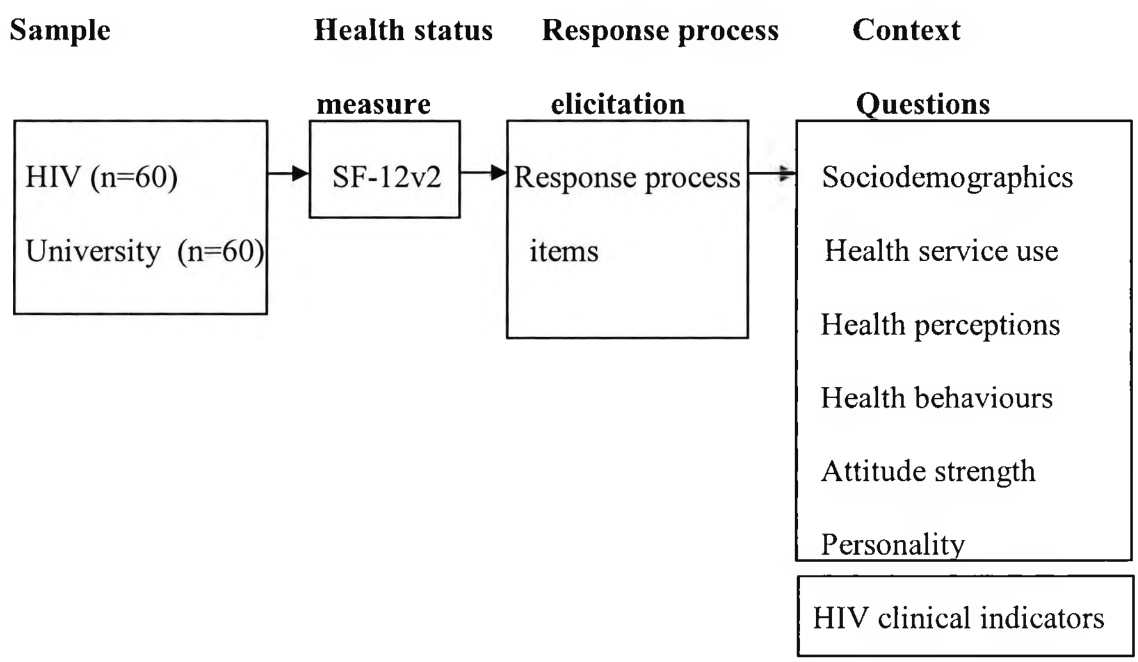
A quantitative study investigated SF-12v2 response in university and HIV samples. Although there were some differences in the psychometric performance of the SF-12v2 in the two samples, the scales were generally associated with contextual variables as predicted by earlier research (Part One). Response process elements (strategy, easiness and usefulness) could be measured meaningfully and were also associated with SF-12v2 scores, but patterns varied according to sample, possibly relating to the effects of health on response processes employed. A path model demonstrated that personality and objective health influenced physical and mental health scale scores, although direct and mediated pathways differed by outcome and sample.

3.1 Introduction

This study used quantitative survey data and techniques to conduct bivariate and multivariate statistical analyses to investigate the psychometric properties of the SF-12v2 scores, and associations between contextual influences, response processes and SF-12v2, culminating in a path model, linking all three. As with the qualitative study, a nominally healthy sample (university students and staff) and a health problem group (people with HIV) were recruited and the results compared. Figure 3.1 demonstrates the design of the study. Power calculations demonstrated that an optimal size of sixty respondents in each sample was necessary in order to detect group differences (See Appendix 16; Tabachnik and Fidell, 2001; Ware et al, 2004). A survey questionnaire was administered including a health status measure, response process elicitation items, and contextual items asking about a range of areas that might influence response (Appendix 17).

Different methods were used to recruit university and HIV samples; most university respondents were recruited by distribution of questionnaires within the university. In contrast, the HIV sample was recruited by advertisement and predominantly by post, along with personal recruitment. A reminder mail-out was sent to those who expressed an interest by post, using the mailing as an opportunity to conduct "snowball" sampling and recruit additional respondents known to the addressee. Data received was coded and input on a secure computer for analysis. The next section outlines the methods used in the study, including an introduction to the methodology employed, the development of the materials, piloting, ethical considerations, recruitment data collection and analysis strategy.

Figure 3.1: Quantitative study design



3.2 Methods

Development of the 'Health perceptions study' questionnaire

The survey was based on a self-completion questionnaire, designed to investigate the relationship between response processes, contextual factors, and actual response to the SF-12v2 health status measure. The areas included in the questionnaire were:

Response processes

Each SF-12v2 item was followed by three additional questions about the respondents' perceptions of the task, asking about the main response strategy used when answering (described below), along with perceptions of the task: easiness of each item to complete (considered to relate to the cognitive complexity), and usefulness of the item as a rating health (a proxy for the attitude strength indicator of importance). Easiness and usefulness were operationalised by the researcher, since there are no standard questions for either concept.

Previous research has shown that a single self-completion question can summarise response strategies employed when answering items from a personality measure (Gordon and Holden, 1996). The findings from the qualitative study were used to inform the design of an equivalent response strategy item for the SF-12v2, focusing on the retrieval/judgement stages: that is, formulating the answer to a question.

Two versions of the response process question were developed (Table 3.1), summarising the strategies revealed by the qualitative (cognitive) techniques. The wording of both questions was approved by members of the PhD steering group (Appendix 1), and colleagues with experience of social research techniques. As can be seen in Table 3.1, version A was the more complex, a multiple-choice question including seven closed response options, plus options for other and not sure. In contrast, version B included only three closed response options, plus options for other and not sure, and was forced-choice, requiring the respondent to select a single main process.

The questions were piloted with two City University students recruited by word-of-mouth. Both questions were understood and accepted, but for pragmatic reasons of questionnaire size and respondent burden, the simpler question was selected for inclusion in the self-completion questionnaire. To expand on this decision, a response process item was required for each of the twelve SF-12v2 items; the more detailed the list of strategies provided, the longer would be the questionnaire, and the more complex the process of selection would have been for the respondent. Therefore, although a longer list of options would have led to more finely differentiated strategies, it was considered that there was an attendant risk that this could have led to fewer returned questionnaires, more missing strategy data and non-optimal choices being made over the twelve strategy items. Strategy version B was found to be acceptable with respondents and appeared to include, albeit more crudely than version A, a comprehensive list of strategies, from general impressions to specific experiences and comparisons with others, along with the option for respondents to include a strategy of their own.

Table 3.1: Two possible response process questions

A. Did you consider any of the following when you were deciding how to answer this health question? *(tick as many as apply)* • Other people's health or well-being • Your current health or well-being • Your usual health or well-being • Your past health or well-being • A hypothetical idea of health or well-being (for example, perfect health, or as healthy as you would like to be) • Specific experiences • A general picture of yourself • *Not sure: the answer just came to me* • *Other (please specify)* **B. Which of the following comes closest to the way you answered this health question? Was your answer mainly based on… *(please choose a single option)*** • comparing yourself to others • a general picture of yourself • specific experiences • *Not sure: the answer just came to me* • *Other (please specify)*

Contextual factors

These were variables hypothesised to relate to health status, and therefore to potentially influence response processes. The relationship between contextual variables and health status was outlined in Part One (pp. 59-66). The contextual variables included were:

Attitude strength – a series of items relating to attitude strength dimensions, reflecting the centrality of self-health attitudes to the respondent (Sehulster, 1994), including intensity (concern about health, and change in concern), accessibility (frequency of thought about health), and identification (experience of health problems among family or friends). These items were written specifically for the questionnaire by the researcher.

General health – items on perceptions of general health and changing health, related also to self-health attitudes and processes such as adaptation (Heyink, 1993): retrospective health (over the course of the year), health compared to one year earlier, and health expectations for the year ahead. Health compared to one year earlier was taken directly from the SF-36 health status measure (Ware et al, 1993). The item on health expectations was similar in structure to this, but was written specifically for the questionnaire, with the timeframe adapted in terms of future health. Retrospective health was also written specifically for the questionnaire, and was similar to the item assessing general health in the SF-12v2, but with a more specific timeframe. Finally, the self-reported presence of long-standing illnesses, including a request to write the name of any illness, and to indicate whether the illness limited the performance of any activities. These questions were drawn from the Health Survey for England (Department of Health, 2001).

Health service contact and use – including self-reported level of contact with a GP surgery, hospital casualty or outpatient department, and experience as a hospital daycare patient, or as a hospital inpatient (measures of objective circumstances and the attitude strength dimension of direct experience). These questions were also based on the interview questions included in the Health Survey for England (Department of Health, 2001).

Health behaviours - Smoking behaviour (frequency and amount of cigarette smoking) and drinking behaviour (the CAGE questionnaire on problem drinking, plus frequency of drinking and amount of alcohol consumed). The CAGE questionnaire comprises four items asking the respondent whether they had ever felt like cutting down on their drinking, been annoyed by other people's criticism of their drinking, felt bad or guilty about their drinking, or had a drink first thing in the morning to steady their nerves or get rid of a hangover (Mayfield et al, 1974). The CAGE items were summed to form a single scale score. A scale score was calculated only if all items were answered, resulting in casewise deletion for analyses. The recommended cut-off score of two or more was adopted to indicate problem drinking (Mayfield et al, 1974). The frequency and amount questions were drawn from the Millennium Cohort Study parental interview questions (National Centre for Social Research, 2003).

Extraversion and Neuroticism – Two scales of the English Big Five Inventory (BFI) were included (John et al, 1991). Three of the eight Extraversion and three of the eight Neuroticism items were reverse scored and were recoded before they were summed to form the scales. Scores for a scale were not calculated if there were any missing data among the constituent items (casewise deletion).

Sociodemographic items – Age, sex, ethnicity and work and domestic role status were included, because they have been shown to be associated with health status and to allow characteristics of the samples to be compared. The questions were written specifically for the questionnaire, with the exception of ethnicity, which was taken from the 2001 Census of England household form (Office for National Statistics, 2001).

The HIV sample received four additional *questions on HIV/AIDS and treatment*, derived from questions used in the Internet and HIV study (Elford et al, 2004), including year of diagnosis, anti-HIV medication, recent viral load test and CD4 count. Plasma HIV viral load and CD4 T-cell counts are the two principal immunologic/virologic measures used to monitor disease progression and treatment efficacy for HIV/AIDS and it was therefore considered crucial to include them.

Two versions of the self-completion questionnaire were developed, one for each sample. Both were printed on double-sided A4 sheets, corner stapled, the front of which was a covering letter, followed by the questions on subsequent sheets. The covering letter was written on City University headed paper and explained the background to the survey, its general purpose and importance, assured confidentiality and anonymity, and gave an estimate of the time required to complete the questionnaire. The letter included contact details for the researcher.

Both questionnaires were eight pages in length, although the HIV sample questionnaire included additional questions on disease status, continued on the back page (the HIV sample questionnaire is included in Appendix 17. It differed from the university sample questionnaire only by the inclusion of the questions of HIV/AIDS and treatment). Previous research has shown that general health status questions should be included at the beginning of a survey in order to avoid potential influences from other questions (Bowling et al, 1999). Therefore, all respondents completed the SF-12v2 at the outset. Generally, sections were ordered to provide a simple flow of topics for the respondent.

Piloting of materials

The questionnaire was reviewed by members of the PhD Steering group (Appendix 1) and Dr Simon Barton, Medical Director of the HIV/GUM Directorate, Chelsea & Westminster Hospital, London. The questionnaire was also checked with colleagues before being piloted for understandability with four students (from the City University Graduate Social Research Methods programme) and four people with HIV (from Chelsea & Westminster Hospital). Selection of pilot participants was by word-of-mouth in the university and the hospital.

Ethical considerations, including confidentiality and anonymity

Since respondents were expected to provide personal information about themselves and their health, before either study could be conducted, a new research proposal had to be submitted to the ethics committee of City University for scrutiny, following the same guidelines used for the earlier qualitative study (Appendix 5). The submission was accepted in the winter of 2004. In order to preserve confidentiality and anonymity, the survey was entitled 'Health Perceptions Study', and did not explicitly identify HIV positive as the study group, only mentioning HIV in a specific filter section of four questions on clinical markers (Appendix 17). This prevented respondent HIV status being revealed accidentally during the mail-out phase of the study. All questionnaires were anonymous with no identifier linking returned questionnaires to the person they were mailed to. Data from questionnaires were input by the researcher on password-protected computer equipment. Only the researcher had access to the processed computer data and completed questionnaires were stored in a locked filing cabinet. The results of the study were not disclosed or published in a form by which any individual respondent could be identified. The computer data and questionnaires will be destroyed at the end of the study.

Survey recruitment

Different procedures were used in the two samples, due to the specific recruitment issues for each sample. Specifically:

University Procedure

Recruitment occurred in three ways, each carried out in turn in order to achieve the target sample of 60 respondents. The initial approach attempted to recruit from a dissertation meeting. The required sample was not achieved, and therefore a second approach recruited from within the Sociology Department. Since the sample was still less than required, respondents were additionally recruited from student e-mail lists (departmental and course mailing lists and the City University society of research students membership list). As a result of the sequential use of the three approaches, a large enough sample of respondents had been recruited for the analyses to begin. The details of each recruitment approach were as follows:

Recruitment from dissertation meeting

The researcher attended an MSc Social Research Methods dissertation briefing meeting and revision sessions. Potential respondents were told about the aims of the study, that it was anonymous and that it would take about 10 minutes to complete. The questionnaires were distributed and respondents were informed that they should return the questionnaire to a returns box, stored in a visible place during the questionnaire distribution. Copies of the "Helplines" sheet, providing contact details for support services available to all respondents, were placed next to the box.

Recruitment from Sociology Department

At other times the returns box, blank questionnaires and copies of the "Helplines" sheet were kept in a visible place in the City University Sociology Department administration office, frequented by all staff and students.

Recruitment from mailing lists

Details of the survey were sent to four e-mail lists: students on the Advanced Social Research Methods and Statistics MSc course; Sociology PhD students; Psychology PhD students; and the City University society of research students list. Copies of the questionnaire and "Helplines" were mailed to potential respondents.

HIV recruitment

Initially, it was hoped that recruitment could be conducted entirely within the Chelsea & Westminster Hospital HIV clinic. Previous work has indicated that those attending hospital outpatient clinics in the United Kingdom are broadly representative of the HIV population, since almost all HIV positive people attend such clinics for their treatment (Elford et al, 2004). However, due to the timing and complexity of the NHS ethical committee system, which would cause significant delay to the project, it was decided to try another approach to recruitment, involving a mail-out survey publicised in a national magazine for HIV positive people. However, a low response meant that the target sample of 60 respondents was not achieved, and therefore other approaches were adopted. Firstly, a reminder was sent to those who had already expressed an interest in the study, in order to increase response; the reminder letter was also used as an opportunity to recruit additional respondents through "snowball" sampling. Secondly, respondents were recruited face-to-face from the Kobler HIV Clinic, by making patients aware of the magazine advertisement and the survey. The Chelsea & Westminster HIV/GUM Directorate only offered this approach after the mail-out survey had been initiated, as a means to increase the number of respondents. However, only a single day was made available because of organisational issues in the clinic, particularly in relation to the large number of other research projects being undertaken in the clinic at the time. The details of these approaches are provided below:

Recruitment by advertisement

The first recruitment method involved the use of an advertisement describing the study briefly, including entry criteria and the procedure (Appendix 18). Potential respondents were requested to write to or email the researcher with their name and address in order to receive a questionnaire. They were also invited to contact the researcher if they wanted more information about the study. The advertisement emphasised confidentiality and anonymity and stated that there was no mention of HIV status in the questionnaire. Entry into a prize draw was used as an incentive to participate, although the initial information included in the advertisement made it clear that respondents should indicate whether or not they wished to take part in this.

The advertisement was used in three ways:

- Published as a quarter page in the January 2005 edition of Positive Nation magazine.

Positive Nation is a British magazine dealing with issues of interest to people living with HIV and AIDS. It was first published in 1995 and has since attained a readership of 50,000 people every issue, ten times a year. The magazine aims to appeal to a general

HIV positive population. The magazine is widely available free of charge in genitourinary clinics (health service centres specialising in sexual health), HIV centres and community organisations, and can also be delivered to readers by post. Positive Nation magazine was approached and agreed to take the recruitment advertisement for their readership.

- Copies of the advertisement were reproduced as an A3 poster to be placed on the walls of genitourinary clinics.
- An A4 recruitment "flyer" was produced, one half of which comprised the advertisement and the second half a detachable form including name, address and request for inclusion in the prize draw. Each flyer was stapled to a reply-paid envelope.

Both the poster and flyer forms of the advertisement were distributed to two HIV clinics (Kobler and Victoria Clinics, part of the HIV/GUM Directorate of Chelsea & Westminster Hospital), and two support organisations (Lighthouse West London, part of the Terence Higgins Trust HIV charity, and London East Aids Network, an HIV/AIDS support organisation). These materials were to be displayed in waiting rooms from January 2005.

Reminder / additional recruitment

Since there was no means of ascertaining whether those who expressed an interest in taking part returned a questionnaire, a courteous reminder was sent out in March 2005 to the 39 people who had registered their details by this time (Appendix 19). The letter was printed on City University headed paper and stressed the importance of returning the questionnaire, and included a return date. This letter was additionally designed to increase the sample by a form of "snowball" sampling, with the registered person asked if they could recruit someone else who matched the study entry criteria. The reminder pack included a spare questionnaire with a personal details form attached, including name, address, and interest in participating in the prize draw, together with a reply paid envelope. New recruits could register to take part in the survey by post or by e-mail.

Face-to-face recruitment

The final method of recruitment involved the researcher spending one day at the Kobler HIV Clinic in March 2005. The researcher explained the nature of the study to individual patients in the waiting area, and the health perceptions study questionnaire (together with the advertisement, personal details form, the "Helplines" sheet, and a reply-paid envelope) was given to anyone interested in taking part. Potential respondents could either complete the questionnaire while waiting or post it back in the reply-paid envelope.

Slow recruitment resulted in the extension of the final deadline for receipt of questionnaires, from 1st March 2005 to 23rd May 2005. Returned anonymous questionnaires were assigned an identity number as soon as they were received by the researcher, and this number was used for matching questionnaires to records on a Microsoft Access database designed for data entry. Paper copies of questionnaires were stored in a locked filing cabinet at City University.

University response

Using information available about distribution methods, it was possible to calculate an approximate response rate for two of the three methods of data collection (Table 3.2):

Table 3.2: Approximate response rate: University sample

Distribution method	Population N	Response N (%)
Dissertation meeting and revision sessions	70	36 (51.4%)
City University society of research students mailing list	258	15 (5.8%)

From this crude analysis, which does not take into account the overlap between different methods of recruitment, it was clear that the face-to-face data collection carried out in the dissertation meeting resulted in the best response.

A second analysis of response involved a description of the sources of sample recruitment (Table 3.3). Again, using this method, it was found that most respondents were recruited form the dissertation meeting and revision sessions. Most questionnaires were completed during an MSc dissertation briefing meeting and revision sessions.

Table 3.3: Returned completed questionnaires: University sample

Distribution method	N (%)
Dissertation meeting and revision sessions	36 (56.3%)
Posted in returns box (Sociology admin office)	13 (20.3%)
Interest generated from mailing lists	15 (23.4%)
TOTAL	**64 (100.0%)**

HIV Response

An approximate response rate from each of the methods of distribution could be calculated from data available on the distribution of Positive Nation magazine and the number of patients registered as attendees at the Kobler Clinic on the day when face-to-face recruitment was undertaken (Table 3.4). These figures do not include the possibility of overlap between sources of recruitment.

Table 3.4: Approximate response rate: HIV sample

Distribution method	Population N	Response N (%)
Estimated Positive Nation magazine readership per issue	50000	57 (0.1%)
Kobler Clinic attendees	40	15 (37.5%)

The approximate response rates show that, as with university recruitment, a far higher survey response could be achieved by face-to-face recruitment.

In addition, a description of the sources of sample recruitment was produced, and from this was calculated the proportion of responses received from each method of recruitment. 86 sets of personal details were documented on a password-protected Microsoft Word table used for recording information required for questionnaire mail-out and the selection of the prize draw winner. Of the 86 sets of personal details, 71 related to interest in survey participation; 49 (57.0%) had arrived by post and 22 (25.6%) via e-mail. In addition, 15 (17.4%) sets were provided by those personally recruited in the Kobler Clinic.

The researcher's initials were written in the corner of each reply-paid envelope, primarily to facilitate the return of completed questionnaires within the University.

However, the initials were written in three different colours: one colour was used on envelopes sent to those who posted or emailed their interest to the researcher; a second colour was used for the reminder mail-out; and a third colour was used for envelopes handed out in person at the Kobler clinic. 54 questionnaires were returned from the initial mail-out (76.1% of the 71 posted or emailed expressions of interest). One of these was not completed, with a comment explaining that the respondent had found SF-12v2 questions PF02, PF04, RP02 and RP03 confusing and that they did not therefore wish to carry on with the questionnaire. Therefore, the figure was reduced to 53 completed questionnaires (74.6% of 71). Of the reminder mailing, 4 questionnaires were returned (5.6% of 71), although it was not possible to distinguish new recruits from those who had been reminded. Combining the returned initial and reminder questionnaires resulted in an overall mail-out response of 58 (81.7%) of the expressions of interest (71), falling to 57 (80.3%) if only completed questionnaires were considered. 19 questionnaires were distributed by the researcher in the Kobler Clinic, of which 15 were returned, either to the researcher in the clinic or by post (78.9% of the Kobler Clinic questionnaires).

The final breakdown of completed questionnaires according to distribution method is shown in Table (3.5). The figure of 72 relates to the total of returned, completed questionnaires.

Table 3.5: Returned completed questionnaires: HIV sample

Distribution method	**N (%)**
Mail-out #1	53 (73.6%)
Mail-out #2	4 (5.6%)
Kobler Clinic	15 (20.8%)
TOTAL	**72 (100.0%)**

University and HIV sample sociodemographic composition

The sociodemographic variables included in the quantitative study were: gender; age; ethnicity; education level; and role occupation. Descriptive characteristics for the university sample are included in Tables 3.6 and for the HIV sample in Tables 3.7. The samples differed to some extent on most of the measured variables. The university sample was generally younger, comprised more women, was better educated, and more ethnically diverse (See Appendix 20).

Gender

Most university sample respondents were female whereas, in contrast, the majority of HIV respondents were male (Tables 3.6.1 and 3.7.1).

Age

Although both samples included a wide range of ages (university sample: 22 to 61 years; HIV sample: 27 to 71 years), the university sample was generally younger than the HIV sample (university sample: mean: 35.0 years; median: 32 years; mode: 29 years; HIV sample: mean: 43.9 years; median: 42 years; mode: 39 years) (Tables 3.6.1 and 3.7.1).

Ethnicity

The university sample included twice as many ethnic groups as the HIV sample (five versus ten). However, in both samples, 'white' ethnic group categories were predominant, selected by more than three-quarters of respondents in both samples. 'White-Other' respondents represented the largest single ethnic group among the university sample, whereas 'White-British' was the largest group in the HIV sample, comprising more than half of all respondents (Tables 3.6.1 and 3.7.1).

Education

Not surprisingly, all respondents in the university sample, recruited mainly from students and lecturing staff, were educated to tertiary level. Almost three quarters of respondents indicated that they had received post-graduate education. Respondents from the HIV sample had more diverse educational experiences, with education ranging from primary school to postgraduate level, although most respondents had received some form of education to further education level or beyond (Tables 3.6.2 and 3.7.2).

Role status

Work and domestic roles were assessed using a multiple-choice question. The array reported in the two samples varied considerably (Tables 3.6.3 and 3.7.3). While roles reported in the university sample were restricted to employment, study and caring for others, in the HIV sample, additional roles included long-term sick, unemployed and retired. The most commonly selected role in the samples also differed, with full-time employment selected by the majority of the university sample, and long-term sick the most frequently chosen option in the HIV sample. In both samples the total number of roles reported ranged between one and three, although the proportions differed between samples. The majority of respondents in both samples selected only a single role, but this was the case with a much larger proportion of the HIV sample (84.7% versus 57.1%). In the university sample, fifteen different single or combined roles were identified. Most single and combined roles involved working or studying. In the HIV sample, sixteen different role combinations were identified. As in the university sample, employment and study, singly or in combination, were reported. However, long-term sick and unemployed appeared among the major role combinations. Of note, 'long-term sick only' was the modal role in the HIV sample, and long-term sick was also included in one of the larger role combinations (along with employed part-time).

Table 3.6.1: Description of University sample sociodemographic characteristics: gender and age (N=60-64)

		N (%)	Mean (SD)
Gender:	Male	25 (39.7)	
	Female	38 (60.3)	
Age in years, mean (SD):			35.00 (10.9)
Age bands:	20-29	25 (41.7)	
	30-39	17 (28.3)	
	40-49	11 (18.3)	
	50-59	5 (8.3)	
	60-69	2 (3.3)	

Table 3.7.1: Description of HIV sample sociodemographic characteristics: gender and age (N=71-72)

		N (%)	Mean (SD)
Gender:	Male	58 (81.7)	
	Female	13 (18.3)	
Age in years, mean (SD):			43.92 (9.1)
Age bands:	20-29	3 (4.2)	
	30-39	23 (32.4)	
	40-49	27 (38.0)	
	50-59	15 (21.1)	
	60-69	2 (2.8)	
	70-79	1 (1.4)	

Table 3.6.2: Description of University sample sociodemographic characteristics: ethnicity and education (N=60-64)

		N (%)
Ethnicity:	White-Other	28 (44.4)
	White-British	18 (28.6)
	Black-African	4 (6.3)
	Black-Caribbean	3 (4.8)
	Asian-Other	3 (4.8)
	White-Irish	2 (3.2)
	Chinese	2 (3.2)
	Mixed-White/Asian	1 (1.6)
	Asian-Indian	1 (1.6)
	Other	1 (1.6)
Education:	Further education	7 (11.1)
	Higher education	11 (17.5)
	Post-graduate	45 (71.4)

Table 3.7.2: Description of HIV sample sociodemographic characteristics: ethnicity and education (N=71-72)

		N (%)
Ethnicity:	White-British	43 (59.7)
	White-Other	17 (23.6)
	Black-African	6 (8.3)
	White-Irish	4 (5.6)
	Other	1 (1.4)
Education:	Primary school	1 (1.4)
	Secondary education	10 (13.9)
	Further education	16 (22.2)
	Higher education	32 (44.4)
	Post-graduate	13 (18.1)

Table 3.6.3: Description of University sample sociodemographic characteristics: role status (N=60-64)

		N (%)
Role status:	Full-time employment	33 (51.6)
	Full-time study	25 (39.1)
	Part-time employment	14 (21.9)
	Part-time study	14 (21.9)
	Caring for home/family	9 (14.1)
Number of roles reported:	One role	36 (57.1)
	Two roles	22 (34.9)
	Three roles	5 (7.9)
Role combinations:	Employed full-time only	19 (30.2)
	Studying full-time only	13 (20.6)
	Employed full-time + Studying part-time	8 (12.7)
	Employed full-time + Studying full-time	4 (6.3)
	Employed part-time + Studying part-time	3 (4.8)
	Other (single or combined roles)	16 (25.4)

Table 3.7.3: Description of HIV sample sociodemographic characteristics: role status (N=71-72)

		N (%)
Role status:	Long-term sick	29 (40.3)
	Full-time employment	23 (31.9)
	Part-time employment	12 (16.7)
	Part-time study	5 (6.9)
	Caring for home/family	5 (6.9)
	Unemployed	5 (6.9)
	Full-time study	4 (5.6)
	Retired	2 (2.8)
Number of roles reported:	One role	61 (84.7)
	Two roles	9 (12.5)
	Three roles	2 (2.8)
Role Combinations:	Long-term sick only	24 (33.3)
	Employed full-time only	22 (30.6)
	Employed part-time only	6 (8.3)
	Unemployed only	4 (5.6)
	Employed part-time + long-term sick	3 (4.2)
	Other (single or combined roles)	13 (18.1)

Sexual orientation

One factor that was not included in the survey was sexual orientation. This was because it was considered a sensitive issue that might deter respondents from taking part. Although it would not necessarily have been an unusual question for a questionnaire distributed to people with HIV, it was felt to be potentially intrusive for the university sample, whose respondents were only aware that this was a study of "Health Perceptions". Nevertheless, evidence from the Kobler clinic patient cohort indicates that the majority of the HIV positive sample would have been homosexual, whereas the university sample would have been expected to be heterosexual. The influence of gender and sexuality may have influenced the results obtained, and therefore subsequent findings need to be considered in relation to measured sociodemographic differences, as well as the unmeasured potential influence of sexuality.

Prize draw

71 of the 86 people who provided personal details indicated that they wished to take part in the £100 prize draw (82.6%). The list of contestants was unrelated to participation in the survey. The contestant names were written on separate slips of paper that were placed in a box. The principal PhD supervisor selected the winner from among the entries and oversaw the prize payment from the ESRC Research Training Support Grant (RTSG), managed by City University on behalf of the researcher. Payment was by City University cheque, made out to the winner and sent to their postal address. The covering letter stated that they had won the Health Perceptions Study prize draw, with no mention of HIV/AIDS.

Recoding text answers

Free text was provided on some questions, where an ‘other’ response was allowed; specifically, response strategy, education level and economic activity questions. These were recoded as required.

Data quality checks

Questionnaire data were input into an MS Access data entry database. The researcher compared all database records with the original questionnaires, in order to check for data entry errors. When errors were identified, these were corrected on the database. Distributions of all variables were used as a second check of data entry errors. No out-of-range values were identified. Analyses were carried out to check for inconsistent entries (crosstabulating: long-standing illness versus limiting longstanding illness; health service use (ever) versus most recent health service use versus amount of contact in last year; smoking status versus amount smoked; drinking status versus amount drunk), although none was identified. All survey data were transferred to SPSS for analysis.

SF-12v2 scores

Algorithms in specialist software supplied by QualityMetric Incorporated were used to score the SF-12v2, resulting in eight scales and two overarching physical and mental health summary scales. The software converted the scales into norm-based scores (NBS), with a standardised mean and standard deviation, relative to the 1998 United States general population norms (Ware et al, 2004). Since the US norms have been linear transformed to T-scores, with a mean of 50 and a standard deviation of 10, scores that are above or below 50 can be interpreted as being above or below the population norm. Computed NBS scales were transferred to SPSS for analysis.

Statistical software

Three statistical software packages were used for the analyses:

- Univariate and bivariate descriptive and inferential statistical analyses were performed using SPSS version 12.0 for Windows (SPSS, 2003).
- Differential item functioning was assessed using MPlus version 4.1 statistical modelling program (Muthén and Muthén, 2006).
- Path modelling was carried out using the graphics component of the AMOS 4 structural equation modelling package (Arbuckle and Wothke, 1999).

Strategy for statistical analysis

Analyses included univariate, bivariate and multivariate techniques. Initial analyses were carried out to show the data distributions and the scale properties. Classical and modern psychometric analyses were conducted to investigate scale properties, particularly in relation to the SF-12v2. Additional analyses were carried out to investigate the relationship between the response process, contextual factors and SF-12v2 scores. Finally, statistical path modelling was used to bring together response process and contextual influence in relation to the SF-12v2 scores. All analyses were conducted to allow comparisons between the results obtained in the two samples. Missing data were handled on a casewise basis, with respondents excluded if they had missing observations for any variable included in an analysis.

There were three fundamental methodological issues that were considered prior to the analyses; the appropriateness of parametric testing with health status data; the use of standard asymptotic tests for crosstabulation analyses with the available research samples; finally, multiple comparisons and significance level. These are discussed in more detail:

The use of parametric versus non-parametric statistics

Parametric tests are generally preferred to non-parametric tests when:

- dependent variable data are measured on an interval scale
- there is homogeneity of variance on a dependent variable across all levels of the independent variables
- the underlying population data for the dependent variable are normally distributed, generally considered to be the case when the sample data is normally distributed

However, these criteria are not always met in analyses. For example, scores achieved on scales are often treated as having interval-like qualities although they are measured using questionnaire item responses that have simply been summated. Health status scale scores in a general population tend to be skewed towards good rather than poor health. Analyses of the SF-12v2 data for both university and HIV samples showed that the data are negatively skewed, indicating the predominance of high scores (good health) over low scores (poor health). This is also reflected in the population data available for the SF-12v2 (Ware et al, 2004). Nevertheless, parametric techniques are considered to be robust and reliable, even when the criteria are not met (Pagano, 1998). In addition, parametric techniques are more powerful than equivalent non-parametric tests and also permit complex multivariate analyses. In initial bivariate analyses carried out for this thesis, both parametric and non-parametric tests were carried out in parallel. The results from both approaches were largely identical, and for clarity, only parametric analyses are reported.

The use of exact tests versus standard asymptotic tests in crosstabulation analyses

Standard asymptotic analysis compares a calculated test statistic against a hypothetical distribution, such as chi-square. However, such testing requires a large dataset, and in asymptotic crosstabulation analysis there should be enough observations to ensure that the expected value of every cell is above one and is greater than five in at least eighty percent of cells (Siegel and Castellan, 1988). In contrast, analyses carried out using the university and HIV samples result in tables that are sparse and unbalanced. In this situation, asymptotic tests may produce misleading results, although variables may be recoded to decrease the number of cells, thereby increasing the expected values in the remaining cells (Siegel and Castellan, 1988). Exact tests provide an alternative analytical approach (Mehta and Patel, 1996). In crosstabulation analysis using an exact test, significance is calculated directly from the data rather than in relation to a test distribution, through the calculation of the probability of having a table with the observed, or a more extreme, pattern of data, due to chance in relation to all possible patterns of given the observed marginal totals. For these reasons, exact test results are reported for the crosstabulation analyses carried out in this thesis. For two-by-two tables, Fisher's exact test is reported while for a larger table, an extension to this test, the Freeman-Halton test, is reported (Mehta and Patel, 1996). Additional analyses were carried out using recoded variables in order to reduce the number of cells and show more clearly the nature of the relationships between variables.

Multiple comparisons and significance level

A significance level of five percent was required for rejection of the null hypothesis in these analyses. However, a large number of comparisons were carried out and it is generally considered that, in these circumstances, a more stringent level of significance should be sought, due to the increased likelihood that significant results may be achieved by chance (Bland and Altman, 1995). As in other exploratory research (for example, Goodwin and Friedman, 2006), one of the main objectives was to identify patterns of results rather than proving statistical significance in each case. Therefore, it was also important to avoid Type II error, missing a true effect. In addition, less powerful non-parametric tests were carried out in parallel to the parametric tests, where possible, and, in general, the results from both sets of analyses were in agreement. Nevertheless, it is possible that some of the significant results reported resulted from Type I errors, rejecting the null hypothesis when it is true, and small effects need to be considered with caution.

Components of analyses

1. Item descriptives

Frequencies were calculated for all scale items (SF-12v2, Big 5, Easy items, Useful items) and other items (response strategy, sociodemographics and clinical indicators, general health, attitude strength, health behaviours and health services contact) to examine item distributions (see Appendix 20).

2. Psychometric evaluation of measures (SF-12v2; Big Five Neuroticism and Extraversion; Easiness; Usefulness)

A range of techniques were used to investigate the psychometric properties of the scales, including item and scale analyses, tests of reliability and validity, and, for the SF-12v2, statistical modelling of differential item functioning between samples:

- Completeness of data (an indicator of data quality) (Ware et al, 2004)
- Item facility (mean score for an item) (Rust and Golombok, 1999)
- Reliability (Cronbach's alpha) (Cronbach, 1951)
- Construct validity: factor structure (principal component analysis) (Ware et al, 1993)
- Construct validity: convergent and discriminant validity (item-item correlation; item-scale correlation; item component correlation) (Ware et al, 1993)
- Scale distribution analyses (kurtosis; skewness; Shapiro-Wilk test) (Kerr et al, 2002)
- Comparisons of sample SF-12v2 scale scores with population norms and published HIV sample data (z test) (Delate and Coons, 2000; Pagano, 1998)
- Between-item variability for Easiness and Usefulness scales (repeated measures analysis of variance and Friedman non-parametric analysis of variance: indicating that ratings differed according to item) (Nevo, 1985; Siegel and Castellan, 1988)

- Relationship between Easiness and Usefulness scale scores and overall ratings (parametric and Friedman non-parametric analyses of variance) (Pagano, 1998; Siegel and Castellan, 1988)
- Differential Item Functioning (DIF), using Multiple-Indicator-Multiple-Cause (MIMIC) modelling to assess item invariance between the two sample in relation to SF-12v2 latent physical and mental health factors (Muthén, 2002). This was a form of statistical modelling, and the model fit was determined using methods outlined below

3. Analyses of the SF-12v2 and contextual factors

Relationships were tested between SF-12v2 scores and contextual variables (Sociodemographics; General health; Attitude strength; Personality; Health behaviours; Health service contact) using both parametric and non-parametric techniques:

- Two-category contextual variables (such as gender) were assessed using the parametric t-tests and non-parametric Mann-Witney U tests (Pagano, 1998; Siegel and Castellan, 1988)
- Three or more category contextual variables (such as education level) were tested with parametric and Friedman non-parametric analyses of variance (Pagano, 1998; Siegel and Castellan, 1988). Post-hoc pairwise comparisons between groups in the parametric ANOVA were carried out using Tukey HSD post hoc pairwise comparisons (Winer, 1971)
- The relationship between SF-12v2 scales and other continuous variables (such as age) was assessed using Pearson and Spearman correlations (Pagano, 1998)

4. SF-12v2 and response process questions

The relationship between individual SF-12v2 and response process questions (that is, response strategy, easiness and usefulness rating) were examined, using exact tests due to the characteristics of the data, described earlier:

- Fisher's exact (for two-by-two tables) or χ^2 test (larger tables) to assess independence between variables, in order to determine the likelihood that two variables are related (Siegel and Castellan, 1988)
- Strength of association between unordered variables, or when one variable was unordered, was assessed using the phi (for two-by-two tables) and Cramer's V statistics (larger tables); for two ordered variables, the equivalent analysis was undertaken with the Tau-b statistic (Everitt, 1992)

SF-12v2 and response process items were recoded for some analyses, in order to more clearly identify patterns of association

5. Response process modelling

Finally, a path analytic approach (Wright, 1934) was adopted to investigate the relationship between some of the individual contextual variables, response process strategy and the SF-12v2 physical and mental summary scores, PCS and MCS. A simple model, based an earlier published model (Brief et al, 1993), investigated the comparative influences of personality and objective circumstances on subjective well-being, including attitude strength and strategy as mediating factors. Four versions of the model were tested, separately for the two samples and the PCS and MCS scales, using statistical techniques outlined below to assess the relationship between the model and the data.

Statistical modelling: "fit" statistics

In statistical modelling, including CFA and path analysis, a model is tested to identify whether it adequately accounts for the data. To do this, three standard model-fitting statistics are employed: χ^2, RMSEA and CFI. Each tests the model fit in a different way, and the pattern of results informs the selection or rejection of the model. All goodness-of-fit measures are to some extent a function of sample size and degrees of freedom. Most take into account both the fit of the model and also model complexity in terms of the degrees of freedom of the model. χ^2 tests the null hypothesis that the model adequately accounts for the data, in terms of whether the covariances among items specified in the model are significantly different to the observed covariances in the data. If there is a significant difference between the model and the data, the proposed model should be rejected. However, the χ^2 test is sensitive to sample size, and for this reason is seldom reported in isolation. RMSEA (Root Mean Square Error of Approximation) is an estimate of discrepancy between the data and the model, per degree of freedom in the model. The closer the data to the model, the smaller the value of RMSEA. CFI (Comparative Fit Index) compares the hypothesised model with a baseline model of independence among observed variables, adjusted for degrees of freedom. CFI ranges between 0 - 1, with a larger value indicating a better fit. A good model fit is indicated by a non-significant chi-square, an RMSEA of less than 0.05 and a CFI (Comparative Fit Index) over 0.95 (Byrne, 2001). As these statistics are sensitive to sample size, they are only used to indicate model fit.

The next stages in this study, described in the following section, involved the substantive statistical analyses, starting with the classical and modern psychometric analysis of the SF-12v2 survey data from the two samples, followed by the analysis of the contextual factors, response processes, and the final path model linking the elements of the analysis together to investigate a simple model of the response process.

4 Psychometric properties of the SF-12v2

Overview

The following section refers to a psychometric analysis of the SF-12v2. First, a classical psychometric framework was used to investigate the reliability and validity of the SF-12v2 in the university and HIV samples. Second, a modern psychometric approach modelled item response in order to identify the extent of any variation according to sample (differential item functioning).

Results from classical psychometric analyses suggested that the SF-12v2 scales were psychometrically reliable and valid. Facility scores indicated that there were more positive responses in the university sample, especially for physical health, and more variation in scoring in the HIV sample. Reliability coefficients for the scales where it was possible to compute Cronbach's alpha (at least two items) were good.

Construct validity was generally supported, with evidence for the existence of physical and mental health dimensions. However, there were some differences between the samples in terms of the underlying structure of the SF-12v2. Tests of convergent/discriminant validity indicated that: i. there were more significant inter-item correlations between the twelve items in the HIV sample, suggesting that both physical and mental health were more strongly related in that sample; ii. items were more strongly correlated with the hypothesised physical or mental summary scale, but there were also notable correlations with the alternative scale, especially in the HIV sample. Principal component analyses of the scales showed different patterns between samples, with item loading on two components as hypothesised for the university samples (physical health and mental health), but a less clear-cut structure in the HIV sample.

As expected, scales were not normally distributed in the university sample, being instead skewed toward higher scores, indicating better health. There was less evidence of skewness in the HIV sample, signifying more varied health status among these respondents. There were lower mean mental health scores in the university sample than the general US population norms, but comparable physical health. The HIV sample had significantly lower scores than the US population norms for both physical and mental health. These findings could relate to the physical or mental health effects of HIV, or be a consequence of the composition of the two samples, which differed according to sociodemographic variables. Significantly lower scores were shown in the HIV sample on all scales compared to the university sample.

In order to investigate whether items comprising the scales were invariant between samples, a differential item function (DIF) analyses was conducted. The results suggested that the Bodily Pain item (BP02) performed differently between the two samples: BP02 responses were not as severe as could be expected in the HIV sample, given the overall physical health scores achieved. This could be due to the adaptation to the effects of chronic disease, or differentially worse effects of HIV on the other aspects of physical health measured. However, the sample sizes were small for DIF analyses and the results are only tentative.

Can the reliability and validity of a standard health status measure, the SF-12v2, be demonstrated for both healthy and health problem groups, using a classical psychometric approach?

The key concepts of classical test theory are reliability and validity (Rust and Golombok, 1999). A reliable measure is measuring something consistently, while a valid measure is measuring what it is supposed to measure. The following section includes item analyses and an investigation of reliability and validity in relation to the SF-12v2.

Item completeness

The distributions of individual SF-12v2 items for the university and HIV samples are included in Appendix 20. Results indicated that the SF-12v2 had good item completion. The majority of respondents completed all twelve items comprising the eight SF-12v2 scales (Table 4.1). In the university sample, ten items were missing in total (missed by five respondents), giving a completeness of data score of 98.7% for the dataset; and in the HIV sample, two items were missing (two respondents), resulting in a completeness of data score of 99.8%, both higher than the 90% score recommended by the test developers (Ware et al, 2004). The requirement for complete data on all scales led to a sample size 59 for the MCS and PCS summary scales in the university sample and of 70 in the HIV sample.

Table 4.1: Proportion of respondents with scale scores: University and HIV samples

Scale name	Item Total	University sample N (% of sample)	HIV sample N (% of sample)
PF	2 items	64 (100.0%)	72 (100.0%)
RP	2 items	61 (95.3%)	72 (100.0%)
BP	1 item	63 (98.4%)	70 (97.2%)
GH	1 item	64 (100.0%)	72 (100.0%)
VT	1 item	64 (100.0%)	72 (100.0%)
SF	1 item	64 (100.0%)	72 (100.0%)
RE	2 items	60 (93.8%)	72 (100.0%)
MH	2 items	63 (98.4%)	72 (100.0%)
PCS	**6 items**	**59 (92.2%)**	**70 (97.2%)**
MCS	**6 items**	**59 (92.2%)**	**70 (97.2%)**

Legend:
PF: Physical Functioning / RP: Role-Physical / BP: Bodily Pain / GH: General Health / VT: Vitality / SF: Social Functioning / RE: Role-Emotional / MH: Mental Health / PCS: Physical Component Scale / MCS: Mental Component Scale

Item analysis

Item facility

Facility of an item in a person-based test, such as SF-12v2, refers to the mean item score, reflecting whether or not, on average, responses are located at either extreme of the options available. Items that elicit extreme responses are generally considered to function poorly (Rust and Golombok, 1999). The facility score (mean) of each SF-12v2 item in the university and HIV samples is shown in Table 4.2. In this analysis, four items were recoded (GH01, BP02, MH03 and VT02: specified by (r) in the Table), so that high scores represented good health. In both samples, most facility scores were close to the middle category or just above, indicating a tendency towards positive ratings of physical and mental health, on average. None of the items attained an extreme facility score, although physical health facility scores for the university sample approached the ceiling level (representing no physical limitations). For all items, the facility scores in the university sample were higher than in the HIV sample, indicating better reported health status. The most extreme limitation response was not selected by any university respondents for four of the five physical health items (PF02, PF04, RP02, RP03, GH01) and one of the mental health items (RE02). In addition, standard deviations were larger in the HIV sample, indicating a wider range of scores: for every item, all response categories were selected in the HIV sample. These results can also be seen from the frequencies of the individual items presented in Appendix 20.

Table 4.2: SF-12v2 item comparisons: University and HIV samples

Item	University sample Mean (SD)	HIV sample Mean (SD)	Possible range	$\chi 2$ Significance (Exact tests)	Cramer's V
Physical health					
PF02	2.92 (0.27)	2.40 (0.74)	1-3	$\chi^2(2)=24.11$***	0.42
PF04	2.78 (0.42)	2.19 (0.78)	1-3	$\chi^2(2)=24.21$***	0.42
RP02	4.05 (1.04)	3.17 (1.18)	1-5	$\chi^2(4)=19.44$***	0.38
RP03	4.29 (0.93)	3.31 (1.31)	1-5	$\chi^2(4)=21.72$***	0.40
BP02 (r)	4.25 (1.03)	3.71 (1.30)	1-5	$\chi^2(4)=9.09$	0.26
GH01(r)	3.93 (0.72)	3.11 (1.16)	1-5	$\chi^2(4)=21.60$***	0.40
Mental health					
VT02 (r)	3.25 (0.85)	2.71 (1.08)	1-5	$\chi^2(4)=11.52$*	0.29
SF02	4.02 (1.18)	3.22 (1.19)	1-5	$\chi^2(4)=16.67$**	0.35
RE02	3.70 (1.01)	3.15 (1.17)	1-5	$\chi^2(4)=11.46$*	0.29
RE03	4.21 (0.81)	3.51 (1.20)	1-5	$\chi^2(4)=15.97$**	0.35
MH03 (r)	3.24 (1.01)	2.79 (1.17)	1-5	$\chi^2(4)=8.10$	0.25
MH04	3.78 (0.95)	3.07 (1.03)	1-5	$\chi^2(4)=18.62$***	0.37

Legend:
Moderate activity (PF02) / Climb several flights of stairs (PF04) / Accomplished less (RP02) / Limited in kind of activities (RP03) / Pain impact (BP02) / Health in general (GH01) / Lot of energy (VT02) / Social impact of health/well-being (SF02) / Accomplished less (RE02) / Did work less carefully (RE03) / Felt calm (MH03) / Felt downhearted (MH04)

Comparison of patterns of item response between samples

Responses to individual items in the two samples were compared using χ^2 exact tests, with Cramer's V tests for the strength of association between sample and response (Table 4.2). For most items, there was a strong and significant association. However, the relationship did not reach significance for two items: BP02 (Bodily Pain) and MH03 (felt calm), indicating that, although HIV respondents reported worse health status than university respondents for both items, the response distributions of the university and HIV samples did not differ enough to reach significance at the five percent level.

In summary, SF-12v2 responses varied, particularly in the HIV sample, although there was less variation for physical health items in the university sample, indicating generally good physical functioning. Differences in response patterns between the samples resulted in significant associations for ten of the twelve items.

Reliability

The reliability of the two-item scales and the overall physical and mental summary scales was measured using Cronbach's alpha, a test of internal consistency based on the average inter-item correlation (Cronbach, 1951). The developers of the SF-12v2 cautioned against conducting internal consistency analysis of the scales for two reasons. First, all scales consist of only one or two items. Second, since items were selected because they possessed unique reliable variance in estimating physical or mental health, internal consistency estimates of reliability are considered to underestimate SF-12v2 scale reliability (Ware et al, 2004). With these caveats it mind, the results of the reliability tests showed that most scales with two items demonstrated reasonable internal consistency, close to the α=0.70 level considered acceptable (Nunnally, 1978). In the university sample, acceptable internal consistency was found for three scales: RP (α=0.89), RE (α=0.67), MH (α=0.75). However, one scale had a low alpha coefficient: Physical Functioning (α=0.39), indicating that responses to the two items were not strongly related in this sample. In the HIV sample, good internal consistency was demonstrated for all two-item scales: PF (α=0.85), RP (α =0.83), RE (α=0.70), and MH (α =0.71). Internal consistency was also carried out for the two profile scales, each comprising six items. Cronbach's alpha coefficients were good for both PCS (university sample: α= 0.74; HIV sample: α=0.89) and MCS (university sample: α= 0.85; HIV sample: α=0.87), suggesting that, despite the small numbers of items, the scales were generally homogeneous.

The results from both samples suggested that most scales demonstrated good internal consistency, despite the small number of items.

Validity

Construct validity was assessed according to tests of convergent and discriminant validity and an assessment of the underlying component structure of the SF-12v2 scales.

Convergent / discriminant validity

Convergent validity involves establishing that items and scales that should be related are strongly associated (i.e. physical health items correlated with other physical health items and scales). In contrast, discriminant validity involves demonstrating that those items and scales that should not be related are not strongly associated (i.e. weaker correlations between physical and mental health items and scales).

Tables 4.3 and 4.4 show the inter-item and item-summary scale correlation matrices for the SF-12v2 items in the university and HIV samples. Very different patterns were identified in each sample. In the university sample, 63.6% (42/66) of the sixty-six pairwise correlation coefficients were significant, compared to 94.5% (63/66) in the HIV sample. In general in the university sample, with some exceptions (see the role limitation and vitality items), physical health items correlated more strongly with one another than with mental health items, and vice versa. In the HIV sample, there were generally higher correlations, and no clear distinction between physical and mental health items in terms of the strength of association between items.

Convergent and discriminant validity were also demonstrated between items and mental and physical health summary scales (MCS and PCS), since items correlated more strongly with the scales to which they were proposed to relate than with the alternative scale. The correlations testing convergent validity were generally higher in the HIV sample than the university sample, particularly for physical health. Additionally, as expected from the inter-item correlations, strong and significant correlations were identified between most items and both physical and mental health summary scales in the HIV sample, and the coefficients for item RP02 were of a similar magnitude for both the PCS and MCS (0.60 and 0.57, respectively).

Findings from both the inter-item and the item-scale correlations suggest that the HIV sample were more likely to report poor functioning in multiple domains in comparison to more discrete problems in the university sample.

Table 4.3: SF-12v2 correlation matrix: University sample

	PF02	PF04	RP02	RP03	BP02 (r)	GH01 (r)	VT02 (r)	SF02	RE02	RE03	MH03 (r)	MH04
PF02	-											
PF04	.27*	-										
RP02	.36**	.17	-									
RP03	.35**	.30*	.81***	-								
BP02 (r)	.42**	.06	.49***	.45***	-							
GH01 (r)	.28*	.17	.31**	.20	.28*	-						
VT02 (r)	.22	.20	.60***	.60***	.45***	.34**	-					
SF02	-.10	.04	.36**	.41**	.31**	.17	.53***	-				
RE02	.27*	.35**	.36**	.38**	.06	.19	.48***	.37**	-			
RE03	.08	.05	.30**	.22	.27*	.08	.35**	.32**	.51***	-		
MH03 (r)	-.05	-.03	.17	.28*	-.08	.20	.50***	.56***	.52***	.32**	-	
MH04	-.01	.28*	.21	.31**	.14	.13	.36**	.59***	.59***	.40**	.61***	-
PCS	**.62***	**.36**	**.63***	**.60***	**.72***	**.50***	.36**	-.01	-.05	-.09	-.22	-.16
MCS	-.07	.09	.27*	.34**	.11	.17	**.62***	**.79***	**.73***	**.59***	**.85***	**.82***

*: p<0.05; **: p<0.01; ***p<0.001

Legend:
Moderate activity (PF02) / Climb several flights of stairs (PF04) / Accomplished less (RP02) / Limited in kind of activities (RP03) / Pain impact (BP02) / Health in general (GH01) / Lot of energy (VT02) / Social impact of health/well-being (SF02) / Accomplished less (RE02) / Did work less carefully (RE03) / Felt calm (MH03) / Felt downhearted (MH04)

Table 4.4: SF-12v2 correlation matrix: HIV sample

	PF02	PF04	RP02	RP03	BP02 (r)	GH01 (r)	VT02 (r)	SF02	RE02	RE03	MH03 (r)	MH04
PF02	-											
PF04	.74***	-										
RP02	.57***	.38**	-									
RP03	.76***	.58***	.71***	-								
BP02 (r)	.76***	.65***	.57***	.70***	-							
GH01 (r)	.60***	.51***	.52***	.63***	.57***	-						
VT02 (r)	.50***	.40***	.59***	.65***	.37**	.62***	-					
SF02	.47***	.30**	.57***	.72***	.49***	.54***	.58***	-				
RE02	.43***	.34**	.60***	.61***	.45***	.55***	.58***	.63***	-			
RE03	.44***	.48***	.54***	.58***	.48***	.39**	.38**	.55***	.54***	-		
MH03 (r)	.26*	.11	.52***	.42***	.11	.38**	.52***	.47***	.47***	.22	-	
MH04	.41***	.25*	.53***	.52***	.29*	.49***	.64***	.66***	.75***	.44***	.56***	-
PCS	**.88*****	**.81*****	**.60*****	**.78*****	**.89*****	**.72*****	.50***	.46***	.39**	.44***	.12	.26*
MCS	.24*	.11	.57***	.54***	.18	.47***	**.69*****	**.77*****	**.80*****	**.56*****	**.73*****	**.86*****

*: p<0.05; **: p<0.01; ***p<0.001

Legend:
Moderate activity (PF02) / Climb several flights of stairs (PF04) / Accomplished less (RP02) / Limited in kind of activities (RP03) / Pain impact (BP02) / Health in general (GH01) / Lot of energy (VT02) / Social impact of health/well-being (SF02) / Accomplished less (RE02) / Did work less carefully (RE03) / Felt calm (MH03) / Felt downhearted (MH04)

Exploratory and Confirmatory factor analyses

Factor analysis is a method for summarising the relationships between observed variables, based on the assumption that much of the variability can be explained by fewer underlying factors (Tabachnik and Fidell, 2001). There are two basic types of factor analysis, both of which were used in the psychometric analyses described in this thesis. Exploratory factor analysis (EFA) is a method of data reduction in which variation shared between observed variables is inferred to relate to underlying factors. The structure of the EFA is determined by the relationships between the variables and is unconstrained by any hypothetical relationships between items and latent factors. This contrasts with confirmatory factor analysis (CFA), in which the relationships between items and latent factors are specified *a priori* and tested formally using statistical procedures. CFA is commonly used in modern psychometrics, where explicit measurement models are specified in order to investigate complexity in item response (Bjorner et al, 2003; Fleishman and Lawrence, 2003). Initially, an EFA was undertaken in order to emulate analyses carried out during the development of the SF measures, assessing the relationships between the scales in the two samples (Ware et al, 1993). The CFA approach was used later in the analysis of differential item functioning.

Exploratory factor analysis

Principal component analysis (PCA), an EFA technique, was used to further test the construct validity of the SF-12v2. Following the procedure outlined by Ware et al (1993), two principal components were extracted from the correlations among the eight SF-12v2 scales. These were rotated orthogonally to produce components that were unrelated to one another. An orthogonal solution was considered acceptable since the summary physical and mental health scales were only weakly correlated in both samples (University sample: r=0.25; HIV sample: r=0.13).

Table 4.5: Hypothesised associations between SF-12v2 scales and results from principal component analysis: University sample

	Hypothesised association		Rotated Principal Components			Relative Validity	
Scale	Mental	Physical	Mental	Physical	h^2	Mental	Physical
PF		X	-.03	.74	.55	0.00	1.00
RP		X	.39	.70	.64	0.15	0.89
BP		X	.14	.70	.51	0.02	0.89
GH		X	.11	.58	.35	0.01	0.62
MH	X		.91	.03	.82	1.00	0.00
RE	X		.71	.24	.56	0.50	0.11
SF	X		.84	.07	.71	0.71	0.00
VT	X		.64	.54	.70	0.41	0.53

Legend:
X : Hypothesised association

h^2: Proportion of the total variance of each scale explained by the two components

PF (Physical Functioning) / RP (Role Physical) / BP (Bodily Pain) / GH (General Health) / MH (Mental Health) / RE (Role Emotional) / SF (Social Functioning) / VT (Vitality)

The orthogonal simple structure for the *university sample* is summarised in Table 4.5, along with the hypothesised association of scales to the major dimensions (Ware et al, 2004). The two-component solution accounted for 60.5% of the total variance, comprising 32.7% and 27.7% of the variance respectively. The column labelled h^2 shows the proportion of the total variance in each scale explained by the two components. With the exception of the GH scale (0.35), more than half of the variance in each scale was accounted for by the two components. This means that for the GH scale, much of the variance is unexplained by physical and mental health defined by the two components. The components were labelled Mental Health (MH) and Physical Health (PH) since they were close to the hypothesised structure. The MH scale was most strongly associated with the first component and weakly associated with the second. In contrast, the PF scale was strongly associated with the second component and weakly with the first. Other scales also loaded most strongly on the hypothesised components.

The final two columns of the Table show the Relative Validity of each scale, or proportionately how much less valid each scale is as a measure of subjective physical or mental health, relative to the two most valid scales (the scales which are most strongly correlated with each principal component) (McHorney, 1993). Relative validity is calculated by dividing the common factor variance of each scale (the square of the scale-component correlation) by the scale with the greatest common factor variance. In the university sample, the scales with highest relative validities tended to be those that were hypothesised to be associated with the component. The PH component was best measured by the PF scale and mental health with the MH scale. The relative validity results for the university sample show that RE was half as valid as the most valid mental health scale, MH, while VT was 41% as valid. In contrast, physical scales were almost as valid as the PF scale (with the exception of GH). The VT scale was slightly more than one half less valid than the PF scale, indicating that it was performing as a measure of both physical and mental health. This is acceptable, since the VT and GH scales are considered to relate to both physical and mental health, although more strongly to the hypothesised components (Ware et al, 2000).

Table 4.6: Hypothesised associations between SF-12v2 scales and results from principal component analysis: HIV sample

	Hypothesised association		Rotated Principal Components			Relative Validity	
Scale	Mental	Physical	Mental	Physical	h^2	Mental	Physical
PF		X	.23	.88	.82	0.06	0.91
RP		X	.66	.63	.83	0.53	0.47
BP		X	.18	.92	.87	0.04	1.00
GH		X	.54	.59	.64	0.35	0.41
MH	X		.91	.03	.84	1.00	0.00
RE	X		.71	.44	.70	0.61	0.23
SF	X		.78	.34	.70	0.73	0.14
VT	X		.78	.30	.70	0.73	0.11

Legend:
X : Hypothesised association

h^2: Proportion of the total variance of each scale explained by the two components

PF (Physical Functioning) / RP (Role Physical) / BP (Bodily Pain) / GH (General Health) / MH (Mental Health) / RE (Role Emotional) / SF (Social Functioning) / VT (Vitality)

The equivalent analysis for the *HIV sample* is summarised in Table 4.6. The two-component structure accounted for more of the total variance (76.2%). Both components accounted for more of the variance than the equivalent in the university sample (41.8% and 34.4%, respectively), and more than half of the variance in every scale was explained by the two components. However, the hypothesised association was not fully reflected in the correlations between scales and the two components. As predicted, the mental health scales loaded most strongly on the MH component. Three of the four physical health scales loaded most strongly on the appropriate PH component. The role functioning (RP and RE) and GH scales loaded strongly on both components. In terms of relative validity, the most valid physical component scale was BP, albeit that it was less than 10% more valid than the PF scale. Both RP and GH were considerably less valid physical health scales, both more than 50% worse than the most valid, and RP was a slightly more valid indicator of MH. For mental health, higher relative validities tended to occur for scales that were hypothesised to be associated with the mental health component. The least valid of these (RE) was 39% as valid as the MH scale, but no other mental health scale was notably less valid than the MH scale. The two role functioning scales and GH were valid for physical and mental health components, although only RP had strong validity for both components, and GH is considered to measure physical and mental health (Ware et al, 2000).

The two component orthogonal simple structure of the scales accounted for a large proportion of the scoring variance in both samples. The components were labelled Mental Health (MH) and Physical Health (PH) since they were close to the hypothesised structure. However, the structure was more distinct in the university sample; in the HIV sample, more of the items loaded on both physical and mental health components, suggesting that physical and mental health items were more strongly associated in this sample, confirming the findings of the earlier correlations.

In summary, both the tests of convergent / discriminant validity and principal component analyses indicated that the correlational structure of the SF-12v2 differed to some extent between the samples, and that the structure found in the university sample was closer to that suggested by the test developers. Nevertheless, these findings are only speculative: the analyses were based on two small samples and would require confirmation with larger samples.

Scale analyses

Scale distributions

All scales were standardised by norm-based scoring in relation to the distribution of 1998 US general population norms (Ware et al, 2004). The US general population mean score for scales has been linear transformed to have a mean of 50 and a standard deviation of 10. Therefore, scores can be interpreted in relation to average health status, with those scores above and below 50 respectively higher and lower than the population mean.

Table 4.7: Scale descriptive statistics: University and HIV samples

	PF	RP	BP	GH	VT	SF	RE	MH	PCS	MCS
University sample										
Mean	53.92	49.67	49.84	50.47	50.27	46.63	44.71	46.35	53.54	44.29
25th ptile	50.03	43.36	47.25	44.74	47.75	36.37	39.30	40.16	50.22	39.31
50th ptile	56.47	52.57	57.44	55.52	47.75	46.47	44.90	46.25	54.62	46.98
75th ptile	56.47	57.18	57.44	55.52	57.81	56.57	50.49	52.35	57.76	51.77
SD	4.76	8.52	10.51	7.77	8.60	11.87	8.80	10.76	7.12	11.17
N	64	62	63	64	64	64	60	63	59	59
HIV sample										
Mean	44.42	40.93	44.34	41.57	44.81	38.62	37.44	39.31	45.12	38.04
25th ptile	39.29	34.14	37.06	29.65	37.69	26.27	28.12	29.49	33.57	30.23
50th ptile	47.88	38.75	47.25	44.74	47.75	36.37	36.51	40.16	50.27	37.28
75th ptile	56.47	47.96	57.44	52.83	57.81	46.47	44.90	46.25	54.66	47.94
SD	12.20	10.57	13.23	12.50	10.87	12.01	11.61	11.50	12.07	11.60
N	72	72	70	72	72	72	72	72	70	70

Legend:
PF (Physical Functioning) / RP (Role Physical) / BP (Bodily Pain) / GH (General Health) / VT (Vitality) / SF (Social Functioning) / RE (Role Emotional) / MH (Mental Health) / PCS (Physical Component Scale) / MCS (Mental Component Scale)

Table 4.7 shows descriptive statistics for the distributions of the SF-12v2 scales for the two samples. These figures identify generally higher scores among university respondents, as well as a narrower variation in scores compared to the HIV sample.

In a general population, SF measure scale scores tend to be negatively skewed, with high scores (indicating better health status) more likely than low scores (Ware et al, 2004). More university than HIV respondents attained the maximum possible score (ceiling) and fewer received the minimum score (floor). At least 46% of university respondents achieved the ceiling score (on three of the four physical health scales (PF, RP and BP), and one of the mental health scales (SF)). This level was not reached in the HIV sample for any scale. The floor score was attained on one physical health scale (BP) and three of the four mental health scales (VT, SF, and MH) in the university sample. However, less than 5% of university respondents had such a low score on any scale. In contrast, in the HIV sample the floor score was attained on all scales (except the summary scales), and by more than 10% of respondents on two physical health scales (PF and GH) and one mental health scale (VT).

For both mental and physical health, the distributions were generally negatively skewed, particularly in the university sample, indicating the preponderance of higher scores already highlighted. The significance of scale skewness was tested by dividing skewness statistics by their respective standard errors. Skewness was more evident in the university sample, particularly with regard to physical health scales. In the university sample, significant negative skewness was identified for all physical health scales (PF: $z=-5.80$, $p<0.001$; RP: $z=-2.50$, $p<0.05$; BP: $z=-4.83$, $p<0.001$; GH: $z=-2.10$, $p<0.05$), and two mental health scales (SF: $z=-3.13$, $p<0.005$; MH: $z=-2.70$, $p<0.01$). The two summary scales were also significantly negatively skewed: PCS ($z=-2.39$, p<0.05) and MCS (z=-2.35, p<0.05). In the HIV sample, significant negative skewness was identified for only two of the physical health scales (Physical Functioning: z=-2.27, p<0.05; Bodily Pain: -2.06, p<0.05) and none of the mental health scales.

The kurtosis of the distribution was also investigated, dividing kurtosis statistics by their respective standard errors (Zkurt) in order to test for significance. Significant positive kurtosis was demonstrated for two of the physical health scales in the university sample (Physical Functioning: z=3.69, p<0.001; Bodily Pain: z=2.22, p<0.05), indicating a peak to both distributions. In contrast, in the HIV sample all scales were negatively kurtotic, indicating generally flatter scoring distributions. However, significant negative kurtosis was only demonstrated for the PCS distribution (z=-1.96, p=0.05). Finally, the Shapiro-Wilk test was applied to scales. The results corroborated the previous analyses: in both samples, the scale distributions deviated significantly from normal. In the university sample, all distributions (with the exception of the MCS) were found to be significantly different from normal (p<0.05 and below). Similarly, in the HIV sample, all distributions were found to be significantly different from normal (p<0.01 and below).

Comparison with normative data

One-sample z tests were used to compare scale means for both samples with those from the United States general population, the normative data available at the time the analyses were conducted. The results, which can be interpreted in terms of the deviation of the sample means from the normative distribution, are shown in Table 4.8.

Table 4.8: Comparison of SF-12v2 University and HIV samples and the United States population normative data: z test results

	US norms Mean (SD)	University sample Mean (SD)	HIV sample Mean (SD)	University Z score	HIV Z score
PF	50 (10)	53.92 (4.76)	44.42 (12.20)	3.14**	-4.73***
RP	50 (10)	49.67 (8.52)	40.93 (10.57)	-0.26	-7.70***
BP	50 (10)	49.84 (10.51)	44.34 (13.23)	-0.13	-4.74***
GH	50 (10)	50.47 (7.77)	41.57 (12.50)	0.38	-7.15***
VT	50 (10)	50.27 (8.60)	44.81 (10.87)	0.22	-4.40***
SF	50 (10)	46.63 (11.87)	38.62 (12.01)	-2.70**	-9.66***
RE	50 (10)	44.71 (8.80)	37.44 (11.61)	-4.10***	-10.66***
MH	50 (10)	46.35 (10.76)	39.31 (11.50)	-2.90**	-0.59
PCS	50 (10)	53.54 (7.12)	45.12 (12.07)	3.03**	-3.81***
MCS	50 (10)	44.29 (11.17)	38.04 (11.60)	-4.00***	-9.72***

* $p<0.05$ ** $p<0.01$ *** $p<0.001$

Legend:
PF (Physical Functioning) / RP (Role Physical) / BP (Bodily Pain) / GH (General Health) / VT (Vitality) / SF (Social Functioning) / RE (Role Emotional) / MH (Mental Health) / PCS (Physical Component Scale) / MCS (Mental Component Scale)

The tests revealed that mean scores from both samples deviated from the normative data to some extent, although particularly in the HIV sample. In the university sample, mean scores for three physical health scales (RP, BP and GH) and one mental health scales (VT) were not significantly different from the normative data. Significantly higher scores were observed for one physical health scale (PF), and lower for three mental health scales (SF, RE, MH). These results were reflected in the results for the summary scales: mean PCS was significantly higher and MCS significantly lower than the normative data. In contrast, in the HIV sample, all scales (except MH) were significantly lower than the normative scores.

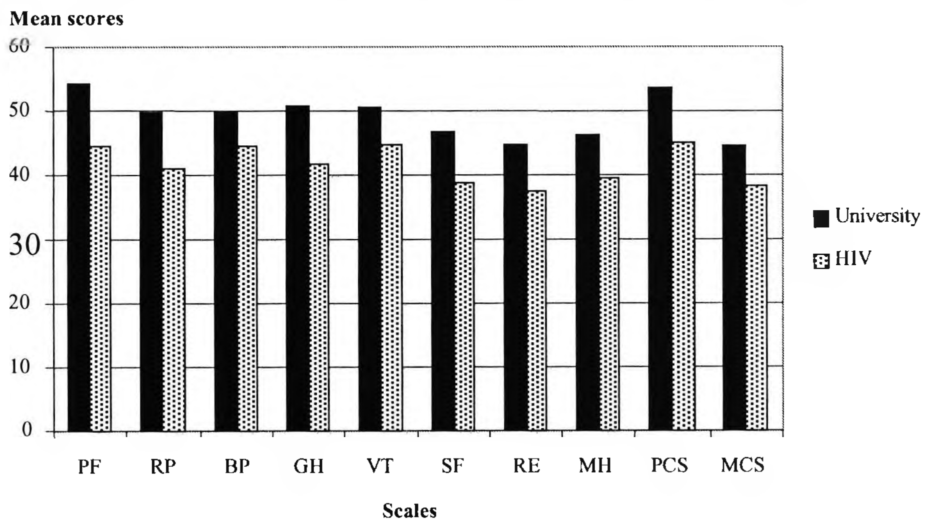
Since the population norms included a wider age range (18-75+ years) than either sample, analyses were repeated using normative data for specific age bands, in order to control to some extent for age-related differences in health status. In the university sample, more than half the respondents were aged less than 34 years. 42% of the sample was aged between 25-34 years (the largest single age group) and therefore the normative age band 25-34 years was selected for comparison. The pattern of results did not differ greatly from that found in comparison with general population norms, except for a reduction in physical health differences: only one physical heath scale, RP ($z=-2.32$, $p<0.05$), was significantly lower in the university sample. For mental health scales, mean scores were significantly lower for SF ($z=-2.80$, $p<0.01$), RE ($z=-4.80$, $p<0.001$) and MCS ($z=-2.88$, $p<0.01$). The mental health results shown, using both the general and age-banded norms, demonstrate poorer mental health in the university sample, compared with the population data, possibly a consequence of the sociodemographic characteristics of the sample. 49% of the HIV respondents were aged between 35 and 44 years and therefore sample scores were compared with normative scores for this age group. The earlier results were replicated, indicating consistently worse scores among the HIV sample.

A final comparison was carried out between the current HIV survey data and published results from a US HIV positive community sample (Delate and Coons, 2000). Unfortunately, the published study used the original SF-12 measure, which, due to lack of scoring precision, only provides scores on the-PCS and MCS summary scales. There was no significant difference between the two samples for the PCS, indicating that the physical health scores of the two samples were comparable. However, the published mean score for the MCS was significantly lower than in the current HIV sample ($z=-4.15$, $p<0.001$). Some caution is required in these comparisons, first because the published data were based on an older version of the SF-12, which, according to the test developers, does not differentiate health states as successfully as version 2 (Ware et al, 2004); second the comparison data are five years older than the data from the current study, during which time treatment regimens and expectations may have changed.

University and HIV sample scale comparisons

Having compared sample and normative data, a next step was to investigate the relationship between sample and scale scores. Figure 4.1 shows the comparative scale scores of HIV and student and staff samples. Lower mean scores, indicating worse subjective health, were observed in the HIV sample for all scales: PF ($t(94.29)=6.10$, $p<0.001$); RP ($t(132)=5.22$, $p<0.001$); BP ($t(129.07)=2.67$, $p<0.01$); GH ($t(120.50)=5.05$, $p<0.001$); Vitality VT ($t(132.27)=3.26$, $p=0.001$); SF ($t(134)=3.91$, $p<0.001$); RE ($t(128.90)=4.09$, $p<0.001$); MH ($t(133)=3.66$, $p<0.001$); PCS ($t(114.52)=4.91$, $p<0.001$); MCS $t(127)=3.10$, $p<0.01$).

Figure 4.1: SF-12v2 scale score comparison: University and HIV samples



In summary, the scoring distributions for mental and physical health scales in both samples deviated from normal, although this is a common finding for SF scales (Ware et al, 2004). University respondents reported generally good physical health status, in line with normative population data. In contrast, for mental health status the sample generally scored below population norms. HIV respondents reported poorer physical and mental health status in comparison with the normative data. HIV sample MCS and PCS scales were also compared with published SF-12v1 data from another sample of HIV positive people, with results suggesting similar physical health but significantly worse mental health in the current sample. While differences from published norms reflect the particular health concerns of the respective samples, they may be influenced by the sociodemographic profiles of each sample, including the age and gender composition. Finally, the overall scale scores differed significantly between the samples, with lower scores attained in the HIV sample compared to the university sample. The concern that the differences in scores between the two samples reflected the influence of differential item functioning will be addressed in the following section.

The last section provided evidence for the reliability and validity of the SF-12v2 scales in the two samples, and demonstrated significantly lower scores in the HIV sample on all SF-12v2 scales. The next research question utilised a modern psychometric framework to investigate whether, despite this, items performed differently in relation to the underlying constructs they were proposed to be measuring according to an exogenous variable, in this case, sample membership:

Are responses to items in a health status measure invariant, regardless of health experiences?

Differential Item Functioning (DIF)

Scores achieved on scales should be equivalent between groups being compared in order that results obtained are meaningful, rather than a possibly spurious product of the measure used. For this to be the case, respondents in different groups should have an equal probability of responding to an item if they are at the same level of the underlying construct being measured. However, evidence from a range of settings outlined in Chapter One indicate that non-scale factors can lead to differences in interpretation and response to survey items, and that items do not necessarily function identically between respondents. In addition to cognitive analyses of survey response processes, statistical modelling can be used to estimate the impact of Differential Item Functioning (DIF) on scale scores.

For example, a population study based on the original SF-12 indicated that DIF was particularly evident for mental health items (Fleishman and Lawrence, 2003). Initial differences according to race and age were influenced by DIF, whereby higher mental health (MCS) scores for black respondents were reduced substantially and a pattern of better mental health (MCS) among older than younger respondents was reversed. In contrast, lower mental and physical health (MCS and PCS) for women and those with less education were reduced in magnitude but not eradicated following the inclusion of DIF.

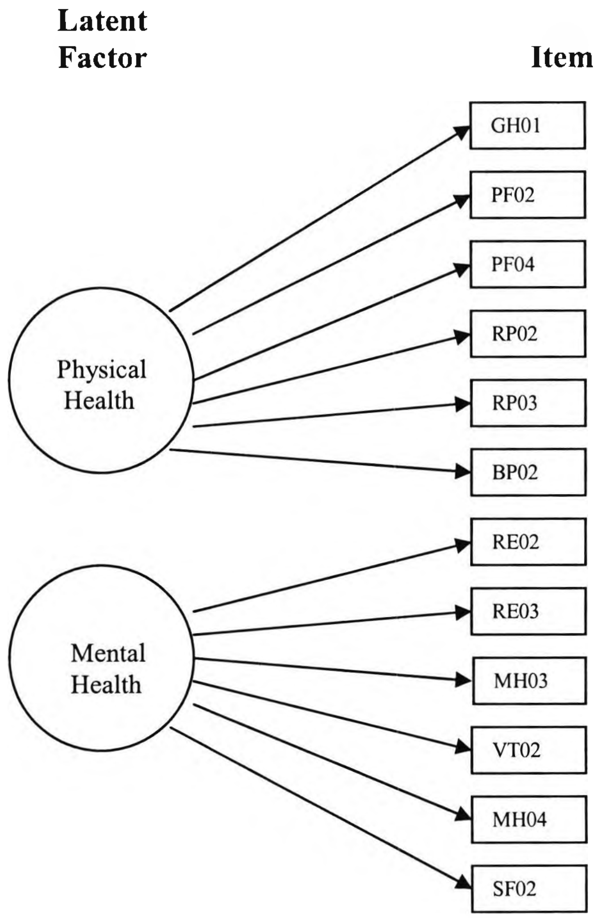
Estimating DIF effects using the MIMIC model

The approach adopted for estimating DIF effects in this research is Multiple-Indicator-Multiple-Cause (MIMIC) modelling, which is an extension of Confirmatory Factor Analysis (CFA) (Muthén, 2002). Although there is a range of approaches available for estimating the effects of DIF (Teresi, 2001), such as analyses of contingency tables, logistic regression procedures and methods based on item response theory (IRT), MIMIC models have an advantage in being easily able to accommodate the two-dimensionality of the SF-12v2.

Confirmatory Factor Analysis

In the earlier psychometric analyses, Exploratory Factor Analysis (EFA) was used to summarise the data using statistical criteria, without reference to prior theory (Ware et al, 1993). The principal component analyses (PCA) identified that two dimensions generally related to physical and mental health, although more clearly in the university sample, and could explain most of the variance in the SF-12v2 scales in the two datasets. A CFA analysis, in contrast, requires a clearly specified measurement model. Specifically, an *a priori* model is postulated in which one or more latent factors are hypothesised to relate to observed variables, and the "fit" of this model, whether it adequately describes the data, is subsequently tested statistically.

Figure 4.2: Two-factor latent model for the SF-12v2



Legend:
Health in general (GH01) / Moderate activity (PF02) / Climb several flights of stairs (PF04) / Accomplished less (RP02) / Limited in kind of activities (RP03) / Pain impact (BP02) / Accomplished less (RE02) / Did work less carefully (RE03) / Felt calm (MH03) / Lot of energy (VT02) / Felt downhearted (MH04) / Social impact of health/well-being (SF02)

In the measurement model for the SF-12v2, two latent factors were proposed: physical and mental health, which are hypothesised to influence responses to the twelve observed items. Here the CFA was applied to the combined dataset from the university and HIV samples, which would form the basis of the later modelling of DIF. The CFA model proposed two latent health factors: *Physical Health*, comprising the six primarily physical health items (GH01 PF02 PF04 RP02 RP03 BP02); and *Mental Health*, comprising the remaining, mental health items (RE02 RE03 MH03 VT02 MH04 SF02) (illustrated in Figure 4.2).

The fit of this two-factor model was tested to identify whether it adequately accounted for the combined dataset. A good model fit is indicated by a non-significant chi-square test result, an RMSEA (Root mean Square Error of Approximation) of less than 0.05 and a CFI (Comparative Fit Index) over 0.95 (Byrne, 2001), described earlier. As these statistics are sensitive to sample size, they were only used to indicate model fit. In this context, the overall model fit for the combined university and HIV samples data was reasonable for a small sample: $\chi^2(21)= 78.50$, $p<0.001$; CFI=0.965; RMSEA=0.146 and therefore this model was adopted for further analyses.

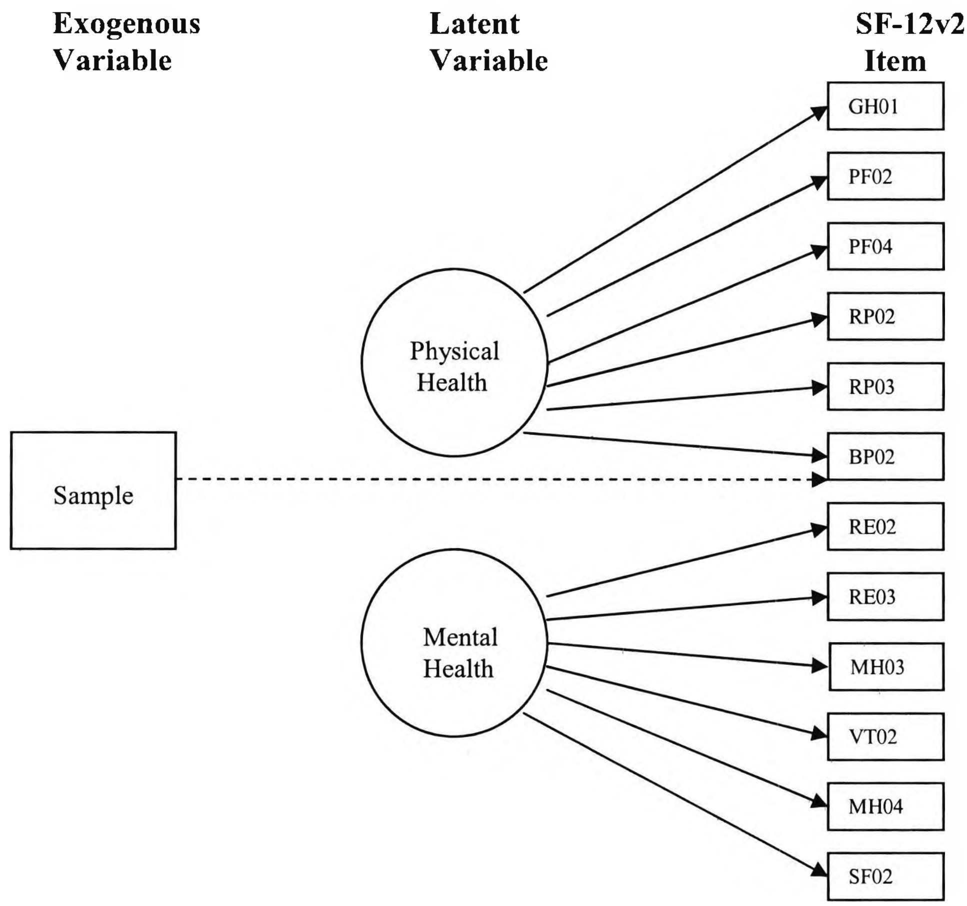
The proposed latent physical and mental health factors do not correspond exactly to the SF-12v2 physical and mental health component scales, PCS and MCS, produced by published scoring algorithms (Ware et al, 2004). For simplicity, the CFA model restricted pathways, so that the physical health latent variable was associated only with physical health items and the mental health latent variable with the mental health items. In contrast, the SF-12v2 PCS and MCS are the weighted product of all twelve items, so that both scales incorporate contributions from physical and mental items, to a greater or lesser extent. Nevertheless, as is evident from the names of the scales, PCS and MCS were derived predominantly from either physical or mental items. In order to test the relationship between a scale comprising only physical or mental items and the PCS and MCS, two scales were derived by summing only SF-12v2 physical health or mental health items. These scales were correlated with the PCS and MCS, and the results revealed strongly significant associations (PCS with physical health scale, $r=0.94$, $p<0.001$; MCS with mental health scale, $r=0.95$, $p<0.001$), suggesting that there was a strong correspondence between the SF-12v2 component scales and the hypothesised latent factors.

Using a MIMIC model to identify DIF

A MIMIC model incorporates one or more observed exogenous variables that are assumed to influence the underlying latent factors, and estimates the extent to which each exogenous variable directly influences a constituent scale item without mediation by the latent factors (Muthén, 2002). In the first stage of the current analysis, already described, a two-factor latent model for the SF-12v2 was identified. In the second stage, the effects of the exogenous variable (university or HIV sample membership) on the two latent health factors were estimated (a No DIF model). This involved regressing physical health and mental health latent factors on the endogenous variable, so that the result revealed the influence of sample membership on the latent factors. The third stage incorporated *direct* effects of the endogenous variable on items, representing systematic differences in item response according to sample membership, controlling for the latent factors (a DIF model). Finally, a χ^2 difference test was used to formally compare the statistical fit of the two models, to establish whether there was a significant improvement in model fit when DIF was estimated (Muthén and Muthén, 2006).

The DIF model (the third stage) included only significant direct effects of sample membership on SF-12v2 items. Initially, the effects of the exogenous variable on ten of the twelve SF-12v2 items were modelled simultaneously. Only ten items were included as, in order to identify the model, one variable for each latent factor had to be assigned as having no DIF. These reference variables were identified as having the smallest DIF effects on the basis of a comparison between the No DIF model and a series of twelve models including a direct effect of the endogenous variable on each item. The physical and mental health items with the smallest amount of DIF were RP02 and VT02, respectively.

Figure 4.3: MIMIC Model



Legend:
Health in general (GH01) / Moderate activity (PF02) / Climb several flights of stairs (PF04) / Accomplished less (RP02) / Limited in kind of activities (RP03) / Pain impact (BP02) / Accomplished less (RE02) / Did work less carefully (RE03) / Felt calm (MH03) / Lot of energy (VT02) / Felt downhearted (MH04) / Social impact of health/well-being (SF02)

Non-significant effects were removed iteratively; starting with the least significant, until only significant direct effects of the endogenous variable on SF-12v2 items remained in the DIF model. This was carried out using the z value of the parameter estimate for each direct effect (that is, comparing the value of the test statistics (z) with values from the known theoretical probability distribution) to identify those that were non-significant. None of the direct effects was significant at the standard five percent level. Therefore a ten percent significance level was adopted, in order to be able to identify notable patterns in the results. Following this approach, the final DIF model included only the single direct effect between sample membership and item BP02 (Bodily Pain) ($z=1.70$, $p<0.10$) (illustrated in Figure 4.3).

Comparing the No DIF and DIF models, the DIF model was a slightly better fit, indicating that the inclusion of the exogenous variable improved the overall model fit (No DIF model: $\chi^2(27)=86.98$, $p<0.001$; CFI=0.956; RMSEA=0.131; DIF model: $\chi^2(26)=83.87$, $p<0.001$; CFI = 0.957; RMSEA 0.131). This was confirmed using the χ^2 test for the difference between models, which was significant at the ten percent level: $\chi^2(1)=2.92$, $p<0.10$. The DIF model also explained more of the variance for the physical factor (no DIF model: $R^2 = 0.23$; DIF model: $R^2 = 0.26$), although there was no difference for the mental health factor (both no DIF and DIF models: $R^2 = 0.15$).

The effect of DIF on patterns of results between university and HIV samples

Results from both the no DIF and DIF models supported the earlier findings, based on PCS and MCS scores, that HIV respondents had significantly lower physical and mental health than the university respondents. In the no DIF model, HIV respondents were shown to score slightly less than one standard deviate below the scores of university respondents on the latent physical health factor (standard deviate=-0.96) and more than three quarters of a standard deviate below university respondents on the latent mental health factor (standard deviate=-0.78). However, allowing for DIF revealed a greater difference between the samples on the physical factor, with HIV respondents now scoring more than one standard deviate lower than university respondents (standard deviate=-1.01). There was no difference between no DIF and DIF models on the latent mental health factor. The parameter estimate for the direct effect of sample membership on item BP02 indicated that HIV respondents scored 0.31 standard deviates higher than university respondents on item BP02, given the underlying physical factor scores. Thus, these respondents appeared to rate bodily pain less severely than would be expected, taking into account underlying physical health status assessed with other items.

Conclusions from DIF analyses

The findings from the MIMIC modeling show that HIV respondents had worse perceived physical and mental health scores than university respondents, on average, both before and after adjusting for DIF related to membership of the HIV sample. In fact, the DIF effect appeared to suppress the magnitude of the discrepancy in physical health between the two samples. Comparing the findings for sample differences indicates that the ratings of Bodily Pain among the HIV respondents were not as severe as would be expected given the overall physical health determined from the other items. BP02 ratings inflated the physical health score for the HIV sample; when the DIF model adjusted for this, the mean difference in physical health between the samples widened.

Other findings lend support to these results. The smallest mean scale difference between the two samples was identified for the Bodily Pain scale, derived from item BP02, and there was no significant difference between samples in the responses provided on BP02. An additional analysis was carried out to confirm the limited influence of BP02 on the relationship between sample and physical health, measured using the SF-12v2 PCS. A correlation of $r=-0.39$ ($p<0.001$) between sample and PCS was not attenuated when BP02 was added as a control variable ($r=-0.35$, $p<0.001$); in contrast the strength of the correlation was greatly reduced when other SF-12v2 physical health items were controlled. This finding confirms that BP02 did little to differentiate HIV and university respondents on physical health.

There were some limitations of the analyses resulting from the limited size of the combined university and HIV samples. First, a sample size of 129 respondents is small for confirmatory factor analyses and this may have had some influence on the results attained. Nevertheless, it provided an opportunity to investigate patterns that could indicate differential item functioning, which could be subsequently analysed using a larger sample. Second, an assumption in this type of analysis is that the underlying factor structure is the same for both HIV and university samples, so that DIF is tested as the mean difference in item responses between samples for the same level on the factors. A larger sample would allow this assumption to be tested by performing a multi-group factor analysis on the covariance matrices of the two samples to assess similarities and differences in the factor structure of the SF-12v2 for the university and HIV samples.

Summary

In both the HIV and university samples, the SF-12v2 was found to be reliable and valid, demonstrating internal consistency, and convergent and discriminant validity. However, there were differences between samples, with a number of notable associations between physical and mental health scales in the HIV sample, indicating that diverse aspects of health were related. There is also other evidence that the psychometric properties of SF measures differ according to health and sociodemographic characteristics of respondents (Seymour et al, 2001; Ware et al, 2004). This should not be surprising since the findings of validation studies are not necessarily transferable between samples, although varying psychometric properties does not necessarily imply that a measure is invalid.

Indeed, there was evidence of scale construct validity in comparisons with population, HIV data and between samples: scale scores were significantly lower in the HIV sample compared to both the general population norms (as expected from previous findings using the SF-12v1: Delate and Coons, 2000), and the nominally healthy university sample. It should be noted that the mental health scores of the university sample were also poor in comparison to population figures, which could reflect the sociodemographic composition of the sample; for example, the majority of university respondents were female and there is evidence that females report poorer health status than men (Franks et al, 2003).

One consequence of the two-sample approach to this study was that scoring ranges differed between the two samples. The HIV sample reported a wider range of SF-12v2 responses than the university sample. This might indicate that the university sample was generally healthy according to the SF-12v2 scales, and also illustrate the heterogeneity in the effects of HIV on health, both findings particularly in physical health. The restricted range of SF-12v2 scores in the university sample to some extent limits discussion of patterns of results since few reported health limitations.

Examining SF-12v2 responses for evidence of differential item functioning, comparing the university and student samples, indicated some variation in item response between the two samples. Patterns of scoring differed between HIV and university samples, particularly for the Bodily Pain item, which received a better rating in the HIV sample than would be expected, considering responses to the other physical items, and which inflated the overall physical score attained by this sample in comparison to the university sample. However, this analysis was undertaken with a sample that was smaller than normally used and would need replication. In any case, DIF effects were weak and did not alter the substantive scale differences between samples, which indicated the expected relationship between HIV status and health scores.

The third research question integrates the SF-12v2 within a broader framework:

What is the relationship between contextual factors (namely, individual factors, such as health experiences, perceptions, and behaviours, personality and sociodemographics), response processes and the answers given to questions about health status?

This question is addressed in the following sections, beginning with a description of the relationship between the contextual variables and response in the two samples in order identify variables that are significantly associated with the SF-12v2. This is followed by similar analyses for response processes. These analyses provide a framework for the development of a path model for SF-12v2 scales, incorporating context and response strategy.

5 Contextual factors and the SF-12v2

Overview

The notion that individual contextual factors may influence both response processes and final response was introduced in Part One (pp. 59-66). Evidence was provided for potential associations involving the cognitive framework for self-health attitudes (Petty and Krosnick, 1995; Sehulster, 1994; Markus, 1977, 1983); health experiences (May and Warren, 2001; Warnecke et al, 1996); personality (Goodwin and Engstrom, 2002; Korotkov and Hannah, 2004); sociodemographic factors, including age, gender, ethnicity, education (Franks et al, 2003); health behaviours (Schmitz et al, 2003; Friedman et al, 1999); and employment status (Bartley, 1993). In addition, the identification of predicted relationships between context variables and health status is conceived to provide support for the construct validation of the SF-12v2, from a classical psychometric perspective (Rust and Golombok, 1999).

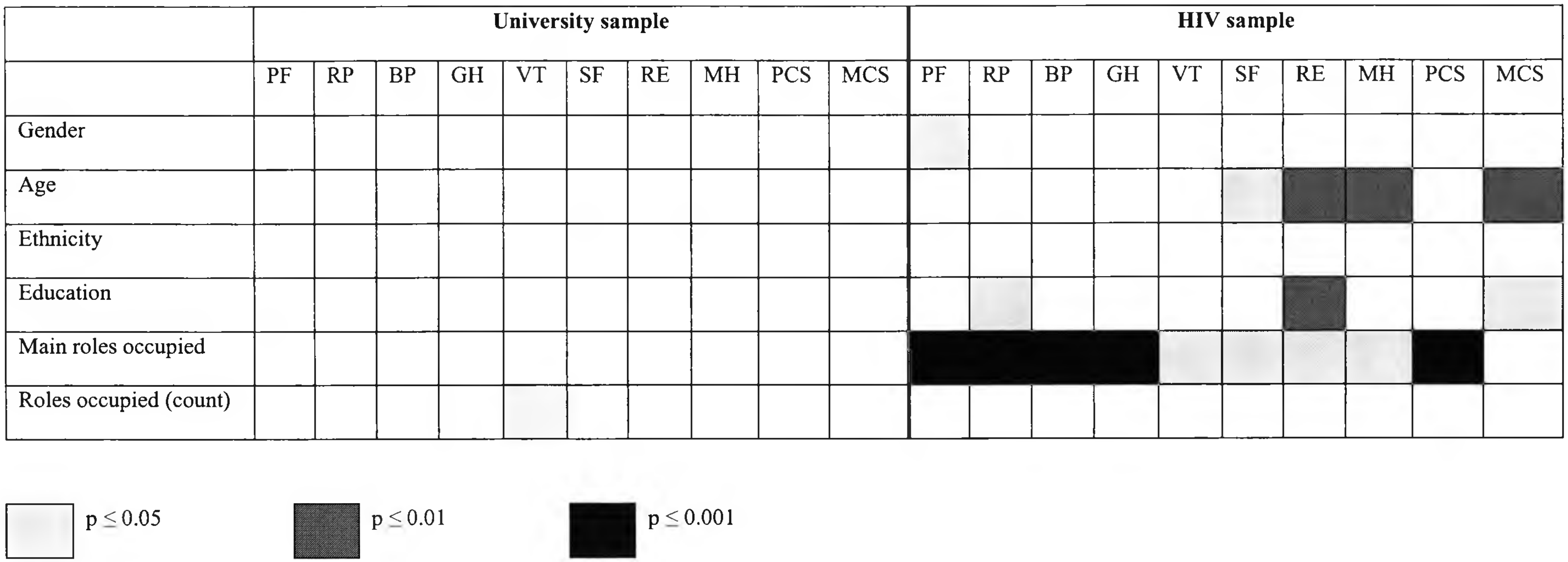
The contextual factors identified, relating to life experiences, enduring characteristics, and aspects of the cognitive structure of health perceptions, were investigated in the quantitative study. This involved a comparison of the distribution of each contextual variable in the two samples, and the analysis of relationships between contextual factors and SF-12v2 scale scores.

The main findings for the two samples are outlined below, with detailed results for the relationship between contextual factors and the SF-12v2 included in Appendix 21. In general, contextual variables were associated with health status as predicted from the literature, outlined in Part One. More significant relationships were observed between contextual variables and heath status scores in the HIV sample, reflecting a wider variation in SF-12v2 scores, the generally poorer health of a large proportion of HIV respondents, and the greater experience of both health problems and services.

Sociodemographic characteristics

The sociodemographic variables included in the quantitative study were: gender; age; ethnicity; education level; and role occupation. Descriptive characteristics for the university sample were presented earlier (pp. 149-155). The principal finding from these analyses was that the samples differed to some extent on most of the measured variables. In comparison with the HIV sample, the university sample was generally younger, comprised more women, was better educated, and more ethnically diverse (See Appendix 20).

Figure 5.1: Sociodemographic characteristics and the SF-12v2



Legend:
PF: Physical Functioning / RP: Role-Physical / BP: Bodily Pain / GH: General Health / VT: Vitality / SF: Social Functioning / RE: Role-Emotional / MH: Mental Health / PCS: Physical Component Scale / MCS: Mental Component Scale

Sociodemographic characteristics and the SF-12v2

Due to differences in composition of the two samples, these factors were not directly comparable in terms of age, ethnic composition, education and work activities. Nevertheless, there were fewer significant associations between sociodemographic factors and health status among university respondents than among HIV respondents (Figure 5.1).

Scores on physical and mental health scales did not differ by gender in the university sample. However, in the HIV sample, scale scores were generally lower for women, significantly so for Physical Function (PF). In both samples, there was correlational evidence that mental health scores increased with age, whereas, generally, physical health scores declined. However, most associations were weak and non-significant in these samples, with the exception of HIV mental health scale results (SF, RE, MH, MCS). This is a common finding in research studies (Campsmith et al 2003; Franks et al, 2003; Mannheimer et al 2005; Murri et al 2003), although there is evidence that it could relate to lower expectations among older respondents (Blaxter, 1990; Fleishman and Lawrence, 2003).

The analysis of ethnicity was complex as both samples contained predominantly white respondents. Although non-white respondents reported poorer health status, this group was too heterogeneous to allow meaningful analyses. In neither sample were there clear differences between the various white respondent groups available for analyses.

Similarly, educational status differed between samples: the university sample comprised almost exclusively respondents within a narrow range of educational experience, since all had received at least further education. Within this narrow range, there were no significant differences in scores. In the HIV sample, there was a much wider range of educational experience reported (from primary to postgraduate level). Worse health status was associated with having the poorest education, significantly so for both role scales (RP and RE) and the mental health summary, a finding that could relate to the association between lower cognitive ability and satisficing (Krosnick, 1999), or lower expectations about their health among those from lower social classes (Blaxter, 1990; Calnan, 1987).

Respondents in the two samples reported very different role occupation. In the university sample, the three most common combinations of roles, including most respondents, were: working full-time only, studying full-time only, and working full-time and studying part-time. There were no significant differences in health status between these roles. In the HIV sample, two contrasting roles predominated: long-term sick only and working full-time only. Comparisons between these activities resulted in a number of significant scale differences, with lower scores reported by the long-term sick. Additional analyses were carried out investigating number of reported activities and health status. In both samples, there were generally no consistent patterns. In both samples, scores on the vitality scale (VT) were higher among those who performed more activities (significantly so in the university sample and borderline significant in the HIV sample). This finding suggests either that those with poorer energy levels are less likely to seek multiple roles (a selection effect) or that those who suffer a decline in energy are no longer able to sustain multiple roles (a causation effect). While there is no conclusive evidence for either proposition, a large proportion of HIV the sample in a single role classified themselves as long-term sick, possibly supporting a causation hypothesis in this sample.

General health

Items related to perceptions of past, present and expectations for future health, including overall general health, general health in the past year, change in health during the year, health expectations in the year ahead, long-standing and long-standing limiting illnesses. Overall, a larger proportion of the university sample reported better general health than the HIV sample, currently and in the recent past. In addition, health expectations for the near future were generally lower among HIV respondents. However, in both samples, there was considerable variation in general health perceptions, particularly for HIV respondents. Many more HIV respondents reported long-standing and limiting illnesses. The illnesses reported differed, with HIV/AIDS the most commonly reported illness in the HIV sample. The finding that, when asked to report any long-standing illnesses, a number of this sample did not acknowledge their HIV status could reflect actual perceptions that HIV is not a long-standing illness, or an editing process, whereby HIV is discounted from the answer (Bowling et al, 1999). The most frequently reported long-standing illness in the university sample involved musculoskeletal problems.

General health

Question 1 of the SF-12v2 asked, 'In general, would you say your health is: Excellent / Very good / Good / Fair / Poor'. In the main, university respondents reported that they perceived their health to be good (Table 5.1). Modal responses were 'very good' and 'good', with the remainder of the sample selecting either 'excellent' or 'fair'. None of the respondents selected 'poor'. Conversely, in the HIV sample all options were selected by some respondents, with a modal answer of 'good', the middle category in the response continuum. The distribution of other responses was near symmetrical, with a slight tendency towards the more negative options ('fair' and 'poor') compared to the positive options ('very good' (14, 19.4%) and 'excellent' (4, 5.6%)). The differences between the two samples was reflected in a significant association between sample and response (Exact test: $\chi^2(4)=21.60$, $p<0.001$; Cramer's V=0.40).

Table 5.1: General health perceptions: University and HIV samples

	University sample N (%)	HIV sample N (%)
Excellent	7 (10.9)	4 (5.6)
Very Good	27 (42.2)	14 (19.4)
Good	27 (42.2)	30 (41.7)
Fair	3 (4.7)	16 (22.2)
Poor	0 (0.0)	8 (11.1)
TOTAL	**64 (100.0)**	**72 (100.0)**

General health in the previous year

Another, similar, question specified that the reference period should encompass the previous year: 'Over the past year, would you say your health has on the whole been…'. Once again, the university sample gave a more positive rating of their general health over the year compared to HIV sample (Exact test: $\chi^2(4)=22.84$, $p<0.001$; Cramer's V=0.41) (Table 5.2). The majority of university respondents indicated their health had been generally good or better. More than half considered that, over the course of the year, their health had been 'very good', and over one third rated it as 'good'. In comparison with the other general health rating, fewer respondents selected the most positive category, 'excellent' and a larger proportion of respondent chose negative health categories, 'fair' and 'poor'. Ratings of health over the past year and the general health question were significantly correlated ($r=0.63$, $p<0.001$). In the HIV sample, as in the university sample, there was a shift towards a more negative rating of general health over the previous year. Although the modal response remained 'good', fewer respondents selected this category, and almost as many rated their health over the year to be ‘fair’ (22, 12.5%). There was an increase in the proportion of respondents rating their health as ‘poor’ and fewer answered ‘excellent’. Despite this, there was a strong association between the two general health questions in the HIV sample ($r=0.70$, $p<0.001$), suggesting that the processes leading to a response to the two items were likely to be similar.

Table 5.2: General health perceptions over the previous year: University and HIV samples

	University sample N (%)	**HIV sample N (%)**
Excellent	1 (1.6)	2 (2.8)
Very Good	33 (52.4)	16 (22.2)
Good	23 (36.5)	23 (31.9)
Fair	5 (7.9)	22 (30.6)
Poor	1 (1.6)	9 (12.5)
TOTAL	**63 (100.0)**	**72 (100.0)**

Changes to general health in the previous year

Respondents were also asked to evaluate any change in their current general health in comparison with their health one year before, ‘Compared to one year ago, how would you rate your health in general now?’. The modal response for both university and HIV samples was ‘about the same as one year ago’. However, the distribution of other responses differed between the samples (Table 5.3). In the university sample, over a quarter of respondents indicated perceived improvement in their health and under a fifth considered their health to be worse, with no respondent selecting the option ‘much worse now than one year ago’. In the HIV sample responses were divided equally between those who felt their health had worsened or improved and all options were selected, including the most extreme. However, there was no significant association between the sample and response to this item. The result remained non-significant when the categories were recoded (better, same, worse). Change is dependent on the state the respondent was in at the earlier time: for example, the varied pattern of results in the HIV sample could reflect changes relating to disease management and treatment regimens during the year.

Table 5.3: Changes to general health in the previous year: University and HIV samples

	University sample N (%)	**HIV sample N (%)**
Much better now than 1 year ago	4 (6.3)	8 (11.1)
Somewhat better now than 1 year ago	13 (20.6)	10 (13.9)
About the same	34 (54.0)	36 (50.0)
Somewhat worse now than 1 year ago	12 (19.0)	14 (19.4)
Much worse now than 1 year ago	0 (0.0)	4 (5.6)
TOTAL	**63 (100.0)**	**72 (100.0)**

Expectations for changes to general health in the next year

Health expectations were also measured, with the question, 'Compared to your health in general now, what do you expect your health to be like in a year's time?'. The majority of the respondents in both samples felt that their health would be stable ('about the same as it is now') (Table 5.4). Most other respondents, particularly in the university sample, had a positive view of their future health, expecting it to improve. Few university respondents felt that their health would worsen, and no one expected it to be 'much worse in a year than now'. In contrast, a larger proportion of respondents in the HIV sample expected their health to worsen, answering either 'somewhat worse in a year than now' or 'much worse in a year than now'. This is a noteworthy finding, since expectations have been linked both to self-schema and to adaptation to illness, in terms of the relationships between beliefs about the potential and current self (Heyink, 1993;

Markus, 1983). However, there was no significant association between the sample and response to this item in the original coding and this result was confirmed when the categories were recoded (better, same, worse).

Table 5.4: Expectations for changes to general health in the next year: University and HIV samples

	University sample N (%)	**HIV sample N (%)**
Much better in 1 year than now	8 (12.7)	6 (8.5)
Somewhat better in 1 year than now	18 (28.6)	19 (26.8)
About the same in 1 year as now	35 (55.6)	38 (53.5)
Somewhat worse in 1 year than now	2 (3.2)	7 (9.9)
Much worse in 1 year than now	0 (0.0)	1 (1.4)
TOTAL	**63 (100.0)**	**71 (100.0)**

Long-standing illnesses

Respondents were asked, 'Do you have any long-standing illness, disability or infirmity (that is, anything that has troubled you over a period of time, or that is likely to affect you over a period of time)?'. A much larger proportion of HIV respondents (the majority) reported currently having a long-standing illness compared to the university sample (Table 5.5) (Exact test: $\chi^2(4)=40.23$, $p<0.001$; Phi=0.55). Nevertheless, more than a third of the university sample reported a long-standing illness. A follow-up question focused on whether any reported long-standing illnesses were limiting ('If you have any long-standing illness or disability, does it limit your activities in any way?'). The majority of respondents in the HIV sample with a long-standing illness indicated that they had experienced limitations, compared to less than half of those in the university sample (Exact test: $\chi^2(1)=5.71$, $p<0.05$; Phi=0.26).

Table 5.5: Presence of long-standing (limiting) illnesses: University and HIV samples

	University sample N (%)	HIV sample N (%)
Longstanding illness(es)	23 (36.5)	64 (88.9)
Limiting illness(es)	*10 (43.5)*	*45 (71.4)*
No limiting illness(es)	*13 (56.5)*	*10 (28.6)*
SUBTOTAL	*23 (100.0)*	*64 (100.0)*
No longstanding illnesses	40 (63.5)	8 (11.1)
TOTAL	**63 (100.0)**	**72 (100.0)**

Respondents were asked to specify their long-standing illness or illnesses. In the university sample, twenty respondents named a single illness, one respondent detailed two illnesses and a further two did not identify their illness. Therefore, a total of twenty-two illnesses were reported in the university sample by twenty-one respondents. In the HIV sample, most respondents reported a single illness (35), with others detailing two (17), three (5), four (3) and five (2) illnesses each. Two respondents failed to name their illness. In total, 106 long-standing illnesses were reported by sixty-two HIV sample respondents.

The specified illnesses were grouped by the researcher according to chapter headings of the International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10) (2006). The ICD-10 is a coding system developed by the World Health Organization (WHO) to enable the classification of health problems for administrative, clinical and epidemiological purposes (WHO, 2006). The reported long-standing illnesses, grouped under ICD-10 headings, are presented in Appendix 22 (university and HIV samples, Tables A22:1 and A22:2, respectively).

Table 5.6 provides a comparative summary of the total number of illnesses coded under ICD-10 headings for the two samples. A range of diseases were reported, although most were grouped under a few headings. Overall, only three chapter headings containing ten or more illnesses ('diseases of the musculoskeletal system and connective tissue', 'certain infectious and parasitic diseases' and 'mental and behavioural disorders'). However, patterns of reported illness differed between the samples.

In the university sample, 'diseases of the musculoskeletal system and connective tissue' comprised almost half of reported diseases (10), of which half were back problems (5). Of those university respondents with a limiting illness (10), 'diseases of the musculoskeletal system and connective tissue' once again predominated (8). In the HIV sample, more than half of the specified illnesses were coded as 'certain infectious and parasitic diseases' (63), with 'mental and behavioural disorders' the only other notable grouping (11). 'Certain infectious and parasitic diseases' included mainly reports of HIV and AIDS (56). 90.3% of the HIV sample respondents who named a long-standing illness reported HIV or AIDS (56 out of 62 respondents). This percentage fell to 77.8% of the total HIV sample (56 out of 72). Since the sample comprised people with HIV, the finding that not all respondents reported HIV may indicate that some did not consider HIV/AIDS to be a long-standing illness, or discounted HIV when answering, assuming that the question was asking about other illnesses (Bowling et al, 1999). Most of the HIV sample respondents who indicated that they had a limiting illness reported HIV or AIDS (39 out of 45, 86.7%). However, multiple illnesses were listed by a large proportion of the HIV respondents (27 out of 62 (43.4%) who named long-standing conditions) and therefore it is not possible to relate health limitations directly to HIV/AIDS. Nonetheless, as HIV is associated with damage to the immune system, resulting in a wide range of opportunistic infections and tumours, multiple illnesses would be expected in this sample.

Table 5.6: Longstanding illnesses reported categorised according to ICD-10 chapter headings: University and HIV samples

ICD-10 chapter heading	University sample N (%†)	HIV sample N (%†)
Certain infectious and parasitic diseases (A00-B99)	0 (0.0)	63 (59.4)
Neoplasms (C00-D48)	0 (0.0)	1 (0.9)
Diseases of the blood and blood-forming organs and certain disorders involving the immune (D50-D89)	1 (4.5)	1 (0.9)
Endocrine, nutritional and metabolic diseases (E00-E90)	3 (13.6)	4 (3.8)
Mental and behavioural disorders (F00-F99)	2 (9.1)	11 (10.4)
Diseases of the nervous system (G00-G99)	0 (0.0)	5 (4.7)
Diseases of the circulatory system (I00-I99)	2 (9.1)	2 (1.9)
Diseases of the respiratory system (J00-J99)	1 (4.5)	6 (5.7)
Diseases of the digestive system (K00-K93)	0 (0.0)	6 (5.7)
Diseases of the skin and subcutaneous tissue (L00-L99)	0 (0.0)	1 (0.9)
Diseases of the musculoskeletal system and connective tissue (M00-M99)	10 (45.5)	2 (1.9)
Diseases of the genitourinary system (N00-N99)	2 (9.1)	3 (2.8)
Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified (R00-R99)	0 (0.0)	1 (0.9)
Injury, poisoning and certain other consequences of external causes (S00-T98)	1 (4.5)	0 (0.0)
TOTAL ILLNESSES REPORTED	**22 (100.0)**	**106 (100.0)**

† Percentage of total reported illnesses in each sample

Figure 5.2: General health and the SF-12v2

	University sample										HIV sample									
	PF	RP	BP	GH	VT	SF	RE	MH	PCS	MCS	PF	RP	BP	GH	VT	SF	RE	MH	PCS	MCS
General health																				
General heath in year																				
Health change in year																				
Health expectations in year																				
Longstanding illness																				
Longstanding illnesses (count)	-	-	-	-	-	-	-	-	-	-										
Musculoskeletal vs. other LSI											-	-	-	-	-	-	-	-	-	-
Limiting illness																				

p ≤ 0.05 p ≤ 0.01 p ≤ 0.001

Legend:
PF: Physical Functioning / RP: Role-Physical / BP: Bodily Pain / GH: General Health / VT: Vitality / SF: Social Functioning / RE: Role-Emotional / MH: Mental Health / PCS: Physical Component Scale / MCS: Mental Component Scale

General health and the SF-12v2

Respondents health status varied according to ratings of their current, past and future perceptions of their general health, along with any self-defined long-standing illnesses, and this was particularly so in the HIV sample, resulting in a number of significant associations (Figure 5.2).

In terms of general health, linear trends were identified in both samples for all scales, with lower mean scores related to poorer general health. In the HIV sample, this pattern was significant for all scales, and most pairwise comparisons were also significant (university sample: significant results for RP, VT and PCS). Although the physical summary score was significantly related to general health in both samples, it should be noted that this summary scale includes the GH scale, derived from the general health item used.

Respondents also rated their general health perceptions more specifically in terms of the previous year. Once again, scale scores were linearly related to general health during the year, particularly among HIV respondents, with significant results for most scales (except MH and MCS), and in terms of pairwise comparisons between general health ratings. In the university sample, the results were significant for PF, GH, VT, MH and PCS. However, there is a close association between general health and general health in the previous year and therefore, these results may reflect that relationship.

There were lower mean scores among those respondents who rated their current health worse than it was a year before. In the HIV sample, this pattern resulted in significant findings for most scales (except VT, MH and MCS), with more notable results for physical health scales. Although the same pattern was observed in the university sample, differences in scale means between response categories were smaller, and there was only one significant result (BP).

Health expectations differed according to sample, so that in the university sample, only two respondents expected their health to get worse during the year. Due to the small number, these respondents were excluded from this analysis. Comparing only those who expected their health to stay the same or get better, current health status tended to be higher among those who expected their health to remain the same, although the difference was only significant for two scales (GH and RE). In contrast, although most HIV respondents expected their health to remain the same or improve, it was also possible to include in the analyses those who expected their health to worsen. Results were significant for most scales (except MH): the group with expectations of worsening health had the lowest mean scores for all scales and the highest scores were observed for those who expected their health to remain the same.

Patterns of scoring among those reporting long-standing illness varied. In the university sample, scores were similar for those with or without a long-standing illness, although physical scores for those who reported a long-standing illness were generally lower, reflected in a significant difference between the two groups for the GH scale. A large majority of HIV respondents indicated that they had a long-standing illness, and these respondents reported significantly lower health status on most scales than those who had not indicated long-standing illness (except BP and PCS). In the university sample, the majority of long-standing illnesses reported were classified as musculoskeletal conditions using ICD-10 classification (WHO, 2006). The mean scores of respondents reporting musculoskeletal long-standing illnesses were lower on all physical scales and the one mental health scale (VT), compared with the scores of those reporting other illnesses, although only significant for two physical health scores (BP and PCS). Many respondents in the HIV sample reported more than one condition, and a comparison was carried out according to the number of conditions reported. The pattern of results was generally linear: worse heath status was reported among those with three or more conditions compared to those who listed only one, significantly so for a number of scales (GH, VT, MH, PCS, MCS). However, reporting of long-standing illness was dependent on definition, subjective decisions regarding inclusion and exclusion criteria, and potentially also satisficing considerations, such as motivation to list multiple conditions. Therefore a crude count may be subject to a range of extraneous influences beyond the presence of illnesses.

In both samples and for most scales, respondents with a limiting illness had worse health status than those with non-limiting long-standing illnesses. However, differences were greater and significant for more scales in the HIV sample (significant results: university sample: BP, VT, PCS; HIV sample: PF, RP, BP, GH, SF, RE, PCS).

These results appear to indicate either that the SF-12v2 scales can differentiate between subtle distinctions in health perceptions regarding the present, the past and the future, or that there are additional factors which similarly influences the rating of both the general health and SF-12v2 items.

Attitude strength

The health-related attitude strength indicators were hypothesised to reflect more complex health attitudes held by those with serious illnesses. The HIV respondents indicated that they thought more about their health, had more experience of serious illness among close friends and family, and were more concerned about their health than those respondents recruited from the university. There was also an indication that the source of attitude strength may differ between the samples. Having experience of others with serious illnesses was related to age in the university sample whereas it was independent of aging in the HIV sample.

Thinking about health (accessibility)

The accessibility of information is a concept drawn from attitude strength research that may be seen to have relevance to the investigation of health perceptions, such as those included in the SF-12v2. Accessibility was assessed indirectly by asking respondents about the frequency of thought about health: Would you say you think about your health… All of the time / Most of the time / Some of the time / A little of the time / None of the time.

The modal response in both samples, selected by almost half of all respondents, was 'some of the time' (Table 5.7). However, a larger proportion of the HIV respondents thought about their health more than this. Consequently, more of the university respondents thought about their health less of the time. Despite this there was no significant association between sample and frequency of thought about health.

Table 5.7: Thinking about health: University and HIV samples

	University sample N (%)	HIV sample N (%)
All of the time	1 (1.6)	5 (7.2)
Most of the time	13 (20.6)	18 (26.1)
Some of the time	31 (49.2)	34 (49.3)
A little of the time	18 (28.6)	11 (15.9)
None of the time	0 (0.0)	1 (1.4)
TOTAL	**63 (100.0)**	**69 (100.0)**

Usefulness (importance)

The attitude strength dimension of importance was assessed indirectly with questions on item usefulness for assessing the respondent's health. Usefulness results are discussed in terms of response process in the next section.

Experience of others' illness (direct experience/knowledge/identification)

A question was included on experiences of health problems among family of friends, relating to the attitude strength components of direct experience, knowledge and identification. The question asked, 'Have you had any experiences of serious illness among people close to you, such as family or friends'. HIV respondents reported much more experience of serious illness in others than the university respondents (Table 5.8). While the distribution of responses in the university sample was symmetrical, centred on the middle option, 'some experience', the majority of HIV respondents felt that they had considerable experience of illness among others ('quite a bit' and 'a lot of experience'). Despite this, there was a range of perceptions in both samples. Differences between the samples just failed to reach significance at the five percent level (Exact test: $\chi^2(4)=8.60$, $p=0.07$; Cramer's $V=0.25$). In order to investigate whether differences in the age profiles of the two samples might influence response, age was correlated with response. The result was significant in the university sample ($r=0.30$, $p<0.05$), but non-significant in the HIV sample, supporting the view that people with HIV have generally more experience of serious illness than a general population, at least until later in life.

Table 5.8: Experience of illness among family and friends: University and HIV samples

	University sample N (%)	HIV sample N (%)
A lot of experience	7 (11.1)	18 (25.0)
Quite a bit of experience	14 (22.2)	23 (31.9)
Some experience	21 (33.3)	18 (25.0)
A little experience	15 (23.8)	9 (12.5)
No experience	6 (9.5)	4 (5.6)
TOTAL	**63 (100.0)**	**72 (100.0)**

Concern about health (intensity)

The intensity dimension of attitude strength was assessed with two questions on concern about health. The first gauged respondents' current level of concern ('At the moment, how concerned are you about your health?'). All responses were represented in both samples (Table 5.9). However, the majority of HIV respondents indicated that they were concerned about their health ('extremely', 'very', or 'fairly'). In the university sample, most respondents felt they were 'not too concerned', although almost a third selected the middle option, 'fairly concerned'. The responding patterns in the two samples differed significantly (Exact test: $\chi^2(4)=14.71$, $p<0.01$; Cramer's V=0.33).

Table 5.9: Concern about health: University and HIV samples

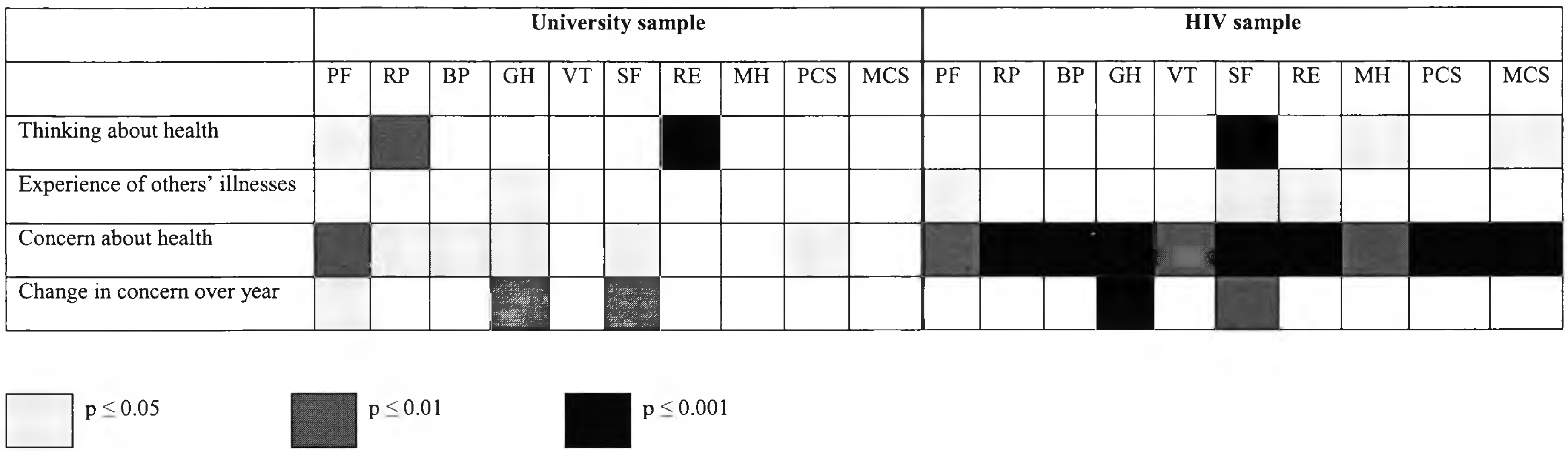
	University sample N (%)	**HIV sample N (%)**
Extremely concerned	1 (1.6)	6 (8.6)
Very concerned	4 (6.3)	12 (17.1)
Fairly concerned	20 (31.7)	31 (44.3)
Not too concerned	35 (55.6)	20 (28.6)
Not at all concerned	3 (4.8)	1 (1.4)
TOTAL	**63 (100.0)**	**70 (100.0)**

A second question asked respondents to estimate whether there was any change in their health concerns during the previous year ('In general, do you think that you are more concerned, less concerned, or have about the same level of concern about your health now as a year ago?'). The modal response was 'about the same concern', selected by the majority of university respondents and almost half of the HIV respondents (Table 5.10). However, in both samples a sizable proportion of respondents were 'more concerned than a year ago', and this was particularly evident in the HIV sample. Nevertheless, responses given by the samples were not significantly different.

Table 5.10: Change in concern about health: University and HIV samples

	University sample N (%)	HIV sample N (%)
More concerned	19 (30.2)	29 (41.4)
Same concern	39 (61.9)	34 (48.6)
Less concerned	5 (7.9)	7 (10.0)
TOTAL	63 (100.0)	70 (100.0)

Figure 5.3: Attitude strength and the SF-12v2



Legend:
PF: Physical Functioning / RP: Role-Physical / BP: Bodily Pain / GH: General Health / VT: Vitality / SF: Social Functioning / RE: Role-Emotional / MH: Mental Health / PCS: Physical Component Scale / MCS: Mental Component Scale

Attitude strength and the SF-12v2

Patterns of results are included in Figure 5.3, with full details in Appendix 21. SF-12v2 scale means were generally linearly related to the amount of time spent thinking about health, with lowest scores among those who responded that they thought about their health 'all of the time' or 'most of the time', and highest for those responding 'a little of the time' or 'none of the time'. Significant results were identified in both samples, but only for mental health scales in the HIV sample (university sample: PF, RP, RE; HIV sample: SF, MH, MCS).

When asked the amount of experience they had of illness among others, there were no clear patterns in scale scores in the university sample, despite a significant relationship for the GH scale. In the HIV sample, patterns also varied between responses, although those respondents who felt they had a lot of experience always had the lowest mean scores, significantly so for three scales (PF, SF, RE).

Current concern about health was associated with SF-12v2 scales, so that the scores of those who considered themselves not at all or not too concerned about their current health were higher than the other response groups. In the university sample, there were significant differences for all physical health scales and one mental health scale (SF). In the HIV sample, the results for all scales were strongly linear, with highest scores for those who were not at all or not too concerned and lowest for those who were very or extremely concerned, leading to significant results for all scales and for most pairwise comparisons between ratings. When asked about any changes in the level of concern during the course of the year, those respondents in both samples who reported the same level of concern generally had better health status whereas those who were more concerned had poorest health, a significant result shared in both samples for GH and SF scales. In addition, in the university sample the relationship was significant for the PF scale.

SF-12v2 scales differed according to attitude strength indicators, indicating that poorer health ratings were associated with more experience of health problems among others, concern about health, and a greater preoccupation with health, more commonly found in the HIV sample than then university sample.

Neuroticism and Extraversion scales

Psychometric properties of the Big Five Neuroticism and Extraversion scales were good, indicating that they are robust scales for use in these samples (Appendix 23). Scale data showed that the university samples had higher Extraversion scores and lower Neuroticism scores than the HIV sample, the Neuroticism finding in particular reflecting the results obtained in studies of health and personality (Goodwin and Engstrom, 2002; Korotkov and Hannah, 2004).

Item analyses

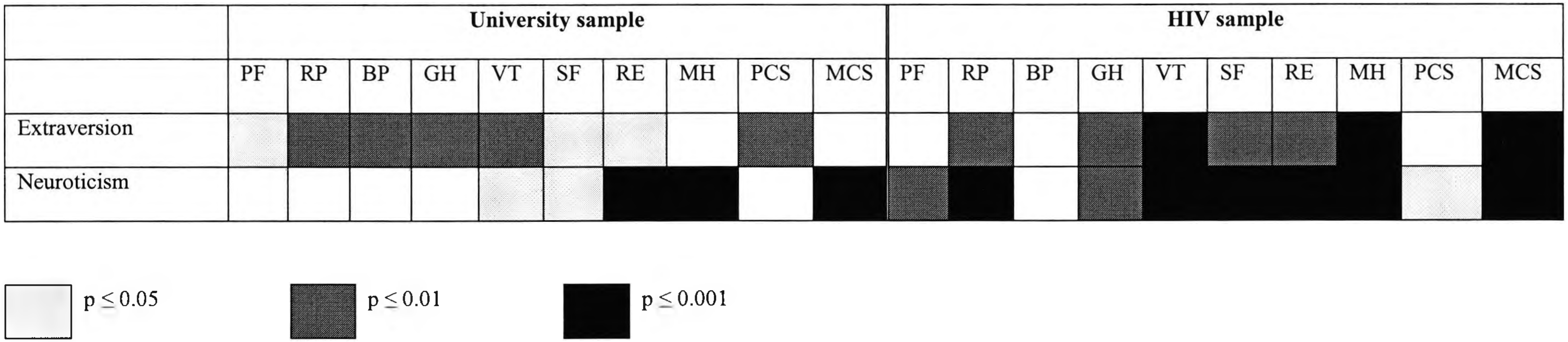
The distributions of the sixteen Extraversion and Neuroticism items were examined: first, item facility; and second, testing item discriminant and convergent validity. Information of item completeness, and item analyses are available in Appendix 23. The results indicated that responses to items were not badly skewed and items correlated with the two scales as predicted. Cronbach's alpha was calculated as a measure of internal consistency. In both samples, alpha coefficients were good for both the Extraversion (university: α=0.85 HIV: α=0.82) and the Neuroticism scales (university: α=0.84; HIV: α=0.76). Principal component analysis was carried out on the Extraversion and Neuroticism items. Two components were extracted from the correlations among the items, and these were rotated orthogonally to produce components that were unrelated to one another. Half or less of the scoring variance was explained by the two components, indicating that a two-component solution may not represent the full solution for the data. Nevertheless, the items loaded on the components as predicted, suggesting that they related to Neuroticism and Extraversion.

Analyses of the university sample data indicated that respondents scored more highly on the Extraversion than Neuroticism scale, on average (Table 5.11). Both scales had a wide range of scores, and distributions were not appreciably skewed or kurtotic, although the Shapiro-Wilk test suggested that the Extraversion scale deviated from normal, being somewhat negatively skewed and having outlying scores. In contrast, Neuroticism was higher than Extraversion in the HIV sample, and the distributions were near normal. Comparing the two samples, significantly higher Extraversion ($t(128)=2.11$, $p<0.05$) and lower Neuroticism ($t(131)=-2.97$, $p<0.05$) were shown for the university compared to the HIV sample.

Table 5.11: Extraversion and Neuroticism scale descriptive statistics: University and HIV samples

	University sample		HIV sample	
	Extraversion	**Neuroticism**	**Extraversion**	**Neuroticism**
Mean	26.69	22.81	24.31	26.04
25^{th} ptile	24.00	18.00	20.00	22.00
50^{th} ptile	27.00	22.00	24.00	26.00
75^{th} ptile	30.00	28.00	29.00	31.00
SD	6.15	6.46	6.65	6.11
N	**59**	**62**	**71**	**71**

Figure 5.4: Personality and the SF-12v2



Legend:
PF: Physical Functioning / RP: Role-Physical / BP: Bodily Pain / GH: General Health / VT: Vitality / SF: Social Functioning / RE: Role-Emotional / MH: Mental Health / PCS: Physical Component Scale / MCS: Mental Component Scale

Personality and the SF-12v2

Neuroticism and Extraversion have been associated with health, with poorer health among those with higher ratings of Neuroticism and better health reported among those with higher ratings of Extraversion (Goodwin and Engstrom, 2002; Korotkov and Hannah, 2004). In this research, the personality dimensions Extraversion and Neuroticism were strongly correlated with most SF-12v2 scales, although the strength and direction of associations varied according to the sample (Figure 5.4). In both samples, Extraversion was positively associated with all scales. In the university sample, these associations were significant and consistently strong for the physical health scales, and significant for most of the mental health scales (except MH and MCS). In the HIV sample, Extraversion was significantly associated with two physical health scales (RP and GH) and all mental health scales.

In contrast, the Neuroticism scale was negatively correlated with all SF-12v2 scales. In the university sample, there was a clear distinction between physical and mental health scales, with significant associations only apparent between Neuroticism and all the mental health scales. In the HIV sample, most of the associations were significant (except BP), although those involving mental health scales were stronger. These results indicate that respondents with higher levels of Extraversion and lower Neuroticism were likely to report better perceived health status, although the strength of associations differed, with a stronger distinction between physical and mental health associations in the university sample.

Health behaviours

Questions on two specific health behaviours, drinking and smoking, were included in the questionnaire, and the detailed analyses of these questions, including psychometric tests are included in Appendix 24. Most respondents in both samples reported that they were regular drinkers, and levels of regular drinking and units consumed were higher in the HIV sample. Regular and heavier drinking were associated with perceived problem drinking, particularly in the university sample. There were notable gender differences between the samples. In the university sample, men and women reported similar drinking behaviour, although more women indicated that they had drinking concerns, and drinking above recommended levels. In the HIV sample, frequent drinking, heavier consumption, concern over problem drinking, and drinking above recommended levels were all more common among male respondents. The majority of respondents were not current smokers, with most university respondents never having smoked. Smoking was more common in the HIV sample than in the university sample, with mean cigarette consumption also slightly higher.

Problem drinking

The four-item CAGE scale is a screening tool for problem drinking (Mayfield et al, 1974). The internal consistency of the CAGE differed between the two samples. An alpha of above 0.70 is generally considered acceptable (Nunnally, 1978). In the university sample, Cronbach's alpha was low (α=0.43), whereas in the HIV sample it was acceptable (α=0.77). This difference indicates that item endorsement was less consistent across the four items in the university sample compared to the HIV sample. Other item analyses (Appendix 24) showed that most items were not strongly inter-correlated in the university sample, and that two items in particular, "Annoyed" and "Eye-opener", were only weakly associated with the CAGE scale. Nevertheless, item heterogeneity was considered acceptable for two reasons: first, internal consistency is related to the number of items in a scale, and the CAGE comprises only four items; second, as a screening tool to detect problem drinking, variation in endorsement patterns over the four items is not as important as whether the standard cut-off score of two or more positive answers could identify problem drinkers. Most respondents in both samples had a CAGE score of 0 (indicating that they did not select "yes" for any item). Only just over a fifth of respondents in each sample scored above the recommended cut-off score (university: 15, 23.8%; HIV: 16, 22.2%).

Alcohol consumption

Most university and HIV respondents reported that they were "regular drinkers" (that is, drinking at some time during an average week). However estimated average weekly drinking (in units of alcohol) was skewed towards lower consumption in both samples, and the modal amount in an average week was 0 units. Consumption was higher in the HIV sample than the university sample, although this difference was non-significant.

Alcohol consumption and the CAGE

In both samples, current drinkers were more likely to report problem drinking behaviour on the CAGE. In addition, levels of consumption were also associated with problem drinking. However, these relationships were stronger (and significant) in the university sample. Non-significant findings may indicate that HIV respondents were reporting past problem drinking, which had little relationship with current consumption. Alternatively, it could perhaps indicate that that they did not perceive themselves to be problem drinkers despite their levels of consumption, or that problem drinking was irregular (a view supported by the finding that over a quarter of infrequent drinkers (less than weekly) scored above the CAGE cut-off score). However, it could also signify that the alcohol questions were answered inaccurately by these respondents.

Gender and drinking behaviour

There were notable differences between drinking reports provided by male and female respondents, according to the sample. Drinking behaviour was quite similar between the sexes in the university sample, with both male and females likely to both drink regularly and consume similar quantities of alcohol. One consequence was that some females exceeded recommended maximum drinking levels of 14 units per week (3, 8.1%) (Royal Colleges, 1995), and a larger proportion of females than males provided two or more positive responses on the CAGE measure. The association between regular drinking and the CAGE was strongest among female university respondents who drank regularly, indicating that the females who drank regularly were more likely to have concerns about their drinking behaviour. In the HIV sample, gender differences were marked, with more regular and heavier drinkers among male respondents. This was reflected in a notable proportion of males who reported that they drank above the maximum level recommended by the Royal Colleges, 21 units (6, 10.7%) (Royal Colleges, 1995), and more males exceeding the CAGE cut-off score.

Smoking behaviour

Detailed information on smoking status and estimated average daily cigarette consumption is shown in Appendix 24. Most respondents were not current smokers, although the HIV sample included more respondents who had smoked in the past. Mean cigarette consumption was slightly, but not significantly, higher in the HIV sample. However, the distribution for cigarette consumption was very skewed towards no cigarettes in a day.

Figure 5.5: Health behaviours and the SF-12v2

	University sample										HIV sample									
	PF	RP	BP	GH	VT	SF	RE	MH	PCS	MCS	PF	RP	BP	GH	VT	SF	RE	MH	PCS	MCS
Problem drinking (CAGE)																				
Drinking frequency																				
Weekly consumption (units)																				
Smoking status																				
Daily consumption (cigs)																				

p ≤ 0.05 p ≤ 0.01 p ≤ 0.001

Legend:
PF: Physical Functioning / RP: Role-Physical / BP: Bodily Pain / GH: General Health / VT: Vitality / SF: Social Functioning / RE: Role-Emotional / MH: Mental Health / PCS: Physical Component Scale / MCS: Mental Component Scale

Health behaviours and the SF-12v2

Significant associations between health behaviours and SF-12v2 scores are shown in Figure 5.5. In the university sample, there was no consistent relationship between CAGE and SF-12v2 scale scores, and none of the analyses was significant. However, in the HIV sample, problem drinkers had lower mean scores, significantly so for a range of physical and mental scales (PF, RP, GH, SF, RE, MCS). In the university sample, mean SF-12v2 scores were unrelated to regular, infrequent and non-drinking groups, leading to non-significant results. In the HIV sample, health scores were always better among regular drinkers, and there was a single significant result (GH). Although these findings could be spurious, it is notable that this pattern was identified for all scales. Possible explanations include healthier respondents being able to socialise more readily, consuming more alcohol as a consequence; or that those in poorer health were not able to consume alcohol. There was some evidence for the former hypothesis, with a smaller proportion of regular drinkers reporting that their social lives were affected by their health (SF-12v2 Social Functioning item). Finally, the association between self-reported average weekly alcohol consumption and health status was generally weak. There were two significant, positive correlations in the university sample, one with a physical and the other a mental health scale (RP and SF), whereby higher consumption was correlated with better health status scores on these scales. None of the associations was significant in the HIV sample.

In terms of smoking status, in the university sample, higher mean scores were attained by respondents who reported that they used to smoke in comparison to those who still smoked or who had never smoked. This pattern was statistically significant for one scale (GH). In contrast, in the HIV sample, those who had never smoked had the highest scores, followed by ex-smokers and then smokers with the lowest scores.

However, scale differences were non-significant. Associations between cigarette consumption levels and health status were weak and non-significant.

In summary, the relationship between current health behaviours and health status was weak, and did not appear to reflect health deficits resulting from reported levels of smoking and drinking, which were generally low in both samples, although patterns differed by sample and gender of respondent. The relationship between ever having experienced problem drinking and health status apparent in the HIV sample did not relate to the current levels of drinking reported, and therefore it is not possible to know whether they relate to problem drinking or measurement issues.

Health services contact

Health service contact would clearly be expected to have a direct relationship with health, at least in terms of a recent contact. However, direct experience is also a dimension of attitude strength (Krosnick and Smith, 1994). Across a range of services, HIV respondents were more likely than university respondents to have accessed health care or advice, to have used services more recently, and more regularly in the previous year. These differences between samples were particularly apparent for hospital services.

GP contact

There was no significant difference between the samples in contact with a General Practitioner (GP): almost all respondents reported that they had received care or advice at some time (Table 5.12). Most of university and HIV respondents who had received advice had done so during the previous year, and, accordingly, there was again no significant difference between the samples. However, those in the HIV sample reported more contacts during that year, and this difference just failed to reach the five percent significance level (Exact test: $\chi^2(2)=5.68$, $p=0.06$; Cramer's $V=0.23$).

Table 5.12: GP contact: University and HIV samples

	University sample N (%)	**HIV sample N (%)**
Ever had contact		
Yes	60 (95.2)	67 (97.1)
Within the past year	49 (83.1)	55 (82.1)
Once	*17 (34.7)*	*15 (27.3)*
2-5	*25 (51.0)*	*21 (38.2)*
5 or more times	*7 (14.3)*	*19 (34.5)*
SUBTOTAL	***49 (100.0)***	***55 (100.0)***
1 – 5 years ago	9 (15.3)	8 (11.9)
More than five years ago	1 (1.7)	4 (6.0)
SUBTOTAL	**59 (100.0)**	**67 (100.0)**
No	3 (4.8)	2 (2.9)
TOTAL	**63 (100.0)**	**69 (100.0)**

Hospital casualty or outpatient contact

Most respondents reported that they had at some time during their lives attended a hospital casualty or outpatient department, although a significantly larger proportion of the HIV sample had done so (Exact test: $\chi^2(1)=17.18$, $p<0.001$; Phi=0.36) (Table 5.13). In terms of most recent contact among those who had ever attended, patterns differed significantly between the samples, with more of the HIV sample respondents reporting that they had attended during the previous year (Exact test: $\chi^2(2)=19.14$, $p<0.001$; Cramer's V=0.42). Among those who reported having attended during the previous year, respondents from the HIV sample were significantly more likely to have attended more frequently (Exact test: $\chi^2(2)=10.01$, $p<0.01$; Cramer's V=0.36).

Table 5.13: Outpatient contact: University and HIV samples

	University sample N (%)	**HIV sample N (%)**
Ever had contact		
Yes	43 (68.3)	66 (95.7)
Within the past year	20 (46.5)	56 (84.8)
Once	*10 (50.0)*	*13 (23.2)*
2-5	*10 (50.0)*	*25 (44.6)*
5 or more times	*0 (0.0)*	*18 (32.1)*
SUBTOTAL	***20 (100.0)***	***56 (100.0)***
1 – 5 years ago	14 (32.6)	8 (12.1)
More than five years ago	9 (20.9)	2 (3.0)
SUBTOTAL	**43 (100.0)**	**66 (100.0)**
No	20 (31.7)	3 (4.3)
TOTAL	**63 (100.0)**	**69 (100.0)**

Hospital daypatient contact

A third of university sample respondents reported that they had been treated as a hospital daypatient at some time, compared to almost two thirds of the HIV sample (Exact test: $\chi^2(1)=10.96$, $p=0.001$; Phi=0.29) (Table 5.14). A significantly larger proportion of the HIV respondents had attended during the previous year (Exact test: $\chi^2(2)=6.01$, $p=0.05$; Cramer's V=0.30). The majority of respondents who had attended recently from both samples had done so between two to five times in the year, and there was no significant difference between samples.

Table 5.14: Daypatient contact: University and HIV samples

	University sample N (%)	**HIV sample N (%)**
Ever had contact		
Yes	21 (33.3)	44 (62.0)
Within the past year	4 (19.0)	20 (45.5)
Once	*1 (25.0)*	*6 (30.0)*
2-5	*3 (75.0)*	*12 (60.0)*
5 or more times	*0 (0.0)*	*2 (10.0)*
SUBTOTAL	***4 (100.0)***	***20 (100.0)***
1 – 5 years ago	8 (38.1)	16 (36.4)
More than five years ago	9 (42.9)	8 (18.2)
SUBTOTAL	**21 (100.0)**	**44 (100.0)**
No	42 (66.7)	27 (38.0)
TOTAL	**63 (100.0)**	**71 (100.0)**

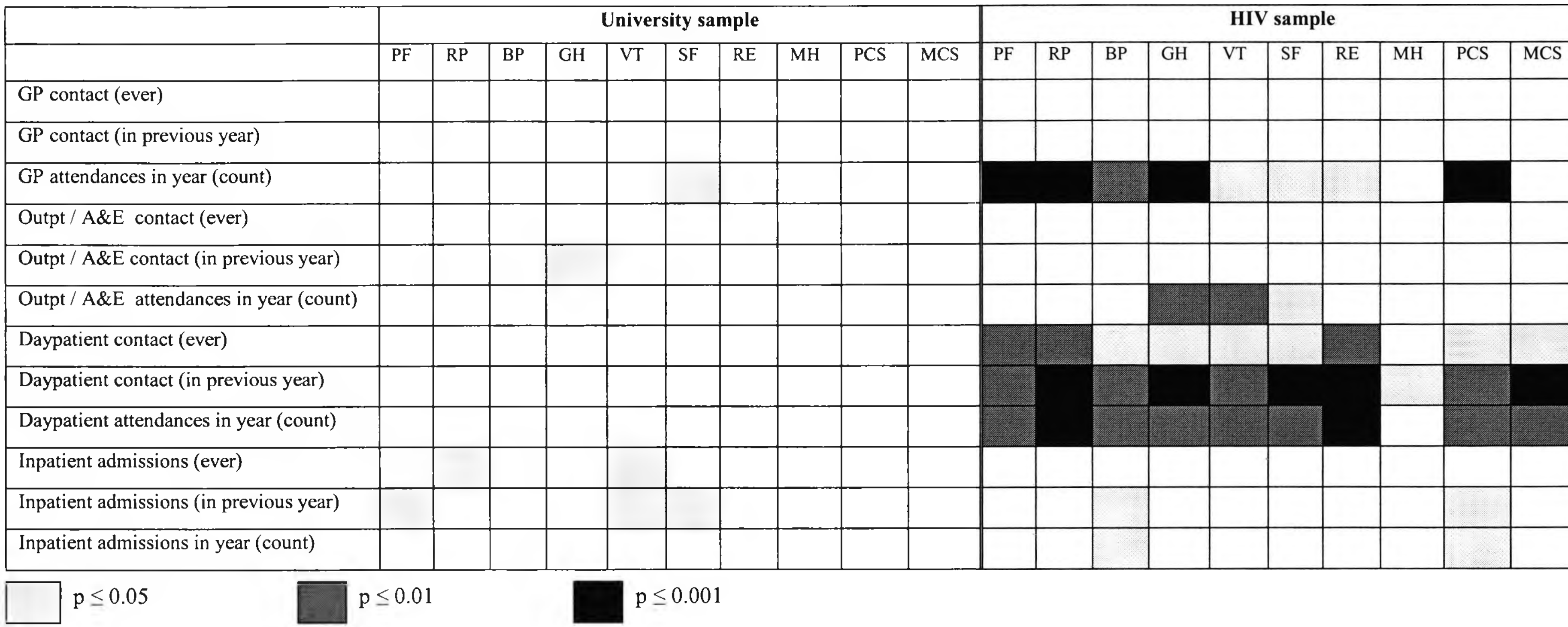
Hospital inpatient contact

The majority of respondents from both samples had inpatient experience, although significantly more of the HIV respondents had been inpatients at some time (Exact test: $\chi^2(1)=5.49$, $p<0.05$; Phi=0.20) (Table 5.15). Two thirds of the university sample respondents had been an inpatient more than five years before (25, 65.8%). The remainder were divided between those who had attended within the previous year, whereas the majority of HIV respondents had experience within the previous five years, resulting in a significant difference between the samples (Exact test: $\chi^2(2)=14.17$, $p=0.001$; Cramer's V=0.39). Most respondents who had been an inpatient within the previous year had only a single admission, and the samples did not differ significantly.

Table 5.15: Inpatient contact: University and HIV samples

	University sample N (%)	**HIV sample N (%)**
Ever had contact		
Yes	38 (60.3)	56 (78.9)
Within the past year	7 (18.4)	20 (35.7)
Once	*4 (57.1)*	*13 (65.0)*
2-5	*3 (42.9)*	*7 (35.0)*
5 or more times	*0 (0.0)*	*0 (0.0)*
SUBTOTAL	***7 (100.0)***	***20 (100.0)***
1 – 5 years ago	6 (15.8)	21 (37.5)
More than five years ago	25 (65.8)	15 (26.8)
SUBTOTAL	**38 (100.0)**	**56 (100.0)**
No	25 (39.7)	15 (21.1)
TOTAL	**63 (100.0)**	**71 (100.0)**

Figure 5.6: Health service contact and the SF-12v2



Legend:
PF: Physical Functioning / RP: Role-Physical / BP: Bodily Pain / GH: General Health / VT: Vitality / SF: Social Functioning / RE: Role-Emotional / MH: Mental Health / PCS: Physical Component Scale / MCS: Mental Component Scale

Health service contact and the SF-12v2

Significant relationships between SF-12v2 and health service contact are shown in Figure 5.6. Almost all respondents in both samples had been in contact with a GP surgery at some time, and most had attended during the year before. There were no differences in health status between those respondent who had attended in the previous year compared to others. In both samples, scores were generally lower among those who reported attending more than five times in the previous year compared to others. In the university sample, few respondents had attended this often, and the differences in scores according to number of attendances were significant only for a single scale (although not when two or more attendances were combined as a group) (SF). In the HIV sample, those who had attended more than five times also had significantly lower scores on all physical and most mental health scales (except MH and MCS).

The majority of university respondents had at some time attended an outpatient or casualty department. A comparison between these and the remainder showed no significant differences in any scales. Since almost every HIV respondent had attended an outpatient or casualty department, it was not possible to run an equivalent analysis. Among those who had received outpatient or casualty department care within the year, mean scores on all scales were lower, compared with other respondents. However, only one of these comparisons was significant in the university sample (GH) and none in the HIV sample. Analyses according to number of attendances within the previous year showed that scores were generally lowest for those reporting the most outpatient or casualty department attendances in the year. However, none of the statistical analyses was significant in the university sample, while one physical and two mental health scales were significantly associated with attendances in the HIV sample (GH, VT and SF).

There was no relationship between ever having been a daycare patient and SF-12v2 scores in the university sample. However, since only four respondents had received daycare in the previous year, no further analyses were possible with this sample. In the HIV sample, most respondents had received daycare at some time, and there was a strong and significant relationship between experience of daycare and most scale scores (except MH). Those HIV respondents who had attended during the previous year had significantly lower scores on every scale. The strong and significant relationship between daycare and health status in the HIV sample continued for number of daycare attendances in the year. Although patterns varied, generally linear, significant relationships were found between number of attendances and scale scores (except MH).

Counter intuitively, mean SF-12v2 scores were higher among those university respondents who been a hospital inpatient at some time, significantly so for two scales, physical and mental (RP and VT). For the HIV sample, in contrast, scores were lower among respondents with experience of inpatient care, although none significantly. The results from the university sample were in part explained by the fact that most of those who had been an inpatient reported that their most recent admission had occurred more than five years before questionnaire completion, and were therefore less likely to influence current health ratings. The few respondents who had been inpatients during the previous year had lower mean scores in comparison to others, significantly so for three scales, physical and mental (PF, VT and SF). However, it is unclear why those with no inpatient experience should have lower scores than others who had had an earlier inpatient stay, although it is possible that these results could be due to adaptations to the effects of illness, or non-optimal responding during questionnaire completion. In the HIV sample, recent admission was associated with poorer health status, although significant for only two physical health scales (BP and PCS). No analyses were carried out on number of admissions during the year for university respondents, since there were so few who had been admitted within the year. In the HIV sample, there was generally a linear relationship between number of admissions and heath status, with lowest scores among those who had been admitted two or more times during the year, and the relationship was significant for two physical health scales (BP and PCS).

The relationships between health service use and SF-12v2, particularly for hospital care, indicated that poorer subjective health status was associated with more recent and regular contact with health services, a pattern that was more common in the HIV sample.

Clinical indicators in the HIV sample

Clinical indicators for the HIV sample are shown in Table 5.16. Most respondents had spent a number of years with an HIV diagnosis. The majority were receiving anti-HIV medication, and self-reported information on the most recent virological markers suggested that, for most respondents, HIV was under control.

Duration of diagnosis

There was a wide range in the number of years since respondents had been diagnosed HIV positive (between 1 and 25 years). The mean number of diagnosed years was 9.01 (median: 7.00; mode: 5).

Anti-HIV medication

Over three quarters of respondents indicated that they took anti-HIV medication (57, 80.3%).

CD4 count

The most recently reported CD4 count (cells/mm^3) in this sample ranged from 130 to 1400, with a mean of 449.99 (median: 399.0; mode: 600). A CD4 count less than 200 has been identified as leading to an increased risk of serious infection, while a value above 500 in an HIV positive person is considered normal (Centers for Disease Control and Prevention, 1993; Gill et al, 2002). The majority of respondents had an intermediate CD4 level (200-499) (40, 57.1%), while few were in the risk group (CD4 less than 200) (5, 7.1%) and the remainder had the level of a healthy HIV positive person (greater than or equal to 500) (25, 35.7%).

Viral load

Most respondents reported that their most recent viral load test produced undetectable results (51, 71.8%), suggesting that the level of HIV in their bodies was under control. A viral load level of 10,000 copies/ml and above is commonly considered to be clinically meaningful in relation to treatment initiation or change (Gill et al, 2002). Ten respondents (14.1%) had a viral load at this level.

Table 5.16: Clinical indicators in the HIV sample (N=70 - 71)

		N (%)	Mean (SD)
Years since HIV diagnosis, mean (SD):			9.0 (6.5)
Years (banded):	1 – 4	22 (31.0)	
	5 – 7	16 (22.5)	
	8 – 15	18 (25.4)	
	16 – 25	15 (21.1)	
Currently taking anti-HIV medication:		57 (80.3)	
CD4 count, mean (SD):			450.0 (240.2)
CD4 count (banded):	< 200	5 (7.1)	
	200 – 499	40 (57.1)	
	≥ 500	25 (35.7)	
Viral load (banded):	Undetectable (<250, <400)	51 (71.8)	
	400 – 1000	6 (8.5)	
	1001 – 10000	4 (5.6)	
	10001 – 100000	7 (9.9)	
	> 100000	3 (4.2)	

Figure 5.7: Clinical indicators in the HIV sample and the SF-12v2

	HIV									
	PF	RP	BP	GH	VT	SF	RE	MH	PCS	MCS
CD4 level										
Viral load level										
Anti-HIV therapy										
Length of HIV diagnosis										

p $\leq$ 0.05 p $\leq$ 0.01 p $\leq$ 0.001

Legend:
PF: Physical Functioning / RP: Role-Physical / BP: Bodily Pain / GH: General Health / VT: Vitality / SF: Social Functioning / RE: Role-Emotional / MH: Mental Health / PCS: Physical Component Scale / MCS: Mental Component Scale

Clinical indicators in the HIV sample and the SF-12v2

There was a linear relation between CD4 level and scale scores, which was particularly apparent for physical scales, with poorer health reported by those with lower CD4. This pattern was significant for three physical health scales (PF, BP and PCS) (Figure 5.7).

There was no significant relationship between level of reported viral load and SF-12v2 score, although mean scale scores were generally higher among those with undetectable viral load, indicating better health status among those who had less HIV in their blood.

The majority of respondents reported that they were on an anti-HIV therapy at the time of the survey. However, although mean SF-12v2 scores for most scales were marginally higher among those on anti-HIV medication, there were no significant differences on any scale.

Length of time since diagnosis was divided into quartiles, since it has been suggested that relationship between stage of disease and health status may be complex and non-linear (Murri et al, 2003). Mean scores were generally lower among those with longest period since diagnosis. However, for most scales, the highest scores were not found for respondents with the most recent diagnosis; rather it was among those who were diagnosed between five to fifteen years before, depending on the scale. Possible explanations for these findings could include health improvements resulting from disease management and treatment regimens that need to be established over time, or a gradual adaptation to HIV symptoms. However, none of the scale differences was significant.

Despite the non-significant results, the patterns of association for disease markers were in directions that could be predicted in relation to clinical progression of HIV/AIDS.

Summary

Health status response exists within a context, and therefore the relationship between contextual factors and health status ratings was examined in the two samples. Associations between contextual variables and SF-12v2 were generally stronger for the HIV sample, due to the response variations reported here. Health status was found to be lower among those with more health problems, greater use of health services, and with health playing a more central part in their lives. Clinical markers of HIV were also generally related to health status, significantly so for CD4 and physical health. In both samples, SF-12v2 scores were generally associated with the contextual variables as predicted, providing evidence for the construct validity of the SF-12v2.

These results revealed expected differences in variables that were predicted to relate to the centrality of health in respondents' lives (Bowling, 1995), such as general health, health service use and attitude strength indicators. The HIV sample included more respondents who reported poorer general health. Most respondents considered their health (good or bad) to be stable over the course of the previous year and expected it to remain so over the forthcoming year, although more HIV respondents expected it to worsen. Expectations have been linked both to self-schema and to adaptation to illness, in terms of the relationships between beliefs about the potential and current self can only be speculated: respondents may not have considered HIV to be a long-standing illness; however, since respondents were recruitment into a survey of people with HIV, they may have considered that they should report only illnesses other than HIV, indicating a context effect (Fayers and Sprangers, 2002). The majority of HIV respondents did, however, report that they had a limiting long-standing illness. In terms of direct experience, reflected in health service use, HIV respondents were generally more likely to have accessed primary through tertiary health services more recently and regularly than university respondents.

Other indicators considered to relate to attitude strength revealed that HIV respondents thought more about their health, had more experience of illness among friends and family, and were more concerned about their health than university respondents, although there was variation within samples. These findings indicate that, as predicted, the HIV sample would possess health perceptions that were directly experienced (general health and health service contact), more intense (concern), accessible (thought about frequently), and were identified with (experience with others) (Markus, 1977; Petty and Krosnick, 1995).

Personality has been linked to both health status and memory coding and recall biases (Larson, 1992; Ruo et al, 2003; Schroeder and Costa, 1984; Watson and Pennebaker, 1989). The university sample included respondents who were significantly higher in Extraversion and lower in Neuroticism than the HIV sample. This is an interesting finding since personality is generally considered to be dispositional, and stable over many years (Gustavsson et al, 1997). Therefore, it would not be expected that a particular life experience, such as living with HIV, or working in a university, would have affected personality, unless the mechanism was selective, whereby those possessing particular personality characteristics were more likely to follow a certain life course. However, it should be noted that there is some evidence that major life events, such as illness, may lead to reactive personality changes (Terracciano et al, 2006). Unfortunately, the current study cannot investigate the development of these relationships over time.

The relationship between sample and health behaviours was not notable: most respondents were regular drinkers, although the amount consumed was low for most respondents and few were problem drinkers, according to the CAGE measure. Most respondents were not current smokers and the mean cigarette consumption was low.

It is also worth noting that the samples differed in composition, which could have influenced both SF-12v2 and the distributions of other variables. In contrast to the university sample, the HIV sample was predominantly male, older, included respondents of more diverse educational background, and performed different roles. For example, one possible implication of the age and gender differences related to the lower mental health scores attained in the university sample: other studies have shown both that poorer subjective mental health status is found among women than men, and younger compared to older (Franks et al, 2003).

In addition, although it would be expected that the HIV sample comprised a larger proportion of gay men in comparison to the university sample, it was not possible to investigate whether some of the findings identified in this section may have been associated with sexual orientation, as this information was not collected. However, other research has shown differences between gay men and the general population in terms of health behaviours, including smoking and drinking, as well as psychological variables associated with Neuroticism, such as depression and anxiety. In terms of the health behaviours, there is evidence that gay men are at greater risk of smoking and problem drinking compared to the general population (McKirnan and Peterson, 1989; Stall et al, 1999). In addition, there is also evidence that gay men experience more mental health problems that would be classified under the Neuroticism trait, including depression and anxiety disorders (Gruskin and Gordon, 2006; Sandfort et al, 2006). Mechanisms proposed to explain the association between sexual orientation, drinking and smoking have included gay socialisation patterns (Bux, 1996; Stall et al, 1999). However, another explanation proposes that so-called "minority stress", resulting from prejudice, stigma, discrimination, internalising negative social attitudes and concealment of sexual identity, may lead to increases in both the symptoms of Neuroticism, and drinking and smoking behaviours (Sandfort et al, 2006). While it was not possible to investigate these relationships using the current data, it is important to highlight that difference between samples in terms of problem health behaviours and personality may relate in part to factors associated with sexual orientation.

Finally, the focus of the current research was on the identification of important contextual factors, and an investigation of relationships between the contextual factors and the SF-12v2 scales, based on patterns that had already been identified in the literature. It should be noted that a number of interesting associations between contextual factors and the SF-12v2 were identified in these analyses. However, it would be beyond the scope of this project to examine these relationships more deeply, although it is intended that they will be the investigated in future research.

Having analysed some of the contextual influences on response, the next section includes the results of an investigation into response process, specifically the retrieval and judgement stages, which could potentially mediate some of the contextual influences on SF-12v2 response.

6 An evaluation of response processes for the SF-12v2

The three components of the response process investigated were: strategy, easiness of completion and usefulness of each SF-12v2 item. Strategy referred to the subjective perception of how a judgement was made (Turner and Fiske, 1968). Previous research has indicated that reported strategy can be used to investigate cognitive processes undertaken in order to generate a response (Gordon and Holden, 1996; Turner and Fiske, 1968). In relation to health, there is evidence that general perceptions are more commonly reported in the absence of health problems (Rothman and Schwartz, 1998), or when a question is vague (Schwartz et al, 1997), suggesting that it is commonly reported when heuristic strategies are employed. Easiness and usefulness were hypothesised to relate indirectly to the cognitive processes involved in generating a response. Easiness was considered to be a proxy of the level of effort required to generate an answer, whereby those with serious health problems or no problems would consider items to be more easy to answer than those with variable health problems, who would need to employ more complex judgement strategies in order to respond (Jenkinson et al, 1996). Usefulness was considered to reflect the subjective importance of each item to the respondent's conceptualisation of their health, a dimension of attitude strength and therefore related to the structure of the self-attitude, influencing cognitive processes undertaken (Krosnick and Smith, 1994).

6.1 An analysis of the use of response strategies for the SF-12v2

Overview

Response strategies were investigated according to strategies reported for individual items and by strategy use reported by respondents over the twelve items. Results from both samples indicated that the most commonly reported strategy involved respondents using 'A general picture of yourself' when answering, followed by the strategy of recalling 'Specific experiences'. The two were used in combination by almost a quarter of university and HIV respondents. Other strategies were reported to a lesser extent, and differed by sample and item. In terms of the use of strategy by university compared to HIV respondents, 'A general picture of yourself' was more prevalent in the university sample. In the HIV sample, there was more use of 'Specific experiences', reflecting greater health experiences. Overall, the results suggest that response strategy differs according to SF-12v2 item and sample. In addition, scales were derived from the total number of times respondents indicated they had used the four most common strategies. There was a strong negative association between the general picture and specific experiences scales, indicating an inverse relationship between the two strategies.

The use of the general perception strategy was more commonly associated with better health ratings on the SF-12v2 at item and scale level, with poorer health associated with a greater use of a recollection of specific situations. It is argued within the thesis that this pattern reflects cognitive processes adopted when responding to a health status questionnaire. It would be expected that considerable health problems would be more salient for the respondent asked to rate their health, whereas others would be more likely to make use of a more heuristic, general strategy (Rothman and Schwartz, 1998). However, factors such as adaptation and discounting of health problems could lead to the use of a response strategy by a respondent that does not reflect any objective evaluation of their health. Finally, it is not possible to reject another interpretation; that the relationship between strategy and health status could result from the "degrees of thoroughness" of question interpretation and response, with the reporting of general perceptions and better health the product of a satisficing approach, and, conversely, specific experiences and worse health ratings the result of optimizing. However, interpretation of the association between strategy and response is made more complex because knowledge and motivation have been shown to relate to optimizing (Krosnick, 1999), and those with health problems might be expected to consider more carefully their specific health experiences when generating responses. In addition, since respondents often selected more than one strategy, this would seem to provide evidence against a satisficing interpretation. To conclude, the precise nature of the relationship between strategy and response cannot be disentangled from a cross-sectional self-completion survey and requires further investigation.

Analyses of response strategy

After each SF-12v2 item, an additional question asked the respondent to indicate a single, main strategy used when answering the question. Three response strategies were listed ('comparing yourself with others', 'A general picture of yourself' and 'Specific experiences'), along with 'not sure: the answer just came to me' and an option to specify any other strategy employed. In terms of analyses, response strategy items were investigated to identify both overall use and the breakdown of strategies according to individual SF-12v2 items. For the analysis of patterns of response strategy use according to respondent, combinations of strategies reported were identified. Finally, scales were derived for each of the main response strategies, summing the number of times a strategy was reported during completion of the SF-12v2. These were investigated according to scale properties and interrelationships between strategies.

Data completeness

The proportion of respondents with complete data on response strategy was calculated. All response strategy items were completed by the majority of respondents. Fourteen of the response strategy items were missing in total from the university sample (nine respondents), giving a completeness score of 98.2%. In the HIV sample, two response strategy items were missing (one respondent), giving a completeness score of 99.8%.

'Other' strategies

Most respondents used the response strategies listed in the strategy item and did not indicate that they had used any other strategies when answering. However, two additional strategies were mentioned in relation to the general health item (GH01):

- Comparison with past. Comparing perceptions of current health with those of a previous health state (indicated by two university respondents and one HIV respondent)
- Ruling out response options. Making use of the response options by ruling out categories until left with a single category (indicated by one university respondent)

In addition, two respondents used the 'Other' option to describe a strategy that would be more appropriately included in one of the predefined categories. One university sample respondent explained that they based their answer to item RP03 on the fact that they had 'been ill'. This response was recoded as 'A specific situation or experience'. An HIV sample respondent wrote for item GH01, 'I simply feel well'. This was recoded as 'A general picture of yourself'.

Item-level response strategy analyses

Total use of strategies in HIV and university samples

The overall percentage of each response strategy reported in the two samples was calculated (Table 6.1). In the university sample, a total of 754 strategies were reported over the twelve questions by all respondents (64 * 12 = 768, minus 14 missing responses). The equivalent figure in the HIV sample was 862 strategies (72 * 12 = 864, minus 2 missing responses). In both samples, two response strategies predominated; 'A general picture of yourself' and 'Specific experiences', accounting for 93.4% (704) of strategies reported in the university sample and 88.8% (765) in the HIV sample. The only other notable reported strategy was 'Comparing yourself to others' in the HIV sample. These patterns were replicated when analyses were rerun excluding those respondents with missing data on any item.

Table 6.1: Overall proportions of response strategies for the SF-12v2: University and HIV samples

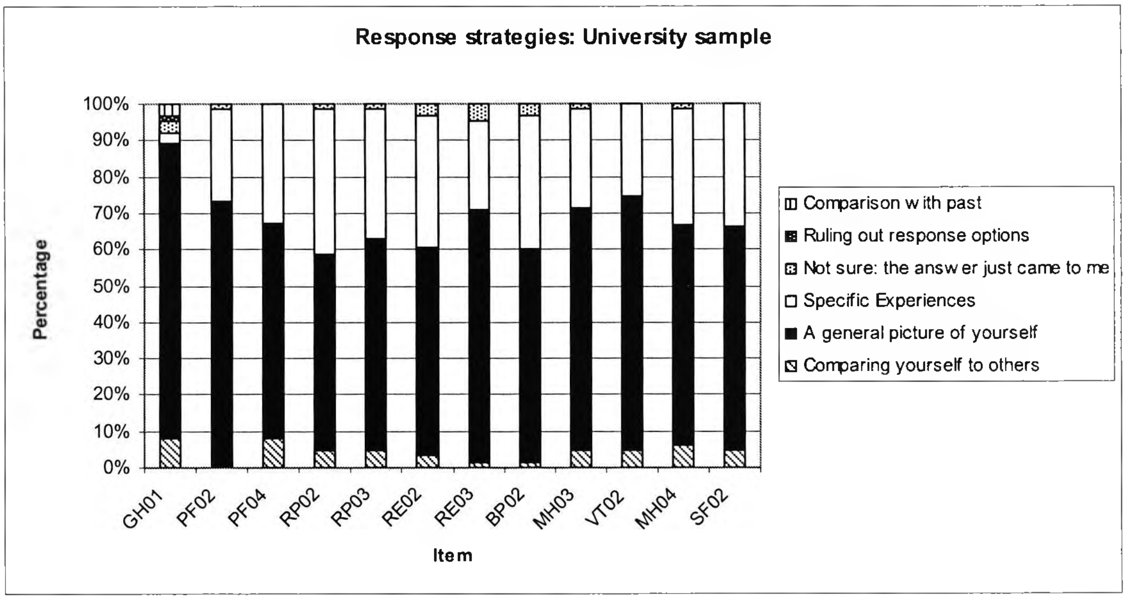
Response strategy	University sample N (%†)	HIV sample N (%†)
A general picture of yourself	484 (64.2)	479 (55.6)
Specific experiences	220 (29.2)	286 (33.2)
Comparing yourself to others	33 (4.4)	79 (9.2)
Not sure: the answer just came to me	14 (1.9)	17 (2.0)
Comparison with past	2 (0.3)	1 (0.1)
Ruling out response options	1 (0.1)	0 (0.0)
TOTAL	**754 (100.0)**	**862 (100.0)**

† Percentage of total strategies reported in the two samples

Use of individual SF-12v2 item strategies in HIV and University samples

Figures 6.1 and 6.2 show the strategies for each SF-12v2 item in the two samples. The modal response strategy for every item was 'A general picture of yourself', selected by at least 34 (54.0%) university and 35 (48.6%) HIV respondents. The overall proportions of each reported response strategy outlined earlier mask variation between items for the two samples. For example, a range of strategies was reported for the general health item (GH01). It is the only SF-12v2 item that does not refer to any form of timescale, and it was the item with the largest proportion in both samples reporting that they had considered 'A general picture of yourself'. Nevertheless, there was a significant difference in the proportions of strategies reported by each sample, with more university respondents using 'A general picture of yourself' (university sample: 52, 81.3%; HIV sample: 47, 65.3%), and fewer considering 'Specific experiences' (university sample: 2, 3.1%; HIV sample: 14, 19.4%) (Exact test: $\chi^2(5)=10.69$, $p<0.05$; Cramer's V= 0.28). The next item, PF02, one of two (physical functioning) items referring to events 'during a typical day', had a similarly large proportion of university sample respondents reporting the strategy 'A general picture of yourself' (47, 73.4%), compared with fewer HIV respondents (33, 45.8%), who were more likely to report 'Specific experiences' (university sample 16, 25.0%; HIV sample: 27, 37.5%) and 'Comparing yourself to others' (university sample: 0, 0.0%; HIV sample: 11, 15.3%) (Exact test: $\chi^2(3)=15.85$, $p<0.001$; Cramer's V=0.34). Despite there being a comparatively larger proportion of university respondents reporting the strategy 'A general picture of yourself' for most items (except item MH04), and consequently, among HIV respondents, a greater likelihood of the use of strategies such as 'Specific experiences' and 'Comparing yourself to others', there were no further significant differences in response strategies between the two samples.

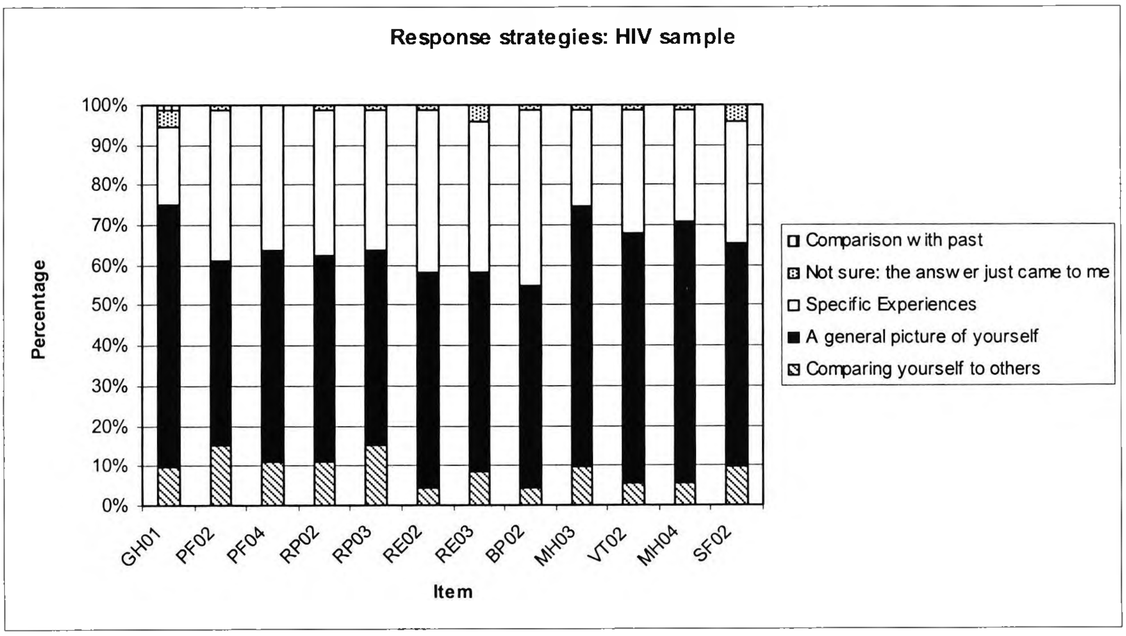
Figure 6.1:



Legend:

Health in general (GH01) / Moderate activity (PF02) / Climb several flights of stairs (PF04) / Accomplished less (RP02) / Limited in kind of activities (RP03) / Pain impact (BP02) / Accomplished less (RE02) / Did work less carefully (RE03) / Felt calm (MH03) / Lot of energy (VT02) / Felt downhearted (MH04) / Social impact of health/well-being (SF02)

Figure 6.2:



Legend:

Health in general (GH01) / Moderate activity (PF02) / Climb several flights of stairs (PF04) / Accomplished less (RP02) / Limited in kind of activities (RP03) / Pain impact (BP02) / Accomplished less (RE02) / Did work less carefully (RE03) / Felt calm (MH03) / Lot of energy (VT02) / Felt downhearted (MH04) / Social impact of health/well-being (SF02)

Analyses of patterns of strategy use by respondents

The analyses to this point have been based on the total of reported strategies over the twelve items for the two samples. In order to investigate patterns of strategy use by individual respondents, the next set of results is based on analyses of restricted university and HIV samples, comprising only those who had completed all twelve of the response strategy items. This reduced the university sample to a total of 55 respondents (12*55=660 strategy responses) and the HIV sample to 71 respondents (12*71=852 strategy responses).

Response strategies used

Table 6.2 shows reported strategy use by respondents over the twelve items: (A) the number and proportion of responses given by the restricted samples for each of the six strategies; (B) the number and proportion of respondents who reported using each strategy at least once; and (C) the mean number of reports of strategies among these respondents (the product of A / B). Proportions of strategies reported (A) were similar to those in Table 6.1, based on strategy responses for the complete samples.

In terms of strategy use by individual respondents (B), almost all had considered 'A general picture of yourself', and most had considered 'Specific experiences' for at least one item. There were some differences between samples in reporting of other strategies, although none was significant. The mean number of reports by respondents varied considerably according to strategy (C). On average, when mentioned, the strategies 'A general picture of yourself' and 'Specific experiences' were used more often than any of the others. There were sample differences in mean reports for all strategies, although large standard deviations indicated considerable variations, particularly in the HIV sample (see, in particular, 'Not sure').

Table 6.2: Reported strategy use by respondents: University and HIV samples

Response strategy	University sample			HIV sample		
	Strategy responses N (%) (A)	Respondents N (%) (B)	Mean (SD) (C)	Strategy responses N (%) (A)	Respondents N (%) (B)	Mean (SD) (C)
A general picture of yourself	432 (65.5)	54 (98.2)	8.00 (3.46)	474 (55.6)	64 (90.1)	7.41 (3.52)
Specific experiences	180 (27.3)	37 (57.8)	4.86 (2.85)	286 (33.6)	48 (67.6)	5.96 (3.74)
Comparing yourself to others	32 (4.8)	17 (30.9)	1.88 (0.86)	74 (8.7)	30 (42.3)	2.47 (2.00)
Not sure: the answer just came to me	14 (2.1)	11 (17.2)	1.27 (0.47)	17 (2.0)	6 (8.5)	2.83 (4.02)
Ruling out response options	1 (0.2)	1 (1.8)	-	0 (0.0)	0 (0.0)	-
Comparison with past	1 (0.2)	1 (1.8)	-	1 (0.1)	0 (0.0)	-
N (%)	**660 (100)**	**55 (100)**		**852 (100)**	**71 (100)**	

Response strategy combinations

So far, analyses have focused on reports of single response strategies. This section deals with the variety of different strategies used in combination by respondents in the restricted samples (complete strategy data) over the twelve items (Table 6.3). The Table shows that over a quarter of respondents in both samples reported using both 'A general picture' and 'Specific experiences' and no other strategies. In addition, a large proportion of respondents in both samples reported using only the single strategy 'A general picture', particularly among university respondents. The proportions of respondents reporting other strategy combinations also varied between samples, particularly with the inclusion of 'Not sure' among university respondents and 'Comparing yourself to others' and 'Specific experiences' among HIV respondents. Indeed, no university respondents reported that they considered only 'Specific experiences' for all twelve items. These differences resulted in a significant association between sample and response strategy combination (Exact test: $\chi^2(6)=12.53$, $p<0.05$; Cramer's V=0.32).

Table 6.3: Combination of strategies reported by respondents when completing the SF-12v2: University and HIV samples

Response strategy combinations	**University sample N (%)**	**HIV sample N (%)**
A general picture	13 (23.6)	11 (15.5)
Specific experiences	0 (0.0)	6 (8.5)
A general picture + Specific experiences	15 (27.3)	19 (26.8)
A general picture + Specific experiences + Comparing yourself to others	10 (18.2)	18 (25.4)
A general picture + Comparing yourself to others	4 (7.3)	10 (14.1)
A general picture + Specific experiences + Not sure: the answer just came to me	7 (12.7)	2 (2.8)
Other strategy combinations	6 (10.9)	5 (6.9)
N (%)	**55 (100.0)**	**71 (100.0)**

The analyses in Table 6.3 only indicates whether a particular strategy was reported by a respondent on at least one item, and therefore does not take into account the proportion of answers involving that strategy. Further analyses focused on the proportions of each of the different strategies reported across the twelve items; that is, not simply whether a strategy was employed when answering but also the number of times it was used by a respondent over the course of the measure, from 0 to 12. The largest of these combinations involved the use of a single strategy for all twelve items, reported above. However, at this level, much more varied use of strategies was identified. In the university sample, there were 34 combinations comprising different usage patterns for strategies over the twelve items, and the equivalent figure in the HIV sample was 38. These results indicate that respondents differed in the proportion of each strategy they reported across the twelve items, since most combinations included only one to three respondents.

The relative use of general and specific strategies varied widely between respondents. In terms of the balance of strategies, Table 6.4 details the proportion of general, specific and other strategies reported. As can be seen, similar, large proportions of respondents in both samples indicated that they used a general strategy for more than three quarters of the items, with an additional sizeable proportion of respondents in the two samples indicating the use of a general strategy for more than half of responses. This could indicate the use of a satisficing approach to response, whereby 'A general picture' was selected without reference to relevant health experiences. Nevertheless, most respondents did not select only 'A general picture' and a notable percentage of respondents from both samples, but particularly the HIV sample, reported the use of specific experiences for fifty percent or more responses.

Table 6.4: The proportions of general and specific strategies reported by respondents when completing the SF-12v2: University and HIV samples

	University sample N (%)	**HIV sample N (%)**
> 75% A general picture	22 (40.0)	30 (42.3)
> 75% Specific experiences	3 (5.5)	14 (19.7)
50% - 75% A general picture	13 (23.6)	12 (16.9)
50% - 75% Specific experiences	10 (18.2)	10 (14.1)
50% - 50% A general picture / Specific experiences	3 (5.5)	1 (1.4)
Other strategy combinations (neither main strategy comprising 50% of reports)	4 (7.3)	4 (5.6)
N (%)	**55 (100.0)**	**71 (100.0)**

Response strategy scales

Summative scales were derived, based on the total number of times each of four most commonly reported strategies was reported over the twelve items by respondents with complete strategy data (range 0-12). Thus, each respondent had four scale scores, each being a count of the number of times a strategy was used. Scale comparisons between university and HIV respondents are shown in Table 6.5. Strategy scale means did not differ significantly according to sample. However, the four scales deviated from the normal distribution (all Shapiro-Wilk tests significant at $p<0.001$), reflecting the effect of the predominance of reports of a general picture strategy over the twelve items. With the exception of the general picture scale, all scales were positively skewed, significantly so for some scales, with a large proportion of zero and low scores (university sample: Comparison with others: $z=4.74$, $p<0.001$; Not sure: $z=6.59$, $p<0.001$; HIV sample: Comparison with others: $z=9.16$, $p<0.001$; Specific experiences: $z=2.27$, $p<0.05$; Not sure: $z=27.09$, $p<0.001$). In addition, there was evidence of both positive kurtosis (university: Not sure: $z=5.74$, $p<0.001$; HIV: Comparison with others $z=15.83$, $p<0.001$; Not sure: $z=110.90$, $p<0.001$) and negative kurtosis (HIV sample: Specific experiences: $z=-2.13$, $p<0.05$).

Table 6.5: Summative response strategy scales (range 0-12): University and HIV samples

	Comparing to others	General picture	Specific Experiences	Not sure
University sample				
Mean	0.58	7.85	3.27	0.25
25th ptile	0.00	5.00	0.00	0.00
50th ptile	0.00	8.00	2.00	0.00
75th ptile	1.00	11.00	6.00	0.00
SD	0.99	3.59	3.27	0.55
N	**55**	**55**	**55**	**55**
HIV sample				
Mean	1.04	6.68	4.03	0.24
25th ptile	0.00	3.00	0.00	0.00
50th ptile	0.00	7.00	3.00	0.00
75th ptile	1.00	10.00	7.00	0.00
SD	1.78	4.01	4.15	1.34
N	**71**	**71**	**71**	**71**

Response strategy scale correlations

Bivariate correlations were computed for the strategy scale scores (Tables 6.6 and 6.7). All significant associations were negative, indicating that the greater use of one response strategy was related to fewer uses of another. This finding clearly relates to the nature of the scales, since a high score on one scale would be associated with correspondingly lower scores on the other response strategy scales. However, despite this, only one result was noteworthy in both samples: the very strong association between general picture and specific experiences scales. All other correlations were weaker and significance varied between samples. In the university sample, there were significant associations between the general picture scale and the two other summative strategy scales. In the HIV sample, there was a significant relationship between specific experiences and the comparison with others scale. Equivalent analyses carried out using scales restricted only to response strategies reported for either physical or mental health SF-12v2 items did not modify these findings, suggesting that the results reflected a common pattern of association between strategies, regardless of the nature of the health items concerned.

Table 6.6: Response strategy scale correlation matrix: University sample

	Comparison with others	General picture	Specific experiences	Not sure
Comparison with others	-			
General picture	-.30*	-		
Specific experiences	.05	-.95***	-	
Not sure	-.14	-.27*	.16	-

Table 6.7: Response strategy scale correlation matrix: HIV sample

	Comparison with others	General picture	Specific experiences	Not sure
Comparison with others	-			
General picture	-.07	-		
Specific experiences	-.33**	-.87***	-	
Not sure	-.09	-.21	-.08	-

* p<0.05 **p<0.01 ***p<0.001

Analyses of response strategy and the SF-12v2

Response strategies were investigated in relation to the summated scales over twelve items for the two main strategies, 'A general picture of yourself' and 'Specific experiences'. Second, the relationship between response process and the SF-12v2 was assessed at an item level. In terms of the response process scales, higher SF-12v2 scores (indicative of better health) related to a greater use of general perceptions when answering. Over a number of items, the use of a general picture strategy was associated with better perceived health.

Response strategies and SF12v2 scales

The two main response strategies reported for the SF-12v2 in both studies were based on 'A general picture of yourself' and 'Specific experiences'. Summated scales based on respondents' total reported use of each of these strategies were correlated with the SF-12v2.

For the general picture strategy scale, most of the correlation coefficients were positive, reflecting a greater use of this strategy among those with higher SF-12v2 scores, indicating better functioning. In the university sample, associations were significant for three mental health scales (SF: 0.38, $p<0.01$; MH: 0.39, $p<0.01$; MCS: 0.34, $p<0.01$). In the HIV sample, there were a greater number of significant results, perhaps relating to the wider variation of scoring in this sample. Results were significant and positive for most physical scales (PF: 0.33, $p<0.01$; RP: 0.32, $p<0.01$; GH: 0.32, $p<0.01$; PCS: 0.28, $p<0.05$) and all mental health scales (VT: 0.35, $p<0.01$; SF: 0.34, $p<0.01$; RE: 0.33, $p<0.01$; MH: 0.27, $p<0.05$; MCS: 0.30, $p=0.01$).

The equivalent analyses conducted with the specific experiences strategy resulted in negative associations between strategy and the SF-12v2, demonstrating that those who reported a greater use of specific experiences to generate a response were more likely to have lower perceived health scores.

These findings indicate the greater use of the strategy of considering general perceptions among those reporting better functioning, and conversely, a greater consideration of specific experiences by those with poorer functioning. Together with the strong inverse relationship between general picture and specific experiences strategies, these results seem to indicate that the use of these two strategies alone appears to be able to identify notable differences in SF-12v2 response styles and subsequent response in the samples

Response strategy and individual SF-12v2 item responses

The association between reported response strategies (recoded general picture, specific experiences and other strategy combined) and item responses were examined for each of the SF-12v2 items (full results for these analyses are shown in Appendix 25). Over the course of the twelve items a pattern was identified, with those respondents reporting better physical or mental functioning on an item being more likely to indicate that they had used the strategy of considering a general picture of themselves when answering (Table 6.8). Significant results reported in the Table refer to a relationship between good functioning (the SF-12v2 item dichotomised) and strategy, with some exceptions: general health (GH01), feeling calm and peaceful (MH03), and energy or vitality (VT02), for which extreme response categories were combined. The pervasiveness of the relationship between strategy and response in both samples for most items indicates that positive functioning was judged as a general perception whereas health problems were more commonly recalled in relation to specific experiences. The fact that this was evident in disparate samples, reporting very different levels of health problems, suggests that this finding relates to general psychological processes. The only item for which there was clear evidence of another relationship between strategy and response was item MH03 (feeling calm and peaceful). One possible explanation for the U-shaped relationship is that, in contrast to most of the other physical and mental health items, this item draws on general personality perceptions at the extremes of response.

Table 6.8: Relationship between strategy and SF-12v2 response: University and HIV samples

	University		HIV	
	Strategy/response	**Significance**	**Strategy/response**	**Significance**
GH01	+		+	†
PF02	+		+	†
PF04	+		+	
RP02	+	†	+	
RP03	+		+	†
RE02	+		+	†
RE03	+		+	
BP02	+		+	†
MH03	U	†	U	†
VT02	+	†	+	†
MH04	+	†	+	
SF02	+	†	+	†

+ 'General picture of yourself' strategy associated with better functioning

U Curvilinear relationship between 'General picture of yourself' strategy and functioning

† Significant relationship, p<0.05 or less

Legend:
Health in general (GH01) / Moderate activity (PF02) / Climb several flights of stairs (PF04) / Accomplished less (RP02) / Limited in kind of activities (RP03) / Accomplished less (RE02) / Did work less carefully (RE03) /Pain impact (BP02) Felt calm (MH03) / Lot of energy (VT02) / Felt downhearted (MH04) / Social impact of health/well-being (SF02)

6.2 A psychometric analysis to identify an easiness scale for the SF-12v2

Overview

In both samples, respondents were generally likely to rate the SF-12v2 items and the overall measure to be easy to answer. Psychometric analyses were undertaken to test the item characteristics. These results suggest that respondents in both samples can meaningfully provide ratings of the easiness of the SF-12v2 to complete, both in relation to individual SF-12v2 items and also an overall rating of the SF-12v2 as a whole. Ratings were shown to vary by item but also to form an internally consistent scale. Most respondents considered the SF-12v2 to be easy to answer, although more respondents in the HIV sample rated to measure 'very easy' to answer, compared to the university sample. However, easiness scores in the two samples were not significantly different.

The relationship between SF-12v2 and easiness was generally curvilinear at the item level, indicating that responses at the extreme of the response continuum is considered easier than having to refer to more central responses, possibly reflecting the systematic cognitive processing required to answer a question which is considered not straightforward (Jenkinson et al, 1996; Rothman and Schwartz, 1998).

Item analyses

The distributions of the twelve easy items were examined (item distributions: Appendix 20): first, item facility; second, convergent validity. Appendix 26 provides a detailed description of the results. Facility scores indicated that, on average, respondents in both samples rated all items easy to answer.

Response distributions in each sample for the perceived easiness of the SF-12v2 items to answer are shown in Figures 6.3 and 6.4. Analyses according to response revealed that HIV respondents were more likely than university respondents to rate items 'very easy' to answer: the modal response for most items in the university sample was 'fairly easy' (9/12, 75%) and in the HIV sample 'very easy' (10/12, 83%).The only significant (or borderline significant) associations between sample and item rating of easiness related to a greater proportion of university respondents who considered the Mental Health scale items more difficult (Exact tests: MH03: $\chi^2(2)=5.48$, $p=0.06$; Cramer's V: 0.20; MH04: $\chi^2(3)=6.77$, $p=0.06$; Cramer's V: 0.22), as well as the Vitality scale item (Exact test: VT02: $\chi^2(2)=6.21$, $p<0.05$; Cramer's V: 0.21). However, a notable proportion of respondents in both samples rated the role limitation items either fairly or very difficult to answer. Other items rated difficult by a number of respondents differed by sample.

Reliability was satisfactory when items were combined to form a scale (university sample: $\alpha = 0.85$; HIV sample: $\alpha = 0.86$). In addition, items correlated strongly with a summated easiness scale, derived from the twelve items.

Figure 6.3:

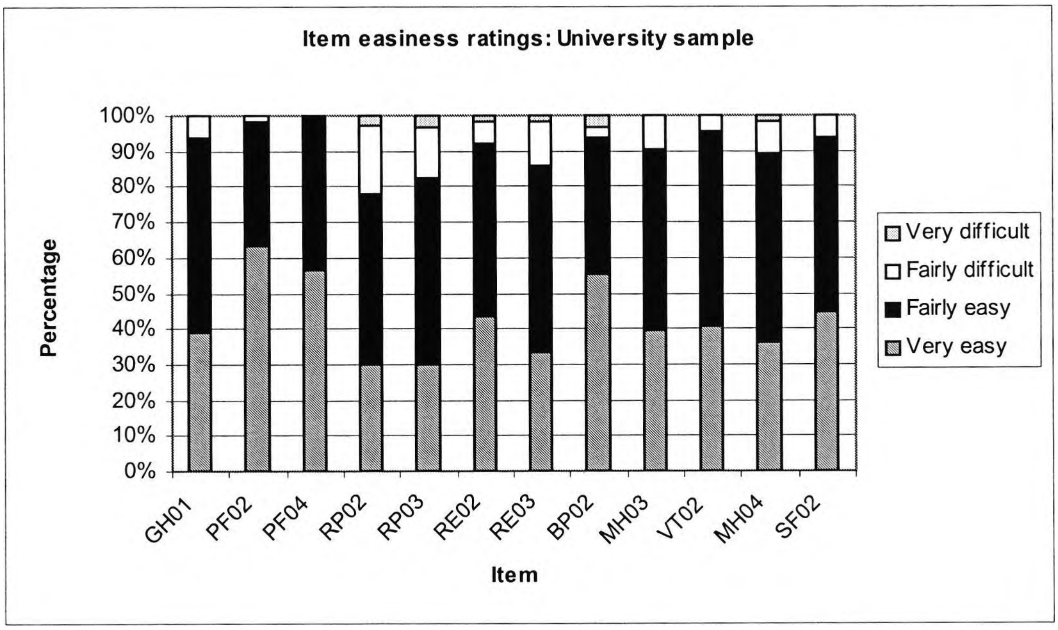
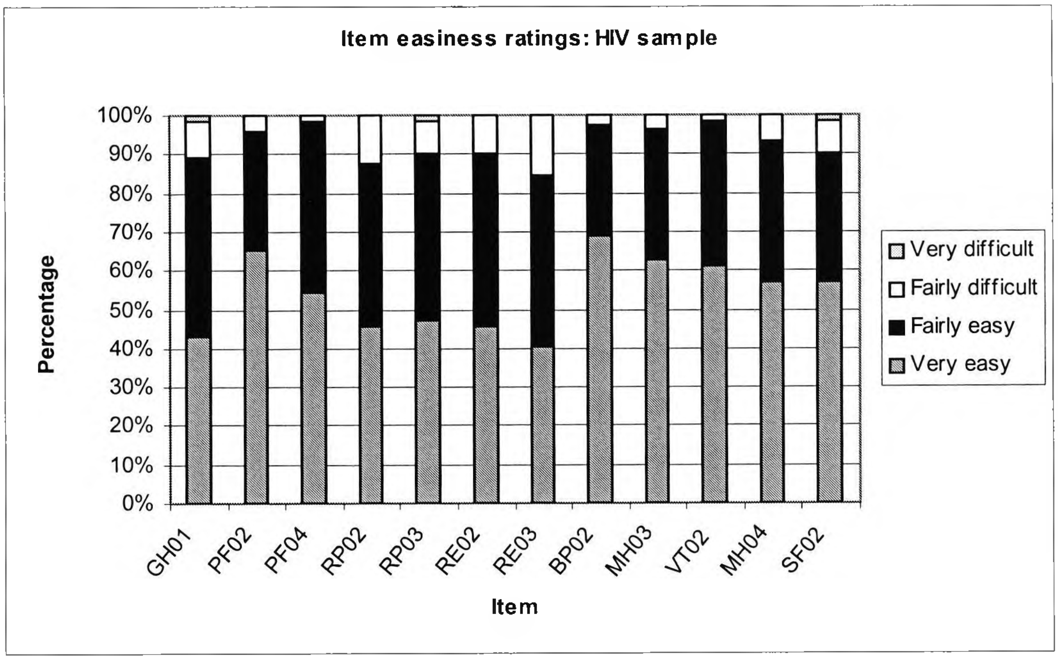


Figure 6.4:



Legend:
Health in general (GH01) / Moderate activity (PF02) / Climb several flights of stairs (PF04) / Accomplished less (RP02) / Limited in kind of activities (RP03) / Pain impact (BP02) / Accomplished less (RE02) / Did work less carefully (RE03) / Felt calm (MH03) / Lot of energy (VT02) / Felt downhearted (MH04) / Social impact of health/well-being (SF02)

Scale analyses

When summed, the derived easiness scale in both samples deviated significantly from normal, relating to the high proportion of respondents who considered items easy to answer. Although scores of the scale were higher in the HIV sample (indicating that more HIV respondents rated items easy), scores did not differ significantly between samples (Table 6.9).

Table 6.9: Easy item scale descriptives: University and HIV samples

	HIV sample	**University sample**
Mean	19.88	18.42
25th ptile	16.00	14.00
50th ptile	20.00	18.00
75th ptile	23.25	23.00
SD	4.88	4.72
N	**58**	**71**

Overall rating of ease of completion

After completion of all twelve items, respondents were asked to give an overall impression of the easiness or difficulty of the SF-12v2 to answer. Most respondents felt that the SF-12v2 was easy to answer. In the university sample, over 96.9% of respondents rated the SF-12v2 easy to answer ('very easy': 13, 20.3%; 'fairly easy': 49, 76.6%). Only two respondents answered 'fairly difficult' (3.1%) and none answered 'very difficult'. In the HIV sample, a similarly large proportion of respondents felt that, overall, the SF-12v2 was easy to answer (94.4%), with the remainder selecting 'fairly difficult' (4, 5.6%). However, 'very easy' and 'fairly easy' were selected by almost equal numbers of respondents ('very easy': 33, 45.8%; 'fairly easy': 34, 48.6%). The larger proportion of HIV respondents who rated the SF-12v2 'very easy' to answer compared to university respondents resulted in a significant association between sample and the overall rating of easiness for the SF-12v2 (Exact test: $\chi^2(2)=11.26$, $p<0.05$).

Relation between easy item scale and perceptions of overall easiness

Tests were carried out to investigate whether the composite easy item scale related to overall perceptions of SF-12v2 easiness (Table 6.10). In both samples, a linear trend was identified: the more easy the SF-12v2 was considered overall, the more individual SF-12v2 items received ratings of 'fairly easy' or 'very easy' (resulting in lower scores on the easy scale). This trend was found to be significant (university sample: $F(2, 55)=17.46$, $p<0.001$; HIV sample: $F(2, 68)=26.89$, $p<0.001$). Tukey HSD post hoc pairwise comparisons indicated significant differences in the easy item scale. For the university sample, all three comparisons were significant: very easy-fairly easy ($p<0.001$), fairly easy-fairly difficult ($p<0.05$), and very easy-very difficult ($p<0.001$); and for the HIV sample, there were two significant comparisons: very easy-fairly easy, very easy-fairly difficult (both $p<0.001$). The relationship between the easy item scale and overall perception of easiness remained significant in both samples following the exclusion of the 'fairly difficult' overall rating category, which was selected by few respondents.

Table 6.10: Relationship between easy item scale scores and perceptions of overall easiness of the SF-12v2: University and HIV samples

	University sample		**HIV sample**	
Overall rating	**Total**	**Mean (SD)**	**Total**	**Mean (SD)**
Very easy	11	14.45 (2.54)	33	15.24 (3.44)
Fairly easy	45	20.80 (4.09)	34	20.74 (3.82)
Fairly difficult	2	29.00 (5.66)	4	25.00 (1.83)

Relationships between the easy item scale and SF-12v2 scales

Associations between the summative easy item scale, comprising easiness ratings for all SF-12v2 items, and the SF-12v2 scales are shown for both samples in Table 6.11. In the university sample, the easy item scale was negatively related to all SF-12v2 scales, significantly so for six of the ten coefficients (two physical health and four mental health scales). Since higher scores on the SF-12v2 are suggestive of better functioning while higher scores on the easy item scale indicate that fewer of the twelve items are rated easy to answer, a negative relationship indicates that those with better functioning were more likely to rate SF-12v2 items easy to answer. In contrast, there were no significant correlations between the easy scale and the SF-12v2 scales in the HIV sample, indicating that ratings of easiness were unrelated to health scores among these respondents. However, the relationship between composite easiness ratings and SF-12v2 scales was not linear, and therefore further analyses were carried out into associations at the individual item level.

Table 6.11: Correlations between the easy scale and the SF-12v2: University and HIV samples

Scale	University sample	HIV sample
PF	-0.30*	0.09
RP	-0.32*	-0.03
BP	-0.25	0.14
GH	-0.21	0.13
VT	-0.29*	0.02
SF	-0.26*	-0.15
RE	-0.36**	-0.07
MH	-0.22	-0.10
PCS	-0.25	0.14
MCS	-0.26*	-0.15

* <0.05 ** <0.01

Legend:
PF (Physical Functioning) / RP (Role Physical) / BP (Bodily Pain) / GH (General Health) / VT (Vitality) / SF (Social Functioning) / RE (Role Emotional) / MH (Mental Health) / PCS (Physical Component Scale) / MCS (Mental Component Scale)

Analyses of individual easiness ratings and SF-12v2 item responses

The relationship between easiness rating and response was influenced by the proportions of respondents who selected each of the SF-12v2 item response categories (full results: Appendix 26). For example, fewer of the respondents from the university sample selected poor health options in comparison to HIV respondents. In terms of investigating relationships, those responses selected by larger numbers of respondents, such as a 'no limitations' category, had a greater influence on results than other categories; while, in contrast, proportions in smaller categories may appear more noteworthy since percentage calculations are based on fewer respondents. Nevertheless, Table 6.12 summarises the patterns of findings over the twelve items for each sample, and indicates whether a relationship has been shown to be significant. In both samples, there are a number of curvilinear relationships, either J-or U-shape. The comparatively greater number of U-shaped relationships in the HIV sample may result from the wider range of responses in this sample in contrast to the predominance of good functioning in the university sample: that is, all curvilinear relationships may be indicative of an underlying U-shaped relationship. Some relationships appeared linear, whereby the items were more likely to be rated easy to answer among those with good functioning (for example, PF items in the university sample). Few of the analyses resulted in significant results, particularly for the HIV sample. This may reflect the complex relationships between the two variables and the small proportions of respondents in many cells, even after recoding. Following from this, the question arises, what do these relationships mean? A response style explanation could be presented in which those who provide extreme responses on one item are likely to do so on another (resulting in a U-shape). However, as mentioned, not all the relationships between the items outlined were U-shaped. There is evidence to suggest that those with strong attitudes are more likely to be able to access them more easily (Liberman and Chaiken, 1996). In addition, schema theory proposes that people have automatic schemata about routine or more central aspects of their self-concept. Since those with extreme perceptions of their health (good or poor functioning) may have stronger attitudes, or may have health schema, it is possible that they find the task easier to carry out than others who have to consider their response in more detail (Rothman and Schwartz, 1998). This may also relate to the earlier observation that HIV respondents are more likely to rate the SF-12v2 easier to complete than those from the university sample, regardless of their overall health scores.

Table 6.12: Relationship between easiness ratings and SF-12v2 response: University and HIV samples

	University		HIV	
	Strategy/response	**Significance**	**Strategy/response**	**Significance**
GH01	+	†	U	
PF02	+	†	U	
PF04	+	†	U	†
RP02	J	†	U	
RP03	J	†	U	†
RE02	U	†	U	
RE03	J	†	U	
BP02	U		U	
MH03	U		U	
VT02	U		+	
MH04	U		U	
SF02	U		U	

+ Linear association between easiness and functioning

A 'U' entry represents a U-shaped association between easiness and functioning

A 'J' entry represents a J-shaped association between easiness and functioning

† Significant relationship, $p<0.05$ or less

Legend:
Health in general (GH01) / Moderate activity (PF02) / Climb several flights of stairs (PF04) / Accomplished less (RP02) / Limited in kind of activities (RP03) / Accomplished less (RE02) / Did work less carefully (RE03) / Pain impact (BP02) / Felt calm (MH03) / Lot of energy (VT02) / Felt downhearted (MH04) / Social impact of health/well-being (SF02)

In summary, these results suggest that respondents in both samples can meaningfully provide ratings of the easiness of the SF-12v2 to complete, both at the level of individual items and also overall. Ratings were shown to vary by item but also to form an internally consistent scale. In addition, individual ratings of the easiness of items to complete were associated with overall ratings of easiness of the SF-12v2. Ratings differed between samples, but not significantly. However, overall, more HIV respondents rated the SF-12v2 easy to answer. The relationship between SF-12v2 and ratings of easiness was non-linear, indicating that respondents with poor or good health were more likely to consider the item easier to complete than other respondents.

6.3 A psychometric analysis to identify a usefulness scale for the SF-12v2

Overview

In this study, the attitude strength concept of importance was operationalised in terms of perceived usefulness of an item to measure health: items rated useful were considered to relate more strongly to self-health attitudes (Sehulster, 1994). In addition, it has been argued that relevant items are likely to have greater face validity, which may influence the empirical validity of a measure (Holden and Jackson, 1979). Psychometric analyses were undertaken to test the characteristics of the items on perceived usefulness of the SF-12v2 for measuring respondents' health. In both samples, respondents generally considered the SF-12v2 items to be useful. HIV respondents were more likely to rate the items useful than the university sample. However a notable proportion of both samples rated items not useful. Responses to the usefulness items were more varied than easiness. Psychometric analyses indicated that the usefulness items were internally consistent as a scale and that items correlated with a scale total and an overall rating of usefulness (significantly so in the HIV sample). These findings indicate that usefulness can be measured for the SF-12v2 measure and that ratings differ between samples. In addition, usefulness appeared to relate to the SF-12v2, although there was considerable diversity in ratings of usefulness for any SF-12v2 response. Those who considered that they had poorer health were more likely to rate the SF-12v2 useful, a finding which relates to the initial hypothesis that importance relates to more pertinent issues.

Item analyses

Data distributions are provided in Appendix 20. The distributions of the twelve items of the SF-12v2 were examined: first for item facility, and then for item convergent validity. These data, along with information on data completeness, are included in

Appendix 27. None of the items attained an extreme facility score in either sample, and there was more variation in the usefulness ratings than had been observed in the ratings of easiness. In the university sample, all facility scores indicated that, on average, respondents were likely to rate the items as useful rather than not useful. In the HIV sample, the facility scores for all items were lower, indicating that items were more commonly rated useful among these respondents.

In comparison to the ratings of SF-12v2 item easiness, usefulness ratings were more varied, suggesting more diverse opinions. Figures 6.5 and 6.6 show the item distributions for each item in the two samples. In both samples, all response options were selected for every item. The modal response in the university sample for all items was 'fairly useful', selected by more than half of respondents (except item RP02: 'fairly useful' selected by 46.0% of the sample). In the HIV sample, modal response for items was divided between 'very useful' (5 of the 12 items, 41.7%) and 'fairly useful' (7/12, 58.3%). All items received a rating of 'not that useful' or 'not at all useful' by a sizeable minority. The associations between rating and sample were significant for three physical health items (Exact tests: PF04: $\chi^2(3)=9.13$, $p<0.05$; Cramer's V: 0.26; RP02: $\chi^2(3)=8.93$, $p<0.05$; Cramer's V: 0.26; BP02: $\chi^2(3)=12.68$, $p<0.01$; Cramer's V: 0.31) and significant (or borderline significant) for the ratings of four mental health items (Exact test: RE02: $\chi^2(3)=14.48$, $p<0.001$; Cramer's V: 0.33; MH03: $\chi^2(3)=10.35$, $p<0.05$; Cramer's V: 0.28; MH04: $\chi^2(3)=17.07$, $p<0.001$; Cramer's V: 0.35; SF02: $\chi^2(3)=6.78$, $p=0.07$; Cramer's V: 0.22).

Internal consistency of the usefulness items was good for both samples (university sample: $\alpha=0.88$; HIV sample: $\alpha=0.93$). The twelve items used to rate the perceived usefulness of each SF-12v2 item were summed to produce a single scale and there were strong correlations between items and the overall usefulness scale.

Figure 6.5:

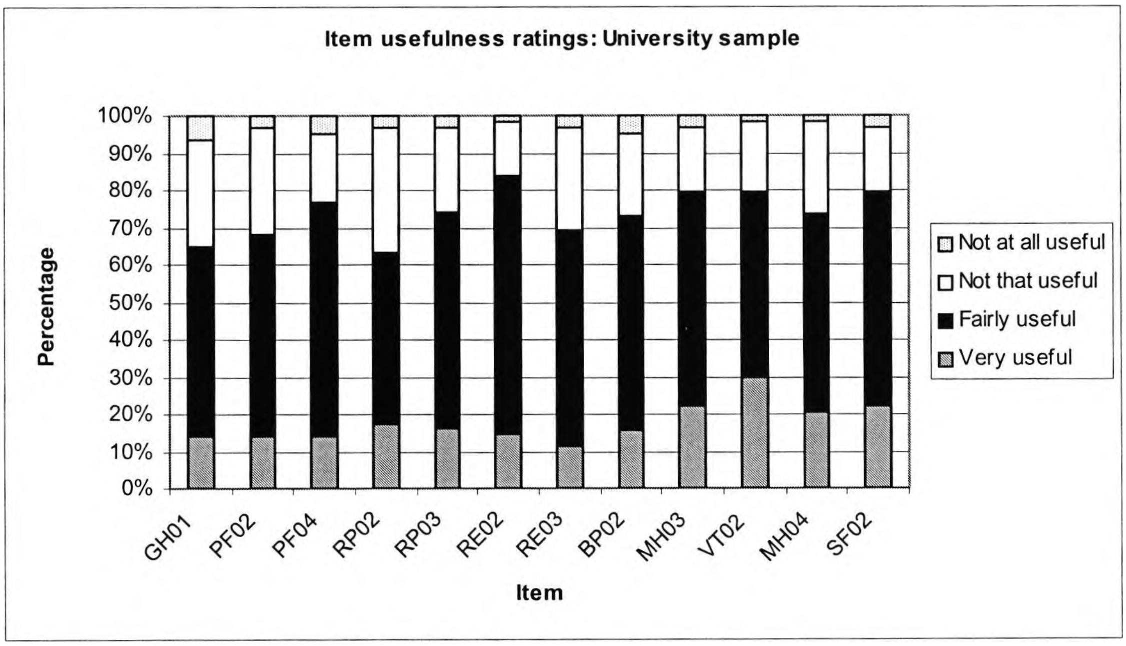
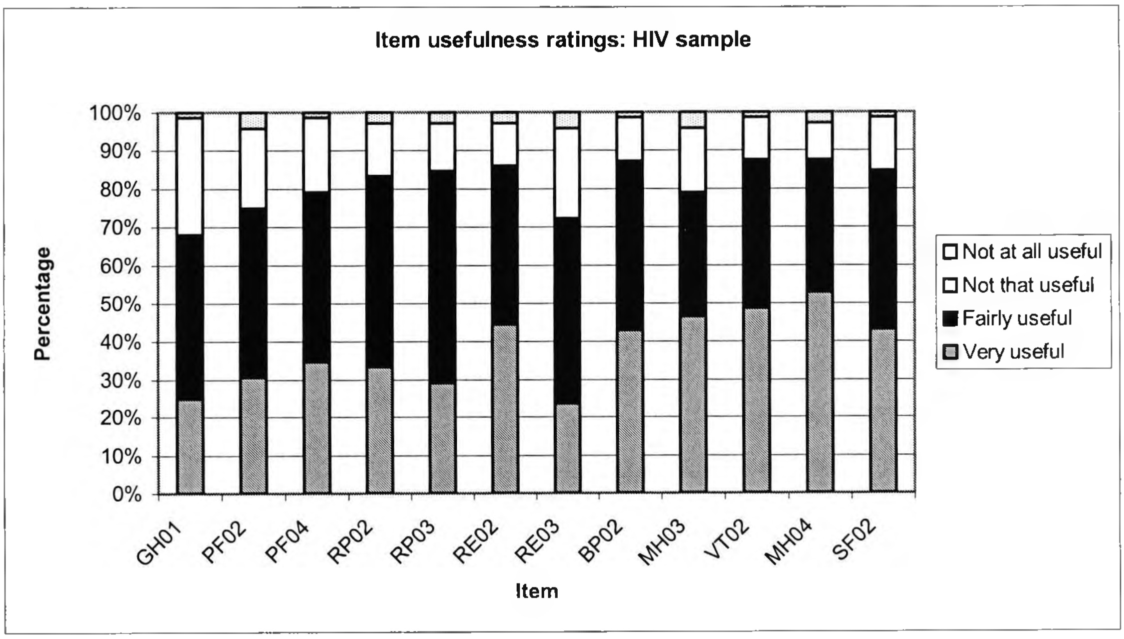


Figure 6.6:



Legend:
Health in general (GH01) / Moderate activity (PF02) / Climb several flights of stairs (PF04) / Accomplished less (RP02) / Limited in kind of activities (RP03) / Pain impact (BP02) / Accomplished less (RE02) / Did work less carefully (RE03) / Felt calm (MH03) / Lot of energy (VT02) / Felt downhearted (MH04) / Social impact of health/well-being (SF02)

Scale analyses

The distribution of both scales differed significantly from normal (Table 6.13). The mean score for item usefulness was higher in the university sample compared to the HIV sample, reflecting fewer items rated useful among university respondents. This difference was statistically significant ($t(125)=2.86$, $p<0.01$).

Table 6.13: Useful item scale descriptive statistics: University and HIV samples

	University	**HIV**
Mean	25.33	21.99
25th ptile	22.00	16.75
50th ptile	25.00	21.00
75th ptile	28.00	25.00
SD	5.89	7.05
N	**57**	**70**

Overall rating of usefulness of completion

After completion, respondents were asked to give an overall impression of the usefulness of the SF-12v2 for assessing their current health. Most respondents in both samples considered the SF-12v2 to be a useful for assessing their health. In the university sample, 79.4% rated the SF-12v2 useful ('very useful': 6, 9.4%; 'fairly useful': 45, 70.%). The remaining respondents rated the SF-12v2 'not that useful' (13, 20.3%). In the HIV sample, a similar proportion of respondents considered that, overall, the SF-12v2 was a useful measure for assessing their health ('very useful' (22, 30.6%); 'fairly useful' (36, 50.0%)). Most of the remainder considered it 'not that useful' (12, 16.7%) and two respondents answered 'not at all useful' (2.8%). It is clear that respondents in both samples considered the SF-12v2 to be a useful instrument, although this belief was expressed more strongly in the HIV sample, resulting in a significant association between sample and overall rating of usefulness (Exact test: $\chi^2(3)=11.75$, $p<0.01$; Cramer's V: 0.29).

Relation between useful item scale and perceptions of overall usefulness

In both samples, there was a significant linear relationship between the usefulness scale, comprising the summed ratings of the usefulness of individual SF-12v2 items, and an overall rating of SF-12v2 usefulness (university sample: $F(2, 54)=28.50$, $p<0.001$; HIV sample: $F(3, 66)=50.97$, $p<0.001$) (Table 6.14). The less useful the SF-12v2 was considered overall, the more individual SF-12v2 items received ratings of 'not at all useful' or 'not that useful' (resulting in higher scores on the usefulness scale). Post hoc parametric statistical comparisons using Tukey HSD revealed significant differences in the usefulness scale. In the university sample significant pairwise comparisons were identified between: very useful-fairly useful ($p<0.001$), fairly useful-not that useful ($p<0.001$), and very useful-not that useful ($p<0.001$). In the HIV sample, when the judgement category 'not at all useful' (selected by only a single respondent) was excluded, the result remained significant ($F(2, 66)=56.21$, $p<0.001$), and all pairwise comparisons were significant at $p<0.001$.

Table 6.14: Relationship between usefulness item scale scores and perceptions of overall usefulness of the SF-12v2: University and HIV samples

	University sample		HIV sample	
Overall rating	**Total**	**Mean (SD)**	**Total**	**Mean (SD)**
Very useful	6	16.50 (5.17)	21	15.90 (3.82)
Fairly useful	40	24.78 (3.39)	36	21.82 (3.56)
Not that useful	11	25.33 (5.89)	12	31.08 (5.21)
Not at all useful	0	0 (0.00)	1	47.00 (0.00)

In summary, as with the easiness items, these results suggest that respondents can rate their subjective perceptions of the qualities of the SF-12v2, such as usefulness, both in relation to individual items and the SF-12v2 as a whole. The scale was internally consistent and related to overall ratings provided by respondents. The ratings in this research differed according to sample: HIV respondents were more likely to consider the items and the overall scale useful for assessing their current health. This appeared to relate to the importance of these health issues to them in comparison to a healthy respondent, to whom health questions are irrelevant.

Relationships between the useful item scale and SF-12v2 scales

Correlations were computed to investigate evidence for any associations between the composite usefulness item scale, comprising usefulness ratings for all items, and SF-12v2 scale scores (Table 6.15). In both samples, most correlations were negligible, indicating none or only a weak relationship between usefulness ratings and scale scores. For the university sample, there were no significant correlations. In the HIV sample, there was only one significant correlation (RE), a positive association, indicating that those with worse health status were more likely to rate the SF-12v2 items as useful. However, this result could be spurious, since it was a single significant association out of ten correlations carried out. Nevertheless, all associations were positive, supporting the earlier finding that those with poorer health were more likely to consider the SF-12v2 scales to be useful for measuring their current health.

Table 6.15: Correlations between the usefulness scale and the SF-12v2: University and HIV samples

Scale	University	HIV
PF	0.15	0.12
RP	0.03	0.09
BP	0.01	0.17
GH	-0.21	0.16
VT	-0.16	0.02
SF	0.06	0.09
RE	-0.03	0.25*
MH	0.03	0.01
PCS	-0.01	0.13
MCS	-0.02	0.08

* $p<0.05$

Legend:
PF (Physical Functioning) / RP (Role Physical) / BP (Bodily Pain) / GH (General Health) / VT (Vitality) / SF (Social Functioning) / RE (Role Emotional) / MH (Mental Health) / PCS (Physical Component Scale) / MCS (Mental Component Scale)

Analyses of individual usefulness ratings and SF-12v2 item responses

Detailed analyses of the associations between each usefulness item and the response to the matching SF-12v2 item are shown in Appendix 27. Relationships between usefulness ratings and response were influenced by the proportions of respondents who selected each of the SF-12v2 item response categories. Few university respondents indicated that they experienced poor health, limiting the conclusions that can be drawn about the relationship between health status and usefulness ratings. However, some patterns were discernible in both samples (Table 6.16). An apparent J-shaped relationship was common, whereby those respondents who considered that they had problem with their physical or emotional health were more likely to rate the item as a useful means of assessment. This finding indicates that the health problems included among the twelve items of the SF-12v2 were considered to be useful health constructs, in particular by those who rated themselves as having worse health on these items.

While the J-shape relationship was more apparent for mental health items, it was also identified for physical health items. More of the J-shaped relationships were identified in the HIV sample, although this may have resulted from the wider range of SF-12v2 responses. Most of the associations between usefulness ratings and response were non-significant, with slightly more significant results attained in the HIV sample. Even though the J-shaped relationship was commonly identified, complex patterns of results were also noticeable. For example, in the university sample, one of the J-shaped relationships was reversed whereby those with better health were more likely to rate the item useful (RE02). In addition, there was wide variation in the ratings of usefulness, even within the same SF-12v2 response options; with some respondents rating an item ‘Very useful’ while others rated the same item ‘Not at all useful’. This heterogeneity hinders attempts to identify clear relationships between usefulness and response but does highlight the range of perceptions of the usefulness of items, even among those respondents who apparently share the same health status.

Table 6.16: Relationship between usefulness ratings and SF-12v2 response: University and HIV samples

	University		HIV	
	Usefulness	**Significance**	**Usefulness**	**Significance**
GH01				
PF02	J		J	
PF04			J	
RP02				
RP03	J		J	
RE02	JR	†	J	†
RE03			J	
BP02			J	†
MH03	J	†	J	†
VT02	J	†	J	†
MH04	J	†	J	†
SF02	J	†	J	†

A 'J' entry represents a J-shaped association between usefulness and functioning

A 'JR' entry represents J-shaped (reversed) association between usefulness and functioning

† Significant relationship, p<0.05 or less

Legend:
Health in general (GH01) / Moderate activity (PF02) / Climb several flights of stairs (PF04) / Accomplished less (RP02) / Limited in kind of activities (RP03) / Accomplished less (RE02) / Did work less carefully (RE03) / Pain impact (BP02) / Felt calm (MH03) / Lot of energy (VT02) / Felt downhearted (MH04) / Social impact of health/well-being (SF02)

These analyses showed that respondents could rate usefulness reliably. The SF-12v2 was rated more useful by the HIV sample. This relates to the suggestion that health is considered more important among those with more health experiences (Hays et al, 2002), and that those with health problems are more likely than most other respondents to rate health important (Bowling, 1995). This is supported by the finding from the current research that SF-12v2 scores are related to usefulness, whereby poorer health status was associated with perceived greater usefulness. However, there was considerable diversity in ratings of usefulness, even within a SF-12v2 response category.

Summary

The response process was assessed in terms of the reported main response strategy adopted, along with perceptions of the task: easiness of each item to complete (considered to relate to the cognitive complexity), and usefulness of the item as a rating health (a proxy for the attitude strength indicator of importance).

The most commonly reported response strategy was 'A general picture of yourself', followed by 'Specific experiences'. Adoption of particular strategies differed by sample and item. For example, more respondents reported a general picture when rating their overall health, which is in line with findings reported elsewhere, indicating that vague questions are more likely to trigger heuristic strategies rather than the recall of specific events (Schwartz et al, 1987). The general perception was more commonly reported by university respondents, suggesting that heuristics were used in the absence of health problems (Rothman and Schwartz, 1998). In contrast, specific experiences were more commonly reported by HIV respondents. In both samples, when summed, the general picture strategy was strongly inversely associated with the use of recollected specific experiences, suggesting that the greater use of one of these strategies rules out the use of the other. There were positive associations between the cumulative use of general perceptions and health status scores, indicating that this strategy was more common among those respondents with better perceived health. This pattern was stronger in the HIV sample, for which the SF-12v2 scores were more varied, including more respondents with poor health. Conversely, the use of specific experiences was associated with poorer health in both samples. These findings indicate that strategy can be used to differentiate respondents according to the nature of cognitive processing undertaken.

Ratings for ease of completion also differed by item, respondent and sample. However, the SF-12v2 was considered easy to complete by most respondents in both samples, with more HIV respondents considering it to be very easy. Rating of easiness could be measured in a scale with good psychometric characteristics. The relationship between ease of completion and SF-12v2 was not linear and there was evidence that respondents who gave an extreme SF-12v2 response, indicating limitation or no limitations, were more likely to rate items as easy to answer than those who gave other responses. This suggests that those who considered themselves healthy and those with serious chronic illnesses were more likely to regard health questions easier to answer than those with variable health problems, for whom a complex judgement was required, weighing up the nature and impact of disability (Jenkinson et al, 1996). The curvilinear relationship was more apparent in the HIV sample and for mental health items in the university sample, where the range of SF-12v2 responses were greater.

Personal importance has been considered to be one of the key defining features of attitude strength, reflecting both the nature of stored evaluations and influencing cognitive processing (Krosnick and Smith, 1994). Personally relevant attitudes have been shown to be qualitatively different from low relevance attitudes (Liberman and Chaiken, 1996). In research on health status, personal relevance has been related to the response strategy adopted, leading to greater use of systematic processing (Rothman and Schwartz, 1998). In addition, it has been argued on psychometric grounds that item meaningfulness or importance are crucial for the validity of a measure (Bjorner et al, 2003; Messick, 1995; Tulsky and Rosenthal, 2002), and that validation should include domains measured that are important to patients rather than the researcher (May and Warren, 2002). Studies comparing the way in which health is rated have shown that items differ in terms of importance according to health state (Carver et al, 1999; Rothwell et al, 1997).

There is no single question which is generally used to gauge attitude importance, and therefore it was operationalised in relation to health in terms of "how useful" the respondent considered each item and the overall SF-12v2 for measuring their current health, capturing the key element of personal relevance (Krosnick and Smith, 1994). Ratings of usefulness for measuring respondents' health differed by item and sample, with more variation in ratings than ease of completion, and could be summed to form a psychometrically reliable scale. The majority of university and HIV respondents rated the SF-12v2 useful. However, it has been suggested that health is more important among those with more health experience (Hays et al, 2002), and there is evidence that, when asked to rate what was important to their lives, those with health problems are more likely than most other respondents to rate health important (Bowling, 1995). Similarly, in this research, the SF-12v2 was rated more useful among those with lower scores, particularly in the HIV sample. The relationship between ratings of usefulness and SF-12v2 score was complex, due to the diversity of ratings of usefulness in relation to levels of SF-12v2 item response. However, there appeared to be a pattern whereby those who rated their health as poor were more likely to rate items useful: due to the composition of the two samples, this was particularly apparent among HIV respondents. This finding may suggest that the issues covered by the SF-12v2 were considered more useful to those with health problems, although it could also indicate that there was a shared bias to report health problems and usefulness, or that respondents felt some pressure to rate items as useful, although this appeared not to be uniform across health states. In addition, although the items were generally considered useful, this does not indicate that respondents would not have wished for other items to be included.

The relationships demonstrated thus far between the SF-12v2 scale scores, context and response process provide support for both the construct validity of the SF-12v2 and the value of quantitatively investigating response processes. Finally, the constituent elements of the quantitative study were drawn together: contextual factors, and response process, and SF-12v2 scores, in a single model.

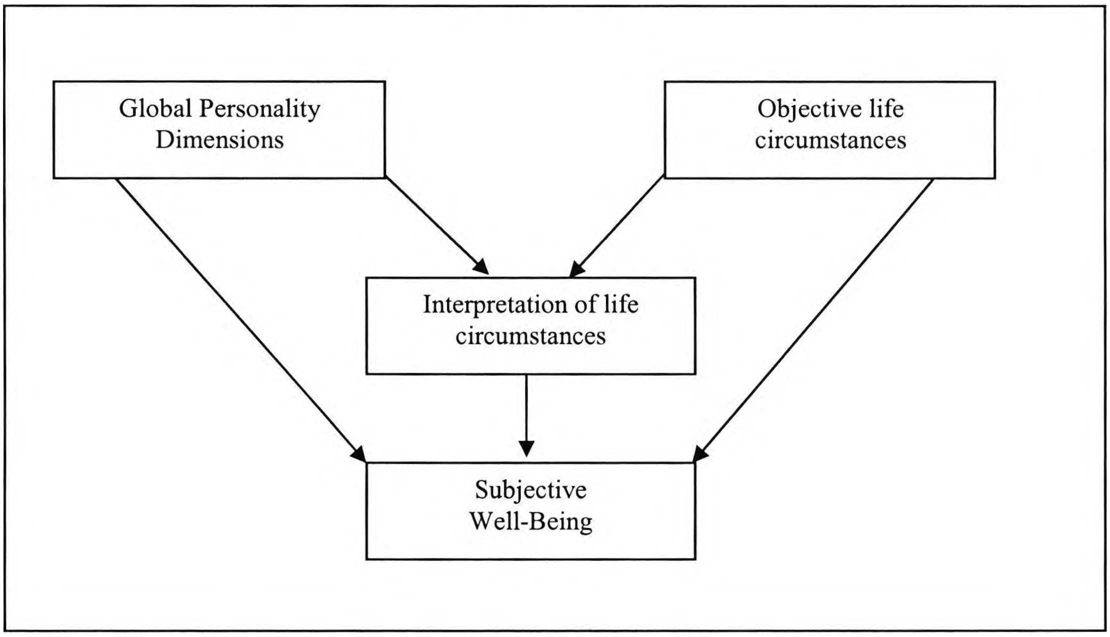
7 A contextual path model for the SF-12v2

Introduction

The previous sections have shown that the SF-12v2 is related both to contextual factors, and also to variables measuring aspects of the response process. The final analysis in this thesis tested a response process path model, linking contextual factors, response strategy and SF-12v2 scores. The same pathways were included in all models in order to investigate whether paths vary according to sample and mental or physical health summary scale (PCS and MCS).

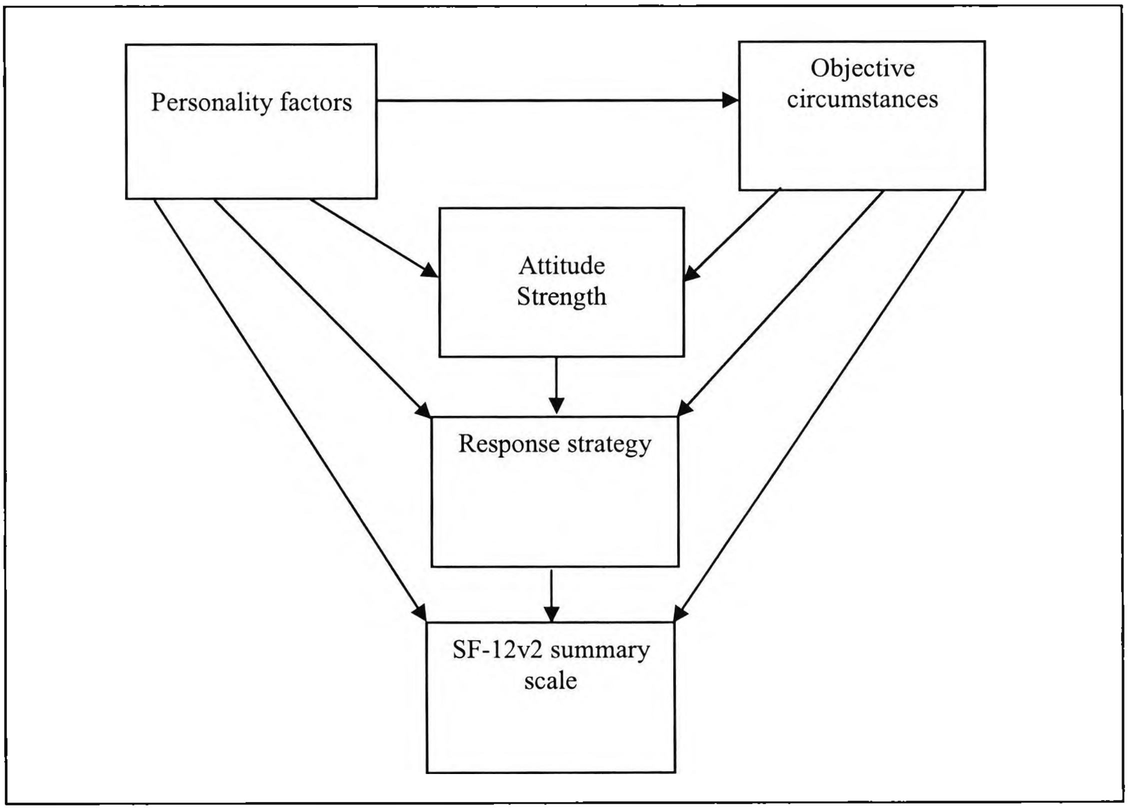
Multiple regression techniques are often considered for use in situations where the relationship between a number of variables and an outcome are to be investigated. However, while such an analysis can take into account the correlations that exist between predictor variables, it cannot be used to infer the likely causal relations between measures, both direct and indirect. In this instance, path analysis is more appropriate, as it allows variables to be placed in a likely causal sequence, which can be justified empirically and theoretically (Wright, 1934). Therefore, path modelling was adopted in order to test a *contextual model* for health status response, in which pathways are hypothesised to exist from contextual factors through to response process and response.

Figure 7.1: Path model of top-down and bottom-up influences on ratings of Subjective Well-Being (Brief et al, 1993)



The development of a simple path model was considered to be one of the aims of the thesis from the outset, linking as it does context, response process and response. The path model approach adopted was based on one developed by Brief et al (1993) to investigate the relative importance of top-down (personality) and bottom-up (experiential) influences on subjective well-being ratings (Diener, 1984) (Figure 7.1). The published model by Brief et al indicated that personality (negative affect, related to Neuroticism) and objective life circumstances (use of health services) were *both* associated with subjective well-being, although their influence was mediated by interpretation of life circumstances (measured in terms of subjective health perceptions). While information on a range of contextual variables was collected in the current project, and some proved to be significantly related to the outcome measures, it was felt that a simple, contextual model, developed from that of Brief et al, and including both top-down and bottom-up influences, should be tested, before refining the pathways in future research.

Figure 7.2: Proposed *contextual model* of top-down and bottom-up influences on the SF-12v2



With the data available in the current quantitative study on a range of contextual factors, it was possible to extend this work, investigating the role of Neuroticism and health service use on perceived physical and mental health, but including additional *mediating* measures, attitude strength and response strategy (Figure 7.2). Brief et al's model included general health perceptions in the role of interpretation of life events. However, the current analysis was able to operationalise interpretation in terms of actual judgement processes, using the summative scale based on number of reported uses of the general perceptions strategy, which was strongly inversely related to the use of specific experiences. Response strategy was hypothesised to mediate all other variables included in the model. Attitude strength was considered to be the product both of events and temperamental factors, and an influence on response processes (Markus, 1983; Petty and Krosnick, 1995), and therefore a measure of attitude strength was also included in the model. Finally, the earlier model did not include possible linkages between top-down and bottom-up pathways. However, since both are self-report, there is a possibility that Neuroticism could influence the reports of objective life circumstances (Larson, 1992). Therefore, this was addressed by including direct pathways from Neuroticism to all other variables.

To recapitulate the model pathways:

- Neuroticism, health service use and response strategy had direct paths to the SF-12v2 scale
- Attitude strength was a product both of Neuroticism and health service use and influenced response strategy
- Response strategy mediated all other variables
- There were direct pathways between Neuroticism and all other variables

Methods

The variables included in the model were the Big Five Neuroticism scale, attitude strength, health service use during the year, general perception strategy summated scale and PCS and MCS scales. All predictor variables were significantly associated with the SF-12v2 in one or both samples (See Chapter 5). Expanding on the approach adopted by Brief et al (1993), two groups of variables were summarised using principal component analysis (PCA), so that a single component was identified for health service use and attitude strength in each dataset:

Health service use (items on: number of times attended in the previous year: general practice, outpatient or casualty department, daycare services, or had an inpatient stay).

Attitude strength (items on: level of concern, change in concern, and amount of time thinking about health). The remaining item (experience of others) was excluded because it failed to load on the same attitude strength component as the others.

Component loadings differed somewhat between samples, indicating that patterns of response to health service and attitude strength items varied according to sample. In terms of the health service component, factor loadings were stronger for GP and outpatient attendance in the university sample (GP: 0.69; outpatient: 0.72; daycare: 0.58; inpatient: 0.65), whereas the reverse was found in the HIV sample (GP: 0.57; outpatient: 0.58; daycare: 0.72; inpatient: 0.74), reflecting the differences in health service contact reported in the two samples. For attitude strength, factor loadings were similar in magnitude between samples for health concerns and changes in concern, although thinking about health loaded slightly more strongly on the attitude strength component in the HIV sample (concern: 0.85; change in concern: 0.74; thinking: 0.76) than the university sample (concern: 0.85; change in concern: 0.75; thinking: 0.69), indicating that thoughts and concern were more closely related among HIV respondents. The proportion of variance explained differed according to the component, although not greatly between samples. More than half of the item variance could be explained by an attitude strength component (university sample: 58.9%; HIV sample: 61.4%), whereas just under half was explained by a health service use component (university sample: 43.9%; HIV sample: 43.0%). Principal component scores for health service use and attitude strength were derived from these analyses, and were used in subsequent path modelling.

Figure 7.3: University sample contextual model for the Physical Component Scale (PCS)

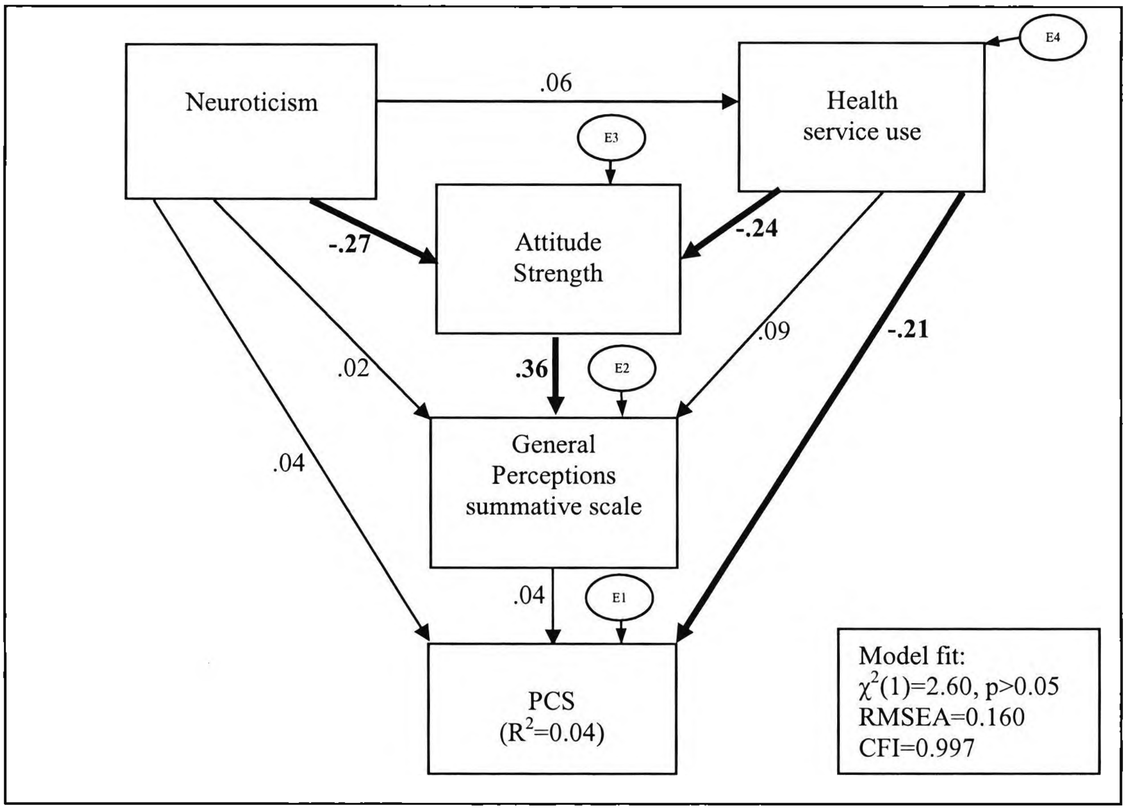


Figure 7.4: University sample contextual model for the Mental Component Scale (MCS)

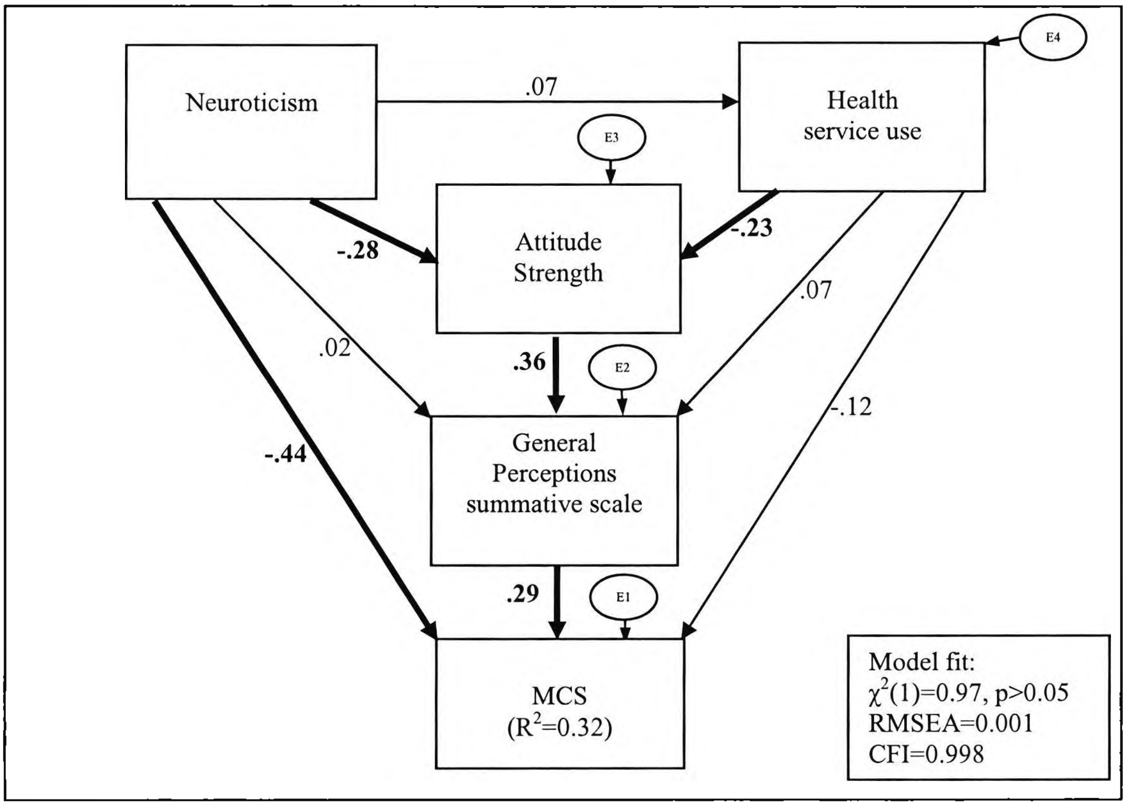


Figure 7.5: HIV sample contextual model for the Physical Component Scale (PCS)

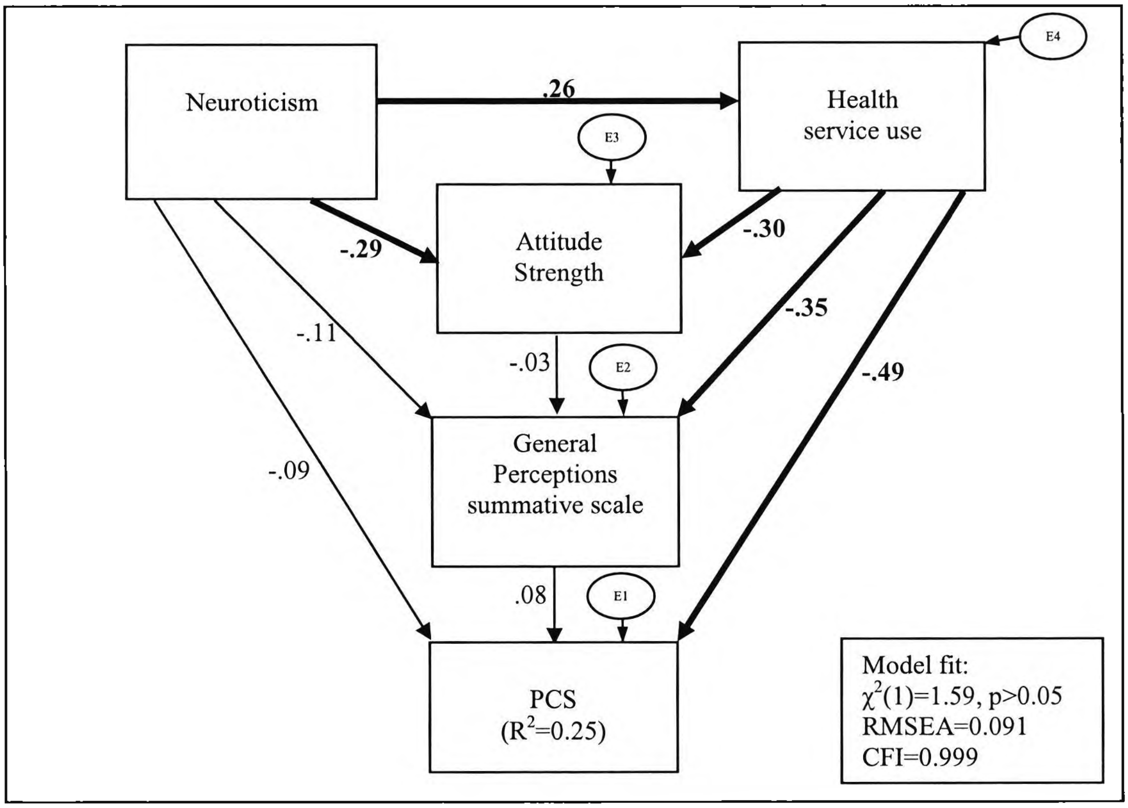
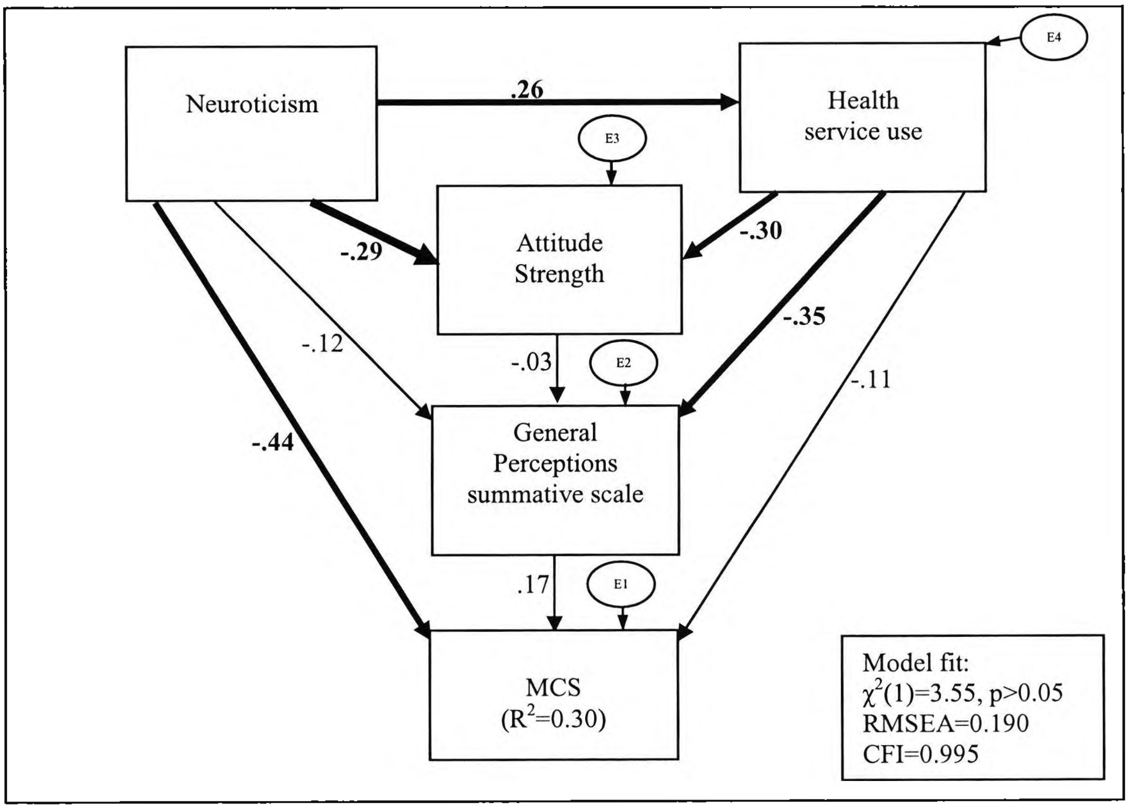


Figure 7.6: HIV sample contextual model for the Mental Component Scale (MCS)



Results and discussion

Four path model diagrams are shown in Figures 7.3 to 7.6, one for each sample with either the mental health score (MCS) as the outcome or physical health score (PCS) as the outcome. The models included measurement error terms (e) for all variables except Neuroticism, which precedes other variables in the temporal sequence included in the model.

A good model fit is indicated by a non-significant chi-square, an RMSEA of less than 0.05 and a CFI (Comparative Fit Index) over 0.95 (Byrne, 2001). In terms of the four models tested, the fit statistics varied, but provided some evidence that the model formulation performed reasonably well in both samples for the PCS (university sample, $\chi^2(1)=2.60$, $p>0.05$; RMSEA=0.160; CFI=0.997); HIV sample, $\chi^2(1)=1.59$, $p>0.05$; RMSEA=0.091; CFI=0.999) and also the MCS (university sample, $\chi^2(1)=0.97$, $p>0.05$; RMSEA=0.001; CFI=0.998); HIV sample, $\chi^2(1)=3.55$, $p>0.05$; RMSEA=0.190; CFI=0.995). However, more importantly, the model only explained a small proportion of the scoring variance for the SF-12v2 scales: PCS (university sample: 4%; HIV sample: 25%); MCS (university sample: 32%; HIV sample: 30%). The particularly low figure for the university PCS scale is likely to relate to the generally good physical health scores and limited use of health services in this sample.

Focusing on the contextual models themselves, significant pathways varied according to sample and outcome, although some general patterns were observed. Both personality and objective factors related to the outcomes, with Neuroticism directly linked to the mental health scale (MCS) and health service use to the physical health scale (PCS), indicating that the direct top-down or bottom-up influences included in the model varied according to outcome.

However, it was notable that in the HIV sample there appeared also to be a direct relationship between Neuroticism and the measure of objective circumstances used, the component derived from frequency of health service contact in the previous year, indicating that reported health service mediated Neuroticism to some extent. There are a number of possible explanations for this result. These include response bias related to personality factors (Larson, 1992), or the more frequent use of health services among respondents with a Neurotic personality trait. In addition, these findings could also relate to reactive personality changes resulting from chronic illness, with increased anxiety following health problems requiring health service care. Personality change has been linked to traumatic events in previous longitudinal research (Terracciano et al, 2006). Unfortunately, it was not possible to identify causation in this cross-sectional research.

Attitude strength mediated both personality and objective circumstances in the two samples: both higher levels of Neuroticism and a greater reported use of health services were linked to increased health concerns and more frequent thoughts about health. This pattern of results would be expected from self-schema and attitude strength theories, whereby the structure of attitudes may be modified by experiences but will also be influenced by other cognitive and affective components (Markus, 1983; Petty and Krosnick, 1995).

The relationship between the contextual variables, the general perception strategy scale and outcomes varied, indicating that a cumulative scale for the use of a general perception strategy when answering the SF-12v2 was not a universal measure of response strategy. In the university sample, attitude strength was strongly related to the general perception scale, whereby higher attitude strength scores (greater levels of health concerns and thinking about health) were related to lower strategy scores (fewer reported uses of the general perception strategy). However in the HIV sample, there was a significant pathway between objective circumstances and the general perception scale, indicating that greater reported use of health services was related to lower scores on the general strategy scale.

Finally, the association between the general perceptions scale and the outcome differed for physical and mental health. The general perception scale was related to MCS score in both samples, although significantly so only for the university respondents, with the greater use of general perceptions associated with better mental health. However, there was no relationship between general perceptions and physical health, measured using the PCS. These findings were replicated in analyses using a general perception scale based on ratings of either physical or mental health variables, indicating the results were robust.

Overall, these findings suggest that the contextual path model needs to be reconsidered, particularly in relation to physical health in a healthy sample. Another caveat relates to the operationalisation of variables, which may have influenced the results obtained: for example, the health services component was based on recalled attendances in a year rather than any objective marker, attitude strength comprised a summary of items measured, and response process was based on a count of the number of times a general strategy was reported, rather than a full exposition of the strategies used during SF-12v2 completion.

Nevertheless, the model presented indicates that health status scores are related to personality and health attitudes, in addition to objective health indicators. There was a complex interplay between these factors, and response processes differed for outcomes and between samples. In addition, use of the general perception strategy, which was earlier shown to relate to the SF-12v2 at scale and item level, was significantly associated with other variables included in the model, suggesting that response to the general strategy scale does relate to objective life circumstances and health concerns.

In conclusion, this model needs to be further developed in the future in order to gain a better understanding of the relationship between contextual influences, response processes and response. Most importantly, although there was evidence of some comparability between samples in the pathways to physical and mental health scale response, the contextual framework for response differed between the samples, and therefore models need to be developed separately for the two samples. This would involve moving from the simple, theoretical model presented to more finely-detailed models, grounded in the empirical results obtained from the bivariate contextual analyses. In doing this, significant pathways identified for each sample would be retained, with the inclusion of additional variables that related to the outcomes for university and HIV samples. For both samples, Extraversion was significantly associated with SF-12v2 scale scores, significantly so for physical health in the university sample and mental health for the HIV sample. As with Neuroticism, additional modelling could be carried out with the addition of this trait. In the HIV sample, there were a greater number of significant bivariate associations between contextual variables and mental and physical health, which could be investigated further, including the association between the sociodemographics variables age and education, and reported problem drinking and mental health. In the original model, concepts were included in relation to hypothesised pathways between context and final response. However, the operationalisation of these concepts could differ for each sample. For example, rather than include the use of health services, self-reported long standing limiting illness might prove to be a more sensitive measure of health experience in the university sample. In addition, pathways between variables other than the outcome could be further specified to broaden the model, with the inclusion of additional variables identified through the examination of the inter-correlations between the contextual and response variables. Clearly, model refinements based on empirical results would need a theoretical rationale. Nevertheless, they would be extensions to the original model and would be developed within the contextual model framework.

Summary

The final stage of the quantitative study involved using path modelling to integrate all aspects discussed so far, in order to investigate relationships between contextual variables, response process and SF-12v2 score for the two samples. Previous research had provided a model in which disposition and events were linked through interpretation to influence subjective perceptions (Brief et al, 1993). A contextual model was tested, based on a hypothesised relationship between personality, reported health service use, attitude strength, use of general perceptions (response strategy) and SF-12v2 scale score. Although this was a simple model, which did not explain a large amount of the scoring variance, it did identify that there were pathways between the contextual variables included and response, particularly Neuroticism and subjective mental health and use of health services and physical health, suggesting that pathways between contextual factors and score may be similar for both healthy and HIV samples, but differ according to the outcome measured. However, the relationships between these factors and mediating variables, including attitude strength and response process, were complex and not fully realised in the model tested.