DEVELOPMENT OF A DESCRIPTIVE SYSTEM FOR PATIENT EXPERIENCE

A thesis submitted for the degree of Doctor of Philosophy

by

Jeshika Singh

Department of Clinical Sciences, Brunel University London

2018

ABSTRACT (310 words)

Efficient allocation of public resources requires identification, measurement and quantification of costs and benefits of alternative programs. Patient reported outcomes (PROs) are routinely incorporated into economic evaluations of health technologies, but patient experience is often overlooked. This thesis aims to develop a descriptive system for patient experience that can be valued and used to inform economic evaluation.

The generation and selection of items is key in the development of any PRO measure. The thesis provides a contemporary overview of recommended methods and those actually used by instrument developers. Frequently a staged approach is used to establish dimensions first, using exploratory factor analysis, followed by item selection using item response theory (IRT), Rasch or structural equation modelling (SEM).

I demonstrate the use of different methods for item selection and its underlying mechanics, followed by comparison of the methods. An existing patient dataset, the Inpatient survey (2014) that collected information on nearly 70 aspects of healthcare delivery from NHS users was used.

Logistic regression analyses were applied with respondents' rating of overall patient experience specified as dependent variable. Advanced statistical analyses focussed mostly on patients who had an operation or procedure. Latent construct or dimensions were derived and measurement model was confirmed using confirmatory factor analysis. IRT and factor analysis were employed in each one-factor model for item selection.

Regression analyses identified many significant variables but most overlapped conceptually. An 11 and 8 factor model for patients with A&E and planned admissions respectively was determined. A generalised partial credit model and a factor analysis model identified different items to include in each dimension. Broadly the items identified by different methods related to respect, comfort and clear communication to patients.

This thesis presents descriptive systems for patient experience that is amenable to valuation. It also demonstrates that different patient experience instruments are generated based on patient population used and item selection technique adopted.

Table of Contents

List of tables

List of figures

Table of Appendices

List of Abbreviations

PRO: Patient Reported Outcome

PROM: Patient Reported Outcome Measure

QoL: Quality of Life

QALY: Quality Adjusted Life Year

HRQoL: Health Related Quality of life

NICE: National Institute for Health and Care Excellence

EFA: Exploratory Factor Analysis

CFA: Confirmatory Factor Analysis

IRT: Item Response Theory

SEM: Structural Equation Modelling

SF-20: 20-Item Short Form Health Survey

MOS: Medical Outcomes Study

SF-36: MOS 36-Item Short-Form Health Survey

AQLQ: Asthma Quality of Life Questionnaire

AQoL: Assessment of Quality of Life

SF-12: A 12-Item Short-Form Health Survey

RAQoL: Rheumatoid Arthritis Quality of Life

OHIP: Oral Health Impact Profile

WHOQOL-BREF: Short version of World Health Organization Quality of Life

ALSAQ-40: Amyotrophic Lateral Sclerosis Assessment Questionnaire

McSad: Direct utility measure for major, unipolar depression

PORPUS: Patient-Oriented Prostate Utility Scale

OAB-q: Overactive Bladder Symptom and Health-related Quality of Life Questionnaire

HIT-6: Six-Item Short-Form Survey for Measuring Headache Impact

QoLIAD: Quality of Life Index for Atopic Dermatitis

PIQoL-AD: Parents' Index of Quality of Life in Atopic Dermatitis

PSORIQoL: Psoriasis Index of Quality of Life

VisQol: Vision-related Quality of Life

PFDI-20: Pelvic Floor Distress Inventory

PFIQ-7: Pelvic Floor Impact Questionnaire

CPCHILD: Caregiver Priorities and Child Health Index of Life with Disabilities

EORTC QLQ-C15-PAL: EORTC QLQ-C30 'core questionnaire' for palliative care

CPQ11–14: Child Perceptions Questionnaire for 11-14-year-old children

QualiPause toolkit: Quality of life in climacteric and postmenopausal women

ICECAP-O: ICEpop CAPability measure for Older people

ICECAP-A: ICEpop CAPability measure for Adults

PROMIS: Patient-Reported Outcomes Measurement Information System

DUI: Diabetes Utility Index

L-QoL: QoL in Systemic Lupus Erythematosus

CHU-9D: Child Health Utility Index

CORE-6D: Short version of a measure of common mental health problems (CORE-OM)

PBM: Preference-based Measure

EORTC QLQ-C30: the European Organization for Research and Treatment (EORTC)

MF-SAF: Myelofibrosis Symptom Assessment Form

MobQues28: Mobility limitations in children with cerebral palsy

ThyPRO: Thyroid-Related Patient-Reported Outcome

DEMQOL: Questionnaires for the person with dementia

DEMQOL-Proxy: Questionnaires for dementia carer

NEWQOL-6D: Epilepsy-specific QALY measure

NEI VFQ-25: National Eye Institute Visual Functioning Questionnaire

STQOLI: Specific Thalassemia Quality of Life Instrument

CAT-QoL: Child Amblyopia Treatment Questionnaire

CP QOL-Teen: Cerebral Palsy Quality of Life Questionnaire-Teen

P-PBMSI: Preference-Based Multiple Sclerosis Index

IUI: Incontinence Utility Index

MSIS-8D: Multiple Sclerosis Impact Scale

Acknowledgement

I would like to thank my supervisors Dr Subhash Pokhrel and Professor Doug Coyle for their encouragement, guidance and support. I am sincerely grateful to you for helping me get here. Since my part time PhD has spanned over six years and many changes, I am only one supervisor short from having one for each year. I would not have embarked on a PhD without the steer from Professor Joanne Lord and Professor Martin Buxton. And I would not have been able to complete this thesis without continued support, faith and insights from Dr Louise Longworth.

I am grateful to Professor Stephen Hanney for his constant encouragement, many jokes and the steady supply of a chocolate-dipped Viennese finger every Tuesday and a Nutella doughnut every Friday! I would like to thank my wonderful friends for providing a solid support system and my colleagues (from HERG and PHMR) for their friendship. A special thank you to Frauke for the final sense check. Lastly I would like to tell my family (Safu, Santafu, Tatu, Mamu, Dada, Bhauju, Miko and Timko) that they are my strength, especially Dada, and I owe everything to my dear Tatu and Mamu.

Declaration of Authorship

I, Jeshika Singh, declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

[DEVELOPMENT OF A DESCRIPTIVE SYSTEM FOR PATIENT EXPERIENCE]

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Either none of this work has been published before submission, or parts of this work have been published as:

(2017) Singh, J. and Longworth, L. A Review of Methods for Item Selection in the Development of a Descriptive System. A report to EuroQol Group (project 2016300)

Singh J, Pokhrel S, Coyle D, Longworth L. Development Of A Descriptive System For Measuring Patient Experience. Value in Health.20(9):A516.

Singh J, Longworth L, Coyle D, Pokhrel S. A Review Of Methods For Item Selection In The Development Of A Descriptive System For Quality Of Life Instrument. Value in Health.20(9):A687

Signed: ……

Date: ……02/02/2018……

Chapter 1 Introduction

1.1 Resource allocation decisions

Given the demand for health and limited resources, there is an opportunity cost to every decision taken and an important challenge regarding the allocation of resources across competing interventions and technologies. The libertarian and egalitarian perspective form the fundamental ideologies about provision of health care. The crux of distinction between the two perspectives is the differing maximand in their social welfare functions. In a libertarian system health care is part of the reward system of society and access to care is determined by willingness and ability to pay. In an egalitarian system, the dominant ethic is equal opportunity of access for those in equal need. A collective decision has to be made about which of the two ideological positions to adopt to govern the provision of health care in a given political community. In the United Kingdom (UK), an egalitarian framework is the one predominantly used to make decisions about healthcare priorities, and cost containment and equality of access are strong tenets.

Priorities in an egalitarian system are determined by social judgements about need. Need is defined as the patient's relative ability to benefit in relation to opportunity cost (Williams, 1974). Individual needs are arbitrated by a third party who weighs different needs of people against one another, so that collective values are placed before the values of a particular interest group within it. Often representative samples of the population are asked to make marginal trade-offs between different healthcare benefits to make social value judgements.

One of the first institutions to adopt this framework in the UK is the National Institute for Health and Care Excellence (NICE), which was launched in 1999 and one of its core functions is to ensure the taxpayers' money is invested in the National Health Service (NHS) so that population health is maximised (Chalkidou, 2009). NICE assesses healthcare technologies in terms of clinical effectiveness and cost effectiveness.

NICE uses patient reported outcome measures (PROMs) to describe the health consequences of healthcare as perceived by the patient. One of the most commonly used PROMs in the UK is the EQ-5D, which is a generic measure of health related quality of life. The EQ-5D describes health status using five dimensions or domains: mobility, self-care, usual activity, pain & discomfort and anxiety & depression. Each dimension consists of one item or question, and five responses levels that range from no problem to severe or extreme problem. Combination of these levels and dimensions yields a number of health profiles or health states. Numeric valuation is applied to health states based on preferences for being in that state relative to perfect health (1) and dead (0) using different valuation techniques. These preferences weights or utility values are elicited from a representative sample of the general population in the UK. Utility score is combined with survival data to generate Quality Adjusted Life Years (QALYs). Expressing health outcomes in a single metric, i.e. QALY, enables comparisons to be made across different treatments and different health conditions. This facilitates consistency and predictability in decision-making.

Efficient allocation of public resources requires identification, measurement and quantification of costs and benefits of alternative programs. Initially PROMs were only applied to NICE appraisal of new health technologies but in 2009 NHS started collecting PROMs data routinely for four procedures: hip replacement, knee replacement, groin hernia and varicose veins for measurement and management purpose. This practice aimed to improve medical practice by making clinical activity more transparent as well as facilitating systematic appraisal of the success of the NHS in improving health. One can argue that use of patient reported outcomes (PROs) to compare patient experience related to non-health benefits will make prioritisation decisions about quality of care interventions more transparent and systematic. This thesis aims to develop a descriptive system for patient experience that can be used to aid decision-making.

An integral component of healthcare is quality of care or health delivery characteristics, which has a direct impact on patient 'experience' and patient 'satisfaction'. Patient 'experience' and 'satisfaction' are often used interchangeably but have different meanings (Beattie et al., 2015, Coulter et al., 2009, Sitzia and Wood, 1997b). Patient experience is related to events that occurred and the extent to which needs were met. Patient satisfaction is related to patient expectations, which is a complex concept with various determinants (for example, prior experience). It generates a 'discrepancy' model where satisfaction is 'relative' and determined by the perceived discrepancy between the actual experience and actual expectation (Sitzia and Wood, 1997a). In addition patient characteristics, such as age and educational attainment, and psychosocial determinants, for example gratitude bias, affect patient satisfaction (Sitzia and Wood, 1997a). There is a strong argument that patient experience has more validity than satisfaction data, since it is based on a patient's experience of

healthcare and not external factors (Beattie et al., 2015, Coulter et al., 2009, Sitzia and Wood, 1997b).

A series of policies have been set up to improve the quality of care in the NHS. A report published in 2016 by the King's fund state that these policies proceeded following well publicised lapses in care of patients by the NHS, concerns about performance gaps in the NHS amongst the voters and the recognition that NHS 'could do more to improve quality and patient safety' (Ham et al., 2016). In 2012, the Department of Health specified 'making sure that people have a positive experience of care in the NHS' as one of the five objectives included in the New Care Objectives (DH, 2012b). This includes re-stated commitments to improve the patient experience through, for example, reduced waiting times and eliminating the use of mixed-sex accommodation in hospitals. The NHS Outcomes, Performance and Productivity policy document recognises that patients may benefit from health care not only in the form of health-related quality of life (HRQL) but also in the form of 'the humanity of the care' they receive (OHE, 2008). It notes the importance of non-health characteristics of care such as: speed of access to advice or treatment; participation in decisions; respect accorded and dignity preserved; availability of comprehensible information about treatment including provision of support for self-care and attention to physical and environmental needs (OHE, 2008). In 2014, the government set out 'Hard truths: the journey to putting patients first' as a response to the report of the Francis Inquiry into Mid Staffordshire (DH, 2014a, DH, 2014b). There is currently a strong argument that to improve quality of care a fundamental shift is required, whereby performance is improved using reform from within the NHS rather than using external pressures (Alderwick et al., 2015). This entails reduction of waste and inefficiency.

Cost-effectiveness analyses focus on value-for-money. Identifying and measuring standardised PROs related to quality of care interventions will enable interventions to be assessed based on effectiveness. PROs based on health related QoL of the patient are routinely incorporated into economic evaluations of health technologies and promote transparent decision-making. However economic analysis of quality of care interventions is not conducted because identification and measurement of patient experience is not possible. Patient experience has been overlooked, firstly in assessment of health technologies because of focus being entirely on QoL outcome only and secondly in not applying cost effectiveness analysis in decisions regarding quality of care intervention.

1.2 Extra-welfarist approach

Systematic identification, measurement and valuation of alternative interventions, and the subsequent comparative analyses of these costs and benefits are called economic evaluations. There are two main perspectives to economic evaluation: the welfarist and extra-welfarist approach (Brouwer et al., 2008). The welfarist approach aims to maximise societal welfare and places considerable emphasis on the value individuals place on outcomes, because individuals are considered to be the best judges of their own welfare. It ascertains the total amount the individual would be willing to pay for the programme and directly compares with the costs in order to assess whether the program is worthwhile (Brouwer et al., 2008). This approach is consistent with economic theory and economic evaluation gives us the result we would have obtained from the market had there been one operating.

Another approach, referred to as the extra-welfarist perspective adopts a narrower health sector perspective, which may be close to that adopted by health care decision makers. The aim of the extra-welfarist approach is to maximise health effects in a resource-constrained health system, and may reflect both individual and societal preferences (Brouwer et al., 2008). It shifts the evaluative space from maximisation of utility to the maximisation of health. For example, NICE uses an extra-welfarist approach in economic evaluation of health technologies. It assumes that the role of health services is to increase the overall health of the society. It considers health care resource only and compares the resources consumed with the health improvement obtained in terms of natural units or health effects, which are valued using health state preference scores from the general public. Hence cost effectiveness analysis conducted for the NHS, to aid resource allocation, is based on the ratio of incremental cost per QALY that capture health gain.

Within the extra-welfarist paradigm, a single standardised descriptive system is used to assess the impact of different interventions on health related quality of life (HRQoL). In the UK, NICE recommends using EQ-5D instrument to measure HRQoL and expressing health gains in terms of QALYs in health technology assessments (HTAs) (NICE, 2013b). Use of QALYs enables comparison of healthcare technologies across different disease conditions and patient groups. The supporters of welfarism use willingness to pay expressed as monetary value as a common numeraire to achieve comparability.

A large number of studies have assessed characteristics of healthcare delivery or the 'process' aspect of health that are non-clinical (Mooney, 1998). These studies have using stated preference techniques such as

contingent valuation to estimate willingness to pay (WTP) for the overall healthcare service; and discrete choice experiment (DCE), frequently containing a cost component to estimate willingness to pay (WTP) value (Clark et al., 2014, de Bekker-Grob et al., 2012, Diener et al., 1998, Smith and Sach, 2010). This WTP estimate reflects the values placed by the respondents (patients and members of general public) on the attributes, attribute levels and overall service described. The WTP value can be incorporated into cost benefit analysis to produce the net benefit of the intervention.

Other than expressing the value of delivery characteristics in monetary value, many studies have estimated 'process utility' or the utility from process of care attributes. A systematic review conducted to examine empirical estimate of process utility identified fifteen studies between 1996 and 2012 (Brennan and Dixon, 2013). The included studies explored care characteristics from three different settings: treatment, screening and preventative care. The hypothetical health states used in the studies were designed to describe the interventions being examined and comparison between estimates was very difficult. The review suggested further research in many areas, including a comparative study of alternative methods, the need for testing the validity of results through psychometric approaches and comparative studies with other patient reported measures.

Perhaps 'process' was too general a term? Process utility encapsulates values arising from a wide range of healthcare characteristics and it is possible that narrowing it down to one aspect of care would help in finding utility estimates that are comparable. I was involved in a systematic review of the literature carried out in 2014 to estimate the value associated

specifically with convenience in health care delivery, independent of health outcome (Higgins et al., 2014). The study found twenty-seven studies reporting some evidence of convenience-related process utility, in the form of either a positive utility or a positive WTP value. There were broadly two categories of convenience valued in the studies identified from the literature, those relating to the administration of an intervention, such as dosing or mode of administration and those examining access to an intervention, such as distance to travel. The attributes used, attribute levels and the wording varied from study to study, even when the concepts being described were identical (Higgins et al., 2014). Higgins et al also observed that the methods used to estimate WTP value differed across studies. Of the WTP studies identified, only one study on convenience employed open-ended WTP and the rest used a DCE format that required respondents to select a value from a set of predefined levels. It is true that WTP estimates enable calculation of net benefit arising from different aspect of healthcare delivery. And theoretically the elicitation of value in monetary form allows comparison of different attributes and the overall state or scenario across interventions and sectors. However, in reality comparing estimates from one study to another is very challenging as the study methods differ substantially.

Another concern with use of DCE and WTP values is that they are estimated using bespoke description with limited external validity, and attributes or vignette to be valued are constructed on a case-by-case basis. One of the advantages of using a bespoke vignette for description of healthcare is the richness in data and specificity. However this also limits its use. Psychometric criteria such as validity and reliability are important for any measurement and assessing a bespoke vignette using these criteria would be challenging (Brazier et al., 2007) (pp 68). For instance a vignette constructed for a contingent valuation study would only describe one state

and not a distribution of health states that the patient may go through. The extent to which a set of vignettes accurately represents the distribution of possible combinations or has 'construct validity' is a quantitative assessment that cannot be examined in a bespoke descriptive system. This would also result in difficulties comparing vignettes from different studies. It should be noted that while a majority of DCE studies use a bespoke descriptive system, there are exceptions. The EQ-5D-5L, which is a generic measure of health status and able to classify 3125 unique health states, included DCE in the valuation protocol as a preference elicitation technique to value health states (Krabbe et al., 2014, Oppe et al., 2014); and it was employed in the study that estimated EQ-5D-5L value set for England (Devlin et al., 2017). The study derived utility values and not willingness to pay estimates. More importantly it used a standard descriptive system.

There are concerns with the use of a WTP approach on normative grounds as well. It requires individual level assessment of benefits and values to be expressed in monetary terms that may be subjective to income or affordability rather than true preference. Finally it is time consuming to conduct.

An extra-welfarist framework has not been applied to quality of care and my thesis explores if it is possible to do so. In order to enable a comparison of healthcare interventions based on patient experience, a standardised patient experience measure is necessary.

1.3 Standardised descriptive system

An instrument using a structured format with multilevel items and characteristics for measurement that allow consistent administration to

groups of patients and across different time points is defined as a standardised descriptive system. A standardised instrument is produced with the aim of achieving order in a given context and providing a common basis to compare status or experience. An alternative approach is to create bespoke descriptions of patient experience.

Bespoke vignettes were used more commonly in the past to describe health but over time there has been an increase in used of standardised measures of health such as the EQ-5D and SF-36. This perhaps coincides with valuations of health states or measurement of utility to inform economic evaluation of health technologies.

Using a standardised descriptive system has two key advantages. Firstly it allows assessment of validity and reliability, for example construct validity is a key consideration during construction of a standardised descriptive system using the classical or modern test theory approach that is described in more detail in later chapters. Secondly a standardised instrument is able to capture responsiveness or measure 'significant' changes over time or across intervention. For example, comparison of health states is crucial in assessing the impact of a new technology in a patient before and after the intervention and/or across intervention and control arms. A similar approach will benefit the assessment of quality of care interventions.

A standardised descriptive system can be generic or condition specific. When an instrument is generic or not disease-specific, it is able to measure HRQoL in a more holistic manner and incorporate the side effects or complications of treatment, which may be unrelated to the condition itself. Another important advantage of a generic measure is that it allows

comparison of technologies or interventions across a wide range of diseases. A potential disadvantage of a generic measure is that they are less responsive to health changes than condition-specific measures. The most frequently used generic preference based measures EQ-5D (3 and 5 levels), the Health Utilities Index version 3 (HUI3) and the Short Form 6 dimensions (SF-6D). Similarly there are a number of condition-specific measures that are preference based such as cancer-specific preference based measure (EORTC-8D), asthma quality of life questionnaire (AQL-5D) and QALY measure epilepsy (NEWQOL-6D).

Condition specific measures are not applicable for all patients but may have an important role for economic evaluation, for which generic measures are inappropriate, insensitive or unresponsive (Brazier et al., 2012). However naming the condition, the exclusion of side effects and comorbidities and focusing effects limit their use in economic evaluation. Whether a reduction in comparability should be accepted depends on the extent of any gain in validity and responsiveness. This will depend on the condition and measure in question. No distinction was made between generic and condition specific measures while examining methods used to develop HRQoL instruments. However an important consideration throughout this thesis was that the instrument should be amenable to valuation in the future.

1.4 Patient Reported Experience Measures (PREMs)

A number of surveys and indicators are available to obtain insights into quality of care from patients and they are also known as patient reported experience measures (PREMs). The evidence scan report by the Health Foundation provides a review of the surveys used, including administration

methods, timeframes and question type (Health, 2013). For the purpose of this study, I have focussed on instruments that are standardised and repeatedly administered. Beattie et al. conducted a systematic review to identify and critique measures of patient experience in hospitals (Beattie et al., 2015). They conducted the search in databases such as MEDLINE, CINAHL, PsychINFO, Web of Knowledge and grey literature in November 2013. A total of 26 papers examining 11 international instruments were included.

Three PREMs from the UK were identified: Picker Patient Experience Questionnaire (PPE-15), National Health Survey (NHS) Inpatient Survey and Scottish Inpatient Patient Experience Survey (Beattie et al., 2015). The NHS Inpatient Survey has the most extensive history amongst the three, with original development work tracing back to 1991 when the original Picker Adult Inpatient survey was reported (Picker, 2012, Boyd, 2007, DeCourcy et al., 2012, Sizmur and Redding, 2012). The NHS Inpatient Survey is formed of 70 items or questions and it has been administered annually to NHS users since 2002 in England and Wales (CQC, 2018). Scottish Inpatient Patient Experience Survey consists of 30 items, it was first administered in 2010 and it is currently run every two years.

The PPE-15 is a 15-item patient experience questionnaire designed for use in inpatient care settings (Jenkinson et al., 2002a, Jenkinson et al., 2003, Reeves et al., 2002a). It is a short form version of the Picker Adult In-Patient Questionnaire. These items had good face validity and when summed to an index they showed a high degree of construct validity and internal consistency (Jenkinson et al., 2002a). These questions were described as core questions and authors assured the score derived from it

were easy to interpret and actionable. Further details about the methods used in this study are discussed in chapter five.

1.5 Development of a descriptive system for patient experience

The data from the current patient experience surveys allow comparison of quality of services across NHS trusts in the UK and over time, and help to target improvement but these instruments are lengthy and they are not amenable to valuation. All aspects of quality of care are important but it is possible that there are preferences across its dimensions and levels. And given limited resources in the NHS it will be useful to understand the underlying trade-offs between different aspects of care. Valuing patient experience will allow better targeting of services based on what is most highly valued and assessment of cost-effectiveness of competing strategies to improve quality of care in a systematic manner. While there are several measures of patient experience available in the literature, it is not possible to elicit preferences using them as they are lengthy and its measurement properties are not established.

The aim of my thesis was to develop a descriptive system for patient experience that is amenable to valuation and can be used to inform economic evaluation of healthcare interventions. This thesis is built on two core concepts derived from health economics and psychometrics, which is a branch of psychology and it is concerned with the theory and technique for measurement of psychological variables, which I describe in detail in chapter three. I used an existing secondary dataset, which assesses patient experience across nearly seventy variables to derive a brief measure that is amenable to valuation and focussed on item selection using different

statistical techniques based on measurement theory. It should be noted from the onset that actual utility measurement is beyond the scope of my study.

The generation and selection of items is essential in the development of any PRO measure. I have developed a descriptive system using an existing instrument – the NHS Inpatient Survey (CQC, 2018) and item generation is not required in this study. The empirical studies in this research are concerned with item selection and reduction, using a two stage based on psychometric assessment and a direct approach using regression analysis. A reduced or short form of the inpatient survey will enable data collection on patient experience to be streamlined to a core set of dimensions, which are distinct but related and is able to summarise patient experience similarly to the existing measure.

Two key considerations in the thesis were that the methods used for construction of the descriptive system were investigated in detail and carefully selected, and that the final instrument is amenable to valuation. Finally this application of an extra-welfarist approach to quality of care is exploratory and requires a stage-wise approach over the years to systematically unravel and understand. My research is a step in that direction.

Chapter two describes the aims and conceptual framework of the thesis. The objectives, research questions and methods to support the aim are summarised in the conceptual framework. In chapter three and four, the key concepts and methods used to develop an instrument are presented based on a focussed review of guidance and standards, and a systematic

review of studies reporting development of a HRQoL respectively. The two chapters provide a contemporary overview of methods recommended and used by instrument developers for both generic and condition specific measures.

I demonstrate the use of different methods for item selection and its underlying mechanics, followed by comparison of the methods. Item selection can be conducted directly, where items are selected from the items generated (or item bank) using regression analysis or qualitative studies. But more often a staged approach is used to establish dimensions first, using exploratory factor analysis, followed by item selection using item response theory (IRT), Rasch or structural equation modelling (SEM). Five methods were identified from the review and employed in this thesis to develop a measure of patient experience using the Inpatient survey. They are regression analysis, exploratory factor analysis (EFA), confirmatory factor analysis (CFA), structural equation modelling (SEM) and item response theory (IRT).

The National Patient Survey Programme (NPSA) was created to monitor patient experience across inpatient, outpatient, A&E, community mental health and maternity services in England but in this study I focussed only on inpatient or hospital stay. In the NHS Inpatient Survey 2014 dataset, more than 64,000 respondents assessed nearly seventy aspects of inpatient stay. I demonstrate use of three statistical approaches to identify dimensions and items to describe patient experience in those who had an operation or procedure during their Inpatient stay. Firstly regression analysis for direct item selection, secondly exploratory factor analysis and IRT and finally EFA followed by CFA. SEM was applied to explore ordering of items based on salience.

Statistical models were applied in the inpatient dataset to identify dimensions and items for patients who had an operation or procedure. In the first two approaches, dimensions based on latent construct were derived using EFA. The dimensions were further assessed and confirmed using confirmatory factor analysis (CFA). Item selection for each one-dimensional model was conducted using structural item response theory (IRT) and underlying variable approach (factor analysis). Finally SEM was applied in the multidimensional model to determine number of dimensions to include in the final model. For comparison logistic regression analyses were applied with respondents' rating of overall patient experience specified as dependent variable. In addition to application, the strength and limitations of these techniques and the underlying mechanics in each method was discussed to gain an understanding of the methods and sequential ordering of the analyses.

Chapter 2 Aim and Outline of Thesis

Chapter 1 highlighted that while PROs are now routinely incorporated into economic evaluations of health technologies; little attention has been given to patient experience to inform prioritisation of quality of care interventions. Several measures of patient experience are currently available but they are lengthy and it is not possible to combine its dimensions levels to generate plausible profiles for valuation. Also valuation of process characteristics, which is closely related to quality of care, using DCE and contingent valuation is widely available in the literature. But the interventions valued this way are difficult to compare across studies, as they tend to be case specific and methods differ substantially across studies although they are estimated in monetary value.

The approach adopted by many decision makers, including NICE, to ensure comparability across intervention (and to conduct economic evaluations) has been to determine the effect of each health care programme on the health state of each individual affected by the programme. And then to generate a social preference function defined over the relevant health states as the common unit of measure (Torrance 1976, Rosser and Watts 1978). The overarching aim of this thesis is to develop a brief instrument to measure impact on patient experience of health care intervention with the view of providing a commensurable unit across different types quality of care intervention.

This study will be a novel attempt to apply an extra-welfarist framework to explore patient experience, whereby a standardised descriptive system with same items and levels will be administered to patients to measure the

impact of quality of care intervention. It will be narrower in terms of focus when compared to the welfarist (WTP) approach, but it will enable measurement of patient experience as an individual (or a group) over time and across interventions. The study will focus on inpatient stay and it will be applicable to interventions improving patient experience in hospitals. Also consideration will be made to develop a brief measure that is amenable to valuation.

A value set or social preference function can then be elicited for patient experience profiles, based on combinations of items and item levels in the measure. A group of subjects, such as patients or the general population, are used to elicit preferences and the aggregate score across the subjects determine the overall social preference function. Besides valuation, it is still desirable for an instrument to be brief because it minimises measurement burden on the respondent, making it more acceptable and feasible. It also benefits the provider, as they are cheaper to administer, easier to compute and analyse. Note than valuation is beyond the scope of this study. Further deliberation is required regarding methods of valuation such as description of anchors, the time horizon for the healthcare states, the valuation procedure and the population group to obtain values from. These will not be addressed in this thesis in detail.

2.1 Aim and objectives

This thesis aims to develop a descriptive system for patient experience that is amenable to valuation. An existing patient dataset, the Inpatient survey (2014), that collected information on nearly 70 aspects of healthcare delivery from NHS users was used in this thesis to develop a brief measure of patient experience for inpatient setting.

An objective within the study is to provide a contemporary overview of recommended methods and those actually used by instrument developers. A staged approach is generally adopted during the development of an instrument whereby dimensions are established first, followed by item selection for each dimension using IRT methods (including Rasch), factor analytic techniques and SEM. Use of different item selection methods may generate very different descriptive systems. The second objective of this study is to demonstrate application of different methods for item selection and describe its underlying mechanics, followed by comparison of the methods and results.

2.2 Conceptual framework and thesis outline

A conceptual framework illustrating the research conducted in the thesis is provided in Figure 1, this includes the primary research questions driving each study. A summary of each chapter is provided below.

Chapters 3 and 4 provide a review of the literature and highlight key concepts that shaped this thesis. Chapter 3 introduces the readers to the keys concepts in patient reported outcome measures, methods used to develop a measure and evaluation of psychometric properties, including assessment of the measurement model to establish construct validity based on guidance documents.

Chapter 4 presents a systematic review of methods used by instrument developers to develop a descriptive system for measuring health related quality of life and patient experience. The chapter focuses on item

selection. The methods to be employed in the thesis are also discussed in this chapter.

Chapters 5, 6 and 7 analyse the responses of the NHS Inpatient Survey published in 2014 using methods that have been identified in the review chapters. Chapter 5 details this dataset and presents the multivariate regression analyses carried out. Logistic regression analyses were fitted with respondents' rating of overall patient experience specified as the dependent variable.

Advanced statistical analyses were carried out on patients who had an operation or procedure during their inpatient stay in chapter 6. Latent construct or dimensions were derived using EFA, which automatically ensures unidimensionality. Confirmatory framework was applied to the measurement model to revise factors and factor items till a good fit was achieved.

In chapter 7, item selection for each dimension was carried out using IRT and underlying variable approach, which is an extension of factor analysis. Use of SEM in multidimensional model was also explored. The descriptive systems produced to measure patient experience are presented in this chapter.

Chapter 8 provides an overview of the thesis, discusses limitations, methodological contribution and conceptual contributions, highlight areas of further research and policy implications followed by the conclusion.

Figure 1: Conceptual framework

<table>
<tr><td colspan="4">Title: Development of a Descriptive System for Patient Experience</td></tr>
<tr><td>Background</td><td colspan="2">Chapter 3: Key concepts and principles of measurement

<u>Research Questions</u>
What are PROMs?

How are they developed?

What are the measurement properties essential to a measure?

How to assess adequacy of a measurement model?</td><td>Chapter 4: Systematic review of methods used for item selection

<u>Research Questions</u>
How was item generation carried out?

How was item selection carried out?

Why one statistical method was chosen over another for item selection?</td></tr>
<tr><td>Empirical analysis (using Inpatient Survey 2014)</td><td><u>Chapter 5</u>: to identify items of Inpatient survey most highly associated with overall patient experience

Method used: regression analysis

<u>Research Questions</u>

Which items within inpatient survey explains overall patient experience rating?

Are these items consistent across subgroups?

Can these items be used in a measure?</td><td><u>Chapter 6</u>: to estimate and assess dimensions of patient experience

Methods used: EFA and CFA model

<u>Research Questions</u>

What are the dimensions of patient experience?

Are the dimensions estimated unidimensional?

How are the items related to each dimension?</td><td><u>Chapter 7</u>: item selection using different methods

Method used: IRT, FA and SEM models

<u>Research Questions</u>

If and why item selection using different methods yield different measures?

Can the new measures be used for valuation?</td></tr>
<tr><td>Discussion</td><td colspan="3"><u>Chapter 8: discussions and conclusion</u></td></tr>
</table>

Chapter 3 Key concepts and approach to developing a PROM

This chapter provides an introduction to what PROMs are, how they are developed and the measurement properties essential to a measure, including assessment of measurement model for item selection, which is the focus of my study. It is based on a focussed review of key standards and guidance documents for the development of a PROM.

3.1 Patient Reported Outcome Measures (PROMs)

Hundreds of standardised outcome measures have been developed to measure health-related quality of life (QoL) and health status, with a range of approaches used in the development. These instruments can be used for different purposes: to measure QoL or health status of an individual at a point in time, to discriminate between individuals or groups, to evaluate change over time among individuals or groups, to predict future status or a combination of above (Kirshner and Guyatt, 1985, FDA, 2009). The choices made at each stage of constructing a QoL measure will differ depending on the purpose of the instrument (Kirshner and Guyatt, 1985).

The importance of measuring QoL from the patient perspective to inform patient management and policy decisions is well accepted by clinicians and policymakers, and PROMs are widely used to inform decision making (Guyatt et al., 1993). In health care, it is common practice to ask patients to describe their health by indicating the level most applicable to them on each item of the measure; for example a healthy person is able to report no problem in all five dimensions (items) of the EQ-5D. It is administered to individuals across known groups (for example, control and intervention)

and/or over time (before and after intervention), and it provides patient reported assessment of effectiveness and efficiency.

Ideally an instrument should be brief and include items covering all relevant issues that are of interest to the study (Fayers and Machin, 2013). An instrument that only captures the key concepts is easier and quicker to complete and minimises measurement burden on the respondent, making it more acceptable and feasible (Boyle and Torrance, 1984, Kirshner and Guyatt, 1985). It also benefits the provider as they are cheaper to administer than lengthier versions, easier to compute and analyse (Boyle and Torrance, 1984, Kirshner and Guyatt, 1985). Also if one were to elicit preference values, there is a natural limit in terms of the number of items that a respondent can consider during a valuation task (Brazier et al., 2007). It is important that items to include in an instrument are selected with great care.

3.2 Development of PROM

PROMs are based on hypothetical concepts, constructs or latent variables, which represent the quality of life issues the developers intend to capture in their measure. This could be a single broad concept or concepts designed to assess multiple domains within a broad concept. Often these are unobservable and measured through a set of items or questions, and this forms the 'descriptive system' of the instrument. Development of a descriptive system is based on the validity and reliability of the items in measuring what is intended to be measured.

Some measures may have one item for each construct or dimension like the EQ-5D while others have multiple items per dimension such as the

AQoL-4D, which has three items per dimension. In addition, the level of measurement for each item differs across different outcome measures, for example it could be based on severity or frequency. Finally the scoring of the measure may be ordinal, interval or ratio scales and the procedure for deriving scale score may differ. For example some measures use raw scores, while others are transformed using weighting or standardisation. Additionally for measures used in economic evaluations, the preferred scoring system is based on preference-weights or utilities.

The development of a descriptive system consists of defining what is being measured, generating a pool of potential items and selecting items from the pool for the final questionnaire before use for measurement purpose. Once the concepts being measured have been determined and the completeness of the concepts contained in the items have been confirmed, the development of a descriptive system is largely concerned with item selection and testing. Item selection is based on the review of validity, reliability and ability to detect change.

Many instrument developers identify or develop a “conceptual framework” which refers to the description or diagram of the relationships between the items in a PRO instrument and the concepts measured, to inform the development of a descriptive system (Aaronson et al., 2002, FDA, 2009). While this can be based on theory, increasingly instrument developers use a variety of measurement models to operationalise the conceptual framework based on observed responses. The measurement models determine how items are associated with each dimension and how dimensions are associated with each other. This is tested using classical and modern test approaches.

The final framework is based on assessment of psychometric properties of the measure. Several guidelines and standards have been published to bring rigor and consistency in instrument development. Most of them identify essential measurement properties required for validation and the development process.

3.3 Focussed review

The key articles and textbooks used in this chapter were identified in the systematic search detailed in chapter 4. The focus here is on presenting the key concepts considered and approaches adopted when developing an instrument, whilst chapter 4 describes the methods reported by instrument developers when developing a PROM.

A systematic search was carried out to identify methods used to develop a PROM by instrument developers using the search strategy developed by consensus based standards for the selection of health measurement instruments (COSMIN) group. The search conducted is detailed in Appendix 1. A total of 553 records were identified from the search and an additional 51 were identified from other sources. The eleven guidelines, textbook chapters and quality standards which provided guidance to instrument developers are included in the focussed review. The review helped me gain an understanding of methods advocated and measurement properties examined during the development of a PROM.

The eleven key texts reviewed are listed in Table 1. I describe the development process and the measurement properties essential to a PRO measure in this chapter. All instruments were included, generic and

condition-specific, and irrespective of whether they were preference based or not. A narrative summary of development phases and psychometric properties is described below based on majority and/or consensus findings.

Table 1: Key studies identified in focussed review of guidance and standards

Authors	Year	Title
Food and Drug Administration (FDA)**(FDA, 2009)**	2009	Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labelling Claims
European Medicines Agency (EMA)**(EMEA, 2005)**	2005	Reflection paper on the regulatory guidance for the use of health related quality of life measures in the evaluation of medicinal products
Mokkink et al.**(Mokkink et al., 2010)**	2010	The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments
Aaronson et al. *on behalf of Scientific Advisory Committee **(SAC)*** **(Aaronson et al., 2002)**	2002	Assessing health status and quality-of-life instruments: attributes and review criteria
Terwee et al. **(Terwee et al., 2007)**	2007	Quality criteria were proposed for measurement properties of health status questionnaires
Johnson et al. *on behalf of EORTC Quality of Life Group* **(Johnson et al., 2011)**	2011	Guidelines for Developing Questionnaire Modules
Streiner and Norman **(Streiner and Norman, 2008)**	2008	Health measurement scales: a practical guide to their development and use
Patient Reported Outcome Measurement Information System (PROMIS) **(PROMIS, 2013)**	2013	Instrument Development and Psychometric Evaluation Scientific Standards
Brazier et al. **(Brazier et al., 2007)**	2007	Measuring and Valuing Health Benefits for Economic Evaluation
Fayers and Machin **(Fayers and Machin, 2013)**	2013	Quality of Life: The Assessment, Analysis and Interpretation of Patient-Reported Outcomes
Krabbe **(Krabbe, 2016)**	2016	The Measurement of Health and Health Status: Concepts, Methods and Applications from a Multidisciplinary Perspective

3.4 Phases of instrument development

The instrument development process begins with the developers establishing the objective of the measure and the target population, including scope, spectrum and contents of the questionnaire (Brazier et al., 2007, Fayers and Machin, 2013). The development process from here forward can be divided into sequential phases (Fayers and Machin, 2013, Johnson et al., 2011). A summary of the main activities/objectives at each phase and the measurement properties to examine in each phase of developing a descriptive system is presented in Table 2.

Three things to note before proceeding to a narrative of these phases are as following. First that item selection refers to item reduction and takes place only after the relevance of the item has been established and it takes place after deletion of items that appear unimportant during item generation phase. Secondly one should not delete items solely on the basis of very strong or very weak correlation (Fayers and Machin, 2013). At all stages, face validity and clinical sensibility should be considered. Finally taking note of whether the items included in the instrument are causal or indicator in nature is important as the descriptive system containing causal or indicator items have separate considerations in terms of construction. Applying psychometric criteria to causal variables may lead to instruments that are either suboptimal or invalid (Fayers and Hand, 2002). Also when causal variables are involved, simple scoring approaches do not work as there is no common latent factor and the causal variables have independent influence on QoL.

Table 2: Phases in Developing a Descriptive System

Development Phase	Activities	Measurement property assessed
Item generation	• Determine the objective of the measure and the target population • Use qualitative methods (literature review, interviews and/or focus groups) with patients having relevant condition and relevant healthcare professionals to generate an exhaustive list of all quality of life issues that are relevant to the domains of interest • Convert list of items into questions that are brief, clearly worded, easily understood, unambiguous and easy to respond to • Attach the time period to which the question refers to	Content validity
Pre-testing questionnaire	• Administer the questionnaire to obtain indication of level for each item, together with rating of relevance and importance. • Conduct structured interview with each patient after completion of the questionnaire to determine completeness and acceptability of the items included. • Use findings about relevance, importance, wording or translation of the item to make amendments. • Conduct preliminary testing of the latent relationship between items and dimensions, if possible	Content validity; Construct validity; Internal consistency
Field testing questionnaire	• Administer the questionnaire to subjects representative of target population, and establish acceptability by asking debrief questions after completion of the questionnaire. • Examine latent relationship between items and dimensions to determine item selection. • Determine and confirm the acceptability, validity, sensitivity, responsiveness, reliability and general applicability of the instrument to cultural groups or other subgroups	Construct validity; Internal consistency; Test-retest; Responsiveness

3.4.1 Phase 1: Item generation

The first phase of developing a PRO measure is to generate an exhaustive list of all quality of life issues that are relevant to the domains of interest by using the literature and qualitative studies with experts and patients. Literature searches of relevant journals and bibliographic databases are carried out to identify relevant concepts and subsequently reviewed by a number of healthcare professionals (with expertise and experience in the area of interest) and patients using interviews and focus groups. The EORTC (Johnson et al., 2011) guideline on developing questionnaire suggests interviewing three to five health care providers in the early stages followed by samples of 5-10 patients from each different treatment group or disease stage. Content validity is examined in this stage by identifying candidate items for deletion and relevant items not included in the list based on qualitative research methods. After ensuring understanding and completeness of the concepts to include in the questionnaire in the first phase, the issues are converted into items or questions. The questions should be brief, clearly worded, easily understood, unambiguous and easy to respond to. In most PRO instruments individual questions are qualitative and elicit responses in binary or labelled category format.

3.4.2 Phase 2: Pre-testing questionnaire

The aim of pre-testing is to identify missing and redundant issues, improve wording and testing the hypothesised scale structure (if large sample is available). The questionnaire is administered to obtain response score for each item, together with rating of relevance and importance. For example developers may ask target population or relevant heath care professionals to indicate the importance of that item as experienced, in terms of quality of life. Structured interviews are conducted with each patient after completion of the questionnaire to get information on completeness and

acceptability of the items included. Findings about relevance, importance, wording or translation of the item are used to make amendments to the items.

While the focus in this stage is to identify missing and redundant items and improve wording, some developers may also test the hypothesised scale structure by asking respondents to complete the questionnaire and examining the responses using statistical techniques. If an adequate sample size is recruited, the hypothesised scale structure is examined using classical test theory statistics such as inter-item correlations, item-scale correlations and internal consistency reliability (PROMIS, 2013). The pre-testing usually involves between 10 and 30 patients selected presenting the range of patients in the target population. Items that performed poorly may be noted in this stage but usually items selection decisions are carried out based on assessment in a larger sample in the subsequent stage.

3.4.3 Phase 3: Field-testing questionnaire

The objective of field-testing is to determine and confirm the acceptability, validity, sensitivity, responsiveness, reliability and general applicability of the instrument in subgroups. This phase requires respondents to complete the questionnaire and responses are examined iteratively to examine various psychometric properties. Note that field-testing of questionnaire is an evolving process and may continue being conducted several years after the instrument has been developed. For example, the appropriateness of a generic instrument has to be established for different conditions and countries (e.g. language versions).

Field-testing of the instrument is carried out amongst a large sample of patients representing heterogeneous groups and covering full range of the target population. It is only in this stage that redundant or inappropriate items are identified based on assessment of psychometric properties. Psychometrics is concerned with assessing if the instrument is a reliable and valid form of measurement and is described in detail in next section. Nevertheless before excluding any item from the measure, face validity and clinical sensibility should be considered.

3.5 Measurement Properties

The psychometric or measurement properties considered critical for PRO instruments are validity, reliability and ability to detect change (FDA, 2009). Validity is the extent to which an instrument measures what it is meant to measure, while reliability refers to the reproducibility and consistency of measurement (Krabbe, 2016). Ability to detect change may not be considered a psychometric property but it is crucial if one were to use the findings of the measure for the purpose of evaluation. Other important characteristics of an instrument include interpretability; respondent and administration burden and cross cultural validity (Mokkink et al., 2010). Evaluation of psychometric properties include content validity assessment, construct validity testing using adequacy of measurement model and hypothesis testing, internal consistency and test-retest reliability assessment. A description of the psychometric properties are provided below but the focus of the thesis will be on assessing the adequacy of the measurement model, which is part of construct validity and explained in detail later on.

Content validity is invariably important to all measurement scales as it assesses how well the instrument captures all of the important aspects of health that the developers are intending to measure. Content validity determines whether the items cover all aspects of the concept important to patients and that saturation has been reached i.e. there is no new relevant or important information emerging (FDA, 2009). According to the FDA, content validity should be presented by documenting all item generation techniques: theoretical approach; population studied; source of items; selection, editing and reduction of items; cognitive interview summaries or transcripts; pilot testing; importance ratings and quantitative techniques for item evaluation.

Internal consistency is a measure of the extent to which items in a dimension or a measure are correlated and are measuring the same underlying concept (Fayers and Machin, 2013, Terwee et al., 2007). After determining the number of (homogeneous) dimensions, Cronbach's alpha is calculated for each dimension separately to measure inter-relatedness of items. A low Cronbach's alpha indicates a lack of correlation between the items in a scale indicating they are poorly related and cannot be combined to summary score. A very high Cronbach's alpha indicates high correlations among the items in the scale, and may indicate redundancy of one or more items. Although internal consistency is often regarded as a distinct concept it is closely related to construct validity as both methods make use of within-scale between-item correlations (Fayers and Machin, 2013). Another method used to measure internal consistency is the standard error of measurement, which is obtained from a crossed design analysis of variance (ANOVA). Internal consistency helps to avoid redundancy and ensure that items do not duplicate information collected with other items that have equal or better measurement properties (FDA, 2009).

Test-retest reliability or reproducibility is very important. Unstable scores on a repeated administration during which the respondent's condition did not change indicates inconsistency of the instrument and produces invalid assessment. On the other side, ability to detect change or responsiveness is also very important (FDA, 2009, Mokkink et al., 2010). If an instrument is intended to be evaluative and it has items that are not sensitive and does not change when there is known change in the concepts of interest, it will not be useful in evaluation (Fayers and Machin, 2013). Responsiveness also depends on response range and variability. A highly skewed distribution of item responses or a high percentage of patients responding at the floor or ceiling lowers the ability of the instrument to detect change. Also an item in which patients note that none of the response choices applies to them or one where all patients give the same answer (no variance) are not likely to detect differences even when known.

Hypothesis testing is often used to determine construct validity, where hypotheses based on known group differences, change over time or between measures are specified in advance and tested, under construct validity (Fayers and Machin, 2013, Terwee et al., 2007). Comparison of known groups and changes over time are tests for validity, but also a test of sensitivity or responsiveness. And it is very important, as Fayer et al state that, "a scale that cannot distinguish between groups with known differences, either because it lacks sensitivity or because it yields results that are contrary to expectations, is hardly likely to be of value for many purposes" (Fayers and Machin, 2013). Hypothesis testing between measures refers to convergent and discriminant validity between instruments. Other validity criteria included in the guideline are feasibility and acceptability, which is assessed from response rate; also distribution of responses, in particular avoidance of floor and ceiling effect, indicates that the item is well targeted to the population in question.

Apart from hypothesis testing, and usually before it, construct validity is determined by assessing the adequacy of the measurement model. Construct validity is concerned with the appropriateness of inferences made on the basis of observed and latent variables (Krabbe, 2016). This can be broken down to two parts, firstly confirming that the conceptual model is adequate and secondly confirming that the measurement model corresponds to the postulated theoretical construct using correlations (Fayers and Machin, 2013). The first part provides a descriptive assessment about the content represented in the draft measure and the second part provides more definite information about the measurement characteristics of the measure. The adequacy of a measurement model is evaluated by examining evidence that: 1) the scale/dimension measures a single conceptual domain or construct; 2) multiple dimensions measure distinct domains; 3) the scale adequately represents variability in the domains and 4) scoring procedures are justified (Aaronson et al., 2002).

3.5.1 Classical and modern test theory

The tests carried out to assess the measurement model are based on classical test theory and modern test theory. These tests examine the hypothesised item structure or relationship between observed items and construct. Classical test theory is largely based on either summated scales, in which the scores on multiple items are added together, or linear models such as factor analysis models. In contrast, models that are based on item response models fall within modern test theory. Multi-trait scaling analysis and principal component analysis are based on correlation and assesses whether the postulated scale is consistent with the response data. Factor analysis is used either as an automatic procedure to explore the patterns amongst the correlations, called exploratory factor analysis (EFA) or as a

confirmatory method (CFA) for testing whether the correlations correspond to the a priori structure of items. CFA does not allow the factors in the dataset to be freely estimated, it involves imposing a measurement model to the data. The assumption of unidimensionality, that all the items in the scale are measuring the same latent variable or construct, is also often tested using CFA as they reveal how items contribute to the underlying variable. Assessment of the measurement model allows testing and validating the association between factors and the relationship between items and factors. This removes some of the arbitrariness of using an exploratory approach and enables testing the model using further hypothesis. For example, association between two factors may be very high and this implies interaction.

Once the assumption of unidimensionality is confirmed, item response theory (IRT) can be used for examining scale structure, item selection and item calibration. IRT specifies the conditional distribution of the complete response pattern as a function of the latent factors and makes the assumption that responses to different variables are independent for given latent factors (conditional independence). It assumes that respondents with a particular level of QoL have a certain probability of responding positively to each question and that this probability is dependent on the 'difficulty' of the item in question.

The PROMIS guidance describes IRT as "a family of models that describe, in probabilistic terms, the relationship between a person's response to a question and his or her standing (level) on the PRO latent construct that the scale measures" (PROMIS, 2013). Finally differential item functioning (DIF) is assessed using methods based on IRT or logistic regression. DIF is

observed when the probability of item response differs across comparison groups such as gender, country or language despite having the same underlying true ability. Another modern trend in constructing descriptive system is the use of dynamic computer based questionnaires that only ask as many items as required for obtaining pre-specified precision and are called computer adaptive tests (CATs) (PROMIS, 2013).

Structural equation modelling (SEM) is a more general technique that encompasses factor analysis and regression techniques. The structural part of the model is estimated using a generalized least squares method or weighted least squares method and it illustrates how the latent variables of interest are related. While SEM may not work in a one-dimensional model because of high interaction between items, it can be used for dimension selection as it indicates contribution to overall multi-dimensional model. And those with very little contribution to the presumed underlying factor can be considered redundant.

3.6 Summary of findings from the focussed review

Eleven guidance documents were included in this review. Item generation involved review of the literature including existing instruments and qualitative research with healthcare professionals and patients. Although it is possible that there are more guidance documents available that I have not reviewed in this chapter, the review based on the included papers were sufficient to get an understanding of the stages involved in instrument development and psychometric assessments conducted.

The guidelines describe two development phases after item generation; these are pretesting and field testing of the full questionnaire. Pretesting

involves item refinement based on rating of relevance, importance and obtaining feedback on completeness and acceptability of items. Field-testing involves administration of the questionnaire to a larger sample of the target population and assessment of psychometric properties to enable item selection.

This chapter provided an overview of methods used in the development of a descriptive system and the measurement properties considered essential. The empirical component of my thesis (chapter 5, 6 and 7) will focus on phase 3 of the development process in which fieldwork is carried out. A full questionnaire is administered to the target respondents and descriptive system is generated based on statistical analysis of the responses.

This chapter focussed on stages of instrument development and psychometric testing more generally. In the next chapter the focus is on methods actually used. A systematic review of the literature was carried out to identify methods used to develop descriptive systems and while both item generation and item selection is reported. The focus of this thesis will be on the latter. Any information regarding criteria used in item selection was noted with the view of obtaining sufficient information to be able to analyse empirical dataset based on the review.

Chapter 4 Systematic review of methods used by instrument developers

Chapter 3 highlighted the assessment of measurement model. It is key to establish the relationship between observable and latent variables in order to conduct item selection and item selection forms the crux of empirical analyses carried out in subsequent chapters. It should also be noted that there are several statistical techniques available to determine the significance of the items to each dimension and none of the standards or guidance recommend one method over another.

4.1 Introduction

A key part of development of a new descriptive system is generating a pool of potential items and selecting items from the pool for inclusion in the instrument. Various methods for item selection are available and advances in technology over recent years have enabled a range of methodologies for item selection to become more accessible to instrument developers.

Several guidelines and standards have been published to bring rigor and consistency in instrument development. Most of them identify essential measurement properties and criteria to assess them. However there are many methods of analysis recommended and it is difficult to ascertain the ordering of these analyses (if any) and how it may affect item selection.

This study provides a contemporary overview of methods recommended in guidelines for item selection in instrument development and methods used by instrument developers. The review includes both disease specific and

generic QoL studies. The review is not exhaustive but aims to provide an analytical view on how these methods are used in developing a descriptive system, in particular item selection.

4.2 Methods

4.2.1 Systematic Review of Empirical Studies

Search Strategy

Academic papers reporting development of descriptive systems were identified in Medline, PsycINFO, Embase and Global Health databases using OvidSP and Scopus in September 2016 and January 2017 respectively. Supplementary searches were conducted by reviewing the bibliographies of included studies.

The studies identified from the literature were not included in isolation. If a study led to further studies, efforts were made to capture them to enable completeness. Relevant forward linkages were identified where possible, for example the studies assessing construct validity for ICECAP measures at a later date were included; similarly backward linkage was explored, for example if a measure was being extracted from an existing instrument, the studies reporting the development of the original questionnaire was also examined. The focus however was always on the descriptive system only, in particular item selection. Scale calibration and valuation was not explored in the study.

Inclusion and Exclusion Criteria

Papers reporting methods used for the development of descriptive systems specific to quality of life were included. Exclusion criteria included reviews,

papers not reporting patient reported outcome measures and papers not reporting the development of a descriptive system. No date restrictions were placed on the searches, but the review was restricted to English-language publications. One researcher (JS) independently screened titles and abstracts. Ten per cent of the studies and any study the first researcher was unsure about were screened by second independent reviewer (LL). Full text was obtained for all studies, which had been included by one or both reviewers.

4.2.2 Data Extraction

The following information was extracted from the included studies: study information, methods used and measurement properties assessed. Study information included name of authors, journal, publication year and the country of study, name of instrument, disease area and number of items in the new instrument. Methods extracted included information on empirical study design, method of analysis used for item generation and item selection. In addition, measurement properties assessed during the development of the descriptive system were noted.

4.3 Results

The review included a total of 61 articles describing development of an instrument or scale and it is illustrated in Figure 2. A total of 553 studies were identified through database searching using COSMIN search strategy, which is detailed in Appendix 1. Search was conducted using OvidSP and Scopus platform in September 2016 and January 2017 respectively. An additional 51 studies were identified through other sources, such as reference lists for preceding papers reporting earlier work and articles citing the included paper for any subsequent assessment of the descriptive system, where possible.

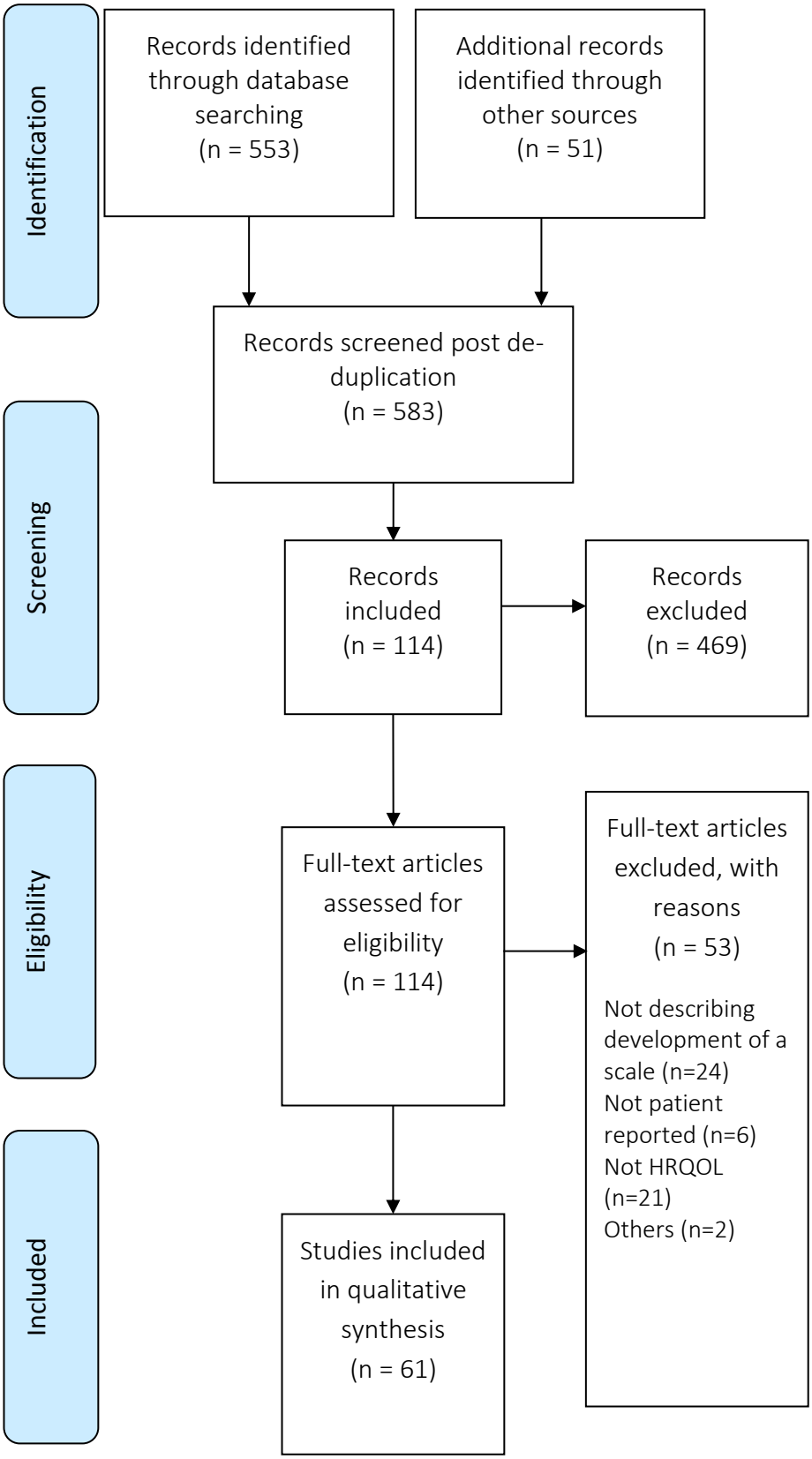
It is evident that the search strategy does not include development of all descriptive systems used to measure PRO.

Descriptive systems included in the review

The review included 61 studies covering 13 generic and 41 condition-specific QoL measures. One third of the measures reviewed were developed in UK (n=18), followed by Australia (n=8), USA (n=7), Canada (n=7), across multiple countries (n=8), Denmark (n=3), Netherlands (n=2), Greece (n=1) and Spain (n=1). Established generic measures such as the SF-20 and the EuroQoL index that were introduced as early as 1988 and 1990 respectively have been included in this review alongside their various versions (EQ-5D, SF-36 and SF-6D) and more recent instruments such as ICECAP (2006), AQoL (1999) and CHU-9D (2009). Disease or condition specific instruments such as AQLQ, EORTC QLQ-C30 and OHIP measure QoL in patients with asthma, cancer and oral health problems have also been included. A large number of condition specific measures focussed on item selection to develop a short form that was amenable to valuation.

The sections below provide a narrative summary of the methods used in the included studies, and detailed data extraction information is in Appendix 2.

Figure 2: Summary using PRISMA flow diagram



4.3.1 Stage One: Item Generation

There are two main approaches to developing a descriptive system, generating items de novo and using items from an existing measure.

4.3.1.1 Items de novo

A total of 19 studies included in the review used a bottom up approach whereby focus groups and interviews were conducted with relevant populations to generate items (ALSAQ-40, AQLQ, AQoL-6D, CAT-QoL, CHU-9D, CP QOL-Teen, CPCHILD, ICECAP-A, ICECAP-O, L-QoL, OAB-q, PORPUS, QoLIAD, PIQoL-AD, PSORIQoL, RAQoL, STQOLI, The Endometriosis Health Profile-30, VisQol, PROMIS and QualiPause toolkit). The sample sizes used in these studies were diverse, and ranged from 6 asthma patients for AQLQ to 74 children for CHU-9D and over 1000 participants for PROMIS. Qualitative data collected from focus groups and interviews were analysed using thematic analysis, critical incident technique and framework analysis, with few allowing for the analysis to be conducted iteratively.

4.3.1.2 Items from existing instruments

Five instruments relied on examination of several existing measures to identify items and create an item pool (EQ-5D, SF-36, P-PBMSI, item bank for knee pathology, PROMIS and AQoL-4D). Items for the EuroQoL index were derived from detailed examination of the descriptive content of existing health status measures including the Quality of Well Being Scale, the Sickness Impact Profile, Nottingham Health Profile, the Rosser Index and measures used by members of the EuroQol Group at the time of development. Similarly the developers of SF-36 used the 149-item Functioning and Well-Being Profile (FWBP) to develop the SF-36. The FWBP consists of items from existing measures such as General Psychological Well-Being Inventory, various physical and role functioning measures, the

Health Perceptions Questionnaire, and other measures that proved to be useful during the Health Insurance Experiment. There were two instruments that combined the item pool of an existing measure with additional items identified using focus groups (AQoL-6D) and items suggested by clinicians (HIT-6).

The development of an item bank for Patient-Reported Outcomes Measurement Information System (PROMIS) included extensive reviews and included over 10,000 items identified from existing measures. The PROMIS item bank aims to provide a foundation for developing short-form instruments and enabling computerized adaptive testing (CAT). The article by DeWalt et al. reported a step-wise qualitative item review process for over 10,000 items gathered from existing PROs (DeWalt et al., 2007). It included: identification of extant items, item classification and selection, item revision, focus group exploration of domain coverage, cognitive interviews on individual items, and final revision before field testing. A total of 138 interviews with patients were conducted for the following domains: physical functioning, fatigue, pain, emotional distress, and social role participation. This was followed by qualitative ratings and although the total numbers of participants who took part in this task is not reported, the study states that the rating of items in the physical function domain included 734 participants.

4.3.1.3 Items from an existing instrument (short form)

Almost half the studies (n=29) selected an existing instrument, which was lengthy, and referred to as a long form, to create simplified versions or short forms (SF). These studies focussed on item reduction, which is described in more detail in the next section.

4.3.2 Stage Two: Item Selection

A variety of methods were used for item selection in the studies reviewed and are summarised in table 3. Few studies relied on only one method for item selection; the majority used a combination of techniques. Some studies determined items based on relevance and importance, assessed from interviews and rating. There were a couple of instruments where item selection was not carried out, and a few that were determined by instrument developers. The most common approach was to apply statistical analyses to datasets containing completed questionnaires.

Statistical analyses such as multivariate regression analysis, EFA, principal component analysis (PCA), CFA, Rasch analysis, SEM and IRT were applied to response data; these techniques allow assessment of construct validity. In particular three statistical methods emerged and are discussed in this section. These were regression analysis, EFA and IRT. It should be noted dimensions were established using factor analysis or PCA before selecting items in the studies that used IRT or Rasch. Other psychometric properties were also examined (during or post item selection) and are reported in next section.

Table 3: Summary of Methods Used in Included Studies

Instrument Name	Condition	Final number of items	Item Generation 1: using qualitative studies 2: using items from existing measure(s)	Item Selection
SF-20 (Stewart et al., 1988)	Generic	20 items	2	By authors
EQ-5D (Williams, 1990, Brooks, 1996)	Generic	5 items	2	By authors
SF-36 (Ware and Sherbourne, 1992, Gandek and Ware, 1993)	Generic	36 items	2	By authors
AQLQ (Juniper et al., 1992, Juniper et al., 1997)	Asthma	32 items; 36 items	1	1) IS; 2) PCA
SF-12 (Ware Jr et al., 1996)	Generic	24 items	2	RA
RAQoL (De Jong et al., 1997)	Rheumatoid arthritis	30 items	1	Qual.
Short-form of OHIP (Slade, 1997)	Oral health	14 items	2	PCA and RA
WHOQOL-BREF (WHO, 1998)	Generic	26 items	2	Correlation analysis
ALSAQ-40 (Jenkinson et al., 1999)	Amyotrophic lateral sclerosis/motor neurone disease	40 items across five areas	1	EFA
AQoL (AQoL-4D) (Hawthorne et al., 1999)	Generic	15 items across 5 dimensions; NB: reduced to 4 later	2	PCA, EFA and SEM
McSad (Bennett et al., 2000)	Major, unipolar depression	6 items	2	By authors
PORPUS (Krahn et al., 2000)	Prostate cancer	10 items	1	IS
The Endometriosis Health Profile-30 (Jones et al., 2001)	Endometriosis	30 items across five scales	1	PCA

Instrument Name	Condition	Final number of items	Item Generation 1: using qualitative studies 2: using items from existing measure(s)	Item Selection
SF-6D (SF-36) (Brazier et al., 1998)	Generic	6 items	2	By authors
SF-6D (SF-12) (Brazier and Roberts, 2004)	Generic	6 items	2	By authors
OAB-q (Coyne et al., 2002)	Overactive bladder	25 items	1	EFA
HIT-6 (Kosinski et al., 2003)	Headache	6 items	1&2	IRT
Short-form for emotional scale of EORTC QLQ-C30 (Bjorner et al., 2004)	Cancer patients in palliative care	2-item emotional functioning scale	2	IRT
QoLIAD; PIQoL-AD; PSORIQoL (McKenna et al., 2004, Whalley et al., 2004)	Atopic dermatitis; parents of child with atopic dermatitis; psoriasis	25 items; 28 items and 25 items	1	Qual. and Rasch
VisQol (Misajon et al., 2004)	Vision-impairment	6 items	1	EFA, SEM and IRT
PFDI-20; PFIQ-7 (Barber et al., 2005)	Pelvic floor disorders	20 items & 7 items	2	RA
Knee pathology (item bank) (Comins et al., 2013)	Knee conditions	157 items	2	NA*
Short-form of three EORTC QLQ-C30 scales (Petersen et al., 2006)	Cancer patients in palliative care	3-item physical scale; 2-item fatigue scale; 1 item nausea scale; 2 item cognitive scale	2	IRT
CPCHILD (Narayanan et al., 2006)	Children with severe cerebral palsy	36 items across 6 dimensions	1	IS
EORTC QLQ-C15-PAL (Groenvold et al., 2006)	Cancer patients in palliative care	15 items	2	IS
Short form of $CPQ_{11–14}$ (Jokovic et al., 2006)	Oral health in children (age 11-	4 short forms were indicated	2	1) IS 2) RA

Instrument Name	Condition	Final number of items	Item Generation 1: using qualitative studies 2: using items from existing measure(s)	Item Selection
	14)	(two with 16 items and two with 8 items each)		
QualiPause toolkit (Brazier et al., 2005)	Menopause	22 items across 6 domains	1	EFA and authors
ICECAP-O (Coast et al., 2008a, Grewal et al., 2006)	Capability wellbeing (>65 years)	5 items	1	Qual.
PROMIS (item bank) (DeWalt et al., 2007, Reeve et al., 2007)	Physical functioning, fatigue, pain, emotional distress, and social role participation	Items customised to offer most precision for a given individual	1	IRT
DUI (Sundaram et al., 2009)	Diabetes	5 items	2	IS, EFA and Rasch
AQoL-8 (Hawthorne, 2009)	Generic	8 items	2	IRT*, Rasch and RA
L-QoL (Doward et al., 2009)	Systemic lupus erythematosus	25 items	1	Rasch
CHU-9D (Stevens, 2009, Stevens, 2011, Stevens, 2012)	Generic (children)	9 items	1	Qual.
CORE-6D (Mavranezouli et al., 2011)	Common mental health problems	6 items across 2 domains	2	Rasch and PCA
AQL-5D (Young et al., 2010)	Asthma	5 items	2	PCA and Rasch
MobQues28 (Dallmeijer et al., 2011)	Mobility in children with cerebral palsy	28 items	2	Rasch
PBM for DEMQOL & DEMQOL-Proxy (Mulhern et al., 2013)	Dementia	5 items in DEMQOL & 4 items in DEMQOL-Proxy	2	EFA and Rasch

Instrument Name	Condition	Final number of items	Item Generation 1: using qualitative studies 2: using items from existing measure(s)	Item Selection
NEWQOL-6D (Mulhern et al., 2012)	Epilepsy	6 items	2	EFA and Rasch
AQoL-6D (Richardson et al., 2012b)	Generic	20 items across 6 dimensions	1&2	EFA and SEM
AQoL-8D (Richardson et al., 2014)	Generic with psychosocial focus	35 items across 8 dimensions	2	EFA
Short form of NEI VFQ-25 (Kowalski et al., 2012)	Vision related	9 items across 6 domains	2	EFA and Rasch
STQOLI (Lyrakos et al., 2012)	Thalassemia	41 items across 4 domains	1	PCA
ICECAP-A (Al-Janabi et al., 2012, Al-Janabi et al., 2013)	Capability wellbeing (adults)	5 items	1	Qual.
CAT-QoL (Carlton, 2013b, Carlton, 2013a)	Amblyopia (children)	11 items	1	Qual.
CP QOL-Teen (Davis et al., 2013)	Cerebral palsy (adolescents)	72 items across 7 dimensions	1	NA*
AQoL-7D (Richardson et al., 2012a)	Vision related	26 items across 7 dimensions	2	By authors
P-PBMSI (Kuspinar et al., 2014)	Multiple sclerosis	5 items	2	Rasch
IUI (Cuervo et al., 2014)	Urinary conditions	5 items	2	PCA and IRT
MSIS-8D (Goodwin and Green, 2015)	Multiple sclerosis	8 items	2	EFA and Rasch
PBM for Myelofibrosis (Mukuria et al., 2015)	Myelofibrosis	8 items	2	EFA and Rasch
Short form of ThyPRO (Watt et al., 2015)	Thyroid	39 items	2	IRT
Note: IRT: item response theory; PCA: principal component analysis; FA: factor analysis; EFA: exploratory FA; SEM: structural equation modelling; IS: importance survey; RA:				

Instrument Name	Condition	Final number of items	Item Generation 1: using qualitative studies 2: using items from existing measure(s)	Item Selection
regression analysis; Qual.: qualitative study; NA*: not applicable as item selection was not conducted; IRT*: IRT based Mokken scale analysis				

4.3.2.1 Determined by scale developers

Item selection was not reported in detail for the generic instruments introduced in the late eighties and nineties. The EuroQoL index first introduced in 1990 included six dimensions and the items were selected so as to cover as many as possible of the domains most frequently covered by others. The instrument was modified to EQ-5D with five dimension based on further development work in 1991 (Williams, 1990, Brooks, 1996). The twenty items in SF-20 is made up of eighteen items from RAND Health Insurance Experiment (HIE) and two single item measures included from similar measures (Stewart et al., 1988). The items in SF-36 were selected to reproduce the "parent" scale, the medical outcomes study (MOS), and other psychometric standards; the authors report that the specific strategies used for item selection varied across the domains (Ware and Sherbourne, 1992). The actual analyses conducted were not reported.

The SF-6D preference-based measure (PBM) was derived from the SF-36, which includes 35 items across eight dimensions. The number of dimensions was reduced from eight to six by excluding general health and combining the two role limitation dimensions as one. The final six dimensions contained one item each but captured contents of 11 items

from the SF-36. A multidisciplinary team of researchers carried out item selection for each dimension based on judgements about following criteria: i) avoidance of redundancy, ii) preference given to negative items and iii) relative value of items and responses based on IQOLA study (Brazier et al., 1998). Another PBM, the SF-6D (SF-12), was derived from SF-12, which contains 12 items across 8 dimensions. Dimensions were reduced from 8 to 6 as previously; one item was chosen for each dimension based on findings from two studies which used Rasch, correlation and regression analyses (Brazier and Roberts, 2004). The developers of AQoL-7D grouped together 20 items of AQoL-6D with 6 items of VisQoL to get a vision related measure (Richardson et al., 2012b).

4.3.2.2 Item selection not conducted

There were two studies where the focus was on creating an item bank. The first study on knee pathology did not attempt item selection (40). The second (PROMIS) built an item bank so that items from it can be used for constructing a short form or enabling CAT that customises items according to individual characteristics to offer most precision (DeWalt et al., 2007, Reeve et al., 2007). PROMIS utilised datasets containing PRO responses with over 10,000 respondents (please see Appendix Table A2 for more detail) from the US, including members of the general population and patients, to identify candidate items for the item bank. The following analyses were conducted: evaluation of data quality, descriptive statistics, item response theory model assumptions, model fit, differential item functioning, and item calibration for item banking to allow for subsequent item selection by scale developers. Finally, the item selection carried out for the utility measure McSad is not reported (Bennett et al., 2000).

4.3.2.3 Determined using rating exercise with target population

Some developers used an importance rating approach during item selection (AQLQ, PORPUS, CPCHILD, EORTC QLQ-C15-PAL, Short form for CPQ_{11-14} and DUI). The AQLQ developers asked 150 patients to identify the frequency of all the 152 items in the item pool they had experienced during the last year. For each item indicated as experienced, the patient was asked to rate the importance of that item. Results were expressed in terms of 'frequency' (the proportion of patients experiencing a particular item), 'importance' (the mean importance score attached to each item), and 'impact'. Impact was generated as a product of 'frequency' and 'importance'. The items were ranked according to their impact score and, in general, the highest scoring items were selected (Juniper et al., 1992, Juniper et al., 1997). A similar exercise was carried out to generate the short form for CPQ_{11-14} (Jokovic et al., 2006). Items were rated on importance during item selection for PORPUS (by both patients and medical practitioners), CPCHILD (caregivers of children with CP) and DUI (by patients and experts). The rating dataset was analysed using a weighted score for PORPUS and CPCHILD (Krahn et al., 2000, Narayanan et al., 2006); and factor analysis was used in DUI (Sundaram et al., 2009). In addition to rating 'appropriateness', patients and healthcare professionals assessed 'relevance' and 'importance' during the development of EORTC QLQ C15-PAL (Bjorner et al., 2004). Participants were asked whether they perceived any of the items as inappropriate or upsetting and whether there were any additional issues not included in the questionnaire that are relevant for evaluating the outcome of palliative care. The ratings of each item were linearly transformed to a 0–100 scale and used in determining item selection.

4.3.2.4 Determined using qualitative studies with target population

The two wellbeing capability measures, ICECAP-O and ICECAP-A, relied on qualitative data analysis to transform the attributes into descriptive systems (Coast et al., 2008b, Al-Janabi et al., 2012). The attributes identified in ICECAP-O were refined and reworded using some of the informants who participated in the previous study developing attributes. Semi-structured interviews were conducted (N=19), to refine the terminology used to describe each attribute and levels within it. An iterative approach was taken to ensure the meaning of the terms were tested, altered and refined. Semi-structured interviews were also conducted to determine one item per attribute for ICECAP-A (N=18). This involved asking participants how lists of the specific concepts related to each attribute could best be summarised and determining wording. Analysis was conducted iteratively and a changing coding framework was used to identify themes that "represented what was ultimately important in individuals' lives." The scale developers used interviews with the target population to determine wording, levels (response scale) and item presentation. The selection of items for inclusion in RAQoL reflected the frequency with which issues were raised by the interviewees (De Jong et al., 1997). These items were tested in a pilot study and items were removed based on internal consistency, correlation with other items (too low/high) and distribution of responses (skewed).

Finally, item selection in CHU-9D and CAT-QoL relied on the initial qualitative studies, in other words it did not separate item generation and item selection (Carlton, 2013b, Carlton, 2013a, Stevens, 2009, Stevens, 2011). The themes identified from the interviews were used as items of the descriptive system, and in the next stage studies were carried out to attach response levels and refine wording

Item selection using statistical models

Nearly 70% of the included studies used statistical models to analyse responses included in the item pool generated in phase one or existing instrument and select items. While some studies conducted empirical studies to collect response data, others relied on existing datasets. Three different types of statistical models were applied: 1) regression analysis, 2) factor analysis and 3) IRT models. These methods were applied in isolation as well as in combination and part of psychometric evaluation (see section 3.5.1). Factor analysis falls within the classical test theory and includes EFA, CFA and SEM. IRT is based on modern test theory and includes Rasch analysis.

4.3.2.5 Determined using regression analysis

Item selection based on regression analysis was found in four studies (SF-12; PFDI-20 and PFIQ-7; Short-form of OHIP and CPQ_{11-14}). The SF-12 was first introduced in 1996 as a subset of SF-36 that produced a score and explained at least 90% of the variance in SF-36 physical and mental health summary; other criteria were that is highly comparable to the eight-scale profile, reproduces the average scores for the summary measures and is brief (Ware Jr et al., 1996). Forward-step regression analysis was applied in the dataset (N=3363) and ten items were sufficient to reproduce both the PCS-36 and MCS-36 scores with an R^2 above 0.90 but two additional items were selected to represent all eight concepts (Ware Jr et al., 1996). Instrument developers of PFDI-20 and PFIQ-7employed regression analysis to find items for short form that best predicted the scale score on the respective long form (Barber et al., 2005). When different items appeared equivalent, a choice was made on item content by developers.

For the analysis of OHIP, the total OHIP score, obtained by summing the coded Likert-type responses from all questions, formed the dependent variable and each question was an independent variable (Slade, 1997). A controlled stepwise procedure was carried out to the full OHIP dataset (n=1217) and items making the greatest contribution to total R^2 were added sequentially. Please note that no more than two items from each conceptual dimension (previously determined by PCA) were allowed to enter the model. Similarly developers of CPQ_{11-14} used a forward stepwise procedure to identify the best predictors of the overall score and created two versions of the short form: using four and two items from each domain (Jokovic et al., 2006).

Only one study relied on correlation analysis. The developers of WHOQOL-BREF chose items that were "correlated most highly with the total score, calculated as the mean of all facets" (WHO, 1998).

4.3.2.6 Determined using factor analysis techniques

The factor analysis techniques use the underlying response variable approach, which assumes that each observed variable is generated by an underlying unobserved continuous variable. It utilises a correlation or covariance matrix to compare the estimated correlation matrix with the observed correlation matrix. The difference between the observed and the expected correlation matrix is found in the residual correlation matrix. Factor analysis is used to determine the underlying dimensions of data and distil large amounts of data into simpler structures using multivariate descriptive methods. These models can be used for item reduction and assessment of construct validity.

Use of exploratory factor analytic methods, namely EFA and PCA, was reported in 18 studies included in the review as illustrated in Table 2. It was the primary analysis in eight studies. For example, in the ALSAQ-40 items were reduced from 78 items to 40 using EFA (Jenkinson et al., 1999). Only factors with an eigenvalue over 1 were initially retained and items with a loading of <0.5 for any of the factors were excluded; this yielded 13 factors. In addition internal reliability was assessed for the items constituting each scale until only those with largest contribution to the scale remained. The developers of AQoL-8D reported that "a combination of restrictive and unrestrictive factor analyses" was used to create the descriptive system but did not provide any detail about what it entailed (Richardson et al., 2014). Similarly, the study reporting preference elicitation for the QualiPause toolkit described application of EFA and psychometric criteria to create a descriptive system with 22 items across 6 domains (Brazier et al., 2005). The "most robust item(s) for all domains" were selected, resulting in a classification system with 7 items but how robustness was assessed is not reported.

Finally, developers of AQLQ also applied PCA to the dataset after removing skewed items, those with frequency < 40% and item total correlations less than 0.40 (Juniper et al., 1992, Juniper et al., 1997). Items loading by less than 0.4 on the first factor were removed and varimax rotations elicited 3, 4, 5, and 6 factors. Three clinicians reviewed these groupings and selected the one that made the most sense. A descriptive system with 36 items was formed using this method. Please note that applying impact score technique described earlier generated 32 items for AQLQ measure (Juniper et al., 1992, Juniper et al., 1997).

Three other measures, namely OAB-q, The Endometriosis Health Profile-30 and STQOLI applied PCA and psychometric criteria during item selection. Decision rules for item reduction in OAB-q were: (1) >60% of participants denied the occurrence or impact of the item; (2) low item to total correlations (<0.40); or (3) inadequate factor loading on any factor (<0.40) or >0.40 on more than one factor to be excluded (Coyne et al., 2002). For the Endometriosis Health Profile-30, an 87-item questionnaire was administered to 1000 women. The extraction method used was PCA, with varimax rotation (Jones et al., 2001). Psychometric tests were also carried out. Finally in STQOLI a first selection of items was made from the descriptive response distribution for each item, followed by assessment of psychometric properties including PCA (Lyrakos et al., 2012). Also note that PCA was fitted to data collected from adolescents with cerebral palsy (n=87) and primary caregivers (n=112) to determine domain structure of CP QOL-Teen (Davis et al., 2013); however the developers used PCA to confirm scale structure and not to conduct item selection.

SEM was applied in the development of descriptive systems for AQoL-4D, AQoL-6D and VisQoL. SEM is a multi-level model that combines observed items into latent constructs and secondly estimates the relationship between the latent constructs (including overall underlying measure). It is also used to confirm the hypothesised structure of the model. Item selection for AQoL-4D entailed an iterative process comprising PCA, EFA and SEM to responses obtained from hospital patients, and community members (n=996). They specifically used SEM to determine the explanatory power of the derived model. Similarly item selection for AQoL-6D was based on a combination of EFA and SEM in the complete dataset (n=709) (Richardson et al., 2012b). The development of descriptive system

for VisQoL also involved SEM, and it was used as the final step to confirming the model (Misajon et al., 2004).

4.3.2.7 Determined using IRT

A two-staged approach was often adopted in which EFA and PCA was used to identify or confirm hypothesised scales in datasets, followed by application of the IRT model to reduce the number of items in each dimension. The IRT specifies the conditional distribution of the complete response pattern as a function of the latent factors and makes the assumption that responses to different variables are independent for given latent factors (conditional independence) (Moustaki, 2003). This approach is particularly developed within a single latent factor, and the response function is in either logit or probit form.

All IRT models require an assumption of unidimensionality to hold, which implies: 1) items represent only one latent variable; 2) local independence and 3) no differential item functioning (DIF). The general IRT models, such as the graded response and generalized partial credit models, which model the data at hand; and the more restrictive IRT models called Rasch generally, and include partial credit and rating scale model. Rasch focuses on the theoretical requirements for a good measurement, this implies fitting data to the model and excluding misfitting items that do not address the latent variable. Items with disordered response levels are excluded from the model because it suggests that the individual could not distinguish between response levels; items displaying DIF are also excluded because it indicates that the item systematically performs differently across different sample characteristics.

In seven studies EFA or PCA was used to identify hypothesised scales in the dataset, before using Rasch to extract items per dimension (PBM derived for DEMQOL, DEMQOL-Proxy, Myelofibrosis and EORTC QLQ-C30, NEWQOL-6D, AQL-5D, CORE-6D and DUI). Rasch analyses were used to select one or two items to represent each dimension for the preference based measures. The studies that involved Rasch (L-QoL, MobQues28, QoLIAD, PIQoL-AD, PSORIQoL and P-PBMSI), but did not use EFA or PCA upfront, used other techniques to hypothesise dimensions in the dataset (and is described below).

The L-QoL, QoLIAD, PIQoL-AD, PSORIQoL and the MobQues28 applied Rasch to the dimensions elicited from the qualitative studies (Dallmeijer et al., 2011, Doward et al., 2009, McKenna et al., 2004, Whalley et al., 2004). The P-PBMSI was extracted from an existing measure called PGI; each patient's response on the PGI was mapped to the ICF domains independently by four raters and Rasch analysis was carried out in each domain (Kuspinar et al., 2014). CORE-6D developers used successive Rasch analyses for item selection in each domain of the existing measure (CORE-OM) and used PCA later as "an extra post hoc test" to confirm the unidimensionality of the new scale (Mavranezouli et al., 2011).

The PBMs derived for DEMQOL and DEMQOL-Proxy, Myelofibrosis, vision (from NEI VFQ-25), cancer (from EORTC QLQ-C30 & MF-SAF), Epilepsy (NEWQOL-6D), Asthma (AQL-5D), multiple sclerosis (MSIS-8D), Diabetes (DUI) and one generic instrument developer (AQoL-8) noted the need for an abbreviated version of the existing instrument for the instrument to be amenable to valuation. These studies focussed on item reduction and the final measure included only four to nine items. Some of the studies allowed only one (best-performing or representative) item per dimension to be

selected (DEMQOL and DEMQOL-Proxy, cancer, epilepsy and asthma). Some studies validated the item selection by fitting the measurement model on a second subset of the target population (NEWQOL-6D, cancer, DUI, Myeloma, AQL-5D, CORE-6D, L-QoL, PFDI-20 and PFIQ-7).

IRT models were reported in only four included studies (Short-form of EORTC QLQ-C30 emotional functioning scale, and three other scales, PROMIS based instruments and ThyPRO). Data from European cancer studies were analysed using generalised partial credit (GPCM) model to select items in emotional and fatigue scale of the EORTC QLQ-C30; and the more restrictive partial credit model (PCM) for physical functioning, nausea and cognitive scale (Bjorner et al., 2004, Petersen et al., 2006). IRT models were fitted to the scales of ThyPRO dataset (n=907) for item selection and the 85-item long form was reduced to a 39-item short form (Watt et al., 2015). PROMIS (discussed earlier in this section) used the Graded Response model for both item and scale analysis and for item calibration to enable CAT (DeWalt et al., 2007, Reeve et al., 2007).

Finally one study reported use of IRT based Mokken analysis, in addition to Rasch and regression analysis (Hawthorne, 2009). Data from the AQoL validation database (n=996) were reanalysed to identify the least fitting items, which were removed and AQoL-8 was created.

4.3.3 Psychometrics Properties Examined During Development Phase

The review I conducted captured psychometrics properties examined during the development of descriptive system. Some instrument developers did not report evaluation of psychometric properties in the seminal article introducing it and are not reported here (QualiPause toolkit,

SF-20, EQ-5D, SF-36, McSad, SF-6D (SF-36), SF-6D (SF-12), AQoL-7D and Knee pathology item bank). The psychometric criteria applied during the development of the remaining descriptive system is detailed in Appendix 3 and a summary is provided below. Psychometric assessment of validity using hypothesis testing requires administration of multiple instruments or qualitative research, while assessment of reliability and responsiveness assessment require repeated administration. The focus of my thesis is on a single empirical dataset and item selection using statistical analyses. Additional data collection was not conducted.

The most frequently cited evaluation in the included studies was content validity; this was assessed from qualitative studies with relevant population and experts during item generation, selection and validation. Also content validity is assumed as given if the items are derived from an existing measure. For example, MSIS-8D is derived from a 29-item Multiple Sclerosis Impact Scale (MSIS-29); the development and psychometric properties of MSIS-29 are already reported in the literature elsewhere (Hobart et al., 2001, Riazi et al., 2002).

Guidelines, reviewed in chapter 3, advise scale developers to conduct pilot and field studies to examine the descriptive statistics of the responses including missing values, distribution of responses (floor and ceiling effect) and internal consistency examined using inter item correlations. More than half the studies included in the review reported using the range and distribution of responses and internal consistency during item selection. During the development of RAQoL, two pilot studies were conducted (n=50 each). Items were excluded from draft descriptive system based on internal consistency, correlation with other items (too low/high) and distribution of responses (skewed) (De Jong et al., 1997). On the other hand CHU-9D data

(n=247) were assessed to examine practicality (including response rates, completion rates and time to complete), item presentation and validity but this was conducted post item selection (Stevens, 2009, Stevens, 2011).

Construct validity was assessed either using hypothesis testing such as known group, convergent and discriminant validity; or during item selection by use of statistical measurement model such as IRT and FA. Both techniques were frequently cited in the literature. Test-retest of overall score and/or response to items was reported in only one fifth of the studies. Other measurement properties that were reported in a few studies were concurrent validity (for example IUI, short form of OHIP, shortened scales of EORTC QLQ-C30 and WHOQOL-BREF) and responsiveness (for example AQL-5D, EQ-5D and AQoL-7D).

4.3.4 Selection of Methods

Based on the empirical review of methods used for development of a descriptive system, I have summarised key methods for item generation and item selection below.

Item generation

Guidelines presented earlier (chapter 3) highlighted the use of target populations and relevant clinical experts during the development phases to improve the face validity and content validity of the instrument. Scale developers using qualitative studies for item generation often mentioned this point, but the descriptive systems derived from item reduction would also have benefitted from this attribute. Developers of the original instruments presumably used qualitative studies with relevant population.

I noted that a majority of generic measures relied on item generation using items from existing measures and this could be because of focus on completeness. Selecting items from a variety of or all existing measures allows for a wide range of items to be captured and increases the likelihood of including all aspects of health. The more recent generic PRO measures such as CHU-9D and ICECAP however rely on qualitative studies. This could be because of their specific focus on children and capability wellbeing that has less evidence compared to generic QoL measure and a need to be custom built. Amongst condition specific measures, the use of qualitative studies and use of items from existing measures was mixed. It should also be noted that a large number of instrument developers focussed on creating a short form from an existing instrument that is amenable to valuation.

Item selection

The methods used for item selection varied substantially. Guidelines recommended evaluation of measurement properties to determine item selection, and this includes assessment of content validity, construct validity (hypothesis testing and measurement model), internal consistency and test-retest reliability. The majority of instrument developers included in the review examined observed item responses to make decisions on item selection, but there were a few in which observed responses were not reported at all or relied completely on qualitative studies (with focus being content validity only).

Few of the methods identified by the guidance and used by the instrument developers are interchangeable in terms of purpose. Firstly, item selection focussing on different measurement properties is likely to generate different instruments. For example, developers of AQLQ noted that item

selection using PCA that relies on construct validity yielded very different results to item impact method that uses frequency and important rating and focuses on content validity. Is there an inherent ordering in terms of examination of measurement properties? This was not clear from the review.

Based on the development phases described in guidelines and summarised above, content validity should be established before item selection is carried out and item selection takes place during the field testing when the items have been administered to a large sample. But what if item selection was carried out based completely on qualitative studies, without having the questionnaire completed by respondents to examine responses? This is acceptable if the items are causal and psychometric criteria are not applicable. However it is preferable to confirm the measurement model first if the instrument developers are looking to attach scores or elicit utility values.

Even within statistical methods used to establish construct validity, there are several techniques that can be used for the same purpose. For example, both EFA and PCA can be used to determine dimensions; SEM, Rasch and inter-item correlations can be used to identify items that are redundant. Do these techniques always result in excluding the same items? Only one study included in this review compared two different techniques and further evidence needs to be obtained to be able to answer this question. Nevertheless, it is important that the developers explain why one method was chosen over another.

Finally there were a few studies that used regression models to determine the items most important to the overall construct being measured. This approach was not mentioned in the guidelines. Both factor analysis and IRT use regression methods to confirm the measurement model, however this is carried out only after the hypothetical structure of the data has been established to identify dimensions and the dependent variable is the underlying construct being measured.

4.4 Discussion

PROs are used to measure health and the impact of healthcare activities, and are often used to inform decision-making. If a measure is to be used to inform decisions, it needs to be appropriate and developed using robust methods. Over the years many guidance and standards have been put in place to facilitate this. This review of methods helped me to identify preferred methods for developing a descriptive system for patient experience.

4.4.1 Summary of findings

The search conducted in the study was based on the COSMIN search strategy. In the documents reviewed, item generation involved two distinct routes; those developing the item de novo and those using existing measures. The latter can be broken down into those using more than one measure and those focussing on a single measure (to generate short forms). Item selection was determined using statistical analyses in datasets containing completed questionnaires. A staged approach was used to establish dimensions first, with EFA or PCA to ensure distinct dimensions, before finalising items in each dimension, using IRT or factor analysis techniques. Note that instruments developed using non-quantitative

techniques and regression analysis bypassed this approach and went straight to item selection. Some instrument developers used confirmatory approaches (CFA or SEM) to further assess that the dimensions within the instrument were related constructs. These measurement models establish construct validity and enable valuation in the future.

Few of the statistical methods identified are interchangeable in terms of purpose, but may generate very different instruments. The developers seldom explained why one statistical model was chosen over another. For example, EFA and PCA are used to determine structurally independent factors in the dataset, but may generate different results. Similarly both IRT and SEM can be used for item reduction. An understanding of the underlying mechanics and discussions about strengths and limitations of these methods is necessary in order to make decision on why one method may be more appropriate than the other.

4.4.2 Gaps in the literature

There are different statistical techniques available to 1) establish dimensions and 2) to select items within dimensions. What is the impact of choosing one method over another on the performance of the instrument? Do the different analyses generate different descriptive systems? Why? The answer to these questions could not be verified from this review.

For example, both PCA and EFA can be used to summarise the dataset into a smaller subset of components or factors. However the underlying mechanics between these two methods are different. PCA is seeking components to maximize the variance of a linear combination of the

variables, whereas EFA seeks to account for the covariances or correlations among the variables (Bartholomew et al., 2008). Is one more appropriate or preferable than the other? Also one perhaps needs to take into account the nature of variables and available software while conducting the analysis. Similar can be said about use of IRT, SEM, Rasch and other analysis to identify and exclude misfitting items from the dimensions. Comparisons of a few of these methods are available in the literature (Cappelleri et al., 2014, Petrillo et al., 2015). But the guidance documents, reported in chapter 3, and the articles reviewed on instrument development did not report why one statistical model was chosen over another in their articles. There is a need for more empirical studies and wide-ranging evidence to be made available for those interested in instrument development.

A possible approach I could have adopted to obtain information on the rationale for the methods used was direct correspondence with the authors or instrument developers. However investigation of methods used and associated rationale for it became a secondary aim of this thesis. I examined use of IRT and factor analysis for item selection in the Inpatient dataset to add to discussions about the best method to use in the development of any new descriptive system. Nevertheless it should be noted that there is no clarity on how to finalise the instrument if the developers generate very different instruments using different statistical techniques. Conducting qualitative studies to determine the validity of the items may be one option but this addresses face validity rather than construct validity. Another option would be to collect evidence on psychometric properties using the two instruments and comparing the findings in terms of responsiveness, reliability and range of responses. While the second is robust it is onerous in terms of resources. It is not

possible to analyse this empirically in the remit of my thesis but a discussion on this issue is presented in my discussion chapter.

In chapters 5, 6 and 7, I present stage-wise analysis of the Inpatient Survey 2014 dataset using the statistical techniques identified in the review. In chapter 5, I describe the Inpatient Survey in detail and conduct regression analysis to gain an understanding of the dataset. Note that instrument developers have used regression analysis for item selection, but guidelines and standards have not recommended it as a method. Chapter 6 focuses on assessing dimensionality of the dataset using factor analysis. Item selection cannot take place until distinct dimensions or constructs have been determined. Chapter 7 describes item selection using IRT and factor analysis techniques.

Chapter 5 NHS Inpatient Survey – Data Familiarisation and Regression Analysis

5.1 Introduction

Following the background, review and aims presented in previous chapters, I introduce the primary dataset of my thesis in this chapter and the first set of analyses. There are different statistical techniques available 1) to establish dimensions and 2) to select items within dimensions. Some instruments reviewed in chapter 4 used regression analysis to select items, such as the SF-12 and OHIP (Ware Jr et al., 1996, Slade, 1997). Ware el al bypassed dimension selection and applied item selection directly using regression analysis. This is the approach used in this chapter. In the two subsequent chapters a stage wise approach will be adopted, whereby dimensions will be estimated first and then item selection per dimension.

In terms of patient experience datasets, NHS England has one of the largest patient survey programmes in the world called the National Patient Survey Programme. It monitors patient experience related to inpatient, outpatient, A&E, community mental health and maternity services across all NHS trusts by obtaining feedback about healthcare from recent users of the NHS; and this is conducted every year. The instruments ask patients to assess events or processes that they experience within the healthcare service, for example views on cleanliness of hospital ward or if the discharge was delayed. The Inpatient survey is currently administrated by Care Quality Commission (CQC), which also acts as the main regulator of the NHS (CQC, 2016). I have used the NHS Inpatient dataset in this thesis to investigate aspects related to patient experience during hospital stay using different statistical techniques to create a brief instrument to measure patient experience.

This chapter first provides a detailed narrative of the survey questionnaire used in the Inpatient survey to help familiarise readers with the dataset. Secondly descriptive statistics and multivariate regression analysis are conducted to identify significant processes and events related to overall patient experience rating.

Instrument developers have used regression method in the past for item selection, such as the SF-12 which was derived from SF-36 using regression analysis, and it is intuitive to use items significantly associated with overall patient experience rating in an instrument developed to measure patient experience. Instrument developed have used this method (Barber et al., 2005, Jokovic et al., 2006, Slade, 1997, Ware Jr et al., 1996) but guidelines examined (see chapter 3) did not advocate it. The focus here is on getting a good understanding of the data and test the method in the context of instrument development. The methods recommended for the development of a descriptive system by guidance and standards are employed in chapter six and seven.

5.1.1 Inpatient survey

The NHS Inpatient dataset was used in this thesis to investigate aspects related to patient experience during hospital stay using advanced statistical techniques. The main reasons for selecting this dataset are: it is contemporary with a very large sample of NHS users that allow response data to be examined in detail. Secondly it was developed using rigorous methods and it is comprehensive, covering all aspects of care during a hospital stay.

The Inpatient Survey questionnaire was developed after rigorous qualitative research and pilot studies. The Picker institute created the questionnaire using the framework of patient experience it had developed in the USA (Gerteis et al., 1993). Further consultation with experts, systematic review of literature, in-depth interviews and focus groups with patients in the UK were conducted to determine key issues encountered in healthcare to inform the questionnaire (Reeves et al., 2002b, Cleary et al., 1993). The Picker framework of patient experience consisted of seven specific dimensions of patient care: information and communication, coordination of care, respect for patient preferences, involvement of family and friends, and continuity and transition. The framework was expanded to include 'processes and events' within each dimension to form the current questionnaire.

5.1.2 Overall patience experience rating

The inpatient survey is detailed and comprises over seventy questions related to patient experience and additional background questions (NHS trust code, length of stay, age and gender) across seven settings of a hospital stay: admission; doctors and nurses; hospital and ward; leaving the hospital; operations and procedures; and patient care and treatment. The questionnaire follows a chronological format to include all key processes or events a patient is likely to undergo in an inpatient stay.

A majority of questions in the Inpatient survey questionnaire have responses that can be ordered in terms of magnitude of utility or disutility obtained. Some of these questions ask patients to factually report what happened, for example: 'Were you offered a choice of food?' While others ask for personal views on the service experienced, for example: 'In your opinion, were there enough nurses on duty to care for you in hospital?'

Additionally there are nominal questions that do not have a logical ordering in terms of impact on patient experience, for example: 'During your stay in hospital, did you have an operation or procedure?' Finally the questionnaire includes an overall satisfaction question to elicit patient experience on a rating scale with 10 being the highest score and representing 'I had a very good experience' and 0 being the lowest score, representing 'I had a very bad experience'. However, the CQC does not rely on this self-assessed rating score to compare performance of NHS trusts. Instead it produces a separate index measure called the overall patient experience score (CQC, 2016).

The CQC generates an overall score by placing the ordered responses onto a scale and aggregating across the items. The item assessing the performance of the trust across different aspect are scored, for example if a patient answers 'yes' to a question about unnecessary delay in discharge from the hospital, a score of 0 is allocated and 10 if 'no'. The questions used for filtering respondents to relevant or applicable question (such as: 'Was your most recent hospital planned in advance or an emergency?' was asked before asking questions specific to A&E), or those used for information purpose (such as background characteristics) are not scored. The CQC further summarises the overall score for each of the seven settings of hospital stay by combining items. This enable detailed scoring on different sections of patient experience and indication of specific areas that need improvement. However focus here is on dimensions rather than individual items.

There have been two studies to generate a brief measure by determining core dimensions, specifically in an inpatient setting (Jenkinson et al., 2002b, Sizmur and Redding, 2009). One study used regression analysis to

determine the dimensions that are most likely to influence satisfaction with care (Sizmur and Redding, 2009). In the study a mean score was assigned to the seven dimensions of care based on the Picker framework of patient experience, by grouping together items a priori. A value of zero was assigned to indicate patient reporting no problem in the area and 100 indicating maximum problems reported to score each dimension. This approach focussed on dimensions in relation to overall patient experience and this approach did not examine each item individually in relation to overall patient experience.

In a study by Jenkinson et al. a total of fifteen items were identified from the Picker inpatient questionnaires based on analysis of survey data of patients who had attended acute care hospitals in five European countries (Jenkinson et al., 2002b). The items were selected based on face validity, internal consistency reliability and high correlation with parent instrument. One of the inclusion criteria in the study was that items were applicable to as many respondents as possible (e.g. questions on emergency admissions will not be applicable to in-patients who had planned admissions). Items not applicable to a large proportion of respondents were excluded. While this criterion enables development of a generic instrument, it is possible that this approach excludes items that are important. For example, how much information was given about your condition or treatment while you were in the A&E department is not applicable for those who had planned admission but will have a significant impact on patient experience of an A&E patient.

In this chapter, I use self-assessed satisfaction rating as an indicator of patient's overall assessment rather than the derived overall patient experience score. Secondly instead of assigning items into seven sections

and using these dimensions as explanatory variables I include all items assessing process and events as explanatory variables. Finally I use subgroups to explore how items related to overall patient experience differed by subgroups (namely route of admission and whether or not they had an operation or procedure). I use econometric methods to analyse the dataset, specifically ordered logit regression which allows the dependent variable to be ordered and does not require the response levels to be equidistant. The regression models aim to identify healthcare events and processes most significantly associated with patient experience rating in an inpatient stay.

5.2 Methods

The details of the Inpatient questionnaire is summarised in this section followed by several issues that were identified before analysis could be carried out. The dataset contained missing data, 'I don't know' responses, 'not applicable' responses and questions that were not applicable to all. The measures taken to address them are presented here, followed by description of multivariate regression techniques used to analyse the dataset.

5.2.1 Dataset

This thesis focuses on the Inpatient Survey 2014, which was administered via post to patients aged 16 years or older across the UK who had been admitted to an NHS hospital with at least one overnight stay between September 2013 and January 2014. The survey had a response rate of 49% resulting in 62,443 completed and returned questionnaires. A majority of non-response was due to the patient not returning the questionnaire and not providing a reason; there were also patients who died, were too ill or

were not eligible to fill in question and those who could not be contacted because the patient relocated.

As the dataset that I have used is an anonymised secondary dataset obtained from publicly available source, I followed the best practice guidelines prevalent at the time in Brunel University London. The data was collected by the NHS CQC as part of their patient survey program. I downloaded the data from the UK data archive using my university affiliations and I obtained it free of cost. The raw dataset was downloaded to Stata13 from the UK Data archive (http://data-archive.ac.uk/) in September 2014. A detailed table presenting the questions and the distribution of responses is in Appendix 4. The terms and conditions for the use of this data are available in following web page that I have fully met: https://www.ukdataservice.ac.uk/get-data/how-to-access/conditions

Hospital users from a total of 156 trusts participated in the survey and the response rate varied across the trusts. The type of care provided to patients is likely to be similar within a trust and I applied cluster analysis (partition method) to adjust for this cluster effect in all the regression models by breaking the observations into trust groups, which were non-hierarchical and non-overlapping.

5.2.2 Coding

I consulted the scoring system used by CQC to recode the data for ordering of responses (CQC, 2014). The coding ensured response levels were in the same direction in order to ease interpretation. A lower score is better on all items and the highest score is assigned to worst level. The survey comprised a number of questions with 'I don't know' or 'I can't remember' or 'Unsure' as a response option and ranged from less than one per cent to

nineteen per cent. These respondents might have legitimately not known the answer to the question or were undecided about it. One option to handle 'don't know' answer is to treat is as missing but that may result in loss of information. In this study I assumed these responses as 'neutral' or middle category. For example I coded a 'No', indicating information was not provided, as a 1; I coded 'Don't know' as a 2 and a 'Yes' as a 3.

5.2.3 Inapplicable questions

Not all the questions in the survey were applicable to all respondents. This was because the question asked differed according to the inpatient journey (Appendix 5). There were a few filter questions, following which respondent was told to skip a question or a set of questions based on response provided. Although it is common to have irrelevant or inapplicable items for subgroups of respondents, there is no standard protocol to deal with it (Carpita and Manisera, 2011). One option is to specify them as missing values if the inapplicable responses display a non-random pattern of missing data. However there were some variables in which the inapplicable category accounted a large percentage of response and treating them as missing data would bias the results. Another option is to impute the response. One way is to categorise inapplicable response as another fixed category across all questions or employing imputation procedures to fill in missing values. However, if a large proportion of the response is indicated as inapplicable, imputation is not possible and this was the case in twelve questions described below. The final approach is to divide the sample into subgroups, excluding irrelevant sections for each subgroup.

There were a total of twelve questions which were not applicable to all patients and belonged to two main sections: 1) the admission section that

had separate sets of questions for inpatients admitted via an emergency route and for planned admissions and 2) the section on operation and procedures which was not applicable to patients who did not have any procedure. All patients were divided across these two key questions and imputation was not possible. In order to take this into account, the dataset was split into four subgroups that captured all possible combinations: 1) emergency admission with operation 2) emergency admission without operation 3) planned admission with operation and 4) planned admission without operation.

There were additional filter questions with single follow-up question(s) that was not relevant to all, these were:

- Questions on accident and emergency (A&E) were not applicable to urgent cases although they were not planned admission either.
- Questions on pain management were not relevant to those who did not report pain.
- Question on anaesthetic was not relevant for those who didn't require anaesthetics.
- Question related to stay in subsequent ward of the patient relevant for those who used more than one ward but not for those who stayed in one ward only.
- Questions about medication on discharge were not applicable to all patients, as some of them were not prescribed any medicine.

The above filter questions that included only a single follow up question were collapsed to form one item. Finally the survey also contained few questions that allowed indication of irrelevance or inability to assess the element of care as a response option. For example, in the questions 'Did

you get enough help from staff to eat your meals?' a majority of respondents answered 'I did not need any help'. I categorised these responses as neutral category on the assumption that when the question was asked they opted to not provide any assessment, which can be inferred as the middle ground. Also from the perspective of process and events carried out, the event did not take place and categorising it positively would not be accurate depiction. Another approach would have been to collapse the response categories. From a disutility perspective, 'I did not need any help' could be inferred as 'I received enough help'. However when a majority of patients indicate irrelevance or inability, collapsing categories with positive response creates a skew in the distribution of responses and results in loss of information.

Both use of subgroups to address inapplicable questions and coding of inapplicable response as a neutral category enabled complete analysis of dataset. It also increases specificity of findings and will be discussed in more detail at the end of the chapter.

5.2.4 Missing data

Descriptive statistics were generated to examine missing responses in the dataset. The missingness in a dataset can be categorised as 'missing completely at random', 'missing at random' and 'missing not at random'. To examine this, I created a dummy variable to categorise respondents with and without missing responses and applied logistic regression analysis to check if the background variables and overall rating between the two groups differed.

There are two ways of handling missing responses: deletion, which includes complete case analysis and pair wise deletion, and imputation, which comprises single imputation techniques and model-based method such as multiple imputations (Faria et al., 2014, Roderick, 1988). Multiple imputation assumes the data are missing at random, that is the probability of missingness does not depend on the unobserved value of the missing variable, but it can depend on any of the other variables in the dataset e.g. age or sex. Given the ordinal nature of the dataset, I decided to discard imputation and focus on complete case analysis.

5.2.5 Multivariate regression analysis

The dependent variable in my analysis was overall patient experience indicated by the respondents using a global rating scale. It is possible to treat a variable with 11 categories as continuous and apply standard linear regression. This approach assumes that the points between each of the scale levels are equidistant or that the difference between successive levels of patient experience is equal. I used ordered regression analysis instead as it allows ordering of responses but does not assume equidistance between response levels. This approach is closer to the true nature of the dataset. I applied clustering by trust to all multivariate analysis to reflect the data structure.

A majority of respondents evaluated their patient experience as good or very good. In order to generate a less skewed distribution of the outcome variable, different categorisations of the variable were explored by collapsing categories based on percentage. Dependent variable consisting of binary response and ordered levels were examined by fitting logit and probit distribution to determine the model with best fit and interpretation. I chose ordered logit model to analyse the dataset, and is the focus here.

All healthcare events and processes during inpatient stay were included in the analyses. The background questions used in the dataset (length of stay, age and gender) and nominal questions such as whether the patient was in critical care, pain and they self-completed the questionnaire were used as covariates. Bivariate analyses were carried out before proceeding to multivariate analyses.

An ordered logit model describes an indirect relationship between patient experience (y_i) and the healthcare elements (x_i). A latent continuous variable γ_i is assumed and it is described as a function of observed and unobserved variables as presented in equation 1; with vector x_i representing the set of explanatory variables and β being the vector coefficient of estimated parameters; ε_i was the error term with its mean and variance normalised to zero and one (Greene and Hensher, 2010). The ordered logit model assumes the relationship:

$\boldsymbol{\gamma_i = \beta x_i + \varepsilon_i}$, *i* = 1, … , n **… Equation 1**

The latent regression model described an underlying continuous but unobservable, preference for overall patient experience γ_i. It was assumed that the respondent in the survey does not provide overall patient experience y_i^*, but rather a censoring of γ_i into different ranges by indicating an ordinal category closest to their own true preferences (Greene and Hensher, 2010). The probability model has to have the error term specified. Two non-linear parametric specifications of the error (ε) term are logit and probit. I fitted logit distribution in my analysis. Both logit and probit have an 'S' shape distribution and are similar in appearance although the logit model gives more weight to the tails of the distribution or logit curve approaches the axes slower than the probit curve. Assuming distribution of ε as cumulative standard normal distribution gives an

ordered probit model and assuming ε to have cumulative standard logistic distribution (mean zero, variance $\pi^2/3$) gives an ordered logit model. The logit model was estimated by the method of maximum likelihood estimation (MLE).

The threshold parameters, denoted here as π, partitions the latent variable into a series of regions corresponding to the various ordinal categories as shown in Figure 3. The threshold values π_j is unknown, as the value of the index necessary to push from one level of patient experience to next is unknown and the threshold values differs from one person to another. Let $\pi_{i(j)}(f)$ be the probability that , given f, a response falls in category j for variable i. The probability of the response categories can be described as following:

Figure 3: Probability of the response category

Categories	0	1				
Response probabilities	$1-\pi_i(f)$	$\pi_i(f)$				
Categories	1	2	…	j	…	m_i
Response probabilities	$\pi_{i(1)}(f)$	$\pi_{i(2)}(f)$	…	$\pi_{i(j)}(f)$	…	$\pi_{i(m_i)}(f)$

The response probabilities of the categories will sum to one, regardless of the number of response categories. If an item is binary, the response in category one would be a linear function of the fs and described as $1-\pi_{i(1)}(f)$. However if there are more than two categories, the approach taken is to divide the categories into two groups with categories $(1,2,\dots,j)$ in one group and $(\mathrm{j}+1,\mathrm{j}+2,\dots,m_i)$ in the other group; and to report into which of the two groups the response fell. This reduces the polytomous variable to a binary variable.

And the probability of the response falling into the first and second groups respectively can be written as following:

$$\boldsymbol{\gamma_{i(j)}(f) = \Pr(x_i \leq j) = \pi_{i(1)}(f) + \pi_{i(2)}(f) + \ldots + \pi_{i(j)}(f)} \quad \textbf{... Equation 2}$$

and

$$\boldsymbol{1 - \gamma_{i(j)}(f) = \Pr(x_i > j) = \pi_{i(s+1)}(f) + \pi_{i(s+2)}(f) + \ldots + \pi_{i(m_i)}(f)} \quad \textbf{... Equation 3}$$

Where, x_i denotes the category into which the ith variable falls. The probabilities $\gamma_{i(j)}(f)$ are referred to as cumulative response probabilities. It assumes that the binary logit model holds for all divisions of the m_i categories into two groups. The model can be written in terms of logit, fitted probabilities are obtained by using the inverse logit transformation

$$\boldsymbol{\gamma_i = \frac{\exp^{\beta x_i}}{1+\exp^{\beta x_i}}} \quad \textbf{... Equation 4}$$

The cumulative probability for the category would then be expressed as:

$$\boldsymbol{\mathbf{Prob}\,[\gamma_i \leq j|\, x_i] = \text{Prob}\left[\varepsilon \leq \mu_j - \beta' x_i\right] = \frac{\exp(\mu_j - \beta' x_i)}{1+\exp(\mu_j - \beta' x_i)}} \quad \textbf{... Equation 5}$$

The ordered logit responses are considered linear in β'

$$\boldsymbol{\log[\gamma_i \leq j|\, x_i] = \mu_j - \beta' x_i} \quad \textbf{... Equation 6}$$

The ordered logit model is also known as the proportional odds model because the parallel regression assumption implies the proportionality of the odds of not exceeding the j-th category

$$\boldsymbol{\mathbf{odds}[\gamma_i \leq j|\, x_i] = \frac{\gamma_{i(j)}(f)}{1-\gamma_{i(j)}(f)}} \quad \textbf{... Equation 7}$$

This implies that the coefficients for each category in the explanatory variable must be equal across all levels of dependent variable i.e. the coefficients that describe the relationship between, say, the lowest versus all higher categories of the response variable are the same as those that describe the relationship between the next lowest category and all higher categories.

The proportional odds assumption or assumption of parallel regression is tested using Brant test. If the proportional odds assumption is violated, the model can be re-estimated by collapsing adjacent categories of dependent variable to improve. However if the model still has non-proportional odds, generalised ordered logit or multinomial logit can be considered. On estimation of the model, I employed several assessments on the appropriateness, adequacy and usefulness of the model. These can be described in three stages: 1) determining the importance of each independent variable in the model using statistical tests of the significance of the coefficients; 2) testing the overall goodness of the fit of the model and the ability of the model to discriminate between the categories of dependent variable 3) finally, if possible, validating the model by checking the goodness of fit and discrimination on a different set of data from that which was used to develop the model (Bewick et al., 2005). Note that validation was not carried out in this study.

Wald statistics and likelihood ratio are used to test the significance of individual coefficients in the model. In the Wald test, the null hypothesis states that the coefficients of interest are simultaneously equal to zero. If the test fails to reject the null hypothesis the variables are excluded. The Likelihood Ratio test for a particular variable compares the likelihood of obtaining the data when the parameter is zero with the likelihood of

obtaining the data evaluated at the MLE of the parameter. I used the Wald test to check the significance of particular parameters before exclusion from the model and the Likelihood Ratio test to compare the fit of models for example reduced versus full and across subgroups. The goodness of fit tests measure how well the model describes the response variable, or how close the values predicted by the model are to the observed values. In order to assess the fit of the predictions by the model to the observed data, compared to no model, the overall model chi squared can be used (usually reported as "pseudo R^2").

Another commonly used test for assessing the goodness of fit of a model is Hosmer–Lemeshow test. The test is similar to a chi-squared goodness of fit test with additional advantage of partitioning the observations into groups of approximately equal size and therefore there are less likely to be groups with very low observed and expected frequencies (Greene and Hensher, 2010). The model specification was tested using the Regression Equation Specification Error Test (RESET) based on the idea that if a model is properly specified, one will not find any additional independent variables that are significant except by chance. It indicates omission of important variables. I applied both Hosmer-Lomeshow test and RESET test to check model fit.

5.3 Results

5.3.1 Descriptive statistics

In the Inpatient Survey 2014, over 62,000 respondents assessed nearly seventy aspects of inpatient stay. A total of 40% of completed questionnaires in the 2014 Inpatient Survey had a missing response for at least one or more variables and the remaining 34,976 respondents had a

complete set. Also examination using logistic regression found the group of respondents with any missing response was statistically significantly different than the complete case sample in terms of sex, age and patient experience rating. This is not surprising given the sample size, as any slight difference would be statistically significant. However the differences were not substantial.

I attempted imputation of missing responses using a chained predictive mean matching (PMM) technique that was conditional on four baseline variables (age, sex, patient experience rating and NHS trust). However, when distribution of imputed values was checked with the observed data for comparability it was found that additional response levels imputed were non-integer (for example, additional categories such as 1.5 and 2.5 were created) and created additional response categories. Rounding the value would create further bias. It is possible to continue regression analysis without rounding to whole number but this would create problems in conducting factor analysis in subsequent chapters. I decided to exclude respondents with missing values from the analyses.

Bivariate analysis was carried out to examine independence of explanatory variables from dependent variable. I checked for correlation using Spearman's rank correlation coefficient and rho significance level (Appendix 6). The coefficients indicated mild or moderate correlation with the dependent variable barring Q67 on being treated with respect and dignity, which was quite high at around 0.6 across all subgroups. All items assessing processes and events were found to be associated with overall rating of patient experience. I conducted multivariate analysis using the four subgroups: 1) emergency admission with operation 2) emergency

admission without operation 3) planned admission with operation and 4) planned admission without operation.

5.3.2 Ordered logistic regression models

I used ordered logit models with overall patient rating as outcome variable for all four-subgroups. The parallel lines assumption was violated, regardless of categorisation of the outcome variable. This means that the ordered logit coefficients were not equal across the levels of the outcome, and this could be due to large number of exploratory variables considered. The full model examining A&E patients with operation or procedure consisted of 60 independent variables; the model examining A&E patients without operation or procedure included 54 explanatory variables; the model examining planned admission with operation or procedure included 62 explanatory; finally the full model of patients with planned admission and without operation or procedure included 57 independent variables. The full models failed the specification test. I suspected it was due to large number of explanatory variable, some of which were likely to be irrelevant. The full ordered logit models fitted to the subgroups are presented in Appendix 7, Appendix 8, Appendix 9 and Appendix 10.

Reduced models were estimated to include only statistically significant variables in each subgroup model. A summary of the variables, which were most significant in the models across the four subgroups, is presented in table 4. When a coefficient of a logit model is positive, this translates into an odds ratio of greater than one and indicates that an increase in the independent variable by one unit, ceteris paribus, is associated with increased odds of observing a higher category of Y (i.e. poor patient experience) and vice versa. For example, the odds of patients reporting poor overall experience is nearly four times larger in patients not treated

with respect and dignity compared to patients always treated with respect and dignity.

When a coefficient is negative, the predicted odds ratio is less than one indicating the odds of reporting poor overall experience is lower than the reference category. Only two items in the dataset had odds ratio less than one. These were privacy when being examined or treated and member of staff answering patients' questions about the operation or procedure. Note that the distributions of responses in these two questions were skewed with only a very small proportion of respondents answering 'no'. All the items in the survey had positive coefficient apart from two-control variables and these were pain and age. Across all subgroups, patients who experienced pain were more likely to report poor overall experience than patients who did not experience pain during the hospital stay.

Table 4: Ordered logit model across subgroups

		Emergency With Operation	Emergency without Operation	Planned with Operation	Planned without Operation
		Odds ratio	Odds ratio	Odds ratio	Odds ratio
Q4	Were you given enough privacy when being examined or treated in the A&E Department?	1.17	1.12		
Q6	How do you feel about the length of time you were on the waiting list before your admission to hospital?			1.15	
Q9	From the time you arrived at the hospital, did you feel that you had to wait a long time to get to a bed on a ward?	1.32	1.36	1.31	1.22
Q14	While staying in hospital, did you ever use the same bathroom or shower area as patients of the opposite sex?		1.04		
Q15	Were you ever bothered by noise at night from other patients?	1.19	1.22	1.27	
Q16	Were you ever bothered by noise at night from hospital staff?			1.22	1.92
Q17	In your opinion, how clean was the hospital room or ward that you were in?	1.51	1.6	1.63	
Q18	How clean were the toilets and bathrooms that you used in hospital?	1.13	1.12	1.1	
Q19	Did you feel threatened during your stay in hospital by other patients or visitors?	1.38			
Q21	How would you rate the hospital food?	1.3	1.3	1.31	1.38
Q24	When you had important questions to ask a doctor, did you get answers that you could understand?	1.09			
Q25	Did you have confidence and trust in the doctors treating you?	1.37	1.54	1.63	1.69
Q27	When you had important questions to ask a nurse, did you get answers that you could understand?	1.09		1.07	
Q28	Did you have confidence and trust in the nurses treating you?	1.7	1.74	1.68	1.63
Q29	Did nurses talk in front of you as if you weren't there?			1.09	

		Emergency With Operation	Emergency without Operation	Planned with Operation	Planned without Operation
Q30	In your opinion, were there enough nurses on duty to care for you in hospital?	1.34	1.25	1.4	1.42
Q31	Sometimes in a hospital, a member of staff will say one thing and another will say something quite different. Did this happen to you?	1.28	1.24	1.29	1.35
Q32	Were you involved as much as you wanted to be in decisions about your care and treatment?	1.33	1.36	1.55	1.68
Q33	How much information about your condition or treatment was given to you?	1.33	1.35	1.18	1.22
Q35	Do you feel you got enough emotional support from hospital staff during your stay?	1.14	1.15	1.13	
Q36	Were you given enough privacy when discussing your condition or treatment?			1.1	
Q37	Were you given enough privacy when being examined or treated?			0.85	
Q39	Do you think the hospital staff did everything they could to help control your pain?	1.2	1.18	1.21	
Q40	How many minutes after you used the call button did it usually take before you got the help you needed?	1.06		1.07	
Q44	Beforehand, did a member of staff answer your questions about the operation or procedure?			0.93	
Q45	Beforehand, were you told how you could expect to feel after you had the operation or procedure?	1.11		1.22	
Q48	After the operation or procedure, did a member of staff explain how it had gone in a way you could understand?	1.07		1.2	
Q49	Did you feel you were involved in decisions about your discharge from hospital?	1.11	1.1	1.08	
Q50	Were you given enough notice about when you were going to be	1.25	1.24	1.3	1.43

		Emergency With Operation	Emergency without Operation	Planned with Operation	Planned without Operation
	discharged?				
Q51	How long was the delay in discharge?	1.07	1.05	1.09	
Q55	Did a member of staff explain the purpose of the medicines you were to take at home in a way you could understand?		1.05		
Q59	Did a member of staff tell you about any danger signals you should watch for after you went home?	1.07	1.07	1.06	
Q61	Did the doctors or nurses give your family or someone close to you all the information they needed to help care for you?		1.08	1.1	1.1
Q62	Did hospital staff tell you whom to contact if you were worried about your condition or treatment after you left hospital?	1.07	1.08	1.09	
Q66	Were the letters sent between hospital doctors and your family doctor (GP) written in a way that you could understand?	1.05	1.07	1.06	
Q67	Overall, did you feel you were treated with respect and dignity while you were in the hospital?	3.98	4.2	4.26	5.04
Q69	During your hospital stay, were you ever asked to give your views on the quality of your care?	1.13	1.13	1.09	
Q70	Did you see, or were you given, any information explaining how to complain to the hospital about the care you received?	1.15	1.06	1.08	1.22
	Critical care stay	1.26	1.24	1.11	
	Pain	0.64	0.7	0.6	
	Patient completed questionnaire		1.35	1.34	1.44
	Age group		0.89		
	Length of stay	1.01			
	Fit of the model				
	Number of observations	9,665	12,964	17,414	3,411
	Wald chi2	7270	8543	12791	1652
	Prob > chi2	0	0	0	0

		Emergency With Operation	Emergency without Operation	Planned with Operation	Planned without Operation
	Pseudo R2	0.26	0.25	0.24	0.23
	Log likelihood	-13600	-19184	-22067	-4823
	Number of obs	9,665	12964	17414	3411
	Wald chi2	7270	8544	12792	1652
	RESET	Failed	Failed	Failed	Failed

The item considered most important across all subgroups was whether the patient felt that they were treated with respect and dignity in the hospital. This was followed by whether the patient had confidence and trust in the nurses treating them and the cleanliness of the ward in most cases. Other items that were considered important were confidence and trust in the doctors, being involved in decision making and having enough nurses on duty. It should be noted that the model on patients with planned admission and without any operation or procedure was slightly different to the other models. Being bothered by noise at night from hospital staff and being given enough notice about when they are going to be discharged were featured as very important considerations towards their patients' experience. The reason for this is that these patients primarily comprised of older patients who could not complete the questionnaire without assistance, required care services and not necessarily had huge clinical needs as they did not have any operation or procedure carried out.

Most of the variables that were included as statistically significant in the models, and those that were not were consistent with my expectation. However the aspect of the patient feeling threatened during their stay in hospital by other patients or visitors, which came up only for patients with emergency admission that underwent an operation or procedure was peculiar and difficult to explain. It is possible that patients with emergency admission felt threatened by other patients because there are sometimes possibly violent drunks, or their visitors, in A&E. However it is difficult to understand why those who went to A&E but did not have an operation didn't feel threatened unlike those who did have an operation.

It should also be noted here that the items could be grouped together based on the area being addressed or there were themes. For example, the

variable being treated with dignity and respect is very closely related to receiving emotional support from healthcare staff and being given enough privacy. All of these items are a manifestation of the attitude adopted by the healthcare professionals rather than the efficacy of the treatment. Similarly confidence and trust in staff was broken down to look specifically at trust in doctors and in nurses separately. Also noise and cleanliness are related to physical comfort of the patient. Again while the information obtained is quite specific, the underlying construct is common across the items. This concept will be examined in detail in next chapter.

Table 5: Ten most important variables from the ordered models

Emergency With Operation	Emergency without Operation	Planned with Operation	Planned without Operation
Overall, did you feel you were treated with respect and dignity while you were in the hospital?	Overall, did you feel you were treated with respect and dignity while you were in the hospital?	Overall, did you feel you were treated with respect and dignity while you were in the hospital?	Overall, did you feel you were treated with respect and dignity while you were in the hospital?
Did you have confidence and trust in the nurses treating you?	Did you have confidence and trust in the nurses treating you?	Did you have confidence and trust in the nurses treating you?	Were you ever bothered by noise at night from hospital staff?
In your opinion, how clean was the hospital room or ward that you were in?	In your opinion, how clean was the hospital room or ward that you were in?	In your opinion, how clean was the hospital room or ward that you were in?	Did you have confidence and trust in the doctors treating you?
Did you feel threatened during your stay in hospital by other patients or visitors?	Did you have confidence and trust in the doctors treating you?	Did you have confidence and trust in the doctors treating you?	Were you involved as much as you wanted to be in decisions about your care and treatment?
Did you have confidence and trust in the doctors treating you?	From the time you arrived at the hospital, did you feel that you had to wait a long time	Were you involved as much as you wanted to be in decisions about your care and	Did you have confidence and trust in the nurses treating you?

Emergency With Operation	Emergency without Operation	Planned with Operation	Planned without Operation
	to get to a bed on a ward?	treatment?	
In your opinion, were there enough nurses on duty to care for you in hospital?	Were you involved as much as you wanted to be in decisions about your care and treatment?	In your opinion, were there enough nurses on duty to care for you in hospital?	Were you given enough notice about when you were going to be discharged?
Were you involved as much as you wanted to be in decisions about your care and treatment?	How much information about your condition or treatment was given to you?	From the time you arrived at the hospital, did you feel that you had to wait a long time to get to a bed on a ward?	In your opinion, were there enough nurses on duty to care for you in hospital?
How much information about your condition or treatment was given to you?	How would you rate the hospital food?	How would you rate the hospital food?	How would you rate the hospital food?
From the time you arrived at the hospital, did you feel that you had to wait a long time to get to a bed on a ward?	In your opinion, were there enough nurses on duty to care for you in hospital?	Were you given enough notice about when you were going to be discharged?	Sometimes in a hospital, a member of staff will say one thing and another will say something quite different. Did this happen to you?
How would you rate the hospital food?	Sometimes in a hospital, a member of staff will say one thing and another will say something quite different. Did this happen to you?	Sometimes in a hospital, a member of staff will say one thing and another will say something quite different. Did this happen to you?	From the time you arrived at the hospital, did you feel that you had to wait a long time to get to a bed on a ward?

5.4 Discussion

The Inpatient survey consists of a large number of items and levels, and it is not amenable to quantification using preferences. The first aim in this

study was to familiarise the readers with the Inpatient Survey 2014 dataset, which forms the core dataset of my thesis. The second aim was to estimate ordered regression models to identify variables that are able to estimate the probability of reporting poor overall patient experience, with the view of item selection for patient experience measure. In terms of key findings the ordered logit regression models I fitted in the dataset were able to identify most salient items in patient experience. Owing to the nature of the dataset and need to use subgroups, four models were developed instead of one all-encompassing measure for inpatient stay.

A limitation of the analysis was use of complete case analysis. It should be noted that the data was not missing in random and multiple imputation was not possible. Nevertheless the sample included for regression analysis contained nearly 35,000 respondents. A detailed discussion about key methods, findings and comparison to existing literature of the study is presented below.

5.4.1 Use of subgroups

Given the large number of inapplicable questions, I divided the data into four subgroups. This allowed all relevant variables to be included in the dataset according to each subgroup, increasing completeness as well as specificity of the models. Important variables that may influence patient experience were not excluded for being relevant to all and similarly 'inapplicable' response was retained as neutral category where possible to avoid loss of information. The disadvantage of the approach I took is that it results in four models and not one all-encompassing model of patient experience, which is applicable to all. In the next chapter I have focussed on patients with an operation or procedure (via both planned and A&E admission).

Regression models were fitted to complete case data and standard error was adjusted by cluster of trusts, for following four subgroups: 1) A&E admission with operation, 2) planned admission with operation, 3) A&E admission without operation and 4) planned admission without operation. The factors commonly identified as most strongly associated with overall patient experience across all subgroups were: being treated with respect and dignity, trust and confidence in the nurses, cleanliness of the hospital room/ward and trust and confidence in doctors treating the patient. Of the control variables, the item assessing pain experienced was statistically significant across all subgroups. The patients who reported no pain were more likely to report good overall experience than those who experienced pain.

5.4.2 Selection of dependent variable

The overall satisfaction question in this survey was at the end of the questionnaire after the patient has been asked to provide views on all elements of hospital stay corresponding to usual practice of care in hospital and it is most likely influenced by thinking about all those aspects of care in an objective manner. Alternatively the patient experience score used by the CQC could have been used which aggregates responses to all the questions in the survey and computes a composite score. The overall score incorporates assessment of each item in the survey. However, the weight assigned to item responses is subjective and one would have to make assumptions about how a patient differentiates between item levels. Also the purpose of the CQC was perhaps to summarise the responses for monitoring performance of trusts but the focus here is to understand patient experience and I opted to use self-assessed rating of the patients.

5.4.3 Use of ordered logit models

I examined binary dependent variable, in which I collapsed outcome variable into two categories to estimate Y=0 vs. Y=1, to ensure an equal distribution of the outcome variable but this did not improve model fit. I presented ordered logit models instead as they are better able to capture the richness in the dataset. Proportionality of odds assumption was violated in the models estimated. It is possible to use a multinomial logit model that does not require such an assumption but it does not preserve the inherent ordering of the rating scale and results in a loss in the efficiency of the estimators. Alternatively generalized ordered logit model preserves the ordering however it is very sensitive to low frequency counts and would require adjacent categories to be collapsed; also having a large number of explanatory variables make it cumbersome in terms of interpretation.

Step-wise regression using backward or forward approach could have been conducted using automated procedure or manually to exclude variables till a parsimonious model was generated. However this method relies heavily on the R^2 value, amongst other limitations, and is not encouraged in the literature (Babyak, 2004, Thompson, 1995). Another method to derive a parsimonious model is the least angle regression (LARs) (Efron et al., 2004). However it estimates linear regression, which may not be appropriate with the inpatient dataset, and secondly variable selection appears to have problems with highly correlated variables. Similar to LARs is the lasso (least absolute shrinkage and selection operator) regression analysis method. The approach I took in this study was to conduct full regression models for each subgroup but present reduced models including variables that were found most statistically significant only.

5.4.4 Item selection using regression analysis

Instrument developers (e.g. SF-12 by Ware et al. 1996) have conducted regression analysis for item selection, but this method is not advocated in current guidance or standards for development of a descriptive system. Nevertheless I adopted the method because it was intuitive to examine explanatory variables related to overall patience experience using regression analysis and select the most salient variables based on it. However on completion of this study, I think that it is better to have a wide ranging items that capture the full questionnaire (Inpatient survey questionnaire) in its entirety rather than selecting only those statistically significantly related to overall rating.

The final probability models resulted in a reduced number of items (25, 29, 32 and 14 items) for the four subgroups selected as statistically significant. However having so many items will make the instrument very difficult to value. Eliciting values or preferences is important because it allow different dimensions and levels of health care to be explicitly differentiated by target population group, for example being treated with dignity compared to cleanliness of the ward or different levels of cleanliness. In order to create a value set, it is necessary to generate health care states that are a combination of the specific level of each item. For example, the EQ-5D-3L, which has 5 dimensions with 1 item each and 3 levels on each item, generates 3^5 = 243 unique health states; when the number of levels for each dimension was expanded from 3 to 5 levels, the number of unique health states increased to 3125. Adding an additional dimension (e.g. EQ-5D bolt on) to EQ-5D-3L increases the number of possible health states to 729. Similarly adding a sixth dimension to EQ-5D-5L will increase number of health states to 15,625. The higher the number of items in the descriptive system, the larger the profile combining a level from each item and the larger the number of possible discrete states. The increased information

burden on the respondent and those undertaking preference elicitations will find valuation very challenging. For an instrument to be amenable to valuation, it is crucial that it is brief.

One may argue that a regression analysis employing a higher p-value threshold or another method for step-wise reduction could have been used. However the Reflect on the use of regression analysis for item selection should be avoided. The items selected using regression models are not 'unidimensional'. In other words they may be capturing the same latent concept and an overlap of items may cause problem in scoring and valuation of the instrument.

The underlying construct of the dataset across of four subgroups is examined in next chapter by employing factor analysis techniques. Grouping together similar items to derive dimensions or domains or scales from the dataset before applying item selection is a more efficient way to estimate a descriptive system. This staged approach is also the most commonly used method of instrument development based on the review conducted earlier. In the next chapter I employ factor analytic techniques to identify and define a small number of separate factors (dimensions) that make up a patient experience measure. It will describe how the items contained in each dimension group together and reduce a large number of inter-related observations to a smaller number of common dimensions.

5.4.5 Item associated with good patient experience rating

The regression analyses I carried out indicated that the most important aspect of healthcare delivery is respect and dignity. The probability of a very good patient experience rating increased manifold when they were

treated with respect and dignity, and vice versa. The odds ratio of this variable ranged from 3 to 5, and it was statistically significant across all subgroups. This was followed by confidence and trust in the nurses treating the patients and the cleanliness of the ward in most of the subgroups.

While there is no study that has specifically examined items most closely associated with patient experience, there is one study which used regression method and applied it in the NHS inpatient dataset but focussed on dimensions (domains) rather than specific questions (Sizmur and Redding, 2009). In the study a mean score was assigned to the seven dimensions of care based on the Picker framework of patient experience, by grouping together items a priori and scoring each dimension. The multivariate linear regression analysis conducted in this study, using overall rating of care as the dependent variables, found that physical comfort, emotional support and respect for patient preferences were most strongly and significantly associated with overall patient experience.

The approach adopted by Sizmur and Redding is simple and the findings are neat. It assumes that patient experience has seven dimensions of care, it implicitly assumes that if a question is not applicable to a patient then it should be scored as no problem (note that CQC scoring guide does the same). The focus here is on relative importance of these dimensions and a linear regression model examines this. While the information on importance of dimensions is important, the dimensions still need to be expressed in explicit items for it to be used in a preference-based study. In my study I have accepted the complexities of the dataset and employed two key techniques to address it, these are use of subgroups and ordered

regression. Also I have not made any assumptions about the dimensions in the dataset. Chapter 6 of my thesis will examine data dimensionality.

In a study by Jenkinson et al. a total of fifteen items were identified from the Picker inpatient questionnaires based on analysis of survey data of patients who had attended acute care hospitals in five European countries (Jenkinson et al., 2002b). The items were selected based on face validity, internal consistency reliability and high correlation with parent instrument. One of the inclusion criteria in the study was that items were applicable to as many respondents as possible (e.g. questions on emergency admissions will not be applicable to in-patients who had planned admissions). Items not applicable to a large proportion of respondents were excluded. While this criterion enables development of a generic instrument, it is possible that this approach excludes items that are important. For example, how much information was given about your condition or treatment while you were in the A&E department is not applicable for those who had planned admission but will have a significant impact on patient experience of an A&E patient.

Figure 4: Responsiveness domains and questions used

Dignity • being shown respect • having physical examination conducted in privacy	Autonomy • being involved in deciding on your care or treatment if you want to • having the provider ask your permission before starting treatments or tests
Confidentiality of information • having your medical history kept confidential • having talks with health providers done so that other people who you don't want to have hear you can't overhear you	Surroundings or environment • having enough space, seating and fresh air in the • having a clean facility (including clean toilets) • having healthy and edible food
Choice • being able to choose your doctor or nurse or other person usually providing your health care • being able to go to another place for health care if you want to	Social support • being allowed the provision of food and other gifts by relatives while in hospital • being allowed freedom of religious practices
Prompt attention • having a reasonable distance and travel time from your home to the healthcare provider • getting fast care in emergencies • short waiting time for appointments and consultations, and getting tests done quickly • short waiting list for non-emergency surgery	Communication • having the provider listen to you carefully • having the provider explain things so you can understand • having time to ask questions

Note: Adapted from figure 1 in the WHO MCS Study (Valentine et al., 2008)

A study by WHO, conducted within the Multi-Country Survey Study on Health and Health Systems Responsiveness (the ''MCS Study''), examined the relative importance of eight dimensions across 41 countries and a total of 105,806 respondents completed the responsiveness questionnaire (Valentine et al., 2008). The domain most frequently selected as important was prompt attention, followed by dignity. Dignity comprised of being shown respect and having physical examination conducted in privacy.

The Inpatient survey included separate questions on 'dignity and respect' and 'privacy during examination'. In my study, privacy was a concern only for patients with planned admission that underwent an operation or procedure. The reason behind this could be that patients are not concerned about privacy when they require urgent care or when they are in admitted in the hospital without having to undergo any operation or procedure. Alternative explanation could be that privacy is already being sufficiently addressed in the other subgroups and is not a consideration which evaluating overall patient experience.

While there has been many studies to investigate the importance of the dimensions in an inpatient stay or generally, there has not been any study to examine items related to overall patient experience rating using regression analysis. Note that later in this thesis (chapter 7) I present item selection using IRT and factor analysis in the Inpatient survey dataset (2014). A direct comparison of items selected using different methods can be found there.

5.5 Conclusion

The Inpatient Survey provides a comprehensive assessment of healthcare delivery from a patient perspective. It is part of NHS's National Patient Survey Programme and the CQC, the national regulator in the UK, reports results for each trust every trust from the Inpatient survey. The survey was developed based on extensive research and provides important information in terms of monitoring of NHS trusts and improving quality of care. However the survey is too lengthy to be used in a preference elicitation study.

In this chapter I focussed on identifying items associated with overall patient experience indication using regression analysis. I used advanced methods to deal with the complexities of the large dataset and there are four models to explain patient experience. The findings from this study are nuanced and provide a unique insight into aspects of healthcare delivery that different patient groups value. The items identified in this study are however too many and too similar to be used directly in a descriptive system.

Chapter 6 Dimensions of patient experience

6.1 Introduction

In the previous chapter I used ordered logit regression models to identify healthcare events and processes most significantly associated with overall patient experience in the Inpatient Survey 2014 dataset for item selection. The regression models estimated for the four subgroups were able to identify the most statistically significant items in the four subgroups, but the item selection using this method was problematic. The descriptive system generated included items that overlapped conceptually. The items could perhaps have been grouped together to form dimensions. However dimensions generated in this way would not have represented the original instrument or the inpatient dataset, as the focus was on the relationship of the variables with the overall patient experience rating.

I now adopt a staged approach, in which dimensions are selected first and then items. It is consistent with instrument development guidelines and based on measurement theory. In this chapter, I focus on dimension selection. Factor analytic techniques are used to systematically identify and assess dimensions of patient experience using Inpatient Survey 2014. In the next chapter item selection will be carried out for each dimension. In order to enable a preference-based measure of patient experience, it is important that only one item is selected per dimension.

6.1.1 A multidimensional instrument

An instrument may have one or multiple dimensions. And a dimension may contain one or more items and these items are multilevel. Here I have considered patient experience to be multidimensional, i.e. including more

than one latent construct. A multidimensional measurement of patient experience is likely to be holistic and it is able to capture more information than an instrument measuring a single latent construct.

The information produced by a multidimensional measure can be summarised using a scoring that weights all levels and dimensions equally. However this approach does not allow items and item levels to be differentiated. A weighted score can be applied to overcome this issues and preferred method in the field of health outcomes is to develop a value set based on preferences we may have as a society. This is the approach used in the UK for HTA, see NICE reference case (NICE, 2013b).

As previously mentioned, while selecting dimensions for a descriptive system that is designed for valuation, it is important to have structural independence between dimensions to avoid nonsensical corner states (Dowie, 2002, Feeny, 2002). One technique for identifying structurally independent dimensions with little correlation between them is exploratory factor analysis (EFA) and this is the method applied in this chapter.

6.1.2 Determining number of dimensions

Another important consideration is the number of dimensions to include. In order to generate a value set, the dimensions and levels in the measure are combined together to generate unique profiles or states. For example the EQ-5D-3L has five dimensions with one item in each dimension and three levels per item and this creates a total of $3^5 = 243$ possible health states. A function can be estimated for valuing these states of the descriptive system. For example multi-attribute utility theory is employed

to exact the functional form and the sample of states to be valued in health (Brazier et al., 2007). Choice based methods such as time trade off and standard gamble have been employed to value health states of the EQ-5D and SF-36.

Individuals can only process between five and nine pieces of information at a time in a valuation task (Brazier et al., 2002). Hence an important criterion while developing an instrument that is amenable to valuation is the number of dimensions to include in the instrument. Also an instrument containing multiple dimensions (or items) with multiple levels or response categories will generate thousands of combinations for valuation. This improves the responsiveness of the instrument as it allows for a wider spectrum of states related to patient experience to be described. However, the valuation of a large descriptive system is very challenging in practical terms as a very large number of states will have to be valued to estimate the utility function across all possible combinations.

Brazier et al. (p 74) suggest instrument developers should aim for an instrument that is "amenable to valuation by respondents with a minimum loss of descriptive information and subject to constraints that responses to the original instrument can be unambiguously mapped into it" (Brazier et al., 2007). Determining the dimensions, items and levels of a descriptive system is crucial, and the selection should be done carefully. The dimensions estimated from EFA are automatically unidimensional or measuring a single latent construct.

This chapter focuses on deriving dimensions of patient experience based on the Inpatient dataset that is latent but interpretable using factor

analysis method. Factor analysis models are based on the premise that the correlated variables most likely have a common dependence on one or more variable. And factor analysis models are used to explain this common dependence amongst the observed variable.

Two types of factor analysis are applied here. Firstly the EFA is used to derive a smaller set of factors that represents the correlations in the dataset. Secondly a confirmatory factor analysis (CFA) framework is used to assess the relationship between the underlying unobserved variable and observed variables. CFA allows prior expectations about the data to be investigated and this was utilised by grouping together items as indicated by the EFA. The final decision about number of dimensions and its content are based on model adequacy or fit. Findings on overall fit of the model and relationship between dimensions and items are used to revise the number and content of dimensions.

6.2 Methods

The methods section first describes the dataset briefly followed by general factor analysis technique, estimation of factor models, interpretation of findings, goodness of fit and assessment of unidimensionality.

6.2.1 Dataset using subgroups

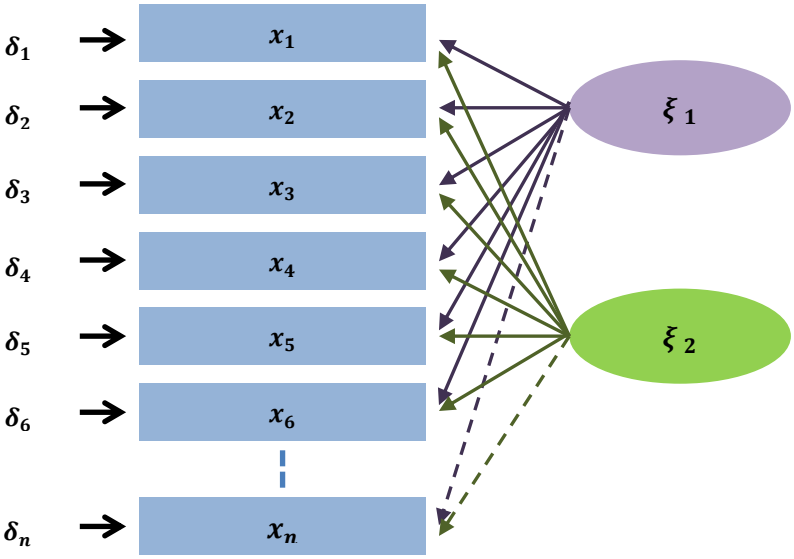
The Inpatient Survey 2014 dataset was used in this study. Based on the findings of the data structure, particularly inapplicable questions (see chapter five for detail), I continued factor analyses in the subgroups rather than the full dataset. Factor analyses were carried out in all four subgroups but this chapter will focus only on the two subgroups in which factor analytic models were fitted satisfactorily: these are patients who

underwent operation or procedure through emergency admission and planned admission. Similar to chapter 5, respondents with missing data were excluded. While it is possible that this approach may bias the results, imputation and list wise deletion were not adopted because the responses are ordinal in nature and sample sizes are large (>12,500 respondents in each subgroup).

6.2.2 Exploratory factor analysis

The key concept in factor analysis is that latent variables (ξ) can be represented by an observed or manifest variable (x). The diagram below (Figure 5) illustrates a simple EFA model with 2 factors.

Figure 5: An example of EFA



Key
ξ = latent variable (factor)
x = observed variable (item)
δ = measurement error
Arrow (path) = relationship between factor and item

Linear relations are postulated to hold between the factors (ξ) and observed variables (x) as expressed in model equation 4, where τ denotes

the regression intercept, λ denotes factor loading and δ is the measurement error (Bartholomew et al., 2008). The constants (τ) does not play a role in fitting the model and are often dispensed by assuming that observed variables (x) are measured about their means. The model is described using following equation, n is the item number and λ is the factor loading:

$$\mathbf{x_n = \tau_n + \lambda_n \xi + \delta_n} \qquad \textbf{... Equation 8}$$

The factor loadings denote the covariance between the latent variable and the observed variables (or correlations if the observed variables are standardised). The factor loadings indicate how much the variable has contributed to the factor and the larger the factor loading, the more the variable has contributed to that factor. Kline et al describe factor loadings as being akin to weights in multiple regression analysis, and they represent the strength of the correlation between the variable and the factor (Kline, 1994). The EFA generates a model of the covariance matrix of the observed x-variables, depending on a set of parameters. These are: 1) factor loadings 2) variances of the unique factors and 3) variances and covariances of the latent variables. The estimation of the parameters is based on identifying values for them such that the fitted covariance or correlation matrix is as close as possible to the sample covariance or correlation matrix of the observed variables.

6.2.3 Factor analysis model

There are different methods for model estimation based on the nature of the dataset, for example ordinary least squares is used to estimate regression containing continuous dependent variable. The Inpatient survey consists of non-linear data. To deal with such data, the literature suggests estimation of thresholds, means and variances by maximum likelihood

method; estimation of polychoric correlations by conditional maximum likelihood and estimation of parameters for the structural part of the model using a generalised least squares estimation (Bartholomew et al., 2008, Muthén, 1984). And I adopted this approach to estimate factor analysis models.

Extracting the smallest number of factors that explains the largest amount of variation (among the observed variables) involves checking the 'eigenvalue'. The eigenvalue for each factor indicates how much variance in the observed indicators is being explained by that latent factor. Because factor analysis is carried out to attain a parsimonious model, only those latent factors with sufficiently high eigenvalues are included in the model. The most common practice is to retain factors that have eigenvalues above 1, however it is not a rigid rule and decision can be based on interpretability rather than just statistical value.

An alternative method of determining the appropriate number of factors to retain is to consider the relative size of the eigenvalues rather than the absolute size and inspecting a scree plot can do this. A scree plot shows the eigenvalues on the y-axis and the number of factors on the x-axis. It always displays a downward curve and the point where the slope of the curve is levelling off or the "elbow" indicates the number of factors that should be generated by the analysis. Nevertheless determining the appropriate number of factors to retain by inspecting a scree plot is subjective and open to different interpretations.

Similar to the previous chapter in which ordinal regression analyses was conducted; factor analysis models assume that each ordinal variable x_1 is

generated by an underlying unobserved variable x_i^*. The EFA model is now implied on the x^* variable:

$$\boldsymbol{x_i^* = \lambda_{i1}\xi_1 + \lambda_{i2}\xi_2 + \ldots + \lambda_{iq}\xi_q + \delta_i,\ \ i = 1, 2, \ldots, p} \quad \text{... Equation 9}$$

Since only ordinal information is available for x_i^*, the mean and variance of x_i^* are not identified and are set to zero and one respectively. The latent factors (ξ_1,..., ξ_q) and errors (δ_1,..., δ_p) are independent and assumed to be normally distributed.

The underlying unobserved variable $x_i^*, \ldots, x_p^*$ has a multivariate normal distribution with zero means, unit variances and correlation matrix $P = (\rho_{ij})$, where:

$$\boldsymbol{\rho_{ij} = \sum_{l=1}^{q} \lambda_{il}\lambda_{jl}} \quad \textbf{... Equation 10}$$

The parameters of the model are the thresholds:

$$\boldsymbol{\tau_a^{(i)}, i = 1, 2, \ldots, p,\ \ a = 1, 2, \ldots, m_{i-1}} \quad \text{... Equation 11}$$

And the estimated factor loadings or covariance between latent variables and observed variables is presented as following:

$$\boldsymbol{\lambda_{ij}, i = 1, 2, \ldots, p,\ \ j = 1, 2, \ldots, q} \quad \textbf{... Equation 12}$$

Dimensions generated using EFA are automatically unidimensional. The percentage of total variance explained by the first factor in an EFA is often regarded as an index of unidimensionality. The interpretation of dimension or factor is based on items that have a large (positive or negative) loading

on that factor. It is also important to check the communality of the standardised observable variable which is the squared multiple correlation coefficient or the proportion of the variance explained by the common factors. The sum of communalities is the variance explained by the factor model. However when a communality greater than one is observed, it is considered a 'Heywood case' with estimated unique variance of 0 or less for observed variable. A possible solution is to fit the model using one less factor. The number of factors to include in a model depends on two factors: meaningful interpretation of factors and model fit.

Interpretation of loadings for multiple factors is difficult when loadings do not have a simple structure. Simple structures can sometimes be found by rotating the solution. Rotation does not alter the fit of the model or the communalities; it alters the loadings and changes the interpretation of factors. There are two methods for selection of rotated solutions: orthogonal rotation (e.g. varimax method), which keeps the axes orthogonal; and oblique rotation (e.g. oblimin, geomin), which allows the new axes to be non-orthogonal or factors to be correlated. I opted for orthogonal rotation, as it is preferable to have uncorrelated factors in any descriptive system to enable subsequent quantification without double counting or overlap.

6.2.4 Measurement model assessed using CFA

The measurement relation between the latent factor (ξ) and the items or indicators and between the factors is called the measurement model. Confirmatory factor analysis (CFA) framework was used to identify the association between the items and the dimensions, and between dimensions identified by EFA. This helps to remove arbitrariness of using an EFA, which does not use any prior theory.

A good measurement model consists of unidimensional dimensions, which ensure each dimension is measuring a distinct construct, and there is no conceptual overlap. Secondly that the constructs are still sufficiently related to each other and the combinations of the dimensions do not produce incoherent profiles. Having items and dimensions that fit the conceptual framework will also facilitate valuation of the instrument in the future as the combinations or states generated will be cohesive.

Item loadings were assessed and revised using CFA findings. CFA postulates constraints to the model and helps to determine whether the hypothesised model is identified by the data. Within the CFA framework, I only allowed the items identified in EFA to be indicators of latent variables. Also the scale of the latent variable was defined and the error terms were assumed to be uncorrelated with each other. The EFA model was revised by omitting poorly associated dimensions and items until good model fit is obtained using the confirmatory framework. A good measurement model can also be used to establish construct validity of the descriptive system. Note that more information on how the CFA is estimated is presented in chapter seven, where it is applied for item selection as opposed to confirming hypothetical construct of the dataset.

6.2.5 Goodness of fit

It is critical that the subset or the dimensions estimated is able to provide an approximate representation of inpatient stay dataset. Goodness of fit of the model was employed to assess whether the subset can be mapped into the original dataset. The most common techniques to do this are: variance explained by the factors, reproduced correlation matrix, goodness of fit

tests and standard errors of factor loadings. Goodness of fit includes the log likelihood ratio test and adjusted chi-square test statistics to take into account non-normality. A non-significant chi-square estimate or a chi-square: df ratio of less than 3:1 (Kline, 2005) is considered goof fit. RMSEA and SRMR values less than .05 suggest good fit and values up to .08 indicate reasonable errors of approximation in the population (Browne and Cudeck, 1992). Also CFI and TLI values above 0.90 indicate adequate fit (Hu and Bentler, 1999).

6.3 Results

6.3.1 Dimensions estimated by EFA

All factor analysis work was carried out in Mplus 7 Software. EFA was fitted in each subgroup and corresponding eigenvalue and screeplot were examined to determine the number of factors. A cut-off of an eigenvalue ≥1 yielded 10 or 11 factors in the four subgroups (Appendix 11). The scree plots had long tails and the elbow started at five-factor model. However this sloped off afterwards (Appendix 12).

I estimated EFA for to up to 16-factor model for each subgroup using weighted least squares mean and variance (WLSMV) estimator and orthogonal rotation. A summary of the model fit for each subgroup estimates from 6 to 15 factor model is illustrated in Table 6. Only three models across two subgroups had the form specified by the covariance matrix and displayed good fit amongst all the models identified. The remaining EFA models were either not estimated or generated statistically significant chi-square value indicating poor fit.

Table 6: Summary of Model Fit

	Emergency Admission with operation or procedure			Planned Admission with operation or procedure		
	Chi-square	Degrees of freedom	P-Value	Chi-square	Degrees of freedom	P-Value
6-factor model	25325.74	1122	0	33279.35	1219	0
7-factor model	20770.25	1074	0	26940.21	1169	0
8-factor model	17340.21	1027	0	22669.01	1120	0
9-factor model	14661.03	981	0	0.234	1072	1
10-factor model	N/A			0.193	1025	1
11-factor model	0.169	892	1	N/A		
12-factor model	N/A			N/A		
13-factor model	N/A			N/A		
14-factor model	6349.19			N/A		
15-factor model	N/A			N/A		
	Emergency Admission without operation or procedure			Planned Admission without operation or procedure		
	Chi-square	Degrees of freedom	P-Value	Chi-square	Degrees of freedom	P-Value
6-factor model	26137.39	855	0	5016.41	940	0
7-factor model	21644.04	813	0	N/A		
8-factor model	18170.76	772	0	3517.72	853	0
9-factor model	N/A			3033.72	811	0
10-factor model	12245.39	693	0	N/A		
11-factor model	10099.01	655	0	N/A		
12-factor model	N/A			N/A		
13-factor model	N/A			N/A		
14-factor model	6198.97	547	0	N/A		
15-factor model	N/A			N/A		

The EFA models fitted in the two subgroups consisting of patients with operation or procedure during their hospital stay had good fit. The factor models for patients without any operation or procedure, using emergency and planned admission route, had a statistically significant chi-square value indicating poor fit. This could be due to relatively smaller sample size. The majority of these patients were older and perhaps had care needs that were not necessarily clinical. The focus from here on will only be on patients who underwent an operation or procedure during their inpatient stay, which also reflects the majority of inpatients in the NHS and medical patients.

Only the 11-factor model for patients with emergency admission and operation or procedure had good fit. But two EFA models for patient with planned admission and operation or procedure were specified as having satisfactory fit. These were the 9 and 10 factor models, and based on further examination of goodness of fit tests, the 10-factor model was selected. This ensured that the 11-factor and 10-factor EFA model for the two subgroups had good model fit with RMSEA and SRMR value less than 0.05.

The full EFA model results are reported in detail in Appendix 13 and Appendix 14 . It presents the fit of the EFA models and factor items reporting a varimax rotated loading of over 0.4 included in the models. The items grouped together for each factor were further interpreted based on what underlying factor it could be suggesting. Some factors were made up of items that fell chronologically and were quite specific, so easy to interpret. Others were broad and contained disparate items, and were more difficult to interpret.

The eleven factors in the 11-factor model identified for patients with emergency admission and operation or procedure covered following: information on operation or procedure; healthcare professionals talking in front of the patient as if they weren't there; information about condition or treatment; privacy; physical comfort and feeling safe; emotional support; being asked for feedback; cleanliness; information about medication; provisions for after leaving hospital; and aspects of discharge. While some factors were very specific and made up of two items (these were factor 2,

4, 7 and 8), others were broader and contained disparate items (for example, factor 3 and 5).

The 10-factor model for patients with planned admission comprised of waiting time; information about treatment, operation or procedure; being asked for feedback; comfort, trust and communication; aspects of discharge; cleanliness; emotional support; information about medication; support for after discharge. Note that one of the factors in this model only had items with loading less than 0.4 and was excluded. A summary of the final 11-factor and 9-factor EFA model for patient with emergency and planned admission is presented in Table 7.

Table 7: Dimensions described for patients with emergency and planned admission using EFA

Emergency admission (11-Factor)	**Planned admission (9-Factor)**
	Factor 9: Time on waiting list
Factor 4: Information on operation or procedure	Factor 3: Information about treatment, operation or procedure
Factor 7: Doctor and nurses talking in front of the patient as if they weren't there	
Factor 2: Communication, trust and feeling involved	Factor 1: Comfort, trust and communication
Factor 8: Privacy when being treated	
Factor 3: Noise	
Factor 6: Emotional support from hospital staff	Factor 5: Emotional support from hospital staff
Factor 11: Being asked for feedback	Factor 8: Being asked for feedback
Factor 1: Cleanliness	Factor 2: Cleanliness
Factor 5: Information about medication	Factor 4: Information about medication
Factor 10: Provisions for after leaving hospital	Factor 7: Provisions for after leaving hospital
Factor 9: Aspects of discharge	Factor 6: Aspects of discharge

The patients included in the EFA analysis only differed in terms of route of admission however the dimensions derived across the two subgroups differed substantially. Eight of the dimensions described in the two subgroups were very similar. The length of time on the waiting list before being admitted in the hospital was identified as a factor explaining variance in inpatient data of patients with planned admission. And in patients with emergency admission, the following additional factors were generated: nurses talking in front of patients as if they weren't there, noise in the hospital and privacy when being treated.

6.3.2 Dimensions assessed using CFA

CFA models were fitted in the dimensions estimated using EFA for patient with operation or procedure for assessment. The CFA employs cross loadings, local correlation and theoretical consistency. Findings from it were used to revise the EFA model. In the model for emergency admission, the factors from the 11-factor EFA model were broadly the same but with addition of four items across three factors. The CFA models estimated for the 11 factors, including model fit and factor scores is presented in Appendix 15 for illustrative purpose. The final CFA model estimated for those with planned admission revised the EFA model considerably. An 8-factor model was supported by the CFA. The variable on length of time on the waiting list before admission to hospital was not included. Also the variable assessing amount of information provided to the patient about their condition or treatment was excluded. Four additional variables were included instead. The final CFA models are presented in Table 8 and Table 9, and the additional items included in the CFA are highlighted in italic in the tables. Model fit statistics for the final CFA models are summarised in Table 10. The final models for the two subgroups exhibited good fit with RMSEA<0.5. CFI and TLI were equal to 0.95 in subgroup with emergency admission. CFI was equal to 0.95 and TLI was 0.94 in subgroup with

planned admission. The diagrams illustrating the CFA for the two subgroups are presented in Appendix 16 and Appendix 17.

Table 8: 11-Factor Model identified for patients with emergency admission and operation or procedure using CFA

Factor	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
F1	In your opinion, how clean was the hospital room or ward that you were in?	How clean were the toilets and bathrooms that you used in hospital?				
F2	*While you were in the A&E Department, how much information about your condition or treatment was given to you? **	When you had important questions to ask a doctor, did you get answers that you could understand?	Did you have confidence and trust in the doctors treating you?	*Sometimes in a hospital, a member of staff will say one thing and another will say something quite*	Were you involved as much as you wanted to be in decisions about your care and treatment?	How much information about your condition or treatment was given to you?
F3	Were you ever bothered by noise at night from other patients?	Were you ever bothered by noise at night from hospital staff?				

Factor	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
F4	Beforehand, did a member of staff explain the risks and benefits of the operation or procedure in a way you could understand?	Beforehand, did a member of staff explain what would be done during the operation or procedure?	Beforehand, did a member of staff answer your questions about the operation or procedure?	Beforehand, were you told how you could expect to feel after you had the operation or procedure?	After the operation or procedure, did a member of staff explain how the operation or procedure had gone in a way you could understand?	
F5	Did a member of staff explain the purpose of the medicines you were to take at home in a way you could understand?	Did a member of staff tell you about medication side effects to watch for when you went home?	Were you told how to take your medication in a way you could understand?	Were you given clear written or printed information about your medicines?	Did a member of staff tell you about any danger signals you should watch for after you went home?	
F6	Did you find someone on the hospital staff to talk to about your worries and fears?	Do you feel you got enough emotional support from hospital staff during your stay?				
F7	Did doctors talk in front of you as if you weren't there?	*Did nurses talk in front of you as if you weren't there? **				

Factor	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
F8	Were you given enough privacy when discussing your condition or treatment?	*Were you given enough privacy when being examined or treated? **				
F9	Did hospital staff take your family or home situation into account when planning your discharge?	Did hospital staff discuss with you whether you would need any additional equipment in your home, or any adaptations made to your home, after leaving hospital?	Did hospital staff discuss with you whether you may need any further health or social care services after leaving hospital?			
F10	Did you feel you were involved in decisions about your discharge from hospital?	Were you given enough notice about when you were going to be discharged?				
F11	During your hospital stay, were you ever asked to give your views on the quality of your care?	Did you see, or were you given, any information explaining how to complain to the hospital about the care you received?				

Table 9: 8-Factor Model identified for patients with planned admission and operation or procedure using CFA

Factor	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10
F1	*Were you ever bothered by noise at night from other patients? **	*Were you ever bothered by noise at night from hospital staff? **	*Did doctors talk in front of you as if you weren't there? **	Did you have confidence and trust in the nurses treating you?	Did nurses talk in front of you as if you weren't there?	In your opinion, were there enough nurses on duty to care for you in hospital?	Did a member of staff say one thing and another said something quite different. Did this happen to you?	Were you given enough privacy when discussing your condition or treatment?	Were you given enough privacy when being examined or treated?	Overall, did you feel you were treated with respect and dignity while you were in the hospital?
F3	Did a member of staff explain the risks and benefits of the operation/ procedure in a way you could under-stand?	Did a member of staff explain what would be done during the operation/ procedure?	Did a member of staff answer your questions about the operation/p rocedure?	Were you told how you could expect to feel after you had the operation/p rocedure?	Before the operation/ procedure, did a staff explain how they would put you to sleep/control your pain in a way you could understand?	After the operation/ procedure, did a staff explain how it went in a way you could under-stand?				

Factor	Item 1	Item 2	Item 3	Item 4
F2	In your opinion, how clean was the hospital room or ward that you were in?	How clean were the toilets and bathrooms that you used in hospital?		
F4	Did a member of staff explain the purpose of the medicines you were to take at home in a way you could understand?	Did a member of staff tell you about medication side effects to watch for when you went home?	Were you told how to take your medication in a way you could understand?	Were you given clear written or printed information about your medicines?
F5	Did you find someone on the hospital staff to talk to about your worries and fears?	Do you feel you got enough emotional support from hospital staff during your stay?		
F6	Did you feel you were involved in decisions about your discharge from hospital?	Were you given enough notice about when you were going to be discharged?		
F7	Did hospital staff discuss with you whether you would need any additional equipment in your home, or any adaptations made to your home, after leaving hospital?	Did hospital staff discuss with you whether you may need any further health or social care services after leaving hospital?		
F8	*Did a member of staff tell you about any danger signals you should watch for after you went home? **	During your hospital stay, were you ever asked to give your views on the quality of your care?	Did get information on how to complain to the hospital about the care you received?	

Table 10: Model fit information of CFA models

CFA for patients with emergency admission and operation or procedure	
RMSEA (Root Mean Square Error Of Approximation)	
Estimate	0.045
90 Percent C.I.	0.044 0.046
Probability RMSEA <= .05	1.000
CFI/TLI	
CFI	0.956
TLI	0.947
CFA for patients with planned admission and operation or procedure	
RMSEA (Root Mean Square Error Of Approximation)	
Estimate	0.047
90 Percent C.I.	0.046 0.047
Probability RMSEA <= .05	1.000
CFI/TLI	
CFI	0.943
TLI	0.935

6.4 Discussion

The study presented in this chapter examined the hypothetical structure of the inpatient dataset using factor analytic techniques. Based on the understanding gained about the Inpatient dataset in chapter 5, non-linear nature of the variables and use of subgroups to describe the dataset were important considerations in the factor analysis conducted. Factor analysis using weighted least squares mean and variance estimator was used to identify dimensions and an orthogonal rotation was specified. Factor analysis method aggregated items into dimensions using the hypothetical structure of the dataset. These factors were assessed for unidimensionality and association of items with dimensions and between dimensions.

6.4.1 Unidimensionality of dimensions

The dimensions generated automatically using EFA and without prior theory are inherently unidimensional. EFA was successfully fitted in only two of the four subgroups. In patients undergoing operation and

procedure, an eleven-factor model was estimated and a ten-factor model was estimated for patients with emergency and planned admission. The model comprised items that were mostly chronological. This is not surprising because the Inpatient questionnaire was designed by CQC to cover seven settings: admission; doctors and nurses; hospital and ward; leaving the hospital; operations and procedures; and patient care and treatment.

EFA was successful in reducing the large Inpatient survey dataset into a smaller number of dimensions, containing one to seven items. A similar method used to summarise data matrix in fewer dimensions without loss of information is PCA. Instrument developers use it to assess data dimensionality. It is not possible to conduct PCA analysis in Mplus 7. I used SPSS software to estimate PCA in the subgroups. Given the ordinal nature of the dataset, CATPCA was used. I also specified WLSMV estimation and varimax rotation. While the principal components derived using PCA were similar to factors estimated using EFA, model fit information was lacking and assessment of overall model was not possible. Also in order to confirm the dimensions generated using PCA the results would have to be entered in a factor analysis framework.

6.4.2 Number of dimensions

Exploratory analysis relied on eigen values and scree plot to determine the number of dimensions to select. Both techniques are quite subjective, for example using an eigen value of 1.5 results in half the number of dimensions (4 or 5) compared to using an eigen value of 1. Similarly examining the scree plot, the elbow started for all the subgroups between three and five factors. The selection of items in each factor using exploratory approach can be a little arbitrary as well. The factor loading

commonly used in literature for item selection per dimension varies from 0.3 to 0.5, and depending on the cut off chosen the number of items per dimension will differ. EFA allows the user or researcher to input their judgement and hence is flexible to purpose, but the disadvantage of this method would be that it is subjective. I have used CFA to reassess the factors and factor items.

Given that the overall aim of this thesis is to develop an instrument, which is amenable to valuation, it would have been preferable to have a brief instrument and a smaller number of factors. I opted to rely on the fit of the model and assessment using confirmatory framework (which is explained in more detail in the next section). Further refinement of the model will have to be conducted later using qualitative or quantitative methods. This includes rating of the dimensions based on importance or frequency and estimating the contribution of each dimension to the overall construct being measured using SEM.

6.4.3 Assessment of measurement model using confirmatory framework

It was important to assess the dimensions and items estimated using EFA in a confirmatory framework that evaluates the overall measure and not just individual dimensions or constructs. It should be highlighted here that use of model based approach to factor analysis is a recent development. Methodological development and advances in technology over recent years have enabled a range of methodologies to become more accessible to instrument developers. This was apparent in the review conducted in chapter 4, in which the majority of instrument developers employed

statistical techniques to establish unidimensionality but did not assess the overall fit of the model.

Use of confirmatory model is likely to be important if one is developing a multidimensional instrument. It helps to establish construct validity of the descriptive system and ensure good measurement properties. The EFA models were re-estimated using CFA framework and both items and dimensions generated were revised until good model fit was achieved. The final measurement model for patient who underwent an operation or procedure and had an emergency admission included 33 items across 11 dimensions. In the patients who underwent an operation or procedure but had a planned admission, the final model comprised of 31 items across 8 dimensions.

The dimensions estimated using EFA for patients who had an emergency admission did not change much when it was examined in CFA framework. But the EFA model in second subgroup of patients, who had a planned admission, was revised considerably and the final model included only eight factors. The factor that was dropped during confirmatory analysis was assessment of time on the waiting list before admission to hospital. While this variable is very important to patients who have planned admission and may affect overall assessment of the NHS, one can argue that it is not indicative of the quality of care received at the hospital and it precedes the actual inpatient stay. Hence not including it in a measure of patient experience in a hospital is reasonable.

6.4.4 Comparing factor analysis models to logistic regression models

Regression models and factor analytic techniques can be used to derive a smaller subset of variables from a large dataset, but the purpose and mechanics involved in the two approaches are different. Hence it is not surprising that the results are different.

Regression models developed in previous chapter used patient rating scale of overall experience as the dependent variable. The logistic regressions employed identified the variables that were most statistically significant in estimating the probability of good patient experience. The factor analysis models are focussed on generating a subset of variables by grouping together items to form latent factors based on variance or covariance of observed variables. Factor models can be described as regression models with observed variables as the dependent variables and latent variables representing independent variables. In both approaches logit distribution was used.

The focus of the two techniques is very different. Regression analysis places emphasis on the observed dependent variables and determines statistical significance of the items based on how it is associated with the dependent variable. On the other hand, factor analysis model does not have an observed dependent variable. The focus is on representing the full dataset based on interdependencies of variables in the dataset; this is an important consideration when developing a brief instrument from an existing instrument.

In the previous chapter the items selected by regression analysis as being most statistically significant were: being treated with respect and dignity in

the hospital, the patient having confidence and trust in the nurses treating them and the cleanliness of the ward. However if these items were to be ranked based on the magnitude of the coefficient or the p-value, and only the top ten items were to be selected (see Table 5 in chapter 6) one would miss out on important aspects of care which were included in the original instrument such as information and communication about procedure or treatment carried out; medication; and provisions for after leaving the hospital. In order to incorporate all of the items one would need an instrument with over 20 items, which again defeats the purpose of this study.

It should be noted that factor analysis generates dimensions consisting of one or multiple items, whereas regression analysis directly identifies items. Adopting a staged approach in which dimensions summarising the dataset are estimated in the start followed by item selection allowed for unidimensionality to be established within each dimension.

6.4.5 Dimension of care identified in the literature

A large number of studies have identified dimensions of patient care using both qualitative and quantitative framework. Over fifty years ago, Avedis Donabedian published his seminal article on measuring the performance of healthcare titled "Evaluating the Quality of Medical Care" and introduced ‘logic, evidence, and scientific’ inquiry to this area (Berwick and Fox, 2016, Donabedian, 1966). Thousands of articles have been published since then to define and measure quality of care, with an increasingly large number of them focussed on patient-centeredness and use of PROs (Berwick and Fox, 2016, Doyle et al., 2013, Murray and Frenk, 2000, Reeves and Seccombe, 2008, Sitzia and Wood, 1997a).

There are two key pieces of research that have identified the dimensions of healthcare (non-clinical) using several studies and at a large scale. The first stems from WHO's multi-national investigation on: "What makes for a good health system? What makes a health system fair? And how do we know whether a health system is performing as well as it could?" (WHO, 2000). The report identified three goals for the health system and one of them was responsiveness. Health systems' responsiveness included eight domains categorised across two categories: respect-for persons (interpersonal) and client-orientation (structural domains)(Valentine et al., 2008). The first covered the domains: dignity, autonomy, confidentiality and communication (Donabedian, 1980). The second category comprised of following domains: choice of care provider, prompt attention, quality of basic amenities and access to social support networks (during inpatient care) (Campbell et al., 2000). The domains and questions within it were presented in chapter 5 (see

Figure 4).

The second set of dimensions is called the Patient Experience Framework, it was published by the NHS National Quality Board in 2012 to improve NHS trusts (DH, 2012a). The eight dimensions from the NHS Patient Experience Framework are described in

Table 11. These dimensions are used by the CQC to monitor the performance of NHS trusts (CQC, 2017). The domains described by the WHO's responsiveness model and the Patient Experience Framework are very similar, I will focus on the latter to make comparison with my findings using factor analysis as they are more specific to the UK healthcare.

Table 11: Dimensions from NHS Patient Experience Framework

Dimensions	Examples provided
Respect for patient-centred values, preferences, and expressed needs	e.g. awareness of cultural and quality-of-life issues; the dignity, privacy and independence of patients and service users
Coordination and integration of care	e.g. every professional involved in care pathway having access to care records
Information, communication and education	e.g. being informed about available options in a clear way; having the opportunity to discuss concerns
Physical comfort	e.g. pain management; cleanliness of wards; quality of food
Emotional support	e.g. reassurance; being listened to; being able to ask questions
Involvement of family and friends	e.g. level of partner involvement during childbirth
Transition and continuity	e.g. knowing what to expect at each stage of planned care journeys; "seamless" care
Access to care	e.g. receiving care as close to home as possible, length of referral time/journeys; "seamless" care

Note: Adopted from https://www.gov.uk/government/publications/nhs-patient-experience-framework

I examined the dimensionality of the Inpatient survey data using exploratory factor analysis technique (no prior theory was used). Three dimensions from the Patient Experience Framework were not identified in my analyses, these were access to care; coordination and integration of care and involvement of family and friends. However these items did not emerge as one of the top ten statistically significant variables in regression analyses presented in chapter 5, with the exception of one question on the patient's view on there being enough nurses on duty in the hospital (Q30). The other factors fall within the five dimensions of patient care identified from the literature and are listed in Table 12. One of the factors I estimated (highlighted using an asterisk in Table 12) is broad and includes elements from more than one dimension of the Patient Experience Framework. The most representative item for this dimension will be selected in chapter 7.

Table 12: Comparing CFA Dimensions to dimensions from the NHS Patient Experience Framework

Dimensions	Emergency admission	Planned admission
Respect for patient-centred values, preferences, and expressed needs	Doctor and nurses talking in front of the patient as if they weren't there; Being asked for feedback; Privacy when being treated	Being asked for feedback
Information, communication and education	Information on operation or procedure; Information about medication; Communication, trust and feeling involved*	Information about treatment, operation or procedure; Information about medication; Comfort, trust and communication*
Physical comfort	Noise; Cleanliness	Cleanliness
Emotional support	Emotional support from hospital staff	Emotional support from hospital staff
Transition and continuity	Provisions for after leaving hospital; Aspects of discharge	Provisions for after leaving hospital; Aspects of discharge

It is possible that the Inpatient survey was built using the dimensions of care from the literature, but it has been over ten years since it was first administered in the UK. Several questions have been added, removed and amended over time. For example the design of 2016 Inpatient Survey questionnaire changed from the 2015 version (CQC, 2017). The new questionnaire had one question amended, one question removed and three new question added. The number of items have changed substantially over time and so has the hypothetical construct of the survey. The dimensions from the Patient Experience Framework did not match the hypothetical construct of the Inpatient Survey 2014.

6.4.5 Comparing subgroups

The analyses I conducted excluded patients who did not have an operation or procedure. Factor models could not be established for these patients and this could be because of small sample size. I was able to capture the

different aspects of healthcare valued by patients with planned admission and emergency admission within this group.

For patients who had an operation or procedure, an 11-factor model was determined for those who had an emergency admission and an 8-factor model was determined for patients with planned admission. The items across the two models were broadly consistent. Two factors that emerged in patients with emergency admission but was missing in patients with planned admission, was privacy and noise. Note that privacy variable was found to be statistically significant only in patients with planned admission but not in patients with emergency admission when regression models were fitted in the same subgroup earlier (see chapter 5). The two findings are contradictory, however it should be noted that two methods have different purposes.

6. Conclusion

The focus of this chapter was to identify and confirm the overall hypothetical scale structure of a large secondary dataset (Inpatient Survey 2014). I excluded patients who did not have an operation or procedure from the analyses. However I was able to capture the difference in the hypothetical construct of the data for patients with emergency and planned admission using an 11-factor model and an 8-factor model respectively. The dimensions established for the patients with emergency admission were: doctors and nurses talking in front of the patients as if they were not there; being asked for feedback; privacy when being treated; information on operation or procedure; information about medication; communication, trust and feeling involved; noise at night; cleanliness of the facility; emotional support from hospital staff; provision for after leaving the hospital and aspects of discharge. The dimensions for patients

with planned admission were: being asked for feedback; information about treatment, operation or procedure; information about feedback; comfort, trust and communication; cleanliness of the facility; emotional support from hospital staff; provision for after leaving hospital and aspects of discharge. In terms of methods, firstly use of EFA in this chapter ensured unidimensionality; secondly use of CFA to examine the relationship between the latent factors estimated and items within it confirmed the measurement model.

Chapter 7 Item selection for each dimension of patient experience

7.1 Introduction

In the previous chapters regression analysis and factor analysis have been applied in the inpatient dataset with the view of developing an instrument from the original questionnaire used in the survey. Use of regression analysis helped in understanding the dataset, the items selected from this method however did not summarise the dataset and the items overlapped conceptually or were not unidimensional. In chapter 6, I applied factor analysis models in the same NHS Inpatient Survey data to determine the dimensionality in the dataset. In addition, the interrelationship between items and dimensions was assessed using confirmatory factor based model. And the results were used to finalise the number of dimensions and content. A total of two to ten items were included per dimension.

Using all the items selected in each dimension as final will result in overlap of items in each unidimensional scale. This is likely to interfere in valuation task as the items in each dimension are measuring the same latent construct and selecting combination of these items to generate healthcare profiles is likely to result in nonsensical states. In this chapter the goal is to select one item per dimension, to reduce the instrument further and produce a brief measure of patient experience. Also selecting one item per dimension will ensure that the items are distinct and do not overlap, and ease future scoring system and/or valuation.

The Inpatient dataset comprises of binary and ordinal response items, and two approaches are available for item selection. The first approach is IRT-based and uses an item response function to select the item providing most information for each latent construct. The second approach is the factor analysis model and uses underlying variable theory to study the interrelationship between underlying variables. Note that application of both IRT and factor analysis will be on each subset of the data based on dimensions established earlier. Each subset measures one unique latent construct and one-factor models are fitted. The EFA and CFA model are identical in one-factor models. While application of IRT and factor analysis to identify items that represent the overall dataset is key here.

7.1.1 Possible dimension selection

In addition to determining the hypothetical structure of the dataset and establishing dimensions with statistical techniques, like the ones adopted here, it is important to make a careful selection of dimensions. The most widely used preference based measures have between five and nine dimensions.

The 8-dimension and the 11-dimension measure I have generated here will be considered as having too many dimensions and just proceeding to an item selection will not resolve it.

Further selection needs to be undertaken, and this requires a balance between comprehensiveness and pragmatism. I could use my judgement on what I believe are the most important dimensions or eliminate the ones that I think are not important based on interpretation of the factors. Secondly opinions (or rating) studies using relevant population such as the

general public or patients can be used to assist selection process. Finally statistical techniques can be used.

I have used an advanced statistical model called SEM for dimension selection, which employs full response dataset and is an extension of the factor analysis approach. SEM enables structural parts or regression equations to be added into the measurement models. It is intuitively appealing for use in the development of a descriptive system as it allows causality to be examined in detail. The review of the literature described in chapter 4 listed SEM as a method used for development of a descriptive system, however the instrument developers were vague about it and did not report sufficient detail on how it was used. In addition to item selection, SEM application and discussion is provided in this chapter.

7.2 Methods

7.2.1 Unidimensional dataset

An 11-factor model and 8-factor model established for patients that underwent an operation or procedure through emergency admission and planned admission respectively are revisited in this section. Each factor comprised of two or more items associated with a distinct latent construct. One-factor models were constructed using the two methods described below for each dimension. An item was selected to represent each factor.

7.2.2 IRT

The assumptions made within the IRT approach are: 1) the latent variables are independent and normally distributed with mean zero and variance one and 2) the responses to the ordinal items are independent, conditional on the latent variables (conditional independence). There are different

types of IRT models. The unidimensional latent variable model for binary and polytomous variable is called a two-parameter model. A commonly used two-parameter model is the graded response model (GRM) (Samejima, 1968). It is called the GRM because it secures the order of participant responses.

The two-parameter IRT involves modelling the probability of a randomly selected individual giving a positive response to an item as a function of the latent variable. This is done in terms of a set of probabilities $\{\pi_i(f)\}$ and it is an adaptation of the logistic regression model described in chapter five (see equation 2 and 3 in chapter 5). For the binary variable, a logit model is estimated which expresses the logit of the probability of a response in category one as a linear function of the fs. The polytomous variables are also modelled as dichotomous whereby the probability of the response falling into the first and second groups respectively is written as following:

$$\boldsymbol{\gamma_{i(j)}(f) = \mathbf{Pr}(x_i \leq j) = \pi_{i(1)}(f) + \pi_{i(2)}(f) + \ldots + \pi_{i(j)}(f)} \quad \textbf{... Equation 13}$$

and

$$\boldsymbol{1 - \gamma_{i(j)}(f) = \mathbf{Pr}(x_i > j) = \pi_{i(s+1)}(f) + \pi_{i(s+2)}(f) + \ldots + \pi_{i(m_i)}(f)} \quad \textbf{... Equation 14}$$

Where, x_i denotes the category into which the i-th variable falls. The probabilities $\gamma_{i(j)}(f)$ are referred to as cumulative response probabilities. On the assumption that the binary logit model holds for all divisions of the m_i categories divided into two groups, the model can be written in terms of logit as following

$$\boldsymbol{log\left[\frac{\gamma_{i(j)}(\mathbf{f})}{1+\gamma_{i(j)}(\mathbf{f})}\right] = \alpha_{i(j)} + \sum_{s=1}^{q} \alpha_{is} f_s} \quad \textbf{... Equation 15}$$

Where j= $1, \dots, m_i - 1; i = 1, \dots, p$. For a positive factor loading α_{is} the higher the value of an individual on the latent variable f_s , the higher the probability of that individual responding in the higher categories of item i. In other words, a given change in the value of f_s will produce a larger change in the probability of a positive response when this parameter is larger than when it is small. In educational testing, this is referred to as the discrimination parameter. Increasing the parameter α_{is} increases the probability for all values of f_s and so it is referred to as the difficulty parameter. A special case of the unidimensional model is obtained when all the discrimination parameters are equal. Such a one-parameter logistic model is called Rasch model (Rasch, 1960) and the rating scale model falls within it.

The two-parameter model also contains one intercept parameter $\alpha_{i(j)}$ for each category. The ordering of the categories implies that the intercept parameters are also ordered:

$$\boldsymbol{\alpha_{i(1)} \leq \alpha_{i(2)} \leq \cdots \leq \alpha_{i(m_i)}} \quad \textbf{... Equation 16}$$

However the factor loadings α_{is} is identical across categories of the same variable. This means that the discriminating power of the item is not dependent on where the split into two groups is made. The π s are obtained from the γ s by

$$\boldsymbol{\pi_{i(j)}(f) = \gamma_{i(j)}(f) - \gamma_{i(j-1)}(f) \qquad (j = 2, \dots, m_i)} \quad \textbf{... Equation 17}$$

Where $\gamma_{i(1)}(f) = \pi_{i(1)}(f)$ and $\gamma_{i(m_i)}(f) = 1$. The $\pi_{i(j)}(f)$ is referred to as the category response function. In psychometrics literature it is referred to

as the item characteristics curve and it shows how the probability of a correct response increases with say ability.

This type of IRT model is known as the 'difference' model, in which the probabilities are set as differences between cumulative probabilities. It is based on the assumption that all items fitted in the model have the same number of response categories. The inpatient dataset consists of items with different levels and a difference model cannot be used in the dataset.

7.2.3 Generalised Partial Credit Model

Another commonly used IRT approach is the 'divide-by-total' models in which probabilities are set as ratios of values divided by the sum of these values across response categories. They allow response categories to vary across items and they are more suited to the dataset I am using in this thesis.

One of the earliest polytomous IRT models to use divide by total approach is the partial credit model (Masters, 1982). It is an extension of the one-parameter logistic model (Rasch model). Master's partial credit model treats polytomous responses as ordered performance levels, assuming that the probability of selecting the kth category over the [k – 1]-th category for item j follows a conditional probability

$$\frac{\pi_{ik}}{\pi_{i(k-1)} + \pi_{ik}} = \frac{\exp \sum_{h=0}^{k} \gamma_{i(j)}}{\sum_{k=0}^{m_j} \exp \sum_{h=o}^{k} \gamma_{i(j)}} \quad \textbf{... Equation 18}$$

where the numerator is the individual response outcomes and the denominator is the sum of all the possible outcomes. i=1,2,…, N refers to

individual respondents, N refers to total number of respondents in the sample, j=1,2,…, J refers to items and h=1,2,…, k refers to the number of response categories.

Muraki introduced generalisation of the partial credit model in 1992 but with a parameter for item discrimination added to the model (Muraki, 1992). It is called the generalised partial credit model (GPCM) and inserts a discrimination parameter $\alpha_{is}$for each item as described below

$$\frac{\exp \sum_{h=0}^{k} \alpha_{is}\gamma_{i(j)}}{\sum_{k=0}^{m_j} \exp \sum_{h=o}^{k} \alpha_{is}\gamma_{i(j)}} \qquad \textbf{… Equation 19}$$

The parameters $\gamma_{i(j)}, \pi_{ik}$ and α_{is} can respectively be interpreted as a person's underlying patient experience, the patient experience measured by the response category threshold and an item's ability in discriminating between persons with different underlying experience of hospital stay.

The goodness of fit of the model can be checked in different ways. The IRT is made up of four key assumptions. A global goodness of fit test that compares the observed and expected frequencies across the response patterns is the Pearson chi-squared goodness of fit. It examines the item response function. A statistically significant chi-square statistic indicates poor fit of the model. It should be noted here that while adequacy of the model is important, the goal here is on item selection for every one-factor model and the focus is on examining relative ordering of items in terms of discrimination to select the most representative item.

7.2.4 Factor analysis model

The alternative approach for constructing and fitting a factor analysis model in binary and polytomous items is called the underlying variable approach and it consists of the classical linear factor analysis model, described in detail in previous chapter. In this approach the observed variables are assumed to be realisation of continuous underlying variables. The assumption here is that the variable is unobserved and one can only observe whether or not each variable exceeded a threshold. In order to fit the model there are three sets of parameters to be estimated. They are the thresholds, the polychoric correlations between the underlying variables and the item (factor) loadings.

The origin and unit of measurement of the latent variable is unknown since it is unobserved. In a factor model, the origin of this variable is usually set to zero and the scale of the unobserved variable is set using two alternative ways. Note that both lead to equivalent solutions. The first is 'standardised' latent variable, which assumes that they have zero means and unit variances in the population. When the latent variable is standardised and fitted in a factor model the correlation between the latent variables is estimated.

An alternative way to set the scale of a latent variable is to assign it the same scale as one of the observed items and set its factor loading as equal to one. The variable selected to represents the latent construct is known as the reference variable.

I tested both approaches in the study but opted for the standardised approach to determine association between latent factors. The

interrelationship between the observed items and the underlying latent construct are assessed to order the items based how much of the variance in the unobserved variable is explained by each item. The item with the highest value was considered most central to the underlying variable.

7.2.5 SEM

Finally SEM was used to investigate its use in the development of a descriptive system, as a basis of item selection and dimension selection. SEM falls within the underlying variable approach. It is a framework that brings together simultaneous equation models, factor analysis and path analysis. It adds the structural part to the measurement model to capture the relationship between latent explanatory (ξ) variables and latent dependent variables (say η); and/or among latent dependent variables (η_1 and η_2).

It should be noted that although interest is more on the structural part of the model, the structural part stands on the measurement model that defines the constructs through observed variables. The measurement model needs to be tested first and only when an adequate or satisfactory fit is obtained, the structural part is added. SEM is employed here to estimate the contribution of each underlying unobserved factor to the overall construct of patient experience. This will enable ordering of each factor in terms of relevance.

SEM model can be described using following measurement equations:

$\mathbf{x} = \boldsymbol{\tau}_{\mathbf{x}} + \Lambda_x \boldsymbol{\xi} + \boldsymbol{\delta}$ **… Equation 20**

$\mathbf{y} = \boldsymbol{\tau}_{\mathbf{y}} + \Lambda_y \boldsymbol{\eta} + \boldsymbol{\epsilon}$ **… Equation 21**

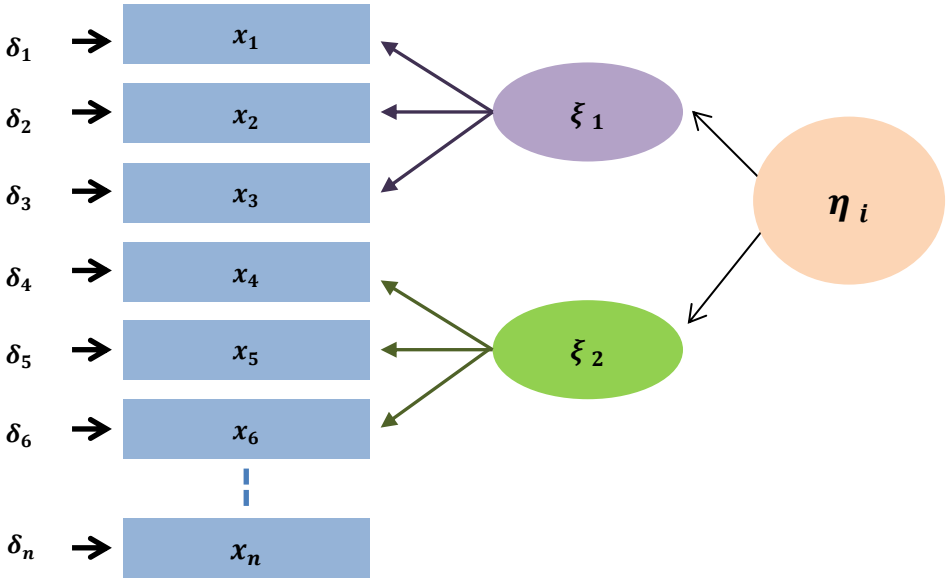
And structural equation:

$\boldsymbol{\eta = \beta_{\eta} + \gamma\xi + \zeta}$ **… Equation 22**

The model assumes that the covariance matrix is diagonal and that errors terms are uncorrelated i.e. error terms δ are uncorrelated with the ξ variables, error terms ϵ are uncorrelated with the η variables, and error term ζ is uncorrelated with the ξ variables and uncorrelated with the measurement errors ($\delta \text{ and } \epsilon$).

Measurement models produced using the underlying variable approach is examined using a single structural equation. It examines the relationship between latent explanatory variable, i.e. the factors estimated, with latent dependent variable measuring the overall construct, say patient experience, using the underlying variable approach. An illustration of the SEM conducted is provided in Figure 6. The same tools introduced earlier to examine model fit and adequacy of factor analysis models applies to SEM too (see section 6.2.5).

Figure 6: Illustration of a SEM



Key
ξ = latent explanatory variable (factor)
η = latent dependent variable (overall construct)
x = observed variable (item)
δ = measurement error
Arrow (path) = relationship between factor and item

7.3 Results

7.3.1 IRT application

To identify the most robust items to use in the descriptive system for patient experience, I fitted GPCM models separately to each of the dimensions estimated in chapter 6. For illustration purpose, I will demonstrate use of IRT in factor two for patients with emergency admission that underwent an operation or procedure (subgroup 1). It consisted of six multi-level items, and are described below:

Q3. While you were in the A&E Department, how much information about your condition was given to you?

- Right amount
- Not enough/too much
- Don't know/can't remember
- Not given any information

Q24. When you had important questions to ask a doctor, did you get answers that you could understand?

- Yes, always
- Yes, sometimes
- I had no need to ask
- No

Q25. Did you have confidence and trust in the doctors treating you?

- Yes, always
- Yes, sometimes
- No

Q31. Sometimes in a hospital, a member of staff will say one thing and another will say something quite different. Did this happen to you?

- Yes, always
- Yes, sometimes
- No

Q32. Were you involved as much as you wanted to be in decisions about your care and treatment?

- Yes, definitely
- Yes, to some extent
- No

Q33. How much information about your condition or treatment was given to you?

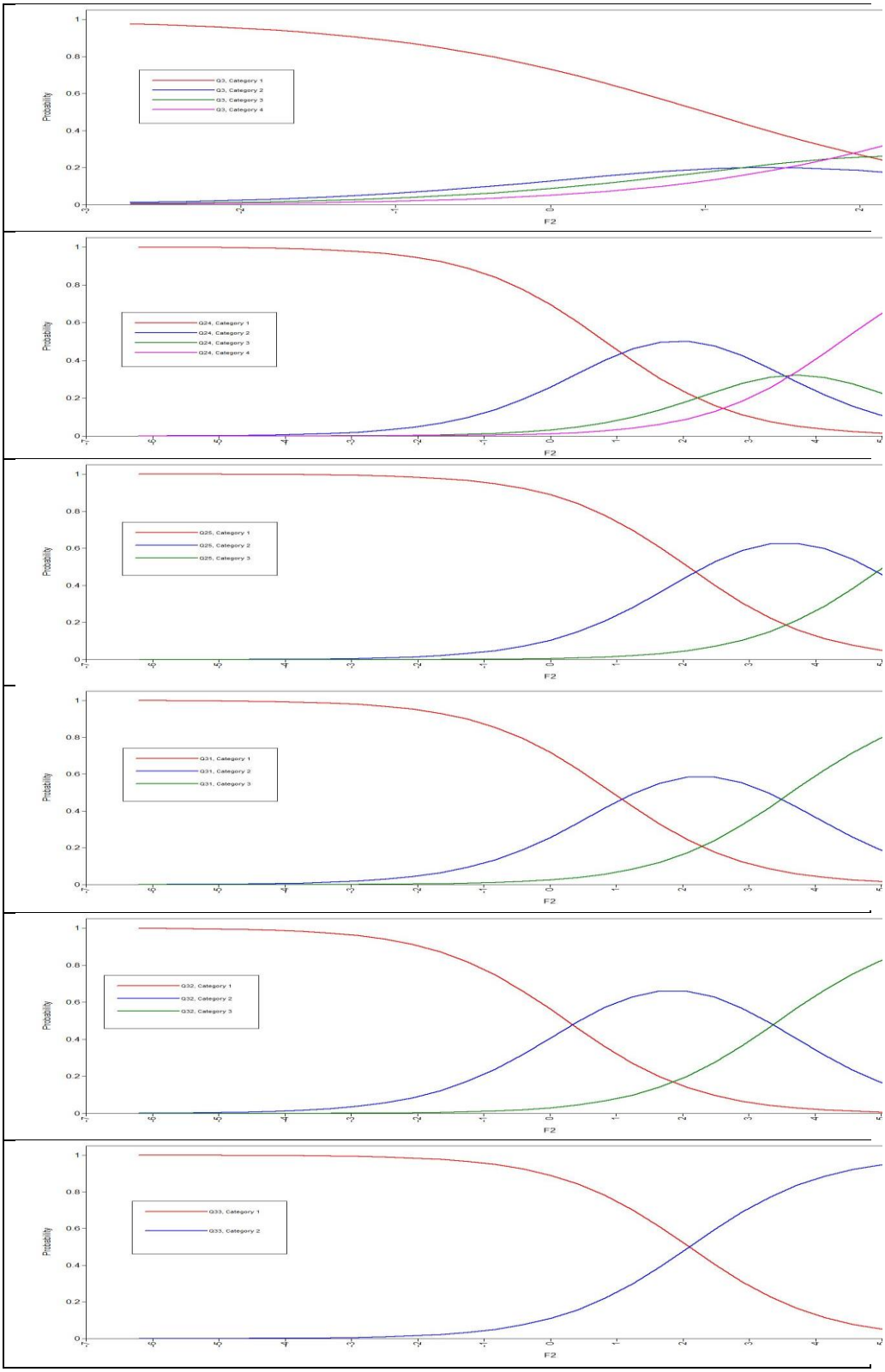
- Right amount
- Not enough/too much

Figure 7 shows curved lines depicting the category response probabilities of these items. The GPCM model for each item describes three threshold parameters and they are:

- The slope parameter describes the item's ability to discriminate; the items with higher slope parameters or steeper trace lines are better at discriminating between good and bad factor two than items with lower slope parameter.

- If vertical straight lines were to be drawn at the population mean or from IRT score of 0 in the horizontal axis, the probability of estimating different item responses can be estimated.
- The IRT score can be regarded as the scale score that would have been observed if there was no floor and ceiling effect and all items had their optimal weight.

Figure 7: Item trace line of factor 2 from subgroup 1



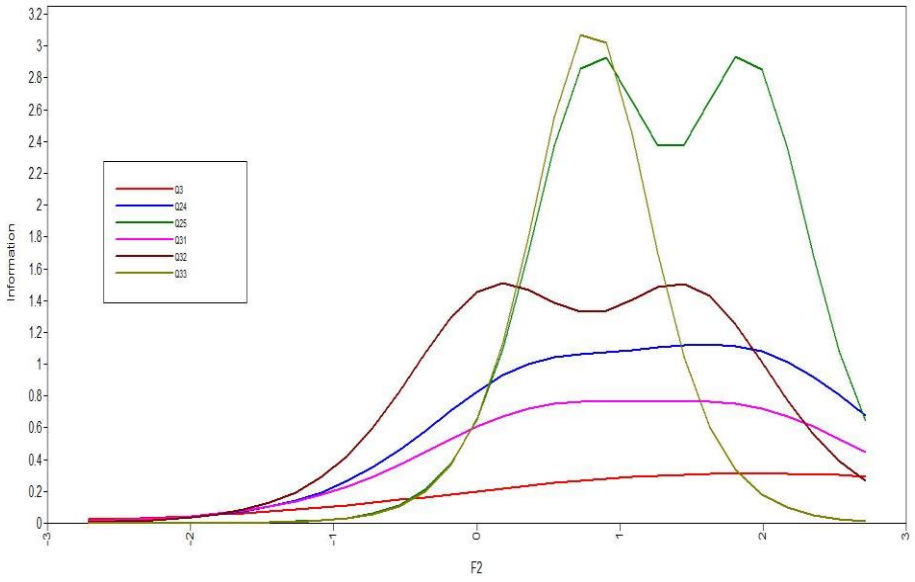
It is possible to estimate IRT models in different ways. In the first set of results, all loadings are set to 1 and the variance of the latent variable is freely estimated. The mean of the latent variable is constrained to 0. In the IRT parameterisation, the latent variance is constrained to 1 and the item discrimination is estimated, but still constrained to be equal across items. The item difficulty parameters are calculated as threshold/discrimination. The findings for factor 2 from subgroup 1 adopting above approach are presented in Table 13.

Table 13: IRT results of factor 2 in subgroup 1

	Thresholds	Item Difficulties
Q3. While you were in the A&E Department, how much information about your condition was given to you?		
Right amount	0.521	1.176
Not enough/too much	0.924	2.086
Don't know/can't remember	1.462	3.300
Not given any information	Omitted	Omitted
Q24. When you had important questions to ask a doctor, did you get answers that you could understand?		
Yes, always	0.350	0.511
Yes, sometimes	1.222	1.784
I had no need to ask	1.668	2.435
No	Omitted	Omitted
Q25. Did you have confidence and trust in the doctors treating you?		
Yes, always	0.851	0.984
Yes, sometimes	1.874	2.166
No	Omitted	Omitted
Q31. Sometimes in a hospital, a member of staff will say one thing and another will say something quite different. Did this happen to you?		
Yes, always	0.390	0.607
Yes, sometimes	1.392	2.168
No	Omitted	
Q32. Were you involved as much as you wanted to be in decisions about your care and treatment?		
Yes, definitely	0.163	0.211
Yes, to some extent	1.305	1.686
No	Omitted	Omitted
Q33. How much information about your condition/treatment was given to you?		
Right amount	0.830	0.962
Not enough/too much	Omitted	Omitted

Item information functions (IIF) were calculated for all items. The IIF is a function of the standard error of the latent score estimate and is a measure of how much information the item provides about the person's score for various levels of the factor. It is illustrated in Figure 8 and the largest information in the area was provided by Q25 and Q33. They provide the most accurate estimation of the overall scale.

Figure 8: Item information curve as a function of Factor 2 in subgroup 1



The final one-factor IRT model estimated for factor 2 is presented in Table 14. The standardised loading is provided and the high values suggest that the single factor model provides a good explanation for all variables especially for item Q25. The question associated with Q25 was on confidence and trust the patient has on the doctors treating the patient. The response categories were based on frequency, namely 'always', 'sometimes' and 'no'.

Table 14: Item discrimination of items from factor 2 in subgroup 1

	Factor 2	Estimate	S.E	P-value
Q3	While you were in the A&E Department, how much information about your condition was given to you?	0.443	0.011	0.00
Q24	When you had important questions to ask a doctor, did you get answers that you could understand?	0.685	0.008	0.00
Q25	Did you have confidence and trust in the doctors treating you?	0.865	0.006	0.00
Q31	Sometimes in a hospital, a member of staff will say one thing and another will say something quite different. Did this happen to you?	0.642	0.008	0.00
Q32	Were you involved as much as you wanted to be in decisions about your care and treatment?	0.774	0.007	0.00
Q33	How much information about your condition or treatment was given to you?	0.863	0.007	0.00

7.3.2 IRT models

Similarly, IRT model were fitted in each factor across the two subgroups: subgroup 1 comprising of patients who had an operation or procedure and had an emergency admission; subgroup 2 comprising of patients who had an operation or procedure and a planned admission. The results from the IRT models are presented in Table 15 for subgroup 1 and Table 16 for subgroup 2.

Table 15: Estimated factor loadings in IRT model in subgroup 1

Factor	Independent Items	Estimate	S.E.
F1	In your opinion, how clean was the hospital room or ward that you were in?	1.000	0.000
	How clean were the toilets and bathrooms that you used in hospital?	0.781	0.006
F2	While you were in the A&E Department, how much information about your condition was given to you?	0.443	0.011
	When you had important questions to ask a doctor, did you get answers that you could understand?	0.685	0.008
	Did you have confidence and trust in the doctors treating you?	0.865	0.006
	Sometimes in a hospital, a member of staff will say one thing and another will say something quite different. Did this happen to you?	0.642	0.008
	Were you involved as much as you wanted to be in decisions about your care and treatment?	0.774	0.007
	How much information about your condition or treatment was given to you?	0.863	0.007
F3	Were you ever bothered by noise at night from other patients?	0.517	0.013
	Were you ever bothered by noise at night from hospital staff?	1.000	0.018
F4	Beforehand, did a member of staff explain the risks and benefits of the operation or procedure in a way you could understand?	0.889	0.004
	Beforehand, did a member of staff explain what would be done during the operation or procedure?	0.917	0.004
	Beforehand, did a member of staff answer your questions about the operation or procedure?	0.794	0.005
	Beforehand, were you told how you could expect to feel after you had the operation or procedure?	0.778	0.006
	After the operation or procedure, did a member of staff explain how the operation or procedure had gone in a way you could understand?	0.708	0.007
F5	Did a member of staff explain the purpose of the medicines you were to take at home in a way you could understand?	0.883	0.004
	Did a member of staff tell you about medication side effects to watch for when you went home?	0.735	0.006
	Were you told how to take your medication in a way you could understand?	0.889	0.004
	Were you given clear written or printed information about your medicines?	0.738	0.006

Factor	Independent Items	Estimate	S.E.
	Did a member of staff tell you about any danger signals you should watch for after you went home?	0.621	0.007
F6	Did you find someone from the hospital staff to talk to about your worries and fears?	0.767	0.040
	Do you feel you got enough emotional support from hospital staff during your stay?	0.787	0.041
F7	Did doctors talk in front of you as if you weren't there?	0.772	0.044
	Did nurses talk in front of you as if you weren't there?	0.902	0.050
F8	Were you given enough privacy when discussing your condition or treatment?	0.800	0.007
	Were you given enough privacy when being examined or treated?	0.999	0.000
F9	Did hospital staff take your family or home situation into account when planning your discharge?	0.554	0.010
	Did hospital staff discuss with you whether you would need any additional equipment in your home, or any adaptations made to your home, after leaving hospital?	0.795	0.010
	Did hospital staff discuss with you whether you may need any further health or social care services after leaving hospital?	0.707	0.010
F10	Did you feel you were involved in decisions about your discharge from hospital?	0.639	0.007
	Were you given enough notice about when you were going to be discharged?	0.985	0.000
F11	During your hospital stay, were you ever asked to give your views on the quality of your care?	0.736	0.249
	Did you see, or were you given, any information explaining how to complain to the hospital about the care you received?	0.754	0.459

Table 16: Estimated factor loadings in IRT model in subgroup 2

Factor	Independent Items	Estimate	S.E.
F1	Were you ever bothered by noise at night from other patients?	0.064	0.012
	Were you ever bothered by noise at night from hospital staff?	0.131	0.012
	Did doctors talk in front of you as if you weren't there?	0.529	0.010
	Did you have confidence and trust in the nurses treating you?	0.404	0.014
	Did nurses talk in front of you as if you weren't there?	0.696	0.009
	In your opinion, were there enough nurses on duty to care for you in hospital?	0.719	0.010
	Sometimes in a hospital, a member of staff will say one thing and another will say something quite different. Did this happen to you?	0.550	0.014
	Were you given enough privacy when discussing your condition or treatment?	0.624	0.011
	Were you given enough privacy when being examined or treated?	0.683	0.010
	Overall, did you feel you were treated with respect and dignity while you were in the hospital?	0.263	0.010
F2	In your opinion, how clean was the hospital room or ward that you were in?	0.706	0.108
	How clean were the toilets and bathrooms that you used in hospital?	0.828	0.113
F3	Beforehand, did a member of staff explain the risks and benefits of the operation or procedure in a way you could understand?	0.173	0.009
	Beforehand, did a member of staff explain what would be done during the operation or procedure?	0.177	0.009
	Beforehand, did a member of staff answer your questions about the operation or procedure?	0.858	0.006
	Beforehand, were you told how you could expect to feel after you had the operation or procedure?	0.896	0.005
	Before the operation or procedure, did a member of staff explain how he or she would put you to sleep or control your pain in a way you could understand?	0.771	0.006
	After the operation or procedure, did a member of staff explain how the operation or procedure had gone in a way you could understand?	0.697	0.007
F4	Did a member of staff explain the purpose of the	0.108	0.01

Factor	Independent Items	Estimate	S.E.
	medicines you were to take at home in a way you could understand?		
	Did a member of staff tell you about medication side effects to watch for when you went home?	0.589	0.011
	Were you told how to take your medication in a way you could understand?	0.846	0.011
	Were you given clear written or printed information about your medicines?	0.678	0.009
F5	Did you find someone on the hospital staff to talk to about your worries and fears?	0.920	0.031
	Do you feel you got enough emotional support from hospital staff during your stay?	0.699	0.024
F6	Did you feel you were involved in decisions about your discharge from hospital?	0.548	0.104
	Were you given enough notice about when you were going to be discharged?	0.72	0.143
F7	Did hospital staff discuss with you whether you would need any additional equipment in your home, or any adaptations made to your home, after leaving hospital?	0.522	0.035
	Did hospital staff discuss with you whether you may need any further health or social care services after leaving hospital?	0.871	0.053
F8	Did a member of staff tell you about any danger signals you should watch for after you went home?	0.245	0.009
	During your hospital stay, were you ever asked to give your views on the quality of your care?	0.831	0.014
	Did you see, or were you given, any information explaining how to complain to the hospital about the care you received?	0.848	0.015

7.3.3 Factor analysis models

One-factor models were fitted to the polychoric correlation matrix of the items identified for each factor in each subgroup. The parameters of the model were estimated using weighted least squares. The estimated factor loadings and standard errors estimated using factor analysis method are presented in Table 17 and

Table 18.

Table 17: Estimated item loadings in factor analysis model in subgroup 1

	Independent Items	Estimate	S.E.
F1	In your opinion, how clean was the hospital room or ward that you were in?	1.000	0.000
	How clean were the toilets and bathrooms that you used in hospital?	0.894	0.013
F2	While you were in the A&E Department, how much information about your condition was given to you?	1.000	0.000
	When you had important questions to ask a doctor, did you get answers that you could understand?	1.474	0.034
	Did you have confidence and trust in the doctors treating you?	1.773	0.041
	Sometimes in a hospital, a member of staff will say one thing and another will say something quite different. Did this happen to you?	1.354	0.034
	Were you involved as much as you wanted to be in decisions about your care and treatment?	1.706	0.039
	How much information about your condition or treatment was given to you?	1.793	0.040
F3	Were you ever bothered by noise at night from other patients?	1.000	0.000
	Were you ever bothered by noise at night from hospital staff?	1.290	0.039
F4	Beforehand, did a member of staff explain the risks and benefits of the operation or procedure in a way you could understand?	1.000	0.000
	Beforehand, did a member of staff explain what would be done during the operation or procedure?	0.986	0.007
	Beforehand, did a member of staff answer your questions about the operation or procedure?	0.874	0.007
	Beforehand, were you told how you could expect to feel after you had the operation or procedure?	0.923	0.008
	After the operation or procedure, did a member of staff explain how the operation or procedure had gone in a way you could understand?	0.936	0.008
F5	Did a member of staff explain the purpose of the medicines you were to take at home in a way you could understand?	1.000	0.000
	Did a member of staff tell you about medication side effects to watch for when you went home?	0.949	0.009
	Were you told how to take your medication in a way you could understand?	0.954	0.008
	Were you given clear written or printed information about your medicines?	0.833	0.009

	Independent Items	Estimate	S.E.
	Did a member of staff tell you about any danger signals you should watch for after you went home?	0.918	0.009
F6	Did you find someone from the hospital staff to talk to about your worries and fears?	1.000	0.000
	Do you feel you got enough emotional support from hospital staff during your stay?	1.182	0.018
F7	Did doctors talk in front of you as if you weren't there?	1.000	0.000
	Did nurses talk in front of you as if you weren't there?	1.031	0.019
F8	Were you given enough privacy when discussing your condition or treatment?	1.000	0.000
	Were you given enough privacy when being examined or treated?	1.040	0.015
F9	Did hospital staff take your family or home situation into account when planning your discharge?	1.000	0.000
	Did hospital staff discuss with you whether you would need any additional equipment in your home, or any adaptations made to your home, after leaving hospital?	0.620	0.015
	Did hospital staff discuss with you whether you may need any further health or social care services after leaving hospital?	0.691	0.016
F10	Did you feel you were involved in decisions about your discharge from hospital?	1.000	0.000
	Were you given enough notice about when you were going to be discharged?	1.081	0.013
F11	During your hospital stay, were you ever asked to give your views on the quality of your care?	1.000	0.000
	Did you see, or were you given, any information explaining how to complain to the hospital about the care you received?	1.209	0.037

Table 18: Estimated factor loadings in factor analysis model in subgroup 2

Factor	Independent Items	Estimate	S.E.
F1	Were you ever bothered by noise at night from other patients?	1.000	0.000
	Were you ever bothered by noise at night from hospital staff?	1.206	0.027
	Did doctors talk in front of you as if you weren't there?	1.309	0.033
	Did you have confidence and trust in the nurses treating you?	1.881	0.039
	Did nurses talk in front of you as if you weren't there?	1.497	0.035
	In your opinion, were there enough nurses on duty to care for you in hospital?	1.477	0.032
	Sometimes in a hospital, a member of staff will say one thing and another will say something quite different. Did this happen to you?	1.487	0.032
	Were you given enough privacy when discussing your condition or treatment?	1.734	0.036
	Were you given enough privacy when being examined or treated?	1.815	0.041
	Overall, did you feel you were treated with respect and dignity while you were in the hospital?	2.091	0.043
F2	In your opinion, how clean was the hospital room or ward that you were in?	1.000	0.000
	How clean were the toilets and bathrooms that you used in hospital?	0.900	0.011
F3	Beforehand, did a member of staff explain the risks and benefits of the operation or procedure in a way you could understand?	1.000	0.000
	Beforehand, did a member of staff explain what would be done during the operation or procedure?	0.949	0.008
	Beforehand, did a member of staff answer your questions about the operation or procedure?	0.839	0.008
	Beforehand, were you told how you could expect to feel after you had the operation or procedure?	0.923	0.008
	Before the operation or procedure, did a member of staff explain how he or she would put you to sleep or control your pain in a way you could understand?	0.595	0.011
	After the operation or procedure, did a member of staff explain how the operation or procedure had gone in a way you could understand?	0.913	0.009
F4	Did a member of staff explain the purpose of the	1.000	0.000

Factor	Independent Items	Estimate	S.E.
	medicines you were to take at home in a way you could understand?		
	Did a member of staff tell you about medication side effects to watch for when you went home?	0.943	0.007
	Were you told how to take your medication in a way you could understand?	0.965	0.006
	Were you given clear written or printed information about your medicines?	0.829	0.006
F5	Did you find someone on the hospital staff to talk to about your worries and fears?	1.000	0.000
	Do you feel you got enough emotional support from hospital staff during your stay?	1.171	0.016
F6	Did you feel you were involved in decisions about your discharge from hospital?	1.000	0.000
	Were you given enough notice about when you were going to be discharged?	1.098	0.013
F7	Did hospital staff discuss with you whether you would need any additional equipment in your home, or any adaptations made to your home, after leaving hospital?	1.000	0.000
	Did hospital staff discuss with you whether you may need any further health or social care services after leaving hospital?	1.128	0.031
F8	Did a member of staff tell you about any danger signals you should watch for after you went home?	1.000	0.000
	During your hospital stay, were you ever asked to give your views on the quality of your care?	0.554	0.013
	Did you see, or were you given, any information explaining how to complain to the hospital about the care you received?	0.697	0.012

7.3.4 Comparison of estimates from IRT and FA models

The items selected based on relative values indicated by the IRT and underlying variable model for each one-factor model for patients who underwent an operation or procedure are summarized in Table 19 for patients with emergency admission and Table 20 for patients with planned admission. In terms of interpretability, the items selected using the two approaches based on relative value of estimate were comparable. The items may not have been exact but the underlying concept was very

similar, and this is not surprising given that the items were grouped together to be unidimensional.

There were some dimensions in which items selected by the two models were considerably different. For instance, IRT chose the question ‘did you have confidence and trust in the doctors treating you?’ while the underlying variable approach selected the question ‘was enough information given to you about your condition or treatment?’ for patients with emergency admission. A positive response to both would comfort the patient but one can be called attitude based while the other is action specific. The second item selection, which differed between the methods used, was related to discharge of patients with emergency admission. IRT model considered the item about need for any additional equipment in your home, or any adaptations made to your home to be the most discriminating. While with factor model the question on whether or not the hospital staff took family or home situation into account when planning discharge was selected. Amongst patients with planned admission, four of the items selected differed and are listed below. There was no clear pattern and it is not possible to judge items selected from one approach as better than the other based on item content.

Table 19: Item selection for patients with emergency admission who underwent an operation or procedure

IRT	Factor analysis model
In your opinion, how clean was the hospital room or ward that you were in?	In your opinion, how clean was the hospital room or ward that you were in?
Did you have confidence and trust in the doctors treating you?	How much information about your condition or treatment was given to you?
Were you ever bothered by noise at night from hospital staff?	Were you ever bothered by noise at night from hospital staff?
Beforehand, did a member of staff explain what would be done during the operation or procedure?	Beforehand, did a member of staff explain the risks and benefits of the operation or procedure in a way you could understand?
Were you told how to take your medication in a way you could understand?	Did a member of staff explain the purpose of the medicines you were to take at home in a way you could understand?
Do you feel you got enough emotional support from hospital staff during your stay?	Do you feel you got enough emotional support from hospital staff during your stay?
Did nurses talk in front of you as if you weren't there?	Did nurses talk in front of you as if you weren't there?
Were you given enough privacy when being examined or treated?	Were you given enough privacy when being examined or treated?
Did hospital staff discuss with you whether you would need any additional equipment in your home, or any adaptations made to your home, after leaving hospital?	Did hospital staff take your family or home situation into account when planning your discharge?
Were you given enough notice about when you were going to be discharged?	Were you given enough notice about when you were going to be discharged?
Did you see, or were you given, any information explaining how to complain to the hospital about the care you received?	Did you see, or were you given, any information explaining how to complain to the hospital about the care you received?

Table 20: Item selection for patients with planned admission who underwent an operation or procedure

IRT	Factor analysis model
In your opinion, were there enough nurses on duty to care for you in hospital?	Overall, did you feel you were treated with respect and dignity while you were in the hospital?
How clean were the toilets and bathrooms that you used in hospital?	In your opinion, how clean was the hospital room or ward that you were in?
Beforehand, were you told how you could expect to feel after you had the operation or procedure?	Beforehand, did a member of staff explain the risks and benefits of the operation or procedure in a way you could understand?
Were you told how to take your medication in a way you could understand?	Did a member of staff explain the purpose of the medicines you were to take at home in a way you could understand?
Did you find someone on the hospital staff to talk to about your worries and fears?	Do you feel you got enough emotional support from hospital staff during your stay?
Were you given enough notice about when you were going to be discharged?	Were you given enough notice about when you were going to be discharged?
Did hospital staff discuss with you whether you may need any further health or social care services after leaving hospital?	Did hospital staff discuss with you whether you may need any further health or social care services after leaving hospital?
Did you see, or were you given, any information explaining how to complain to the hospital about the care you received?	Did a member of staff tell you about any danger signals you should watch for after you went home?

7.3.5 SEM application

SEM was applied to the measurement model to investigate item and dimension selection based on association with the dependent variable. For item selection in each dimension, the latent variable estimated as continuous underlying variable was specified as dependent variable and the observed items as independent variable. The one-factor SEM models using this approach were not identified and standard error were not estimated in any of the dimensions. As a second approach for item selection in each dimension, the overall patient experience measured using rating scale (described in chapter 5) was used as dependent variable. Again the one-factor SEM models fitted in observed variables did not converge. What became evident was the very high correlation between the items specified as independent variables. And that it is inappropriate to use SEM in factors estimated using factor analysis.

For dimension selection, I explored two approaches. In the first instance, I specified the latent factors as latent independent variables and then used the overall latent construct of the measure as the dependent variable. These SEM models, with latent variables on both sides, were not identified as the models did not converge. A reduced model was fitted instead using the items selected from factor analysis and IRT as observed explanatory variables and the latent construct measured by them specified as the dependent variable. The findings from these models were used to order the items in terms of contribution to overall variance in the dependent variable.

The item for each subgroup model is ordered based on the magnitude of the coefficient, with the most significantly associated being listed on the top row and least relevant listed on bottom row, in Table 21 and

Table 22. Each item in the two models was statistically significant and the R^2 value of the model was just over 0.50 in each subgroup. The ordering of factors using SEM is fairly reasonable. There was one factor, which can be considered trivial but was given a large score in the SEM model. This was the question on being given enough notice about discharge. It is possible that being able to plan discharge, including transition and continuity is an important consideration for patients who have just had an operation or procedure. The difference in scores between the items is minimal here and it is difficult to make a judgment on item reduction based on it.

Table 21: SEM result for patients with emergency admission who underwent an operation or procedure

Factor	Estimate	Item selected using IRT
F2	0.78	Did you have confidence and trust in the doctors treating you?
F8	0.764	Were you given enough privacy when being examined or treated?
F1	0.635	In your opinion, how clean was the hospital room or ward that you were in?
F10	0.631	Were you given enough notice about when you were going to be discharged?
F7	0.589	Did nurses talk in front of you as if you weren't there?
F6	0.566	Do you feel you got enough emotional support from hospital staff during your stay?
F4	0.535	Beforehand, did a member of staff explain what would be done during the operation or procedure?
F3	0.519	Were you ever bothered by noise at night from hospital staff?
F5	0.481	Were you told how to take your medication in a way you could understand?
F11	0.388	Did you see, or were you given, any information explaining how to complain to the hospital about the care you received?
F9	0.264	Did hospital staff discuss with you whether you would need any additional equipment in your home, or any adaptations made to your home, after leaving hospital?

Factor	Estimate	Item selected using factor analysis
F2	0.829	Was enough information given to you about your condition or treatment?
F10	0.665	Were you given enough notice about when you were going to be discharged?
F8	0.596	Were you given enough privacy when being examined or treated?
F4	0.546	Beforehand, did a member of staff explain the risks and benefits of the operation or procedure in a way you could understand?
F5	0.478	Did a member of staff explain the purpose of the medicines you were to take at home in a way you could understand?
F7	0.445	Did nurses talk in front of you as if you weren't there?
F6	0.433	Do you feel you got enough emotional support from hospital staff during your stay?
F1	0.385	In your opinion, how clean was the hospital room or ward that you were in?
F9	0.285	Did hospital staff take your family or home situation into account when planning your discharge?
F3	0.269	Were you ever bothered by noise at night from hospital staff?
F11	0.259	Did you see, or were you given, any information explaining how to complain to the hospital about the care you received?

Table 22: SEM result for patients with planned admission who underwent an operation or procedure

Factor	Estimate	Item selected using IRT
F3	0.613	Beforehand, were you told how you could expect to feel after you had the operation or procedure?
F6	0.604	Were you given enough notice about when you were going to be discharged?
F1	0.489	In your opinion, were there enough nurses on duty to care for you in hospital?
F2	0.471	How clean were the toilets and bathrooms that you used in hospital?
F4	0.47	Were you told how to take your medication in a way you could understand?
F8	0.465	Did you see, or were you given, any information explaining how to complain to the hospital about the care you received?
F5	0.442	Did you find someone on the hospital staff to talk to about your worries and fears?
F7	0.362	Did hospital staff discuss with you whether you may need any further health or social care services after leaving hospital?
Factor	Estimate	Item selected using factor analysis
F8	0.658	Did a member of staff tell you about any danger signals you should watch for after you went home?
F6	0.599	Were you given enough notice about when you were going to be discharged?
F1	0.573	Overall, did you feel you were treated with respect and dignity while you were in the hospital?
F3	0.562	Beforehand, did a member of staff explain the risks and benefits of the operation or procedure in a way you could understand?
F2	0.421	In your opinion, how clean was the hospital room or ward that you were in?
F4	0.388	Did a member of staff explain the purpose of the medicines you were to take at home in a way you could understand?
F5	0.387	Do you feel you got enough emotional support from hospital staff during your stay?
F7	0.224	Did hospital staff discuss with you whether you may need any further health or social care services after leaving hospital?

7.4 Discussion

The study presented in this chapter generated an 11-item descriptive system for patients with emergency admission that had an operation or procedure from an 11-factor model developed in chapter 7. And an 8-item measure for patients with planned admission that had an operation or procedure from an 8-factor model was also presented. Each factor measures a single latent construct (recall that this was assessed used EFA and CFA earlier) and item selection conducted in this chapter focussed on selecting one item for each factor.

7.4.1 Comparison of methods used for item selection

IRT and underlying variable approach are capable of handling polytomous dataset and are recommended methods for item selection (Bartholomew et al., 2008). The factor loadings for the factor analysis model are correlations between a normal latent variable and the normal underlying variables, whereas for the IRT logit model the standardised loadings are correlations between the normal latent variable and underlying variables that are not normally distributed (Bartholomew et al. pp 258). In factor analysis, the model is fitted by choosing the parameter values to make the covariance matrix predicted by the model as close as possible to the observed matrix. A similar process is followed in IRT where the items are fitted to estimate parameter values which make the frequency distribution across responses predicted by the model as close as possible to the observed one (Bartholomew et al. pp 216).

The models in both approaches can be estimated using various techniques. However, the IRT analysis software currently available is based on the

maximum likelihood function. The obvious difference between the IRT and factors models used in my study is that the factor loadings for the IRT model were estimated using maximum likelihood and factor analysis models used weighted least squares method. In studies where the matrix of polychoric correlations are used to estimate the models, use of ML is known to produce erroneous standard error estimates and chi square based fit measures when applied to correlation matrices (Cudeck, 1989). The models fitted within both approaches employed a logit distribution and displayed good fit.

Bartholomew and Knott (1999) argue that although models in IRT and underlying variable approach look quite distinct in terms of model fitting procedure and some of the model assumptions, there is equivalence between the two approaches (Bartholomew and Knott, 1999). There is an exact equivalence between the parameter estimated using normit factor analysis and normit IRT (see Bartholomew et al. pp 225). If probit or normit IRT model was used instead of the logit, the results would have perhaps been much closer. However the estimates using logit models were quite different in my study. More importantly the relative values estimated using the two approaches were different and generated different items to be included in the patient experience descriptive system.

Which approach is preferred? IRT determines item discrimination at various levels of the latent variable by graphically examining item response functions. Bartholomew et al. prefer IRT and they describe it as the full information method that utilises the full distribution across all the categories. And the underlying variable approach is considered a partial information model that uses information only from the pairwise distribution of the ordinal variables. It is grounded in the factor analysis

tradition, and provides ease in interpretation, where standardised alphas can be interpreted as correlations. It should also be noted that in the literature factor analysis and IRT are seen as complementary approaches, with IRT better suited for examining item characteristics and factor analysis as more appropriate for multidimensional model testing.

I also considered SEM for item selection but it is not possible in a one-factor model. None of the SEM models using the items in each factor as independent variables converged, regardless of the dependent variable chosen; the reason being high correlation between items. It should be noted that items in a unidimensional model are grouped together because they are associated with the same single latent construct.

7.4.2 Determining number of items in a descriptive system

SEM can be used to gain insight into how much of the overall variance, measured using observed overall rating, was explained by each of the selected variables in the dataset. This information may be particularly useful when the descriptive system contains many items and item reduction is necessary to create a brief instrument. I applied SEM model to the IRT and factor analysis findings on item selection to order the items in terms of contribution to overall variance of the dependent variable, which was patient experience. While this approach is logical, at this stage I feel further data collection is required rather than extending reliance on statistical techniques. My thesis has used only one large secondary dataset and employed a number of advanced statistical models, however in order to finalise the item selection further validity and reliability tests must be conducted.

The face validity of the new instruments should be confirmed using patients and healthcare professionals. Also cross validation of the hypothetical structure of the dataset should be conducted using a second sample is important. It is common practice to use one sample to calibrate the proposed structure of the data and a second independent sample to validate the structure identified using CFA. Inpatient Survey data from another year, say 2015 can be used for this purpose. Finally hypothesis testing to examine convergent and divergent validity of this measure with other measures can be carried out.

7.4.3 Using judgement to determine item selection

In addition to reliance on psychometric assessment, I think an instrument developer has to be more nuanced about the need for single item and apply judgment where necessary. In this study I combined near identical items to avoid the loss of breadth which selecting an item per dimension. The caveat to using judgement is that it introduces some degree of subjectivity, however it is worth apply it. For example the factor on cleanliness comprised of an item assessing the cleanliness of the ward and another item asking about the cleanliness of the toilets and bathrooms. IRT results indicate using the latter item, however this would result in a question that is too specific. And this may be problematic because it would not be able to a capture scenarios in which the toilets are clean toilets but the ward is not, and affects patient experience. The key variable in the factor is cleanliness and should be retained. I have combined items that are almost identical for four factors with slight amendments to the wording of the original questions. These are presented in Table 24.

One could argue that the items from other factors are measuring a single latent variable, such as aspects related to medication and discharge, and

should be combined in a similar manner. I have relied on IRT findings in the other factors because the questions within this are not similar to this extent and cannot be combined by a simple change in wording that broadens the item. The item selection in other factors focussed on finding an item that is most representative of the construct being measured.

Table 23: Items that were combined by amending wording

Original Items	Combined items
In your opinion, how clean was the hospital room or ward that you were in?	In your opinion, how clean were the hospital facilities?
How clean were the toilets and bathrooms that you used in hospital?	
Were you ever bothered by noise at night from other patients?	Were you ever bothered by noise at night from patients or hospital staff?
Were you ever bothered by noise at night from hospital staff?	
Did doctors talk in front of you as if you weren't there?	Did the doctors or the nurses talk in front of you as if you weren't there?
Did nurses talk in front of you as if you weren't there?	
Were you given enough privacy when discussing your condition or treatment?	Were you given enough privacy at the hospital?
Were you given enough privacy when being examined or treated?	

7.4.4 A descriptive system to measure patient experience

The objective of the study was to select one item per dimension, and an important consideration was to produce a brief instrument. Items were selected using two approaches: IRT model from modern test theory and factor analysis model (underlying variable models) from classical test theory. Both methods are sound and produced very similar results, but I have selected IRT findings for the final instrument. The reason being the use of complete item information by IRT compared to factor analysis; and the focus being on differentiation.

The items selected for patients who underwent an operation or procedure using emergency admission was different to those who had planned admission. This may not be surprising given that the number of dimensions were different between the two groups. In patients who had an emergency focus the items considered most important were regarding information about condition or treatment and privacy. While for those who had a planned admission the focus was on discharge and after care following discharge. It should be noted that patients who are likely to have planned admission are older and those with chronic conditions, hence transition and continuity of care would be a critical consideration in this group of patients.

Privacy and noise at night was a concern for patients with emergency focus but not the planned admission patients. This could be because of the unexpected entry to hospital and being wearier because of it. Based on the items selected, it seems like patients with emergency admission place more weight on soft skills of healthcare staff such as communication. For example, confidence and trust in the doctors and not having healthcare staffs talk in front of the patient as if they were not there were important for emergency patients but not for planned admission patients. Perhaps issues about noise, privacy and not having nurses talk in front of them are all related to the reality of being a patient in an A&E ward, i.e. such wards often really are noisier, have less privacy, and might be more likely to have nurses urgently talking in front of the patients. That might explain why these issues are more important to them.

The items that were common across the two subgroups were: cleanliness of the hospital facility; being told how to take medication in a comprehensive manner; emotional support from hospital staff; being given enough notice about discharge time; provision for after leaving hospital and opportunity to provide feedback about care received. Just to reiterate the point I made in chapter 6 (see Table 12), the items selected for the descriptive system for patient experience fall within five of the eight dimensions of care stated in the Patient Experience Framework. While this could be specific to Inpatient stay or the NHS Inpatient Stay 2014 dataset, there is a need to examine the hypothetical constructs of patient experience using contemporary survey responses.

7.4.5 Implications of separate measurement models for subgroups

At the onset, I planned to create a measure of patient experience that can be applied to all patients with an inpatient stay. However a thorough investigation of the dataset, detailed in chapter 5, resulted in knowledge of questions that was not applicable to all and the division of the data into four subgroups. In chapter 5, non-convergence of the dataset for two of the four subgroups was stated. This led to two subgroups being dropped and it comprised of patients who did not have an operation or procedure carried out in the hospital.

While I am aware that the current trend is towards creating a measure that is generic and my decision to use of subgroups was data driven, the insight brought by this approach has been very useful. The similarities and differences between inpatient patients based on two main routes of admission is quite novel and one that would have been missed if I had insisted on creating a universal measure of patient experience.

7.4.6 A measure that is amenable to valuation

A systematic review of measures of patient experience in hospitals identified 11 instruments (Beattie et al., 2015). None of these measures are amenable to valuation. This was a key consideration in this thesis. The two measures of patient experience that I have constructed for patients with emergency admission and planned admission are 11-item and 8-item respectively. Both measures have good construct validity. However the number of items included may be too many still. SEM was applied in this study to order the items elicited using IRT and factor analysis in terms of importance. It is possible to draft a 5-item measure for each subgroup using the top five items associate with the overall construct. A descriptive system for patients with emergency admission, who underwent an operation or procedure includes following items, using IRT findings are:

- Did you have confidence and trust in the doctors treating you?
- Were you given enough privacy at the hospital?
- In your opinion, how clean were the hospital facilities?
- Were you given enough notice about when you were going to be discharged?
- Did the doctors or the nurses talk in front of you as if you weren't there?

Since the descriptive system for patients with planned admission, who underwent an operation or procedure includes following IRT items:

- Beforehand, were you told how you could expect to feel after you had the operation or procedure?
- Were you given enough notice about when you were going to be discharged?

- In your opinion, were there enough nurses on duty to care for you in hospital?
- In your opinion, how clean was the hospital facilities (including toilet)?
- Were you told how to take your medication in a way you could understand?

The two items that are common across the two groups are cleanliness of the hospital facility and being given enough notice about discharge. In addition the overall experience of patients with emergency admission was influenced by confidence and trust in the doctors; privacy; and not having the doctors or the nurses not talk in front of the patient as if they weren't there. While patients with planned admission valued being told what to expect after the operation or procedure; having enough nurses on duty; and being told how to take medication in a comprehensive way.

In patients who had unplanned admission, the interpersonal skills were highlighted as being the most relevant while in patients with planned admission, the more functional items were highlighted such as information and coordination/integration of care. The reason could be that first group of patients did not have time to prepare, were more weary and focussed more on how they were being treated. The latter group had time to prepare themselves for the hospital visit and they were more concerned about the practical aspects of care.

The five items identified by SEM can be used to construct hypothetical healthcare states for preference elicitation studies. However my preference still would be validate item selection and test for further

psychometric properties using patients and healthcare professionals before to inform final selection. This is discussed further in chapter 8.

7.5 Conclusion

In this chapter item selection was conducted using IRT and underlying variable approach. The IRT and factor analytic technique produced slightly dissimilar but comparable results. It should be noted that each dimension was already unidimensional hence the latent construct being captured in each dimension by the items is the same. In terms of which method was better, the IRT approach is preferred.

By ensuring that each item selected in each dimension summarises the latent construct being measured and all the dimensions reflect the hypothetical structure of the responses, the two-staged study presented over two chapters (6 and 7) is able to ensure that the Inpatient Survey 2014 is summarised by a reduced number of items. The items selected broadly fell under respect for the patient, information and communication, physical comfort, emotional support and transition and continuity.

Chapter 8 Discussions

The overarching aim of this thesis was to investigate methods used for the development of a descriptive system and use the findings to construct a brief descriptive system to measure patient experience that is amenable to valuation. The third and fourth chapter of my thesis studied existing literature and guidance documents on methods used to develop a descriptive system of an instrument, and chapters 5 to 7 presented empirical analyses of a large secondary dataset using the methods identified from the literature.

In this chapter I discuss the strengths and limitations of the patient experience measure, then highlight the conceptual and methodological contributions of this thesis. Finally, recommendations for further research are made.

8.1 Overview of the thesis

I chose the NHS Inpatient survey, which is one of the patient surveys that the CQC publishes, and focuses on hospital stay. It is a comprehensive dataset of patient experience and presents recent NHS users with nearly seventy aspects of healthcare delivery to evaluate. This includes an overall rating of patient experience. Three sets of analyses were carried out to determine the descriptive system to measure patient experience. A number of items from the Inpatient survey questionnaire were not applicable to all patients and this was addressed using subgroups.

The first study was focussed on identifying items that were best able to explain overall patient experience rating. The variables that were statistically significant across all subgroups centred around respect, trust and cleanliness. A similar study to this was conducted by Sizmur and Redding (Sizmur and Redding, 2009), in which linear regression analysis was applied to the NHS Inpatient survey dataset to examine the core dimensions of Inpatient data. It found that physical comfort, emotional support and respect for patient preferences were most strongly and significantly associated with overall patient experience. In the regression analyses I carried out, I found respect and dignity to be the most important aspect of healthcare delivery, followed by confidence and trust and cleanliness. These items are comparable to the dimensions identified by Sizmur and Redding.

A particular strength of the regression analysis conducted was that it investigated items related to overall patient experience rating rather than dimensions. Hence the findings I have presented are more specific. Also separate analysis was conducted for four subgroups: based on whether the patients had an operation or procedure; and the route of admission was emergency or planned. This enables nuanced understanding about the patient experience. For example, the comparison of the four regression models highlight that the length of time on the waiting list before admission to hospital was a statistically significant variable in patients who had a planned admission and an operation or procedure. In other subgroups, this variable was not statistically significant. The reason could be that patients with an emergency admission have an urgent or unplanned need for medical care. Whereas patients with planned admission without any operation or procedure were perhaps not too concerned about waiting time because they were not looking forward to a specific procedure that would treat them or provide relief.

The second and third study I conducted was sequential and followed a staged approach as recommended by standards and guidelines (presented in chapter 3). The objective of the study described in chapter 6 was to establish dimensions based on the hypothetical constructs of the Inpatient dataset. The exploratory factor models estimated for patients that did not have an operation or procedure did not have good fit and were excluded. The remaining analyses focussed only on patients who had an operation or procedure.

In patients who had an operation or procedure, an 11-factor model was determined for those who had an emergency admission. Broadly these dimensions were regarding: cleanliness; communication, trust and feeling involved; noise at night; information about operation procedure; information about medication; emotional support from hospital staff; doctors and nurses not talking in front of the patient as if they were not there; privacy when being treated; aspects of discharge; provisions for after discharge; being asked for feedback. An 8-factor model was determined for patients with planned admission who underwent an operation or procedure. These were: comfort, trust and communication; cleanliness; information about treatment, operation or procedure; information about medication; emotional support from hospital staff; aspects of discharge; provisions for after discharge; being asked for feedback. The items across the two models were broadly consistent. Two factors, which were identified in patients with emergency admission but were missing in patients with planned admission, were privacy and noise.

The latent factors identified from this study can be compared to the NHS Patient Experience Framework (DH, 2012a). The eight dimensions from the NHS Patient Experience Framework are described in

Table 11. It can be said that a few of the dimensions stated here were not picked up in the analysis I carried out in the NHS inpatient survey dataset. Items on access to care, involvement of family and friends and coordination and integration of care are available in the original inpatient questionnaire. However these items did not emerge in factor analyses, nor in regression analyses, perhaps with the exception of one question on the patient's view on there being enough nurses on duty in the hospital (Q30).

The remaining dimensions: respect for patient centred values; information and communication; physical comfort; emotional support; and transition and continuity are common themes. The themes are broken down further in the dimensions I estimated, for example the information dimension consisted of obtaining information specific to medication and information on operation or procedure. However if one were to go beyond the two subgroups assessed, the question on operation or procedure are not applicable to some inpatient patients who did not have a procedure carried out.

The dimensions estimated from the study were unidimensional and comprehensive. However each dimension consisted of two or more items. I took the decision to include only one item per dimension to measure patient experience. The main reason being that having two or more items in an instrument measuring the same latent construct may lead to illogical healthcare profiles when they are combined. The third empirical study focussed on item selection for the patient experience measure and descriptive system generated is presented in the next section.

8.2 A descriptive system to measure patient experience

I chose IRT as the preferred method to item selection. The measure for patients with emergency admission included 11 items and those with

planned admission had 8 items. The items for the final instruments of patient experience are listed in Table 24. I am not aware of any patient experience instruments specific to hospital stay that the results of my study can be directly compared to. It is the first application of factor analysis and IRT methods (in a staged manner) to patient experience dataset.

Note that there was a few factors that comprised of two items each and they were near identical in meaning (and wording) barring a single component. These items were combined, for example the two questions on noise at night from other patients and noise from healthcare staff were combined as noise from patients or healthcare staff. This prevented the factor from being too specific and loss of information.

This patient experience instrument will provide a useful measure of effectiveness against which to compare different policies to improve quality of care, or effect patient experience.

Table 24: Items for the final patient experience measure

Patients with emergency admission who underwent an operation or procedure	Patients with planned admission who underwent an operation or procedure
Did you have confidence and trust in the doctors treating you?	Beforehand, were you told how you could expect to feel after you had the operation or procedure?
Were you given enough privacy at the hospital?	Were you given enough notice about when you were going to be discharged?
In your opinion, how clean were the hospital facilities?	In your opinion, were there enough nurses on duty to care for you in hospital?
Were you given enough notice about when you were going to be discharged?	In your opinion, how clean were the hospital facilities?
Did the doctors or the nurses talk in front of you as if you weren't there?	Were you told how to take your medication in a way you could understand?
Do you feel you got enough emotional support from hospital staff during your stay?	Did you see, or were you given, any information explaining how to complain to the hospital about the care you received?
Beforehand, did a member of staff explain what would be done during the operation or procedure?	Did you find someone on the hospital staff to talk to about your worries and fears?
Were you ever bothered by noise at night from hospital staff or patients?	Did hospital staff discuss with you whether you may need any further health or social care services after leaving hospital?
Were you told how to take your medication in a way you could understand?	
Did you see, or were you given, any information explaining how to complain to the hospital about the care you received?	
Did hospital staff discuss with you whether you would need any additional equipment in your home, or any adaptations made to your home, after leaving hospital?	

8.3 Limitations

The studies I conducted were not without limitations and key ones are discussed here.

8.3.1 Literature review

The literature review I conducted was large and conducted systematically but it cannot be considered complete. I examined only PROMs, with a focus on health related quality of life and patient experience. I did not distinguish between generic and condition specific measures; or preference based or not. Extensive reviews have been carried out focussing on condition specific to preference based measures (Brazier et al., 2012, Goodwin and Green, 2016). The review I conducted helped to identify methods used to develop an instrument, but the articles I identified from the literature were not sufficiently detailed in reporting the methods used. Perhaps a narrower focus with no more than twenty instruments that are most frequently used, complimented with correspondence with the instrument developers would have yielded more in-depth discussions and helped me confirm the approach to adopt.

8.3.2 Lack of validation

A limitation of this thesis has been reliance on only one dataset and lack of validation of the findings. The dataset I used is formed of the Inpatient questionnaire that is administered by the NHS trusts for monitoring purpose and the questions are designed to assess different aspects of healthcare delivery. I assumed that the items used in the Inpatient questionnaire were comprehensive, and while this is very likely I did not confirm it using any qualitative studies. The empirical studies I conducted focussed on item selection.

The Inpatient survey 2014 consisted of a large number of inapplicable questions. This had a large and lasting impact on the analyses I conducted. I was true to the dataset and generated four subgroups to accommodate the inapplicable questions, which subsequently would result in four instruments of patient experience for each subgroup. The findings from the subgroup analyses are very useful in understanding the experience of the patients in a nuanced way. However, having more than one instrument of patient experience can be criticised as an artefact of the dataset. I was not able to confirm that the differences between the subgroups are not because of the dataset used in this study.

I used exploratory factor analysis to examine the hypothetical construct of the inpatient dataset followed by confirmatory factor analysis. I used the confirmatory framework to assess the multidimensional model and revise the items included in each dimension. Instrument developers often split the dataset into two or use data from administration of the questionnaire at a different time point to confirm dimensionality (Young et al., 2008, Young et al., 2010, Young et al., 2011). The first option would have been difficult given the subgroups, as I would have divided the data into eight groups and reduced the sample size of the models substantially. However, inpatient questionnaire is administered every year and I could have gained access to data from another year. Owing to time constraints I was not able to validate the dimensional structure using another dataset. Further validation of the findings using data from another sample or qualitative studies with key stakeholders is necessary in the future.

8.3.3 Length of the descriptive system

The instruments I have developed for patients who had an operation or procedure in the hospital is not as brief as I had hoped and it will need further reduction before valuation exercise can be carried out. I proposed SEM as a possible method. It orders the items in terms of ability to explain the overall variance in the overall construct. While I think it is a valid method, I was reluctant to reduce items based on SEM findings only. Further validation of the methods and dataset is necessary at this stage.

8.3.4 Excessive focus on the measurement model

My thesis focussed on the measurement model, specifically construct validity only. However I may have put too much emphasis on the relationship between items and dimensions. A comprehensive evaluation of psychometric properties was not provided in my thesis. Other psychometric properties such as content validity, internal consistency, test-retest validity, hypothesis testing and responsiveness are important to the performance of an instrument. Further research is warranted to examine these properties (Fayers and Machin, 2013, Johnson et al., 2011, Mokkink et al., 2010, PROMIS, 2013).

8.4 Methodological contributions of this thesis

The key methodological contributions of this thesis to existing literature are highlighted below.

8.4.1 Further understanding on the development of a descriptive system

The focussed review and review of the literature I have conducted contributes toward understanding of the development of a descriptive system, namely item selection. There are very few studies that have focussed specifically on the descriptive system. The only other study that I am aware of which has looked at questionnaire development was published in 2013 by the EORTC quality of life group (Johnson et al., 2011).

Chapter 3 of my thesis presented a review of guidelines, textbook chapters and quality standards advocating methods for instrument development and evaluation of measurement properties. A total of eleven key texts were identified to gain an overview of methods used in the development of a descriptive system and the measurement properties considered essential were summarised.

The review highlighted the concept of 'measurement model' which is used to demonstrate how items are associated with each dimension and how dimensions are associated with each other in the descriptive system. The tests conducted to evaluate the measurement model, namely factor analysis from classical psychometric testing and IRT from modern test approach. While it was clear that IRT cannot be used for estimating dimensions because it can be applied only in unidimensional models, the merits of using IRT over factor analysis in item selection was not clear from

the review. My thesis used both for item selection to investigate the methods in more detail.

Chapter 4 presented a large literature review of methods used by instrument developers to generate the descriptive system of health related quality of life instruments. A total of 61 full text articles were included. Item generation involved three approaches: developing items de novo, identifying items from several existing measures or using a single measure to derive a short form measure. The instrument developers in few cases determined item selection without reporting any empirical work, but most reported use of empirical work. The most common approach was to apply statistical analyses to datasets containing completed questionnaires.

Measurement models were used to establish construct validity, however in most cases further hypothesis testing was carried out to assess known group, convergent and divergent validity. Psychometric criteria included descriptive statistics of the responses such as missing values, range and distribution of responses and internal consistency. The majority of instrument developers used a combination of statistical methods. A staged approach was used to establish dimensions first, with EFA or PCA, before finalising items in each dimension, using IRT, Rasch or SEM for item selection. However is there an inherent ordering in the analyses one conducts for items, if yes why? This was not clear from the review. Secondly the review highlighted that few of the statistical methods identified are interchangeable in terms of purpose, but may generate very different instruments. However the developers seldom explained why one statistical model was chosen over another.

I realised that while it is possible to identify methods used to develop an instrument from the literature, there is little discussions in terms of the rationale, strengths and limitations of these methods. Also details about the method used was unclear in the papers reporting instrument development to be able to replicate them, however this could be because of the limited length allowed in journal article. The experienced instrument developers must have knowledge about the methods and rationale for different stages and techniques from practice; however this is not sufficiently detailed in the public domain for a new instrument developer like me. Hence methodological investigation was an important focus of my thesis.

8.4.2 Support use of a staged approach

An important criterion for the measure I am developing is for it to be amenable to valuation. None of the patient experience measures currently available are preference-based (Beattie et al., 2015). For a preference-based measure it is essential that the descriptive system includes items that are distinct from each other and brief. Overlapping items or large number of items will result in a very large number of combinations that are not plausible and cannot be valued. In my thesis I explored direct item selection using regression analysis and a staged approach where dimensions are determined first. It was common practice to establish dimensionality of the dataset before carrying out item selection but why one should do so is not apparent from the literature.

Regression analysis identified items that were able to explain the probability of good overall patience experience rating. And while this a valid research agenda in itself, the method was not able to ensure unidimensionality in the items generated. This could have perhaps been

overcome by grouping together items identified by the regression analysis, based on common latent construct they were measuring. However if one is looking to create a subset of items from a lengthy questionnaire, the items selected from a regression may not be able to summarise the full questionnaire as the focus is on assessing the relationship with the dependent variable rather than identifying correlation structure of the dataset.

8.4.3 Advocate the use of confirmatory framework to assess dimensions

Dimensions of the dataset are determined by examining the hypothetical construct of the dataset. Exploratory and confirmatory factor analysis can be conducted to do this. The primary differences between an exploratory and confirmatory approach are that EFA does not use any prior hypothesis about how items are grouped together but CFA does; secondly EFA automatically results in unidimensional dimensions but a CFA does not.

The approach I took in my thesis is estimating the dimensions using EFA, followed by evaluation of the model using a confirmatory framework. CFA allows the multi-dimensional model estimated using EFA to be assessed in terms of overall model fit. I used the results from CFA to revise the items I had included from EFA (based on factor loading) till a good fit was achieved. The process was fiddly and I had to readjust the items several times till the CFA converged. Also this was only possible in two subgroups, as the factor models for patients who did not have an operation or procedure did not converge.

I would recommend including assessment of dimensions using confirmatory framework as an important step in developing a descriptive

system. Eigen value results, scree plots and factor loadings may be subjective. CFA employs a multi-dimensional model based approach where the fit of the model helps to decide the number of items per dimension and number of dimensions to include in the model.

8.4.4 Advocate the use of IRT in item selection

IRT technique was used to select items for the patient experience measure. It is able to examine responses to candidate items that are binary or ordinal and suggest items to include based on discriminative ability. In a one-parameter model such as Rasch the probability of a correct response is determined by the item's difficulty and the respondents assessment of the latent variable. I used two parameter models in my study where another parameter called the discrimination (slope) parameter was introduced to measure the differential capability of an item. A high discrimination parameter value suggests an item that has a high ability to differentiate subjects, and the probability of a patient experience increases more rapidly as the latent variable increases.

I compared the findings from IRT models to models using underlying variable approach (factor analysis). While the results were very similar and interpretable, it was the actual mechanics of the IRT approach compared to factor analysis approach that made me select the first approach. The IRT is a full model while factor analysis approach is considered a partial information model that uses information only from the pairwise distribution of the ordinal variables (Bartholomew et al., 2008). Also in the literature IRT is considered better suited for examining item characteristics and factor analysis appropriate for multidimensional model testing (Bartholomew et al., 2008).

In addition to examining the statistical findings, an instrument developer should be open to applying their own judgement. It should be noted that in addition to IRT I used minor amendments in wording to combine items in the final descriptive system I presented. Limiting myself to one item per factor in factors that contained near identical items would have created unnecessary loss of information and breadth. While item selection using IRT method was the primary focus, it is important to remember that statistical techniques are 'tools' that are available to us and the finding are often indicative rather than definitive. Use of own judgement and deviations from actual findings during item selection should be clearly reported by instrument developers.

8.4.5 Novel use of SEM in dimension reduction

SEM is an extension of factor analysis and in this chapter I proposed using it for dimension reduction in a multidimensional model. SEM is a statistical technique that allows evaluation of relationship between latent and observed variables. I specified the overall latent construct being measured as the dependent variable and each item as an independent variable to assess the relationship. It is possible to select the variables based on their association with the overall patient experience (latent variable). The difference in item estimates was minimal and I did not think it was appropriate to make a decision based only on SEM findings only. Further testing needs to be carried out, namely cross validation using another sample of Inpatient survey data (from a different year) and studies to assess face validity of the new measures.

8.4.7 Use of 'judgement' in exceptional circumstances

The use of judgement by instrument developers was reported in few of the empirical studies I reviewed in chapter 4. While this practice introduces a

degree of subjectivity, it is important to be able to do so as the final instrument needs to align with what the instruments developers set out to measure.

The item selection I conducted in this thesis was primarily based on objective psychometric evidence, but there were a few factors in which I applied my judgement for item selection. I combined near identical items to avoid the loss of breadth while selecting an item per dimension. The reason being that being too specific sometimes misses out important aspects of the underlying condition.

Undoubtedly the use of judgement by instrument developers should be an exception rather than rule to avoid subjectivity. However, an instrument developer should be free to do so at any stage of development if it is sensible or pragmatic, and well-supported.

8.5 Conceptual contributions of this thesis

The two key conceptual contributions of this thesis are outlined below:

8.5.1 Application of an extra-welfarist framework to quality of care

My thesis takes the concept of extra-welfarist framework that is routinely applied in HTAs in the UK and takes a step towards employing it to quality of care interventions. The measurement and valuation of health outcomes to inform public spending is widely accepted and practised but this is one of the first studies to explore how it might be applied to decisions related healthcare delivery that aim to improve patient experience.

The preliminary descriptive system I have proposed is designed to resemble the EQ-5D health related quality of life descriptive system that is a brief and preference-based. It sets out to provide a classification system that identifies the dimensions of patient experience that are affected by quality of care and generate "healthcare" states upon combination. This enables social values or public preferences to be attached to healthcare states that reflect a broad spectrum of healthcare experiences and inform prioritisation decisions in a systematic way.

8.5.2 Measurement and valuation of patient experience

My thesis takes the concept of measurement and valuation in health to inform public spending, and begins to explore how it might be applied to decisions related to healthcare delivery. PREMs are already being used for monitoring and improvement of quality of care, and the new descriptive systems I have generated can assist measurement of patient experience. However, the vision of this thesis is more ambitious and the design of PREM allows it inform economic analysis of healthcare intervention in the future.

I have not set out the details of the valuation process in this thesis but a few important considerations are discussed in future research section. Having a value set to attach to patient experience scenarios will enable comparison of quality of care interventions and help to inform cost effectiveness analyses.

Some may argue that the NHS should continue to proceed on a piecemeal basis, making decision on quality of care interventions as and when required, because that way useful progress can be made more quickly. However a multidimensional instrument like the one I have proposed provides a more accountable and systematic way of making decision regarding public spending, in which cost effectiveness is an important criterion.

8.6 Further research

The basic concept of this PhD stems from the need to consider value for money in all aspects of the public sector. It begins to explore whether the methods used for HTA can be adopted in other aspects of health. Further discussions on why and how a measure of patient experience that was developed with the view of making it amenable to developing preference weights can be used to inform decision-making is warranted. As the thesis currently stands, the descriptive systems I have developed are still preliminary. Further research is recommended in terms of refining the items. This includes both psychometric testing and validation using experts, and is elaborated below.

8.6.1 Scope of the instrument

It is important to validate the final item selection results using a mix of qualitative and quantitative analysis. Applying confirmatory factor analysis in a different inpatient dataset will cross validate the results. Another approach is to present the results to experts and patients for validation purpose. It should be highlighted that statistical analyses are indicative rather than definite and it is possible that methodological artefacts have been captured, and objective judgements need to be made throughout the instrument development process.

An important decision to be made using experts would be regarding number of measures for patient experience, in other words should it be generic to all patients or specific to subgroups as was found in my thesis. It is important to consider the implications of the two approaches in terms of responsiveness and comparability. Similar to generic and condition specific measure, a measure that is generic (compared to specific sub groups) will allow all patient experience across hospitals to be measured using a commensurable unit and enable wider comparison of quality of care interventions. However a measure that is specific to a subgroup may be more sensitive in capturing aspects most relevant to their patient experience than a generic instrument.

Upon confirming item selection, the focus should turn to refining item response levels and wording to ensure correct interpretation, response burden and acceptability.

8.6.2 Further evaluation of psychometric properties

Establishing the validity and reliability of the instrument is important for it to be widely accepted. Further research needs to be conducted in terms of content validity, hypothesis testing with other measures, internal consistency and test-retest reliability. Finally for the measure to be able to inform economic evaluation, responsiveness or ability to detect change is also crucial. For instance the impact of a new intervention to improve quality of care can be measured by assessing the patient experience before and after the intervention.

8.6.3 Preference elicitation study

While a brief patient experience measure will be appreciated for measurement purpose and comparison of interventions, I specifically wanted to be able to attach preference weights to the levels and dimensions of the measure. The reason being that the response levels and dimensions are valued differently by patients; being able to differentiate across these nuances will help in informing prioritisation decisions. In the future, valuation studies could be carried out.

In health state valuation two anchors are assigned: at the top end, no problems related to any of the dimensions is given a value of 1 and, at the other end, death is given a value of 0. While death is a natural anchor for health, patient experience does not have a natural bottom anchor and some of the conventional valuation techniques such as standard gamble and time trade off may not work. Perhaps a DCE could be used to value the healthcare states. DCE is a stated preference technique consistent with random utility theory (utility maximisation), in which an assumption is made that each respondent will choose the alternative, which gives him or

her the greatest utility in relative terms. In the DCE study individuals will be asked to state their preferences for hypothetical scenarios comprising the dimensions of patient experience I have identified. Regression techniques will be used to establish utility function by modelling the choices made by the respondent.

8.7 Implications: measurement and valuation of PREM

The measurement and valuation of patient experience is a novel and broad concept, and there are four key implications or visions of this thesis if this research were to be taken forward.

Firstly the descriptive systems I have proposed in this thesis provide short forms to the lengthy Inpatient questionnaire currently used in the NHS and can be used to measure patient experience. The methods used to develop them are scientifically robust. Both are also shorter than the instruments currently available to measure patient experience in a hospital setting (Beattie et al., 2015). It should be noted apart from surveys, a one-item instrument called the NHS Friends and Family Test (FFT) is also administered to NHS users (NHS, 2013). It helps service providers to understand if the patients are satisfied and areas for improvement. It is a quick and anonymous way to provide feedback. However the information provided by a one-item anonymised instrument is extremely limited as it is very vague. Further granularity is necessary to be able to highlight attributes of care that were good or bad, and make improvements based on this assessment. The 11-item and 8-item measures of patient experience I have presented are likely to be useful as they are neither too brief not too lengthy.

Secondly the measures I have developed are perhaps the only patient experience measures that are amenable to valuation. And it is possible to elicit societal value set for these measures in the future. Many of the patient experience measures currently available have a simple scoring system that applies equal weight to all dimensions and levels. However it is very likely that the patients and the society differentiate between them. Eliciting preferences for dimensions and dimension levels will enable decision makers to differentiate across healthcare attributes using value judgements of the general public, and prioritise interventions accordingly.

Thirdly it enables use of patient experience values to inform economic evaluation of quality of care interventions. The implementation of PREM to inform economic analyses would entail increased measurement of patient experience to enable detailed study of costs and effects of competing quality of care intervention. This would undoubtedly increase the overall administrative burden, need for analysis and slow down decision making process. However this system is likely to create efficiency gains in the future. It engages patients and the general public, and places their views at the heart of decision making. While intellectually the use of economic analysis to aid decision making is very appealing, the practical aspects such as time, resources and skill sets involved are challenging and need further consideration. Whether or not the policymakers should routinely incorporate economic evaluation of quality of care needs to be thoroughly debated.

Finally it is important to consider the relationship of the patient experience measure to the EQ-5D or SF-6D that is used to generate health related quality of life data and inform economic evaluation. There is absolutely no content overlap between the patient experience measure and the EQ-5D

measure, however the two concepts may be related as overall health outcome is likely to affect the patient's hospital experience. A possible implication of this thesis could be adopting a weighted QALY model for the health sector that combines quality of care and health outcome into a single index. This super QALY would provide an opportunity to consider how quality of care contributes to individual and collective wellbeing provided by the NHS. For example, following a hospital stay there may not be an impact on the patient's health outcome but the patient may still value the inpatient experience for emotional support gained from healthcare staff. Hence the NHS is able to measure the utility provided to the patients in a broader scope, which goes beyond health gain. It will allow process related utility to be incorporated into decision making. Evaluation of an intervention will therefore take into account dimensions that are valued by a patient in a comprehensive manner.

8.8 Conclusion

Efficient allocation of public spending requires consideration of value for money. While the concept of the QALY is well accepted in health care and used to ensure cost effectiveness, a similar approach to quality of care has not been explored before. In my thesis, I developed a descriptive system for patient experience that can be valued and used to inform economic evaluation.

The initial chapters provided a contemporary overview of recommended methods and those actually used by instrument developers. Frequently a staged approach was used to establish dimensions first, using exploratory factor analysis, followed by item selection using item response theory (IRT), Rasch or structural equation modelling (SEM).

Three empirical chapters demonstrate the use of different methods for item selection and its underlying mechanics, followed by comparison of the methods. An existing patient dataset, the Inpatient survey (2014) that collected information on nearly 70 aspects of healthcare delivery from NHS users was used.

Logistic regression analyses were applied with respondents' rating of overall patient experience specified as dependent variable. Regression analyses identified a large number of significant variables but most overlapped conceptually. Advanced statistical analyses focussed only on patients who had an operation or procedure. Latent construct or dimensions were derived using exploratory factor analysis and a multi-dimensional measurement model was confirmed using factor analysis method. An 11 and 8 factor model for patients with A&E and planned admissions respectively was determined using factor analysis. IRT and factor analysis approach were used for item selection. Generalised partial credit model and factor analysis model identified different items to include in each dimension.

In terms of method, this thesis demonstrated that different patient experience measures are generated based on patient population used and item selection technique adopted, and this should be an important consideration in instrument development. The thesis recommends IRT technique for item selection. The question about having a specific measure to cater to each group or a generic measure of patient experience is less straightforward and requires further debate. Use of subgroups was recommended here because that was the approach that best reflected the

inpatient dataset. However using a different response dataset could provide a different insight into patient experience. The thesis also highlights use of confirmatory factor analysis to assess the measurement model describing the instrument.

The two descriptive systems presented in this thesis allows all the key items of patient experience to be incorporated into multi-dimensional frameworks. Further research needs to be conducted to validate the item selection and elicit value sets to be able to use them in cost effectiveness analyses in the future.

References

AARONSON, N., ALONSO, J., BURNAM, A., LOHR, K. N., PATRICK, D. L., PERRIN, E. & STEIN, R. E. 2002. Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res,* 11**,** 193-205.

AL-JANABI, H., FLYNN, T. N. & COAST, J. 2012. Development of a self-report measure of capability wellbeing for adults: the ICECAP-A. *Quality of Life Research,* 21**,** 167-76.

AL-JANABI, H., PETERS, T. J., BRAZIER, J., BRYAN, S., FLYNN, T. N., CLEMENS, S., MOODY, A. & COAST, J. 2013. An investigation of the construct validity of the ICECAP-A capability measure. *Quality of Life Research,* 22**,** 1831-1840.

ALDERWICK, H., HAM, C. & BUCK, D. 2015. Population health systems. Going beyond integrated care. London: The King's Fund.

BABYAK, M. A. 2004. What You See May Not Be What You Get: A Brief, Nontechnical Introduction to Overfitting in Regression-Type Models. *Psychosomatic Medicine,* 66**,** 411-421.

BARBER, M. D., WALTERS, M. D. & BUMP, R. C. 2005. Short forms of two condition-specific quality-of-life questionnaires for women with pelvic floor disorders (PFDI-20 and PFIQ-7). *American Journal of Obstetrics and Gynecology,* 193**,** 103-113.

BARTHOLOMEW, D. J. & KNOTT, M. 1999. *Latent Variable Models and Factor Analysis,* London, Arnold.

BARTHOLOMEW, D. J., STEELE, F., GALBRAITH, J. & MOUSTAKI, I. 2008. *Analysis of multivariate social science data*, CRC press.

BEATTIE, M., MURPHY, D. J., ATHERTON, I. & LAUDER, W. 2015. Instruments to measure patient experience of healthcare quality in hospitals: a systematic review. *Systematic Reviews,* 4**,** 97.

BENNETT, K. J., TORRANCE, G. W., BOYLE, M. H., GUSCOTT, R. & MORAN, L. A. 2000. Development and testing of a utility measure for major, unipolar depression (McSad). *Quality of Life Research,* 9**,** 109-20.

BERWICK, D. & FOX, D. M. 2016. "Evaluating the Quality of Medical Care": Donabedian's Classic Article 50 Years Later. *The Milbank Quarterly,* 94**,** 237-241.

BEWICK, V., CHEEK, L. & BALL, J. 2005. Statistics review 14: Logistic regression. *Crit Care,* 9**,** 112-8.

BJORNER, J. B., PETERSEN, M. A., GROENVOLD, M., AARONSON, N., AHLNER-ELMQVIST, M., ARRARAS, J. I., BREDART, A., FAYERS, P., JORDHOY, M., SPRANGERS, M., WATSON, M. & YOUNG, T. 2004. Use of item response theory to develop a shortened version of the EORTC QLQ-C30 emotional functioning scale. *Qual Life Res,* 13**,** 1683-97.

BOYD, J. 2007. *The 2006 inpatients importance study,* Oxford, Picker Institute Europe, The Acute Co-ordination Centre for the NHS Acute Patient Survey Programme.

BOYLE, M. H. & TORRANCE, G. W. 1984. Developing multiattribute health indexes. *Medical Care,* 22**,** 1045-1057.

BRAZIER, J., RATCLIFFE, J., SALOMON, J. A. & TSUCHIYA, A. 2007. *Measuring and Valuing Health Benefits for Economic Evaluation*, Oxford University Press.

BRAZIER, J., ROBERTS, J. & DEVERILL, M. 2002. The estimation of a preference-based measure of health from the SF-36. *J Health Econ,* 21**,** 271-92.

BRAZIER, J., USHERWOOD, T., HARPER, R. & THOMAS, K. 1998. Deriving a Preference-Based Single Index from the UK SF-36 Health Survey. *Journal of Clinical Epidemiology,* 51**,** 1115-1128.

BRAZIER, J. E. & ROBERTS, J. 2004. The estimation of a preference-based measure of health from the SF-12. *Medical Care,* 42**,** 851-859.

BRAZIER, J. E., ROBERTS, J., PLATTS, M. & ZOELLNER, Y. F. 2005. Estimating a preference-based index for a menopause specific health quality of life questionnaire. *Health & Quality of Life Outcomes,* 3**,** 13.

BRAZIER, J. E., ROWEN, D., MAVRANEZOULI, I., TSUCHIYA, A., YOUNG, T., YANG, Y., BARKHAM, M. & IBBOTSON, R. 2012. Developing and testing methods for deriving preference-based measures of health from condition-specific measures (and other patient-based measures of outcome). *Health Technol Assess,* 16**,** 1-114.

BRENNAN, V. K. & DIXON, S. 2013. Incorporating process utility into quality adjusted life years: a systematic review of empirical studies. *Pharmacoeconomics,* 31**,** 677-91.

BROOKS, R. 1996. EuroQol: the current state of play. *Health Policy,* 37**,** 53-72.

BROUWER, W. B., CULYER, A. J., VAN EXEL, N. J. & RUTTEN, F. F. 2008. Welfarism vs. extra-welfarism. *J Health Econ,* 27**,** 325-38.
BROWNE, M. W. & CUDECK, R. 1992. Alternative Ways of Assessing Model Fit. *Sociological Methods & Research,* 21**,** 230-258.
CAMPBELL, S. M., ROLAND, M. O. & BUETOW, S. A. 2000. Defining quality of care. *Soc Sci Med,* 51**,** 1611-25.
CAPPELLERI, J. C., JASON LUNDY, J. & HAYS, R. D. 2014. Overview of Classical Test Theory and Item Response Theory for the Quantitative Assessment of Items in Developing Patient-Reported Outcomes Measures. *Clinical Therapeutics,* 36**,** 648-662.
CARLTON, J. 2013a. Developing the draft descriptive system for the child amblyopia treatment questionnaire (CAT-Qol): a mixed methods study. *Health & Quality of Life Outcomes,* 11**,** 174.
CARLTON, J. 2013b. Identifying potential themes for the Child Amblyopia Treatment Questionnaire. *Optometry & Vision Science,* 90**,** 867-73.
CARPITA, M. & MANISERA, M. 2011. On the Imputation of Missing Data in Surveys with Likert-Type Scales. *Journal of Classification,* 28**,** 93-112.
CHALKIDOU, K. 2009. Comparative effectiveness review within the U.K.'s National Institute for Health and Clinical Excellence. *Issue Brief (Commonw Fund),* 59**,** 1-12.
CLARK, M. D., DETERMANN, D., PETROU, S., MORO, D. & DE BEKKER-GROB, E. W. 2014. Discrete Choice Experiments in Health Economics: A Review of the Literature. *PharmacoEconomics,* 32**,** 883-902.
CLEARY, P. D., EDGMAN-LEVITAN, S., WALKER, J. D., GERTEIS, M. & DELBANCO, T. 1993. Using patient reports to improve medical care: a preliminary report from 10 hospitals. . *Quality Management in Health Care,* 2**,** 31-38.
COAST, J., FLYNN, T. N., NATARAJAN, L., SPROSTON, K., LEWIS, J., LOUVIERE, J. J. & PETERS, T. J. 2008a. Valuing the ICECAP capability index for older people. *Social Science & Medicine,* 67**,** 874-882.
COAST, J., PETERS, T. J., NATARAJAN, L., SPROSTON, K. & FLYNN, T. 2008b. An assessment of the construct validity of the descriptive system for the ICECAP capability measure for older people. *Quality of Life Research,* 17**,** 967-76.
COMINS, J. D., KROGSGAARD, M. R. & BRODERSEN, J. 2013. Ensuring face validity in patient-related outcome scores--a matter of content. *Knee,* 20**,** 72-8.
COULTER, A., FITZPATRICK, R. & CORNWELL, J. 2009. *The point of care measures of patients' experience in hospital: purpose, methods and uses. The Kings fund*.
COYNE, K., REVICKI, D., HUNT, T., COREY, R., STEWART, W., BENTKOVER, J., KURTH, H. & ABRAMS, P. 2002. Psychometric validation of an overactive bladder symptom and health-related quality of life questionnaire: The OAB-q. *Quality of Life Research,* 11**,** 563-574.
CQC 2014. *Overall Patient Experience Scores: Supporting Information*.
CQC. 2016. *Care Quality Commission* [Online]. http://www.cqc.org.uk/about-us. [Accessed].
CQC. 2017. 2016 Adult Inpatient Survey: Statistical Release.
CQC 2018. NHS Surveys. http://www.cqc.org.uk/publications/surveys/surveys: Care Quality Commission.
CUDECK, R. 1989. Analysis of correlation matrices using covariance structure models. *Psychological Bulletin,* 105**,** 317-327.
CUERVO, J., CASTEJON, N., KHALAF, K. M., WAWERU, C., GLOBE, D. & PATRICK, D. L. 2014. Development of the Incontinence Utility Index: estimating population-based utilities associated with urinary problems from the Incontinence Quality of Life Questionnaire and Neurogenic Module. *Health & Quality of Life Outcomes,* 12**,** 147.
DALLMEIJER, A. J., SCHOLTES, V. A., BECHER, J. & ROORDA, L. D. 2011. Measuring mobility limitations in children with cerebral palsy: Rasch model fit of a mobility questionnaire, MobQues28. *Archives of Physical Medicine & Rehabilitation,* 92**,** 640-5.
DAVIS, E., MACKINNON, A., DAVERN, M., BOYD, R., BOHANNA, I., WATERS, E., GRAHAM, H. K., REID, S. & REDDIHOUGH, D. 2013. Description and psychometric properties of the CP QOL-Teen: a quality of life questionnaire for adolescents with cerebral palsy. *Research in Developmental Disabilities,* 34**,** 344-52.
DE BEKKER-GROB, E. W., RYAN, M. & GERARD, K. 2012. Discrete choice experiments in health economics: a review of the literature. *Health Economics,* 21**,** 145-172.

DE JONG, Z., VAN DER HEIJDE, D., MCKENNA, S. & WHALLEY, D. 1997. The reliability and construct validity of the RAQoL: a rheumatoid arthritis-specific quality of life instrument. *Rheumatology,* 36**,** 878-883.

DECOURCY, A., WEST, E. & BARRON, D. 2012. The National Adult Inpatient Survey conducted in the English National Health Service from 2002 to 2009: how have the data been used and what do we know as a result? *BMC Health Serv Res,* 12.

DEVLIN, N. J., SHAH, K. K., FENG, Y., MULHERN, B. & VAN HOUT, B. 2017. Valuing health-related quality of life: An EQ-5D-5L value set for England. *Health Econ*.

DEWALT, D. A., ROTHROCK, N., YOUNT, S. & STONE, A. A. 2007. Evaluation of item candidates: the PROMIS qualitative item review. *Med Care,* 45**,** S12-21.

DH. 2012a. NHS Patient Experience Framework. *Department of Health and Social Care* [Online].

DH 2012b. Our NHS care objectives: A draft mandate to the NHS Commissioning Board. *In:* HTTP://WEBARCHIVE.NATIONALARCHIVES.GOV.UK/20130104235941/HTTP://WWW.DH.GOV.UK/HEALTH/FILES/2012/07/A-DRAFT-MANDATE-TO-THE-NHS-COMMISSIONING-BOARD.PDF (ed.) *Department of Health.* London: The Stationary Office.

DH 2014a. Hard truths: the journey to putting patients first. Volume one of the government response to the Mid Staffordshire Foundation Trust public inquiry. . *In:* HTTPS://WWW.GOV.UK/GOVERNMENT/PUBLICATIONS/MID-STAFFORDSHIRE-NHS-FT-PUBLIC-INQUIRY-GOVERNMENT-RESPONSE (ed.) *Department of Health.* London: The Stationery Office.

DH 2014b. Hard truths: the journey to putting patients first. Volume two of the government response to the Mid Staffordshire Foundation Trust public inquiry: response to the inquiry's recommendations. . *In:* HTTPS://WWW.GOV.UK/GOVERNMENT/PUBLICATIONS/MID-STAFFORDSHIRE-NHS-FT-PUBLIC-INQUIRY-GOVERNMENT-RESPONSE (ed.) *Department of Health.* London: The Stationery Office.

DIENER, A., O'BRIEN, B. & GAFNI, A. 1998. Health care contingent valuation studies: a review and classification of the literature. *Health Econ,* 7**,** 313-26.

DONABEDIAN, A. 1966. Evaluating the quality of medical care. *The Milbank Memorial Fund quarterly,* 44**,** Suppl:166-206.

DONABEDIAN, A. 1980. *Explorations in quality assessment and monitoring: The definition of quality and approaches to assessment.,* Ann Arbor, Michigan, Health Administration Press.

DOWARD, L. C., MCKENNA, S. P., WHALLEY, D., TENNANT, A., GRIFFITHS, B., EMERY, P. & VEALE, D. J. 2009. The development of the L-QoL: a quality-of-life instrument specific to systemic lupus erythematosus. *Ann Rheum Dis,* 68**,** 196-200.

DOWIE, J. 2002. Decision validity should determine whether a generic or condition-specific HRQOL measure is used in health care decisions. *Health Econ,* 11**,** 1-8.

DOYLE, C., LENNOX, L. & BELL, D. 2013. A systematic review of evidence on the links between patient experience and clinical safety and effectiveness. *BMJ Open,* 3.

EFRON, B., HASTIE, T., JOHNSTONE, I. & TIBSHIRANI, R. 2004. Least angle regression. *Ann. Statist.,* 32**,** 407-499.

EMEA. 2005. *Reflection paper on the regulatory guidance for the use of health related quality of life measures in the evaluation of medicinal products* [Online]. [Accessed 1st September 2016].

FARIA, R., GOMES, M., EPSTEIN, D. & WHITE, I. R. 2014. A Guide to Handling Missing Data in Cost-Effectiveness Analysis Conducted Within Randomised Controlled Trials. *Pharmacoeconomics,* 32**,** 1157-1170.

FAYERS, P. & MACHIN, D. 2013. *Quality of Life: The Assessment, Analysis and Interpretation of Patient-Reported Outcomes*, Wiley.

FAYERS, P. M. & HAND, D. J. 2002. Causal variables, indicator variables and measurement scales: an example from quality of life. *Journal of the Royal Statistical Society: Series A (Statistics in Society),* 165**,** 233-253.

FDA. 2009. *Guidance for industry. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims* [Online]. [Accessed 1st September 2016].

FEENY, D. 2002. Commentary on Jack Dowie, "Decision validity should determine whether a generic or condition-specific HRQOL measure is used in health care decisions". *Health Econ,* 11**,** 13-6; discussion 21-2.

GANDEK, B. & WARE, J. 1993. SF-36 Health Survey manual and interpretation guide. *Boston: The Health Institute, New England Medical Center*.
GERTEIS, M., EDGMAN-LEVITAN, S., DALEY, J. & T., D. 1993. *Through the Patient's Eyes,* San Francisco, CA: Jossey-Bass.
GOODWIN, E. & GREEN, C. 2015. A Quality-Adjusted Life-Year Measure for Multiple Sclerosis: Developing a Patient-Reported Health State Classification System for a Multiple Sclerosis-Specific Preference-Based Measure. *Value in Health,* 18**,** 1016-24.
GOODWIN, E. & GREEN, C. 2016. A Systematic Review of the Literature on the Development of Condition-Specific Preference-Based Measures of Health. *Appl Health Econ Health Policy,* 14**,** 161-83.
GREENE, W. & HENSHER, D. 2010. *Modeling Ordered Choices*, Cambridge University Press.
GREWAL, I., LEWIS, J., FLYNN, T., BROWN, J., BOND, J. & COAST, J. 2006. Developing attributes for a generic quality of life measure for older people: Preferences or capabilities? *Social Science & Medicine,* 62**,** 1891-1901.
GROENVOLD, M., PETERSEN, M. A., AARONSON, N. K., ARRARAS, J. I., BLAZEBY, J. M., BOTTOMLEY, A., FAYERS, P. M., DE GRAEFF, A., HAMMERLID, E., KAASA, S., SPRANGERS, M. A. & BJORNER, J. B. 2006. The development of the EORTC QLQ-C15-PAL: a shortened questionnaire for cancer patients in palliative care. *Eur J Cancer,* 42**,** 55-64.
GUYATT, G. H., FEENY, D. H. & PATRICK, D. L. 1993. Measuring health-related quality of life. *Ann Intern Med,* 118**,** 622-9.
HAM, C., BERWICK, D. & DIXON, J. 2016. Improving Quality in the English NHS: A Strategy for Action. London: The King's Fund.
HAWTHORNE, G. 2009. Assessing utility where short measures are required: development of the short Assessment of Quality of Life-8 (AQoL-8) instrument. *Value Health,* 12**,** 948-57.
HAWTHORNE, G., RICHARDSON, J. & OSBORNE, R. 1999. The Assessment of Quality of Life (AQoL) instrument: A psychometric measure of health-related quality of life. *Quality of Life Research,* 8**,** 209-224.
HEALTH 2013. Evidence scan: Measuring patient experience. *In:* SCAN, E. (ed.). http://www.health.org.uk/publication/measuring-patient-experience: The Health Foundation.
HIGGINS, A., BARNETT, J., MEADS, C., SINGH, J. & LONGWORTH, L. 2014. Does convenience matter in health care delivery? A systematic review of convenience-based aspects of process utility. *Value Health,* 17**,** 877-87.
HOBART, J., LAMPING, D., FITZPATRICK, R., RIAZI, A. & THOMPSON, A. 2001. The Multiple Sclerosis Impact Scale (MSIS-29): a new patient-based outcome measure. *Brain,* 124**,** 962-73.
HU, L. T. & BENTLER, P. M. 1999. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal,* 6**,** 1-55.
JENKINSON, C., BRENNAN, C., FITZPATRICK, R., BROMBERG, M. & SWASH, M. 1999. Development and validation of a short measure of health status for individuals with amyotrophic lateral sclerosis/motor neurone disease: the ALSAQ-40. *Journal of neurology,* 246**,** III16-III21.
JENKINSON, C., COULTER, A. & BRUSTER, S. 2002a. The Picker Patient Experience Questionnaire: development and validation using data from in-patient surveys in five countries. *Int J Qual Health C,* 14.
JENKINSON, C., COULTER, A. & BRUSTER, S. 2002b. The Picker Patient Experience Questionnaire: development and validation using data from in-patient surveys in five countries. *International Journal for Quality in Health Care,* 14**,** 353-8.
JENKINSON, C., COULTER, A., REEVES, R., BRUSTER, S. & RICHARDS, N. 2003. Properties of the Picker Patient Experience questionnaire in a randomized controlled trial of long versus short form survey instruments. *J Public Health Med,* 25.
JOHNSON, C., AARONSON, N., BLAZEBY, J. M., BOTTOMLEY, A., FAYERS, P., KOLLER, M., KULIŚ, D., RAMAGE, J., SPRANGERS, M., VELIKOVA, G. & YOUNG, T. 2011. *Guidelines for Developing Questionnaire Modules* [Online]. [Accessed 1st September 2016].
JOKOVIC, A., LOCKER, D. & GUYATT, G. 2006. Short forms of the Child Perceptions Questionnaire for 11-14-year-old children (CPQ11-14): development and initial evaluation. *Health Qual Life Outcomes,* 4**,** 4.

JONES, G., KENNEDY, S., BARNARD, A., WONG, J. & JENKINSON, C. 2001. Development of an endometriosis quality-of-life instrument: The Endometriosis Health Profile-30. *Obstet Gynecol,* 98**,** 258-64.

JUNIPER, E. F., GUYATT, G. H., EPSTEIN, R. S., FERRIE, P. J., JAESCHKE, R. & HILLER, T. K. 1992. Evaluation of impairment of health related quality of life in asthma: development of a questionnaire for use in clinical trials. *Thorax,* 47**,** 76-83.

JUNIPER, E. F., GUYATT, G. H., STREINER, D. L. & KING, D. R. 1997. Clinical impact versus factor analysis for quality of life questionnaire construction. *Journal of Clinical Epidemiology,* 50**,** 233-238.

KIRSHNER, B. & GUYATT, G. 1985. A methodological framework for assessing health indices. *Journal of chronic diseases,* 38**,** 27-36.

KLINE, P. 1994. *An Easy Guide to Factor Analysis.,* Abingdon-on-Thames, Routledge.

KLINE, P. 2005. *Principles and Practice of Structural Equation Modeling,* New York, Guildford.

KOSINSKI, M., BAYLISS, M. S., BJORNER, J. B., J. E. WARE, JR., GARBER, W. H., BATENHORST, A., CADY, R., DAHL, C. G. H., XF, DOWSON, A. & TEPPER, S. 2003. A Six-Item Short-Form Survey for Measuring Headache Impact: The HIT-6&#x2122. *Quality of Life Research,* 12**,** 963-974.

KOWALSKI, J. W., RENTZ, A. M., WALT, J. G., LLOYD, A., LEE, J., YOUNG, T. A., CHEN, W.-H., BRESSLER, N. M., LEE, P., BRAZIER, J. E., HAYS, R. D. & REVICKI, D. A. 2012. Rasch analysis in the development of a simplified version of the National Eye Institute Visual-Function Questionnaire-25 for utility estimation. *Quality of Life Research,* 21**,** 323-34.

KRABBE, P. F. 2016. *The Measurement of Health and Health Status: Concepts, Methods and Applications from a Multidisciplinary Perspective*, Elsevier Science Publishing Company Incorporated.

KRABBE, P. F., DEVLIN, N. J., STOLK, E. A., SHAH, K. K., OPPE, M., VAN HOUT, B., QUIK, E. H., PICKARD, A. S. & XIE, F. 2014. Multinational evidence of the applicability and robustness of discrete choice modeling for deriving EQ-5D-5L health-state values. *Med Care,* 52**,** 935-43.

KRAHN, M., RITVO, P., IRVINE, J., TOMLINSON, G., BEZJAK, A., TRACHTENBERG, J. & NAGLIE, G. 2000. Construction of the Patient-Oriented Prostate Utility Scale (PORPUS): a multiattribute health state classification system for prostate cancer. *Journal of Clinical Epidemiology,* 53**,** 920-30.

KUSPINAR, A., FINCH, L., PICKARD, S. & MAYO, N. E. 2014. Using existing data to identify candidate items for a health state classification system in multiple sclerosis. *Quality of Life Research,* 23**,** 1445-57.

LYRAKOS, G. N., VINI, D., ASLANI, H. & DROSOU-SERVOU, M. 2012. Psychometric properties of the Specific Thalassemia Quality of Life instrument for adults. *Patient Preference and Adherence,* 6**,** 477-497.

MASTERS, G. N. 1982. A rasch model for partial credit scoring. *Psychometrika,* 47**,** 149-174.

MAVRANEZOULI, I., BRAZIER, J. E., YOUNG, T. A. & BARKHAM, M. 2011. Using Rasch analysis to form plausible health states amenable to valuation: the development of CORE-6D from a measure of common mental health problems (CORE-OM). *Quality of Life Research,* 20**,** 321-33.

MCKENNA, S. P., DOWARD, L. C., NIERO, M. & ERDMAN, R. 2004. Development of needs-based quality of life instruments. *Value in Health,* 7**,** S17-S21.

MISAJON, R., HAWTHORNE, G., RICHARDSON, J. & KEEFFE, J. 2004. Vision and Quality of Life: The Development of a Utility Measure. *Investigative Ophthalmology & Visual Science,* 45**,** 1382-1382.

MOKKINK, L. B., TERWEE, C. B., PATRICK, D. L., ALONSO, J., STRATFORD, P. W., KNOL, D. L., BOUTER, L. M. & DE VET, H. C. 2010. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res,* 19**,** 539-49.

MOONEY, G. 1998. Beyond health outcomes: The benefits of health care. *Health Care Analysis,* 6**,** 99-105.

MOUSTAKI, I. 2003. A general class of latent variable models for ordinal manifest variables with covariate effects on the manifest and latent variables. *British Journal of Mathematical and Statistical Psychology***,** 331-351.

MUKURIA, C., ROWEN, D., BRAZIER, J. E., YOUNG, T. A. & NAFEES, B. 2015. Deriving a Preference-Based Measure for Myelofibrosis from the EORTC QLQ-C30 and the MF-SAF. *Value in Health,* 18**,** 846-55.

MULHERN, B., ROWEN, D., BRAZIER, J., SMITH, S., ROMEO, R., TAIT, R., WATCHURST, C., CHUA, K. C., LOFTUS, V., YOUNG, T., LAMPING, D., KNAPP, M., HOWARD, R. & BANERJEE, S. 2013. Development of DEMQOL-U and DEMQOL-PROXY-U: generation of preference-based indices from DEMQOL and DEMQOL-PROXY for use in economic evaluation. *Health Technology Assessment (Winchester, England),* 17**,** v-xv, 1-140.

MULHERN, B., ROWEN, D., JACOBY, A., MARSON, T., SNAPE, D., HUGHES, D., LATIMER, N., BAKER, G. A. & BRAZIER, J. E. 2012. The development of a QALY measure for epilepsy: NEWQOL-6D. *Epilepsy & Behavior,* 24**,** 36-43.

MURAKI, E. 1992. A Generalized Partial Credit Model: Application of an EM Algorithm. *Applied Psychological Measurement,* 16**,** 159-176.

MURRAY, C. J. & FRENK, J. 2000. A framework for assessing the performance of health systems. *Bulletin of the World Health Organization,* 78**,** 717-731.

MUTHÉN, B. 1984. A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika,* 49**,** 115-132.

NARAYANAN, U. G., FEHLINGS, D., WEIR, S., KNIGHTS, S., KIRAN, S. & CAMPBELL, K. 2006. Initial development and validation of the Caregiver Priorities and Child Health Index of Life with Disabilities (CPCHILD). *Developmental Medicine & Child Neurology,* 48**,** 804-12.

NHS. 2013. Friends and Family Test *https://www.england.nhs.uk/fft/* [Online]. [Accessed January 10, 2018].

NICE 2013b. Guide to the methods of technology appraisal 2013: Process and methods guides. London: National Institute for Health and Care Excellence.

OHE 2008. NHS Outcomes, Performance and Productivity. *In:* HTTP://WWW.OHE.ORG/PUBLICATIONS/ARTICLE/NHS-OUTCOMES-PERFORMANCE-AND-PRODUCTIVITY-42.CFM (ed.) *Office of Health Economics Commission.*

OPPE, M., DEVLIN, N. J., VAN HOUT, B., KRABBE, P. F. & DE CHARRO, F. 2014. A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol. *Value Health,* 17**,** 445-53.

PETERSEN, M. A., GROENVOLD, M., AARONSON, N., BLAZEBY, J., BRANDBERG, Y., DE GRAEFF, A., FAYERS, P., HAMMERLID, E., SPRANGERS, M., VELIKOVA, G. & BJORNER, J. B. 2006. Item response theory was used to shorten EORTC QLQ-C30 scales for use in palliative care. *J Clin Epidemiol,* 59**,** 36-44.

PETRILLO, J., CANO, S. J., MCLEOD, L. D. & COON, C. D. 2015. Using classical test theory, item response theory, and Rasch measurement theory to evaluate patient-reported outcome measures: A comparison of worked examples. *Value in Health,* 18**,** 25-34.

PICKER 2012. *Guidance manual for the NHS Adult Inpatient Survey 2012,* Oxford, Picker Institute Europe.

PROMIS. 2013. *Instrument Development and Psychometric Evaluation Scientific Standards* [Online]. [Accessed 1st September 2016].

RASCH, G. 1960. *Probabilistic model for some intelligence and achievement tests,* Copenhagen, Danish Institute for Educational Research.

REEVE, B. B., HAYS, R. D., BJORNER, J. B., COOK, K. F., CRANE, P. K., TERESI, J. A., THISSEN, D., REVICKI, D. A., WEISS, D. J., HAMBLETON, R. K., LIU, H., GERSHON, R., REISE, S. P., LAI, J. S. & CELLA, D. 2007. Psychometric evaluation and calibration of health-related quality of life item banks: Plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Medical Care,* 45**,** S22-S31.

REEVES, R., COULTER, A., JENKINSON, C., CARTWRIGHT, J., BRUSTER, S. & RICHARD, N. 2002a. *Development and pilot testing of instruments for use in the acute NHS Trust Inpatient Survey Programme,* Oxford, Picker Institute.

REEVES, R., COULTER, A., JENKINSON, C., CARTWRIGHT, J., BRUSTER, S. & RICHARDS, N. 2002b. Development and Pilot Testing of Questionnaires for use in the Acute NHS Trust Inpatient Survey Programme http://www.nhssurveys.org/survey/8: Picker Institute Europe.

REEVES, R. & SECCOMBE, I. 2008. Do patient surveys work? the influence of a national survey programme on local quality-improvement initiatives. *Quality and Safety in Health Care,* 17**,** 437-441.

RIAZI, A., HOBART, J., LAMPING, D., FITZPATRICK, R. & THOMPSON, A. 2002. Multiple Sclerosis Impact Scale (MSIS-29): reliability and validity in hospital based samples. *Journal of Neurology, Neurosurgery & Psychiatry,* 73**,** 701-704.

RICHARDSON, J., IEZZI, A., KHAN, M. A. & MAXWELL, A. 2014. Validity and reliability of the assessment of quality of life (AQoL)-8D multi-attribute utility instrument. *Patient,,* 7**,** 85-96.
RICHARDSON, J., IEZZI, A., PEACOCK, S., SINHA, K., KHAN, M., MISAJON, R. & KEEFFE, J. 2012a. Utility weights for the vision-related Assessment of Quality of Life (AQoL)-7D instrument. *Ophthalmic Epidemiol,* 19**,** 172-82.
RICHARDSON, J. R. J., PEACOCK, S. J., HAWTHORNE, G., IEZZI, A., ELSWORTH, G. & DAY, N. A. 2012b. Construction of the descriptive system for the assessment of quality of life AQoL-6D utility instrument. *Health and Quality of Life Outcomes,* 10.
RODERICK, J. A. L. 1988. Missing-Data Adjustments in Large Surveys. *Journal of Business & Economic Statistics,* 6**,** 287-296.
SAMEJIMA, F. 1968. ESTIMATION OF LATENT ABILITY USING A RESPONSE PATTERN OF GRADED SCORES1. *ETS Research Bulletin Series,* 1968**,** i-169.
SITZIA, J. & WOOD, N. 1997a. Patient satisfaction: A review of issues and concepts. *Social Science and Medicine,* 45**,** 1829-1843.
SITZIA, J. & WOOD, N. 1997b. Patient satisfaction: A review of issues and concepts. *Social Science & Medicine,* 45**,** 1829-1843.
SIZMUR, S. & REDDING, D. 2009. Core domains for measuring inpatients' experience of care. *In:* EUROPE, P. I. (ed.). http://www.picker.org/wp-content/uploads/2014/10/Core-domains-for-measuring-inpatient-experience-of-care.pdf.
SIZMUR, S. & REDDING, D. 2012. *Core domains for measuring inpatients' experience of care,* Oxford, Picker Institute Europe.
SLADE, G. D. 1997. Derivation and validation of a short-form oral health impact profile. *Community Dent Oral Epidemiol,* 25**,** 284-90.
SMITH, R. D. & SACH, T. H. 2010. Contingent valuation: what needs to be done? *Health Econ Policy Law,* 5**,** 91-111.
STEVENS, K. 2009. Developing a descriptive system for a new preference-based measure of health-related quality of life for children. *Quality of Life Research,* 18**,** 1105-13.
STEVENS, K. 2011. Assessing the performance of a new generic measure of health-related quality of life for children and refining it for use in health state valuation. *Applied Health Economics & Health Policy,* 9**,** 157-69.
STEVENS, K. 2012. Valuation of the Child Health Utility 9D Index. *Pharmacoeconomics,* 30**,** 729-47.
STEWART, A. L., HAYS, R. D. & WARE, J. E., JR. 1988. The MOS short-form general health survey. Reliability and validity in a patient population. *Med Care,* 26**,** 724-35.
STREINER, D. L. & NORMAN, G. R. 2008. Health measurement scales: a practical guide to their development and use. *Health measurement scales: a practical guide to their development and use.* Oxford University Press.
SUNDARAM, M., SMITH, M. J., REVICKI, D. A., ELSWICK, B. & MILLER, L.-A. 2009. Rasch analysis informed the development of a classification system for a diabetes-specific preference-based measure of health. *Journal of Clinical Epidemiology,* 62**,** 845-56.
TERWEE, C. B., BOT, S. D., DE BOER, M. R., VAN DER WINDT, D. A., KNOL, D. L., DEKKER, J., BOUTER, L. M. & DE VET, H. C. 2007. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol,* 60**,** 34-42.
THOMPSON, B. 1995. Stepwise Regression and Stepwise Discriminant Analysis Need Not Apply here: A Guidelines Editorial. *Educational and Psychological Measurement,* 55**,** 525-534.
VALENTINE, N., DARBY, C. & BONSEL, G. J. 2008. Which aspects of non-clinical quality of care are most important? Results from WHO's general population surveys of "health systems responsiveness" in 41 countries. *Social Science & Medicine,* 66**,** 1939-1950.
WARE, J. E. & SHERBOURNE, C. D. 1992. The MOS 36-Item Short-Form Health Survey (SF-36): I. Conceptual Framework and Item Selection. *Medical Care,* 30**,** 473-483.
WARE JR, J. E., KOSINSKI, M. & KELLER, S. D. 1996. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Medical Care,* 34**,** 220-233.
WATT, T., BJORNER, J. B., GROENVOLD, M., CRAMON, P., WINTHER, K. H., HEGEDUS, L., BONNEMA, S. J., RASMUSSEN, A. K., WARE, J. E., JR. & FELDT-RASMUSSEN, U. 2015. Development of a Short Version of the Thyroid-Related Patient-Reported Outcome ThyPRO. *Thyroid,* 25**,** 1069-79.
WHALLEY, D., MCKENNA, S., DEWAR, A., ERDMAN, R., KOHLMANN, T., NIERO, M., COOK, S., CRICKX, B., HERDMAN, M. & FRECH, F. 2004. A new instrument for assessing quality

of life in atopic dermatitis: international development of the Quality of Life Index for Atopic Dermatitis (QoLIAD). *British Journal of Dermatology,* 150**,** 274-283.
WHO 1998. Development of the World Health Organization WHOQOL-BREF quality of life assessment. The WHOQOL Group. *Psychol Med,* 28**,** 551-8.
WHO 2000. Improving Performance: Health Systems. *The World Health Report.* http://www.who.int/whr/2000/en/whr00_en.pdf?ua=1: World Health Organization.
WILIIAMS, A. 1990. EuroQol - a new facility for the measurement of health-related quality of life. *Health Policy,* 16**,** 199-208. The EuroQol Group.
WILLIAMS, A. 1974. Measuring the Effectiveness of Health Care Systems. *British Journal of Preventive and Social Medicine,* 28**,** 196-202.
YOUNG, T., YANG, Y., BRAZIER, J. E., TSUCHIYA, A. & COYNE, K. 2008. The first stage of developing preference-based measures: constructing a health-state classification using Rasch analysis. *Quality of Life Research,* 18**,** 253.
YOUNG, T. A., ROWEN, D., NORQUIST, J. & BRAZIER, J. E. 2010. Developing preference-based health measures: using Rasch analysis to generate health state values. *Quality of Life Research,* 19**,** 907-17.
YOUNG, T. A., YANG, Y., BRAZIER, J. E. & TSUCHIYA, A. 2011. The use of Rasch analysis in reducing a large condition-specific instrument for preference valuation: the case of moving from AQLQ to AQL-5D. *Medical Decision Making,* 31**,** 195-210.

Appendix

Appendix 1: Ovid Medline and Scopus Search Results

Search carried out in Ovid database on 13/09/2016

Searches	Results	Type
1	Adult/px [Psychology]	62
2	validation studies.pt.	81104
3	reproducibility of results.sh.	325096
4	reproducib*.ti. or reproducib*.ab.	129361
5	psychometrics.sh.	61924
6	psychometr*.ti. or psychometr*.ab.	32800
7	clinimetr*.ti. or clinimetr*.ab.	701
8	clinometr*.ti. or clinometr*.ab.	23
9	observer variation.sh.	36244
10	observer variation.ti. or observer variation.ab.	927
11	discriminant analysis.sh.	8656
12	reliab*.ti. or reliab*.ab.	368234
13	valid*.ti. or valid*.ab.	511128
14	internal consistency.ti. or internal consistency.ab.	20268
15	cronbach.ti. or cronbach.ab.	3208
16	alpha.ti. or alpha.ab.	806100
17	15 and 16	2748
18	item correlation*.ti. or item correlation*.ab.	478
19	item selection*.ti. or item selection*.ab.	416
20	item reduction*.ti. or item reduction*.ab.	463
21	agreement.tw.	212508
22	precision.tw.	89829
23	imprecision.tw.	5052
24	precise values.tw.	213
25	test retest.ti. or test retest.ab.	18894
26	test.ti. or test.ab.	1149239
27	retest.ti. or retest.ab.	21067

28	26 and 27	19885
29	reliab*.ti. or reliab*.ab.	368234
30	28 and 29	16933
31	stability.ti. or stability.ab.	302622
32	interrater.ti. or interrater.ab.	6359
33	intrarater.ti. or intrarater.ab.	1644
34	intertester.ti. or intertester.ab.	274
35	inter tester.ti. or inter tester.ab.	140
36	intratester.ti. or intratester.ab.	210
37	intra-tester.ti. or intra-tester.ab.	105
38	interobserver.ti. or interobserver.ab.	14071
39	inter-observer.ti. or inter-observer.ab.	4898
40	intertechnician.ab. or intertechnician.ti.	4
41	inter-rater.ab. or inter-rater.ti.	6445
42	intra-rater.ab. or intra-rater.ti.	1361
43	intraobserver.ab. or intraobserver.ti.	5677
44	intra-observer.ab. or intra-observer.ti.	2715
45	inter-technician.ab. or inter-technician.ti.	12
46	intratechnician.ab. or intratechnician.ti.	2
47	intra-technician.ab. or intra-technician.ti.	3
48	interexaminer.ab. or interexaminer.ti.	701
49	inter-examiner.ab. or inter-examiner.ti.	599
50	intraexaminer.ab. or intraexaminer.ti.	370
51	intra-examiner.ab. or intra-examiner.ti.	459
52	interassay.ab. or interassay.ti.	2829
53	inter-assay.ab. or inter-assay.ti.	3910
54	intraassay.ab. or intraassay.ti.	801
55	intra-assay.ab. or intra-assay.ti.	3334
56	interindividual.ab. or interindividual.ti.	13880
57	inter-individual.ab. or inter-individual.ti.	7760
58	intraindividual.ab. or intraindividual.ti.	4657
59	intra-individual.ab. or intra-individual.ti.	3930

60	interparticipant.ab. or interparticipant.ti.	26
61	inter-participant.ab. or inter-participant.ti.	36
62	intraparticipant.ab. or intraparticipant.ti.	13
63	intra-participant.ab. or intra-participant.ti.	26
64	kappa*.ab. or kappa*.ti.	134863
65	coefficient of variation.ab. or coefficient of variation.ti.	18619
66	repeatab*.tw.	25266
67	(replicab* or repeated).tw.	239710
68	(measure* or finding* or result* or test*).tw.	10528952
69	67 and 68	177606
70	generaliza*.ab. or generaliza*.ti.	28653
71	generalisa*.ab. or generalisa*.ti.	3032
72	concordance.ab. or concordance.ti.	32912
73	intraclass.ab. or intraclass.ti.	16946
74	correlation*.ab. or correlation*.ti.	810328
75	73 and 74	16540
76	discriminative.ab. or discriminative.ti.	13069
77	known group.ab. or known group.ti.	690
78	factor analysis.ab. or factor analysis.ti.	29238
79	factor analyses.ab. or factor analyses.ti.	5302
80	factor structure*.ab. or factor structure*.ti.	9037
81	dimensionality.ab. or dimensionality.ti.	7729
82	subscale*.ab. or subscale*.ti.	31433
83	multitrait scaling analysis.ab. or multitrait scaling analysis.ti.	72
84	multitrait scaling analyses.ab. or multitrait scaling analyses.ti.	20
85	item discriminant.ab. or item discriminant.ti.	97
86	interscale correlation*.ab. or interscale correlation*.ti.	96
87	interscale correlation*.ab. or interscale correlation*.ti.	96
88	error*.ab. or error*.ti.	227105
89	measure*.ab. or measure*.ti. or correlat*.ab. or correlat*.ti. or evaluat*.ti. or evaluat*.ab. or accuracy.ab. or accuracy.ti.	6381278

	or accurate.ab. or accurate.ti. or precision.ab. or precision.ti. or mean.ab. or mean.ti.	
90	88 and 89	139614
91	individual variability.ab. or individual variability.ti.	6426
92	interval variability.ab. or interval variability.ti.	598
93	rate variability.ab. or rate variability.ti.	13860
94	variability analysis.ab. or variability analysis.ti.	902
95	uncertainty.ab. or uncertainty.ti.	50849
96	measurement.ab. or measurement.ti. or measuring.ab. or measuring.ti.	593313
97	95 and 96	5567
98	standard error of measurement.ab. or standard error of measurement.ti.	1132
99	sensitiv*.ab. or sensitiv*.ti.	1123910
100	responsive*.ab. or responsive*.ti.	189438
101	limit.ab. or limit.ti.	203916
102	detection.ab. or detection.ti. or minimal detectable concentration.ab. or minimal detectable concentration.ti. or interpretab*.ti. or interpretab*.ab.	697650
103	small*.ab. or small*.ti.	1327199
104	real.ab. or real.ti. or detectable.ab. or detectable.ti.	464162
105	change.ab. or change.ti. or difference.ab. or difference.ti.	1614307
106	103 and 104 and 105	5556
107	meaningful change.ab. or meaningful change.ti.	600
108	minimal important change.ab. or minimal important change.ti.	68
109	minimal important difference.ab. or minimal important difference.ti.	240
110	minimally important change.ab. or minimally important change.ti.	44
111	minimally important difference.ab. or minimally important difference.ti.	218
112	minimal detectable change.ab. or minimal detectable change.ti.	557

113	minimal detectable difference.ab. or minimal detectable difference.ti.	50
114	minimally detectable change.ab. or minimally detectable change.ti.	10
115	minimally detectable difference.ab. or minimally detectable difference.ti.	5
116	minimal real change.ab. or minimal real change.ti.	0
117	minimal real difference.ab. or minimal real difference.ti.	4
118	minimally real change.ab. or minimally real change.ti.	0
119	minimally real difference.ab. or minimally real difference.ti.	0
120	ceiling effect.ab. or ceiling effect.ti.	1286
121	floor effect.ab. or floor effect.ti.	361
122	Item response model.ab. or Item response model.ti.	89
123	Item response theory.ab. or Item response theory.ti.	1832
124	IRT.ab. or IRT.ti.	2025
125	Rasch.ab. or Rasch.ti.	2791
126	Differential item functioning.ab. or Differential item functioning.ti.	1047
127	DIF.ab. or DIF.ti.	1941
128	computer adaptive testing.ab. or computer adaptive testing.ti.	125
129	item bank.ab. or item bank.ti.	351
130	cross-cultural equivalence.ab. or cross-cultural equivalence.ti.	96
131	structural equation model*.ab. or structural equation model*.ti.	10203
132	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 or 45 or 46 or 47 or 48 or 49 or 50 or 51 or 52 or 53 or 54 or 55 or 56 or 57 or 58 or 59 or 60 or 61 or 62 or 63 or 64 or 65 or 66 or 69 or 70 or 71 or 72 or 75 or 76 or 77 or 78 or 79 or 80 or 81 or 82 or 83 or 84 or 85 or 86 or 87 or 90 or 91 or 92 or 93 or 94 or 97 or 98 or 99 or 100 or 101 or 102 or 106 or 107 or 108 or 109 or 110 or 111 or 112 or 113 or 114 or 115 or 116 or 117 or 118 or 119 or 120	3644443

	or 121 or 122 or 123 or 124 or 125 or 126 or 127 or 128 or 129 or 130 or 131	
133	"Quality of Life"/ or "Outcome Assessment (Health Care)"/ or "Outcome and Process Assessment (Health Care)"/ or "Surveys and Questionnaires".mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]	528944
134	descriptive system.mp.	169
135	Classification/mt, st [Methods, Standards]	2140
136	classification system.mp.	14158
137	134 or 135 or 136	16383
138	(address* or biograph* or case report* or comment or directory or editorial or festschrift or interview or lecture* or legal case* or legislation* or letter* or news or newspaper article or patient education handout or popular work* or congresses or consensus development conference or consensus development conference, nih or practice guideline).pt. not animal*.sh. not human*.sh.	353702
139	132 and 133 and 137	444
140	139 not 138	444

Search carried out in Scopus on 09/01/2017

Searches	Results	Type
1	(instrumentation[sh] OR validation studies[pt] OR ''reproducibility of results''[mesh terms] OR reproducib*[tiab] OR ''psychometrics''[mesh] OR psychometr*[tiab] OR clinimetr*[tiab] OR clinometr*[tiab] OR ''observer variation''[mesh] OR observer variation[tiab] OR ''discriminant analysis''[mesh] OR reliab*[tiab] OR valid*[tiab] OR coefficient[tiab] OR ''internal consistency''[tiab] OR (cronbach*[tiab] AND (alpha[tiab] OR alphas[tiab])) OR ''item correlation''[tiab] OR ''item correlations''[tiab] OR ''item selection''[tiab] OR ''item selections''[tiab] OR ''item reduction''[tiab] OR ''item reductions''[tiab] OR agreement[tw] OR	120,367 document results

	precision[tw] OR imprecision[tw] OR ''precise values''[tw] OR test--retest [tiab] OR (test[tiab] AND retest[tiab]) OR (reliab*[tiab] AND (test[tiab] OR retest[tiab])) OR stability[tiab] OR interrater[tiab] OR inter-rater[tiab] OR intrarater[tiab] OR intra-rater[tiab] OR intertester[tiab] OR inter-tester[tiab] OR intratester[tiab] OR intra-tester[tiab] OR interobserver[tiab] OR inter-observer[tiab] OR intraobserver[tiab] OR intra-observer[tiab] OR intertechnician[tiab] OR intertechnician[tiab] OR intratechnician[tiab] OR intra-technician[tiab] OR interexaminer[tiab] OR inter-examiner[tiab] OR intraexaminer[tiab] OR intra-examiner[tiab] OR interassay[tiab] OR inter-assay[tiab] OR intraassay[tiab] OR intra-assay[tiab] OR interindividual[tiab] OR inter-individual[tiab] OR intraindividual[tiab] OR intra-individual[tiab] OR interparticipant[tiab] OR inter-participant[tiab] OR intraparticipant[tiab] OR intra-participant[tiab] OR kappa[-tiab] OR kappa's[tiab] OR kappas[tiab] OR ''coefficient of variation''[tiab] OR repeatab*[tw] OR ((replicab*[tw] OR repeated[tw]) AND (measure[tw] OR measures[tw] OR findings[tw] OR result[tw] OR results[tw] OR test[tw] OR tests[tw])) OR generaliza*[tiab] OR generalisa*[tiab] OR concordance[tiab] OR (intraclass[tiab] AND correlation*[tiab]) OR discriminative[tiab] OR ''known group'' [tiab] OR ''factor analysis''[tiab] OR ''factor analyses''[tiab] OR ''factor structure''[tiab] OR ''factor structures''[tiab] OR dimensionality[tiab] OR subscale*[tiab] OR ''multitrait scaling analysis''[tiab] OR ''multitrait scaling analyses''[tiab] OR ''item discriminant''[tiab]or ''interscale correlation''[tiab] OR ''interscale correlations''[tiab] OR ((error[tiab] OR errors[tiab]) AND (measure*[tiab] OR correlat*[tiab] OR evaluat*[tiab] OR accuracy[tiab] OR accurate[tiab] OR precision[tiab] OR mean[tiab])) OR ''individual variability''[tiab] OR ''interval variability''[tiab] OR ''rate variability''[tiab] OR ''variability analysis''[tiab] OR (uncertainty[tiab] AND (measurement[tiab] OR measuring[tiab])) OR ''standard error of measurement''[tiab] OR sensitiv*[tiab] OR responsive*[tiab] OR (limit[tiab] AND detection[tiab]) OR ''minimal detectable concentration''[tiab]orinterpretab*[tiab] OR (small*[tiab]	

	AND (real[tiab] OR detectable[tiab]) AND (change[tiab]or difference[tiab])) OR ''meaningful change''[tiab] OR ''minimal important change''[tiab] OR ''minimal important difference''[tiab] OR ''minimally important change''[tiab] OR ''minimally important difference''[tiab] OR ''minimal detectable change''[tiab] OR ''minimal detectable difference''[tiab] OR ''minimally detectable change''[tiab] OR ''minimally detectable difference''[tiab] OR ''minimal real change''[tiab] OR ''minimal real difference''[tiab] OR ''minimally real change''[tiab] OR ''minimally real difference''[tiab] OR ''ceiling effect''[tiab] OR ''floor effect'' [tiab] OR ''item response model''[tiab] OR irt[tiab] OR rasch[tiab] OR ''differential item functioning''[tiab] OR dif[tiab] OR ''computer adaptive testing''[tiab] OR ''item bank''[tiab] OR ''cross-cultural equivalence''[tiab]) OR structural equation model* AND (LIMIT-TO (SRCTYPE , "j ")) AND (LIMIT-TO (DOCTYPE , "ar ")) AND (LIMIT-TO (SUBJAREA , "PSYC ") OR LIMIT-TO (SUBJAREA , " MEDI ") OR LIMIT-TO (SUBJAREA , " SOCI ") OR LIMIT-TO (SUBJAREA , " ECON ") OR LIMIT-TO (SUBJAREA , " PHAR ") OR LIMIT-TO (SUBJAREA , " HEAL ")) AND (LIMIT-TO (LANGUAGE , "English "))	
2	TITLE-ABS-KEY (quality of life) OR TITLE-ABS-KEY (outcome assessment) OR TITLE-ABS-KEY (outcome AND process assessment) OR KEY (surveys AND questionnaires) AND (LIMIT-TO (SRCTYPE , "j ")) AND (LIMIT-TO (DOCTYPE , "ar ")) AND (LIMIT-TO (SUBJAREA , "PSYC ") OR LIMIT-TO (SUBJAREA , " MEDI ") OR LIMIT-TO (SUBJAREA , " SOCI ") OR LIMIT-TO (SUBJAREA , " ECON ") OR LIMIT-TO (SUBJAREA , " PHAR ") OR LIMIT-TO (SUBJAREA , " HEAL ")) AND (LIMIT-TO (LANGUAGE , "English "))	693,870 document results
3	(TITLE-ABS-KEY (classification system)) OR (TITLE-ABS-KEY (descriptive system))	341,216 document results
4	((instrumentation[sh] OR validation studies[pt] OR ''reproducibility of results''[mesh terms] OR reproducib*[tiab] OR ''psychometrics''[mesh] OR psychometr*[tiab] OR clinimetr*[tiab] OR clinometr*[tiab] OR ''observer variation''[mesh] OR observer variation[tiab] OR ''discriminant analysis''[mesh] OR	109 document results

	reliab*[tiab] OR valid*[tiab] OR coefficient[tiab] OR ''internal consistency''[tiab] OR (cronbach*[tiab] AND (alpha[tiab] OR alphas[tiab])) OR ''item correlation''[tiab] OR ''item correlations''[tiab] OR ''item selection''[tiab] OR ''item selections''[tiab] OR ''item reduction''[tiab] OR ''item reductions''[tiab] OR agreement[tw] OR precision[tw] OR imprecision[tw] OR ''precise values''[tw] OR test--retest [tiab] OR (test[tiab] AND retest[tiab]) OR (reliab*[tiab] AND (test[tiab] OR retest[tiab])) OR stability[tiab] OR interrater[tiab] OR inter-rater[tiab] OR intrarater[tiab] OR intra-rater[tiab] OR intertester[tiab] OR inter-tester[tiab] OR intratester[tiab] OR intra-tester[tiab] OR interobserver[tiab] OR inter-observer[tiab] OR intraobserver[tiab] OR intra-observer[tiab] OR intertechnician[tiab] OR intertechnician[tiab] OR intratechnician[tiab] OR intra-technician[tiab] OR interexaminer[tiab] OR inter-examiner[tiab] OR intraexaminer[tiab] OR intra-examiner[tiab] OR interassay[tiab] OR inter-assay[tiab] OR intraassay[tiab] OR intra-assay[tiab] OR interindividual[tiab] OR inter-individual[tiab] OR intraindividual[tiab] OR intra-individual[tiab] OR interparticipant[tiab] OR inter-participant[tiab] OR intraparticipant[tiab] OR intra-participant[tiab] OR kappa[-tiab] OR kappa's[tiab] OR kappas[tiab] OR ''coefficient of variation''[tiab] OR repeatab*[tw] OR ((replicab*[tw] OR repeated[tw]) AND (measure[tw] OR measures[tw] OR findings[tw] OR result[tw] OR results[tw] OR test[tw] OR tests[tw])) OR generaliza*[tiab] OR generalisa*[tiab] OR concordance[tiab] OR (intraclass[tiab] AND correlation*[tiab]) OR discriminative[tiab] OR ''known group'' [tiab] OR ''factor analysis''[tiab] OR ''factor analyses''[tiab] OR ''factor structure''[tiab] OR ''factor structures''[tiab] OR dimensionality[tiab] OR subscale*[tiab] OR ''multitrait scaling analysis''[tiab] OR ''multitrait scaling analyses''[tiab] OR ''item discriminant''[tiab]or ''interscale correlation''[tiab] OR ''interscale correlations''[tiab] OR ((error[tiab] OR errors[tiab]) AND (measure*[tiab] OR correlat*[tiab] OR evaluat*[tiab] OR accuracy[tiab] OR accurate[tiab] OR precision[tiab] OR mean[tiab])) OR ''individual variability''[tiab] OR ''interval variability''[tiab] OR ''rate variability''[tiab] OR ''variability analysis''[tiab] OR (	

	uncertainty[tiab] AND (measurement[tiab] OR measuring[tiab])) OR ''standard error of measurement''[tiab] OR sensitiv*[tiab] OR responsive*[tiab] OR (limit[tiab] AND detection[tiab]) OR ''minimal detectable concentration''[tiab]orinterpretab*[tiab] OR (small*[tiab] AND (real[tiab] OR detectable[tiab]) AND (change[tiab]or difference[tiab])) OR ''meaningful change''[tiab] OR ''minimal important change''[tiab] OR ''minimal important difference''[tiab] OR ''minimally important change''[tiab] OR ''minimally important difference''[tiab] OR ''minimal detectable change''[tiab] OR ''minimal detectable difference''[tiab] OR ''minimally detectable change''[tiab] OR ''minimally detectable difference''[tiab] OR ''minimal real change''[tiab] OR ''minimal real difference''[tiab] OR ''minimally real change''[tiab] OR ''minimally real difference''[tiab] OR ''ceiling effect''[tiab] OR ''floor effect'' [tiab] OR ''item response model''[tiab] OR irt[tiab] OR rasch[tiab] OR ''differential item functioning''[tiab] OR dif[tiab] OR ''computer adaptive testing''[tiab] OR ''item bank''[tiab] OR ''cross-cultural equivalence''[tiab]) OR structural equation model*) AND (TITLE-ABS-KEY (quality of life) OR TITLE-ABS-KEY (outcome assessment) OR TITLE-ABS-KEY (outcome AND process assessment) OR KEY (surveys AND questionnaires)) AND ((TITLE-ABS-KEY (classification system)) OR (TITLE-ABS-KEY (descriptive system))) AND (LIMIT-TO (SRCTYPE , "j ")) AND (LIMIT-TO (DOCTYPE , "ar ")) AND (LIMIT-TO (SUBJAREA , "PSYC ") OR LIMIT-TO (SUBJAREA , " MEDI ") OR LIMIT-TO (SUBJAREA , " SOCI ") OR LIMIT-TO (SUBJAREA , " ECON ") OR LIMIT-TO (SUBJAREA , " PHAR ") OR LIMIT-TO (SUBJAREA , " HEAL ")) AND (LIMIT-TO (LANGUAGE , "English "))	

Appendix 2: Methods used to develop descriptive systems

Measure	Place	Condition	Item generation	Item selection
SF-20	USA	Generic	RAND Health Insurance Experiment (HIE)	20 items were selected to represent six health concepts. Eighteen of the 20 items were adapted longer HIE measure and two additional single-item measures (social functioning and pain) were included from similar measures.
EQ-5D	Multi-country	Generic	Detailed examination of the descriptive content of existing health status measures including the Quality of Well Being Scale, the Sickness Impact Profile, Nottingham Health Profile and the Rosser Index. Additional measures used by members of the EuroQol Group were also included.	Items were selected by researchers so as to cover as many as possible of the domains frequently covered by others; based on further development work, the instrument was modified, to produce by October 1991 a standard five-dimensional format
AQLQ	Canada	Asthma	Review of studies with severe asthma patients, existing measures, experience of patients with chronic airflow limitation, discussion with local chest physicians and interviews with patients (N=6)	Respondents (N=150) were asked which of the items had been troublesome to them at any time during the past year and to indicate the importance of each of the identified items on a five point scale.
SF-36	USA	Generic	Researchers	Items in SF-36 scale were

Measure	Place	Condition	Item generation	Item selection
			selected and adapted items from existing measures (General Psychological Well-Being Inventory, various physical and role functioning measures, Health Perceptions Questionnaire, and measures used in Health Insurance Experiment) and other sources and developed new measures for a 149-item Functioning and Well-Being Profile (FWBP).	selected to reproduce the "parent" scale and other psychometric standards. The specific strategies used varied across the domains.
SF-12	USA	Generic	35 items (over 8 dimensions) from SF-36	Two secondary datasets were used (public in NSFHS, N=2474 and patients in MOS, N=889). Forward-step regression analysis was used to identify a subset of 12 or fewer items from the SF-36 and 2 weighting algorithms for estimating physical and mental health component scale.
AQLQ	Canada	Asthma	152 items identified earlier	For the impact method, items that were identified most frequently and that scored the highest were included in the final instrument (N=150). For the psychometric method, factor analysis was performed after highly skewed items had

Measure	Place	Condition	Item generation	Item selection
				been removed (N=?).
RAQoL	UK and Netherlands	Rheumatoid arthritis	Qualitative interviews were conducted in two countries (N=25 each) and items included reflected the frequency with which issues were raised by the interviewees.	Two pilot studies were conducted (N=50 each). Items were removed from draft questionnaire based on internal consistency, correlation with other items (too low/high) and distribution of responses (skewed). Further interviews were conducted during field testing (N=15 each) to examine relevance and acceptability. In addition validation survey was conducted.
Short-form of OHIP	Australia	Oral health	OHIP measure containing 49 items across 7 dimensions	Secondary dataset (N=1217) was used. Internal reliability analysis using Cronbach's alpha, PCA and regression analysis (step-wise) were undertaken to derive a subset.
WHOQOL-BREF	Multi-country	Generic	WHOQOL-100 with 100 items covering 25 facets organised in 6 domains	The most general question from each facet (i.e. the item that correlated most highly with the total score, calculated as the mean of all facets) was chosen.
AQoL (AQoL-4D)	Australia	Generic	Items from 14 existing QoL measures were pooled and synthesised into a model by researchers and reviewed/revised by medical specialists and general practitioners (N=24) in focus	Final item bank was administered to hospital patients, and community members (N=996). PCA; EFA and SEM were applied to dataset to determine dimension and items.

Measure	Place	Condition	Item generation	Item selection
			groups	
ALSAQ-40	UK	Amyotrophic lateral sclerosis/motor neurone disease	In-depth, semi-structured exploratory interviews (N=18)	The data (N=173) were factor analysed (varimax rotation) to determine the underlying dimensions. Items were selected based on analyses to areas measured by the instrument.
McSad	Canada	Major, unipolar depression	Items extracted from Diagnostics and Statistical Manual for Mental Disorders symptom criteria for the diagnosis of major, unipolar depression	Items selected by those involved
PORPUS	UK	Prostate cancer	Empirical studies of HRQoL measurement in prostate cancer and generic, cancer-specific instruments were reviewed. Clinical experts (n=10) and patients (n=80) rated the importance of items for each domain.	Key concepts were selected for each domain using item importance weightings, and a set of predetermined criteria.
The Endometriosis Health Profile-30	UK	Endometriosis	Open-ended exploratory interviews (N=25). A pilot study checked the face validity of the item generated (N=20).	PCA was applied to data (N=1000) identify the most salient dimensions of health-related quality of life. In addition reliability and validity of the questionnaire were assessed.
OAB-q	USA	Overactive bladder	Literature review and focus groups (N=16)	Data included community sample (N=254) and a clinical study (N=736). Subscales identified using EFA. Items were excluded

Measure	Place	Condition	Item generation	Item selection
				if high floor/ceiling responses, low item to total correlations or inadequate factor loading on any factor or on more than one factor.
SF-6D	UK	Generic	35 items (over 8 dimensions) from SF-36	The number of dimensions was reduced from 8 to 6 by the researchers. Excluding general health and combining the 2 role limitation dimensions achieved this. The 6 dimensions are presented as 6 items and capture 11 items from the SF-36.
Short-form of HIT-6	Netherlands	Headache	HIT item pool of 54 items (reported elsewhere) and additional 35 items suggested by clinicians.	10 candidate items from HIT and additional 35 items suggested by clinicians were administered (N=459 by phone and N=601 over the internet). Items were selected and modified based on content validity and models using IRT.
PBM derived from SF-12	UK	Generic	12 items (over 8 dimensions) from SF-12	The number of dimensions was reduced from 8 to 6 by excluding general health and combining the 2 role limitation dimensions. The number of items per dimension was reduced to one based on findings from two other studies (using Rasch, correlation and regession analyses). The 6 dimensions are presented as 6 items and capture 7 items from the SF-12.
QoLIAD	Multi-country	Atopic dermatitis	In-depth interviews in three	Comments from patients in field-test interviews

Measure	Place	Condition	Item generation	Item selection
			countries (N=65)	(N=20) in each of the 5 countries to examine relevance and acceptability. Rasch model applied to completed surveys (N=286 in UK, N=46 in the Netherlands, N=213 in France, N=187 in Germany, N=178 in the US and N=83 in Spain).
QoLIAD; PIQoL-AD; PSORIQoL	Multi-country	Atopic dermatitis (AD) (adults); parents of child with AD; psoriasis (adults)	For both QoLIAD and PIQoL-AD, interviews were conducted with relevant adults (N=65). Similar qualitative interviews were conducted for PSORIQoL (N=62)	The appropriateness and acceptability of the new measures was evaluated by semi-structured interviews (N=20) in each of the five countries. Completed questionnaires for QoLIAD (N=1085), PIQoL-AD (N=979) and PSORIQoL (N=148) were analysed using Rasch for item selection.
Short form of EORTC QLQ-C30 scale (emotional functioning)	Canada	Cancer patients in palliative care	EORTC QLQ-C30 consisting of 30 items; EF scale has four items	Secondary data (N=8242) from 24 European cancer studies conducted in 10 different languages was analysed using IRT (Generalized Partial Credit Model) to select items in EF scale.
PFDI-20 and PFIQ-7	UK	Pelvic floor disorders	Pelvic Floor Distress Inventory and Pelvic Floor Impact Questionnaire	Data from women (N=100) on Pelvic Floor Distress Inventory and Pelvic Floor Impact Questionnaire long forms was used. Subsets regression analysis was used to find the items in each scale that best predicted the scale score on the respective long form.
VisQol	Australia	Vision-	Three focus	Item bank was

Measure	Place	Condition	Item generation	Item selection
		impairment	groups were with 8-9 visually impaired participants each	administered (N=156); EFA, IRT and SEM analyses were conducted.
CPCHILD	Canada	Children with severe cerebral palsy (CP)	Items were generated from primary caregivers using unstructured interviews and review of other outcome measures.	A pilot version of the new questionnaire was constructed and multiple iterations were carried out in which items were added, deleted and modified (N=77). Caregivers' rated the importance of each of the items and those rated poorly were dropped.
EORTC QLQ-C15-PAL	Multi-country	Cancer patients in palliative care	EORTC QLQ-C30 consisting of 30 items	Item selection was based on interviews with patients (N=41) and health care professionals (N=66) in palliative care. They were asked to determine the appropriateness, relevance and importance of items and scales of the QLQ-C30
ICECAP-O	UK	Capability wellbeing for older people	In-depth interviews (N=40 older people)	Semi-structured interview with informants (N=19) interviewed previously.
Item bank for knee patholog y	Denmark	Knee conditions	Exhaustive literature search identified 31 instruments with 87 separate sub-domains and 539 items	Content redundancy and item reduction was carried out to isolate items of unique content.
QualiPau se toolkit	UK	Menopause	Items identified as being important on the grounds of two focus group sessions with peri- and post-menopausal women, literature review, and expert	Based on fulfilment of retest reliability, face validity, construct validity and convergent validity*

Measure	Place	Condition	Item generation	Item selection
			opinion*	
Short form of the EORTC QLQ-C30 scales	Multi-country	Cancer patients in palliative care	EORTC QLQ-C30 consisting of 30 items; 5-item physical (PF) scale; 3-item fatigue (FA) scale; 2-item nausea (NV) scale; 2-item cognitive (CF) scale	The shortening was based on 2,366 (PF) and 10,815 (three other scales) observations, respectively. IRT-based methods were used for the selection of items. For FA scale, the generalized partial credit model (GPCM) was estimated. And for other scales (PF, NV, and CF) the more restrictive partial credit model (PCM) was used.
Short-form of CPQ11–14	Multi-country	Oral health in children (age 11-14)	CPQ11–14 which consists 37 items organised into 4 domain	Item impact was carried out by children (N=83) and items deemed most important (top 4 and 2) from each domain were selected for first two versions. A single model was generated with all items included and a forward stepwise procedure to identify best predictors of the overall score. Top 4 and 2 items from each domain that made the largest contribution to the coefficient of variation were selected to form short forms.
PROMIS	USA	Physical functioning, fatigue, pain, emotional distress, and social role participation	Step-wise qualitative item review process that included: identification of extant items, item classification and selection, item revision, focus group exploration of domain	Datasets containing PRO response with general population (n=7523), disease population with cancer (n=1000), heart disease (n=500), rheumatoid arthritis (n=500), osteoarthritis (n=500), psychiatric conditions (n=500), spinal cord injury (n=500), and

Measure	Place	Condition	Item generation	Item selection
			coverage, cognitive interviews on individual items, and final revision before field testing	chronic obstructive pulmonary disease (n=500) was used. Analyses included evaluation of data quality, descriptive statistics, IRT model assumptions, model fit, differential item functioning, and item calibration for banking.
AQoL-8	Australia	Generic	Items from AQoL-4D	AQoL validation database (N=996) were reanalysed using Mokken IRT and Rasch to identify the least fitting items. Regression models were constructed to ensure items selected closely approximated the original AQoL descriptive system
CHU-9D	UK	Generic (children)	Interviews with children (N=74) were analysed using thematic content analysis to identify dimensions. To identify levels, ranking work was piloted with children (N=10) before main study (N=31)	Data (N=247) were collected to examine practicality (including response rates, completion rates and time to complete), validity (content, correlation with self-assessed health and known group validity), whether the child could self-complete and item presentation.
DUI	USA	Diabetes	ADDQoL items were used and coded according to a framework to assess of patients with diabetes by Polonsky.	A data set containing patient-reported impact ratings for each ADDQoL item (N=385) was analysed using factor analysis. An expert panel (N=7) reviewed the results of the factor analysis and rated importance of items. Rasch analysis was used on the data obtained

Measure	Place	Condition	Item generation	Item selection
				pilot rounds (N1=52, N2=65 and N3=111) for attribute selection and construction of severity levels for each attribute. CFA and Rasch was again applied to data from validation survey (N=396)
L-QoL	UK	Systemic lupus erythematosus	Content was derived from in-depth interviews with relevant patients (N=50).	Rasch analysis was applied to data (N=95) to remove misfitting items. It was reapplied to another survey (N=93) to validate findings.
CORE-6D	UK	Common mental health problems	CORE-OM with 34 items	Rasch analysis was used to reduce the number of items (N=400). The findings of Rasch analysis findings were validated on another random sample (N=400)
AQL-5D	Denmark	Asthma	AQLQ containing 32 items across 4 domains	PCA was used to confirm dimensionality of the AQLQ (N=413). Rasch analyses were conducted in remaining sample for validation (N=1706)
MobQues 28	Netherlands	Mobility in children with cerebral palsy	MobQues47 (version 1.3) instrument with 47 items	Rasch analysis applied to completed questionnaires (N=246)
PBM derived from EORTC QLQ-C30	Canada	Cancer	EORTC QLQ-C30	Factor analysis, Rasch analysis, and other psychometric analyses were undertaken on a clinical trial dataset (N=655)
AQoL-6D	Australia	Generic	Four focus groups (N=22) were conducted and items identified were combined with those obtained during the construction of AQoL-4D.	General public (N=316) and patients (N=304) completed the questionnaire. EFA and SEM were employed to determine dimensions and item selection.

Measure	Place	Condition	Item generation	Item selection
AQoL-8D	Australia	Generic	Items from AQoL (or AQoL-4D) and AQoL-6D and new ones identified from four focus groups with mental health patients	Items were administered to a representative sample of general public (N=195) and mental health patients (N=514). A combination of restrictive and unrestrictive factor analyses was used.
ICECAP-A	UK	Capability wellbeing for adults	In-depth interviews (N=36)	Semi structured interviews were used to select one item per attribute (N=18)
NEWQOL-6D	UK	Epilepsy	NEWQOL	Data from SANAG study (N=1611) was used to determine dimensions using EFA and input from epilepsy clinicians. Rasch analysis was used for item selection and this was validated on a second subset of the data using Rasch again.
PBM derived from DEMQOL and DEMQOL-Proxy	Canada	Dementia	DEMQOL and DEMQOL-Proxy measures (carer proxy report)	Dataset using DEMQOL (N=644) and for DEMQOL-Proxy (N=683) analysed using principal axis factoring with varimax rotation and Rasch analysis
Simplified version of the NEI VFQ-25	USA	Vision related	NEI VFQ-25	Data from patients with central (n=932) and peripheral vision loss (n=2,451) was examined using factor analysis and Rasch models.
Specific Thalassemia Quality of Life Instrument (STQOLI)	Greece	Thalassemia	Literature review and critical incident technique (N=10) analysis was used (N=10) to generate items and identify domains. This version was pre-piloted with	PCA and psychometric criteria was applied to dataset (N=128).

Measure	Place	Condition	Item generation	Item selection
			thalassemia patients (N=10) and experts (N=13)	
CAT-QoL	UK	Child Amblyopia	Semi-structured interviews were undertaken (N=59 children with amblyopia)	A conceptual framework was used to identify themes in the data
CP QOL-Teen	Australia	Adolescents with cerebral palsy.	Interviews were conducted with adolescents with cerebral palsy (N=17) and primary caregivers (N=23) within their home	PCA was fitted to data collected from adolescents with cerebral palsy (N=87) and primary caregivers (N=112)
AQoL-7D	Australia	Vision related	VisQoL and the AQoL-6D	All 20 from AQoL-6D and all 6 from VisQoL were combined
IUI	Spain	Urinary conditions	Used existing measures: Incontinence Quality of Life Questionnaire (I-QOL) and Neurogenic Module	Dimensionality was investigated using PCA (N=691) and item responses were analysed using the Partial Credit Model
P-PBMSI	UK	Multiple sclerosis	Items from RAND-36, the EQ-5D, the Patient Generated Index (PGI), the Perceived Deficits Questionnaire (PDQ), the Six-Minute Walk Test (6MWT) and the EDSS was included	Rasch analysis was applied to dataset (N=189)
MSIS-8D	UK	Multiple sclerosis	Multiple Sclerosis Impact Scale (MSIS-29)	Dimensional structure determined using factor analysis (N=529). Item selection involved Rasch analysis and psychometrics.
PBM for Myelofibr	UK	Myelofibrosis	MF-SAF 2.0 and the EORTC QLQ-	Factor analysis was applied to the data

Measure	Place	Condition	Item generation	Item selection
osis			C30	(N=309). Items selected had low levels of missing data, high correlation with the dimension and responses across severity range. Rasch analyses were used to select an item for each dimension and validated with experts.
ThyPRO	Denmark	Thyroid	ThyPRO consists of 85 items summarized in 13 scales	One scale was retained in full length and one was excluded because of high missing responses. For each of the remaining 11 scales, graded IRT model was used for item selection.

Appendix 3: Psychometric properties of descriptive systems included in the study

Measure	Content validity	Range and distribution of	Internal consistency	Construct validity (hypothesis testing)	Test re-test	Construct validity (measurement model)
AQLQ	Y	Y				Y
ALSAQ-	Y	Y	Y	Y	Y	Y
OAB-q	Y					
VisQol	Y		Y			Y
PORPUS	Y					
CPCHILD	Y	Y	Y	Y		Y
CP QOL-	Y			Y	Y	Y
The Endometriosis	Y	Y	Y	Y		
STQOLI	Y			Y		Y
RAQoL	Y	Y	Y	Y		
ICECAP-O	Y	Y	Y	Y		
ICECAP-A	Y	Y	Y	Y		
CHU-9D	Y	Y	Y	Y		
CAT-QoL	Y					
QoLIAD; PIQoL-AD; PSORIQol	Y	Y	Y	Y	Y	Y
L-QoL	Y	Y	Y		Y	Y
WHOQO	Y	Y	Y	Y	Y	Y
AQoL-8D	Y		Y			Y
PBM for	Y			Y	Y	
PBM for DEMQOL & DEMQOL	Y	Y	?	Y		Y
NEWQOL	Y	?	Y			Y
Short form of	Y					
MSIS-8D	Y	Y	Y	Y	Y	Y
PBM for Myelofibrosis	Y	Y	Y	Y		Y
Short-form for emotiona	Y	Y	Y			Y

Measure	Content validity	Range and distribution of	Internal consistency	Construct validity (hypothesis testing)	Test re-test	Construct validity (measurement model)
Short-form of 3 EORTC QLQ-C30	Y					
Short form of ThyPRO	Y					
AQoL-8	Y		Y		Y	Y
EORTC QLQ-C15-PAL	Y	Y	Y			Y
DUI	Y	Y	Y	Y		
Short form of $CPQ_{11–14}$	Y	Y	Y	Y	Y	Y
IUI	Y		Y	Y		Y
Short-form of OHIP	Y		Y	Y	Y	Y
AQL-5D	Y	Y	Y	Y	Y	Y
AQoL (AQoL-4D)	Y	Y	Y	Y		Y
SF-12	Y	Y	Y	Y		
PFDI-20; PFIQ-7	Y					Y
MobQues28	Y	Y	Y	Y		
P-PBMSI	Y					
CORE-6D	Y	Y	Y			Y
AQoL-6D	Y		Y	Y		Y
HIT-6	Y		Y	Y		Y

Appendix 4: Distribution of responses in the Inpatient Survey 2014

	Question	Response categories	Frequency	Percent
	Department of Health NHS Trust code			
	Length of Stay			
Q1	Was your most recent hospital stay planned in advance or an emergency?			
		Emergency or urgent	35,884	57.47%
		Waiting list or planned in advance	22,129	35.44%
		Something else	1,755	2.81%
		Missing responses	2,675	4.28%
		Total	62,443	100.00%
Q2	When you arrived at the hospital, did you go to the A&E Department (the Emergency Department / Casualty /Medical or Surgical Admissions unit)?			
		Yes	33,211	53.19%
		No	5,200	8.33%
		Inapplicable	22,129	35.44%
		Missing responses	1,903	3.05%
		Total	62,443	100.00%
Q3	While you were in the A&E Department, how much information about your condition or treatment was given to you?			
		Not enough	4,583	7.34%
		Right amount	21,342	34.18%
		Too much	114	0.18%
		Not given any information	2,744	4.39%
		Don't know / can't remember	4,110	6.58%
		Inapplicable	25,250	40.44%
		Missing responses	4,300	6.89%
		Total	62,443	100.00%
Q4	Were you given enough privacy when being examined or treated in the A&E Department?			
		Yes, definitely	24,369	39.03%
		Yes, to some extent	6,444	10.32%
		No	720	1.15%
		Don't know / can't remember	1,725	2.76%
		Inapplicable	25,250	40.44%
		Missing responses	3,935	6.30%
		Total	62,443	100.00%
Q5	When you were referred to see a specialist, were you offered a choice of hospital?			
		Yes	6,967	11.16%
		No, but I would have liked a choice	2,638	4.22%
		No, but I did not mind	15,940	25.53%
		Don't know / can't remember	1,084	1.74%
		Inapplicable	33,211	53.19%
		Missing responses	2,603	4.17%
		Total	62,443	100.00%
Q6	How do you feel about the length of time you were on the waiting list before your			

	Question	Response categories	Frequency	Percent
	admission to hospital?			
		I was admitted as soon as I thought was necessary	19,957	31.96%
		I should have been admitted a bit sooner	3,858	6.18%
		I should have been admitted a lot sooner	2,036	3.26%
		Inapplicable	33,211	53.19%
		Missing responses	3,381	5.41%
		Total	62,443	100.00%
Q7	Was your admission date changed by the hospital?			
		No	21,269	34.06%
		Yes, once	4,148	6.64%
		Yes, 2 or 3 times	792	1.27%
		Yes, 4 times or more	81	0.13%
		Inapplicable	33,211	53.19%
		Missing responses	2,942	4.71%
		Total	62,443	100.00%
Q8	In your opinion, had the specialist you saw in hospital been given all of the necessary information about your condition or illness from the person who referred you?			
		Yes, definitely	21,485	34.41%
		Yes, to some extent	3,677	5.89%
		No	742	1.19%
		Don't know / can't remember	710	1.14%
		Inapplicable	33,211	53.19%
		Missing responses	2,618	4.19%
		Total	62,443	100.00%
Q9	From the time you arrived at the hospital, did you feel that you had to wait a long time to get to a bed on a ward?			
		Yes, definitely	7,798	12.49%
		Yes, to some extent	12,422	19.89%
		No	40,785	65.32%
		Missing responses	1,438	2.30%
		Total	62,443	100.00%
Q10	While in hospital, did you ever stay in a critical care area (Intensive Care Unit, High Dependency Unit or Coronary Care Unit)?			
		Yes	12,835	20.55%
		No	45,234	72.44%
		Don't know / can't remember	3,111	4.98%
		Missing responses	1,263	2.02%
		Total	62,443	100.00%
Q11	When you were first admitted to a bed on a ward, did you share a sleeping area, for example a room or bay, with patients of the opposite sex?			
		Yes	5,803	9.29%
		No	54,785	87.74%
		Missing responses	1,855	2.97%
		Total	62,443	100.00%
Q12	During your stay in hospital, how many wards			

	Question	Response categories	Frequency	Percent
	did you stay in?			
		1	37,899	60.69%
		2	17,839	28.57%
		3 or more	4,850	7.77%
		Don't know / can't remember	823	1.32%
		Missing responses	1,032	1.65%
		Total	62,443	100.00%
Q13	After you moved to another ward (or wards), did you ever share a sleeping area, for example a room or bay, with patients of the opposite sex?			
		Yes	1,612	2.58%
		No	21,015	33.65%
		Inapplicable	37,825	60.58%
		Missing responses	1,991	3.19%
		Total	62,443	100.00%
Q14	While staying in hospital, did you ever use the same bathroom or shower area as patients of the opposite sex?			
		Yes	6,935	11.11%
		Yes, because it had special bathing equipment that I needed	624	1.00%
		No	46,334	74.20%
		I did not use a bathroom or shower	3,352	5.37%
		Don't know / can't remember	2,970	4.76%
		Missing responses	2,228	3.57%
		Total	62,443	100.00%
Q15	Were you ever bothered by noise at night from other patients?			
		Yes	23,511	37.65%
		No	37,516	60.08%
		Missing responses	1,416	2.27%
		Total	62,443	100.00%
Q16	Were you ever bothered by noise at night from hospital staff?			
		Yes	12,024	19.26%
		No	49,173	78.75%
		Missing responses	1,246	2.00%
		Total	62,443	100.00%
Q17	In your opinion, how clean was the hospital room or ward that you were in?			
		Very clean	43,064	68.97%
		Fairly clean	16,868	27.01%
		Not very clean	1,462	2.34%
		Not at all clean	295	0.47%
		Missing responses	754	1.21%
		Total	62,443	100.00%
Q18	How clean were the toilets and bathrooms that you used in hospital?			
		Very clean	37,564	60.16%
		Fairly clean	18,833	30.16%
		Not very clean	2,582	4.13%

	Question	Response categories	Frequency	Percent
		Not at all clean	608	0.97%
		I did not use a toilet or bathroom	2,058	3.30%
		Missing responses	798	1.28%
		Total	62,443	100.00%
Q19	Did you feel threatened during your stay in hospital by other patients or visitors?			
		Yes	2,060	3.30%
		No	59,574	95.41%
		Missing responses	809	1.30%
		Total	62,443	100.00%
Q20	Were hand-wash gels available for patients and visitors to use?			
		Yes	57,123	91.48%
		Yes, but they were empty	927	1.48%
		I did not see any hand-wash gels	1,488	2.38%
		Don't know / can't remember	2,220	3.56%
		Missing responses	685	1.10%
		Total	62,443	100.00%
Q21	How would you rate the hospital food?			
		Very good	12,776	20.46%
		Good	21,217	33.98%
		Fair	16,938	27.13%
		Poor	8,115	13.00%
		I did not have any hospital food	2,444	3.91%
		Missing responses	953	1.53%
		Total	62,443	100.00%
Q22	Were you offered a choice of food?			
		Yes, always	48,384	77.49%
		Yes, sometimes	8,911	14.27%
		No	3,363	5.39%
		Missing responses	1,785	2.86%
		Total	62,443	100.00%
Q23	Did you get enough help from staff to eat your meals?			
		Yes, always	10,637	17.03%
		Yes, sometimes	3,102	4.97%
		No	2,817	4.51%
		I did not need help to eat meals	43,980	70.43%
		Missing responses	1,907	3.05%
		Total	62,443	100.00%
Q24	When you had important questions to ask a doctor, did you get answers that you could understand?			
		Yes, always	38,325	61.38%
		Yes, sometimes	14,071	22.53%
		No	2,809	4.50%
		I had no need to ask	6,193	9.92%
		Missing responses	1,045	1.67%
		Total	62,443	100.00%
Q25	Did you have confidence and trust in the doctors treating you?			
		Yes, always	49,756	79.68%

	Question	Response categories	Frequency	Percent
		Yes, sometimes	9,756	15.62%
		No	1,967	3.15%
		Missing responses	964	1.54%
		Total	62,443	100.00%
Q26	Did doctors talk in front of you as if you weren't there?			
		Yes, often	3,182	5.10%
		Yes, sometimes	11,393	18.25%
		No	46,693	74.78%
		Missing responses	1,175	1.88%
		Total	62,443	100.00%
Q27	When you had important questions to ask a nurse, did you get answers that you could understand?			
		Yes, always	38,017	60.88%
		Yes, sometimes	14,357	22.99%
		No	2,195	3.52%
		I had no need to ask	7,040	11.27%
		Missing responses	834	1.34%
		Total	62,443	100.00%
Q28	Did you have confidence and trust in the nurses treating you?			
		Yes, always	47,556	76.16%
		Yes, sometimes	12,313	19.72%
		No	1,799	2.88%
		Missing responses	775	1.24%
		Total	62,443	100.00%
Q29	Did nurses talk in front of you as if you weren't there?			
		Yes, often	2,441	3.91%
		Yes, sometimes	8,822	14.13%
		No	49,836	79.81%
		Missing responses	1,344	2.15%
		Total	62,443	100.00%
Q30	In your opinion, were there enough nurses on duty to care for you in hospital?			
		There were always or nearly always enough nurses	36,418	58.32%
		There were sometimes enough nurses	18,284	29.28%
		There were rarely or never enough nurse	6,795	10.88%
		Missing responses	946	1.51%
		Total	62,443	100.00%
Q31	Sometimes in a hospital, a member of staff will say one thing and another will say something quite different. Did this happen to you?			
		Yes, often	4,225	6.77%
		Yes, sometimes	14,608	23.39%
		No	42,574	68.18%
		Missing responses	1,036	1.66%
		Total	62,443	100.00%
Q32	Were you involved as much as you wanted to			

	Question	Response categories	Frequency	Percent
	be in decisions about your care and treatment?			
		Yes, definitely	34,596	55.40%
		Yes, to some extent	20,520	32.86%
		No	6,051	9.69%
		Missing responses	1,276	2.04%
		Total	62,443	100.00%
Q33	How much information about your condition or treatment was given to you?			
		Not enough	11,926	19.10%
		The right amount	49,022	78.51%
		Too much	428	0.69%
		Missing responses	1,067	1.71%
		Total	62,443	100.00%
Q34	Did you find someone on the hospital staff to talk to about your worries and fears?			
		Yes, definitely	14,828	23.75%
		Yes, to some extent	13,291	21.29%
		No	8,402	13.46%
		I had no worries or fears	24,664	39.50%
		Missing responses	1,258	2.01%
		Total	62,443	100.00%
Q35	Do you feel you got enough emotional support from hospital staff during your stay?			
		Yes, always	22,532	36.08%
		Yes, sometimes	11,239	18.00%
		No	5,454	8.73%
		I did not need any emotional support	22,055	35.32%
		Missing responses	1,163	1.86%
		Total	62,443	100.00%
Q36	Were you given enough privacy when discussing your condition or treatment?			
		Yes, always	46,106	73.84%
		Yes, sometimes	11,056	17.71%
		No	3,884	6.22%
		Missing responses	1,397	2.24%
		Total	62,443	100.00%
Q37	Were you given enough privacy when being examined or treated?			
		Yes, always	55,561	88.98%
		Yes, sometimes	5,057	8.10%
		No	811	1.30%
		Missing responses	1,014	1.62%
		Total	62,443	100.00%
Q38	Were you ever in any pain?			
		Yes	39,058	62.55%
		No	21,740	34.82%
		Missing responses	1,645	2.63%
		Total	62,443	100.00%
Q39	Do you think the hospital staff did everything they could to help control your pain?			
		Yes, definitely	28,043	44.91%
		Yes, to some extent	8,996	14.41%

	Question	Response categories	Frequency	Percent
		No	2,393	3.83%
		Inapplicable	21,740	34.82%
		Missing responses	1,271	2.04%
		Total	62,443	100.00%
Q40	How many minutes after you used the call button did it usually take before you got the help you needed?			
		0 minutes / right away	5,272	8.44%
		1-2 minutes	13,994	22.41%
		3-5 minutes	10,767	17.24%
		More than 5 minutes	6,184	9.90%
		I never got help when I used the call button	473	0.76%
		I never used the call button	23,537	37.69%
		Missing responses	2,216	3.55%
		Total	62,443	100.00%
Q41	During your stay in hospital, did you have an operation or procedure?			
		Yes	37,762	60.47%
		No	22,911	36.69%
		Missing responses	1,770	2.83%
		Total	62,443	100.00%
Q42	Beforehand, did a member of staff explain the risks and benefits of the operation or procedure in a way you could understand?			
		Yes, completely	30,596	49.00%
		Yes, to some extent	5,212	8.35%
		No	1,228	1.97%
		I did not want an explanation	796	1.27%
		Inapplicable	22,911	36.69%
		Missing responses	1,700	2.72%
		Total	62,443	100.00%
Q43	Beforehand, did a member of staff explain what would be done during the operation or procedure?			
		Yes, completely	28,196	45.15%
		Yes, to some extent	7,079	11.34%
		No	1,660	2.66%
		I did not want an explanation	995	1.59%
		Inapplicable	22,911	36.69%
		Missing responses	1,602	2.57%
		Total	62,443	100.00%
Q44	Beforehand, did a member of staff answer your questions about the operation or procedure?			
		Yes, completely	25,456	40.77%
		Yes, to some extent	5,932	9.50%
		No	1,071	1.72%
		I did not have any questions	5,306	8.50%
		Inapplicable	22,911	36.69%
		Missing responses	1,767	2.83%
		Total	62,443	100.00%
Q45	Beforehand, were you told how you could expect to feel after you had the operation or			

	Question	Response categories	Frequency	Percent
	procedure?			
		Yes, completely	21,747	34.83%
		Yes, to some extent	10,231	16.38%
		No	5,421	8.68%
		Inapplicable	22,911	36.69%
		Missing responses	2,133	3.42%
		Total	62,443	100.00%
Q46	Before the operation or procedure, were you given an anaesthetic or medication to put you to sleep or control your pain?			
		Yes	32,126	51.45%
		No	5,315	8.51%
		Inapplicable	22,911	36.69%
		Missing responses	2,091	3.35%
		Total	62,443	100.00%
Q47	Before the operation or procedure, did the anaesthetist or another member of staff explain how he or she would put you to sleep or control your pain in a way you could understand?			
		Yes, completely	27,308	43.73%
		Yes, to some extent	3,568	5.71%
		No	1,276	2.04%
		Inapplicable	28,078	44.97%
		Missing responses	2,213	3.54%
		Total	62,443	100.00%
Q48	After the operation or procedure, did a member of staff explain how the operation or procedure had gone in a way you could understand?			
		Yes, completely	25,726	41.20%
		Yes, to some extent	7,946	12.73%
		No	3,758	6.02%
		Inapplicable	22,911	36.69%
		Missing responses	2102	3.37%
		Total	62,443	100.00%
Q49	Did you feel you were involved in decisions about your discharge from hospital?			
		Yes, definitely	32,131	51.46%
		Yes, to some extent	17,732	28.40%
		No	9,225	14.77%
		I did not want to be involved	2,038	3.26%
		Missing responses	1,317	2.11%
		Total	62,443	100.00%
Q50	Were you given enough notice about when you were going to be discharged?			
		Yes, definitely	34,207	54.78%
		Yes, to some extent	19,265	30.85%
		No	7,706	12.34%
		Missing responses	1,265	2.03%
		Total	62,443	100.00%
Q51	On the day you left hospital, was your discharge delayed for any reason?			
		Yes	24,516	39.26%

	Question	Response categories	Frequency	Percent
		No	36,228	58.02%
		Missing responses	1,699	2.72%
		Total	62,443	100.00%
Q52	What was the MAIN reason for the delay? (Cross ONE box only)			
		I had to wait for medicines	14,560	23.32%
		I had to wait to see the doctor	3,359	5.38%
		I had to wait for an ambulance	2,339	3.75%
		Something else	3,378	5.41%
		Inapplicable	36,087	57.79%
		Missing responses	2,720	4.36%
		Total	62,443	100.00%
Q53	How long was the delay?			
		Up to 1 hour	3,711	5.94%
		Longer than 1 hour but no longer than 2	6,932	11.10%
		Longer than 2 hours but no longer than	8,168	13.08%
		Longer than 4 hours	5,901	9.45%
		Inapplicable	36,087	57.79%
		Missing responses	1,644	2.63%
		Total	62,443	100.00%
Q54	Before you left hospital, were you given any written or printed information about what you should or should not do after leaving hospital?			
		Yes	41,827	66.98%
		No	18,507	29.64%
		Missing responses	2,109	3.38%
		Total	62,443	100.00%
Q55	Did a member of staff explain the purpose of the medicines you were to take at home in a way you could understand?			
		Yes, completely	34,839	55.79%
		Yes, to some extent	7,431	11.90%
		No	3,718	5.95%
		I did not need an explanation	6,941	11.12%
		I had no medicines	7,446	11.92%
		Missing responses	2,068	3.31%
		Total	62,443	100.00%
Q56	Did a member of staff tell you about medication side effects to watch for when you went home?			
		Yes, completely	15,753	25.23%
		Yes, to some extent	7,608	12.18%
		No	16,273	26.06%
		I did not need an explanation	13,153	21.06%
		Inapplicable	7,446	11.92%
		Missing responses	2,210	3.54%
		Total	62,443	100.00%
Q57	Were you told how to take your medication in a way you could understand?			
		Yes, definitely	31,133	49.86%
		Yes, to some extent	5,990	9.59%

	Question	Response categories	Frequency	Percent
		No	3,710	5.94%
		I did not need to be told how to take my medicine	12,132	19.43%
		Inapplicable	7,446	11.92%
		Missing responses	2,032	3.25%
		Total	62,443	100.00%
Q58	Were you given clear written or printed information about your medicines?			
		Yes, completely	30,654	49.09%
		Yes, to some extent	6,669	10.68%
		No	4,999	8.01%
		I did not need this	9,365	15.00%
		Don't know / can't remember	1,289	2.06%
		Inapplicable	7,446	11.92%
		Missing responses	2,021	3.24%
		Total	62,443	100.00%
Q59	Did a member of staff tell you about any danger signals you should watch for after you went home?			
		Yes, completely	19,785	31.68%
		Yes, to some extent	9,581	15.34%
		No	15,910	25.48%
		It was not necessary	14,983	23.99%
		Missing responses	2,184	3.50%
		Total	62,443	100.00%
Q60	Did hospital staff take your family or home situation into account when planning your discharge?			
		Yes, completely	25,880	41.45%
		Yes, to some extent	8,776	14.05%
		No	7,263	11.63%
		It was not necessary	16,984	27.20%
		Don't know / can't remember	1,689	2.70%
		Missing responses	1,851	2.96%
		Total	62,443	100.00%
Q61	Did the doctors or nurses give your family or someone close to you all the information they needed to help care for you?			
		Yes, definitely	20,933	33.52%
		Yes, to some extent	9,332	14.94%
		No	11,466	18.36%
		No family or friends were involved	7,411	11.87%
		My family or friends did not want or need to	10,946	17.53%
		Missing responses	2,355	3.77%
		Total	62,443	100.00%
Q62	Did hospital staff tell you who to contact if you were worried about your condition or treatment after you left hospital?			
		Yes	43,168	69.13%
		No	12,322	19.73%
		Don't know / can't remember	4,857	7.78%
		Missing responses	2,096	3.36%

	Question	Response categories	Frequency	Percent
		Total	62,443	100.00%
Q63	Did hospital staff discuss with you whether you would need any additional equipment in your home, or any adaptations made to your home, after leaving hospital?			
		Yes	14,962	23.96%
		No, but I would have liked to	3,222	5.16%
		No, it was not necessary to discuss it	42,217	67.61%
		Missing responses	2,042	3.27%
		Total	62,443	100.00%
Q64	Did hospital staff discuss with you whether you may need any further health or social care services after leaving hospital?			
		Yes	27,407	43.89%
		No, but I would have liked them to	4,785	7.66%
		No, it was not necessary to discuss it	28,113	45.02%
		Missing responses	2,138	3.42%
		Total	62,443	100.00%
Q65	Did you receive copies of letters sent between hospital doctors and your family doctor (GP)?			
		Yes, I received copies	35,556	56.94%
		No, I did not receive copies	19,480	31.20%
		Not sure / don't know	5,337	8.55%
		Missing responses	2,070	3.32%
		Total	62,443	100.00%
Q66	Were the letters written in a way that you could understand?			
		Yes, definitely	26,822	42.95%
		Yes, to some extent	7,529	12.06%
		No	792	1.27%
		Not sure / don't know	269	0.43%
		Inapplicable	24,817	39.74%
		Missing responses	403	0.65%
		Total	60,632	97.10%
Q67	Overall, did you feel you were treated with respect and dignity while you were in the hospital?			
		Yes, always	27,407	43.89%
		Yes, sometimes	4,785	7.66%
		No	28,113	45.02%
		Missing responses	2,138	3.42%
		Total	62,443	100.00%
Q68	Overall experience			
		I had a very poor experience	487	0.78%
		1	549	0.88%
		2	710	1.14%
		3	1,066	1.71%
		4	1,253	2.01%
		5	2,878	4.61%
		6	3,022	4.84%

	Question	Response categories	Frequency	Percent
		7	6,532	10.46%
		8	13,860	22.20%
		9	12,473	19.98%
		I had a very good experience	16,233	26.00%
		Missing responses	3,380	5.41%
		Total	62,443	100.00%
Q69	During your hospital stay, were you ever asked to give your views on the quality of your care?			
		Yes	11,078	17.74%
		No	42,988	68.84%
		Don't know / can't remember	6,814	10.91%
		Missing responses	1,563	2.50%
		Total	62,443	100.00%
Q70	Did you see, or were you given, any information explaining how to complain to the hospital about the care you received?			
		Yes	12,011	19.24%
		No	36,649	58.69%
		Not sure / don't know	11,773	18.85%
		Missing responses	2,010	3.22%
		Total	62,443	100.00%
Q71	Who was the main person or people that filled in this questionnaire?			
		The patient	51,160	81.93%
		A friend or relative of the patient	3,530	5.65%
		Both patient and friend/relative together	5,378	8.61%
		The patient with the help of a health professional	276	0.44%
		Missing responses	2,099	3.36%
		Total	62,443	100.00%
	Age group			
		16-35	4,358	6.98%
		36-50	7,507	12.02%
		51-65	15,202	24.35%
		66+	35,376	56.65%
		Total	62,443	100.00%
	Gender			
		Male	28,936	46.34%
		Female	33,507	53.66%
		Total	62,443	100.00%

Appendix 5: Inapplicable questions based on patient journey

Inpatient Experience		
1. Admissions	A. A&E a) Information provision b) Privacy	B. Waiting List or Planned Admission a) Choice b) Waiting to be admitted c) Transition between services
2. Hospital Stay	Hospital and Ward a) Waiting to get to the ward b) Single sex accommodation: when first admitted c) Single sex accommodation: after moving wards d) Single sex accommodation: bathroom areas e) Noise at night f) Cleanliness g) Security h) Food	
	Doctors and Nurses a) Communication b) Confidence and trust c) Availability of staff	
	Patient Care and Treatment a) Involvement in decisions b) Privacy c) Pain Management	
3. Operations and Procedures	A. Operations and Procedures a) Before the Operation or Procedure b) After the Operation or Procedure	B. No Operation or Procedure
4. Leaving the Hospital	a) Preparing to leave hospital b) Delays to discharge c) Medication d) Information provision e) Transition from hospital	
5. Overall impression	Patient Rating Scale	

Appendix 6: Spearman's rank correlation coefficient (ρ) and significance level (rho)

Variable	Emergency With Operation		Emergency without Operation		Planned with Operation		Planned without Operation	
	ρ	rho	ρ	rho	ρ	rho	ρ	rho
q3	-0.11	0.00	-0.10	0.00				
q4	0.35	0.00	0.33	0.00				
q5					0.15	0.00	0.16	0.00
q6					0.22	0.00	0.19	0.00
q7					0.08	0.00	0.06	0.02
q8					0.28	0.00	0.36	0.00
q9	0.35	0.00	0.38	0.00	0.23	0.00	0.29	0.00
q11	0.01	0.50	0.06	0.00	0.05	0.00	0.08	0.00
q14	0.10	0.00	0.13	0.00	0.09	0.00	0.14	0.00
q15	0.27	0.00	0.25	0.00	0.27	0.00	0.26	0.00
q16	0.30	0.00	0.29	0.00	0.28	0.00	0.27	0.00
q17	0.48	0.00	0.48	0.00	0.43	0.00	0.44	0.00
q18	0.44	0.00	0.43	0.00	0.41	0.00	0.43	0.00
q19	0.16	0.00	0.15	0.00	0.08	0.00	0.16	0.00
q20	0.16	0.00	0.17	0.00	0.15	0.00	0.15	0.00
q21	0.45	0.00	0.43	0.00	0.41	0.00	0.43	0.00
q22	0.27	0.00	0.27	0.00	0.24	0.00	0.24	0.00
q23	0.15	0.00	0.14	0.00	0.11	0.00	0.10	0.00
q24	0.40	0.00	0.39	0.00	0.31	0.00	0.35	0.00
q25	0.48	0.00	0.52	0.00	0.36	0.00	0.47	0.00
q26	0.31	0.00	0.31	0.00	0.25	0.00	0.29	0.00
q27	0.44	0.00	0.40	0.00	0.37	0.00	0.37	0.00
q28	0.54	0.00	0.54	0.00	0.49	0.00	0.48	0.00
q29	0.35	0.00	0.33	0.00	0.31	0.00	0.29	0.00
q30	0.46	0.00	0.45	0.00	0.43	0.00	0.44	0.00
q31	0.44	0.00	0.44	0.00	0.39	0.00	0.43	0.00
q32	0.49	0.00	0.50	0.00	0.47	0.00	0.51	0.00
q33	0.44	0.00	0.47	0.00	0.34	0.00	0.41	0.00
q34	0.24	0.00	0.27	0.00	0.20	0.00	0.20	0.00
q35	0.37	0.00	0.36	0.00	0.28	0.00	0.26	0.00
q36	0.41	0.00	0.43	0.00	0.38	0.00	0.38	0.00
q37	0.35	0.00	0.35	0.00	0.28	0.00	0.28	0.00
q39	0.20	0.00	0.14	0.00	0.11	0.00	0.03	0.23
q40	0.09	0.00	0.03	0.00	0.06	0.00	-0.01	0.83
q42	0.33	0.00			0.31	0.00		
q43	0.32	0.00			0.28	0.00		
q44	0.25	0.00			0.24	0.00		
q45	0.38	0.00			0.37	0.00		

Variable	Emergency With Operation		Emergency without Operation		Planned with Operation		Planned without Operation	
q47	0.24	0.00			0.23	0.00		
q48	0.40	0.00			0.37	0.00		
q49	0.45	0.00	0.45	0.00	0.43	0.00	0.40	0.00
q50	0.44	0.00	0.45	0.00	0.40	0.00	0.41	0.00
q51	0.26	0.00	0.24	0.00	0.25	0.00	0.20	0.00
q54	0.27	0.00	0.24	0.00	0.20	0.00	0.17	0.00
q55	0.31	0.00	0.30	0.00	0.22	0.00	0.17	0.00
q56	0.36	0.00	0.32	0.00	0.30	0.00	0.24	0.00
q57	0.27	0.00	0.24	0.00	0.19	0.00	0.12	0.00
q58	0.27	0.00	0.23	0.00	0.20	0.00	0.14	0.00
q59	0.37	0.00	0.33	0.00	0.33	0.00	0.24	0.00
q60	0.28	0.00	0.24	0.00	0.27	0.00	0.18	0.00
q61	0.35	0.00	0.31	0.00	0.32	0.00	0.25	0.00
q62	0.34	0.00	0.34	0.00	0.28	0.00	0.31	0.00
q63	0.08	0.00	0.10	0.00	0.12	0.00	0.11	0.00
q64	0.16	0.00	0.15	0.00	0.15	0.00	0.12	0.00
q66	0.22	0.00	0.24	0.00	0.22	0.00	0.18	0.00
q67	0.61	0.00	0.62	0.00	0.51	0.00	0.54	0.00
q69	0.22	0.00	0.18	0.00	0.18	0.00	0.19	0.00
q70	0.26	0.00	0.22	0.00	0.25	0.00	0.25	0.00
LOS	0.05	0.00	0.02	0.04	0.01	0.43	-0.02	0.51
all_gender	0.12	0.00	0.10	0.00	0.05	0.00	0.08	0.00
age_group	-0.13	0.00	-0.13	0.00	-0.10	0.00	-0.09	0.00
critical	0.12	0.00	0.08	0.00	0.02	0.01	-0.01	0.62
pain	-0.15	0.00	-0.15	0.00	-0.14	0.00	-0.17	0.00
selfreported	0.10	0.00	0.13	0.00	0.06	0.00	0.10	0.00

Appendix 7: Regression model for patients with an operation or procedure and emergency admission

Ordered logistic regression				Number of obs	=	6,578
				Wald chi2(60)	=	6409.80
				Prob > chi2	=	0.00
Log pseudolikelihood = -9070.1385				Pseudo R2	=	0.27

		Odds Ratio	Std. Err.	z	P>z	[95% Conf.	Interval]
Q67	Overall, did you feel you were treated with respect and dignity while you were in the hospital?	**4.09**	0.33	17.32	0.00	3.49	4.80
Q28	Did you have confidence and trust in the nurses treating you?	**1.78**	0.11	9.05	0.00	1.57	2.02
Q17	In your opinion, how clean was the hospital room or ward that you were in?	**1.56**	0.09	7.75	0.00	1.39	1.74
Q19	Did you feel threatened during your stay in hospital by other patients or visitors?	**1.39**	0.16	2.82	0.01	1.10	1.74
Q33	How much information about your condition or treatment was given to you?	**1.37**	0.12	3.72	0.00	1.16	1.62
Q32	Were you involved as much as you wanted to be in decisions about your care and treatment?	**1.36**	0.07	6.38	0.00	1.24	1.49
Q30	In your opinion, were there enough nurses on duty to care for you in hospital?	**1.35**	0.06	6.14	0.00	1.22	1.48
Q9	From the time you arrived at the hospital, did you feel that you had to wait a long time to get to a bed on a ward?	**1.33**	0.05	7.85	0.00	1.24	1.43
Q25	Did you have confidence and trust in the doctors treating you?	**1.33**	0.09	4.11	0.00	1.16	1.52
Q21	How would you rate the hospital food?	**1.32**	0.03	13.54	0.00	1.27	1.37
Q31	Sometimes in a hospital, a member of staff will say one thing and another will say something quite different. Did this happen to you?	**1.24**	0.06	4.49	0.00	1.13	1.36
Q39	Do you think the hospital staff did everything they could to help control your pain?	**1.22**	0.06	4.07	0.00	1.11	1.34
Q50	Were you given enough notice about when you were going to be discharged?	**1.21**	0.05	4.21	0.00	1.11	1.32
Q15	Were you ever bothered by noise	**1.17**	0.06	3.00	0.00	1.06	1.30

		Odds Ratio	Std. Err.	z	P>z	[95% Conf.	Interval]
	at night from other patients?						
Q4	Were you given enough privacy when being examined or treated in the A&E Department?	**1.17**	0.05	3.48	0.00	1.07	1.27
Q35	Do you feel you got enough emotional support from hospital staff during your stay?	**1.17**	0.03	5.34	0.00	1.10	1.24
Q45	Beforehand, were you told how you could expect to feel after you had the operation or procedure?	**1.16**	0.05	3.51	0.00	1.07	1.27
Q48	After the operation or procedure, did a member of staff explain how the operation or procedure had gone in a way you could understand?	**1.15**	0.05	3.06	0.00	1.05	1.26
Q70	Did you see, or were you given, any information explaining how to complain to the hospital about the care you received?	**1.14**	0.04	3.84	0.00	1.07	1.22
Q49	Did you feel you were involved in decisions about your discharge from hospital?	**1.13**	0.04	3.37	0.00	1.05	1.22
Q18	How clean were the toilets and bathrooms that you used in hospital?	**1.13**	0.04	3.45	0.00	1.05	1.21
Q24	When you had important questions to ask a doctor, did you get answers that you could understand?	**1.10**	0.05	2.29	0.02	1.01	1.20
Q69	During your hospital stay, were you ever asked to give your views on the quality of your care?	**1.10**	0.03	3.25	0.00	1.04	1.17
Q62	Did hospital staff tell you who to contact if you were worried about your condition or treatment after you left hospital?	**1.09**	0.04	2.74	0.01	1.03	1.16
Q27	When you had important questions to ask a nurse, did you get answers that you could understand?	**1.09**	0.04	2.24	0.03	1.01	1.17
Q40	How many minutes after you used the call button did it usually take before you got the help you needed?	**1.07**	0.03	2.38	0.02	1.01	1.14
Q51	How long was the delay in discharge?	**1.07**	0.02	3.91	0.00	1.03	1.11
Q59	Did a member of staff tell you about any danger signals you should watch for after you went home?	**1.06**	0.03	2.10	0.04	1.00	1.11
Q66	Were the letters sent between hospital doctors and your family	**1.04**	0.02	2.48	0.01	1.01	1.08

		Odds Ratio	Std. Err.	z	P>z [95%	Conf.	Interval]
	doctor (GP) written in a way that you could understand?						
Q54	Before you left hospital, were you given any written or printed information about what you should or should not do after leaving hospital?	1.10	0.06	1.87	0.06	1.00	1.22
Q61	Did the doctors or nurses give your family or someone close to you all the information they needed to help care for you?	1.05	0.03	1.88	0.06	1.00	1.10
Q36	Were you given enough privacy when discussing your condition or treatment?	0.90	0.05	-1.86	0.06	0.80	1.01
Q44	Beforehand, did a member of staff answer your questions about the operation or procedure?	0.94	0.03	-1.78	0.08	0.88	1.01
Q16	Were you ever bothered by noise at night from hospital staff?	1.12	0.07	1.76	0.08	0.99	1.27
Q20	Were hand-wash gels available for patients and visitors to use?	0.91	0.05	-1.74	0.08	0.81	1.01
Q47	Before the operation or procedure, did the anaesthetist or another member of staff explain how he or she would put you to sleep or control your pain in a way you could understand?	0.96	0.04	-1.03	0.30	0.90	1.03
Q43	Beforehand, did a member of staff explain what would be done during the operation or procedure?	0.96	0.04	-1.02	0.31	0.88	1.04
Q3	While you were in the A&E Department, how much information about your condition?	1.03	0.03	1.00	0.32	0.97	1.09
Q22	Were you offered a choice of food?	1.05	0.05	0.95	0.34	0.95	1.16
Q63	Did hospital staff discuss with you whether you would need any additional equipment in your home, or any adaptations made to your home, after leaving hospital?	0.96	0.05	-0.76	0.45	0.87	1.07
Q11	When you were first admitted to a bed on a ward, did you share a sleeping area, for example a room or bay, with patients of the opposite sex?	0.95	0.08	-0.60	0.55	0.82	1.11
Q58	Were you given clear written or printed information about your medicines?	0.99	0.02	-0.60	0.55	0.94	1.03
Q42	Beforehand, did a member of staff explain the risks and benefits of the operation or procedure in a way you could understand?	0.97	0.05	-0.55	0.58	0.88	1.08

		Odds Ratio	Std. Err.	z	P>z	[95% Conf.	Interval]
Q55	Did a member of staff explain the purpose of the medicines you were to take at home in a way you could understand?	1.02	0.04	0.48	0.63	0.95	1.09
Q29	Did nurses talk in front of you as if you weren't there?	0.97	0.06	-0.43	0.67	0.86	1.10
Q23	Did you get enough help from staff to eat your meals?	0.99	0.03	-0.41	0.68	0.92	1.05
Q57	Were you told how to take your medication in a way you could understand?	1.01	0.03	0.38	0.70	0.95	1.08
Q34	Did you find someone on the hospital staff to talk to about your worries and fears?	0.99	0.03	-0.37	0.71	0.93	1.05
Q60	Did hospital staff take your family or home situation into account when planning your discharge?	0.99	0.02	-0.36	0.72	0.95	1.03
Q37	Were you given enough privacy when being examined or treated?	1.03	0.08	0.33	0.74	0.88	1.19
Q56	Did a member of staff tell you about medication side effects to watch for when you went home?	1.01	0.02	0.25	0.80	0.96	1.06
Q26	Did doctors talk in front of you as if you weren't there?	1.01	0.05	0.12	0.91	0.91	1.12
Q64	Did hospital staff discuss with you whether you may need any further health or social care services after leaving hospital?	1.00	0.05	0.08	0.94	0.92	1.10
Q14	While staying in hospital, did you ever use the same bathroom or shower area as patients of the opposite sex?	1.00	0.02	-0.04	0.97	0.97	1.03
	age_group	1.00	0.00	2.00	0.05	1.00	1.01
	all_gender	0.98	0.05	-0.37	0.72	0.89	1.08
	critical	0.95	0.02	-2.26	0.02	0.91	0.99
	LOS	1.29	0.08	4.33	0.00	1.15	1.45
	pain	0.69	0.07	-3.49	0.00	0.56	0.85
	selfreported	1.13	0.09	1.56	0.12	0.97	1.31

Appendix 8: Regression model for patients with an operation or procedure and planned admission

Ordered logistic regression				Number of obs	=	14,304
				Wald chi2(62)	=	13396.92
				Prob > chi2	=	0.00
Log pseudolikelihood = -17980.296				Pseudo R2	=	0.25

		Odds Ratio	Std. Err.	z	P>z	[95% Conf.	Interval]
Q67	Overall, did you feel you were treated with respect and dignity while you were in the hospital?	**4.20**	0.29	20.48	0.00	3.66	4.81
Q28	Did you have confidence and trust in the nurses treating you?	**1.68**	0.09	9.97	0.00	1.51	1.85
Q17	In your opinion, how clean was the hospital room or ward that you were in?	**1.65**	0.07	12.45	0.00	1.53	1.79
Q32	Were you involved as much as you wanted to be in decisions about your care and treatment?	**1.53**	0.05	11.93	0.00	1.43	1.64
Q25	Did you have confidence and trust in the doctors treating you?	**1.51**	0.10	6.33	0.00	1.33	1.72
Q30	In your opinion, were there enough nurses on duty to care for you in hospital?	**1.38**	0.04	11.02	0.00	1.30	1.46
Q9	From the time you arrived at the hospital, did you feel that you had to wait a long time to get to a bed on a ward?	**1.32**	0.04	9.82	0.00	1.25	1.40
Q21	How would you rate the hospital food?	**1.31**	0.02	20.76	0.00	1.27	1.34
Q15	Were you ever bothered by noise at night from other patients?	**1.29**	0.05	6.92	0.00	1.20	1.39
Q50	Were you given enough notice about when you were going to be discharged?	**1.28**	0.04	8.66	0.00	1.21	1.35
Q31	Sometimes in a hospital, a member of staff will say one thing and another will say something quite different. Did this happen to you?	**1.28**	0.05	6.39	0.00	1.18	1.37
Q39	Do you think the hospital staff did everything they could to help control your pain?	**1.24**	0.04	6.33	0.00	1.16	1.33
Q45	Beforehand, were you told how you could expect to feel after you had the operation or procedure?	**1.21**	0.03	6.89	0.00	1.15	1.28
Q16	Were you ever bothered by noise at night from hospital staff?	**1.21**	0.05	4.70	0.00	1.12	1.31
Q48	After the operation or procedure,	**1.16**	0.03	5.30	0.00	1.10	1.23

		Odds Ratio	Std. Err.	z	P>z [95%	Conf.	Interval]
	did a member of staff explain how the operation or procedure had gone in a way you could understand?						
Q6	How do you feel about the length of time you were on the waiting list before your admission to hospital?	**1.14**	0.04	4.03	0.00	1.07	1.22
Q35	Do you feel you got enough emotional support from hospital staff during your stay?	**1.13**	0.02	7.06	0.00	1.09	1.17
Q18	How clean were the toilets and bathrooms that you used in hospital?	**1.11**	0.03	3.92	0.00	1.05	1.17
Q29	Did nurses talk in front of you as if you weren't there?	**1.11**	0.05	2.15	0.03	1.01	1.21
Q36	Were you given enough privacy when discussing your condition or treatment?	**1.11**	0.04	2.60	0.01	1.03	1.19
Q49	Did you feel you were involved in decisions about your discharge from hospital?	**1.10**	0.03	3.95	0.00	1.05	1.16
Q51	How long was the delay in discharge?	**1.10**	0.01	7.80	0.00	1.07	1.12
Q62	Did hospital staff tell you who to contact if you were worried about your condition or treatment after you left hospital?	**1.09**	0.03	2.93	0.00	1.03	1.16
Q69	During your hospital stay, were you ever asked to give your views on the quality of your care?	**1.09**	0.02	4.07	0.00	1.05	1.14
Q40	How many minutes after you used the call button did it usually take before you got the help you needed?	**1.08**	0.02	3.82	0.00	1.04	1.13
Q27	When you had important questions to ask a nurse, did you get answers that you could understand?	**1.08**	0.03	3.26	0.00	1.03	1.13
Q70	Did you see, or were you given, any information explaining how to complain to the hospital about the care you received?	**1.07**	0.02	3.21	0.00	1.03	1.12
Q61	Did the doctors or nurses give your family or someone close to you all the information they needed to help care for you?	**1.07**	0.02	4.00	0.00	1.03	1.10
Q59	Did a member of staff tell you about any danger signals you should watch for after you went home?	**1.06**	0.02	3.53	0.00	1.03	1.10

		Odds Ratio	Std. Err.	z	P>z [95%	Conf.	Interval]
Q66	Were the letters sent between hospital doctors and your family doctor (GP) written in a way that you could understand?	**1.06**	0.01	4.61	0.00	1.03	1.08
Q44	Beforehand, did a member of staff answer your questions about the operation or procedure?	**0.93**	0.02	-2.77	0.01	0.89	0.98
Q37	Were you given enough privacy when being examined or treated?	**0.81**	0.05	-3.11	0.00	0.72	0.93
Q33	How much information about your condition or treatment was given to you?	1.14	0.08	1.97	0.05	1.00	1.30
Q8	In your opinion, had the specialist you saw in hospital been given all of the necessary information about your condition or illness from the person who referred you?	1.05	0.03	1.73	0.08	0.99	1.11
Q64	Did hospital staff discuss with you whether you may need any further health or social care services after leaving hospital?	1.05	0.03	1.47	0.14	0.99	1.11
Q60	Did hospital staff take your family or home situation into account when planning your discharge?	1.02	0.02	1.46	0.15	0.99	1.05
Q56	Did a member of staff tell you about medication side effects to watch for when you went home?	1.03	0.02	1.42	0.16	0.99	1.06
Q47	Before the operation or procedure, did the anaesthetist or another member of staff explain how he or she would put you to sleep or control your pain in a way you could understand?	1.05	0.04	1.33	0.18	0.98	1.12
Q20	Were hand-wash gels available for patients and visitors to use?	0.95	0.05	-1.15	0.25	0.86	1.04
Q5	When you were referred to see a specialist, were you offered a choice of hospital?	1.02	0.02	1.09	0.28	0.98	1.06
Q43	Beforehand, did a member of staff explain what would be done during the operation or procedure?	1.04	0.04	1.04	0.30	0.97	1.12
Q7	Was your admission date changed by the hospital?	1.04	0.04	1.02	0.31	0.97	1.11
Q22	Were you offered a choice of food?	1.04	0.04	1.00	0.32	0.97	1.11
Q34	Did you find someone on the hospital staff to talk to about your worries and fears?	1.02	0.02	0.97	0.33	0.98	1.06
Q11	When you were first admitted to a bed on a ward, did you share a sleeping area, for example a room or bay, with patients of the	0.95	0.06	-0.82	0.41	0.84	1.08

		Odds Ratio	Std. Err.	z	P>z [95%	Conf.	Interval]
	opposite sex?						
Q19	Did you feel threatened during your stay in hospital by other patients or visitors?	1.09	0.15	0.60	0.55	0.82	1.44
Q58	Were you given clear written or printed information about your medicines?	1.01	0.02	0.56	0.58	0.97	1.05
Q24	When you had important questions to ask a doctor, did you get answers that you could understand?	1.01	0.03	0.46	0.65	0.96	1.07
Q26	Did doctors talk in front of you as if you weren't there?	1.02	0.04	0.42	0.68	0.94	1.10
Q55	Did a member of staff explain the purpose of the medicines you were to take at home in a way you could understand?	0.99	0.02	-0.36	0.72	0.94	1.04
Q63	Did hospital staff discuss with you whether you would need any additional equipment in your home, or any adaptations made to your home, after leaving hospital?	0.99	0.04	-0.32	0.75	0.92	1.06
Q14	While staying in hospital, did you ever use the same bathroom or shower area as patients of the opposite sex?	1.00	0.01	0.23	0.82	0.98	1.03
Q54	Before you left hospital, were you given any written or printed information about what you should or should not do after leaving hospital?	0.99	0.05	-0.22	0.82	0.90	1.09
Q57	Were you told how to take your medication in a way you could understand?	1.00	0.02	0.19	0.85	0.96	1.05
Q23	Did you get enough help from staff to eat your meals?	1.00	0.02	-0.18	0.85	0.96	1.04
Q42	Beforehand, did a member of staff explain the risks and benefits of the operation or procedure in a way you could understand?	1.01	0.05	0.12	0.91	0.91	1.11
	age_group	1.00	0.00	2.62	0.01	1.00	1.01
	all_gender	0.96	0.03	-1.19	0.24	0.90	1.03
	critical	0.98	0.02	-1.00	0.32	0.94	1.02
	LOS	1.13	0.05	2.89	0.00	1.04	1.23
	pain	0.55	0.04	-7.92	0.00	0.48	0.64
	selfreported	1.26	0.08	3.54	0.00	1.11	1.43

Appendix 9: Regression model for patients without an operation or procedure and emergency admission

Ordered logistic regression				Number of obs	=	11,636
				Wald chi2(54)	=	14228.07
				Prob > chi2	=	0.00
Log pseudolikelihood = -17115.288				Pseudo R2	=	0.25

		Odds Ratio	Std. Err.	z	P>z [95%	Conf.	Interval]
Q67	Overall, did you feel you were treated with respect and dignity while you were in the hospital?	**4.20**	0.24	25.00	0.00	3.75	4.70
Q28	Did you have confidence and trust in the nurses treating you?	**1.70**	0.09	10.64	0.00	1.54	1.88
Q17	In your opinion, how clean was the hospital room or ward that you were in?	**1.63**	0.06	13.45	0.00	1.52	1.75
Q25	Did you have confidence and trust in the doctors treating you?	**1.58**	0.07	10.19	0.00	1.44	1.72
Q9	From the time you arrived at the hospital, did you feel that you had to wait a long time to get to a bed on a ward?	**1.35**	0.03	12.15	0.00	1.29	1.42
Q21	How would you rate the hospital food?	**1.31**	0.02	16.02	0.00	1.27	1.36
Q32	Were you involved as much as you wanted to be in decisions about your care and treatment?	**1.29**	0.04	7.82	0.00	1.21	1.38
Q33	How much information about your condition or treatment was given to you?	**1.28**	0.06	5.01	0.00	1.16	1.40
Q31	Sometimes in a hospital, a member of staff will say one thing and another will say something quite different. Did this happen to you?	**1.26**	0.04	7.94	0.00	1.19	1.34
Q30	In your opinion, were there enough nurses on duty to care for you in hospital?	**1.23**	0.04	6.54	0.00	1.16	1.31
Q50	Were you given enough notice about when you were going to be discharged?	**1.22**	0.04	6.78	0.00	1.15	1.29
Q15	Were you ever bothered by noise at night from other patients?	**1.19**	0.04	5.00	0.00	1.11	1.28
Q39	Do you think the hospital staff did everything they could to help control your pain?	**1.16**	0.04	4.88	0.00	1.09	1.23
Q35	Do you feel you got enough emotional support from hospital	**1.14**	0.03	5.99	0.00	1.09	1.19

		Odds Ratio	Std. Err.	z	P>z [95%	Conf.	Interval]
	staff during your stay?						
Q69	During your hospital stay, were you ever asked to give your views on the quality of your care?	**1.13**	0.03	5.07	0.00	1.08	1.18
Q49	Did you feel you were involved in decisions about your discharge from hospital?	**1.11**	0.03	4.09	0.00	1.06	1.17
Q4	Were you given enough privacy when being examined or treated in the A&E Department?	**1.11**	0.03	3.70	0.00	1.05	1.17
Q18	How clean were the toilets and bathrooms that you used in hospital?	**1.11**	0.03	3.97	0.00	1.05	1.16
Q62	Did hospital staff tell you who to contact if you were worried about your condition or treatment after you left hospital?	**1.07**	0.03	2.76	0.01	1.02	1.13
Q61	Did the doctors or nurses give your family or someone close to you all the information they needed to help care for you?	**1.07**	0.02	3.53	0.00	1.03	1.11
Q66	Were the letters sent between hospital doctors and your family doctor (GP) written in a way that you could understand?	**1.07**	0.01	5.14	0.00	1.04	1.09
Q70	Did you see, or were you given, any information explaining how to complain to the hospital about the care you received?	**1.06**	0.03	2.21	0.03	1.01	1.12
Q59	Did a member of staff tell you about any danger signals you should watch for after you went home?	**1.06**	0.02	2.73	0.01	1.02	1.10
Q51	How long was the delay in discharge?	**1.05**	0.01	3.60	0.00	1.02	1.08
Q14	While staying in hospital, did you ever use the same bathroom or shower area as patients of the opposite sex?	**1.04**	0.01	3.17	0.00	1.01	1.06
Q55	Did a member of staff explain the purpose of the medicines you were to take at home in a way you could understand?	1.05	0.02	1.97	0.05	1.00	1.10
Q19	Did you feel threatened during your stay in hospital by other patients or visitors?	1.20	0.11	1.95	0.05	1.00	1.45
Q57	Were you told how to take your medication in a way you could understand?	0.96	0.02	-1.86	0.06	0.92	1.00
Q54	Before you left hospital, were you given any written or printed information about what you should or should not do after leaving hospital?	1.07	0.04	1.75	0.08	0.99	1.15
Q37	Were you given enough privacy	0.91	0.05	-1.73	0.08	0.81	1.01

		Odds Ratio	Std. Err.	z	P>z [95%	Conf.	Interval]
	when being examined or treated?						
Q11	When you were first admitted to a bed on a ward, did you share a sleeping area, for example a room or bay, with patients of the opposite sex?	1.10	0.06	1.71	0.09	0.99	1.21
Q3	While you were in the A&E Department, how much information about your condition?	1.03	0.02	1.53	0.13	0.99	1.08
Q27	When you had important questions to ask a nurse, did you get answers that you could understand?	1.04	0.03	1.52	0.13	0.99	1.09
Q56	Did a member of staff tell you about medication side effects to watch for when you went home?	1.03	0.02	1.36	0.18	0.99	1.07
Q16	Were you ever bothered by noise at night from hospital staff?	1.06	0.05	1.30	0.19	0.97	1.16
Q60	Did hospital staff take your family or home situation into account when planning your discharge?	1.02	0.02	1.11	0.27	0.99	1.05
Q63	Did hospital staff discuss with you whether you would need any additional equipment in your home, or any adaptations made to your home, after leaving hospital?	0.96	0.04	-1.06	0.29	0.88	1.04
Q36	Were you given enough privacy when discussing your condition or treatment?	1.03	0.04	0.92	0.36	0.96	1.11
Q23	Did you get enough help from staff to eat your meals?	1.02	0.02	0.82	0.41	0.97	1.07
Q40	How many minutes after you used the call button did it usually take before you got the help you needed?	1.02	0.02	0.78	0.43	0.97	1.07
Q20	Were hand-wash gels available for patients and visitors to use?	1.03	0.05	0.73	0.47	0.95	1.12
Q24	When you had important questions to ask a doctor, did you get answers that you could understand?	1.02	0.03	0.72	0.47	0.97	1.08
Q26	Did doctors talk in front of you as if you weren't there?	1.02	0.04	0.44	0.66	0.94	1.10
Q58	Were you given clear written or printed information about your medicines?	1.01	0.02	0.40	0.69	0.98	1.04
Q64	Did hospital staff discuss with you whether you may need any further health or social care services after leaving hospital?	1.01	0.04	0.19	0.85	0.94	1.08
Q34	Did you find someone on the hospital staff to talk to about your worries and fears?	1.00	0.02	0.18	0.86	0.96	1.05
Q29	Did nurses talk in front of you as if you weren't there?	0.99	0.05	-0.14	0.89	0.91	1.09

		Odds Ratio	Std. Err.	z	P>z [95%	Conf.	Interval]
Q22	Were you offered a choice of food?	1.00	0.03	0.09	0.93	0.94	1.07
	age_group	1.00	0.00	1.26	0.21	1.00	1.01
	all_gender	0.95	0.03	-1.58	0.11	0.88	1.01
	critical	0.87	0.02	-6.66	0.00	0.84	0.91
	LOS	1.23	0.06	4.58	0.00	1.13	1.35
	pain	0.72	0.04	-5.52	0.00	0.64	0.81
	selfreported	1.37	0.07	6.25	0.00	1.24	1.50

Appendix 10: Regression model for patients without an operation or procedure and planned admission

Ordered logistic regression				Number of obs	=	1,780
				Wald chi2(56)	=	1727.06
				Prob > chi2	=	0.00
Log pseudolikelihood = -2405.3253				Pseudo R2	=	0.24

		Odds Ratio	Std. Err.	z	P>z	[95% Conf.	Interval]
Q17	In your opinion, how clean was the hospital room or ward that you were in?	**1.71**	0.21	4.51	0.00	1.36	2.17
Q21	How would you rate the hospital food?	**1.41**	0.05	8.86	0.00	1.31	1.52
Q30	In your opinion, were there enough nurses on duty to care for you in hospital?	**1.43**	0.12	4.13	0.00	1.21	1.69
Q32	Were you involved as much as you wanted to be in decisions about your care and treatment?	**1.51**	0.14	4.36	0.00	1.25	1.81
Q67	Overall, did you feel you were treated with respect and dignity while you were in the hospital?	**4.33**	0.81	7.80	0.00	3.00	6.26
Q9	From the time you arrived at the hospital, did you feel that you had to wait a long time to get to a bed on a ward?	**1.34**	0.11	3.45	0.00	1.13	1.58
Q25	Did you have confidence and trust in the doctors treating you?	**1.64**	0.24	3.41	0.00	1.23	2.18
Q50	Were you given enough notice about when you were going to be discharged?	**1.34**	0.13	3.06	0.00	1.11	1.62
Q70	Did you see, or were you given, any information explaining how to complain to the hospital about the care you received?	**1.17**	0.07	2.79	0.01	1.05	1.30
Q28	Did you have confidence and trust in the nurses treating you?	**1.54**	0.25	2.64	0.01	1.12	2.13
Q33	How much information about your condition or treatment was given to you?	**1.48**	0.24	2.40	0.02	1.08	2.05
Q31	Sometimes in a hospital, a member of staff will say one thing and another will say something quite different. Did this happen to you?	**1.28**	0.14	2.31	0.02	1.04	1.58
Q49	Did you feel you were involved in decisions about your discharge	**1.16**	0.08	2.20	0.03	1.02	1.32

		Odds Ratio	Std. Err.	z	P>z [95%	Conf.	Interval]
	from hospital?						
Q61	Did the doctors or nurses give your family or someone close to you all the information they needed to help care for you?	**1.12**	0.06	2.07	0.04	1.01	1.25
Q15	Were you ever bothered by noise at night from other patients?	1.18	0.10	1.91	0.06	1.00	1.41
Q18	How clean were the toilets and bathrooms that you used in hospital?	1.14	0.08	1.85	0.06	0.99	1.31
Q69	During your hospital stay, were you ever asked to give your views on the quality of your care?	1.11	0.06	1.83	0.07	0.99	1.24
Q60	Did hospital staff take your family or home situation into account when planning your discharge?	0.91	0.04	-1.82	0.07	0.83	1.01
Q16	Were you ever bothered by noise at night from hospital staff?	1.29	0.19	1.75	0.08	0.97	1.73
Q27	When you had important questions to ask a nurse, did you get answers that you could understand?	1.13	0.08	1.69	0.09	0.98	1.31
Q8	In your opinion, had the specialist you saw in hospital been given all of the necessary information about your condition or illness from the person who referred you?	1.12	0.08	1.67	0.10	0.98	1.28
Q5	When you were referred to see a specialist, were you offered a choice of hospital?	1.10	0.07	1.46	0.14	0.97	1.25
Q34	Did you find someone on the hospital staff to talk to about your worries and fears?	0.92	0.05	-1.44	0.15	0.82	1.03
Q51	How long was the delay in discharge?	1.05	0.04	1.32	0.19	0.98	1.12
Q23	Did you get enough help from staff to eat your meals?	0.93	0.06	-1.26	0.21	0.82	1.04
Q14	While staying in hospital, did you ever use the same bathroom or shower area as patients of the opposite sex?	1.04	0.03	1.19	0.24	0.98	1.11
Q36	Were you given enough privacy when discussing your condition or treatment?	1.11	0.10	1.18	0.24	0.93	1.32
Q26	Did doctors talk in front of you as if you weren't there?	0.87	0.11	-1.14	0.25	0.68	1.11
Q40	How many minutes after you used the call button did it usually take before you got the help you needed?	0.94	0.06	-1.05	0.30	0.84	1.06
Q29	Did nurses talk in front of you as if you weren't there?	1.16	0.16	1.03	0.30	0.88	1.52

		Odds Ratio	Std. Err.	z	P>z [95%	Conf.	Interval]
Q64	Did hospital staff discuss with you whether you may need any further health or social care services after leaving hospital?	1.08	0.10	0.83	0.41	0.90	1.29
Q20	Were hand-wash gels available for patients and visitors to use?	1.11	0.14	0.82	0.41	0.86	1.43
Q39	Do you think the hospital staff did everything they could to help control your pain?	1.09	0.12	0.80	0.42	0.88	1.36
Q6	How do you feel about the length of time you were on the waiting list before your admission to hospital?	0.93	0.09	-0.73	0.47	0.78	1.12
Q37	Were you given enough privacy when being examined or treated?	0.91	0.14	-0.62	0.54	0.66	1.24
Q66	Were the letters sent between hospital doctors and your family doctor (GP) written in a way that you could understand?	1.01	0.03	0.46	0.65	0.95	1.08
Q19	Did you feel threatened during your stay in hospital by other patients or visitors?	1.14	0.38	0.40	0.69	0.60	2.17
Q56	Did a member of staff tell you about medication side effects to watch for when you went home?	1.02	0.06	0.32	0.75	0.91	1.14
Q62	Did hospital staff tell you who to contact if you were worried about your condition or treatment after you left hospital?	0.98	0.07	-0.31	0.76	0.84	1.13
Q54	Before you left hospital, were you given any written or printed information about what you should or should not do after leaving hospital?	0.97	0.10	-0.31	0.76	0.79	1.18
Q57	Were you told how to take your medication in a way you could understand?	0.98	0.07	-0.29	0.78	0.85	1.13
Q58	Were you given clear written or printed information about your medicines?	1.01	0.05	0.23	0.82	0.92	1.12
Q59	Did a member of staff tell you about any danger signals you should watch for after you went home?	0.99	0.05	-0.22	0.82	0.90	1.09
Q63	Did hospital staff discuss with you whether you would need any additional equipment in your home, or any adaptations made to your home, after leaving hospital?	0.98	0.11	-0.21	0.83	0.79	1.21
Q24	When you had important questions to ask a doctor, did you get answers that you could understand?	1.01	0.06	0.10	0.92	0.89	1.14

		Odds Ratio	Std. Err.	z	P>z [95%	Conf.	Interval]
Q35	Do you feel you got enough emotional support from hospital staff during your stay?	1.00	0.06	0.07	0.94	0.90	1.12
Q7	Was your admission date changed by the hospital?	1.01	0.18	0.05	0.96	0.71	1.44
Q22	Were you offered a choice of food?	1.00	0.10	-0.04	0.97	0.82	1.22
Q11	When you were first admitted to a bed on a ward, did you share a sleeping area, for example a room or bay, with patients of the opposite sex?	1.00	0.16	-0.01	0.99	0.73	1.36
Q55	Did a member of staff explain the purpose of the medicines you were to take at home in a way you could understand?	1.00	0.08	0.01	0.99	0.86	1.17
	age_group	0.99	0.00	-1.84	0.07	0.99	1.00
	all_gender	0.96	0.09	-0.39	0.69	0.80	1.16
	critical	0.99	0.05	-0.15	0.88	0.91	1.09
	LOS	1.20	0.19	1.20	0.23	0.89	1.63
	pain	0.85	0.18	-0.77	0.44	0.57	1.28
	selfreported	1.35	0.19	2.17	0.03	1.03	1.77

Appendix 11: Eigen value estimated for the subgroups

EIGENVALUES FOR SAMPLE CORRELATION MATRIX

	1	2	3	4	5
1	16.879	3.605	2.396	1.707	1.511

EIGENVALUES FOR SAMPLE CORRELATION MATRIX

	6	7	8	9	10
1	1.412	1.329	1.179	1.126	1.074

EIGENVALUES FOR SAMPLE CORRELATION MATRIX

	11	12	13	14	15
1	0.978	0.966	0.910	0.864	0.847

	1	2	3	4	5
1	14.028	3.349	1.831	1.576	1.558

EIGENVALUES FOR SAMPLE CORRELATION MATRIX

	6	7	8	9	10
1	1.445	1.293	1.158	1.034	1.015

EIGENVALUES FOR SAMPLE CORRELATION MATRIX

	11	12	13	14	15
1	0.924	0.917	0.886	0.874	0.827

	1	2	3	4	5
1	15.835	3.755	2.366	1.735	1.468

EIGENVALUES FOR SAMPLE CORRELATION MATRIX

	6	7	8	9	10
1	1.406	1.319	1.270	1.223	1.133

EIGENVALUES FOR SAMPLE CORRELATION MATRIX

	11	12	13	14	15
1	1.040	0.997	0.959	0.929	0.920

	1	2	3	4	5
1	14.338	3.759	1.865	1.666	1.636

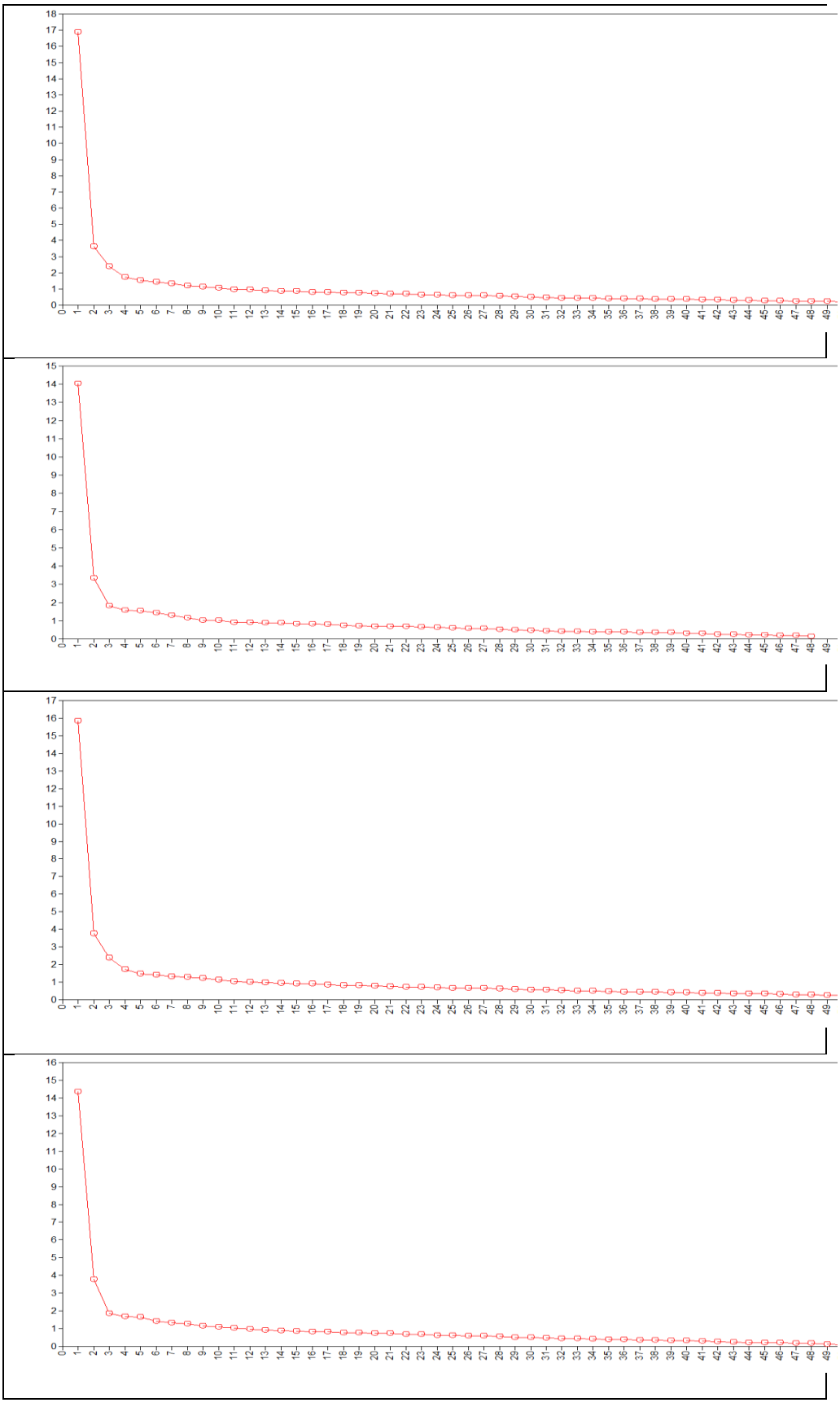
EIGENVALUES FOR SAMPLE CORRELATION MATRIX

	6	7	8	9	10
1	1.412	1.334	1.282	1.147	1.081

EIGENVALUES FOR SAMPLE CORRELATION MATRIX

	11	12	13	14	15
1	1.035	0.965	0.900	0.870	0.856

Appendix 12: Scree plots of the subgroups



Appendix 13: An 11-Factor model (EFA) for patients with an operation or procedure with Emergency admission

Fit of the 11-factor model

```
       MODEL FIT INFORMATION

Number of Free Parameters                       676

Chi-Square Test of Model Fit

          Value                          10253.624*
          Degrees of Freedom                   892
          P-Value                          0.0000

RMSEA (Root Mean Square Error Of Approximation)

          Estimate                           0.026

          90 Percent C.I.                    0.026  0.027

          Probability RMSEA <= .05           1.000



CFI/TLI



          CFI                                0.980
          TLI                                0.968

Chi-Square Test of Model Fit for the Baseline Model

          Value                         478028.071
          Degrees of Freedom                 1431
          P-Value                          0.0000

WRMR (Weighted Root Mean Square Residual)

          Value                              1.814
```

Factors and items identified by EFA

Factor Number	Item 1	Item 2	Item 3	Item 4	Item 5
Factor 1	Beforehand, did a member of staff explain the risks and benefits of the operation or procedure in a way you could understand?	Beforehand, did a member of staff explain what would be done during the operation or procedure?	Beforehand, did a member of staff answer your questions about the operation or procedure?	Beforehand, were you told how you could expect to feel after you had the operation or procedure?	After the operation or procedure, did a member of staff explain how the operation or procedure had gone in a way you could understand?
Factor 2	Did nurses talk in front of you as if you weren't there?				
Factor 3	When you had important questions to ask a doctor, did you get answers that you could understand?	Did you have confidence and trust in the doctors treating you?	Were you involved as much as you wanted to be in decisions about your care and treatment?	How much information about your condition or treatment was given to you?	
Factor 4	Were you given enough privacy when being examined or treated?				
Factor 5	Were you ever bothered by noise at night from other patients?	Were you ever bothered by noise at night from hospital staff?			

Factor Number	Item 1	Item 2	Item 3	Item 4	Item 5
Factor 6	Did you find someone on the hospital staff to talk to about your worries and fears?	Do you feel you got enough emotional support from hospital staff during your stay?			
Factor 7	During your hospital stay, were you ever asked to give your views on the quality of your care?	Did you see, or were you given, any information explaining how to complain to the hospital about the care you received?			
Factor 8	In your opinion, how clean was the hospital room or ward that you were in?	How clean were the toilets and bathrooms that you used in hospital?			
Factor 9	Did a member of staff explain the purpose of the medicines you were to take at home in a way you could understand?	Did a member of staff tell you about medication side effects to watch for when you went home?	Were you told how to take your medication in a way you could understand?	Were you given clear written or printed information about your medicines?	Did a member of staff tell you about any danger signals you should watch for after you went home?
Factor 10	Did hospital staff take your family or home situation into account when planning your discharge?	Did hospital staff discuss with you whether you would need any additional equipment in your home, or any adaptations made to your home, after leaving hospital?	Did hospital staff discuss with you whether you may need any further health or social care services after leaving hospital?		

Factor Number	Item 1	Item 2	Item 3	Item 4	Item 5
Factor 11	Did you feel you were involved in decisions about your discharge from hospital?	Were you given enough notice about when you were going to be discharged?			

Appendix 14: An 10-Factor model (EFA) for patients with an operation or procedure with planned admission

Fit of the 10-factor model

```
MODEL FIT INFORMATION

Number of Free Parameters                       656

Chi-Square Test of Model Fit

          Value                          16103.418*
          Degrees of Freedom                  1025
          P-Value                           0.0000

*   The chi-square value for MLM, MLMV, MLR, ULSMV, WLSM and WLSMV cannot be
used
    for chi-square difference testing in the regular way.  MLM, MLR and WLSM
    chi-square difference testing is described on the Mplus website.  MLMV,
WLSMV,
    and ULSMV difference testing is done using the DIFFTEST option.

RMSEA (Root Mean Square Error Of Approximation)

          Estimate                           0.025

          90 Percent C.I.                    0.025  0.025

          Probability RMSEA <= .05           1.000



CFI/TLI



          CFI                                0.974
          TLI                                0.961

Chi-Square Test of Model Fit for the Baseline Model

          Value                         580171.713
          Degrees of Freedom                  1540
          P-Value                           0.0000

WRMR (Weighted Root Mean Square Residual)

          Value                              2.316
```

Factors and items identified by EFA

Factor number	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7
Factor 1	*Items not identified*						
Factor 2	How do you feel about the length of time you were on the waiting list before your admission to hospital?						
Factor 3	How much information about your condition or treatment was given to you?	Beforehand, did a member of staff explain the risks and benefits of the operation or procedure in a way you could understand?	Beforehand, did a member of staff explain what would be done during the operation or procedure?	Beforehand, did a member of staff answer your questions about the operation or procedure?	Beforehand, were you told how you could expect to feel after you had the operation or procedure?	After the operation or procedure, did a member of staff explain how the operation or procedure had gone in a way you could understand?	
Factor 4	During your hospital stay, were you ever asked to give your views on the quality of your care?	Did you see, or were you given, any information explaining how to complain to the hospital about the care you received?					

Factor number	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7
Factor 5	Did you have confidence and trust in the nurses treating you?	Did nurses talk in front of you as if you weren't there?	In your opinion, were there enough nurses on duty to care for you in hospital?	Sometimes in a hospital, a member of staff will say one thing and another will say something quite different. Did this happen to you?	Were you given enough privacy when discussing your condition or treatment?	Were you given enough privacy when being examined or treated?	Overall, did you feel you were treated with respect and dignity while you were in the hospital?
Factor 6	Did you feel you were involved in decisions about your discharge from hospital?	Were you given enough notice about when you were going to be discharged?					
Factor 7	In your opinion, how clean was the hospital room or ward that you were in?	How clean were the toilets and bathrooms that you used in hospital?					
Factor 8	Did hospital staff discuss with you whether you would need any additional equipment in	Did hospital staff discuss with you whether you may need any further health or social care					

Factor number	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7
	your home, or any adaptations made to your home, after leaving hospital?	services after leaving hospital?					
Factor 9	Did a member of staff explain the purpose of the medicines you were to take at home in a way you could understand?	Did a member of staff tell you about medication side effects to watch for when you went home?	Were you told how to take your medication in a way you could understand?	Were you given clear written or printed information about your medicines?			
Factor 10	Did you find someone on the hospital staff to talk to about your worries and fears?	Do you feel you got enough emotional support from hospital staff during your stay?					

Appendix 15: Factor Score by dimension for patients with operation or procedure and emergency admission

MODEL FIT INFORMATION

Number of Free Parameters	676	
RMSEA (Root Mean Square Error Of Approximation)		
Estimate	0.026	
90 Percent C.I.	0.026	0.027
Probability RMSEA <= .05	1.000	
CFI/TLI		
CFI	0.980	
TLI	0.968	
WRMR (Weighted Root Mean Square Residual)		
Value	1.814	

MODEL RESULTS

		Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
F1	BY				
Q3		-0.124	0.019	-6.424	0.000
Q4		0.019	0.015	1.229	0.219
Q9		0.118	0.017	6.765	0.000
Q11		0.012	0.018	0.678	0.498
Q14		0.101	0.020	4.994	0.000
Q15		0.034	0.014	2.373	0.018
Q16		0.007	0.013	0.562	0.574
Q17		0.896	0.014	62.298	0.000
Q18		0.802	0.015	54.550	0.000
Q19		0.048	0.027	1.787	0.074
Q20		0.239	0.025	9.714	0.000
Q21		0.387	0.015	26.517	0.000
Q22		0.329	0.020	16.817	0.000
Q23		0.136	0.017	8.025	0.000
Q24		0.008	0.012	0.666	0.505
Q25		0.109	0.018	6.059	0.000
Q26		-0.085	0.015	-5.564	0.000
Q27		0.153	0.016	9.294	0.000
Q28		0.297	0.018	16.451	0.000
Q29		0.042	0.011	3.653	0.000
Q30		0.281	0.016	17.166	0.000
Q31		0.128	0.018	7.081	0.000
Q32		0.030	0.012	2.471	0.013
Q33		-0.059	0.016	-3.743	0.000
Q34		-0.066	0.013	-4.954	0.000
Q35		0.010	0.008	1.225	0.221
Q36		0.005	0.012	0.459	0.646
Q37		0.011	0.012	0.859	0.390
Q39		0.039	0.015	2.588	0.010
Q40		0.185	0.016	11.683	0.000
Q42		0.015	0.010	1.462	0.144
Q43		0.034	0.010	3.527	0.000
Q44		-0.007	0.010	-0.723	0.469

Q45	0.053	0.012	4.397	0.000
Q47	-0.042	0.016	-2.700	0.007
Q48	0.022	0.013	1.757	0.079
Q49	0.009	0.009	0.989	0.323
Q50	0.037	0.011	3.390	0.001
Q51	0.058	0.016	3.561	0.000
Q54	0.030	0.015	1.967	0.049
Q55	0.012	0.010	1.211	0.226
Q56	-0.006	0.011	-0.504	0.614
Q57	0.005	0.009	0.608	0.543
Q58	0.041	0.013	3.153	0.002
Q59	-0.039	0.011	-3.415	0.001
Q60	0.006	0.011	0.575	0.565
Q61	-0.009	0.010	-0.827	0.408
Q62	0.011	0.014	0.808	0.419
Q63	0.031	0.011	2.696	0.007
Q64	0.000	0.011	0.031	0.975
Q66	0.008	0.016	0.526	0.599
Q67	0.235	0.016	14.228	0.000
Q69	0.032	0.013	2.431	0.015
Q70	-0.001	0.010	-0.088	0.930
F2 BY				
Q3	0.479	0.031	15.549	0.000
Q4	0.380	0.058	6.567	0.000
Q9	0.312	0.022	14.124	0.000
Q11	-0.103	0.053	-1.928	0.054
Q14	-0.146	0.047	-3.091	0.002
Q15	0.005	0.015	0.345	0.730
Q16	0.025	0.016	1.560	0.119
Q17	0.015	0.011	1.386	0.166
Q18	-0.028	0.012	-2.279	0.023
Q19	0.073	0.037	1.951	0.051
Q20	0.037	0.031	1.224	0.221
Q21	0.095	0.018	5.133	0.000
Q22	0.127	0.028	4.540	0.000
Q23	-0.052	0.021	-2.501	0.012
Q24	0.658	0.020	32.392	0.000
Q25	0.766	0.018	42.442	0.000
Q26	0.343	0.031	11.026	0.000
Q27	0.307	0.022	13.770	0.000
Q28	0.312	0.029	10.687	0.000
Q29	0.055	0.013	4.212	0.000
Q30	0.268	0.020	13.547	0.000
Q31	0.521	0.020	25.595	0.000
Q32	0.602	0.036	16.776	0.000
Q33	0.721	0.022	32.501	0.000
Q34	0.007	0.012	0.641	0.522
Q35	0.006	0.010	0.585	0.559
Q36	0.279	0.076	3.657	0.000
Q37	0.276	0.092	3.017	0.003
Q39	0.124	0.021	5.824	0.000
Q40	0.075	0.021	3.609	0.000
Q42	0.126	0.017	7.377	0.000
Q43	0.009	0.011	0.824	0.410
Q44	-0.009	0.011	-0.830	0.406
Q45	-0.003	0.013	-0.207	0.836

Q47	0.047	0.021	2.235	0.025
Q48	0.284	0.023	12.446	0.000
Q49	0.285	0.083	3.429	0.001
Q50	0.366	0.081	4.513	0.000
Q51	0.159	0.029	5.412	0.000
Q54	0.128	0.025	5.204	0.000
Q55	0.046	0.017	2.742	0.006
Q56	-0.058	0.016	-3.560	0.000
Q57	-0.061	0.014	-4.372	0.000
Q58	0.032	0.015	2.203	0.028
Q59	0.026	0.019	1.413	0.158
Q60	-0.008	0.020	-0.423	0.672
Q61	-0.001	0.013	-0.069	0.945
Q62	0.140	0.022	6.306	0.000
Q63	-0.009	0.010	-0.892	0.373
Q64	0.066	0.022	3.006	0.003
Q66	0.226	0.029	7.776	0.000
Q67	0.386	0.020	19.685	0.000
Q69	0.000	0.013	0.024	0.980
Q70	-0.030	0.015	-1.993	0.046
F3 BY				
Q3	-0.017	0.015	-1.129	0.259
Q4	-0.028	0.015	-1.816	0.069
Q9	0.129	0.019	6.921	0.000
Q11	-0.083	0.024	-3.478	0.001
Q14	0.005	0.017	0.278	0.781
Q15	0.588	0.023	25.011	0.000
Q16	0.544	0.024	22.254	0.000
Q17	-0.019	0.010	-1.951	0.051
Q18	0.009	0.010	0.944	0.345
Q19	0.323	0.032	10.083	0.000
Q20	-0.054	0.025	-2.189	0.029
Q21	0.102	0.015	6.641	0.000
Q22	-0.077	0.019	-4.085	0.000
Q23	0.031	0.014	2.139	0.032
Q24	-0.135	0.018	-7.457	0.000
Q25	0.066	0.023	2.902	0.004
Q26	0.004	0.010	0.456	0.648
Q27	-0.035	0.014	-2.569	0.010
Q28	0.178	0.021	8.518	0.000
Q29	0.044	0.012	3.699	0.000
Q30	0.202	0.019	10.801	0.000
Q31	0.172	0.022	7.877	0.000
Q32	-0.030	0.013	-2.216	0.027
Q33	0.058	0.019	3.058	0.002
Q34	-0.061	0.012	-5.038	0.000
Q35	0.036	0.010	3.673	0.000
Q36	0.090	0.018	4.960	0.000
Q37	-0.007	0.011	-0.630	0.528
Q39	-0.081	0.016	-5.154	0.000
Q40	0.040	0.014	2.749	0.006
Q42	0.003	0.009	0.366	0.714
Q43	0.032	0.009	3.498	0.000
Q44	-0.065	0.012	-5.254	0.000
Q45	0.154	0.015	10.566	0.000
Q47	-0.037	0.014	-2.624	0.009

Q48	0.037	0.013	2.848	0.004
Q49	-0.130	0.018	-7.130	0.000
Q50	0.006	0.008	0.724	0.469
Q51	0.211	0.017	12.193	0.000
Q54	0.160	0.021	7.599	0.000
Q55	-0.048	0.010	-4.694	0.000
Q56	0.262	0.015	17.289	0.000
Q57	-0.025	0.008	-3.257	0.001
Q58	-0.018	0.011	-1.702	0.089
Q59	0.307	0.018	16.861	0.000
Q60	0.031	0.011	2.899	0.004
Q61	0.194	0.016	12.010	0.000
Q62	0.158	0.021	7.582	0.000
Q63	-0.094	0.017	-5.604	0.000
Q64	-0.040	0.012	-3.322	0.001
Q66	-0.035	0.015	-2.306	0.021
Q67	0.143	0.020	7.336	0.000
Q69	-0.051	0.013	-3.967	0.000
Q70	0.037	0.011	3.329	0.001
F4 BY				
Q3	0.131	0.021	6.378	0.000
Q4	0.082	0.019	4.352	0.000
Q9	-0.064	0.015	-4.253	0.000
Q11	0.022	0.018	1.252	0.210
Q14	0.032	0.016	1.967	0.049
Q15	0.019	0.013	1.527	0.127
Q16	0.011	0.013	0.843	0.399
Q17	0.055	0.012	4.441	0.000
Q18	0.064	0.013	4.923	0.000
Q19	0.022	0.028	0.779	0.436
Q20	0.085	0.024	3.528	0.000
Q21	-0.009	0.011	-0.805	0.421
Q22	0.015	0.016	0.963	0.335
Q23	0.024	0.014	1.701	0.089
Q24	0.132	0.017	7.583	0.000
Q25	0.033	0.014	2.424	0.015
Q26	0.030	0.012	2.623	0.009
Q27	0.049	0.012	4.026	0.000
Q28	-0.029	0.011	-2.753	0.006
Q29	0.016	0.010	1.695	0.090
Q30	-0.053	0.013	-4.165	0.000
Q31	-0.100	0.015	-6.875	0.000
Q32	0.110	0.015	7.489	0.000
Q33	0.130	0.019	6.751	0.000
Q34	0.088	0.014	6.481	0.000
Q35	0.024	0.009	2.561	0.010
Q36	-0.004	0.011	-0.347	0.729
Q37	0.014	0.012	1.184	0.237
Q39	0.040	0.014	2.797	0.005
Q40	-0.024	0.013	-1.818	0.069
Q42	0.797	0.010	81.980	0.000
Q43	0.926	0.008	116.726	0.000
Q44	0.783	0.009	91.361	0.000
Q45	0.687	0.010	65.646	0.000
Q47	0.441	0.014	30.928	0.000
Q48	0.489	0.013	37.489	0.000

Q49	0.058	0.013	4.580	0.000
Q50	-0.039	0.010	-3.882	0.000
Q51	-0.076	0.015	-5.084	0.000
Q54	0.017	0.013	1.316	0.188
Q55	-0.004	0.008	-0.435	0.663
Q56	0.124	0.014	8.701	0.000
Q57	-0.034	0.009	-3.779	0.000
Q58	-0.031	0.011	-2.884	0.004
Q59	0.160	0.015	10.959	0.000
Q60	0.011	0.010	1.119	0.263
Q61	0.029	0.010	2.834	0.005
Q62	0.079	0.015	5.268	0.000
Q63	-0.017	0.009	-1.867	0.062
Q64	-0.031	0.011	-2.788	0.005
Q66	0.067	0.015	4.352	0.000
Q67	-0.015	0.011	-1.392	0.164
Q69	-0.026	0.012	-2.165	0.030
Q70	0.006	0.010	0.635	0.525
F5 BY				
Q3	0.097	0.016	5.902	0.000
Q4	0.066	0.016	4.131	0.000
Q9	0.015	0.014	1.117	0.264
Q11	-0.062	0.019	-3.277	0.001
Q14	0.013	0.015	0.903	0.367
Q15	-0.034	0.013	-2.695	0.007
Q16	0.005	0.012	0.421	0.673
Q17	-0.014	0.009	-1.502	0.133
Q18	0.018	0.009	1.891	0.059
Q19	0.019	0.025	0.764	0.445
Q20	0.032	0.023	1.425	0.154
Q21	-0.014	0.011	-1.283	0.200
Q22	-0.001	0.016	-0.079	0.937
Q23	0.052	0.014	3.748	0.000
Q24	0.026	0.011	2.274	0.023
Q25	-0.033	0.011	-2.956	0.003
Q26	-0.022	0.009	-2.338	0.019
Q27	0.088	0.012	7.121	0.000
Q28	0.035	0.011	3.251	0.001
Q29	0.056	0.012	4.825	0.000
Q30	0.017	0.011	1.464	0.143
Q31	-0.016	0.011	-1.499	0.134
Q32	-0.013	0.010	-1.235	0.217
Q33	-0.013	0.011	-1.166	0.244
Q34	-0.012	0.009	-1.345	0.179
Q35	-0.028	0.008	-3.312	0.001
Q36	0.029	0.011	2.681	0.007
Q37	0.019	0.011	1.848	0.065
Q39	0.033	0.013	2.474	0.013
Q40	0.037	0.012	3.018	0.003
Q42	-0.003	0.009	-0.311	0.756
Q43	-0.037	0.008	-4.357	0.000
Q44	0.018	0.009	2.088	0.037
Q45	0.057	0.011	5.203	0.000
Q47	0.087	0.014	6.102	0.000
Q48	0.020	0.011	1.822	0.068
Q49	0.015	0.008	1.976	0.048

Q50	0.035	0.009	3.804	0.000
Q51	-0.080	0.014	-5.673	0.000
Q54	0.439	0.016	27.430	0.000
Q55	0.855	0.009	100.565	0.000
Q56	0.617	0.012	51.795	0.000
Q57	0.927	0.009	106.004	0.000
Q58	0.692	0.009	75.072	0.000
Q59	0.415	0.015	27.376	0.000
Q60	0.046	0.010	4.386	0.000
Q61	0.150	0.014	10.405	0.000
Q62	0.296	0.016	18.296	0.000
Q63	-0.050	0.009	-5.462	0.000
Q64	0.065	0.012	5.401	0.000
Q66	0.150	0.015	10.245	0.000
Q67	0.027	0.011	2.351	0.019
Q69	0.012	0.011	1.109	0.267
Q70	0.009	0.010	0.922	0.357
F6 BY				
Q3	-0.058	0.015	-3.837	0.000
Q4	-0.042	0.013	-3.180	0.001
Q9	-0.039	0.014	-2.887	0.004
Q11	-0.151	0.025	-5.980	0.000
Q14	-0.104	0.021	-5.042	0.000
Q15	-0.055	0.014	-3.984	0.000
Q16	0.066	0.015	4.421	0.000
Q17	-0.004	0.007	-0.561	0.575
Q18	-0.048	0.010	-4.725	0.000
Q19	0.027	0.020	1.334	0.182
Q20	0.122	0.022	5.564	0.000
Q21	0.030	0.011	2.679	0.007
Q22	0.078	0.017	4.723	0.000
Q23	0.275	0.014	19.634	0.000
Q24	0.085	0.014	6.270	0.000
Q25	0.014	0.011	1.338	0.181
Q26	-0.011	0.010	-1.138	0.255
Q27	0.382	0.015	25.901	0.000
Q28	0.356	0.016	22.783	0.000
Q29	0.236	0.024	9.731	0.000
Q30	0.125	0.013	9.321	0.000
Q31	0.000	0.011	-0.007	0.995
Q32	0.061	0.011	5.341	0.000
Q33	0.033	0.012	2.843	0.004
Q34	0.637	0.013	47.784	0.000
Q35	0.724	0.014	52.134	0.000
Q36	0.066	0.015	4.383	0.000
Q37	0.079	0.019	4.219	0.000
Q39	0.225	0.014	15.539	0.000
Q40	0.228	0.013	17.269	0.000
Q42	0.004	0.009	0.455	0.649
Q43	0.012	0.008	1.531	0.126
Q44	0.153	0.011	13.396	0.000
Q45	0.020	0.009	2.108	0.035
Q47	-0.029	0.013	-2.203	0.028
Q48	-0.023	0.011	-2.100	0.036
Q49	0.031	0.009	3.508	0.000
Q50	-0.023	0.008	-2.794	0.005

Q51	-0.002	0.013	-0.188	0.851
Q54	-0.106	0.015	-7.004	0.000
Q55	0.025	0.008	2.932	0.003
Q56	0.017	0.009	1.897	0.058
Q57	0.062	0.010	6.451	0.000
Q58	-0.010	0.010	-1.009	0.313
Q59	0.003	0.008	0.393	0.694
Q60	0.081	0.012	6.853	0.000
Q61	0.066	0.011	5.873	0.000
Q62	-0.003	0.012	-0.237	0.813
Q63	-0.021	0.008	-2.532	0.011
Q64	0.013	0.009	1.405	0.160
Q66	-0.099	0.014	-6.915	0.000
Q67	0.243	0.014	17.395	0.000
Q69	0.061	0.013	4.492	0.000
Q70	0.006	0.008	0.800	0.424
F7 BY				
Q3	-0.024	0.017	-1.406	0.160
Q4	-0.057	0.016	-3.465	0.001
Q9	0.082	0.016	4.974	0.000
Q11	0.311	0.025	12.370	0.000
Q14	0.222	0.021	10.390	0.000
Q15	-0.003	0.011	-0.260	0.795
Q16	0.072	0.016	4.408	0.000
Q17	0.005	0.008	0.569	0.569
Q18	0.019	0.009	2.102	0.036
Q19	0.150	0.031	4.876	0.000
Q20	0.008	0.023	0.338	0.735
Q21	-0.070	0.013	-5.380	0.000
Q22	-0.053	0.018	-2.945	0.003
Q23	-0.118	0.015	-7.719	0.000
Q24	-0.027	0.012	-2.295	0.022
Q25	-0.009	0.011	-0.829	0.407
Q26	0.582	0.022	26.725	0.000
Q27	0.066	0.013	5.164	0.000
Q28	0.107	0.015	7.383	0.000
Q29	0.788	0.022	35.127	0.000
Q30	0.038	0.012	3.142	0.002
Q31	0.175	0.016	11.251	0.000
Q32	0.028	0.011	2.505	0.012
Q33	0.024	0.012	1.920	0.055
Q34	-0.027	0.009	-2.831	0.005
Q35	-0.064	0.012	-5.287	0.000
Q36	0.006	0.010	0.539	0.590
Q37	0.022	0.011	1.905	0.057
Q39	0.054	0.015	3.512	0.000
Q40	0.005	0.013	0.403	0.687
Q42	0.044	0.011	4.115	0.000
Q43	0.020	0.008	2.358	0.018
Q44	0.033	0.010	3.359	0.001
Q45	-0.033	0.010	-3.153	0.002
Q47	0.024	0.015	1.653	0.098
Q48	-0.019	0.012	-1.617	0.106
Q49	0.033	0.011	3.169	0.002
Q50	0.014	0.009	1.538	0.124
Q51	0.063	0.016	3.972	0.000

Q54	-0.065	0.016	-4.028	0.000
Q55	0.039	0.010	3.776	0.000
Q56	-0.011	0.010	-1.115	0.265
Q57	0.037	0.009	4.016	0.000
Q58	-0.011	0.011	-1.015	0.310
Q59	-0.007	0.010	-0.686	0.492
Q60	0.010	0.010	0.951	0.342
Q61	-0.043	0.011	-3.894	0.000
Q62	0.018	0.014	1.326	0.185
Q63	0.021	0.010	2.161	0.031
Q64	0.008	0.011	0.755	0.450
Q66	-0.012	0.015	-0.771	0.441
Q67	0.092	0.013	7.109	0.000
Q69	0.019	0.011	1.724	0.085
Q70	0.048	0.013	3.673	0.000
F8 BY				
Q3	0.098	0.023	4.187	0.000
Q4	0.356	0.029	12.158	0.000
Q9	0.066	0.019	3.527	0.000
Q11	0.246	0.026	9.534	0.000
Q14	0.250	0.022	11.575	0.000
Q15	0.386	0.034	11.505	0.000
Q16	0.385	0.033	11.746	0.000
Q17	-0.019	0.009	-2.150	0.032
Q18	0.022	0.010	2.224	0.026
Q19	0.234	0.037	6.242	0.000
Q20	0.063	0.025	2.498	0.012
Q21	0.067	0.015	4.351	0.000
Q22	0.099	0.021	4.622	0.000
Q23	0.024	0.015	1.611	0.107
Q24	-0.048	0.016	-3.092	0.002
Q25	0.011	0.017	0.627	0.531
Q26	0.044	0.014	3.166	0.002
Q27	-0.094	0.014	-6.507	0.000
Q28	-0.029	0.012	-2.422	0.015
Q29	-0.022	0.008	-2.806	0.005
Q30	0.054	0.016	3.363	0.001
Q31	0.020	0.015	1.358	0.174
Q32	0.018	0.015	1.185	0.236
Q33	-0.005	0.017	-0.317	0.751
Q34	0.022	0.010	2.265	0.023
Q35	0.106	0.016	6.578	0.000
Q36	0.576	0.033	17.409	0.000
Q37	0.680	0.038	17.805	0.000
Q39	0.064	0.017	3.702	0.000
Q40	0.026	0.014	1.801	0.072
Q42	0.033	0.011	2.931	0.003
Q43	0.029	0.010	3.047	0.002
Q44	-0.008	0.010	-0.773	0.439
Q45	0.004	0.011	0.399	0.690
Q47	0.065	0.017	3.749	0.000
Q48	-0.008	0.013	-0.565	0.572
Q49	0.000	0.009	0.040	0.968
Q50	0.055	0.014	3.969	0.000
Q51	0.040	0.017	2.409	0.016
Q54	-0.043	0.016	-2.618	0.009

Q55	0.065	0.014	4.707	0.000
Q56	-0.027	0.011	-2.514	0.012
Q57	0.077	0.014	5.651	0.000
Q58	0.049	0.013	3.750	0.000
Q59	-0.118	0.015	-7.693	0.000
Q60	0.007	0.011	0.699	0.485
Q61	0.004	0.010	0.342	0.733
Q62	-0.098	0.017	-5.848	0.000
Q63	0.051	0.014	3.753	0.000
Q64	-0.005	0.011	-0.416	0.677
Q66	0.041	0.018	2.278	0.023
Q67	0.078	0.016	4.795	0.000
Q69	0.006	0.012	0.531	0.595
Q70	0.017	0.011	1.478	0.139
F9 BY				
Q3	-0.060	0.015	-4.042	0.000
Q4	-0.074	0.014	-5.225	0.000
Q9	0.036	0.013	2.821	0.005
Q11	0.058	0.017	3.340	0.001
Q14	0.042	0.014	2.959	0.003
Q15	-0.049	0.013	-3.882	0.000
Q16	-0.012	0.011	-1.132	0.258
Q17	0.025	0.008	2.950	0.003
Q18	0.009	0.008	1.141	0.254
Q19	-0.020	0.022	-0.905	0.365
Q20	0.029	0.019	1.467	0.142
Q21	-0.024	0.010	-2.340	0.019
Q22	0.010	0.014	0.733	0.464
Q23	0.102	0.013	7.975	0.000
Q24	-0.005	0.010	-0.534	0.594
Q25	0.104	0.013	8.276	0.000
Q26	0.016	0.008	1.919	0.055
Q27	-0.046	0.010	-4.566	0.000
Q28	-0.011	0.009	-1.292	0.196
Q29	-0.038	0.009	-4.252	0.000
Q30	-0.079	0.011	-6.901	0.000
Q31	0.039	0.010	3.831	0.000
Q32	-0.020	0.009	-2.120	0.034
Q33	0.072	0.012	5.777	0.000
Q34	0.048	0.010	4.819	0.000
Q35	0.075	0.011	6.641	0.000
Q36	0.034	0.010	3.546	0.000
Q37	0.062	0.013	4.667	0.000
Q39	0.126	0.013	9.511	0.000
Q40	0.070	0.011	6.170	0.000
Q42	0.013	0.008	1.560	0.119
Q43	-0.003	0.007	-0.451	0.652
Q44	-0.005	0.008	-0.567	0.571
Q45	-0.038	0.010	-3.950	0.000
Q47	0.048	0.013	3.738	0.000
Q48	0.024	0.010	2.378	0.017
Q49	-0.003	0.008	-0.411	0.681
Q50	-0.023	0.008	-2.843	0.004
Q51	0.011	0.012	0.894	0.371
Q54	0.141	0.017	8.519	0.000
Q55	-0.049	0.008	-5.785	0.000

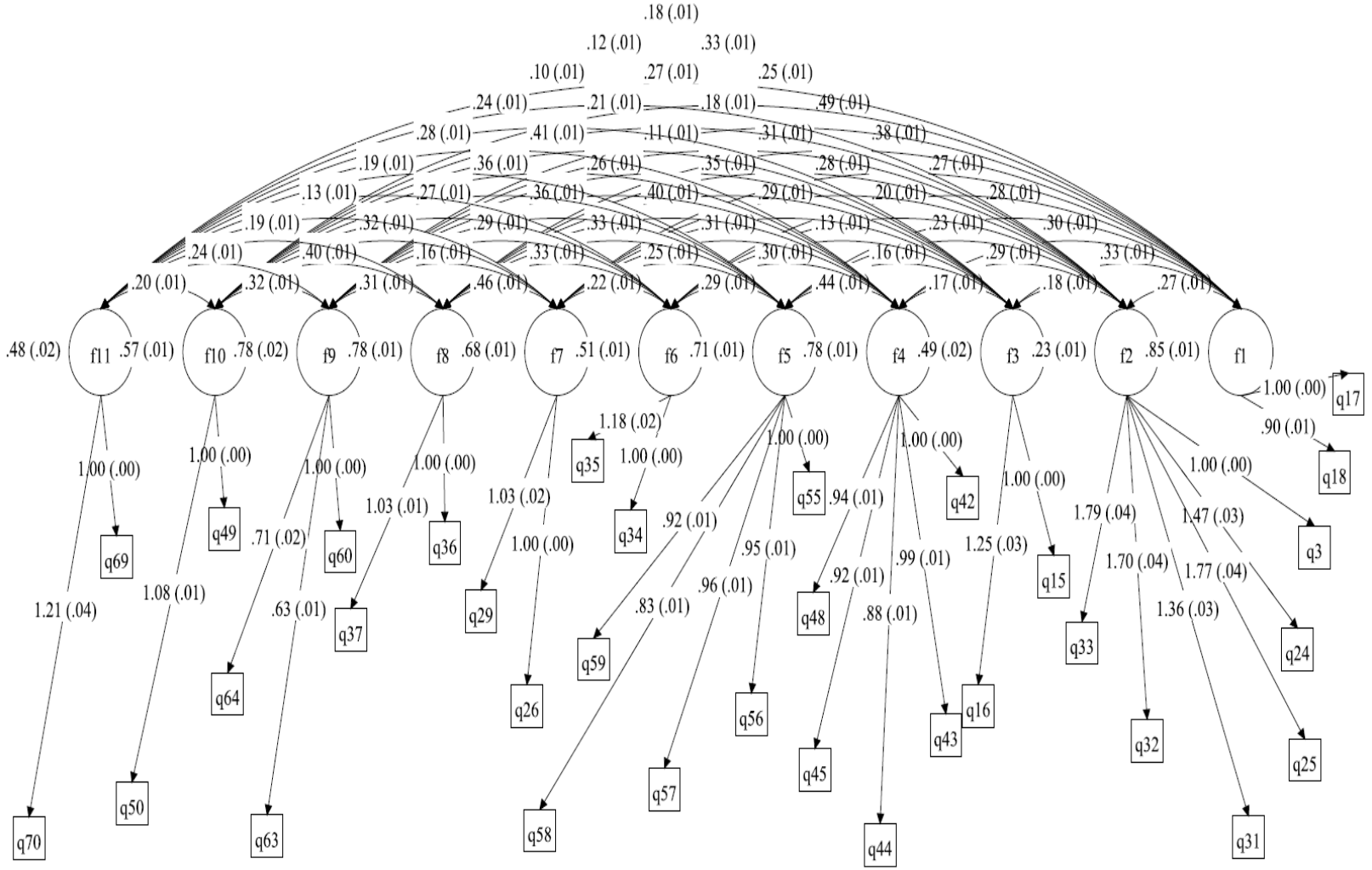
Q56	0.029	0.009	3.175	0.001
Q57	-0.029	0.007	-4.386	0.000
Q58	0.039	0.010	4.069	0.000
Q59	0.135	0.014	9.650	0.000
Q60	0.535	0.012	44.906	0.000
Q61	0.438	0.013	33.686	0.000
Q62	0.153	0.016	9.672	0.000
Q63	0.783	0.012	66.771	0.000
Q64	0.653	0.012	53.260	0.000
Q66	0.045	0.013	3.456	0.001
Q67	0.047	0.011	4.450	0.000
Q69	-0.037	0.012	-3.186	0.001
Q70	0.006	0.009	0.708	0.479
F10 BY				
Q3	-0.031	0.017	-1.863	0.063
Q4	0.005	0.016	0.309	0.758
Q9	0.010	0.015	0.698	0.485
Q11	0.037	0.019	1.966	0.049
Q14	0.063	0.018	3.485	0.000
Q15	-0.011	0.011	-1.051	0.293
Q16	-0.086	0.018	-4.647	0.000
Q17	0.018	0.009	1.914	0.056
Q18	0.025	0.010	2.639	0.008
Q19	-0.045	0.029	-1.527	0.127
Q20	0.024	0.025	0.961	0.337
Q21	0.010	0.012	0.816	0.414
Q22	-0.022	0.017	-1.334	0.182
Q23	0.002	0.014	0.156	0.876
Q24	-0.007	0.015	-0.450	0.653
Q25	-0.042	0.016	-2.603	0.009
Q26	0.020	0.012	1.716	0.086
Q27	-0.005	0.011	-0.444	0.657
Q28	-0.018	0.011	-1.697	0.090
Q29	-0.005	0.008	-0.576	0.564
Q30	0.035	0.013	2.709	0.007
Q31	0.020	0.013	1.530	0.126
Q32	0.170	0.020	8.328	0.000
Q33	0.047	0.018	2.555	0.011
Q34	0.081	0.015	5.374	0.000
Q35	0.080	0.015	5.332	0.000
Q36	0.071	0.016	4.429	0.000
Q37	0.039	0.016	2.522	0.012
Q39	-0.085	0.015	-5.627	0.000
Q40	-0.044	0.013	-3.261	0.001
Q42	-0.015	0.009	-1.604	0.109
Q43	-0.011	0.008	-1.387	0.165
Q44	-0.010	0.009	-1.102	0.271
Q45	0.092	0.013	6.974	0.000
Q47	-0.037	0.015	-2.553	0.011
Q48	0.103	0.016	6.471	0.000
Q49	0.595	0.032	18.587	0.000
Q50	0.576	0.031	18.522	0.000
Q51	0.204	0.019	10.817	0.000
Q54	0.014	0.014	1.000	0.317
Q55	-0.002	0.009	-0.284	0.776
Q56	0.070	0.014	5.069	0.000

Q57	-0.034	0.009	-3.609	0.000
Q58	-0.012	0.011	-1.092	0.275
Q59	0.110	0.018	6.207	0.000
Q60	0.298	0.019	15.618	0.000
Q61	0.259	0.019	13.617	0.000
Q62	0.116	0.019	6.115	0.000
Q63	-0.021	0.008	-2.579	0.010
Q64	-0.050	0.013	-3.791	0.000
Q66	0.049	0.016	3.030	0.002
Q67	0.046	0.014	3.316	0.001
Q69	-0.035	0.012	-2.831	0.005
Q70	0.001	0.009	0.113	0.910
F11 BY				
Q3	0.017	0.016	1.048	0.295
Q4	-0.013	0.015	-0.828	0.408
Q9	0.048	0.015	3.238	0.001
Q11	0.025	0.018	1.390	0.165
Q14	0.015	0.016	0.914	0.361
Q15	-0.016	0.011	-1.351	0.177
Q16	0.008	0.012	0.703	0.482
Q17	-0.001	0.008	-0.160	0.873
Q18	-0.006	0.009	-0.623	0.533
Q19	-0.091	0.029	-3.125	0.002
Q20	0.095	0.026	3.674	0.000
Q21	0.160	0.014	11.864	0.000
Q22	0.159	0.019	8.166	0.000
Q23	0.027	0.015	1.836	0.066
Q24	0.023	0.012	1.860	0.063
Q25	-0.021	0.012	-1.733	0.083
Q26	0.037	0.011	3.199	0.001
Q27	0.008	0.011	0.679	0.497
Q28	-0.030	0.012	-2.628	0.009
Q29	0.008	0.009	0.969	0.332
Q30	0.020	0.012	1.667	0.095
Q31	-0.053	0.013	-4.182	0.000
Q32	0.017	0.011	1.545	0.122
Q33	0.020	0.013	1.528	0.127
Q34	0.074	0.013	5.788	0.000
Q35	0.018	0.009	1.961	0.050
Q36	0.014	0.012	1.245	0.213
Q37	0.027	0.014	1.979	0.048
Q39	-0.001	0.014	-0.053	0.958
Q40	0.045	0.013	3.366	0.001
Q42	-0.005	0.009	-0.501	0.616
Q43	-0.037	0.009	-4.017	0.000
Q44	-0.036	0.010	-3.613	0.000
Q45	0.045	0.011	4.049	0.000
Q47	0.056	0.015	3.800	0.000
Q48	0.059	0.013	4.585	0.000
Q49	-0.014	0.009	-1.538	0.124
Q50	0.002	0.009	0.195	0.845
Q51	0.002	0.014	0.165	0.869
Q54	0.139	0.018	7.861	0.000
Q55	-0.047	0.009	-4.928	0.000
Q56	0.046	0.011	4.299	0.000
Q57	-0.061	0.010	-6.248	0.000

	Q58	0.027	0.010	2.750	0.006
	Q59	0.114	0.014	8.037	0.000
	Q60	0.006	0.010	0.608	0.543
	Q61	0.008	0.010	0.811	0.417
	Q62	0.212	0.017	12.405	0.000
	Q63	-0.012	0.010	-1.230	0.219
	Q64	0.005	0.011	0.401	0.688
	Q66	0.154	0.015	10.186	0.000
	Q67	0.024	0.012	2.000	0.046
	Q69	0.710	0.023	30.512	0.000
	Q70	0.784	0.023	33.493	0.000
F2	WITH				
	F1	0.519	0.013	38.801	0.000
F3	WITH				
	F1	0.360	0.025	14.632	0.000
	F2	0.379	0.036	10.428	0.000
F4	WITH				
	F1	0.208	0.015	14.071	0.000
	F2	0.493	0.016	30.119	0.000
	F3	0.070	0.027	2.644	0.008
F5	WITH				
	F1	0.266	0.012	21.426	0.000
	F2	0.424	0.013	33.370	0.000
	F3	0.137	0.019	7.125	0.000
	F4	0.474	0.010	49.607	0.000
F6	WITH				
	F1	0.360	0.013	26.737	0.000
	F2	0.461	0.021	21.752	0.000
	F3	0.157	0.025	6.189	0.000
	F4	0.307	0.014	22.180	0.000
	F5	0.380	0.012	31.174	0.000
F7	WITH				
	F1	0.364	0.018	20.260	0.000
	F2	0.416	0.019	22.203	0.000
	F3	0.269	0.022	12.042	0.000
	F4	0.152	0.018	8.623	0.000
	F5	0.111	0.016	6.733	0.000
	F6	0.143	0.019	7.643	0.000
F8	WITH				
	F1	0.464	0.037	12.418	0.000
	F2	0.389	0.070	5.533	0.000
	F3	0.090	0.021	4.362	0.000
	F4	0.236	0.042	5.679	0.000
	F5	0.174	0.032	5.425	0.000
	F6	0.256	0.027	9.324	0.000
	F7	0.340	0.039	8.699	0.000
F9	WITH				
	F1	0.217	0.014	15.217	0.000
	F2	0.284	0.024	12.040	0.000

	F3	0.090	0.019	4.729	0.000
	F4	0.201	0.013	15.344	0.000
	F5	0.364	0.012	30.803	0.000
	F6	0.297	0.015	19.471	0.000
	F7	0.025	0.016	1.585	0.113
	F8	0.177	0.019	9.332	0.000
F10	WITH				
	F1	0.270	0.048	5.662	0.000
	F2	0.352	0.078	4.496	0.000
	F3	0.230	0.025	9.241	0.000
	F4	0.360	0.041	8.691	0.000
	F5	0.385	0.032	11.878	0.000
	F6	0.245	0.039	6.339	0.000
	F7	0.156	0.044	3.562	0.000
	F8	0.192	0.060	3.218	0.001
	F9	0.228	0.021	11.083	0.000
F11	WITH				
	F1	0.235	0.015	15.694	0.000
	F2	0.312	0.014	22.109	0.000
	F3	0.163	0.016	9.913	0.000
	F4	0.348	0.012	28.777	0.000
	F5	0.420	0.011	36.841	0.000
	F6	0.271	0.015	18.099	0.000
	F7	0.051	0.018	2.833	0.005
	F8	0.128	0.027	4.803	0.000
	F9	0.350	0.012	28.489	0.000
	F10	0.315	0.025	12.628	0.000

Appendix 16: CFA model for patients with an operation or procedure and emergency admission



Appendix 17: CFA model for patients with an operation or procedure and planned admission