The development and validation of a patient-reported outcome measure of health-related quality of life for patients with pressure ulcers: PUQOL Project

Claudia Anna Gorecki

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds, School of Medicine & Health

September 2011

Intellectual Property and Publication Statements

The candidate confirms that the work submitted is her own, except where work which formed part of jointly-authored publications has been included. The contribution of the candidate and the authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

Part of this research has been carried out by a project team which has included: Professors Jane Nixon (JN), Julia Brown (JB), Andrea Nelson (EAN), Drs Michelle Briggs (MB), Carol Dealey (CD), Mrs. Susanne Coleman (SC), Mrs. Elizabeth McGinnis (EM), Mrs. Lyn Wilson (LW), Mrs. Nikki Stubbs (NS), and a health outcomes methodology group, including: Professor Donna Lamping (DL) and Drs Stefan Cano (SCa), Sarah Smith (SS), Katerina Hilari (KH), Sara Schroter (SSc), Yasmene Alavi (YA), and Jennifer Petrillo (JP). My own contributions, fully and explicitly indicated in the thesis, have been summarised by chapter below. The other members of the groups and their contributions are also summarised.

Chapter 2 - Published: Gorecki C, Brown J, Briggs M, Nelson EA, Dealey C, Schoonhoven L, Defloor T, Nixon J. (2009) Impact of Pressure Ulcers on Quality of Life in Elderly patients: A systematic review. Journal of the American Geriatrics Society, 57:1175-1183.
Gorecki C, Brown J, Briggs M, Nixon J. (2010) The sensitivity and specificity of 5 search strategies in identifying patient-reported quality of life data. Journal of Advanced Nursing, 66(3):645-652.

Contributions: Claudia Gorecki (CG) led protocol development and implementation, searches, data extraction and analysis, results interpretation and the writing of the manuscript. JB contributed to the conception and protocol development, acted as third reviewer, interpreted results and manuscript writing. MB contributed to the analysis methodology, results interpretation and manuscript writing. EAN contributed to protocol development, systematic review methodology, results interpretation and manuscript writing. Lisette Schoonhoven contributed to conception and protocol development, applied inclusion criteria to studies published in Dutch, results interpretation and manuscript writing. CD and Tom Defloor contributed to protocol development and manuscript writing. JN contributed to the conception and design of the project and protocol development, applied inclusion criteria, contributed to results interpretation, manuscript writing, project supervision and conduct.

Acknowledgements: The systematic review was part of a wider collaboration from the European Quality of Life Pressure Ulcer Project (EQUIP) Collaborative Group. Christina Lindholm, as a member of the EQUIP collaborative group, was invited to review and comment on the protocol, data synthesis and final report. Joan-Enric Torra I Bou, Smith & Nephew, Barcelona, Spain (Spanish); Katrien Vanderwee, University of Gent, Belgium (Dutch); Ruud Halfens, University of Maastricht, the Netherlands (German);Marco Romanelli, University of Pisa, Italy (Italian); Yves Passadori, Service de N.R.R, France (French); Hiromi Sanada, University of Tokyo, Japan (Japanese); and Christina Lindholm, Hogskolan Kristianstad University, Sweden (Danish) reviewed papers published in languages other than English.

Chapter 3 - Published: Gorecki C, Lamping DL, Brown J, Madill A, Firth J, Nixon J. (2010) Development of a conceptual framework of health-related quality of life in pressure ulcers: a patient-focused approach. International Journal of Nursing Studies, 66(3):645-652.

Contributions: CG led protocol development and implementation, data collection and analysis, interpretation of results and the writing of the manuscript. DL and JB contributed to study conception, protocol development, interpretation of results and manuscript writing. Anna Madill contributed to the analysis design, interpretation of results and manuscript writing. Jill Firth acted as independent researcher, contributed to data analysis, results interpretation and manuscript writing. JN contributed to study conception, protocol development, acted as third researcher resolving discrepancies, interpretation of results and manuscript writing.

Acknowledgements: EAN undertook quality assessment of the first few transcripts. EAN, EM, MB, CD, LW, NS, SC and DL reviewed the conceptual framework domains and final components. SCa, SSc, JP, YA and DL reviewed the final conceptual framework. The following colleagues identified eligible patients: LW, EM, NS, Margaret Fox, Mike Tudor, Richard Buckland, Valerie Henderson, Jeannie Donnelly, Jenny Stephens, Donna Chaloner, Vicky Gilchrist, Jacky Dawson and Nicky Whitfield).

Chapter 4 - Publication: Gorecki C, Lamping D, Alvari Y, Brown J, Nixon J (under review) A review of patient-reported outcome measures used in pressure-ulcer research. Journal of Rehabilitation Psychology.

Contributions: CG led protocol development and implementation, data extraction and analysis, interpretation of results and the writing of the manuscript. DL contributed to protocol development, psychometric evaluation, interpretation of results and

manuscript writing. YA acted as independent researcher, contributed to data extraction and analysis, interpretation of results and manuscript writing. JN contributed to protocol development, acted as third researcher resolving discrepancies, interpretation of results and manuscript writing. JB contributed to protocol development, interpretation of results and manuscript writing. DL, YA, SCa SSc and Kritika Samsi reviewed and commented on the search strategies.

Acknowledgements: Dr Sylvia Feyler (Leeds, UK) translated the Freiburger Questionnaire of Quality of Life into English.

Chapter 5

Published: Gorecki C, Lamping D, Nixon J, Brown J, Cano S (In Press) Applying a mixed methods in scale development: evaluation of health-related quality of life items in the Pressure ulcer Quality of Life (PU-QOL) measure. QOL Research Journal: DOI 10.1007/s11136-011-9980-x

Contributions: CG led protocol development and implementation, data collection, extraction and analysis, interpretation of results and the writing of the manuscript. DL contributed to protocol development, psychometric evaluation, interpretation of results and manuscript writing. JN and JB contributed to protocol development, interpretation of results and manuscript writing. SCa contributed to the design of the study, protocol development, psychometric evaluation, results interpretation, manuscript writing and study supervision. The project team (LW, NS, EM, JN, EAN, CD, SC) reviewed the list of items (item pool). The Health outcomes methodology group (DL, SCa, SSc, SS, KH, YA, JP) rreviewed the amended item pool and content mapping.

Presentations

National Conference

Invited: Gorecki C. Measuring QOL in pressure ulcers. Conference of the Tissue Viability Society, UK, April 2011 (oral)

Gorecki C, Closs J, Nixon J, Briggs M. Patient-reported pressure ulcer-associated pain: A mixed-methods systematic review. British Pain Society Annual Conference, UK, April 2010 (poster)

Invited: Gorecki C, Brown J, Lamping D, Nelson EA, Briggs M, Dealey C, Schoonhoven L, Nixon J. Pressure Ulcers and Quality of Life: Systematic review and preliminary results from a qualitative study. Tissue Viability Annual Conference, UK, April 2008

International Conference

Gorecki C, on behalf of the PUQOL Project Team. Final version of PU-QOL: a patient-reported outcome measure of health-related quality of life for patients with pressure ulcers. European Pressure Ulcer Advisory Panel (EPUAP) meeting, Porto, Portugal, September 2011 (Oral)

Invited: Gorecki C. Application of mixed methods in early rating scale development. International Rasch Expert Group Meeting, June 2011, Dubrovnik, Croatia (Oral)

Symposium: Gorecki C, Brown J, Lamping D, Nixon J; on behalf of the PUQOL Project Group. Development of a patient-reported outcome measure of quality of life for use with patients with pressure ulcers. RCN International Nursing Research Conference, UK, May 2011 (Oral)

Gorecki C, Lamping DL, Nixon J, Brown J, Cano S. The benefits of mixed methods in scale development I: The added value of Rasch analysis in pre-testing. The 17th Annual Conference of the International Society for Quality of Life Research (ISOQOL), London, UK, October 2010 (Poster)

Gorecki C, Nixon J, Lamping DL, Brown J, Cano S. The benefits of mixed methods in scale development II: Selecting optimal mode of administration. The 17th Annual Conference of the International Society for Quality of Life Research (ISOQOL), London, UK, October 2010 (Oral)

Invited: Gorecki C. EPUAP Novice Award Lecture 2010: What constitutes health-related quality of life in pressure ulcers and how do we measure it? European Pressure Ulcer Advisory Panel (EPUAP) meeting, Birmingham, UK, September 2010 (Oral)

Gorecki C, Brown J, Lamping D, Madill A, Firth J, Nixon J, On behalf of the PUQOL project group. Health-related quality of life in pressure ulceration: Development of a conceptual framework. The 16th Annual meeting of the International Society for Quality of Life Research (ISOQOL), New Orleans, USA, October 2009 (Oral)

Gorecki C, Brown J, Lamping D & Nixon J. Using cognitive interviewing to improve a newly developed health-related quality of life patient-reported outcome for people with pressure ulcers. The 16th Annual meeting of the International Society for Quality of Life Research (ISOQOL), New Orleans, USA, October 2009 (Poster)

Gorecki C, Brown J, Lamping D, Nixon J; on behalf of the PUQOL (Pressure Ulcer Quality of Life) Project Group. Development of a patient-reported outcome measure of quality of life for use with patients with pressure ulcers, European Pressure Ulcer Advisory Panel (EPUAP), The Netherlands, September 2009 (Oral)

Gorecki C, Lamping D, Brown J, Alavi Y, Nixon J. A systematic review of patient-reported outcome measures used to measure quality of life in patients with pressure

ulcers. European Pressure Ulcer Advisory Panel (EPUAP), The Netherlands, September 2009 (Oral)

Gorecki C, Brown J, Nelson EA, Briggs M, Dealey C, Schoonhoven L, Nixon J. A systematic review of Pressure Ulcers and Quality of Life. International Society of Quality of Life (ISOQOL) Annual Conference, Toronto, Canada, October 2007 (Poster)

Gorecki C, Brown J, Briggs M, Nixon, J. Evaluation of 5 search strategies to locate subjective patient-reported HRQL data. International Society of Quality of Life (ISOQOL) Annual Conference, Toronto, Canada, October 2007 (Poster)

Gorecki C, Brown J, Nelson EA, Briggs M, Dealey C, Schoonhoven L, Nixon J. A systematic review of Pressure Ulcers and Quality of Life. European Pressure Ulcer Advisory Panel (EPUAP) Meeting, UK, August 2007 (Oral)

Acknowledgements

This work was made possible as part of a NIHR Programme of Research Grant and I want to acknowledge and express my gratitude for the financial support.

I would like to thank my supervisors, Professors Jane Nixon, Julia Brown and Donna Lamping, and Dr Stefan Cano for their encouragement and guidance during my research and preparation of the thesis. I am grateful to them for believing in me, for motivating and supporting me, and for providing much valued personal and professional advice and a nurturing working environment. I am particularly grateful to Jane for her expertise in the clinical area, Donna for providing me with a good foundation in health outcomes measurement, Julia for her expert advice on the design and conduct of the research, and Stefan for his expert advice on methodology and psychometrics. I will never be able to express my gratitude and respect.

Throughout the period of my research, I was fortunate to have the following people on my project team: Professor Andrea Nelson; Dr Michelle Briggs; Dr Carol Dealey; Mrs. Susanne Coleman; Mrs. Elizabeth McGinnis; Mrs. Lyn Wilson; and Mrs. Nikki Stubbs. I would like to express my appreciation to all the members for providing both practical and expert advice. I would also like to acknowledge the emotional support and professional help from the Outcomes Methodology Group: Dr Sarah Smith, Dr Yasmene Alavi, Dr Sara Schroter, Dr Katerina Hilari and Dr Jennifer Pertrillo.

I am indebted to all the clinical research team members from the 34 hospitals and community care services that participated in this research and made it a part of their busy daily schedules. Without their help recruiting patients and collecting data this research would not have been possible.

I wish to express my gratitude to all the patients who participated in this research, for giving up their time and for sharing their personal thoughts and experiences. I hope that their efforts will contribute to changes in wound care and consequently improvements in the quality of life of future people with pressure ulcers.

At a personal level, I would like to acknowledge the patience, advice and emotional support of my family, and closed friends Carl Palay and Maria Crosby. Thank you for your continued encouragement and support of my passion for knowledge and pursuit of my personal and professional goals.

To my mother, Carl and Maria for all your sustained love, tolerance and belief in me over the years, and in loving memory of my teacher and mentor, Professor Donna Lamping, I dedicate this work.

Abstract

This thesis aimed to establish the impact of pressure ulcers (PUs) on health-related quality of life (HRQL) and determine the need for a PU-specific patient-reported outcome (PRO) instrument. The thesis comprises literature reviews, development of a conceptual framework and development and evaluation of a PU-specific PRO (PU-QOLI) for future research and clinical practice. Methodological developments were woven into the research.

This thesis demonstrates that PUs impact HRQL and there is need for standardised methods for assessing health outcomes important in PUs. The PU-QOLI was developed using qualitative and quantitative psychometric methods and meets international standards for rigorous measurement. Individual scales can be selected from 10 outcomes of PU-symptoms, physical functioning, psychological well-being, and social participation, intended for interview-administration. It is appropriate for use in adults with any type of PU and suitable for all UK healthcare settings.

This research makes important contributions to the PU and wider health measurement fields. The findings demonstrate that mixed methods, including Rasch measurement, were suitable for developing a PRO instrument, the PU-QOLI provides a means for the comprehensive assessment of PU impact and for quantifying the benefits of PU interventions from the patient's perspective, thus far lacking in the area, and the use of PROs will provide evidence-based information to allow health authorities to select the most effective healthcare for patients and to audit and monitor the quality of care given in the PU field.

This work was the first step towards establishing measurement precision. Further research is needed to improve some of the measurement properties of PU-QOLI scales, assess responsiveness, confirm study findings in an independent sample, investigate feasibility of use in specific subgroups and economic evaluation, and develop proxy measures and language translations. Long-term goals include developing the PU-QOLI as a clinical tool intended for individual-patient decision making.

Table of Contents

List of Tables and Figures

Key to Abbreviations

CAT	Computer adaptive testing
CPC	Category probability curve
CRF	Case record forms
CTRU	Clinical Trials Research Unit
CTT	Classical Test Theory
CWIS	Cardiff Wound Impact Schedule
DIF	Differential item functioning
DoH	Department of Health
EBP	Evidence Based Practice
EORTC	European Organisation for Research & Treatment of Cancer
EPUAP	European Pressure Ulcer Advisory panel
EQ-5D	A measure of health outcome; developed by the EuroQol group
FDA	US Department of health & Human services Food & Drug Association
HRQL	Health-Related Quality of Life
ICC	Item characteristic curve
ITC	Item-Total Correlation
MOS	Medical Outcomes Study
NHS	National Health Service
NICE	National Institute for Health & Clinical Excellence
NIHR	National Institute of health research
NPUAP	National (US) Pressure Ulcer Advisory panel
PRO	Patient-Reported Outcome
PSI	Person Separation Index
PU	Pressure Ulcer
PU-QOL	Pressure Ulcer Quality of Life instrument
QoL	Quality of Life
RCT	Randomised Controlled Trial
SCI	Spinal Cord Injured
SD	Standard deviation
SEM	Standard error mean
SF-36	Short-Form health survey with 36 questions
TVN	Tissue Viability Nurse
TRT	Test Retest Reproducibility
TVT	Tissue Viability Team
VAS	Visual Analogue Scale
WHO	World Health Organisation

Chapter 1

BACKGROUND

1.1 Introduction

Chronic wounds are a major health problem and challenge to patients, healthcare professionals and the healthcare system, particularly as the age of the population increases. Pressure ulcers (PUs) are common chronic wounds that occur as localised injury to the skin and/or underlying tissue usually over a bony prominence, as a result of pressure, or pressure in combination with shear (1). PUs range in size and severity of tissue layer affected, with the majority occurring below the waist; particularly vulnerable areas are the sacrum, buttocks and heels (2). With widespread prevalence and incidence in all health settings (3), PUs, often a complication of serious acute or chronic illness, are a significant health problem associated with increased morbidity (4), mortality (5), healthcare costs and hospitalisation, and identified as a UK National Health Service (NHS) quality indicator and priority through Department of Health (DoH) policy (6, 7, 8).

PUs impact quality of life (QoL) and can severely compromise all areas of patient functioning (Chapter 2). Intensive interventions for preventing and treating PUs pose additional patient burden and further affect QoL (9). Therefore, evaluating outcomes such as QoL is particularly important and relevant in PUs where the condition and associated interventions pose patient burden.

This chapter provides a brief summary of key areas of the literature including the definition and classification of PUs, the theoretical basis for the value of patient-reported outcomes (PRO), measurement of PROs and methodological developments.

1.2 Pressure Ulcer Definition and Classification

PUs, also known as pressure or bed sores, decubitus ulcers or dermal ulcers, are a common type of impaired skin integrity that results in erythema and destruction of skin layers due to pressure alone or pressure in combination with shearing forces (1). Areas of necrosis and ulceration occur where skin or deeper soft tissues compress between bony prominences and hard surface. PUs usually develop from laying or sitting in one position for prolonged periods (10). For those immobile, PUs will most likely form on or around the heels, hip-bone and/or lower back/sacrum. However, unlike other chronic wounds where ulceration is constrained to the limbs (e.g. leg or diabetic foot ulcers), PUs can develop anywhere on the body (11).

PUs range in severity and various PU classifications have been used over the years (i.e. PU grade or stage; often used interchangably in the literature). To provide a standardised method of record-keeping and a common description of PU severity for the purposes of clinical practice, audit and research, a recent international collaboration agreed a PU classification system using the term *category* (1) (Table 1.1). This classification posits that PU severity ranges from non-blanchable skin erythema of intact skin (category 1) and superficial skin loss (category 2) to, in serious cases, tissue destruction involving skin, subcutaneous fat, muscle and bone damage (category 3/4) (1). For the purpose of this research, the classification has been adapted to enable classification of normal skin and unstageable PUs, and the term grade and stage will be used when referring to earlier studies, depending on classifications used.

Table 1.1 Pressure Ulcer Classification System

Category	Description
Category 1 Non-blanchable erythema of intact skin	Intact skin with non-blanchable erythema of a localised area usually over a bony prominence. Discolouration of the skin, warmth, oedema, hardness or pain may also be present. Darkly pigmented skin may not have visible blanching.
Category 2 Partial thickness skin loss or blister	Partial thickness loss of dermis presenting as a shallow open ulcer with a red pink wound bed, without slough. May also present as an intact or open/ruptured serum or sero-sanginous-filled blister.
Category 3 Full thickness skin loss	Full thickness tissue loss. Subcutaneous fat may be visible but bone, tendon or muscle not exposed. Some slough may be present. May include undermining and tunnelling.
Category 4 Full thickness tissue loss	Full thickness tissue loss with exposed bone, tendon or muscle. Slough or eschar may be present. Often includes undermining or tunnelling.
Category U Unstageable	Full thickness skin loss where actual depth of the ulcer is completely obscured by slough (yellow, tan, grey, green or brown) and/or eschar (tan, brown, black) in the wound bed.

The intial sign of pressure is skin redness (non-blanchable erythema); skin does not blanch (whiten) under light finger pressure. Whilst commonly classed as category or grade 1 PU, it is regarded in Europe as a precursor to PU development, as the erythema may resolve without skin destruction. However, grade 1 ulcers are also an early warning sign for healthcare providers and patients to apply adequate interventions to avoid further deterioration (12). Grade 1 PUs have been identified as a risk factor for PU development, with approximately one third of grade 1 PUs deteriorating to a grade 2 ulcer (13, 14). Grade 2 PUs are reportable as clinical incidents across the UK NHS and Grade 3/4 PUs represent serious clinical events, increasingly subjected to investigation for abuse, neglect and litigation (15).

To date, PU research has been encumbered by methodological issues (e.g. differences in data collection and study methodology) and variability in describing and reporting PUs (16). This may be due to the variability in classification systems. PUs have been previously classified using terms *grade* or *stage,* which implies a natural progression from 1 to 4, however not all PUs progress in this way. Further, identifying and assessing grade 1 or equivalent PUs is difficult, resulting in inflated prevalence rates in some health settings and lower/under-reported rates in others. Reports of inter-rater reliability have been high in detecting grade 2 PUs, but there is also high disagreement between detection and accurate PU severity classification when skin is assessed by multiple clinical staff (17, 18).

1.2.1 What are the risk factors and who is at risk?

The primary PU risk factor is immobility (19). The likelihood of PU development in immobile patients is increased by interactions between the intensity and duration of mechanical load and factors affecting skin tolerance, namely increased age, impaired circulation, nutritional deficiencies, skin perfusion (e.g. diabetes, vascular disease), incontinence, altered consciousness, and exposure to pressure and shear (5, 20, 21, 22, 23, 24, 25, 26, 27). These factors are common in patients with serious acute and chronic illness and high risk groups include elderly[1], medical, cardio/vascular surgical (13), orthopaedic (28), intensive care (29), end-stage terminally ill (30), long-term care and community care populations (e.g. spinal cord injured (SCI), brain injured, neuromuscular disease – multiple sclerosis, neurologically compromised – demented, who may be both insensate and immobile, spending extended time lying supine) (23).

Despite a number of contributing or confounding factors associated with PUs, available risk assessment scales yield conflicting results regarding their usefulness as predictors of PU development. Various scales exist for identification of people at increased risk of PUs (31, 32, 33, 34, 35, 36, 37) but there is currently no uniform characterisation of PU risk factors. Risk assessment scales were developed *ad hoc* based on opinions about the relative importance of risk factors rather than rigorous evidence (38). The most valid way of determining risk factors is through use of statistical regression models that weigh factors which best predict PU development (39) (i.e. accurately detect PU occurrence and link it to risk variables). Currently, no risk assessment scales have been developed in this way (40) and many patients are falsely identified as at risk or

[1] Age alone is not a risk factor for PUs but rather problems common to elderly people such as altered molecular and cellular characteristics of aged skin and various associated comorbidities.

not at risk (41), suggesting the scales lack sensitivity and specificity. Further, attempts to synthesise meaningful risk factor data were performed in sample sizes not permitting multivariate analyses (PU risk factor systematic review – unpublished). There is need for improved methods for accurately identifying those at risk and targeting appropriate interventions that prevent PU development and improve outcomes important to patients (i.e. QoL).

1.3 Epidemiology

PUs are widespread and often an underestimated health problem. They are common in hospital, nursing home facilities and homecare patients (42). Prevalence is defined as the number of patients with a PU in a specific population at a specific time, usually evaluated on a one-time, cross-sectional basis while incidence is the number of patients that develop a new PU after, for example, hospital admission (43). Prevalence rates for PUs in several countries report rates between 5.3% and 69% in acute care settings (28, 44, 45, 46, 47, 48); 2.6% and 25.9% among nursing home residents (46, 49, 50); and 0.31% and 6.3% in home care patients (42, 51). Approximately one in 10 hospital and one in 20 community patients are affected by PUs (3). In the UK it has been estimated that between four and 10% of all patients admitted to hospital will develop at least one PU during their hospital stay (52).

1.4 Economic impact

Due to high prevalence, the overall prevention and management of PUs is both a national and international healthcare issue. Despite advances in medicine, surgery, nursing care and self-care education, some PUs are considered unavoidable and remain a major cause of morbidity and mortaility. In addition to the cost of human suffering, substantial costs to the healthcare system demonstrate the importance of preventing PUs and providing cost-effective PU management. The cost of PUs is estimated at £1.4-2.1 billion annually; equivalent to 4% of total NHS expenditure (6). Costs arise from increased length of hospital stay/community nursing care; treatments; complications such as serious infection (53); risk assessment of all acute and community nursing patients on admission; and litigation, which is predicted to increase due to both general societal trends (aging population) and proposed changes in the law leading to investigation of severe PUs by government agencies to detect institutional and professional neglect of vulnerable adults (15).

1.5 Prevention and treatment

Due to the complexity of PUs, healthcare providers face the challenge of providing effective preventative and treatment interventions for PUs. The choice of intervention depends on the purpose, for example, pressure damage prevention, skin protection from moisture or removing necrotic tissue to promote healing. NHS practice guidance is focused upon identifying patients at risk through risk assessment of all patients on admission to acute hospitals and community nursing services, implementing preventative care (e.g. specialist mattresses, turning, skin care) and using interventions to halt damage and promote healing (e.g. mattresses, dressings, nutritional supplements) (1, 19, 54, 55, 56).

Preventative interventions include: positioning and support surfaces (e.g. mattresses, cushions, heel elevation, turning) and nutritional supplements (56, 57). PU treatment can be divided into non-operative, including: positioning and support surfaces; nutritional supplements; topical treatments (e.g. dressings, Maggot Debridement Therapy (MDT), negative pressure wound therapy – Vacuum Assisted Closure (VAC)); adjuvant therapies; and hospitalisation, and operative treatments (e.g. plastic surgery involving flap repair) (1, 19, 55, 58, 59, 60, 61):

- *Positioning and support surfaces*

Repositioning and support surfaces are a widely used modality in the prevention and treatment of PUs. They are intended for patient positioning to avoid compression of soft tissues against bony prominences or hard surfaces and to reduce pressure (e.g. pressure-relieving mattress/cushion to manage tissue load or distribute weight) (1, 56).

- *Nutrition*

Adequate protein, carbohydrates, vitamins and minerals are considered important for wound healing (1), specifically the growth of granulation tissue, although the effectiveness of nutrition has not been demonstrated in randomised controlled trials (RCTs) (62). Being under or over–weight have been associated with PU development (63).

- *Topical treatments*

Topical treatments, mainly dressings, are intended to manage odour, exudate and pain, promote wound healing, protect the skin from further damage due to incontinence (e.g. ointments, creams), and for bacterial and infection management. MDT, also known as maggot therapy, is a topical treatment for hard to heal wounds where sterile fly larvae are applied to the wound for two or three days within special dressings

to keep them from migrating. Other topical treatments involve wound management to accelerate endogenous healing or facilitate the effectiveness of other therapeutic methods, such as debridement to remove necrotic tissue, excessive bacteria and/or dead cells (e.g. VAC therapy); wound cleansing to remove loose impediments to wound healing; and moisture balance to facilitate granulation and reepithelialisation (1, 60, 61).

- *Surgery*

Surgical treatment of PUs can be for preparing the wound for healing (e.g. removal of prominences without excessive incision to alleviate pressure due to compression of soft tissue between the skeleton/support surfaces) or definitive wound closure (e.g. direct closure, skin grafts, skin and musculocutaneous flaps) (1).

- *Adjuvant therapies*

Electrical stimulation (e.g. electrotherapy), hyperbaric oxygen therapy, and negative pressure intended for exudate and debris removal, increasing wound perfusion, and formation of granulation tissue, are examples of adjuvant therapies for PUs. Few studies have assessed the efficacy of these methods and therefore they tend to be used when healing with conventional methods fails (1).

In the PU field there is limited evidence of treatment effectiveness because of the poor quality of RCTs conducted (56, 60, 61), a lack of standard outcome measures and the use of clinician- rather than patient-reported outcome measures (see section 1.13). However, interventions are still required and used for PU prevention and treatment but they can cause patients considerable pain and discomfort, affect health, functioning and QoL, and cause significant treatment burden (9, 64). As PUs are painful (65), malodorous (66), and can exudate profusely (67), especially during the early inflammation phase, frequent dressing changes may be required. The intensity of PU-symptoms and the severity and duration of PUs and of receiving PU interventions contribute to treatment burden. One study found that for hospice patients, comfort superseded PU prevention and wound care, particularly if patients were dying or had conditions causing them to lie in a single position (64). Therefore not only PU prevention and effective PU management, but assessment of PU interventions against outcomes important to patients (i.e. treatment burden, QoL) is needed. For example, even though we can measure the level of PU exudate or odour, the experience of these symptoms and the impact they may or may not have is subjective and individual. Furthermore, due to differences in healing patterns between acute and chronic wounds, complete wound closure may not be a realistic outcome in many chronic PUs.

Therefore, alternative endpoints such as reduction of exudate, improved pain management or an overall improvement in health outcomes may be more appropriate in the assessment of PUs.

1.6 The history of health outcome measurement

Health outcome measurement is a relatively new concept that has grown steadily over the past 40 years. Traditionally health was thought of as a physical or mental state with assessments focusing on the presence or absence of disease. With an increase in chronic diseases where cure may not be possible but the disease may cause severe disability, treatment focus has shifted from cure and prolonging life - 'life quantity' - to alleviating symptoms and returning patients to pre-disease levels of functioning; essentially 'life quality' (68). Methodological advances in measuring patient-reported health outcomes have led to a departure from the traditionally perceived biomedical model to a broader perspective of health and disease that perceives health as not merely the absence of disease but complete physical, mental and social well-being (69). Stemming from this emerged a biopsychosocial model (70) that integrates the psychological and environmental (the 'social') into the biomedical model of health. This fundamental shift in health-related thinking led to an increasing interest in health measurement, outcomes such as health status, functioning, well-being and patient satisfaction, and in the use of such outcomes in healthcare research, practice and policy (71). The field of health is reliant on health outcome measurement to provide a strong evidence-base incorporating both patient perspectives and cost analyses.

Models that define and measure health outcomes are principally based on the World Health Organisation's (WHO) definition of health where the patient perspective is central to the concept of health, not simply the presence or absence of disease (72, 73). The WHO distinguished between the aetiology and processes (pathology) of non-fatal chronic diseases on human life (International Classification of Impairments, Disabilities, and Handicaps [ICIDH]) at three levels: impairment (any loss or abnormality of psychological, physiological or anatomical structure or function at organ level); disability (physical and psychological functional limitation caused by impairment); and handicap (the effects of disability, and of adaptations to it, on an individual's ability to perform social roles and thus the degree of social disadvantage conferred by the disability) (74, 75). To address shortcomings of the ICIDH (i.e. overlap between dimensions of impairment, disability and handicap; lack of clarity about causal and temporal relationships between these three dimensions; and insufficient attention to the role of the environment) (76), a new classification model of functioning and activity was developed, the ICIDH-2 (77), and organised into three

dimensions: body level (loss of body functions and structure considered as impairments), individual level (relates to performance of person-level tasks and limitations that cause an individual difficulty performing the activity) and societal level (relates to an individual's participation and involvement in life situations in relation to body functions and structure, health conditions, activities and contextual factors (77); involvement, inclusion of the individual in life activities in the context of where they live, is key (76)).

Wilson and Cleary (1995) proposed a conceptual model of measures of health outcome, integrating both biological and psychological aspects into five levels (Figure 1.1) (78). They proposed a theoretical link between clinical variables and measures of QoL (see Section 1.8), linking the biomedical (i.e. biological, physiological, and clinical outcomes) with the social sciences (i.e. functioning and well-being). In doing so they proposed a wider linear progression than that suggested by WHO. Biological factors directly relate to symptoms; symptoms influence functioning; and functional status directly influences patient perceptions of health outcomes. The concepts proposed are integrated but in addition there are a number of individual and environmental factors that contribute to relationships; factors that cannot be controlled by clinicians or healthcare systems. The model has been widely applied to different populations, including patients with chronic heart failure, HIV infection, and cancer (79, 80, 81).

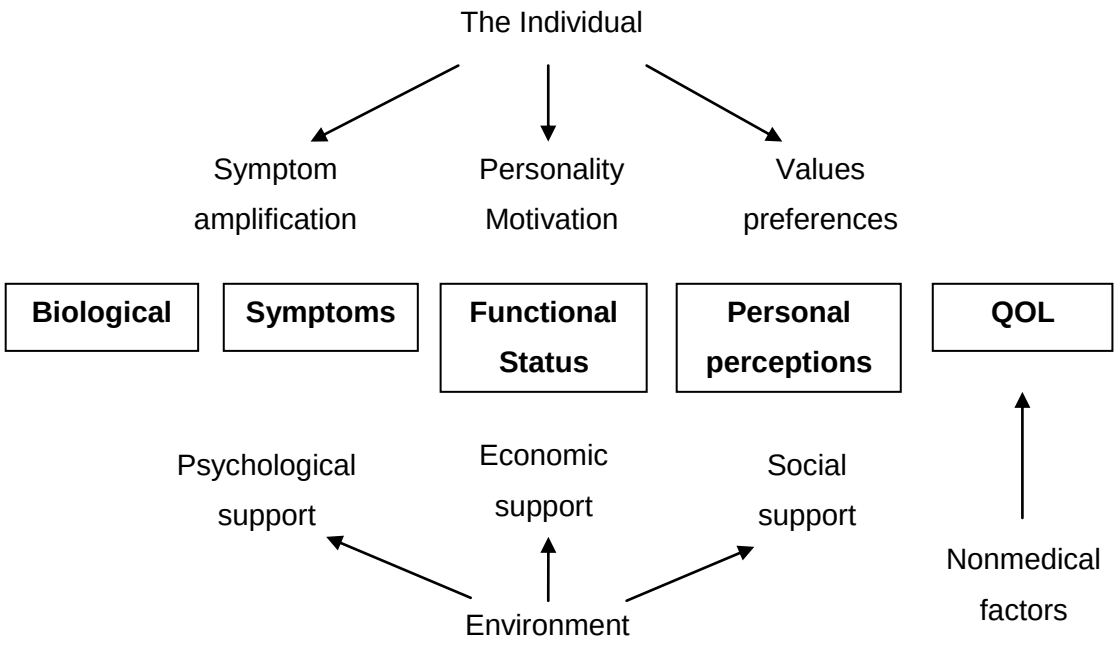


Figure 1.1 Conceptual model of Health Outcomes (78)

Conceptual models are useful for proposing a theoretical link between health problems (e.g. chronic disease, trauma and injury) and the extent these health problems and associated treatments effect patient functioning and well-being (82). In the process of developing a conceptual model, important PROs for measurement, associated to a particular health problem, are identified.

1.7 Patient-reported outcomes

Assessment of PROs has become increasingly important in many conditions (83, 84), and there is now an international consensus that PROs play an important role in clinical practice and research (83, 85, 86, 87). A PRO is a measurement of any aspect of a patient's health status that comes directly from the patient without any interpretation of the patient's response by physicians or others, about how they function or feel in relation to a health condition or its therapy (86). Some variables can be measured directly such as physiologic measurements (i.e. blood pressure) that may assist clinicians in diagnosing or treating an illness. Other variables, such as information about how a patient feels or some treatment effects, are measured indirectly and therefore can only come directly from the patient. Variables that cannot be measured directly (as opposed to observable variables) are called latent variables.

PRO is an umbrella term that has been adopted to describe a set of tools (usually questionnaires) that attempt to measure one or more latent variables related to health. The term does not tell us what is being measured, only that the patient is providing the data. These tools can enhance healthcare provider-patient communication and decision-making and assist in evaluating how well a new drug or medical intervention is working to improve patient care by providing information to physicians and patients needed for selection of the best treatment (see section 1.12). Relying solely on direct measures of health status such as a clinician's assessment of exudate level does not account for an individual's threshold for tolerance of or the amount of bother attributed to the exudate experienced.

1.8 Quality of Life versus Health-Related Quality of Life

A patient's health status can be measured through various concepts including symptomatic outcomes (i.e. pain), effect on ability to carry out daily tasks, and to more complex concepts such as QoL. QoL is a broad concept used (and abused) widely with the assumption that everyone knows what it means. There is no universal definition but it is accepted that QoL is a multi-dimensional concept referring to all aspects of a person's well-being influenced by the persons' perceived level of

satisfaction in a variety of circumstances (i.e. perceived happiness relative to their life in general). It encompasses four primary domains: psychological, physical, social and role functioning (88, 89). Psychological factors include depression, anxiety, and adjustment to illness and treatments; physical factors include pain, discomfort, mobility, sleep, appetite, and nausea; social factors include personal and sexual relationships, and engagement in social and leisure activities; role functioning includes daily activities such as the ability to cope with household duties and fulfilling roles within the family (73, 88, 90, 91). Other domains in QoL models include: occupational (88), health perceptions (92), housing, recreation, environment, independence and spirituality (88, 90, 91, 93, 94, 95) as it is considered that QoL is influenced by factors beyond health status.

QoL is a concept that means different things to different people and takes different forms depending on the specific circumstances and conditions under which the term is measured and applied. When assessed in two people with similar circumstances (i.e. same medical condition), both people may have entirely different evaluations of their QoL, therefore QoL is subjective in nature and best measured by directly asking the person themselves. In healthcare, QoL refers to the impact of a given disease or medical condition on a patients overall normal functioning in addition to the effects of the intervention being evaluated (90), a concept referred to as health-related quality of life (HRQL).

QoL and HRQL are often used interchangeably in the literature but they are distinct concepts. Like QoL, HRQL is a multi-dimensional construct that represents an individual's perceptions of how a given disease or medical condition and its treatment affect, at a minimum, their psychological, physical and social functioning. Several general models of HRQL also include separate domains for adjustment to illness and treatment, psychological symptomology, pain and other physical symptoms (96, 97), general health and health perceptions (73, 91), immobility, sleep disturbance and energy (73, 88, 90, 91, 93, 95). Essentially, HRQL measurements assess physical consequences of disease (symptoms and decreased function), effect of disease on a person's emotional state, feelings, coping behaviours, and self-identify (psychological functioning), and a person's ability to interact with others and participate socially (work, social interaction and relationships, role functioning). The term HRQL will be used throughout this work.

HRQL is viewed as an important outcome in the overall healthcare of patients. If the goal of healthcare is to protect, promote and preserve people's health then we need

standardised methods for assessment of health outcomes. A simple way of assessing HRQL would be to ask a patient how they are feeling but this would produce vastly subjective data. Therefore we need ways to quantify how people are feeling, for example, how a particular disease and/or treatments are affecting the patient. We can do this with rating scales (now known as PRO instruments). PRO instruments measure complex variables broken down into their component parts; manifestations of latent variables are transformed into numbers that can be taken as measurements (87). As HRQL is a multidimensional construct, in practice, HRQL is broken down into core components (domains) that are measured by PRO instruments as separate constructs or variables.

1.9 Types of PRO instruments

There is increasing international consensus advocating the use of scientifically meaningful PRO instruments in clinical trials and healthcare evaluation (83, 84, 86). The best PROs are designed to probe patients in a structured, formal way to give reproducible, meaningful, quantitative assessments of how patients feel and how they function from their perspective (86). PROs may be generic, disease/condition specific and preference based. Questions range from simple dichotomous responses (Yes/No) to those with several possible responses (e.g. Likert scales). The intention is to produce a summed score of responses for a particular variable (e.g. pain).

Generic instruments are designed to measure a range of concepts that can be applied across multiple diseases, outcomes, treatments/healthcare programmes and populations, as well as used with healthy populations (84, 98, 99). Being applicable to various populations makes them useful for broad comparisons of the relative impact of healthcare programs across diseases. A commonly used generic instrument is the medical outcomes study (MOS) short-form (SF-36) health survey (100).

In contrast, disease-specific instruments cover measures that are directly related to a particular disease and used to assess the impact of that disease on patients' health and functioning, with the goal of detecting minimally important effects in individuals (99). Disease-specific instruments often have a better sensitivity to minor changes in HRQL than generic instruments. The two approaches are not mutually exclusive; each has its strengths and weaknesses and are suitable under different circumstances (99).

Although also used to evaluate HRQL outcomes, preference-based measures of health are a means of estimating health state values or preferences for calculating quality adjusted life years (101). These tools are used in economic evaluation alongside

clinical trials to value the benefits of treatment or other healthcare (i.e. cost-based evaluations considering the cost of treatment against time to healing).

1.10 What to consider when selecting a PRO instrument

In practice and research, only structured and psychometrically rigorous instruments should be used. PRO instruments should be designed to minimise measurement error and ensure consistency, ultimately providing a more reliable measurement than one that would be obtained by informal interviews. The Scientific Advisory Committee of the Medical Outcomes Trust identified eight key attributes that a PRO instrument should meet (102); an instrument should be appropriate, acceptable, feasible, interpretable, have precision, reliability, validity and be responsive to change. More recently, the US Department of Health & Human Services Food & Drug Association (FDA) produced guidance to help ensure that PRO instruments are developed to measure what they are intended to measure and that they are supported by solid, scientific rationale (reliable, interpretable, and valid) (86). When selecting an instrument, in the first instance, determining whether an instruments' content is appropriate for addressing a particular research question should be made. For example, what is the area of investigation? If it is HRQL, then what domains are relevant and does the intended instrument represent these domains? Then we might look at the instruments' psychometric properties and make judgements about these properties in respect to specific contexts (103). In healthcare, where we might require scales that are sensitive to changes over time, we would look for evidence of the instruments responsiveness. Key instrument attributes and measurement properties will be discussed in detail in subsequent chapters.

1.11 Measurement of HRQL in patients with PUs

The development of PUs is widely perceived to impact on HRQL although few studies contain empirical data to substantiate this assumption (see Chapter 2 for a review of the literature). Work to date has been mainly qualitative; identifying that PUs cause substantial pain and discomfort and affect sleep, rehabilitation and mobility, and psychological, physical and social aspects of patient's lives (104, 105, 106, 107). PUs require intensive treatment that consequently affects patient health and functioning and causes significant treatment burden (9), particularly the frequency and regularity of dressing changes which in turn affects daily routine, fatigue, restrict mobility, and cause pain or discomfort. Broader changes to health and well-being include loss of independence associated with functional decline, altered eating habits, emotional well-being (from feeling annoyed and frustrated to angry and depressed), social isolation

and reduced activity. The presence of these factors can influence the development of new PUs or exacerbate the severity of or ability to heal existing ulcers.

A lack of quantitative data may be due to no appropriate instruments being available. There has been a growth of HRQL instrument development in chronic wounds, however, there are no PU-specific instruments available (108). Attempts to measure changes in HRQL in people with PUs using quantitative data collection methods (i.e. use of generic PROMs to evaluate HRQL in PUs) have had inconsistent results; some findings indicate reductions in HRQL for patients with PUs compared to those without (109, 110, 111) while others found no differences in HRQL between people with PUs and controls (112).

1.12 HRQL assessment: Why measure it?

Assessment of HRQL outcomes can serve a number of important purposes, many which are not yet being realised in the PU field.

1.12.1 Clinical

Healthcare is increasingly patient-centred with focus moving towards holistic care. An increase in chronic conditions, where cure may be unrealistic, improving HRQL has been an important outcome. In cancer, individual patient level HRQL data has led to improved outcomes for patients as it has been an important way of supplementing informed decisions regarding treatment, acknowledging patient preference for treatments, and improving service provision (113). In PUs, there is some evidence that both short and long term interventions can have a dramatic impact upon key aspects of patient's lives, such as comfort, stability (due to bed/chair height), sleep, and pain. Noteworthy is that in many instances, a patient may not immediately return to their previous levels of functioning once their PU has healed. Therefore important patient outcomes such as HRQL may provide information about the effectiveness of a PU intervention (e.g. an intervention may be perceived as effective in reducing pain, an outcome important to the patient, irrespective of whether the PU has healed); patient-priorities and preferences; acceptable symptom states; and treatment burden (or important improvements with treatment) (83, 114). These patient outcomes are crucial for monitoring treatment success/failure (changes in HRQL) over time and quantifying symptom severity and functional ability that can be used in adjunct to routine clinical assessment to improve quality of care (e.g. a physical function score provides a common reference base for multidisciplinary teams) (113, 115). Further, HRQL outcomes are important when two people with the same clinical criteria have

dramatically different HRQL responses. For this reason, we need instruments that can differentiate between people who have better HRQL from those with worse HRQL.

A study of functional disability in ambulatory patients found that 66% of patient-reported functional limitations were underestimated or not detected by clinicians (116). Importantly, a patient advocate wrote (117) that symptoms are consequences of illness that affect HRQL as does anticipation (hypothetical) of symptoms; but HRQL affects tolerance for symptoms, supporting claims that reduced HRQL may be reported without presence of disease or symptoms. Measuring how a patient feels is as important as objective clinical measures and certainly more important than healthcare provider perceptions of how a patient "ought" to be feeling based on clinical measurements alone.

In addition to informing decisions about treatment selection and healthcare management, assessment of HRQL helps patients make informed decisions based on what others have experienced (i.e. likely effects of treatments); the efficacy and mortality associated with a particular treatment/intervention; and the expected impact on HRQL outcomes (118). Patients can feel empowered from being involved in informing decisions about their care and the role of nurses as patient advocates is enhanced (119).

1.12.2 Research

In the current healthcare environment, rapidly evolving technologies and treatments must be evaluated in light of the impact (and benefits) on patients and cost effectiveness. This is particularly important as patients want information to make informed decisions about their healthcare; healthcare providers are being made accountable for what happens to their patients; and there is growing emphasis on EBP and rising pressure on healthcare providers to provide credible evidence regarding the effectiveness and efficiency of both new and existing treatments and technologies; making it pertinent to appropriately monitor and measure health outcomes and healthcare costs (86, 120). As such, HRQL outcomes can be used in clinical trials as primary endpoints to assess the benefits of one treatment/medical device/management programme over another or when developing new technologies to incorporate the patient perspective; wasted time could be spent on developing new technologies that are later found to be unacceptable to patients. HRQL outcomes are useful when two treatments reveal similar clinical benefits but exhibit differences in HRQL outcomes. When the efficacy of two treatments evaluated in a clinical trial are found to be

equivalent, the treatment associated with better HRQL is more likely to be adopted (121).

1.12.3 Economic

Technology and health practices are continually developing and measures of HRQL are increasing used in studies of cost-effectiveness which produces evidence for those making decisions about resource allocation (24); resource allocation is justified by determining the benefits of treatments against financial costs with consideration of 'human cost'. To inform the development of the most effective PU-interventions and to provide the most effective management of PUs, assessment of the benefits of an intervention needs to be determined (86, 119, 120). In PUs, many practice recommendations are not based on good research evidence and the evidence-base associated with PU intervention effectiveness has not been demonstrated using appropriate methods (122). Effective wound management should not only be about dressing a wound or applying pressure-relieving equipment but also consideration of the effect of the ulcer and its treatment on the patient and whether alternative treatments that better suit patient needs are available.

1.13 Measuring HRQL in PUs: measurement and practical issues

In clinical trials of PU intervention effectiveness, assessment of outcomes other than conventional clinical outcomes has often been neglected. Where HRQL was measured, PRO instruments without established psychometric evidence, single-item questioning methods or non-validated questionnaires were used (123). To have true measurement, instruments need to be reliable and valid. Related, some illnesses may not have appropriate PRO measures available; and instrument development can be a long and arduous process, or validated language translations (PROs are usually developed in dominant languages that might not transcend cross-culturally).

A growing number of studies indicate discordance between patient and healthcare provider reports of HRQL outcomes, with discrepancy between what is important to patients compared to what healthcare providers think are important outcomes (124, 125, 126). Therefore PROs should ideally be patient- not clinician-reported. Training in the application, administration, scoring and interpretation of PRO scores requires time and therefore has administrative burdens and cost implications. Use of some instruments requires permissions and additional cost to use. Another practical issue is that PRO instruments often contain a multitude of items on numerous domains. Completion requires time and effort on the part of the patient, introducing respondent burden and potentially raising sensitive issues that could be distressing to patients.

Many patients who develop a PU often have a range of additional comorbidities adding to the burden. This can affect attrition and response-rates as patients with a poorer outcome prognosis are less likely to report HRQL data (127). Further, the pathophysiology and aetiology of developing a PU, living with a PU, and the burden of various concomitant diseases, contributes to the formation of a complex interaction of factors that need to be accounted for when assessing HRQL in patients with PUs; the HRQL data obtained may not solely reflect PU impact but a combination of factors.

Reviews of the effectiveness of using health outcomes in routine practice report conflicting results; there is limited evidence that their use substantially changes patient management or improves patient outcomes, and a general lack of clarity about the intended applications of PROs in clinical practice (128, 129, 130). An important factor is the selection of appropriate PROs to ensure that the information provided is sufficiently specific and relevant to clinicians and patients to prompt changes to patient management. Clinicians indicate positive attitudes to the feasibility, acceptability and utility of PROs, and their ability to detect psychological, and to a lesser extent, functional problems (128), but mixed reports of the benefits of PRO data limit their use in routine practice. Clinicians seem willing to use PROs but particularly in the PU field, appropriate PROs may not be available.

1.13.1 Methodological developments

Recently there has been an expansion of sophisticated quantitative methods for assessing HRQL. Guidance for the development and evaluation of instruments to obtain HRQL data (86) highlight the importance of conceptually sound, reliable and valid measures that explicitly define the constructs measured (what is measured), the intended population (range and characteristics of people for whom the instrument is suitable for application/use); and the measurement model (scale scores and what values mean in terms of individual measurements). This has resulted in improved measurement properties and aids in the selection of instruments that are fit for purpose and satisfy established scientific criteria. Thus, PRO data are an integral aspect of patient care, policy decision making and healthcare delivery (87).

1.14 Research aims and structure of thesis

The principal aim of this project was to develop a patient-centred measure of HRQL specific for people with PUs (the PU QoL instrument - PU-QOLI). The aim was to produce an acceptable, reliable and valid PRO, suitable for use in clinical trials, epidemiological studies and in routine clinical practice. The perspective of persons

with PUs was central in all stages of development and evaluation. Collaboration was sought from members of the European Pressure Ulcer Advisory Panel (EPUAP) and from various acute and primary care NHS Trusts around the UK.

1.14.1 PUQOL project in the context of Pressure Ulcer Programme

The Pressure Ulcer Programme of Research (PURPOSE) funded by the National Institute of Health Research aims to reduce the impact of PUs on patients through two streams: 1) early identification of patients at risk of PU development and patients at risk of progression to severe PUs; and 2) development of methods to capture outcomes important to patients including HRQL and health utilities for routine clinical use and in clinical trials. The pressure ulcer quality of life (PUQOL) project falls under research stream two of the Programme.

1.14.2 Research Design Overview

This multi-centre study is designed to develop and evaluate the psychometric properties of a PU-specific HRQL instrument for patients with PUs. Guidelines for developing and evaluating health outcome measures were consulted to ensure high quality and standardisation for PU-QOLI development (86, 131, 132). These guidelines recommend that patients and collaboration with experts is utilised through all stages of instrument development and it proposes distinct phases for the development of a PRO measure: 1) conceptual framework; 2) generation of items for the PU-QOLI and pretesting; and 3) item reduction followed by a full test of psychometric properties. Figure 1.2 details these three phases.

Chapter 2 describes the development of a working conceptual framework of PU-specific HRQL based on a comprehensive and systematic review of the international literature on PUs and HRQL including both quantitative and qualitative primary research.

Chapter 3 describes a qualitative study of 30 patients, conducted to explore determinants of HRQL in people with PUs and to refine and further develop the working conceptual framework to produce a final framework, providing a structured and formal method for assessing content of available PRO instruments.

Chapter 4 describes a review undertaken to identify and evaluate currently available PRO instruments used in PU research and other chronic wounds to determine the suitability of available instruments for use in PUs.

Chapter 5 describe the development (phase 2) of the PU-QOLI, specifically the construction (item generation, design and format) and pretesting of the preliminary PU-QOLI using mixed methods (qualitative and Rasch measuremment methods).

Chapter 6 describes the psychometric evaluation (phase 3) of the PU-QOLI in the first field test, intended for establishing feasibility and acceptability; producing a shorter version by selecting items that perform best against established psychometric criteria; and identifying sub-scales and testing scaling assumptions.

Chapter 7 describes the psychometric evaluation of the PU-QOLI in a second field test, intended for carrying out a full psychometric evaluation of the reliability and validity of the final PU-QOLI version in a large independent sample of patients with PUs.

The final chapter summarises the conclusions from the research included in this thesis and discussed some methodological considerations. The implications of the research for clinical practice and future research directions are also discussed.

In addition, methodological developments have been woven into the research, including: an evaluation of optimal search strategy for retrieving HRQL data electronically (Chapter 2); mixed methods for pretesting and the added value of Rasch measurement methods (Chapter 5); and mixed methods for optimal PRO instrument mode of administration (sub-study, Chapter 6).

1.15 Summary of Chapter 1

PUs are a major health problem and challenge to patients, healthcare professionals and the healthcare system. They are prevalent in all health settings, considered a significant health problem associated with morbidity and mortality, and have cost and quality implications for health services. PUs can severely compromise all areas of patient functioning, and the intensive interventions for preventing and treating PUs pose additional patient burden, and subsequently impact HRQL. Therefore, evaluating HRQL outcomes is particularly important and relevant in PUs. The extent and nature of the impact of PUs on HRQL needs to be determined.

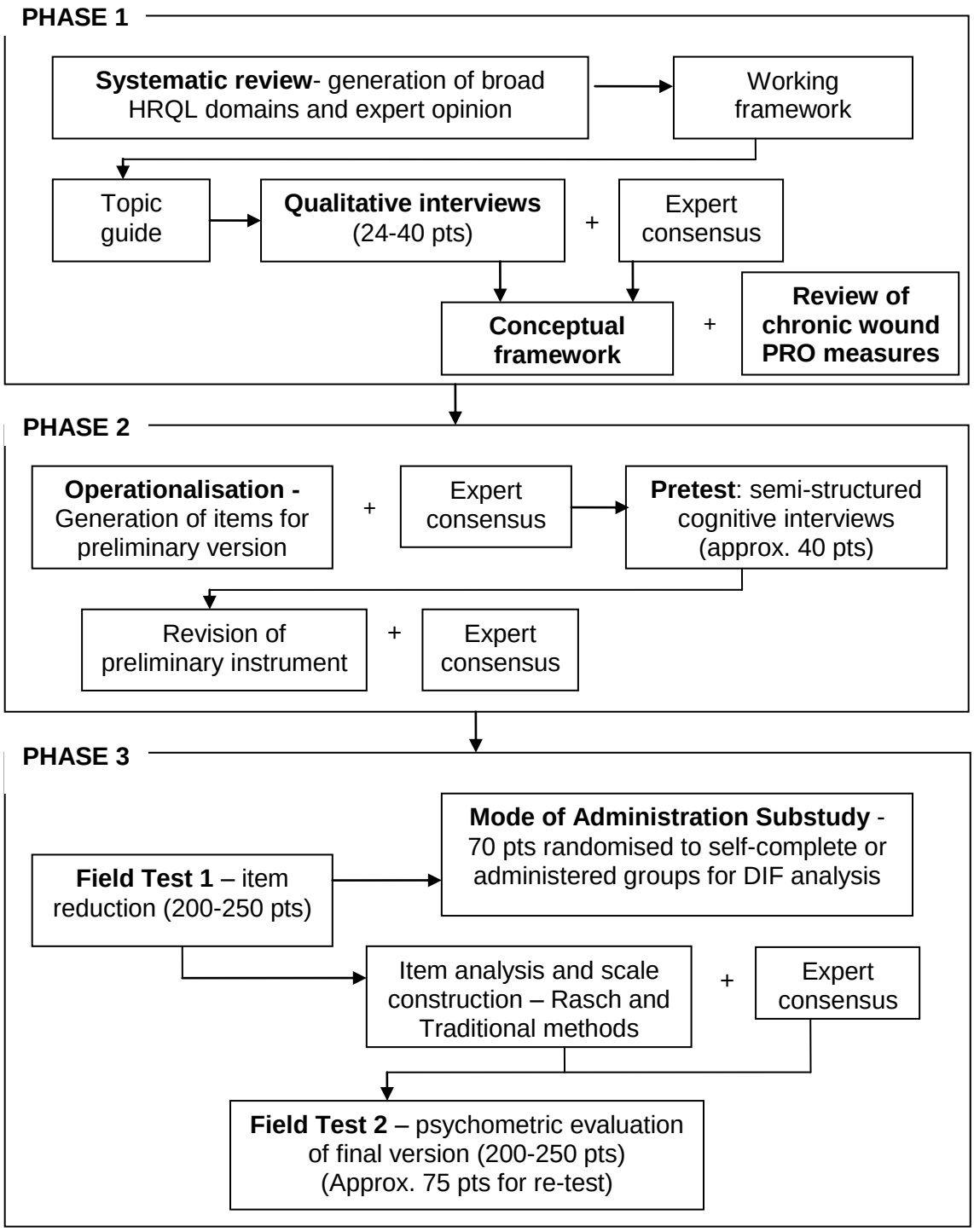


Figure 1.2 Flow diagram of the research design including 3 phases of the development and evaluation of the PU-QOLI

Chapter 2

SYSTEMATIC REVIEW OF THE LITERATURE

A working conceptual framework of health-related quality of life in pressure ulcers

2.1 Overview

Chapter 2 reports the methods and results of a systematic review of the PU and HRQL literature. This chapter includes both qualitative and quantitative research used to identify PU-specific PROs and to develop a working conceptual framework of only PU-specific HRQL outcomes.

2.2 Background

Decisions about the delivery of healthcare and about healthcare policy development need to be informed by 'best' EBP. One key element of EBP is reviewing, appraising and synthesising research in the form of robust systematic reviews. Whilst it is widely established that randomised controlled trials (RCTs) are the 'best' approach to generating evidence of effectiveness, RCTs have provided limited information about HRQL in the PU field. RCTs in this area have relied on objective clinical assessments such as PU healing time, and rarely included HRQL assessment in their methodology, particularly outcomes directly reported by patients, to gain information about patients' perspective of the impact of PU interventions on HRQL (55, 62, 123).

PU development places the individual at risk for a multitude of negative psychological consequences that can affect health, well-being, and HRQL. Understanding the basic mechanisms underlying the psychological consequences of having a PU and the outcomes that matter most to patients can help provide the basis for interventions aimed at preventing or treating PUs that will benefit patient HRQL. A way to assess these benefits would be with psychometrically robust measures. The existing literature is the best starting point to determine if appropriate measures are available, before adapting or commencing development of new ones. It is also a good source for developing a disease-specific conceptual framework that available measures could be assessed against or used to inform the development of new measures if required.

There are established scientific methods for systematic reviewing to produce a collation of the best available evidence, and these methods seek to minimise bias and error. These methods have predominantly focused on synthesis of RCTs with use of meta-

analysis, although qualitative research is increasingly contributing to EBP, with more recent qualitative synthesis methods established (i.e. meta-synthesis) (133, 134). A wide range of research approaches, from both qualitative and quantitative disciplines have been utilised to generate information about HRQL for further knowledge and enhancing healthcare practice. Specifically, assessment of HRQL has been useful in research, clinical practice, health services evaluation and public health, as well as enormously influential in improving service provision and health care management in many disease areas. One such area is oncology, where the assessment of HRQL in patients with breast cancer (135) found that HRQL assessment was important for supplementing informed decisions regarding treatment and patient preference for treatments.

A problem with researching HRQL in PUs is determining which methodology is most effective for obtaining information about the impact of PUs from the patients' perspective. There is ongoing debate between the proponents of qualitative and quantitative research. On the one hand, advocates of qualitative methods discount quantitative methods on the basis that they produce distorted or inconclusive results; results produced by biased samples (136). While, qualitative methods are argued to be more about creativity and less about scientific knowledge, validity and rigour; data obtained are dependent on individual perceptions and on the interactions between the researcher and participant which may introduce bias (137, 138). Both approaches have their own strengths and weakness and are increasingly advocated and used in combination in research (139). Including studies that have used either approach for data collection may provide a broader, deeper interpretation of PRO data in this field.

In this thesis, of primary importance is the conceptualisation of HRQL specific to PU impact. As such there is a need to investigate the impact of PUs on patients' and of the HRQL achieved with different interventions.

2.3 Aim

The aim was to undertake a comprehensive and systematic review of the international literature on PUs and HRQL including both quantitative and qualitative primary research, based on the methods suggested by the United Kingdom National Health Service Centre for Research and Dissemination (140).

2.3.1 Objectives

The objectives were to develop a working conceptual framework of PU-specific HRQL by identifying the impact of PUs and PU interventions on HRQL (e.g. PROs important to patients who have a PU or are receiving PU treatment interventions) and the relative PU impact and burden from the patients' perspective.

2.4 Methods

2.4.1 Search strategy

An iterative process was used to develop an optimal search strategy to ensure sensitivity without yielding too many results. The search strategy (Appendix 2.1) sought to identify all research investigating patient reports about the impact of PUs and PU interventions on HRQL. It incorporated search terms used by other reviews, extensive consultation among the collaborative group and input from a methodological research advisory group. To heighten sensitivity, the search was designed to identify qualitative research that explored patients' subjective experience and quantitative studies that measured HRQL using standardised or study-specific HRQL measures.

In addition, an evaluation of the effectiveness of research-methodology (qualitative) based search strategies with subject-specific (HRQL) search strategies for electronic retrieval of qualitative patient-reported HRQL research data was undertaken (141). This was in order to determine the best combination of search terms to ensure comprehensive and precise retrieval of all qualitative patient-reported HRQL data, regardless of research methodology, without missing key studies or retrieving excessive numbers of irrelevant studies.

Search terms of key words for: PUs (60), PU-symptoms, patient views (142), HRQL (140), RCT (143), qualitative methodology (144), were combined by Boolean operators: (PUs) and (PU symptoms) and (patient views) or (QoL) and (RCT) or (qualitative).

2.4.2 Data sources

Thirteen electronic databases were searched from inception until 4th July 2008: AMED, British Nursing Index (BMI), MEDLINE, EMbase, PsycINFO, Cumulative Index to Nursing and Allied Health Literature (CINAHL), The Cochrane Library, Proquest, Networked Digital Library of Theses and Dissertations, International Theses in Progress, Theses Canada Portal, Australian Digital Theses Program, and Index to Theses. An auto alert function was set up within the electronic databases until September 2008 to identify additional studies meeting the search criteria.

Systematic reviews of beds, mattresses, dressings and other PU interventions were identified through the Cochrane Library database (62) and a citation search performed (60, 145, 146, 147). RCTs published after the census date of each systematic review were also searched and included in the review if they met eligibility criteria (140). This process was undertaken to ensure that any RCT that had addressed a HRQL outcome as a secondary outcome but not reported this outcome in the abstract, was not missed.

To identify unpublished and on-going research, the electronic searches were supplemented by a hand search of specialist journals, relevant conference proceedings and dissertations. Twelve experts were contacted, the PRO and QoL Database and the UK National Research Register searched, a citation search of all included studies performed, and an internet search of web content relating to PU self-help and focus group websites undertaken (see Appendix 2.2 for detailed list).

2.4.3 Criteria for study selection

Primary research was included if it: addressed the impact of PUs and interventions on HRQL including symptoms and evaluation of interventions (e.g. pain, ease of use, comfort)[2] by direct patient-reports; in adult populations; with existing PU ≥ Grade 1 (148); from any setting; and reporting qualitative and/or quantitative data. Thus, studies were excluded if PU-related outcomes were not reported (i.e. methods for obtaining data were not described/no HRQL results reported); samples of patients with mixed chronic wounds studied and no separate analysis of findings for PU patients undertaken[3]; only adverse effects reported (e.g. pain may be an adverse effect of treatment but not measured as a PRO)[4]; outcomes reported by clinicians, nurses or carers; or the paper was unobtainable or missing data could not be obtained from the authors. No language, date or methodology restrictions were applied. Members of the EPUAP were contacted, requesting their support to review abstracts and studies not published in English or readily accessible (see Intellectual Property Section, page ii).

[2] Studies that only asked patients if an intervention was acceptable or satisfactory (yes/no) were excluded. Acceptability, satisfaction and compliance with interventions raise important issues for healthcare but do not necessarily provide information about the impact on HRQL. A yes/no measure of acceptability/satisfaction is not a sensitive measure and not the focus of this review.

[3] Studies with mixed chronic wound patients were obtained to determine if independent PU analysis had been performed and PU results presented. If not presented, authors were contacted for their PU data.

[4] AEs and SAEs were not deemed as PROs (i.e. they are not an outcome that is collected systematically for all patients)

2.4.4 Selection of studies

Abstracts from studies retrieved were screened for relevance by one researcher (Claudia Gorecki; CG). Studies that did not meet the inclusion criteria were excluded at this stage. Studies assessed as potentially relevant or where relevance was ambiguous, were obtained in full for further scrutiny. Two researchers (CG, JN) independently assessed potentially relevant studies and studies not meeting the eligibility criteria were excluded from further analysis. Where there was lack of consensus, disagreements were resolved through discussion with a third reviewer (JB). Where study details were lacking, attempts were made to contact the authors for additional information.

2.4.5 Quality appraisal

The methodological quality and issues associated with assessment and measurement of HRQL in PUs were considered. There is on-going debate about whether all patient-reported testimony should be included irrespective of quality vs. inclusion of only high quality research. A pragmatic choice was made to include studies meeting a minimum criteria to ensure comprehensiveness of patient voices but not at the complete expense of quality. Therefore, the following quality components for study methodologies including qualitative studies, RCTs, cohort studies and cross-sectional studies were developed and applied to studies identified as eligible:

Qualitative studies

Qualitative studies were appraised using a standard quality appraisal form included in the Joanna Briggs Institute Qualitative Assessment and Review Instrument (JBI-QARI) (149). Studies assessed as yes to at least 6 of the 10 quality statements, including yes to two critical methodological questions: there is congruity between i) the research methodology and the methods used to collect data (Q3) and ii) the research methodology and the representation and analysis of data (Q4), were included. These two questions were considered important to ensure that appropriate data collection and analysis methods were used that reflected the study's methodological approach; without this information reported and congruency in the methodology it would be difficult to conclude accuracy of results (i.e. can we conclude the results are valid based on the methods used to collect and analyse the data?).

Randomised Controlled Trials (RCTs)

RCTs were included if there was: clearly stated primary outcome; *a priori* sample size calculation; randomised allocation to treatment; intention to treat analysis; and no more

than 5% patients excluded from analysis of primary outcomes for reasons including loss to follow-up, withdrawal, death and protocol violations (150).

Cohort Studies

Prospective cohort and cross-sectional studies were included if they had no more than 20% of the study sample excluded from analysis for reasons including withdrawal, death, lost to follow-up and missing records (39, 151). Quality was assessed by two independent reviewers (CG and JN); studies failing to meet the quality criteria were excluded from further consideration.

2.4.6 Data extraction

All findings from included studies were extracted by one reviewer (CG). Qualitative findings were extracted into a QARI database as QARI was developed to manage, appraise, analyse and synthesise study findings that utilise any qualitative approach as part of a systematic review of evidence. QARI incorporates data extraction and quality appraisal forms; a data synthesis function; and a reporting function. This software has been internationally peer reviewed and was successfully used in a recent review with a similar research question (152).

For qualitative studies, findings in the form of textual data, emerged themes and author interpretations (149) were extracted. Data extraction involved the reviewer (CG) reading carefully and thoroughly through each study and identifying the findings in the form of either a direct quote from patients or a statement by the author that was supported by patient reported data; to establish the credibility of the authors' statement and interpretation. This process involved identifying and transferring actual patient-reported text from the original paper to produce a data set of findings that could be later categorised and synthesised. Information pertaining to methodology (i.e. phenomenology; textual narrative; grounded theory), methods to obtain data (i.e. semi-structured interview; face-to-face or telephone), participant characteristics, setting and geographical location, cultural description (i.e. elderly stroke patients; SCI), and analytical methods were also summarised. Data from the quantitative studies were extracted into pre-prepared data-extraction tables based on study design (i.e. RCT prevention or treatment, prospective cohort, cross-section/survey, case series). Data extraction involved the reviewer reading each study and identifying findings in the form of individual questionnaire items or domain results. Participant characteristics, setting, geographical location, study design, HRQL outcomes and assessment methods, and both significant and non-significant results specific to PUs were summarised. For

studies employing both research methods, qualitative and quantitative results were extracted into their corresponding methodology databases.

The data extraction process was undertaken three times on separate occasions (i.e. on each occasion the reviewer thoroughly read each paper and extracted findings) to establish a systematic approach for data extraction, minimise error and ensure that relevant findings were not missed. Data extraction ceased when saturation was reached, with no new themes emerging (152). Where duplicate publications of patient datasets were identified, the most detailed report was used for data extraction. A second independent reviewer (JN) extracted data for 20% of the studies, selected at random. Consistency in extraction and agreement on the main themes was compared.

2.4.7 Analytical procedures

A combined synthesis of qualitative and quantitative research was performed, which involved generation of a list of likely factors and their relative importance, and using content analysis to generate common categories and themes from study findings. The principles of the triangulation method for integrating mixed method data (153), which involves determining consistency or inconsistency in findings from each method, were utilised. In the first instance, meta-synthesis (149, 154) was carried out on the qualitative data set of findings, pooling all findings from qualitative research. Each finding was allocated to one defined category. A category was determined by grouping findings that reflected relationships between similar phenomena and variables. Categorising findings was a way of aggregating findings and focusing on findings as a whole rather than individual studies. Categories sufficiently similar in meaning were generated into synthesised themes. Synthesising categories was a way of aggregating grouped findings into specific HRQL themes, providing a summary of the evidence for that particular domain. All findings from the quantitative data set were encoded to the categories and themes generated by meta-synthesis, with new categories added if necessary. This process allowed for comparing findings from both methods for agreement or identifying where findings appear to contradict. The combined emerging categories and themes were reviewed by a collaborative group (see Intellectual Property Section, page ii) and revised until consensus, producing a working conceptual framework of HRQL specific to PUs.

2.5 Results

2.5.1 Studies included in the review

A total of 6086 citations were retrieved, of which 5416 were excluded. 649 (11.29%) were assessed as potentially relevant and obtained in full for further scrutiny. Of these, 53 met the eligibility criteria however, 22 were excluded after quality appraisal (Figure 2.1). A total of 31 studies were included in the review. Sample sizes ranged from 5-468 participants with PUs from acute, community and long-term care settings, across Europe, USA, Asia and Australia (n=2463), between the ages of 17 and 96 years, with approximately 35% aged 65 years or more.

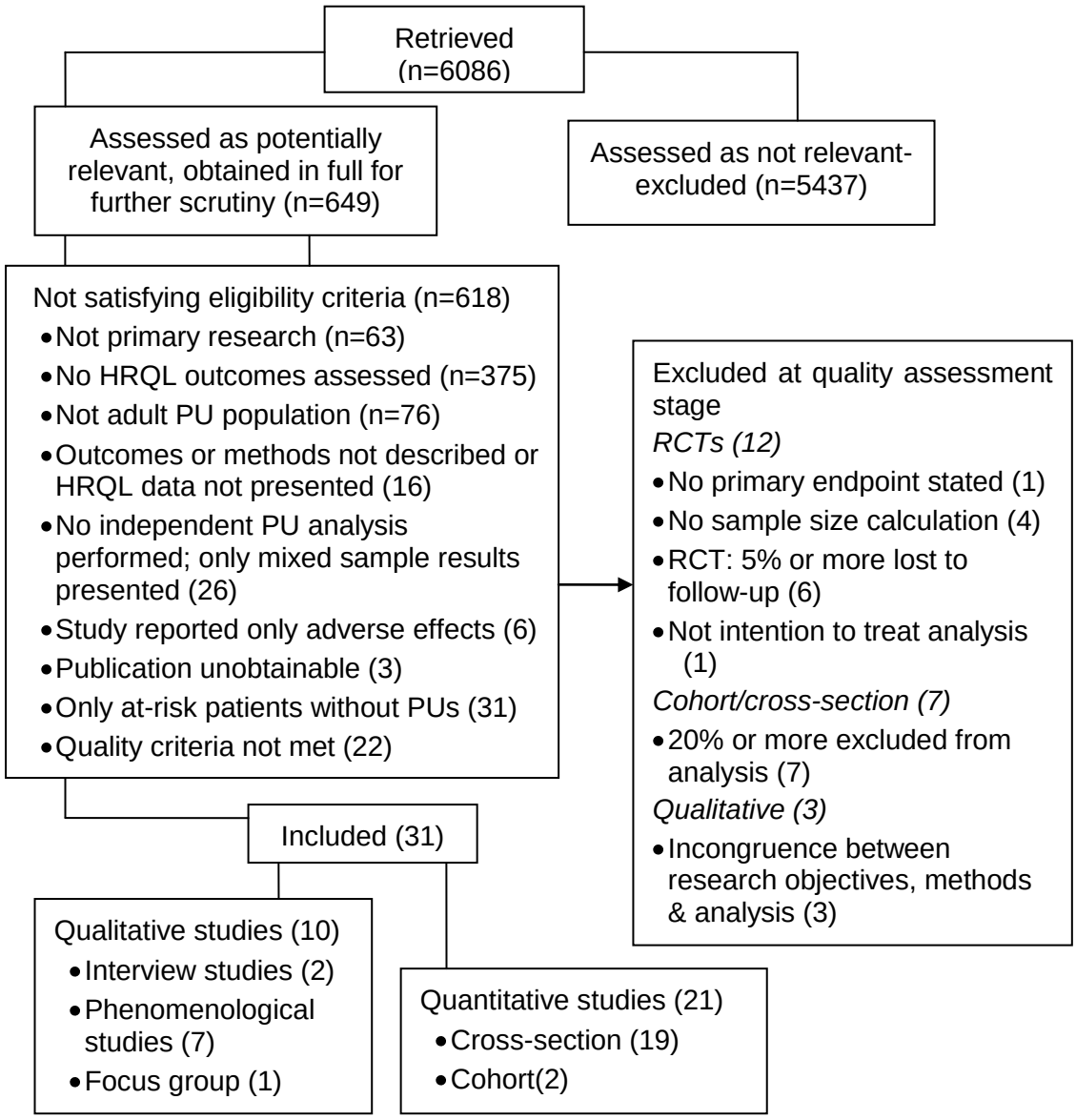


Figure 2.1 Flowchart of studies for inclusion and exclusion

Analysis of the qualitative studies extracted 222 findings which were allocated to one of 46 categories and synthesised into 11 themes (Table 2.1). A total of 71 findings were extracted from the quantitative studies and allocated to one of 20 categories and 7 themes; no additional categories or themes were identified (Table 2.1).

A total of 293 findings, 46 categories and 11 themes emerged from the included studies combined. Table 2.2 classifies the qualitative and quantitative studies by corresponding theme. A narrative summary of HRQL themes generated from the combined synthesis is provided in detail below.

2.5.2 Synthesised HRQL themes

Physical impact and limitations

Eleven studies reported PUs having a significant impact on physical aspects including imposing physical restrictions, lifestyle changes, and the need for environmental adaptations. PUs restricted most day-to-day activities such as performing self-care, shopping, mobility, and ability to move and assume comfortable positions. Patients reported reduced physical activity due to prescribed bed rest and less participation in outdoor activities than before the PU developed. In order to reduce the number of days lost from activity or days spent hospitalised for PU treatment, patients introduced skincare as part of their lifestyle. Skincare caused problems for some patients, although many thought that it was important to comply with recommended skin care.

Patients perceived PU treatment including hospitalisation as an intrusion in their daily lives, describing their PU as inconvenient, troublesome, and interfering their usual schedule. They attributed their PU with loss of appetite and interest in physical activities, insomnia, and the reason for them being less actively engaged in life in general. Physical dependence and need for assisted care resulted in patients being hospitalised, moving to more suitable accommodation, or needing to make their home wheelchair friendly (i.e. some patients with heel ulcers became wheelchair dependent).

Social impact

Social impact was represented in 10 studies in terms of restricted social activity, social isolation, and impact on personal lives. Patients reported that their social life was restricted due to the physical restrictions imposed by the PU and PU treatments. Patients expressed an inability to participate in social activities, explaining that need for wound care restricted social activity and social connectedness, and resulted in them feeling socially isolated, confined and missing family and friends. PUs, as well as hospitalisation for PU treatment, impacted on personal relationships because of

limitations regarding intimacy and sexual relations. The inability to participate in social activities contributed towards reduced social interaction, poorer social integration and adjustment issues, loss of interest in social activities, less positive engagement, and poorer interpersonal relationships. Loss of interest in social activities was also related to various PU symptoms such as pain and odorous exudate and contributed towards self-imposed isolation (i.e. distancing behaviours, detaching from others).

Psychological impact

Psychological impact was reported in 12 studies as coping and acceptance; body image and self-concept changes; desire for regained control and independence; and a variety of emotional problems. Patients found it difficult to deal with their PU and reported avoiding thinking about it which contributed to poor adjustment outcomes. Some coped by comparison with others, to the self in the past, and knowledge that it could be worse, resulting in resignation. Others found hope and support from their spirituality and expressed that the support and help given from family and friends was important and appreciated, especially help with daily activities and wound care.

Physical changes such as scarring and seeing the PU (e.g. “ugly”, “dirty”, “black hole”) impacted self-concept and body image. Patients stated feeling useless and inadequate, and a key issue identified from the studies was patients’ desire and struggle for control and independence. Patients expressed wanting to be involved in decisions about their wound care and wanted help from healthcare providers to become more independent (i.e. self-care). Many patients expressed loss of independence and some experienced self-inflicted tension between what one wanted to do with what one should do to promote PU healing.

Patients commonly reported negative emotions such as low mood, anger, frustration, anxiety and depression. Feeling hopeless was associated with depression and powerlessness, with some patients requiring treatment for their PU-related depression. Feeling shocked described the initial reaction to the PU, and as the PU worsened, dislike and hate were experienced. Preoccupation with healing and anticipation of pain also contributed towards emotional problems. Patients worried about the time it took for the PU to heal, fearing that it would never heal (“I’ve spent a year on that mattress”). In some instances, patients came to accept their situation and the impact the PU had on their HRQL, reducing negative emotions and improving psychological well-being.

Impact of PU symptoms

Fifteen studies reported the impact of PU-related symptoms. Pain was the most significant consequence of PUs. It impacted on every aspect of patients' lives, and was described using evaluative, sensory and affective pain descriptors (107). Evaluative descriptors described constant pain experienced ("never ending", "intense pain"), which varied in location, severity and duration. For some, pain was perceived as a punishment for having a PU. Sensory descriptors included feels like "stabbing", "burning", "throbbing" and "needles". Affective descriptors described emotional feelings about the pain experience such as frustrating, annoying, unbearable, angry and inconvenience. Pain caused discomfort and inability to move, stand, walk and assume comfortable positions. Experience of early symptoms such as pain or discomfort was indicative of pressure or skin damage. Patients felt that they were responsible for communicating any first signs or symptoms and that their healthcare provider was responsible for attending to reports made. Some patients practiced skin protection and pain prevention behaviours such as staying still, adding to immobility problems, or repositioning, which at times was "uncomfortable", "difficult" or "caused pain".

Patients reported that PU symptoms interfered with activities of daily living, contributed to sleep disturbances and impaired appetite. In turn, physical restrictions contributed to decreased social interaction. Patients attributed exudate with more frequent dressing changes and impaired ability to walk because "the exudate would ooze out", contributing to further problems with immobility. Self-imposed isolation was a result of the embarrassment and distress experienced by patients due to foul-smelling PUs.

Impact on general health and consequences

Health and healthcare problems were identified in 12 studies. PU-related health complications and health deterioration, such as infection, delayed wound healing and led to additional problems with fitness and health status. The PU and PU symptoms had implications for the healthcare that patients received. Patients reported extended hospital stays, additional hospital admissions, and restrictions on rehabilitation places. For some, PU pain prevented rehabilitation and receiving treatment for PUs or other medical conditions, as well as making patient handling difficult for healthcare providers and carers. For other patients, the PU was perceived as an acute problem compared to other existing conditions. Many patients have other health conditions and comorbidities and a PU may develop as a consequence of this. Therefore, the PU is not perceived as a significant health problem. For example, palliative care and SCI patients did not report pain as a significant experience of having a PU. This may be because SCI patients have reduced sensitivity or palliative care patients received better pain control.

2.5.3 Synthesised themes: Other impact

Impact on others

Patients from four studies reported their PU impacting on family/friends. These patients believed that their PU-related suffering and poor mood was projected onto those closest to them. In addition, restrictions imposed on patients caused others additional work and worry. Those closest are crucial in providing assistance with skin inspection, PU management and treatment, and assistance with daily activities (i.e. personal hygiene, shopping, cleaning, repositioning). Requiring assistance made patients feel dependent on others and they feared they were being a burden.

Financial impact

Three studies reported PUs contributing towards socioeconomic and financial consequences such as poor living circumstances and work opportunities (inability to work), and financial costs (medical/treatment costs, loss of income).

Need vs. impact of interventions and wound care

Eleven studies reported findings in the theme need for vs. impact of interventions, making it the most commonly occurring theme in both qualitative and quantitative studies. These studies assessed a plethora of treatments for wound management including dressings and pressure-relieving interventions, predominantly patient comfort. Patients also reported concerns about whether their pressure-relief allowed independence, ability to move, ability to get in/out of their bed or chair, whether it was noisy or loud therefore disrupting sleep, and whether they felt safe in bed or sitting on their cushion. Topical treatments were also evaluated (i.e. dressings, creams/lotions, maggot therapy) and some patients reported benefit in reduced pain and exudate, and wound healing. Conversely, for some, benefit was not experienced as analgesia did not reduce or eliminate pain even though it was intended for PU-related pain reduction.

Interventions that assisted wound healing and symptom relief had positive outcomes for patients and improved well-being and HRQL. This was evident from patient reports that wound healing and symptom relief increased physical and social activity, allowing participation in daily activities and social interaction. Alternatively, some patients reported that surgical interventions restricted activity and mobility and contributed to reducing HRQL. These patients reported frustration and emotional problems associated with physical body changes. For example, one patient reported that after surgical intervention their “bottom looked like a meat cleaver was taken to it” (106).

Table 2.1 Synthesised categories and themes identified by both qualitative and quantitative research

Theme	Category	Qualitative	Quantitative
Health-related quality of life			
Physical impact & limitations	Physical restrictions caused by pressure ulcers	✓	✓
	PUs contribute towards lifestyle changes	✓	✓
	Adapt living arrangements to incorporate pressure ulcer impact	✓	
Social impact	Social life restricted due to physical restrictions including treatment	✓	✓
	Social isolation & loss of interest	✓	✓
	Pressure ulcer impact on personal lives	✓	✓
Psychological impact	Develop & use of coping mechanisms	✓	✓
	Support & help from family and friends is important & appreciated	✓	
	Changes to body image & self-concept	✓	✓
	Desire & struggle for control & independence	✓	✓
	Emotional problems (impact on mental health & psychological well-being)	✓	✓
	Preoccupation with pressure ulcer healing & anticipation of pain	✓	
	Acceptance of pressure ulcer & their situation	✓	
Impact of pressure ulcer symptoms	Evaluative pain descriptors	✓	✓
	Sensory pain descriptors	✓	✓
	Affective pain descriptors	✓	✓
	Pain contributed to sleep disturbances	✓	
	Symptoms interfere with activities of daily living & contribute to social isolation	✓	✓
	Wound smell caused embarrassment & distress	✓	
	Wound exudate resulted in increased need for dressing changes	✓	
	Experience of early symptoms indicate skin damage	✓	
	Skin protection & pain prevention behaviours were practiced	✓	✓
Impact on general health & consequences	Health complications & health deterioration caused by pressure ulcer	✓	✓
	Pressure ulcer & symptoms had consequences and implications to healthcare received	✓	
	Comorbid influence & impact on pressure ulcer patients	✓	✓

Table 2.1 Cont.

Theme	Category	Qualitative	Quantitative
Other impacts			
Impact on others	Dependence on others	✓	
	Perceived impact imposed by pressure ulcer on others	✓	
	Fear of being a burden to others	✓	
Financial impact	Perceived financial impact & consequences	✓	✓
Experiences of care			
Need for vs. impact of	Patient evaluations of mattresses & cushions (product evaluations)	✓	✓
Interventions & wound care	Satisfaction & benefit experienced from topical treatment	✓	
	Quality of life & satisfaction with life improved with wound healing & symptom relief	✓	✓
	Lack of resources result in inability to deliver effective wound care	✓	
	Hospitalisation for pressure ulcer had consequences on the individual	✓	
	Incongruence between patient need & clinical/nursing needs	✓	
	Pain caused by interventions and ineffective analgesia	✓	✓
	Reliance on healthcare provider to assist with wound care & daily activities	✓	
	Patients complained about time spent on treatment for pressure ulcer healing	✓	
Healthcare provider-patient	Therapeutic nurse/carer-client interaction & communication	✓	
relationships	Skills & expertise of nurses instilled hope	✓	
	Professional competence of healthcare provider had benefits to patients	✓	
Knowledge about pressure ulcer		✓	✓
Perceived aetiology	Self-responsibility attributed to pressure ulcer development	✓	
	Intrinsic factors caused pressure ulcer to develop	✓	
	Extrinsic factors caused pressure ulcer to develop	✓	
Need for knowledge	Need for knowledge about pressure ulcer development & prevention	✓	
	Need for knowledge of pressure ulcer physiologic processes & wound care/treatment	✓	

√ Represents the category and theme identified by particular research method

Table 2.2 Classification of included studies according to theme

Theme	Reference to relevant Quantitative paper	No. (% of 10 qualitative studies)	Reference to relevant Qualitative paper	No. (% of 21 quantitative studies)
Health-related quality of life				
Physical impact & limitations	(9, 104, 105, 106, 107, 155, 156)	7 (70)	(112, 157, 158, 159)	4 (19)
Impact of pressure ulcer symptoms	(9, 104, 105, 106, 107, 155, 160)	7 (70)	(112, 158, 161, 162, 163, 164, 165, 166)	8 (38.1)
Psychological impact	(9, 104, 105, 106, 155, 156)	6 (60)	(158, 162, 167, 168, 169, 170)	6 (28.6)
Social impact	(9, 104, 105, 106, 155, 156)	6 (60)	(157, 158, 159, 169)	4 (19)
Impact on general health & consequences	(9, 105, 106)	3 (30)	(157, 158, 162, 169, 171, 172, 173, 174, 175)	9 (42.9)
Other impacts				
Impact on others	(9, 104, 105, 106)	4 (40)	0	
Financial impact	(106)	1 (10)	(157, 169)	2 (9.5)
Experiences of care				
Healthcare provider-patient relationships	(9, 107, 155, 156)	4 (40)	0	0
Need versus impact of interventions	(9, 105, 106, 155, 156, 176)	6 (60)	(161, 163, 166, 177, 178)	5 (23.8)
Knowledge about pressure ulcers				
Perceived aetiology	(9, 106, 107)	3 (30)	0	0
Need for knowledge	(105, 106, 160)	3 (30)	0	0

Both community and hospitalised patients reported being dependent on healthcare providers to assist with wound care and daily activities. For some patients, the time spent obtaining information about various treatments, treatment regimens such as wound dressings, and waiting time for treatment or nurse home visit caused additional burden. The main issue for patients requiring home care for their PU was that their whole morning was disrupted because nurses were unable to provide an estimated time when they would be arriving and patients felt that they missed out on social activities because they were required to wait at home for the nurse. Hospitalisation for PU treatment had a different set of consequences. The hospital environment undermined patients' psychological well-being as patients felt captive, disconnected from the outside world, and confined and alienated.

A major issue identified from studies was evident as incongruence between patients' needs and clinical/nursing needs. What was important to patients (i.e. to sleep through the night) was not always the same as the clinical or nursing priority (i.e. turning throughout the night to relief the pressure). Some patients felt that their complaints and symptom reporting were ignored or not attended to by staff. These patients reported that when they complained of pain, their skin may have been inspected but because no damage was observed, nothing was done and consequently a PU developed. Others said that additional pain was experienced with treatments and use of some equipment (dressing changes, friction from being pulled over mattress, use of hoists and pulleys), and this was associated with poor care. Hospitalised patients also mentioned that neglect to draw the curtain during skin inspection or treatment regimen, allowing others to see the wound, left them feeling exposed and humiliated.

Nurse-client relationships

The therapeutic relationship between patient and healthcare provider and the patients' perception of their healthcare providers' competence had important implications for patients. The skills and expertise of healthcare providers instilled hope in patients and contributed to compliance with PU treatment. A healthcare provider's ability to interact with the patient in a holistic way, communicate by explaining the wound care treatment, and teaching patients' how to undertake their own wound care where appropriate, was indicative of a positive therapeutic environment. Agreement about symptoms and treatment regimen through mutual dialogue and decision making was important to patients as it allowed them to regain some control and independence over their lives. Also important was the attitude of the healthcare provider; a positive, friendly attitude contributed to forming a positive therapeutic relationship, and was equated with adequate care. Some patients felt that their healthcare provider had a poor attitude

towards them, making them feel blamed for the PU developing, a nuisance when they asked for assistance or that providing their healthcare was an effort and problem.

Perceived aetiology

Three studies presented patients' beliefs about causative factors. Some attributed self-responsibility, believing that their own actions (i.e. failure to inspect skin, reduced mobility, not reducing risk factors) contributed to the PU developing. Others believed that intrinsic factors (i.e. existing risk factors: incontinence/moisture, inability to move or walk) caused the PU, but for the majority, extrinsic factors were attributed as being the primary cause of PU development. Factors such as incompetent healthcare, inadequate use of equipment, and delays in noticing and treating reports of pain, were the main factors believed to have caused the PU. These patients were angry that their PU developed because of inadequate healthcare and felt that the PU may have been prevented if adequate care was provided.

Need for knowledge

Although some patients were aware of the various risks for PU development and attributed them as being the cause for their PU, some demonstrated a lack of knowledge and understanding about PU development and prevention, suggesting a need for greater patient knowledge. Specifically, patients need more information about PU causes, risks, prevention, the physiological processes, and the various wound care strategies and treatment interventions available. Of the patients who demonstrated knowledge about these factors, many were either SCI patients who had been educated about PU risk or people with a history of having previous PUs.

2.5.4 Existing measures to assess HRQL in PUs

No disease-specific measure of HRQL was identified to assess HRQL outcomes in PUs. The measures used were either generic measures (e.g. Short Form-36 health survey (95)) or measures not developed or validated for use with PUs but intended for other chronic wounds (e.g. Cardiff Wound Impact Schedule (CWIS) (179)). Various non-validated questionnaires (i.e. *ad hoc*) were developed to assess outcomes such as pain, comfort and exudate, although they were study-specific and had not undergone the same rigorous psychometric evaluation as other established PRO measures.

2.6 Discussion

This is the first comprehensive and systematic review of the PU and HRQL literature that has identified HRQL factors that are important to people with PUs. The studies

identified reported the range of HRQL domains that are impacted by PUs and PU interventions. The review shows that PUs impact on a range of HRQL domains, however the size of the impact is not covered by this review. In addition, a range of treatment side effects and quality of care problems were identified.

Not only do PUs impact HRQL, some patients believe that they have become a burden to others, perceiving assistance with wound care and daily activities as an imposition, causing them additional anxiety and worry. The most common belief about the cause of Pus was that they were a direct result of poor quality healthcare. Further, it was evident that unless patients had a history of PUs or belonged to a group at high-risk of PU development, there was lack of knowledge about PUs, suggesting a need for more patient information and education about PU risks and treatments.

This synthesis concurs with reviews undertaken to investigate the impact of leg ulcers on HRQL (152, 180). The most commonly occurring findings for people with leg ulcers and PUs included: restrictions on physical and social aspects of patients' lives (152, 180); experience of emotional problems such as low mood, anxiety, frustration, anger, depression (180) and disparaging self views of physical appearance (152, 180); perceptions of the therapeutic relationship (152); and severe and constant pain. Patients with PUs commonly reported pain from the wound site, during dressing changes, and with use of pressure-relieving interventions and other medical equipment. Not only was pain a central and ever-present feature of having a PU, mirroring findings by Briggs and Flemming (152), pain was debilitating and contributed to restricting physical and social activities, sleep disturbances, difficulties with ambulation and inability to assume comfortable positions. A key issue for patients with PUs was that many complaints of pain made were largely unrecognised and ignored by healthcare providers (9, 107, 181); a surprising finding considering that sensory pain could be an indicator of possible skin damage and a pre-empting of a PU developing in the future. Thus, while it would appear that PUs cause patient's substantial pain, it is underestimated by nurses and other healthcare providers or potentially overlooked because of existing comorbidities.

In contrast to the wider leg ulcer literature, PUs contributed towards lifestyle changes, adapting living environments to incorporate wound care (i.e. nurse visits, hospital admissions) into their life and schedule (9, 156), and an inability to perform normal day-to-day activities. Patients with PUs were concerned about the time taken for wound healing and experienced conflict between what one wanted to do with what one should do to promote healing. Many patients had difficulty accepting their PU while others

developed coping strategies and learnt to accept their PU and situation. For patients able to accept their PU, this had positive implications on their psychological health.

Leg ulcer patients' treatment issues were mainly associated with difficulties with compression (152, 180). For patients with PUs, regardless of type, PU interventions caused patients substantial burden and consequences including physical and social restrictions, continued discomfort/pain, sleep disturbances, loss of appetite, feeling powerless, and additional emotional problems (e.g. low mood, hopelessness, anger). In addition, the impact of hospitalisation was not a feature for leg ulcer patients (152, 180). However, both wound groups expressed the importance of a positive therapeutic relationship with their healthcare providers (152). A positive relationship was perceived as a means of providing healthcare required for PU healing, a way of improving psychological well-being by providing a sense of support and hope, and an opportunity to engage patients in communication about their PU treatment, improving both their self-concept and independence. On the other hand, a poor relationship was detrimental to patients and an additional consequence of the PU experience.

The strength of this review is the robust methods used. Information derived from this review stems from a combination of both qualitative and quantitative research. It has been argued that both methodologies have their own strengths and weakness but they are increasingly advocated and used in combination in research (139). Including studies that have used either method for data collection has provided a complimentary, deeper interpretation of PRO data on this complex phenomena. For example, pain has two elements; physical and emotional. Quantitative measurement provided information about physical aspects of pain such as pain intensity and severity, but provided little about the emotional and psychological impact of pain. Qualitative measurement, on the other hand, provided information about the emotional and psychological consequences of severe pain in a descriptive way. It also provided a greater overall account of the impact PUs have on HRQL. Quantitative measurement was mostly focused around intervention evaluation, specifically in terms of comfort, sleep and pain, failing to identify many of the HRQL issues that are important to patients. These observations highlight the need for quantitative measures that can accurately depict impact of PU and PU interventions on patients' HRQL.

Only studies that presented independent PU results and obtained HRQL issues by direct patient reports were included. Many studies did not differentiate participants by wound aetiology, recruiting mixed wound samples, or obtained outcomes by a healthcare provider or carer. In dementia, a discrepancy between patient-proxy reports

of QoL has been observed (182). Further, bias is introduced when, for example, pain measurement is obtained by the same person who administered a dressing change, then questioned the patient about any pain experienced; the method used in some studies identified. Including only studies with independent PU analyses and outcomes obtained directly by patients ensured the HRQL issues identified from this review were in fact reported by patients with PUs and about the impact of PUs and not other wounds. Further, this review had international collaboration which allowed for translation of non-English papers from settings outside the UK/US; a limitation of other reviews has been the under-representation of non-English speaking cultures and languages. The emergent themes might be transferable across other cultures and languages, although this assumption would need to be validated.

To reduce selection bias, all research designs and studies yielding both statistically significant and insignificant results were included in data analysis. Methodological and clinical heterogeneity were not criteria for exclusion. Although differences in study design and between study interventions and outcome measures were present, it was deemed appropriate to pool data, regardless of known heterogeneity, in order to identify all HRQL outcomes important to patients. Attempts were made to reduce bias, but a limitation of this review was that only one reviewer extracted findings. Data extraction was repeated until saturation and categories and themes were reviewed by the collaborative group to reduce reviewer bias. In addition, the review findings were not qualified in terms of weight given to the majority view (i.e. greater emphasis on findings featured in more studies) as has been done in other qualitative synthesis (152). This was to ensure that all issues important to patients were represented.

A number of general observations about study quality were made. The qualitative studies demonstrated good quality as appraised by QARI (149); seven of 10 studies scored ≥7, indicating congruity between research methodology and research objectives, and between research methodology and analysis, interpretation of results and final conclusions. However, there were limitations associated with the conduct and reporting of quantitative studies. Seven (25%) cohort and cross-sectional studies were excluded as more than 20% of the study sample was excluded from analysis. Some studies had used generic or measures not developed or validated for PUs. The quality of the 12 RCTs was poor, limiting the value RCTs provided in furthering our understanding of the impact of PUs on HRQL: formal sample size calculations were not reported, analysis was not intention to treat, primary end points not stated; or more than 5% of the recruited sample was excluded from analysis, therefore all RCTs were excluded. In addition, methods of administration were poorly described, HRQL was

usually of secondary interest, and all RCTs used visual analogue scales (VAS) or ad-hoc questioning methods rather than utilising existing PRO measures. Consequently, due to poor quality, formal statistical analysis of HRQL outcomes was not performed. Further, the impact of individual PU interventions was not determined as RCTs of PU intervention effectiveness were not designed to assess HRQL outcomes appropriately so comparison between individual interventions was not possible.

There is need for comprehensive, valid and reliable outcome measures to assess HRQL in the PU field. Measuring HRQL would provide information to support decisions for resource allocation, know treatment benefits from the patient perspective, and increase patient compliance with treatment, as the focus would be shifted from wound management to individual holistic healthcare. Further, measuring HRQL would promote patient-centred communication, identify issues that are important to patients on an individual level, and reduce patient suffering, in turn increasing HRQL. Assessing HRQL would provide healthcare providers with greater understanding of specific patient needs and better understanding of the PU experience in order to provide the best possible care to patients. In light of the findings from this review, national guidelines need to be reviewed with consideration of the impact of PUs on HRQL to improve PU healthcare in the future.

2.7 Summary of Chapter 2

A systematic review of the PU and HRQL literature was undertaken to identify and synthesise common issues reported directly by patients about living with a PU. PUs cause patients significant impairment and burden. Major issues identified relate to severe pain, patients' views and concerns ignored by healthcare professionals, early warning signs (e.g. pain) do not prompt action, treatments increase discomfort and pain, and the physical, social and psychological aspects of patient need are not met. This review identified HRQL domains specific to PUs and highlights the potential need for outcome measures that can accurately depict the impact of PUs on HRQL. However, the review findings were not qualified in terms of weight given to the majority view, therefore the next stage of this research is to confirm the relative importance of the findings (working conceptual framework domains) with patients with PUs.

Chapter 3

DEVELOPMENT OF A PU-SPECIFIC CONCEPTUAL FRAMEWORK OF HRQL

3.1 Overview

Chapter 3 reports the methods and results of a qualitative study including interviews with patients with PUs and expert opinion to formulate a final conceptual framework of PU-specific HRQL.

3.2 Background and aims

Evaluating outcomes such as HRQL is particularly important and relevant in PUs where the condition and associated interventions pose substantial burden to patients. Although the management of PUs has received considerable attention in the literature, far less has been given to the impact PUs have on HRQL. A systematic review of the HRQL and PU literature (Chapter 1) identified relatively few studies that have considered HRQL outcomes important to patients with PUs. The review highlighted HRQL outcomes that are unique to the impact of PUs although these outcomes are currently not systematically included as outcomes in clinical trials. Therefore the PU literature is unconvincing in terms of robust evaluation of the impact of PUs on HRQL.

The most appropriate way to assess HRQL is with the use of PRO instruments but PRO instrument development is time consuming and demanding as several stages must be undertaken to ensure that they are carefully designed and fit for purpose. This is particularly important since PRO data is now an integral facet impacting patient care, policy decision making and healthcare delivery, and the adequacy of these decisions depends directly on the scientific quality of the PRO instruments used (87). As such, of primary importance is consideration of conceptualisation (clearly defining constructs or variables measured) and operationalisation (content or items) of the instruments' scales; a minimum prerequisite for acceptance of a PRO instrument (86). A conceptual model should guide what a PRO instrument measures, how it should be measured, and provides a context for interpreting findings (82). Therefore scale construction should be underpinned with a strong conceptual base from the outset utilising three sources: existing literature, patients and clinical experts (86, 131, 132). Grouping items statistically or thematically through expert opinion alone does not ensure that items in a group (scale) measure the same construct or map out a variable in a clinically

meaningful way (82, 87). No amount of statistical manipulation can compensate for poorly conceptualised and operationalised variables.

Developers of PRO instruments have often utilised 'top-down' approaches (i.e. literature review, content from existing measures) for conceptualisation and content development. A risk of this methodology for developing disease-specific measures is that they may not reflect those aspects of HRQL that are most important to patients with the underlying condition. Not utilising patients' perspective in the development of new instruments may pose a threat to the validity and responsiveness of scales to estimate accurately the effect of a disease or to detect clinical change as outcomes relevant to patients may not be included (183, 184).

Various methods have been used to perform qualitative data analysis. Qualitative methods are useful for obtaining direct patient accounts of the illness experience, its impact on human suffering, and the impact of related treatments received (185). Qualitative research produces large amounts of textual data which is explored inductively (i.e. content analysis) to generate categories and likely explanations of the relationships between categories and their properties. Data analysis often takes place alongside data collection to allow the refinement of interview questions based on emerging information, new avenues of inquiry to develop, and confirmation of categories identified in the literature.

There is growing application of qualitative research methods, both in psychology and health sciences research, including the use of numerous qualitative methods to analyse textual data. An analytic method common across many qualitative approaches is thematic analysis; an approach dealing with data that involves the creation and application of 'codes' to data (186). Since FDA guidance (86), qualitative methods have been more scrutinised but unlike quantitative analysis, there are no clearly defined methods of analysis for qualitative data. Approaches to analysis should be determined based on the main focus and aims of the analytical process.

3.2.1 Aims and objectives

This study was conducted to:

- examine what determines HRQL for people with Pus
- identify HRQL domains and their components specific to and important to patients with any grade PU
- identify whether the impact on HRQL is the same in relation to PU severity (i.e. superficial vs. severe), location (i.e. torso vs. limb skin sites) and interventions

- refine and further develop the working conceptual framework to produce a final PU-specific HRQL framework (i.e. conceptualise and operationalise variables)

3.3 Methods

3.3.1 Development of a conceptual framework

A PU-specific conceptual framework of HRQL was developed by combining top-down (existing literature) and bottom-up (qualitative data) approaches.

Top-down approach

In the previous chapter, a top-down approach was used to develop a *working* conceptual framework based on a systematic review of the PU and HRQL literature. The review identified common HRQL components specific to PUs and PU interventions, and general issues associated with having a PU that were grouped into relevant HRQL domains. This produced a list (working conceptual framework) including nine broad provisional domains associated with PU occurrence (physical, social and personal relationships, psychological, general health, symptoms, perceived impact on others, financial) and wound management (need for vs. impact of PU interventions, nurse-client relationships).

Bottom-up approach

The provisional domains formed a topic guide for interviews (Appendix 3.1). The working framework was subsequently revised and extended on in light of the bottom-up data obtained from qualitative interviews, followed by expert opinion and review to finalise the framework. The bottom-up approach involved qualitative work to elicit information pertaining to the *impact* of PUs on HRQL, define specific domain components (content), and confirm the importance of identified outcomes with a group of patients with PUs. Individual face-to-face, semi-structured interviews, guided by the topic guide, with the addition of a series of open-ended questions to elicit any new relevant information (e.g. is there anything else that you want to add about how your PUs may have impacted you?), were undertaken. Specifically, information sought from interviews related to: how having a PU impacted on life from the perspective of the sufferer; which HRQL components are important to patients with PUs and whether some are more important than others; how PU interventions impact HRQL; and information about the relative importance of having HRQL outcomes assessed as part of healthcare. In addition, patients with PUs commonly have severe comorbidities. To ensure reported issues were PU-specific, clarifying questions were asked (e.g. Do you think [that] is only because of your PU or resulting from a combination of things?).

The individual face-to-face method was chosen to encourage participants to reflect and speak openly in personal terms about their PU experience and the impact it may have had on HRQL. The individual method was chosen rather than focus groups to maintain confidentiality, prevent collaboration in responses, and to allow an environment where the participant could freely and openly talk about their experience. It was also considered that focus groups, although perhaps more cost effective and less time consuming, would have been difficult for participants to attend as many are elderly, chronically ill or have mobility problems. Details regarding participant characteristics, the PU (i.e. PU grade, location, duration) and any treatment received were requested verbally either from the patient of from the treating nurse.

3.3.2 Participants

Thirty two patients with PUs from seven acute and primary care settings in England and Northern Ireland during December 2007 to October 2008 consented to participate. It was considered important to include patients from both acute and primary care due to the high prevalence of PUs in both settings. Sites active in tissue viability research with a track record of multi-centre research were invited to participate. A purposive sampling method was devised, with sampling of patients targeted to key factors to reflect the range and diversity of the target population, including age (under and over 70 years), PU severity (superficial and severe PUs) (Table 1.1) and location (torso and limb sites), healthcare setting, and experience of different PU treatments. A minimum of five patients per key factor were consecutively sought and found.

Adult patients with a PU of any grade or location were included if they were aged ≥18 years, from hospital, rehabilitation or community settings and under the care of a specialist tissue viability nurse (TVN) or care-of-the-elderly team, able to reflect on and share their experience and provide informed consent to participate. Patients were excluded if they did not currently have a PU or one that had healed within the previous three months, or who were unconscious, confused, cognitively impaired or unable to speak English. Including patients from acute, community and rehabilitation settings allowed collection of data on a wide range of patient experiences. The 3-month recall period was used to reduce recall bias as longer periods could obscure individual recall.

3.3.3 Recruitment and consent procedure

Specialist nurses at participating hospitals and community services identified and approached eligible patients and provided them with study information that included details about the rationale, design and personal implications of the study, and an ‘agree

to be contacted by the researcher' form. Following information provision, patients had as much time as they needed to consider participation. The TVN, researcher (CG), and the research supervisor were available to answer any questions that patients had about the study. After receiving a signed agreement to be contacted, CG telephoned the patient to arrange a time for the interview. CG provided further information about the study and interview process, and answered any questions raised before gaining verbal consent and arranging an interview at a mutually convenient time. For in-patients who could not be contacted by phone and were expected to be in the hospital during the interview, with the patient's permission, the TVN liaised with CG and the patient and arranged a mutually convenient time for CG to visit the patient on the ward.

CG interviewed patients at their home or on the hospital ward, as determined by the patient's circumstances and preferences at the time of the interview. Before the interview, each patient was given a further verbal explanation of the study, informed that the interview would be audio-recorded but that all identifiable information would remain anonymous, reminded that they could withdraw from the study at anytime without it affecting their care, and then invited formally to participate. Patients were given an opportunity to ask any questions and then if they agreed to take part, they were asked to sign a consent form. For any patients with difficulty writing but who fully comprehended, an audio-recording was taken of the verbal agreement. The right of the patient to refuse consent without giving reasons was respected. Patients remained free to withdraw from the study at any time, again, without giving reasons and without prejudicing any further treatment.

Ethical approval was obtained by the North West Research Ethics Committee prior to study commencement. This study recruited elderly and highly dependent patients considered as vulnerable. Ethical issues surrounding these potentially vulnerable patients were addressed through the study design including a thought out consent process and the use of one-to-one interviews to provide a flexible and supportive interview environment.

3.3.4 Data Collection

Semi-structured interviews were useful for suggesting a topic to patients to discuss freely, and then using specific probes to elicit more focused information. This method allowed for detailed investigation of patient's personal perspective in order to gain understanding and clarity of complex phenomena; the impact of PUs on HRQL. Interviews followed a topic guide that was revised as new information emerged deductively. Issues that gave insights into the impact of PUs and PU interventions on

HRQL were incorporated into discussion of subsequent interviews in response to emerging data, allowing confirmation of new issues with successive patients. The working framework was revised and extended based on qualitative data. This process ensured that the research remained participative. Interviews were undertaken until saturation; the stage where no new information appears to emerge during data coding, that is when no new properties, dimensions, conditions, actions, interpretations or consequences are seen in the data (187, 188). Interviews lasted a mean of 42 minutes, and were audio recorded, transcribed verbatim and content analysed by the researcher (CG). No new information was obtained after interview 25.

3.3.5 Quality Assurance

To ensure quality of data collection and that CG was not leading patients, a researcher with experience of undertaking and supervising qualitative research (see acknowledgements) listened to the first three recorded interviews, read the transcripts and fed-back on performance (e.g. discussed instances in the transcript where CG could have probed or clarified further). Throughout interviews, CG took notes, and where required, fed back issues to patients to ensure comprehension of responses.

3.3.6 Analysis

In the literature, qualitative interviews with patients are promoted as a valid way of generating items for new PRO instruments. However, the method of analysing data generated for this purpose is confusing, with various qualitative approaches used but without consensus about which method is most appropriate. Considering that the study aim was to build on an emerging theory, validate, and possibly extend a framework developed from the literature, qualitative approaches that develop theory from data, such as Grounded theory, were not deemed appropriate; the theory would not be truly grounded as at least some issues related to the impact of PUs on HRQL were known. Instead, qualitative approaches involving inductive reasoning where initial coding starts with a theory were considered. The analysis approach needed to allow the development of a coding schema to be derived from theoretical concepts of HRQL generated from the literature as well as additional HRQL components identified from patient-reported textual data. These two sources would provide the basis for generating new codes or modifying codes developed by induction.

A criticism of inductive methods is that the coding and sorting of emerging categories (codes) may be constrained by existing knowledge of the relationships between some categories and their properties. However, the intention here was to provide support for

work that has been done before and to formulate a final PU-specific conceptual framework of HRQL that incorporated previous research and new patient-reported data. As such, the directed content analysis approach (189) was used as it is intended for extending a conceptual framework. This method uses structured data collection and an explicit analytical procedure informed by *a priori* reasoning based on existing theory (relevant research), thus making it a suitable methodology for the objectives of this study.

In content analysis, a large amount of textual data can be analysed. The process involves interpreting data on a higher logical level by creating codes, categories and themes (i.e. condensing text by aggregating common data) in relation to the context. A category is a group of content that shares a commonality (190). Categories should be exhaustive and mutually exclusive with all relevant data fitting into a category and not excluded due to lack of a suitable category. A category is a thread throughout the codes; a descriptive level of the aggregated content. Linking the underlying meanings in categories creates themes. A theme is a thread of meaning through condensed meaning units, codes or categories, on an interpretative level (191).

Generating a series of main themes and subdividing them into a succession of related components (categories and codes) was the systematic approach of this study. Ongoing analysis of interviews during data collection was undertaken to allow provisional HRQL themes to be refined and to inform the focus and prompts for subsequent interviews. First, transcripts were read while listening to the recording to confirm accuracy of transcription and to get an overview of the data collected. Any first impressions and interpretations were noted including any thoughts about main HRQL components, general feelings about the interview, and audio cues from the patient that would be lost in transcription. This process allowed CG to become immersed in the data and note key ideas and recurrent components.

Following this, two researchers (CG and JF) conducted thematic content analysis manually for textual data from the first four interviews. This involved identifying and examining evidence (text) from the transcripts for any HRQL issues and coding to a provisional coding schema developed using a combined inductive (codes arising from transcripts) and deductive (codes developed from the interview topic guide) approach. The provisional coding schema consisted of six broad themes (first level codes): physical functioning and limitations; social activities and relationships; psychological well-being; self-concept; pain; and exudate and odour. The provisional coding schema was refined to reflect the emerging data during subsequent stages of the analysis; data

collection and coding were conducted iteratively in multiple rounds of interviews so that subsequent data collection was informed by earlier coding and confirmed in later interviews. Note-taking, constant comparison and identifying emerging categories until saturation, allowed the data analysis to be responsive to the data. Specifically, the categories suggested in the transcript of the first interview were identified. Coding categories in the second interview were done with the first interview in mind. Subsequent coding was done with the emerging theory in mind and so on. The emerging theory was developed as new categories emerged from the data. This process of constant comparison ensured that the theory emerged from the data rather than from any preconceived ideas. Further, continuous analysis ensured that the emerging coding schema was complete or saturated, and captured the specific components within the first level codes and emerging HRQL themes.

In the second round of coding, the researchers independently assigned codes to 30% of transcripts. An inclusive approach was taken, adding HRQL components (codes) to reflect the emerging nuances in the data rather than reducing data to a few numerical codes (i.e. first level codes). This type of textual analysis and consensus coding allowed the researchers to identify themes and their specific components inductively as they emerged from the data. Further, any new aspects could be tested with subsequent patients to confirm their relevance and importance (i.e. confirm new issues are important to more than just one patient). Issues mentioned infrequently were discussed by an expert group (see Expert Review section below) and components considered not clinically relevant were either excluded or consolidated with related components (e.g. emotions such as irritated and distressed were consolidated with ‘negative mood’) rather than retained as separate themes in the conceptual framework.

The extraction process involved coding any patient phrase or statement to the appropriate code. If the same text was describing two independent issues, they would be coded to both codes. The two researchers reviewed the independent sets of coding, including any gaps in information or discrepancies, agreed the final framework by consensus and discussed any discrepancies in coding with a third researcher (JN). The intention was to establish a common conceptual understanding of the codes between the researchers and generate a consolidated coding schema. The final schema consisted of main domains divided into sub-domains. Sub-domains comprised a number of components describing slightly different characteristics of the domain. For example, ‘irritated’ and ‘distressed’ were components of sub-domain (variable) ‘mood’. By analysis of the twenty fifth transcript, no new themes or categories emerged in the data and therefore the conceptual framework was considered saturated.

Reliability

Consistent with methods used by others (192), a Kappa statistic (193) was performed to determine the extent of agreement between the two researchers in applying the final coding schema; a Kappa >.70 is considered acceptable for inter-rater reliability (193, 194). Kappa's (SPSS 13.0) were 0.924 and 0.908 for coding main and sub-domains, respectively, showing almost perfect agreement between the researchers (195).

Cross-case analysis

Following manual coding, transcripts were entered into QSR NVivo 7 (2007) and the final coding schema applied to all text data. This was undertaken to enable complex organisation and manipulation of textual data for the purpose of performing cross-case analysis. Such analysis was important to establish the extent to which HRQL domains were consistent across the patient sample as well as to explore any differences. An exploratory cross-case analysis was performed by mapping and interpreting the data from case charts (by patient across all HRQL domains); a method adapted from Framework Analysis (196, 197). The charts were used to search for any patterns and associations in the data; components that occur repeatedly in the data, unique to each patient factor (i.e. age, gender, PU location and severity, and comorbidity) as well as similarities or variations in individual experiences (197, 198).

Expert review

Following data analysis, a multidisciplinary expert group consisting of seven tissue viability specialists, one chronic pain specialist and five outcome methodologists (see acknowledgments section for membership) reviewed the final conceptual framework. As the intension was to develop a HRQL PRO measure, a distinction between HRQL outcomes versus other contributory factors which may affect HRQL (i.e. behaviours, satisfaction with healthcare) was made. The relative importance of outcomes in terms of how commonly they were reported (i.e. the number of times a verbatim was coded as the same code, for example "I feel pain when I move" x12) as well as clinical relevance was determined. A final PU-specific HRQL conceptual framework that adequately defined, conceptualised and operationalised constructs (specific components) was produced.

3.4 Results

3.4.1 Participants

Participants were selected to ensure representation of the PU population, including targeting to factors such as age, gender, setting and comorbidity. Thirty patients consented to participate. They ranged in age from 22 to 94 years (mean age 62.2 yrs),

of which 18 (56%) were male and 19 had other chronic conditions (e.g. 8 with spinal cord injury (SCI) and 3 with multiple sclerosis (MS)). Patients represented hospital/rehabilitation (n=17) and community (n=13) settings, PU severity (12 superficial, 15 severe, 3 mixed severity), duration (from a few days up to 4 years) and skin site. Sacrum (n=15) and heel (n=14) PUs were the most frequently occurring with other sites including the lower back, buttocks, ankles, hips, back of head, and elbows (Table 3.1).

3.4.2. Conceptual Framework

The final PU-specific HRQL conceptual framework includes four domains and 13 sub-domains (Figure 3.1). All domains and descriptive components are described below.

3.4.2.1. Symptoms

Pain and discomfort

Patients with PUs commonly reported some form of pain or discomfort. Pain interfered with sleep, general movement, walking/sitting, daily activities and socialising: "I can't face [activity] because of the pain" (62mG4C)[5]. The pain could be "unbelievable" and they would "do anything to be rid of it", and some even perceived it as a "punishment" for the PU. Patients reported anticipatory pain (e.g. during dressing changes) which caused anxiety, while others coped by comparison with worst pain ever experienced.

Commonly, PU pain was described as "sore", "stabbing", "burning", "throbbing" or "stinging" - words often associated with both nociceptive and neuropathic pain. A large proportion of patients experienced severe and persistent pain that varied in duration (i.e. constant, never ending, intermittent). Most commonly, patients felt pain in and around the PU, shooting up through various parts of the body (e.g. up the leg or back), upon contact, pressure (e.g. when sitting), upon movement, or with various PU treatments (e.g. dressing changes, debridement) or medical equipment (e.g. hoists).

Pain caused discomfort and interfered with one's ability to participate in daily life, move/walk or assume comfortable positions, or engage in sexual activity. Some patients felt that pain made their manual handling and treatment difficult for healthcare providers and carers. The few patients who did not report their PU pain to a healthcare provider claimed it was difficult to describe, feared being a burden or nuisance, thought nothing could be done about it, or perceived it to be the norm.

[5] Annotations refer to participant's age, gender (m or f), PU severity: superficial (G1/2) or severe (G3/4) and healthcare setting: community (C) or hospital (H).

Table 3.1 Participant Characteristics

Age	Gender	PU (Grade) & location	PU duration	PU recurrence	Other condition	Setting	PU on admission	Interventions received
38	Male	(3)left elbow; (3)left hip	3mths	Yes	SCI	Acute rehab, PU treatment	Yes	VT,M,D
41	Male	(4)sacrum	3-4yrs	Yes	SCI, SB	Acute rehab, PU treatment	Yes	M,C,D,VT,LT,NT
45	Male	(3)heel; (4)repaired hip	8wks	Yes,24 yrs	SCI	Acute rehab, PU treatment	Yes	D,M,C,SR
49	Male	(4)left buttock	2yrs	Yes	SCI	Acute rehab, PU treatment	yes	M,C,SR,VT,LT,D
52	Male	(4)right hip	9mths	No	SCI	Acute rehab, PU treatment	Yes	M,C,D,VT
55	Male	(4)left buttock	1.5yrs	No	MS	Community	NA	M,C,D
59	Male	(1)lower back; (3)sacrum; (3)both heels	3mths	Yes	SCI	Acute Vascular, PU treatment	Yes	M,D,WC
60	Male	(4)sacrum	6mths	No	SCI	SCI Rehab	No	M,C,D,VT
61	Male	(2)sacrum	8wks	No	unk	Acute rehab	No	M,D,WC
62	Male	(1)both hips; (2)sacrum	9 yrs	Yes	unk	Acute general surgery	Yes	D,M,C,VT
62	Male	(4)sacrum	3mths	No	Renal	Community	NA	M,D,VT
64	Male	(1)right heel; (1)right hip; (2)buttocks/sacrum	1mth	No	Neurofibromatosis	Acute neurosurgery neurofibromatosis	No	M,D
71	Male	(2)heels	5wks	No	Elderly	Acute neurosurgery	No	D, creams
71	Male	(2)sacrum; (2)buttocks	2 yrs	Yes	Lymp	Acute lymphodeoma	Yes	M,C,D
82	Male	(4)left buttock	18mths	Severe PU	Tumour	Community	NA	M,C,D
88	Male	(3)left heel; (3)right heel	10mths	No	Pneumonia	Community	NA	C,D
88	Male	(2/3)right heel	6mths	No	Elderly	Acute general surgery	Yes	M hospital
89	Male	(2)left heel	7mths	No	Elderly	Long-term care	No	M,D

Age	Gender	PU (Grade) & location	PU duration	PU recurrence	Other condition	Setting	PU on admission	Interventions received
22	Female	(3)sacrum; (4 left thigh	2.5mths	Yes	SCI, lymp	Acute rehab, PU treatment	Yes	M,D,VT,NT
43	Female	(3)sacrum	5wks	No	Vascular	Acute Vascular	No	M,D
52	Female	(1)lower back; (4)sacrum	1yr	Superficial	MS	Community	NA	M,D
52	Female	(1)both heels; (1)sacrum (3)back of head	16wks	No	unk	Acute ICU	No	M,C,D,VT
55	Female	(4)right heel	3yrs	No	SB	Community	NA	D
56	Female	(3)right heel	3mths	Superficial	Blood clots	Community	NA	M,WC,D,LE
59	Female	(2)sacrum	3mths	No	COPD	Community	NA	M,C,D
59	Female	(2)sacrum; (4)both ankles (4)both thighs	13mths	PUs over 12 yrs	MS	Community	NA	M,C,D,LT,NT
64	Female	(3)right heel	4mths	No	Stroke	Acute stroke unit	No	M,C,D
67	Female	(1)heel; (3)heel	4wks 3.5mths	No	Rheumatoid arthritis	Community	NA	M,D,LE
71	Female	(2)left buttock	6mths	No	Incontinent	Community	NA	Cream/lotion
79	Female	(1)right heel; (2)left heel	4wks	No	Elderly	Rehab, fall	No	M,D,LE
82	Female	(2)sacrum	3wks	Yes	Elderly	Acute, fall	No	M,D,lotions
94	Female	(1)left heel	Few days	No	Hip surgery	Community	NA	None

M Mattress; D Dressing; WC Wheelchair Cushion; C Cushion; VT VAC Therapy; NT Nutrition Therapy; LT Larval Therapy; SR Surgical Repair, LE Leg Elevation; wks Weeks; mths Months; yrs Years; SCI Spinal Cord Injury; MS Multiple Sclerosis; Chronic Obstructive Pulmonary Disease COPD; Lymp lymphodeoma; SB Spina Bifida; rehab Rehabilitation ward/facility; ICU Intensive Care Unit; unk Unknown; NA not applicable

Exudate and odour

Exudate and odour were problems, particularly for patients with severe PUs. Both symptoms interfered with daily life, intimacy and being close with others, contributed towards social problems such as self-imposed isolation or choosing not to take part in social activities, and caused emotional distress such as feeling self-conscious or embarrassed. Exudate increased the need for frequent dressing changes and/or caused the dressing to come off, leaving the wound exposed or staining clothes. Some patients reported that frequent dressing changes helped to contain leaking and odour thus reducing symptom bother.

3.4.2.2. Physical Functioning

Mobility and daily activities

Patients reported that PUs restricted movement and mobility and limited activity. These restrictions contributed to the inability to participate in daily activities such as preparing meals, shopping, getting dressed, doing house chores, undertaking self-care and hygiene, and enjoying hobbies and interests. Activity and going out were difficult for patients, restricting their daily life: “I can’t do things like before” (67fG1/3C), “stopped me doing what I want to do” (55fG4C), and “life put on hold” (52mG4H), because of limitations in movement and mobility. Enforced bed rest for PU treatment made travelling and planning trips at the weekend difficult, further restricting participation.

Patients described PUs as inconvenient and troublesome, with PU symptoms interfering with mobility and daily activities. Repositioning or movements were at times uncomfortable, difficult or caused severe pain. Some patients were unable to walk because of oozing exudate. While the majority of responses were negative, some patients found ways around their physical limitations and occasionally perceived them as an opportunity to pursue new hobbies or learn new skills (e.g. using a computer).

General malaise and sleep

PUs and treatments caused problems with quality of sleep, with some patients having to learn to sleep in new positions. For some, appetite was reduced due to the appearance, smell or pain of the PU. Others complained of feeling unwell, tired, fatigued or lacking energy: “I haven’t got the same go in me” (67fG1/3C), making it difficult to carry out normal daily tasks or accomplishing less than normal.

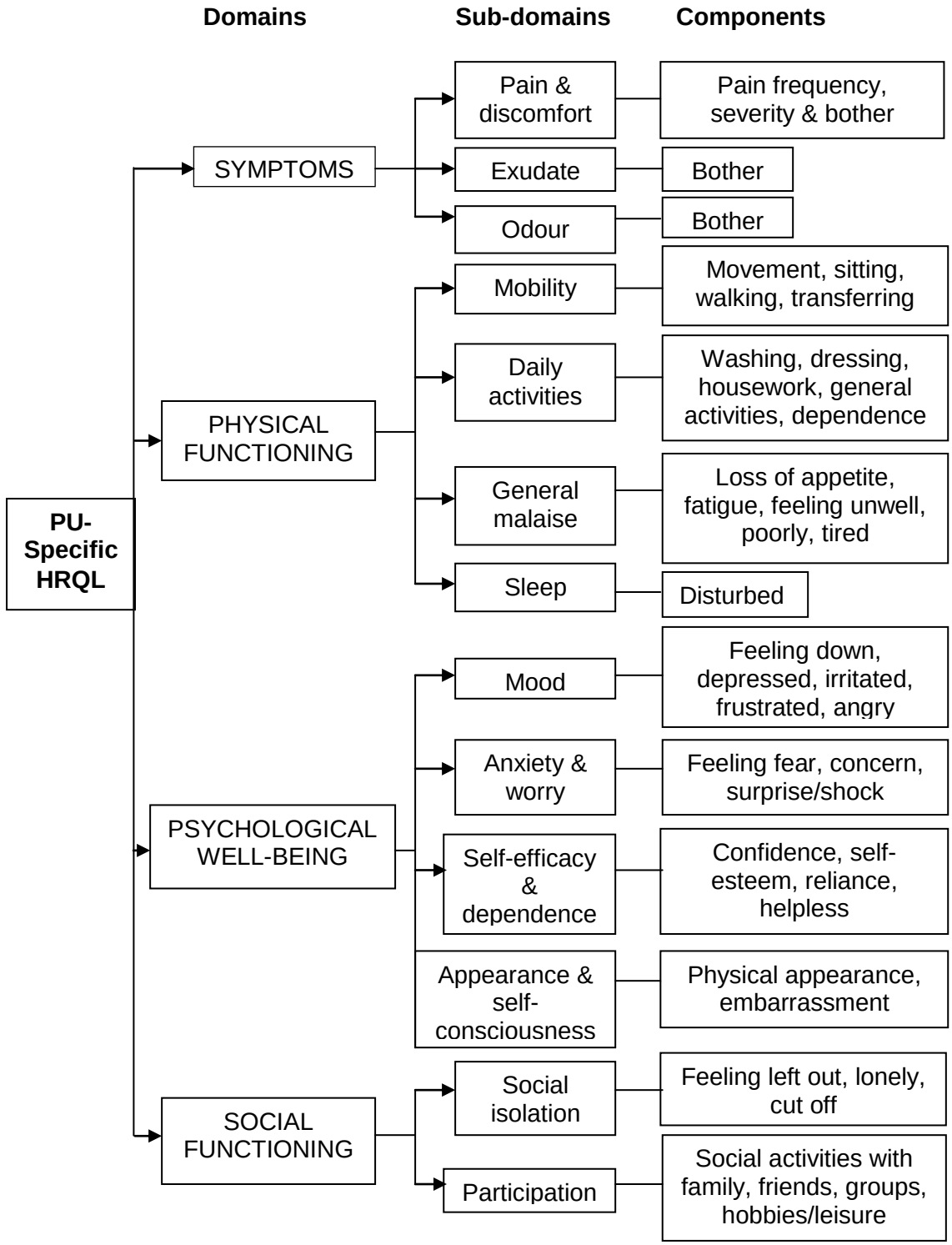


Figure 3.1 PU-Specific HRQL conceptual framework

3.4.2.3. Psychological Well-Being

Mood

PUs and PU treatments have a substantial affect on psychological well-being. All but three patients experienced emotional distress including mood changes such as feeling down, miserable and depressed: “I get really down because I’m fed-up with the time its

taking to heal...it's never ending" (56fG3C). Others reported feeling angry, irritated or annoyed that "I can't do what I want to do" (55mG4C and 59mG1/3H) or that "I have a PU and it needs to be treated" (82mG4C). Feelings of hopelessness and helplessness were usually a result of perceiving that "they've [nurses] tried everything and it's just not getting any better" (59fG2C). Some felt despondent about PU-associated limitations, perceiving this to be out of their hands: "Why the hell has this happened…it's not fair…on top of everything else" (41mG4H). Extreme frustration was also common, particularly when PUs prevented patients from doing things for themselves, moving as normal or doing things they enjoyed (e.g. participating in hobbies or leisure activities). Ineffective treatment or pain relief, the length of time spent undergoing treatment, and feeling there was nothing that could be done further contributed to frustration.

Emotional distress was lower when there was evidence that the PU was improving since this meant regaining independence and being able to do more (i.e. walk again), making patients feel happier and more hopeful and optimistic. Negative affect was also reduced when the patient had realistic expectations about prognosis, such as accepting that the PU would take a long time to heal and require extensive treatment: "you know the dressing has to be changed every day and so on, so it's just a bit of a hassle really but that's the way it is" (62mG4C). For some patients, negative emotions resulted from the perceived failures of healthcare providers.

Anxiety and worry

Anxiety concerning the PU and wound care were both general and specific. Patients worried about infection, medical complications, the PU getting worse or not healing, prognosis, and recurrence. Other worries concerned the dressing coming off (i.e. exposing the PU or causing exudate to leak), time taken to heal, and about being a burden on others. Additional worries related to healthcare and competence of healthcare providers, how the PU was being managed, and anxieties specific to being hospitalised for PU treatment (i.e. length of hospitalisation, possible neglect of the PU, potential for hospital equipment causing pain or PU deterioration).

Self-efficacy and dependence

Loss of independence was associated with loss of control over one's life. Patients felt as if their life had been "robbed", since it now revolved around the PU and PU care. Patients who had different nurses attending to their PU felt that, although it was their body, they could do nothing about who saw it, touched it, attended to it, or when any of this occurred. Patients reported tension between what one should do to promote PU

healing (i.e. medical advice) versus what one would prefer to be doing, which involved weighing up pros and cons for doing something that could be detrimental to the PU: "You know you shouldn't but you do it [going out instead of bed rest] anyway" (41mG4H). Fear of being a burden was a major component of loss of independence and many felt guilty about the additional work imposed on significant others.

Appearance and self-consciousness

The physical appearance and severity of the PU was frightening, disgusting, gruesome, and shocking to patients: "To think I have that on my body…that big black gaping hole…I wasn't expecting that" (52fG1/4C). Others were concerned about the size and depth of the PU, particularly if bone was exposed. Seeing the PU allowed insight into its severity and the monitoring of progress: "It's encouraging to hear that my PU is improving but it's better yet to see the improvement for yourself…seeing improvement makes it all worth it" (22fG3/4H). Observing the PU improve helped reclaim a positive body image. However, a practical issue for many was that their PU was in a place that was not easy for them to view (i.e. sacrum, back of heels). These patients found it troubling listening to others talk about something that was on their body but that they could not see or reach. Hence, using a mirror or taking a photograph were helpful strategies. Others had a disinterest in seeing the PU as it was "ugly and dirty", and "unattractive", and they preferred to remain detached from it or not thinking about the PU as a part of them.

Physical attractiveness and physical changes such as scarring had negative consequences on body image and self-concept. Being treated by others as a PU-patient or as someone with a disability was associated with self-identity problems. The physical appearance and exudate and odour of the PU had a negative impact on body image, leaving patients embarrassed and self-conscious that their "body stinks and is leaking gunge" (22fG3/4H). Others reported feeling ashamed "that [PU/smell] is part of me and others can [see/smell it]" (43fG3H). Becoming totally dependent on others was also embarrassing and affected self-concept, as patients felt useless and inadequate, impacting self-respect, self-esteem, and their confidence. Others felt that their PU had taken on its own identity: "see me and not the PU" (41mG4H), eliciting anxiety about having their identity reduced to being a PU-patient.

3.4.2.4. Social Functioning

Social isolation

Feeling left out or isolated, commonly expressed as feeling "captive", "confined" and "alienated", was an important consequence of having a PU. Many patients were

confined to bed or hospitalised for PU treatment, leaving them feeling disconnected, alone, missing family, friends and pets, and restricting social connectedness and interaction.

Participation

Patients reported a variety of social problems including restrictions in social participation (e.g. where one can go, how long they can stay out). Social lives were disrupted mainly because of physical restrictions including treatment (i.e. time spent on wound treatment regimens, hospitalised or in bed), which kept patients away from friends and loved ones. Related to this was the wait for PU treatment, which further restricted social activities and interfered with going out, making social plans, and loss of control over planned activities. Others mentioned disinterest in socialising: “I don’t want to go out because of [pain or the smell]…don’t enjoy it” (41mG4H and 62mG4C) or described how they were unable to enjoy socialising because they were aware of time restrictions and the need to return home for bed-rest or treatment.

The ability to function in several common roles (e.g. self-care, social/familial, and vocational/educational) was severely impaired because of PUs. Patients reported problems in relationships (marriage, parenting, friendship), with their self-esteem/self-concept and psychological well-being. Many worried about the impact of the PU on relationships with partners, who often had to take on roles that the patient could no longer manage. Patients were also concerned that their relationship would change if they became dependent on their partners for PU care, leaving their partners feeling like a nurse/carer. Difficulties with closeness and interpersonal communication were also reported. For younger adults, having sexual relations, being able to see each other and interference with work or school were additional concerns.

3.4.2.5. Patterns of association

Associations in the data were considered for the following patient factors: age, gender, PU severity and location, and comorbidity (e.g. patients considered elderly, acutely ill or with neurological impairment such as SCI or MS). No gender or PU location differences emerged. For other factors, the reported HRQL issues were largely the same for all patient types, with some minor differences emerging. Specifically, irrespective of age, PU severity, and comorbidity, patients reported physical impairment (e.g. difficulties with movement, daily activities and sleeping) and restricted social participation. About a third who reported social impairment also reported role functioning difficulties. Patients who reported social isolation all had severe PUs of duration greater than 3 months. All patients who reported a problem with pain also

reported that pain contributed to emotional distress and restricted physical and social functioning, whereas smell and exudate were problems mainly for patients with severe PU (i.e. skin breakdown). Older patients were less likely to report pain, fearing being a burden or nuisance. Those with SCI reported pain or discomfort however the sensations experienced differed from those of others depending on the level of injury.

Apart from three, all patients reported impaired psychological functioning, and physical and social limitations irrespective of individual patient factors. Patients who were frustrated reported limitations in physical and social functioning, loss of independence and treatment burden. Loss of independence led to self-concept and identity problems (i.e. questioning self-worth, confidence and importance). Patients who worried about their PU or who were bothered by their physical appearance reported negative emotional problems. Most patients reported some form of coping however not all patients who used coping mechanisms reported less negative psychological impact.

One third of patients had neurological impairment (i.e. SCI or MS) as well as severe PUs of duration greater than 3 months. These patients reported learning to live and function with their existing condition until medical “set-backs” such as PUs caused them substantial impairment and difficulty. On the other hand, patients with acute medical problems (i.e. broken hip, accident/trauma) found that their primary medical condition caused them greater impairment and burden, with the PU perceived to not impact greatly on their lives.

3.5. Discussion

This study produced a PU-specific conceptual framework of HRQL that includes four domains or constructs: PU-symptoms, physical functioning, psychological well-being and social functioning. These constructs are similar to those in generic HRQL models (78, 89, 96), however, this framework also incorporates components specific to PUs. For example, “social isolation” is often excluded from the social domain in generic models, but was found to be important to people with PUs. “Appearance and Self-consciousness” also appear specific to PUs, especially embarrassment and concerns about changes to self-concept. In addition, several components emerged that were not identified by previous research (9, 104, 105, 106, 112, 155, 171, 199). Components added to the psychological well-being domain included extreme frustration, feeling hopeless, helpless and anxious, appearance/self-consciousness and self-efficacy. Malaise and sleep were added to the physical functioning domain.

This is the first study to conceptually map the range and nature of the impact of PUs on HRQL and investigate patterns of association unique to specific patient factors. For people with PUs, irrespective of age, gender, PU severity and location, and comorbidity, the main impact on HRQL is in terms of restrictions in physical and social functioning, particularly mobility and movement, looking after oneself, malaise and sleep disturbance, emotional problems, and pain. Impaired physical and social functioning contributes to emotional and mood problems and is associated with low self-efficacy and independence and negative self-concept. Physical appearance is also associated with self-consciousness and identity problems, reducing psychological well-being. Emotional distress and mood can affect all other components in the HRQL model. For example, self-consciousness about PU appearance, smell or exudate, or anticipation of pain may reduce physical and social activity. Reduced physical or social functioning may, in turn, lead to depression, anxiety and frustration. PU symptoms and anticipation of pain were associated with emotional distress and anxiety and contributed towards limitations in physical and social functioning.

Despite relatively few studies relating to HRQL in individuals with PUs (199), previous research has identified common concerns reported by patients about the experience of living with a PU and the burden of PU interventions on their lives (9, 104, 105, 106, 107, 112, 161, 164, 165, 166). Consistent with previous findings, PUs have negative effects on HRQL (9, 104, 105, 106), restrict activities and mobility (9, 105, 106), contribute towards social isolation (104, 106), emotional problems (9, 104), and cause severe and persistent pain (9, 104, 105, 106). In addition, early warning signs for skin damage (i.e. pain) do not prompt action, treatments increase pain and discomfort, and patients' views and concerns relating to physical, social and psychological aspects are not being met. Evidence from existing qualitative research relates to patients' experience of living with a PU (9, 104, 105, 106) and not specifically about how PUs *impact* HRQL. Lived experience studies give meaning to what it is like to experience having a PU, but patient reports are broader than HRQL issues, including experience of care reports and factors that might be consequences of having PUs but that are not specifically HRQL outcomes; the distinction between concepts is not made. Further, operationalisation of HRQL constructs has not previously been made

Building on previous research, the conceptual framework was developed with qualitative work that incorporates views from a representative sample of patients with PUs. This ensures that the conceptual framework content is characterised by outcomes that are important and relevant to patients. Importantly, poor interview techniques can introduce bias. An interview guide was developed to ensure that

interviews were guided and leading questions avoided. The transcriptions of the first few interviews were checked for context and content before subsequent interviews were undertaken and data analysis begun. Independent coding by two researchers, inter-rater reliability and consensus meetings were undertaken. These precautions, the fact that data was analysed in a systematic and rigorous manner, and that data interpretations were grounded in actual patient data helped minimise the risk of bias.

Patients across all adult age-groups and gender, PU severity and skin site, clinical specialities and healthcare settings were included however the sample was limited to English-speaking British nationals without cognitive impairment. Differences were explored by age, gender, PU severity and location, but no major differences among these subgroups was observed. Inclusion of patients with varying comorbidities assisted in separating out the effect of existing conditions from the effect of PUs to ensure that the conceptual framework was representative of PU-specific impact on HRQL and relevant to all types of patients with PUs. PU symptoms were problematic for all patients, however older patients were less likely to report pain and for those with impaired sensitivity it was assumed pain was not a problem.

3.6. Summary of Chapter 3

This study provides insight into PU-specific components of HRQL and provides qualitative evidence for HRQL outcomes that are important from the perspective of patients with PUs. This work assisted in operationalising PU-specific HRQL constructs; an essential step towards ensuring valid measurement. The final PU-specific HRQL conceptual framework will be used to assess any PROMs identified from a review of existing measures (Chapter 4). The earlier review (Chapter 2) failed to identify PROs specific for PUs. However, the review was not intended to determine whether measures for other similar chronic wounds were available and suitable for PUs. As such, this conceptual framework will provide a structured and formal method to assess the content of available measures against. It will also provide the basis for the development of a new PU-specific PRO measure of HRQL if required.

Chapter 4

SYSTEMATIC REVIEW OF THE LITERATURE

Existing PROs used in PU and other chronic wound research

4.1 Overview

Chapter 4 reports the methods and results of a systematic review of the literature intended to identify and evaluate currently available PRO instruments used in PUs and other chronic wounds. The conceptual framework developed in the previous chapter provides a structured and formal method to assess the content of available instruments to determine suitability for use in PUs or the need for a PU-specific PRO instrument.

4.2 Background

There has been a growth of PRO instruments to evaluate HRQL in many disease areas including chronic skin conditions, with a number of condition-specific measures developed for different chronic wounds. The application of PRO instruments in the evaluation of healthcare has become increasingly important in other disease areas (83, 84) as PROs have commonly been the main dependent variable on which decisions are made that influence patient care. However, in PUs, assessment of outcomes in clinical trials of PU intervention effectiveness have been limited to conventional clinical outcomes and HRQL outcomes rarely assessed (Chapter 2). What is not clear is whether failure to incorporate patients' assessment of outcomes arises because appropriate methods do not exist or whether methods exist but have not been widely adopted or 'fit for purpose' has not been determined (i.e. inadequately validated).

Despite the impact on HRQL, little research has been done to determine the availability of PRO instruments, either generic or disease-specific, and their suitability for use in PU research. A review of the PU and HRQL literature (Chapter 2) failed to identify PROs specific for PUs but the review aims were not to systematically search for PRO instruments. Therefore it was deemed necessary to determine whether PROs for PUs or for other similar chronic wounds were available and suitable for use in PUs.

4.3 Aim

In the absence of any known scientifically robust PU-specific HRQL measure, this pragmatic review is based on the need to identify and evaluate currently available PRO instruments used in research with PU populations and other similar chronic wounds to

determine how useful or appropriate these measures are for use with people with PUs. Specific objectives were to comprehensively and systematically search the international literature on PUs and other chronic wounds to identify generic, PU-specific and chronic wound-specific PRO instruments used to assess HRQL in patients with PUs or other chronic wounds and to assess PRO instruments' content against the PU-specific HRQL conceptual framework (Chapter 3), in addition to evaluating feasibility for use with patients with PUs (i.e. item specificity, response burden).

In the previous chapter, a PU-specific HRQL conceptual framework was derived from a systematic review of the literature and the views of patients with PUs. The conceptual framework includes four HRQL domains divided into 13 sub-domains. Not only will the conceptual framework be important for informing future HRQL research in this field, but the intention was to have a structured and formal method to assess any available measures against for their feasibility and relevance to PUs.

4.4 Methods

4.4.1 Search strategy

Consistent with the methods undertaken in the first systematic review (Chapter 2), an iterative process was used to develop an optimal search strategy to ensure sensitivity without yielding too many results. The search strategy was developed (Appendix 4.1) to retrieve all research related to HRQL instruments used in studies of patients with PUs or other chronic wounds (e.g. leg ulcers, diabetic foot ulcers, fungating ulcers, dehisced surgical wounds). Search terms consisted of key words for the clinical disease area and HRQL and related concepts (QoL, health status, function and well-being), which were combined using Boolean operators as follows: (PU or chronic wounds) and ((HRQL concepts) or (patient-reported) or (questionnaire)). To increase specificity, the searches were refined by publication (not commentary, letter, guidelines or audit), wound type (not buruli, digital ulcers, burns or spider bites), and limited to humans and adults (aged 18 years or above).

4.4.2 Data sources

The following databases were searched from inception until June 2009: MEDLINE, Embase, PsycINFO, CINAHL, BMI, AMED, The Cochrane Library and Web of Knowledge (WOK). The Cochrane and WOK databases were searched using 'PU' or 'pressure sore' and 'QoL' topic words. Auto alerts via OVID database library were set up until data extraction was completed (November 2009) to notify of any additional papers that had been added to the databases since the original search was performed.

To identify unpublished and on-going studies, the electronic searches were supplemented by a hand search of specialist journals, relevant conference proceedings and theses and dissertations. Twelve experts were contacted, the Patient-Reported Outcome and Quality of Life Database (PROQOLID, http://www.qolid.org/) and the UK National Research Register searched, a citation search on all included studies and systematic reviews identified in the searches (147, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210) performed, and an internet search of web content relating to PU self-help and focus group websites undertaken (see Appendix 4.2 for detailed list).

4.4.3 Criteria for study selection

Studies of any design were included if PRO instruments were used to assess HRQL in adult patient populations presenting with any category PU or chronic wound, from hospital, rehabilitation or community health settings within Europe, North America or Australia. Thus, studies were excluded if: HRQL was assessed by a healthcare provider or proxy (i.e. not patient-reported); the PRO was intended primarily for other medical conditions where PUs are secondary outcomes (e.g. Life Situation Questionnaire), assessed mediating or contributory outcomes only (e.g. social support, personality, locus of control, coping or mini-mental for screening cognition), or used single-item rating scales (i.e. VAS); the sample was limited to paediatric populations or wounds caused by trauma (e.g. burns); or the development/use of the PRO was outside Europe, North America or Australia as it would not be culturally appropriate for use in a UK setting. PRO instruments which had not undergone psychometric evaluation for reliability or validity (i.e. *ad hoc* instruments) were also excluded.

4.4.4 Selection of studies

Abstracts from studies retrieved were screened for relevance by one reviewer (CG). Studies that did not meet the inclusion criteria were excluded at this stage. Studies assessed as potentially relevant or where relevance was ambiguous, were obtained in full for further scrutiny. Two researchers (CG, JN) independently assessed potentially relevant studies and those that did not meet the eligibility criteria were excluded from further analysis. Where there was lack of consensus, disagreements were resolved through discussion with a third reviewer (DL). Where study details were lacking, attempts were made to contact the authors for additional information.

4.4.5 Quality appraisal

Individual quality components of PRO instrument methodology were not used as a threshold for inclusion of instruments in this review as the intention was to assess the

content of available instruments. However, one piece of empirical evidence for reliability and validity was a minimum requirement for instrument inclusion. It was proposed that the strength of the psychometric evidence would be appraised by two researchers (CG and YA) if studies met thresholds as described in section 4.4.7. The appraisal was based on a 3-point rating scale where + = some limited evidence in favour (weak evidence); ++ = some good evidence in favour, but some aspects do not meet criteria or not reported (adequate evidence); +++ = good evidence in favour, all major forms of evaluation reported (211).

4.4.6 Data extraction

Two researchers (CG, YA) independently extracted data using structured data collection forms developed to investigate theoretical underpinnings for PRO instruments (83, 212, 213). Data extracted included: descriptive characteristics (name of instrument, mode of administration, modified versions), PRO-specific issues (purpose, intended population, actual population validation performed in, theoretical orientation), content (evidence for an underlying conceptual model, list of all domains and items) and measurement properties (development method, item reduction, psychometric analysis). Extractions were cross-checked for errors, omissions and consistency in extractions. Disagreements or discrepancies were discussed between the two researchers and confirmed with a psychometrics expert (DL). Data pertaining to the development and validation of measures were extracted from the original development paper. Information about the use and application of the measure and further descriptive and psychometric data were extracted from additional papers.

4.4.7 Analytical procedures

The analytical procedure was designed to systematically determine the extent to which PROMs covered the PU-specific framework (see Chapter 3), were relevant to PUs (item specificity), and posed minimal response burden.

Two level content validity: i) Domain level content analysis – determined the percentage of PROM domains that mapped onto the PU-specific conceptual domains and sub-domains; the higher the percentage the better coverage of PU-specific domains (e.g. How many domains map onto PU-specific domains?). To ensure relevant domains were not excluded, where domain names differed from those in the PU-specific framework, item content, as described by the authors, was compared against the PU-specific domain content to determine domain consistency irrespective of different labelling (e.g. activity domain content was consistent with physical

functioning content). For PRO instruments with domains that mapped 75% or more onto PU-specific domains, the percentage of items representing mapped domains was considered (e.g. how many items represent mapped domains?). For instruments with 75% or more items representing mapped domains, item-specificity was determined (level two analysis).

ii) Item level content analysis – item-specificity was determined by the percentage of PRO instrument items that are PU-specific (e.g. to what degree are items relevant to PUs?). Two researchers (CG, JN) independently undertook a critical appraisal of items by making a judgement about whether they represented HRQL outcomes relevant to the impact of PUs. Items were considered PU-specific if they mapped onto the descriptive components of the PU-specific framework (see Chapter 3, section 3.3.2 for specific components). The two sets of assessments were cross-checked for consistency.

Response burden: In addition to content validity and item-specificity, PRO instruments may contain domains and/or items that are not relevant or appropriate for use with patients with PUs. The balance between relevant and non-relevant content (referred to here as response burden) was expressed as the percentage of items representing non-mapped domains (i.e. the inverse of the percentage of items representing mapped domains (above)).

Psychometric properties: For PRO instruments that had at least 75% of PU-specific content (at both domain and item level) and no more than 25% of non-relevant content, it was proposed that psychometric properties, including reliability, validity and responsiveness would be appraised. As none of the identified instruments met these criteria, the psychometric properties were not assessed.

4.5 Results

4.5.1 Studies included in the review

A total of 2616 papers were retrieved, of which 248 were assessed as potentially relevant and 50 met inclusion criteria. From these, 19 potential instruments were identified; however, two were excluded, one at development stage (214) and one provided no evidence for reliability or validity (215). A total of 17 PRO instruments fulfilled the inclusion criteria, including three generic and 14 chronic wound-specific (9 for leg ulcers, 4 for diabetic foot ulcers and 1 for mixed; leg and diabetic foot wounds). No instruments were identified that had been developed and validated in PUs. The

results of the two-level content (domain and item level) and response burden analysis of existing PRO instruments are presented in table 4.1.

Generic measures

There is a plethora of generic instruments with evidence of validity and reliability; however, only five studies investigated HRQL in patients with PUs using three generic measures: Ferrans and Powers Quality of Life Index (QLI) (216), the Medical Outcomes 36-item Short-Form Health Survey (SF36) (112, 172, 217) and the Medical Outcomes 12-item Short-Form Health Survey (SF12) (218). The three generic measures covered six of the 13 (46%) PU-specific HRQL domains, with response burden ranging from 17-68%. None of the generic measures met the 75% criterion of content validity. Four of the 13 PU-specific domains – social isolation, sleep impairment, exudate and odour – were not covered by any generic measures. In addition, the QLI did not include items about mobility or activities of daily living.

Chronic wound-specific measures

The 14 chronic wound-specific instruments were developed for measuring HRQL in leg ulcers (Hyland Questionnaire (219), Venous Leg Ulcer QoL Questionnaire (VLU-QoL) (220), Charing Cross Venous Leg Ulcer Questionnaire (221), Chronic Lower Limb Venous Insufficiency (222), Freiburger Questionnaire of QoL in venous diseases (223), modified version (FLQA-M) (224), Skindex-61 (225), Skindex-29 (226), Venous Insufficiency Epidemiology and Economic Study (227) or diabetic foot ulcers (Diabetic Foot Ulcer Scale (DFS) (228), DFS-Short Form (229), American Academy of Orthopaedic Surgeons Foot and Ankle Outcomes Questionnaire (230), Neuropathy and Foot Ulcer-Specific QoL Instrument (231)). One instrument, the Cardiff Wound Impact Schedule (CWIS) (179), was developed for mixed wounds but only evaluated in leg and diabetic foot ulcer patients. The CWIS is the only chronic wound-specific instrument used in patients with mixed chronic wounds that included six patients with PUs (23% of the total sample) (232).

The percentage of the 13 PU-specific HRQL sub-domains covered by the 14 chronic wound instruments was higher than for generic instruments, ranging from 31-77%. However, as seen with generic instruments, social isolation, sleep, exudate and odour sub-domains were not well covered. Only two instruments, the VLU-QoL and DFS, represented more than 75% of PU-specific HRQL domains with coverage of 10 of the 13 (77%) PU-specific conceptual domains: VLU-QOL with 29/34 items (85%) and the DFS with 38/58 items (66%). As the VLU-QOL met the 75% criterion of items representing mapped domains, item-specificity was determined.

Table 4.1 Chronic Wound-Specific and Generic Quality of Life Instruments: Content analysis

Patient Population	**Venous Leg Ulcer (number of items)**									**Diabetic Foot Ulcer**				**Mixed**	**Generic**		
Pressure ulcer-specific domains[9]	**LFQ (34)**	**VLU-QoL (34)**	**CC-VLU (20)**	**CIVIQ (20)**	**^FLQA (81)**	***FLQA-M (83 +6)**	**Skin-61 (61)**	**Skin-29 (29)**	**Veins-QoL/ Sym (26)**	**DFS (58)**	**DFS-SF (29)**	**AAOS foot (25)**	**NeuroQoL (28)**	**CWIS (33)**	**QLI-SCI (37)**	**SF36 (36)**	**SF12 (12)**
Pain & discomfort from ulcer	•	•	•		•	•	•	•	•	•	•	•	•	•	•	•	•
Odour		•												•			
Exudate/blood		•	•					•						•			
Mobility	•		•	•			•					•	•	•		•	•
Daily activities		•	•	•	•	•	•		•	•	•	•	•	•		•	•
General malaise		•		•			•	•		•	•				•	•	•
Sleep	•	•		•			•	•		•	•			•			
Mood	•	•	•	•	•	•	•	•	•	•	•		•			•	•
Anxiety & worry	•	•	•	•			•	•	•	•	•			•	•		
Self-efficacy & dependence		•						•	•	•	•		•		•		
Appearance & self-consciousness	•	•	•	•			•	•		•	•		•	•	•		
Participation	•		•	•	•	•	•	•	•	•	•	•	•		•	•	•
Isolation										•							
Number (%) of 13 PU-specific domains covered	7 (54%)	10 (77%)	8 (62%)	8 (62%)	4 (31%)	4 (31%)	9 (69%)	9 (69%)	6 (46%)	10 (77%)	9 (69%)	4 (31%)	7 (54%)	8 (62%)	6 (46%)	6 (46%)	6 (46%)
Number (%) of items	-	29	-	-	-	-	-	-	-	38	-	-	-	-	-	-	-

Indicator of content validity

	Patient Population	**Venous Leg Ulcer (number of items)**									**Diabetic Foot Ulcer**				**Mixed**	**Generic**		
	Pressure ulcer-specific domains[9]	**LFQ (34)**	**VLU-QoL (34)**	**CC-VLU (20)**	**CIVIQ (20)**	**^FLQA (81)**	***FLQA-M (83 +6)**	**Skin-61 (61)**	**Skin-29 (29)**	**Veins-QoL/ Sym (26)**	**DFS (58)**	**DFS-SF (29)**	**AAOS foot (25)**	**NeuroQoL (28)**	**CWIS (33)**	**QLI-SCI (37)**	**SF36 (36)**	**SF12 (12)**
	representing mapped domains		(85%)								(66%)							
Indicator of item specificity	Number (%) range of PU-specific items	-	*19-21 (56-62%)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Indicator of response burden	Number (%) of items that do not map onto PU-specific domains	9 (26%)	5 (15%)	4 (20%)	5 (25%)	32 (39%)	45 (47%)	17 (28%)	2 (7%)	7 (27%)	20 (44%)	2 (7%)	8 (32%)	9 (32%)	13 (39%)	25 (68%)	8 (22%)	2 (17%)

-Did not meet criteria for evaluation

^Questionnaire published in German and translated into English

*Range based on 2 independent item assessments

Independent item-specificity assessment by two researchers revealed that 19 and 21 of the 34 items (only 56-62%) were PU-specific (e.g., 'can not wear what I like' is relevant to leg ulcers as they are visible on the legs whereas PUs are usually in areas hidden from view such as the sacrum).

Response burden for all measures ranged between 7-68%. Items representing non-HRQL domains: treatment/healthcare satisfaction; analgesic side-effects; financial-occupation; cognition; time spent on ulcer; change over past year; impact on others; avoidance behaviours; footwear/clothing; non-wound related symptom distress; attitude; support; achievement goals; and happiness and/or spiritual.

4.6 Discussion

This systematic review identified PRO instruments used in studies to measure HRQL outcomes related to PU and PU treatment impact. Where HRQL outcomes have been assessed, generic or chronic-wound specific PRO instruments were used. HRQL is a complex construct in terms of conceptualisation and measurement. Conceptualisation of HRQL specific to PUs has been mapped including four core domains (Chapter 3). This review identified that these domains are common in other generic and chronic wound-specific HRQL models however the content differs at sub-domain and item level. There is inadequate PU-specific sub-domain coverage by generic instruments (between 0-46% coverage). Chronic wound-specific instruments have better coverage (31-77%), although important components such as issues stemming from PU treatments and symptoms, movement, sleep, isolation, body image/self concept, embarrassment and physical appearance were not well represented. Therefore, despite common conceptual domains in chronic wound HRQL models, the available instruments do not adequately represent PU-specific HRQL, questioning their appropriateness for use in PU research.

The most frequently applied PRO instrument used in PU research (3 out of 5 studies), the SF36, is not conceptually comprehensive in its assessment of HRQL outcomes important in PUs. The SF36 has been validated and applied in many disease areas including dermatology, so perhaps it would be the obvious choice by researchers. However, based on conceptual content, some items appear too crude to pick up differences in PUs. At domain level, the SF36 contains 75% of domains (variables) relevant to PUs. However, upon closer inspection of individual items, many were not PU-specific. It is therefore not surprising that attempts to measure changes in HRQL in patients with PUs using the SF36 and other generic instruments have had mixed results. Some findings indicate lower HRQL outcomes for patients with PUs compared

to those without (109) while others found no difference (112). In other clinical areas, generic-instruments were found less powerful in detecting treatment effects compared to disease-specific instruments as they generally reflected a limited perspective (233, 234, 235, 236).

The optimal use of generic instruments would be for population comparisons on broad domain level (i.e. compare PU population with the larger general or chronic wound population). Generic measures are intended to be suitable for a wide range of health conditions and therefore may not be sensitive enough to detect important changes due to a particular disease. Item analysis revealed that generic instruments contained fewer PU-specific content which might make it difficult to detect the true impact of PUs on HRQL and failing to provide information about the real changes within patients with PUs or differences among them. One instrument, the VLU-QoL, emerged as conceptually comprehensive (77%), but only 56-62% of items were judged PU-specific. This finding may not be surprising as the VLU-QoL measure's content was developed based on the SKINDEX model (which is intended to assess dermatological conditions) (226)and focus groups with patients with leg ulcers between the knee and ankle; the most common location of PUs is the sacrum.

As a whole, the PRO instruments presented some degree of response burden for this population either from items representing non-HRQL content or from those not specific to PUs. This poses questions about item relevance and redundancy. Irrelevant items introduce unnecessary response burden. Other general considerations observed during item analysis related to item wording. Ambiguous or not easily understood items leave too much room for individual interpretation. Items can also be problematic in capturing PU-impact if they are too specific (e.g. 'frustrated because you were unable to do things that you wanted' is too specific because frustration is a problem for patients with PUs but not only because of inability to do things they want) or double-barrelled (i.e. 'frustrated by others doing things for you when you would rather do things yourself').

Based on the above observations, it seems that failure to incorporate patients' assessment of outcomes in PU research may be due to the fact that appropriate instruments do not exist for evaluating the impact of PUs on HRQL. In the first instance, when selecting the most appropriate instrument, one needs to determine whether it is fit for purpose (i.e. is the content appropriate to the research question) (83, 87). Regardless of their psychometric properties, PRO instruments can only be considered fit for purpose for the assessment of patient outcomes if they are valid with

minimal response burden (i.e. are constructs and content [items] relevant to the particular disease/outcome of interest?). Other considerations may include determining whether the intended instrument was developed from the patient perspective using patients' words or whether the psychometric properties have been evaluated in the target population. The findings from this review provide further support for the importance of clearly defining and conceptualising the constructs measured when developing new condition/disease-specific instruments.

Reviews of PRO instruments for other disease areas have predominantly evaluated content at broad domain level (constructs) and psychometric properties, without reporting extensive evaluation of content relevance (item specificity) to the particular disease of interest based on conceptualisation and response burden (237, 238, 239). This research is the first to undertake a comprehensive evaluation of the conceptual content of available PRO instruments used in PUs and other chronic wounds based on domain and item content validity, item specificity and response burden to determine appropriateness for use in PU populations. The main strength was consideration and inclusion of available measures, both generic and condition-specific (chronic wound), in an attempt to comprehensively represent HRQL models and map them conceptually to a PU-specific HRQL conceptual framework.

This review had collaboration which allowed for the inclusion of relevant PRO instruments developed in languages other than English (e.g. FLQA developed in German), and inclusion of studies undertaken with PU populations across multiple continents. Thus, the conceptual mapping of instruments included in this review may be transferable to European, North American and Australian cultures and languages, although cross-cultural validity of the PU-specific framework needs to be determined. However, only instruments validated for use within these populations were included. Measures developed in countries outside these regions had been identified but it was considered that those measures would not be culturally appropriate for Western populations. Another limitation of this review is that it is not comprehensive of all available generic instruments; only inclusive of those used previously in PU research. It is possible that other available generic measures may be appropriate for use in PU research but their content and psychometric properties would need to be examined.

No condition-specific PRO instrument for assessing HRQL outcomes in patients with PUs was identified. Some generic and chronic wound-specific instruments were identified but these do not cover content relevant to PUs. Various non-validated questionnaires had been developed to assess outcomes such as pain, comfort and

exudate, although these instruments were study-specific and had not undergone the same rigorous psychometric evaluation as other established generic and disease-specific instruments.

4.7 Summary of Chapter 4

A comprehensive and systematic search of the international literature on PUs and other similar chronic wounds was undertaken to identify and evaluate available PRO instruments based on content validity, item specificity and response burden against an empirically derived PU-specific conceptual framework. This review provides the necessary information for the appropriate selection of PRO instruments when measuring impact of PUs on HRQL. Currently, HRQL outcomes important in PUs are inadequately covered by available instruments despite similarities between conceptual models, and they contain irrelevant content introducing response burden. The findings highlight the need for clear conceptualisation of PRO instruments' content as well as determining fitness-for-purpose when selecting instruments in the future. As no PU-specific PRO instruments exist, and to address some of the limitations of using available instruments in PU research, a new patient-assessed HRQL measure specific for patients with PUs is needed for use in clinical practice and future research.

Chapter 5

PU-QOLI DEVELOPMENT AND PRETESTING

5.1 Overview

Chapter 5 reports the methods and results of the development and pretesting of the PU-QOLI. This chapter includes the methods used to generate content (selection of items) and construct scales, based on a conceptual model. This is followed by a pretest, using mixed methods including qualitative work to ensure that the PU-QOLI was clear, understood and relevant to people with PUs, followed by a Rasch analysis to identify and resolve problems with layout, time frame, response options, framing of items and administration, and confirm content prior to formal psychometric evaluation.

5.2 Background and aims

Scientifically robust PRO instrument development is a timely, costly and demanding process. PROs should possess various attributes that are critical for judging its' strength for measuring health outcomes, including: its conceptual and measurement model; reliability; validity; responsiveness; interpretability; burden; translations; and cultural adaptability (86, 102, 240, 241). To ensure that new instruments meet such standards and are designed and developed for the purpose intended, various stages are undertaken including: 1) item generation (Sections 5.3-5.5); 2) operationalisation (Section 5.6.2) and pretesting (Section 5.7); and 3) field testing (Chapter 6 and 7). A multidisciplinary team consisting of clinical experts and methodologists and input from patients representing the target population is also advised (86).

The first phase of development, item generation, is the stage in which a comprehensive list of items is generated relevant to important issues specific to the topic under investigation. A systematic approach was used to ensure generation of an exhaustive and unbiased item list, utilising three sources: i) patients (Chapter 3), ii) relevant literature (Chapter 2) including generic and disease-specific instruments (Chapter 4), and iii) clinical experts.

5.3 Item generation from patients: semi-structured interviews

An important consideration when developing PRO instruments is conceptualisation and content (120). Developers of new instruments often utilise 'top-down' approaches (i.e. review of literature and content from existing PROs) for conceptualisation and content development. A risk of solely using this methodology for developing disease-specific

instruments is that it may not reflect HRQL aspects most important to patients with the underlying condition and threaten content validity and responsiveness to change as outcomes relevant to patients may not be included (183).

5.3.1 Methods

As previously described (Chapter 3), content (patient words verbatim) from 35 patient interviews were used to generate items. All content was grouped into HRQL domains, with each domain (variable) comprising a number of items describing slightly different components. Grouped items were listed and frequency of reports indicated (i.e. I get frustrated x5). The intention was to be inclusive therefore all verbatims were included if reported more than once. Verbatims were an excellent source for generating items, as items using patients' words and representing variable components across the broad spectrum of PU-specific HRQL outcomes were identified. Items generated from patients are indicated in Appendix 5.1.

5.4 Item generation from existing instruments

Critical appraisal and building on the work of others is recommended by Streiner and Norman (132). Modifying relevant, important or discriminating items rather than constructing new ones could save unnecessary work as items from existing validated instruments have generally undergone repeated testing and proven useful and psychometrically sound. Further, there is no need to be creative; there are a limited number of ways to ask about a specific problem. For example, the impact of a given disease on physical function is likely so instruments in similar disease areas may have developed items for this variable.

5.4.1 Methods

Available generic and chronic wound-specific PRO instruments were identified and reviewed for content and psychometric properties in chapter 4. Items from identified instruments were mapped to the PU-specific conceptual framework (Chapter 3) to determine relevance to PUs. Those considered relevant could be included in the item pool. In addition to literature identified electronically or from hand searching, the internet was searched for patient information and self-help forums relevant to PUs; however, no such information was retrieved. Although a potentially valuable source of generating items, few items were assessed as PU-specific (i.e. worded to assess impact of other conditions and not specifically PU-impact) and those assessed as PU-specific were items already generated from patient reports.

5.5 Source 3: Item generation from experts

Patients with PUs receive specialist care from nurses, tissue viability specialists and other healthcare professionals who have a vast range of experience treating patients with PUs and probable insight into patients' experiences. A group of nursing and TVN specialists were selected to review items generated from patient interviews.

5.5.1 Method

Clinical Group

The clinical expert group consisted of three community and acute TVN specialists (LW, NS, EM), three nurses with extensive experience undertaking chronic wound research predominantly in PUs (JN, EAN, CD), one chronic wound pain specialist (MB) and one nurse with experience of healthcare policy development (SC). The group reviewed the item pool, with instructions to focus on item relevance, content and wording, and clinical importance, until consensus. Items with similar content were highlighted and accuracy of domain categorisation (item grouping) discussed. Additional items were added if necessary. The researcher (CG) made notes.

Items were retained if they were considered: clinically relevant; frequently reported; not too similar (redundant items combined or removed); or pertaining to HRQL (not measuring other constructs, i.e. personality, satisfaction). Final item elimination decisions were based on consideration of item problems in combination. At this stage an important conceptual decision was made to include PU-symptoms: pain, exudate and odour into the PU-specific model. These symptoms are important consequences of having a PU but they were initially not considered HRQL outcomes and therefore excluded from the original item pool. Counting frequency or assessing intensity of symptoms may not be an adequate measure of HRQL (242), but the impact of symptoms and the meaning they have for individuals is an important aspect of HRQL. Therefore, symptoms were considered important for inclusion in a PU-specific PRO instrument. Patient comments pertaining to these outcomes were detected from the qualitative work (Chapter 3) and patient verbatims were used to add items representing these symptoms. Following expert review, the item list was amended.

Outcome Methodology Group

The amended item pool and content mapping was reviewed by a group of seven health outcome methodologists (DL, SSc, SS, SCa, KH, YA, JP) (see Intellectual Property Section, page ii for group membership). The group focused on the construct being measured (e.g. are items for each domain representative of the construct being

measured?) and item wording (i.e. are any items confusing, ambiguous, double-barrelled?). Appendix 5.1 represents domains, final items retained for the first PU-QOLI version and any modifications made to the item pool (e.g. consolidation of similar items or elimination).

5.5.2 Content validity

Using three sources, a pooled item list was produced. Items came from patients but were also consistently identified from the literature and supported by experts. Utilising multiple sources for item generation reduced bias and allowed for an exhaustive item pool. Although some items were not included in the final instrument, it was important for retained items to adequately represent all important domain components to support content validity; item pooling was initially over-inclusive but poor items could be detected and withdrawn during pretesting and psychometric evaluation (Chapter 6). Some theorists have argued that 'content relevance' and 'content coverage' are more accurate descriptors (243). Each domain component should be represented by at least one item (content coverage); depending on the variable, more items might be needed to tap into all relevant aspects of a variable to ensure that important aspects are captured and reflected in the final score. Each item in a group (scale) should represent the content area (content relevant to each domain). Items not related introduce error in the measurement; the item is measuring some other variable and therefore discriminating between respondents on a different variable (87). Items, representing individual domain components, combined, provide content validity for that specific HRQL domain (i.e. physical function).

5.6 Construction of the PU-QOLI

In order to encourage optimal instrument completion (the best responses and completion with no missing data), it is of paramount importance that the way the PU-QOLI is presented is tailored to the characteristics of patients with PUs (i.e. understood by and relevant to the intended population). This section describes the careful stepwise construction of the preliminary PU-QOLI version (Appendix 5.1), based on current detailed guidelines (86, 132, 211, 244). The processes of scale construction, known as operationalisation (logically related item clusters), questionnaire design and pretesting are described.

The pooled item list was transformed into scales to construct the preliminary PU-QOLI (Section 5.6.2, *selecting the questions*). Although this process is seemingly straightforward, careful consideration of the design, layout and instructions, framing of

questions, response format and recall period is needed to reduce cognitive and respondent burden and to ensure that respondents understand each question the way it is intended so that they can accurately formulate responses. Factors influencing scale construction and instrument design are discussed below.

5.6.1 Mode of administration

Mode of administration can be interviewer administered (i.e. face-to-face, telephone) or patient self-completed (e.g. postal survey, during clinic visit). Decisions regarding administration mode selection are assessed against the advantages and disadvantages of both methods. In research, self-completed versions are preferred as they are associated with lower costs; can cover largely dispersed populations (if postal); require less time to collect data but time spent on late returns and follow-up attempts may extend the data collection period; and avoid potential interviewer bias as no interviewer is required (e.g. respondents make less socially desirable responses or more truthful responses with self-completion) (132, 244). Self-completed methods, on the other hand, typically produce lower response rates due to literacy visual or motor impairment, or questionnaires misplaced or forgotten to be returned) (244).

There are also disadvantages to interviewer administration. This method can be time-consuming and costly (i.e. training interviewers, travel to interview) but no waiting time for data returns (ensures returns); produce random variance due to interviewer inaccuracy (e.g. altering item wording by mistake) or systematic bias (e.g. one interviewer consistently records respondents' answers inaccurately); and produce bias due to the social interaction between the respondent and interviewer (e.g. social acquiescence or characteristics of the interviewer may affect both response rates and the nature of responses given) (244). However, the advantages of administration make this method preferable for elderly or chronically ill populations. Specifically, the burden of completion lies with the interviewer, thus, response rates are typically higher, reducing non-response bias; interpersonal interaction (interviewer can provide clarification or explanation); it enables those with reading or writing difficulties to be included in research; it enhances data quality (i.e. less missing data, facilitation with visual aids such as prompt cards for response options and checking data for completeness); and reduces sample bias (ensures that information is actually collected from the target sample) as self-completion often relies on self-selected samples (e.g. those who actually complete and return instruments), with findings not necessarily generalisable to the target population (244).

The current evidence is inconsistent in differentiating the superiority of any one administration method in terms of quantity and quality of response (238, 244). Therefore, the selection of administration mode should be made on a case-by-case basis, taking into account the research population, topic and setting, anticipated response rates, acceptability, time available, and financial budget. Patients with PUs may have difficulty with self-completion (i.e. acutely ill, elderly), but as the PU-QOLI is intended for PU intervention effectiveness research (there is cost-benefit associated with self-completion methods in clinical trials that require large samples (245)), the decision was made to develop a self-completed version in the first instance. The suitability of this method was determined during pretesting.

5.6.2 Design and layout

The design and layout of PRO instruments can effect completion; therefore, careful consideration of item selection, framing and sequencing, time frame, response categories and instructions is needed. Few studies have explored design and layout in the context of health-related topics (244). In the absence of empirical evidence, advice regarding questionnaire design (i.e. length, pagination, colour), supported by theories of cognition, perception and pattern recognition (246), was considered. Based on recommendations, the PU-QOLI was designed as a double-sided A4 size booklet on white paper. Font size 12 was chosen as respondents are largely elderly people with some visual impairment. Questions were grouped into item sets and numbered, not crowded or split between two pages with horizontal response formats attached; and ended with a thank you (247).

Selecting the questions: question blocking and item sequence

Three important issues were considered when selecting items: operationalisation, measurement continuum, and the number of items. Scales are constructed through a process known as operationalisation, where logically related items are grouped or blocked. Blocked items may be prone to order effects but this method allows for the sequencing of items within a scale and focuses the mind to recall events about one topic at a time (248). Following item grouping, the items within each group/scale were ordered in a logical sequence to map out a discernable line of increasing intensity. What is implied is that as people, for example, experience progressively more pain because of their PU, they move along the pain continuum from left to right and if their pain reduces they move from right to left.

Multiple item scales improve validity, responsiveness and precision in measurement as complex variables are broken down to their component parts (items) but there needs to

be a balance between asking many items to allow a detailed assessment of each variable but not too many to risk respondents getting bored, fatigued, and increasing respondent burden (244). This is particularly important if PROs are to be used in longitudinal research or on-going clinical assessment as completing lengthy instruments over and over may threaten compliance (i.e. follow-up data is not returned or loss to follow-up bias is introduced into the analysis). The intention was to keep the overall length of the PU-QOLI as short as possible but ensure that each construct was adequately operationalised with enough items to mark out constructs on a continuum. An inclusive approach was taken to ensure all aspects relevant to the construct and important to patients were included. The choice of items was subsequently pretested (section 5.7). The ordering of items was tested in later stages (Chapter 6) to determine whether a discernable line of increasing intensity had been mapped out by the items.

Framing of questions

Guiding principles for framing questions for new instruments (132, 247, 248) were consulted. Questions that were too specific, hypothetical, employ technical words or jargon, double-barrelled, misleading, vague or ambiguous, contained acronyms, abbreviations or pertained to changes were avoided. The FDA guidance for developing PRO instruments (86) recommends that items should adequately cover the domains under investigation, relate to the instruments objectives, use words familiar to patients, and not be confrontational, upsetting or ambiguous. These recommendations were considered when constructing the PU-QOLI to ensure clearly formulated and precise items.

Question Sequencing

- *Question order*

The ordering of questions has an effect on the responses given (244). Roberson and Sundstrom (249) posit that one of the most important aspects of questionnaire design relates to early items. Opening questions should be easy, non-threatening and salient to respondents. Opening with items relevant to the respondents' circumstances highlights to them the apparent relevance of the questionnaire, encouraging completion. Two descriptive questions pertaining to PUs (i.e. how many PUs do you have and on what part of your body?) were placed on the front page following the instructions to focus patients' thinking about the medical condition of interest. This is particularly important as people who develop PUs usually have a multitude of medical problems. Based on qualitative work with patients with PUs (Chapter 3), PU symptoms were salient for all patients and therefore the decision was made to place symptom questions at the beginning of the PU-QOLI. Beginning with symptom questions set the

context for later questions and ensured a logical sequence of questioning (250). Demographic questions were placed at the end as they were not considered part of the measurement.

- *Sensitive items*

Sensitive or threatening questions may result in misled, understated or exaggerated responses (249). Questions concerning intimacy and body image may be viewed by patients as sensitive. A method of reducing sensitivity is to embed sensitive items within the body of a questionnaire (247). Hence, items pertaining to intimacy and body image were placed toward the middle of the PU-QOLI.

Time-frame

The accuracy with which participants respond to questions is influenced by response or memory error (failure to recall an event within the specified time-frame). Respondents recall and count relevant events when formulating answers to survey questions (251). Setting a too long time-frame may result in under-reporting of events occurring in the past. For HRQL instruments, a short time-frame may not be appropriate as changes in disease impact or treatment burden may not have occurred to warrant assessment. As such, important disease changes/progression and memory error (recall bias) need to be considered when choosing a time-frame. A recall period of the past week was chosen on clinical grounds, as changes in PU severity and symptomology often occur over days and thus a longer recall period would risk not capturing relevant changes to HRQL. Events that occurred over a month ago may no longer be relevant or have been resolved/treated.

Response category format

Having devised a set of questions (scales with multiple items), a response method had to be chosen. There are various response methods and deciding on which method is most appropriate depends largely on the nature of the questions asked. In healthcare research, responses to outcomes are generally continuous rather than categorical. Approaches that attempt to quantify continuous information include Direct Estimation techniques (132). These methods are designed to elicit from participants a direct quantitative estimate of the magnitude of a latent variable, where responses are indicated by a mark on a line such as in visual analogue scale (VAS) or box checking as with Likert scales. A VAS is a line of fixed length, usually 100mm, which is often used to measure pain severity or intensity levels. Using VASs to measure HRQL may not be an optimal method as they assess a variable with a single item, potentially yielding less precise measurement than other methods (132). Likert scales quantify

data on a continuum but rather than marking on a line, a number of discrete response options are offered along a multiple item continuum and respondents are required to select a response most closely representing their experience (85, 132). The Likert scale rating method is commonly used in PRO measurement and was the chosen method for the PU-QOLI.

When constructing Likert scales, the number of response categories to use and how they should be labelled needs to be considered. Deciding on the number of response options to include is a balance between maximising precision and minimising bias and respondent burden (85, 132). Category labels are intended to elicit judgements about the frequency of occurrence of an event or intensity of a feeling. FDA guidance (86) suggests that wording used in response options should: be clear and appropriate to the question stem and intended population; offer a clear distinction between choices; be appropriately ordered and appear to represent equal intervals; avoid potential ceiling or floor effects; and not bias the direction of responses. Based on these recommendations, the response descriptors chosen for PU-QOLI related to the amount of bother of a particular variable rather than frequency (e.g. a symptom might be frequent but might not necessarily cause bother). Patients were required to respond to items in terms of the amount of bother attributed (e.g. “How much have you been bothered by…?”) and each item used a discrete response option category scored with successive integer scores (e.g. 0 = no bother at all to 3 = a lot of bother) that implied a continuum of increasing impact (bother), from less (no bother) to more (a lot of bother). This assumption was tested by examination of threshold ordering in subsequent testing (see Chapter 6). Item response descriptors were kept on the same differentiating scale and in the same order to reduce respondent burden and confusion caused by reversed responses. Numbers were placed beside boxes to reduce ambiguity and reinforce that responses were on a continuum, and a ‘tick box’ instruction was chosen for all items to ensure consistency in the method of responding (252).

Introduction and instructions

Instructions for completing questionnaires need to be clear and unambiguous, using simple, jargon free language that will be easily understood by respondents or interviewers. Consistent with recommendations (247), general information pertaining to what the PU-QOLI is about and instructions about how questions should be answered were placed at the beginning of the PU-QOLI, while instructions specific to individual questions were placed close to the relevant question. Instructions were made brief and clear, and bold font used to highlight important components (e.g. *during*

the past week and *tick all that apply*). A statement to ensure confidentiality was included to encourage accurate reports.

Expert appraisal of preliminary PU-QOLI

Prior to pretesting the draft PU-QOLI was reviewed by an expert group of health outcome methodologists (Section 5.4.1.2) with consideration of question wording, meaning, relevance and format. It was hoped that this would lead to the detection of problems with design or items that could pose potential problems for respondents and threaten response rates or data quality.

Amendments made to the draft PU-QOLI included: reordering of questions (i.e. question order included symptoms then physical limitations, psychological impact and finally social participation as this was considered conceptually logical); inclusion of introducing sentences for each question; single items were separated out as previously they were combined into one grouping; and an overall QoL question, with responses based on other commonly used measures (SF36), added. Amendments produced a preliminary version ready for pretesting (Appendix 5.1).

The preliminary PU-QOLI consisted of 118 items grouped into 10 scales (covering 4 domains and 13 sub-domains): pain (17 items), exudate (12 items), odour (10 items), sleep (7 items), malaise (3 items), mobility (11 items), activities of daily living (ADL; 11 items), mood (9 items), anxiety (8 items), self-consciousness and appearance (9 items), autonomy (5 items), isolation (5 items) and participation (11 items). Three single item questions were included (items that did not fit into scales or were descriptive items). Items considered similar in content were grouped together so that respondents could consider the amount of similarity/difference in wordings. The PU-QOLI was intended as a self-completed instrument and patients rated the amount of bother attributed (e.g. “During the past week, how much have you been bothered by…?”) on a 4-point response scale (e.g. 0=not at all – 3=a lot).

5.7 Pretesting preliminary PU-QOLI

The first part of this chapter described methods used to construct a preliminary PU-QOLI version. This part describes pretesting the PU-QOLI, which is key in PRO instrument development. It is a process for evaluating patients’ understanding of the items, instructions, response options and recall period, determining whether readability is appropriate for the target population, and confirming completeness of concepts covered by items (86). These steps ensure items are appropriately worded, unambiguous and relevant.

The mainstay of pretesting has traditionally been qualitative methods based on cognitive theory (i.e. cognitive interviewing or debriefing). Cognitive theory techniques such as principles of Cognitive Aspects of Survey Methodology (CASM) (253) are applied to investigate and understand response problems in surveys to improve the quality of data collected (254) and the design of questionnaires by evaluating sources of response error (255). Techniques are used to understand how respondents process and respond to questionnaire items (i.e. the underlying cognitive processes employed in reading, comprehending and interpreting questions and formulating answers), with the intention to improve the design of PRO instruments and improve data quality (86, 254, 256).

There are a number of models which are comparable in processes that explain the underlying cognitive processes involved in completing questionnaires. The most general, Tourangeau's (251) four-stage model, posits that respondents follow a sequence when completing questionnaire items, involving comprehension, retrieval, estimation/judgement and response. This four-stage processing model is appropriate for application to HRQL as the model aims to explain the cognitive processes employed by respondents when trying to understand and respond to questions dealing with complex concepts such as QoL.

More recently, guidance for developing PRO instruments (86, 102, 120, 240) has emphasised the importance of pretesting using cognitive methods to ensure that PRO instruments are understood and relevant to the target population. Cognitive testing is important for identifying problems with PRO instrument items (e.g. lack of clarity), layout and administration. Better formatting and phrasing improves comprehension, ensures less time and effort to complete, and increases respondent participation and compliance, thus reducing item non-response and missing data (257) and establishing content validity (86).

Cognitive methods offer a well-established method of improving the precision and content validity of newly developed PRO instruments. However, one avenue which has had less investigation is the use of quantitative methods in the pretesting process. Resultant scales are eventually examined against quantitative criteria, but not usually employed in pretesting due to the traditional psychometric sample size requirements (258, 259) for confident inferences to be made. However, new psychometric methods, such as Rasch measurement methods (260), afford the possibility of being used in pretesting as relatively stable estimates are possible even in small samples ($n \geq 30$) (261) (see section 5.7.10 for details).

Pretesting using Rasch methods provides a useful early view for checking whether items can be combined using the test of fit between the data and Rasch model. Misfit indicates that combining a set of items may not be justified (e.g. an item is not measuring the same variable), but rather than remove the items, the cause of misfit should be explored and explained in the first instance (85, 262). Application of the model early in the development stage of new PRO instruments may provide a useful method for exploring the choice of items in specific scales, item scoring, suitability of response options, and appropriateness of combining items into scales. It may also be useful in demonstrating the extent to which scale scores are free from random error (reliability) and measure the variables they purport to measure (validity). In such contexts, application of the Rasch model can be used as a diagnostic tool for determining how well items, scales and response options work to measure what they are intended to measure (see Chapter 6, Section 6.5.3 for details).

5.7.1 Aims

The aims of this study were two-fold:

1. Pretest preliminary PU-QOLI version, using cognitive interviewing methods, to reduce respondent burden, decrease data errors and non-response due to poor questionnaire design, layout and unclear, misunderstood or irrelevant items, and to ensure that the PU-QOLI was relevant to and understood by people with PUs
2. Rasch analysis to identify and resolve problems with layout, time frame, response options, framing of items and administration, and confirm content (i.e. need for additional items or elimination/rewording of other items) prior to formal psychometric evaluation

Part of the second aim was to test the added value of Rasch measurement methods in detecting problems with items, scales and response options by comparing Rasch findings with the findings from the cognitive interviews for consistency. The perspective of people with PUs was central in all aspects of pretesting.

5.7.2 Study design

Cognitive processes involved in completing the PU-QOLI were investigated through structured face-to-face cognitive interviews. Face-to-face interviews are preferable as they allow observation of non-verbal cues and provide a natural interchange between patient and interviewer. Cognitive interviews were undertaken to elicit feedback from respondent's on their understanding of individual questions, associated response options, and instructions, and to verbalise how they had gone about producing their

answers (263). Emphasis was on comprehension (i.e. clarity, language), retrieval from memory and response judgements (i.e. frequency judgments, logic decisions). Interviews were conducted in patients' homes, clinics or wards, as determined by the patient's circumstances at the time of interview.

5.7.3 Study population

The participant sample was intended to represent the PU population. Therefore patients from 11 acute and community services across England were recruited from April 2009 to September 2009 if they: had a PU of any grade (148), duration or location; were aged ≥18 years; from hospital, rehabilitation or community settings; and able to read and write in English. Patients who: did not have a PU; were unconscious, confused or cognitively impaired; unable to speak English; or deemed ethically inappropriate to approach (e.g. death was imminent) were not eligible.

5.7.4 Recruitment and consent procedure

Participants were purposively sampled according to key factors: age (under 70 years and over 70 years); PU severity (superficial, grade 1–2 and severe, grade 3-4) and location (torso and limb sites); healthcare setting; and experience of different PU treatments, with a minimum of five patients per factor consecutively sought. All patients with PUs were screened for eligibility. Age, gender and PU grades and locations were recorded on screening logs. Refusals were recorded and reasons for refusal were noted if known. The purpose of collecting data on non-participants was to check for sample bias and to assess the generalisability of the participating sample against the general PU population. Recruitment continued on a rolling basis until no new problems with the preliminary PU-QOLI emerged (i.e. saturation).

Members of tissue viability teams (TVT), including the local Principal Investigators, tissue viability nurse specialists, nurse consultants and other members of local clinical teams (i.e. tissue viability and clinical research nurses) at participating sites identified participants. Those meeting eligibility were approached, informed about the study and provided with a project information leaflet detailing the rationale, design, and personal implications of the study, and an 'agree to be contacted by researcher (CG)' form to be either contacted by telephone or visited at the ward.

Following information provision, participants had as much time as they needed to consider participation and return researcher contact forms. Upon receipt, the researcher telephoned the participant and provided further study information, answered

questions, and following verbal consent, arranged an interview for a mutually convenient time. For in-patients who could not be contacted by telephone, the TVT member, with the patient's permission, liaised with the researcher and patient to arrange a time for a ward visit.

Prior to commencing the interview, the researcher gave each participant a verbal explanation of the study, informed them that the interview would be recorded but any identifiable information would remain anonymous, reminded them that participation was completely voluntary and that they could withdraw at anytime without it affecting their care, and asked them to sign a consent form. The study received ethical approval by a UK National Health Service Research Committee. All patients gave written informed consent to participate.

5.7.5 Cognitive interviewing methods: Think-aloud and verbal probing

The mostly commonly applied cognitive interviewing methods are think-aloud and verbal probing techniques. The think-aloud method, derived from psychological procedures described by Ericsson and Simon (264), explicitly instructs participants to "think aloud" or vocalise the thoughts that occur concurrently as they read and provide answers to questions. This method provides information about the cognitive processes respondents employ when completing questionnaires, including comprehension and language (whether the question is worded and understood in the way it is intended), memory (how information is recalled), and problem-solving (whether respondents recall information or simply guess). An alternative to think-aloud is use of verbal probing, where the interviewer asks questions about the processes going on while respondents complete questions, either concurrently or retrospectively. This method elicits respondents' understanding of questionnaire items by asking respondents to say aloud questions in their own words, explain: i) what specific words mean to them and ii) their responses, and identify where difficulty in understanding, interpretation and completion occur (265). Specifically, misunderstandings, incomplete content coverage, and inconsistent interpretations are explored.

Concurrent probing minimises recall bias as the information requested is fresh in the respondent's mind at the time of probing. However, think-aloud speech and concurrent probing may serve to burden or contaminate the cognitive processes used in answering questionnaire items. Switching between tasks can be distracting and may affect responses to subsequent questions. The alternative is debriefing the participant by probing following questionnaire completion. This allows the natural flow of an interview, but risks participants failing to recall what they were thinking when answering

a question, introducing recall bias and hindsight effects (266). To overcome limitations, both methods were combined for the cognitive interviews (see below).

The key differences between the two techniques are (267, 268):

Think-aloud	**Probing**
• Respondent-driven; open-ended format; free from interviewer-imposed bias, but tendency for respondent to wander off-track, spend significant time on one question, delve into irrelevant areas	• Interviewer–driven; avoids discussion that may be irrelevant and non-productive; interviewer can focus on particular areas that appear to be relevant as potential sources of response error
• Lower burden on interviewer as respondent does most of the talking; minimal interviewer training required	• Lower burden on respondent as interviewer does the questionning; minimal respondent training required
• Thinking-aloud places the main burden on respondent, making the interview more difficult for them; need for subject training as many are not proficient in thinking-aloud	• Interview is easier for respondents, however risks artificial dialogue where interviewer simply administers questions and respondent answers, leaving little room for spontaneous dialogue
• Bias in subject information processing: a considerable amount of mental effort is invested into processing questions when thinking-aloud, relative to when simply answering questions.	• Use of probes are a potential for bias, need for careful selection of "non-leading" probing techniques that minimise bias

5.7.6 Conducting cognitive interviews: Procedure and Data collection

Cognitive interviewing techniques were employed to gain understanding of how respondents interpret questions and whether questions are understood as intended. All participants completed the preliminary PU-QOLI without researcher assistance.

Two interviewing techniques were employed; however, the first three participants asked to think-aloud (spontaneous conversation) while completing the PU-QOLI reported that the method made completion difficult. Therefore, the remaining participants were instructed to flag/mark items they found confusing/difficult to understand, upsetting/intrusive or annoying while completing the PU-QOLI. Following

completion, going item-by-item and guided by an interview schedule consisting of open-ended questions and scripted probes to ensure standardisation across administration, the researcher used de-briefing questioning to elicit the cognitive processes used by respondents while completing the PU-QOLI. Probes based on the Tourangeau (1984) cognitive model included: comprehension (i.e. patient interprets questions by determining what s/he believe the question to be asking based on what specific words/phrases mean to them); retrieval (i.e. relevant information searched from long-term memory to enable a response and the strategies used to retrieve the information); estimation/judgement (i.e. information retrieved from memory is evaluated for its relevance to the question and judged for completeness); and response (i.e. the initial response is considered for consistency/acceptability, and the internally generated answer [judgement] mapped onto a response category) (see Table 5.1 for examples).

Probes were carefully developed to ensure unbiased phrasing (i.e. ensure probes were not leading respondents) and sought comments about questionnaire design (i.e. item stem, response options, instructions, format/layout, time frame) and specific items anticipated to be problematic (i.e. "What does *feeling emotionally close* mean to you?"). Spontaneous probes were used when responses led to the interviewer wanting additional information or clarification (i.e. follow-up on issues that emerged or probing for additional information: “Can you tell me more about that”).

Tabel 5.1 Example probes with cognitive components

Cognitive component	**Interview probe**
Comprehension	Can you repeat the question in your own words?
	What does the word X mean to you?
Retrieval	How did you come to think of that?
	Did you have a particular time period in mind?
Judgement	How did you arrive at your response?
	How well do you remember this?
Response	How did you feel about answering this question?
	Were you able to find your first answer to the question from the response options?

During cognitive interviews the researcher took notes, fed-back to patients to ensure comprehension of responses and reviewed recorded interviews, making notes on structured data extraction forms. Interviews lasted for a mean of 46 minutes (range 28 to 112 minutes).

5.7.7 Data analysis schema

An analysis schema was developed based on the Question Appraisal System (QAS-99). The QAS-99 (269) is a coding tool that has been used for pretesting instruments (270, 271) that focuses on the cognitive demands required for answering a question and potentially problematic item characteristics that may lead to response error including: content, instructions, layout, length, time frame and response options (Table 5.2). The intention was to develop a systematic way of categorising the processes underlying responses to PU-QOLI items. Specifically, dominant trends across interviews (i.e. problems that occurred repeatedly) and key findings (i.e. problems identified in a single interview, but considered problematic) were considered.

Table 5.2 QAS-99 categories (269)

Step 1:	READING: Determine any difficulties for interviewers in reading questions uniformly to all respondents
Step 2:	INSTRUCTIONS: Determine any problems with introductions, instructions or explanations from the *respondent's* point of view
Step 3:	CLARITY: Identify problems related to communicating the *intent or meaning* of the question to the respondent
Step 4:	ASSUMPTIONS: Determine any problems with assumptions made to the underlying logic
Step 5:	KNOWLEDGE/MEMORY: Check whether respondents are likely to *not* know or have trouble remembering relevant information
Step 6:	SENSITIVITY/BIAS: Assess questions for sensitivity, wording and bias
Step 7:	RESPONSE CATEGORIES: Assess the adequacy of the range of responses to be recorded
Step 8:	OTHER: Detect any problems not identified in Steps 1-7

5.7.8 Subsequent rounds cognitive testing

To maximise benefit from cognitive interview, multiple interviewing rounds were proposed. Once major conceptual problems were identified and addressed from early interviews, later rounds of interviewing focused more exclusively on the appropriateness of individual questions, testing items in the context of the questionnaire, for example, in terms of clarity, appropriateness of item series, or biases due to question ordering (268). An iterative process was adopted during data collection and analysis, to saturation, using patients' qualitative reports from three rounds of interviews to modify the PU-QOLI. Following the first round (n=10), major problems emerged that required substantial revision (i.e. modifications made to PU-QOLI) prior to subsequent interviews. A second round of interviews were undertaken (n=10) to test the changes made and to provide additional testing of questionnaire aspects not previously explored. A final round (n=15) was undertaken to ensure that

no additional problems were reported or introduced with changes made, and previously identified problems had indeed been rectified. Misunderstanding of item stem, response options or instructions (i.e. vague wording, complicated syntax); unclear wording (i.e. used expressions like should, needs to, must); and negative comments about an item were considered a problem.

5.7.9 Expert appraisal and revisions to PU-QOLI

As there is no standard method for using cognitive interview data to modify PRO instruments (270), the outcome methodologists (Section 5.4.1.2) discussed and resolved aggregated findings (both within and across interviews) after each interview round in an iterative process. This was done on a consensus, item-by-item basis, to decide whether to retain, revise, eliminate or add items or make changes to design and layout, with particular weight given to the same comments by several respondents. Occasionally, a single negative remark led to an item revision (e.g. a remark signalling a serious misunderstanding of an item). A group of clinical experts (Section 5.5.1) also reviewed PU-QOLI revisions to ensure clinical relevance. Expert appraisal assisted in avoiding bias that would be introduced when relying solely on the judgement of one researcher in determining the implications of the cognitive interview findings.

5.7.10 Rasch measurement methods– a 'quantitative de-briefing'

Unlike traditional psychometric methods that are based upon correlational or descriptive analyses, Rasch analysis provides a formal method of testing PRO instruments or rating scales (e.g. health scales considered unidimensional) against a mathematical measurement model developed by Danish mathematician Georg Rasch (260). The Rasch model defines how a set of items should perform to generate reliable and valid measurements (272) and evaluates the legitimacy of summing items to generate measurements (260, 273). In a Rasch analysis, the extent to which observed data (patients' actual responses to scale items) are concurrent with ('fit') predictions of those responses from the Rasch model are examined; whereby the difference between expected and observed scores indicates the degree to which rigorous measurement is achieved (274). The expected response structure for the Rasch model is a probabilistic Guttman pattern, which assumes that for the same person ability, the probability of endorsing an easy item is higher than the probability of endorsing a more difficult item, and vice versa (275). When a rating scale is used to discriminate between persons with different abilities, someone with higher ability is expected to affirm all items endorsed by a person with lower ability in addition to items representative of higher ability.

Rasch measurement methods (260) were used to undertake a preliminary analysis of the PU-QOLI investigating items within the context of the instrument, response options, appropriateness of item series and question ordering (item-fit). The Rasch model for ordered response categories (276) was used and analyses were performed using RUMM2030 (277), including: targeting of sample to items (i.e. is the scale-to-sample targeting adequate for making judgements about the performance of the scales and the measurement of people?), ordering of response options (i.e. are respondents using the response options provided in a manner consistent with the level of the construct being measured?–demonstrated by ordered thresholds (278), item-fit statistical indicators (i.e. fit residuals within +/-2.5 and non-significant chi square values provide evidence that scale items contribute to defining the construct measured (279)), and the spread of item locations (279, 280)). Items should ideally spread out evenly over a reasonable measurement range and be appropriately targeted to the people they are measuring. Items with similar locations on the continuum indicate redundancy. Details of the psychometric criteria are described in Chapter 6, Section 6.5.3.

For the purpose of pretesting, Rasch measurement methods were used as a diagnostic tool at an early stage in instrument development, whereby problems with items, scales and response options could be evaluated against psychometric criteria and findings compared with cognitive interviews for consistency to determine the value of Rasch methods in pretesting.

5.8 RESULTS

Patients (n=134) were screened for eligibility from 6 hospitals, 4 community Trusts and 1 hospice. Patients ranged in age from 32 to 98 years, with 66% aged 70 years or older (7 did not have age recorded); 94 were male (57%); and all were of white ethnicity. PUs varied in severity (representation of all grades including mixed) and location (both torso and limb sites). The screening process is detailed in figure 5.1.

Recruited patients (n=35) ranged in age from 36 to 85 years (mean age 65 yrs) with half (49%) aged 70 years or more. Of these, 16 (46%) were male and 18 (51%) had additional chronic conditions (e.g. SCI, multiple sclerosis). Patients represented different settings (23 hospital/rehabilitation, 12 community), PU severity (13 superficial, 18 severe, 4 mixed severity), duration (2 weeks up to 5 years) and skin site (33 sacrum/buttocks, 13 heel, others on lower back, groin, hips, back of thighs, ankles).

Tables 5.4 and 5.5 present a summary of findings from interviews and Rasch analysis, respectively, and instrument modifications. Flow diagram 5.2 illustrates the integration

of the two methods in an example with one scale to help clarify the iterative nature of the approach and how both methods informed modifications to the PU-QOLI.

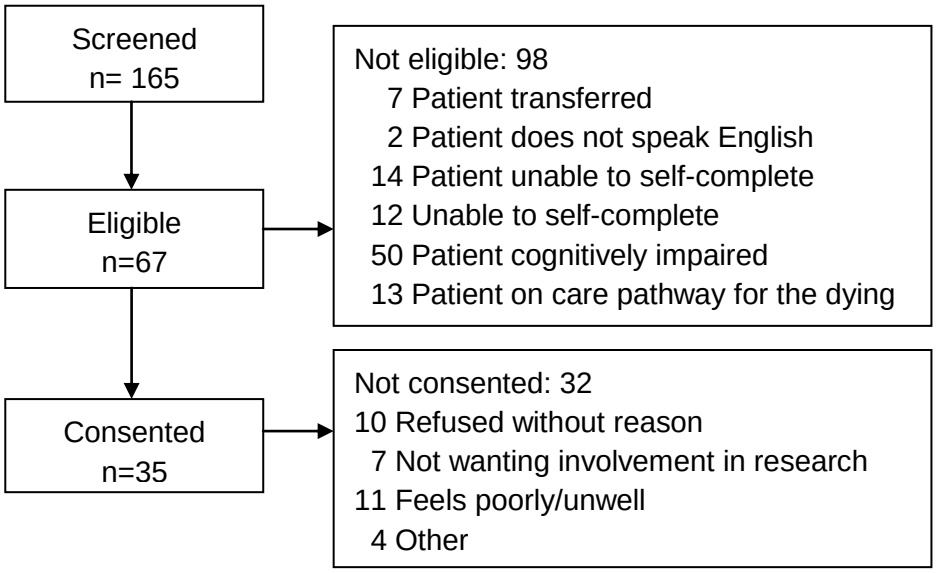


Figure 5.1 Detailed Assessment Process for pretest

5.8.1 Cognitive Interviews

Cognitive interviews identified five problem areas: content, mode of administration, instruction/layout/length, recall period and response options.

Content

Based on patients' comments, 18 items reported to be difficult to understand, confusing or unclear were revised, six items considered redundant were merged and 25 irrelevant items removed (i.e. items that >40% of respondents indicated not relevant; a crude estimate consistent with standard psychometric criteria for acceptability (281, 282)).

For example, patients reported that there were too many items about odour: "How many different words do you need for smell, you could remove a lot of these…" Other words used were too sensitive (dirty smell) or not commonly understood (foisty). Revisions focused on using words 'that patients' use (e.g. pressure sore instead of pressure ulcer). As patients reported that all important aspects of the impact of PUs on HRQL were covered, no new items were added. Review of content resulted in a revised 87-item PU-QOLI (Table 5.3) grouped into 10 scales: pain (11 items), exudate (8 items), odour (6 items), sleep (6 items), vitality (3 items), mobility/movement (11 items), ADL) (9 items), mood (7 items), anxiety (3 items), self-consciousness and appearance (7 items), autonomy (3 items), isolation (4 items), participation (9 items).

Flow diagram 1 Example using the Daily Activities Scale to demonstrate the mixed methods approach, including findings and modifications

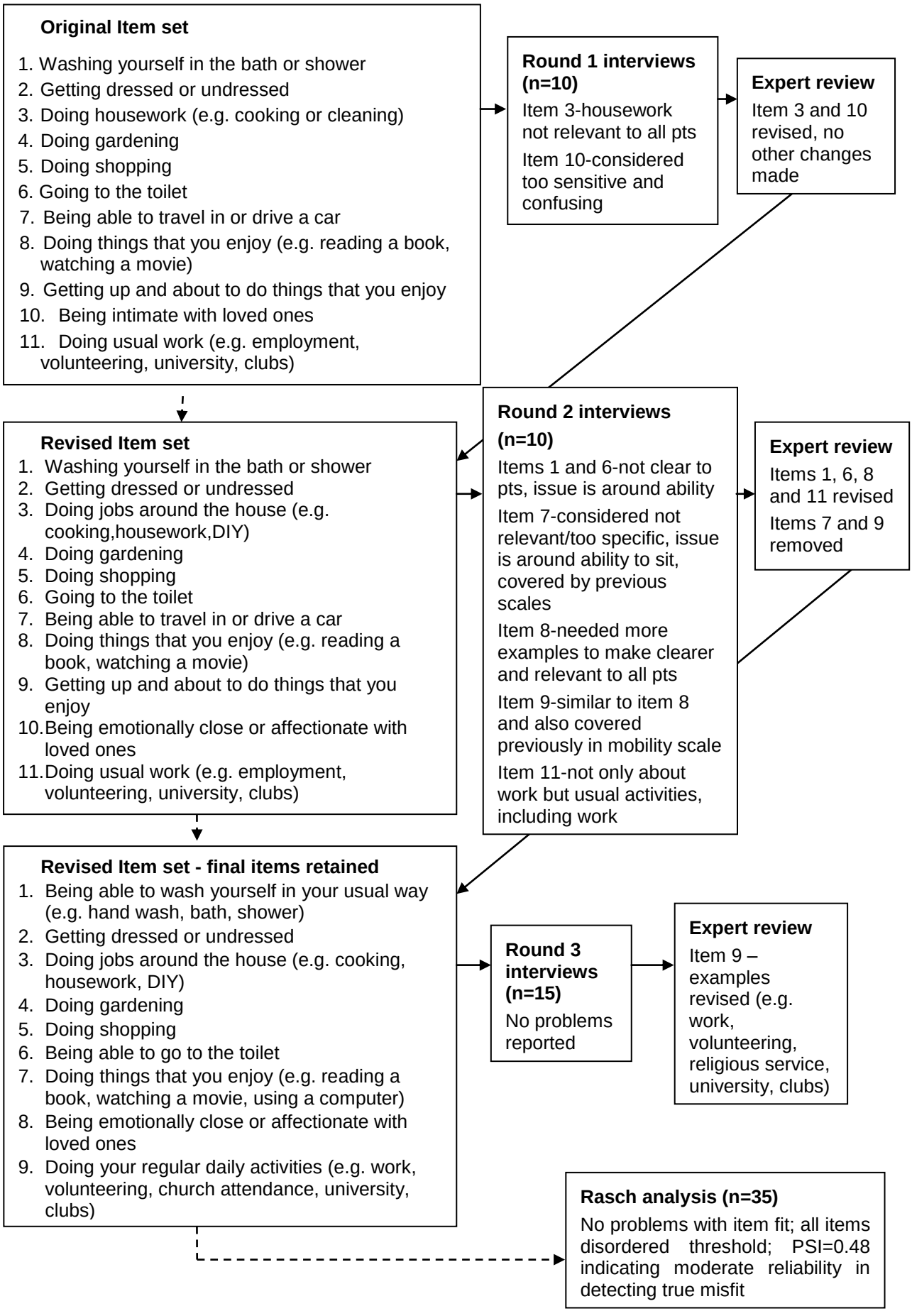


Table 5.3 Summary of item changes at each cognitive interview round based on patient reports and consensus discussion

Item at pretest	Round 1	Round 2	Round 3
Pain Index items			
1) Feeling uncomfortable	No change	No change	Retained
2) Annoying pain or discomfort	No change	No change	Retained
3) Itchiness	No change	No change	Retained
4) Tenderness	No change	No change	Retained
5) Niggling	No change	Removed	
6) Soreness	Removed		
7) Aching	No change	Revised - Dull ache	retained
8) Pins and needles	Removed		
9) Tingling	No change	No change	Retained
10) Throbbing	No change	No change	Retained
11) Nagging	Removed		
12) Shooting	Removed		
13) Stinging	No change	No change	Retained
14) Stabbing	Revised – Stabbing pains	No change	Retained
15) Electric shocks	Removed		
16) Red raw	No change	No change	Retained
17) Burning	No change	No change	Retained
Exudate Index items			
1) Weeping	No change	No change	Retained
2) Oozing	Removed		
3) Running	No change	No change	Retained
4) Sticky	No change	No change	Retained
5) Slimy	Removed		
6) Wet	Removed		
7) Messy	No change	No change	Retained
8) Staining	No change	No change	Retained
9) Causing dressing to come off	No change	No change	Retained
10) Gungy	Removed		
11) Pus	No change	No change	Retained
12) Bleeding	No change	No change	Retained
Odour Index items			
1) An unpleasant smell	No change	No change	Retained
2) A lingering smell	No change	No change	Retained
3) A dirty smell	Removed		
4) A foisty smell	Removed		
5) A stench	Merged – A stench or stink	No change	Retained
6) A stink			
7) A stale smell	Removed		

Item at pretest	Round 1	Round 2	Round 3
8) A pungent smell	No change	No change	Retained
9) A sickening smell	No change	No change	Retained
10) A putrid smell	No change	No change	Retained
Sleep scale			
1) Trouble falling asleep	No change	No change	Retained
2) A restless sleep	Revised – Interrupted sleep (e.g. restless sleep, woken up during your sleep)	No change	Retained
3) Being kept awake	No change	No change	Retained
4) Being woken up during your sleep	Merged with b)		
5) Not getting the amount of sleep that you needed	No change	No change	Retained
6) Having to sleep in one position (e.g. your back or side)	No change	No change	Retained
7) Trouble finding a comfortable position	No change	No change	Retained
Mobility/movement scale			
1) Difficulty sitting up in bed	No change	No change	Retained
2) Difficulty adjusting yourself in bed	No change	No change	Retained
3) Difficulty turning in bed	No change	Revised – difficulty turning or moving around in bed	Retained
4) Difficulty pushing up to a sitting position	No change	No change	Retained
5) Difficulty sitting up in one position for long periods	No change	No change	Retained
6) Difficulty standing for long periods	No change	No change	Retained
7) Difficulty transferring from a bed to a chair	No change	No change	Revised – difficulty transferring (e.g. from a bed to a chair or to a car)
8) Feeling limited in your ability to walk	No change	No change	Retained
9) Feeling limited in your ability to go up and down stairs	No change	No change	Retained
10) Feeling limited in how far you were able to walk	No change	No change	Retained
11) Feeling that your walking was slowed down	No change	No change	Retained
Vitality/Fatigue scale			

Item at pretest	Round 1	Round 2	Round 3
1) Feeling that your appetite has reduced	No change	No change	Retained
2) Feeling unwell or poorly	No change	No change	Retained
3) Feeling that your energy levels have been reduced (i.e. feeling tired or fatigued)	Revised - Feeling that your energy levels have been reduced (i.e. tired, fatigued)	No change	Retained
ADL scale			
1) Washing yourself in the bath or shower	No change	Revised – being able to wash yourself in your usual way (e.g. hand wash, bath, shower)	Retained
2) Getting dressed or undressed	No change	No change	Retained
3) Doing housework (e.g. cooking or cleaning)	Revised – Doing jobs around the house (e.g. cooking, housework, DIY)	No change	Retained
4) Doing gardening	No change	No change	Retained
5) Doing shopping	No change	No change	Retained
6) Going to the toilet	No change	Revised – being able to go to the toilet	Retained
7) Being able to travel in or drive a car	No change	Removed	
8) Doing things that you enjoy (e.g. reading a book, watching a movie)	No change	Revised - Doing things that you enjoy (e.g. reading a book, watching a movie, using a computer)	Retained
9) Getting up and about to do things that you enjoy	No change	Removed	
10) Being intimate with loved ones	Revised – Being emotionally close or affectionate with loved ones	No change	Retained
11) Doing usual work (e.g. employment, volunteering, university, clubs)	No change	Revised – doing your regular daily activities (e.g. work, volunteering, church attendance, university, clubs)	Revised – doing your regular daily activities (e.g. work, volunteering, religious service, university, clubs)
Mood and Anxiety scale			

Item at pretest	Round 1	Round 2	Round 3
1) Feeling frustrated	No change	No change	Retained
2) Feeling fed-up	No change	No change	Retained
3) Feeling annoyed	Merged - feeling annoyed or irritated	No change	Retained
4) Feeling irritated			
5) Feeling bad tempered	Removed		
6) Feeling angry	No change	No change	Retained
7) Feeling miserable	No change	No change	Retained
8) Feeling down	No change	No change	Retained
9) Feeling depressed	No change	No change	Retained
10) Feeling fearful	Removed		
11) Feeling afraid	Removed		
12) Feeling upset	No change	No change	Retained
13) Feeling concerned	Merged - feeling concerned or worried	No change	Retained
14) Feeling worried			
15) Feeling anxious	No change	No change	Retained
16) Feeling surprised	Removed		
17) Feeling shocked	Removed		
Self-conscious & Appearance Scale			
1) Feeling helpless	No change	No change	Retained
2) Feeling a lack of self-confidence	Revised – feeling lack of confidence	No change	Retained
3) Feeling a lack of self-esteem	Removed		
4) Feeling self-conscious	No change	No change	Retained
5) Feeling embarrassed	No change	No change	Retained
6) Feeling physically unattractive	No change	No change	Retained
7) Feeling disinterested in socialising	No change	Removed	
8) Feeling uneasy being close to people	No change	Revised - Feeling uneasy being close to or around other people	Retained
9) Feeling worried about how others will react to your ulcer	No change	Removed	
Isolation scale			
1) Feeling physically dependent on others	No change	No change	Retained
2) Feeling left out	Removed		
3) Feeling isolated	Merged – feeling cut of or isolated from others	No change	Retained
4) Feeling cut off			
5) Feeling lonely	No change	No change	Retained
6) Feeling like you were missing	No change	No change	Retained

Item at pretest	Round 1	Round 2	Round 3
out			
Autonomy & independence Scale			
1) Feeling that people avoided you or treated you differently now	No change	No change	Retained
2) Feeling a lack of understanding from those close to you	No change	No change	Retained
3) Feeling like a burden or nuisance on others	No change	No change	Retained
4) Feeling like you have no control over your life?	Revised – feeling like you have no control over your life or your ulcer	Revised – feeling like you have no control over your life or your sore	Retained
Participation scale			
1) Difficulty going out	No change	No change	Retained
2) Being unable to meet up with others	Merged – difficulty meeting up or seeing family and/or friends	No change	Retained
3) Difficulty seeing family and/or friends			
4) Being unable to participate in family gatherings or activities	No change	No change	Retained
5) Having to plan going out around ulcer care	No change	Revised - Having to plan going out around pressure sore care	Retained
6) Being able to do things spontaneously	Removed		
7) Giving up on hobbies or leisure activities	No change	Revised - having to give up on hobbies or leisure activities	Retained
8) Being restricted to where you could go out	No change	No change	Retained
9) Being restricted to how long you could stay out	No change	No change	Retained
10) Being unable to get away for a holiday or make a trip at the weekend	No change	No change	Retained
11) The amount of time involved in caring for your ulcer	No change	Revised - The amount of time involved in caring for your sore	Retained

Mode of administration

Despite the inclusion criteria of patients being able to self-complete a questionnaire, almost half the sample (n=15; 43%) needed some assistance with completion; of

these, eight were 70 years or older. Older patients were also more likely to not respond to one or more items. Reasons for needing assistance included: too ill/weak to sit up and/or hold a pen (n=5); visually impaired/no glasses (n=2); too tired to finish completion (n=2); and comorbidities (e.g. multiple sclerosis, arthritis) prevented self-completion (n=6). As patients who needed assistance with PU-QOLI completion were generally elderly people or wheelchair users – those at highest-risk of PU development –the mode of administration was changed to interview-administered to ensure suitability across the wide spectrum of PU patients. The equivalence of self-completed and interview-administered versions was therefore explored in a mode of administration sub-study (see Chapter 6, Section 6.6.2).

Instruction, layout and length

Changes were made to the instructions, including: rewording to reflect interviewer administration; reducing content to only contain essential information; and the statement 'We understand that you may have a combination of medical problems, but please try to think about only your pressure sore(s) when answering the questions' added. Statements 'Please tick one box on each line' and '…because of your pressure sore' added to all questions. Changes to the design/layout included: less on each page; larger font; length reduced by eliminating redundant items or merging similar ones; and scales reordered where responses to one question (e.g. feeling isolated) were reported to influence responses to another (e.g. mood)).

Recall period

Interviews confirmed the appropriateness of the 'past week' recall period. In light of comments from patients who had spent time both at home and in hospital, the wording was changed to 'overall impact during the past week'.

Response options

Patients reported difficulty selecting a response option if they: i) did not feel or experience what an item referred to, ii) experienced it but not because of PUs, or iii) felt it applied in the past. Response options were revised to 'no bother', 'a little bother' and so on, to anchor patients to think about bother (impact) rather than frequency. Because some items (e.g. shopping) were not applicable to hospitalised patients and those with mobility problems, a not applicable response option (0=*I did not experience the problem/symptom because of my PU*) was added. The numeric values beside response categories were removed as patients reported these to be a confusing visual distraction. Important aspects were made bold (i.e. 'past week', 'bothered', 'PU').

5.8.2 Preliminary Rasch Analysis

The Rasch analyses did not uncover significant problems with targeting or item-fit (Table 5.4). No items had fit residuals outside the recommended range +/-2.5 and only one scale, 'isolation', had an item with significant chi-square probability. However, consistent with qualitative findings, Rasch analyses did not support the 4-point item response options. Disordered thresholds were found in 74 of 90 items, indicating that the proposed scoring function was not working as intended. A *post hoc* analysis of category probability curves (CPC), which plots scale scores on the x-axis against the probability of endorsing each item response category on the y-axis, suggested that patients were unable to reliably discriminate between options *a little bother* and *quite a bit of bother*. Item locations for all scales were generally spread over a reasonable continuum (e.g. ≥2-logit spread, Table 5.4). However, examination of item locations indicated that some items clustered at similar locations, suggesting redundancy (Table 5.5). As the Rasch analysis was preliminary, it was considered premature at this stage to make changes to the scales until further empirical evidence confirmed these issues.

Table 5.4 Summary of Rasch analysis on all pretest PU-QOLI scales

Scale (items)	**Item Locations Logits range**	**Fit Statistics Fit residuals outside +/-2.5**	**Items with Chi-Square probability significance ≥0.001**	**Person Separation Index**	**Disordered Thresholds**
Pain (11)	-1.564 – 2.460	0	0	0.960	Items 5,6,7, 9,11,12
Exudate (8)	-1.198 – 2.605	0	0	0.473	All 8 items
Odour (6)	-	-	-	-	All 6 items
Sleep (6)	-0.614 – 0.543	0	0	0.611	All 6 items
Mobility (11)	-0.713 – 0.736	0	0	0.636	Items 1,2,3,5 6,8,9,10,11
Vitality (3)	-0.150 – 0.165	0	0	0.191	Items 2,3
ADL (9)	-0.367 – 0.378	0	0	0.479	All 9 items
Mood (7)	-0.778 – 0.595	0	0	0.833	Items 2,4.6,7
Anxiety (3)	-1.164 – 1.903	0	0	0.842	All 3 items
Self-conscious (7)	-0.895 – 0.496	**0**	0	0.363	Items 1,2,4,5 6,7,8,9
Autonomy (3)	-1.358 – 2.002	0	0	-0.037	All 3 items
Isolation (4)	-5.836 – 4.238	0	1	0.972	Items 1,3
Participation (9)	-0.788 – 0.736	0	0	0.734	Items 1,2,3,4 5,6,7,8

- Not powered to perform analysis

Table 5.5 Identified problems with PU-QOLI: Comparison of qualitative reports (patients) with quantitative findings (Rasch)

Scale	**Rasch findings**	**Qualitative patient-reports**
Pain	Item 3 significant p for item-fit (Item may be a different symptom from pain; measuring different construct)	Item 5 removed (patients consistently said redundant) and item 6 wording revised
Exudate	No concerns	No problems reported by patients
Odour	Unable to perform Rasch	No problems reported by patients
Sleep	No concerns	No problems reported by patients
Mobility	Item 3 significant p for item-fit	Item 3 revised based on patient comments (item 'difficulty turning in bed' revised to 'difficulty turning or moving around in bed'). No other problems
Vitality/Fatigue	Power of item-fit too low	No problems reported by patients
ADL	Power of item-fit too low	Item 1 and 6 revised as patients indicated items were too specific and not capturing what the problem actually was
Mood	Item 4 significant p for item-fit	No problems reported by patients
Anxiety	Power of item-fit too low	No problems reported by patients
Self-consciousness	Item 9 approaching criterion	Item 6 removed (patients said redundant and too ambiguous); item 7 revised (ambiguous and sensitive); item 8 removed
Autonomy	Power of item-fit too low	Item 2 revised (ambiguous, needed clarifying)
Isolation	Items 2,3,4 item-fit p=0.000	No problems reported by patients
Participation	Item 9 item-fit p =0.009 approaching significance – (possibly item not measuring participation but another construct)	Item 4 and 5 revised (ambiguous, needed clarifying)
Other findings Response options	Main problem for all scales is disordered thresholds (6/13 scales all items thresholds disordered; 7/13 some items within scale with disordered thresholds)	11/35 items not relevant and thought a 'did not experience symptom/problem b/c of PU' was required otherwise difficult to answer with the current response options

5.9 Discussion

To provide evidence that PRO data collected by PRO instruments is credible, two general criteria should be considered: Does the data stem from an instrument that has a theoretically sound conceptual framework? Do the data meet necessary and sufficient psychometric standards? (283). An item-pool should be developed and tested against a well-founded theoretical model (87) and after thorough familiarity with relevant work in the area (283). To ensure the first criteria was met, the development of PU-QOLI scales was theoretically driven (proceeded from a conceptual framework, Chapter 3) with items derived from patients, using patient words. Items for each construct (or scale) needed to be captured and quantified into rigorous measurements. Items are considered the measurement parts of each variable (construct); items are component parts of complex variables broken down into measurements (87).

Development of PU-QOLI scales was based on first identifying main conceptual domains and then mapping the patient experience. For example, the theory postulates that the development of PUs is associated with PU-specific symptoms (pain, exudate, odour). Symptoms impact on patients' mobility and movement which affects their sleep quality, ability to perform daily activities and reduces vitality (physical functioning domains). These restrictions consequently affect patient's mood and cause anxiety and self-consciousness (i.e. embarrassment; psychological well-being domains), which lead to social isolation and restricted social participation (social functioning domains). In order to devise items into scales representing these domains, key components, as described by patients and clinical experts, were mapped onto sub-domains. Patient transcripts were revisited and other instruments' scales viewed. This process allowed the devising of a preliminary PU-QOLI version consisting of items that captured components representing PU-specific domains representative of the PU patients' perspective (e.g. what is it specifically about PUs that affects mobility – it is inability to: i) stand for long periods, ii) walk long distance, iii) sit for long periods, and so on).

Cognitive interviewing with immediate retrospective probing proved valuable for identifying problems with PU-QOLI content and design including: i) items requiring rewording, clarification or additional examples; ii) problems with instructions and self-completion; iii) inappropriate reference period for patients both hospitalised and at home; and iv) response options. Most importantly, as almost half the sample had difficulty with self-completion, the mode of administration was changed to interview-administered. If mode was not changed then a large proportion of eligible patients would be excluded in future research due to being unable to self-complete the PU-QOLI. The modified PU-QOLI was tested to ensure that no subsequent problems were

introduced after modifications had been made and to confirm acceptability of the final version (Appendix 5.2), which took around 30 minutes to complete.

The Rasch analyses complemented the qualitative results. It revealed no problems with sample-to-scale targeting or item-fit, thus supporting the summing of items to produce scale scores, but the analysis uncovered problems with the scoring function and redundant items. A large proportion of items showed disordered thresholds, the result of too many response options for patients to discriminate between. Both methods independently confirmed the problems with response options, with Rasch analyses detecting the specific point of difficulty discriminating between two response options. The Rasch analysis detected redundant items; patients reported too many items but were unable to consistently decide on which items they considered redundant. It also provided useful information for scale development, specifically support for item-fit, ordering and validity, supporting patients' views that items made contextual sense and were relevant. This important contribution ensures that PU-QOLI scale construction is underpinned with a strong conceptual base, a process central to valid measurement (87, 284). Item estimates (locations) were cross referenced and checked for stability during the full psychometric evaluation (Chapter 6, Section 6.5.3).

This was the first study to use Rasch measurement methods during early PRO development. Rasch methods provided detailed item-level diagnostic information, both corroborating the qualitative work and expanding on the information provided by pinpointing specific psychometric issues, thus demonstrating its value as an additional method to standard pretesting methods for improving PRO instruments. Had the Rasch analyses been undertaken prior to item revisions rather than retrospectively, it may have provided additional information needed for improving the scales at this stage; however it did confirm that the modifications made were appropriate and provided quantitative support for them. Items that present with misfit from the Rasch model should be explored qualitatively in subsequent rounds of interviews with the intention of providing explanations for the misfit (i.e. clinical relevance, importance to patients).

This work adopted an experimental approach that suits the Rasch paradigm, providing cross validation, for example, where patients reported problems with the PU-QOLI, the quantitative preliminary analysis verified qualitative comments made. The ultimate goal for including Rasch analysis at this stage was to provide an early quantitative view of the extent of potential problems/issues with the PU-QOLI. However, despite the issue of disordered thresholds and redundant items being flagged, all findings need to be verified and empirically strengthened in a larger sample (e.g. full psychometric

evaluation in subsequent PU-QOLI testing, Chapters 6 and 7) before any actions to make changes (i.e. reducing responses to three categories) are made.

Other limitations to the methodology need to be mentioned. In the early stages of item generation, one researcher performed the literature search and undertook patient interviews. Despite efforts made to remain open to the HRQL issues emerging during interviews, having been immersed in the literature, some bias may have been introduced. Involving both clinical and PRO outcome methodologists at the conceptual framework development, item generation and pretesting stages assisted in finalising the content map - final item pool was mapped onto conceptual domains - and produced a clear and succinct item list on which to base the preliminary PU-QOLI version. A high proportion of patients unable to take part in the study due to cognitive impairment and despite all ethnicities being eligible, only white British Nationals consented to take part. Therefore item generation from patient transcripts does not include the views of those with cognitive impairment and may not be representative of outcomes important to non-white ethnicities. However, the conceptual framework development, which formed the basis of PU-QOLI scales, was based on mixed ethnic views.

5.10 Summary of Chapter 5

The development of PU-QOLI scales proceeded from a conceptual framework. Items were derived from patients that captured all elements of the impact of PUs on HRQL, quantifying them into measurements. Careful consideration of item selection, framing and sequencing, time frame, response categories and instructions was made, producing a preliminary PU-QOLI version that was pretested with patients with PUs. The use of mixed-methods in PRO development and pretesting was effective for identifying problems with items and questionnaire design early in the development process, and for guiding changes to layout, content and mode of administration. Using both methods together provided useful information about item selection and deletion, particularly for ensuring clinical meaningfulness, importance to patients and good measurement properties. Rasch measurement methods thus provide a complementary method alongside standard qualitative pretesting, for evaluating the strengths and weaknesses of PRO instruments during the early stage of testing.

As scales are the building blocks for the conceptual framework and items within scales are the component parts of complex variables, evidence is required that they are the optimal combination of items that measure the variable they purport to measure. Therefore, the next stage required is quantitative confirmation and psychometric support for the combining of items into scales.

Chapter 6

PRELIMINARY FIELD TEST 1 – PSYCHOMETRIC EVALUATION OF THE PU-QOLI

6.1 Overview

Chapter 6 presents the methods and results of the first of two quantitative field tests undertaken to evaluate the measurement properties of the PU-QOLI scales developed in Chapter 5. A Rasch analysis was performed first on all PU-QOLI scales, followed by traditional psychometrics.

6.2 Background

Conventionally PRO instruments or rating scales have been developed and evaluated according to traditional psychometric standards derived from Classical Test Theory (CTT) (285, 286). CTT comprises a set of principals and related statistical techniques for developing and testing measures (e.g. PRO instruments) against to determine how successful they are at estimating unobservable (e.g. HRQL) variables of interest (287). However, some concerns have been raised about existing PRO instruments developed according to CTT; they may be cumbersome for respondents, burdensome for clinical use, not applicable over the continuum of care or across research settings, suffer from floor and ceiling effects, and/or lack a standardised scoring metric to allow comparisons across health conditions (87, 288, 289, 290).

More recent advances in psychometrics have seen the development and application of two independent psychometric methods: Item Response Theory (IRT) and Rasch measurement theory (see Section 5.7.10 for additional information), to supplement traditional approaches in rating scale development (290). There are some similarities between IRT and Rasch measurement theory; both postulate theories of how scores generated by rating scales relate to measurements of the variables they seek to measure and they both provide mathematical models for which measurement expectations can be formally and rigorously tested against (85). This may be the reason for confusion within the literature that the Rasch model is the IRT one-parameter model, but they both come from different origins and represent different research perspectives. The fundamental difference between the two is that proponents of IRT give primacy to the data, therefore attempting to best explain observed data by fitting it to mathematical models (essentially finding a model that fits the data). On the other hand, Rasch measurement theorists give primacy to the mathematical model,

offering a criterion for fundamental measurement (291). From the Rasch measurement theory perspective, when data does not fit the Rasch model it is further examined to understand why (e.g. response options not working as intended) and, if required, items are removed, data recollected or the construct reconceptualised. The inherent properties of the Rasch model enable it to be used as a diagnostic tool for investigating rating scales, and it is therefore the scientific perspective favoured for use in this work.

In comparison with traditional psychometric methods, the Rasch model (260) provides a powerful framework to address some of the limitations of classical methods through the assessment of a range of additional measurement properties intended to provide information about a scale's performance, and consequently developing instruments with greater precision, reduced response burden and floor/ceiling effects, and reduced cultural biases (292, 293). Ensuring key measurement requirements are met (e.g. unidimensionality, a requirement for construct validity (132)) should be established prior to the more commonly evaluated standards of measurement (i.e. reliability) (290).

Rasch measurement and traditional psychometric methods were used in the development of the PU-QOLI. Use of both methods would allow the selection of scale items that reduce patient burden and eliminate redundant items and include items that are free of bias, fit measurement model expectations and demonstrate unidimensionality (Rasch). It is important to note that some assumptions that underlie scale properties differ between CTT and Rasch measurement theory, however the objective is the same; scale items are the building blocks for the conceptual framework (283) (developed in Chapter 3).

The psychometric properties of the PU-QOLI were evaluated through two-stage field testing including a preliminary field test (evaluation of the preliminary PU-QOLI scales) and a final field test (psychometric evaluation of final PU-QOLI scales) (Chapter 7). The overall strategy and methods for the psychometric evaluation are based on methods used to develop PROs in several areas of medicine and surgery (85, 97, 182, 274, 281, 294, 295, 296). Table 6.1 presents full details of the tests and criteria used in the psychometric evaluation.

6.3 Preliminary Field Test: Analysis of preliminary (87-item) PU-QOLI

To enable the PU-QOLI to be used with confidence in clinical practice and future research it must be shown to meet psychometric standards for reliable and valid measurement. As such, a preliminary field test study was undertaken to perform a psychometric evaluation of the preliminary PU-QOLIs' scales and items, produce a

shorter version if appropriate (i.e. reduce instrument length without losing measurement precision), evaluate the measurement properties of the short PU-QOLI, and empirically investigate optimal mode of administration in a sub-study.

6.3.1 Objectives

The objectives of the psychometric evaluation of the preliminary PU-QOLI were to:

i) confirm the feasibility and acceptability of the instrument;
ii) produce a scientifically robust shorter version PU-QOLI by selecting items that perform best against robust psychometric criteria;
iii) examine the legitimacy of summing items into scales and test scaling assumptions;
iv) carry out a preliminary evaluation of the psychometric properties of the shorter item-reduced PU-QOLI; and
v) determine optimal mode of administration for the PU-QOLI (i.e. can PU-QOLI be developed for use with both self-completed and interview-administered modes or are two mode-specific instruments required?)

6.4 Methods

6.4.1 Design for Preliminary Field Test 1

The evaluation strategy (Table 2) for field test 1 was developed to assess PU-QOLIs: i) response format, item content, response bias, dimensionality and appropriate scale targeting to develop scales and identify items with poor psychometric properties for possible elimination; ii) conduct a preliminary evaluation of PU-QOLIs' scales; and iii) undertake a preliminary evaluation of the acceptability, reliability and validity of the item reduced PU-QOLI.

In addition, to address methodological issues identified during pretesting (Chapter 5), a mode of administration sub-study was undertaken to determine optimal mode of administration for the PU-QOLI. Initially the intention was that the PU-QOLI would be self-completed, however pretesting identified problems with completion rates, questioning the appropriateness of a self-completed instrument for patients with PUs, particularly those aged over 70 years.

The sub-study included response rate, data quality and differential item functioning (DIF) analyses (297) to establish measurement equivalence across two mode of administration groups (self-completed and interview-administered). Two outcomes were possible from the sub-study: 1) one instrument could be developed for both self-

completion or interview-administration or; 2) two mode-specific versions were required. Only an interview-administered version would be developed if the analysis uncovered that two mode-specific versions were required.

6.4.2 Participants

Field test 1 sample

For the PU-QOLI to be relevant to all patients with PUs, a sample of 200-250 patients with PUs were purposively sampled ensuring representation of patients across PU categories (superficial and severe) and location (torso and limb skin sites), settings (acute and community), age (under 70 years and 70 or older) and gender. The intention was to evaluate the PU-QOLI in patients with various experiences including patients treated for PUs in different healthcare settings, receiving different treatments (i.e. pressure-relief, topical), and with various acute and chronic conditions. Accrual was reviewed to ensure balanced representation of patients.

Sub-study sample

A sub-sample (60-100 patients) of field test study participants were recruited to the sub-study.

6.4.3 Sample size

No formal sample size estimation methods for evaluation of PRO instruments were found. The 'rule of thumb' sample size recommendation for psychometric analyses of new summated scales are best done with five to 10 subjects per item, to reduce the effect of chance (131, 286). Following this recommendation, if we take the longest potential summated scale, assessing pain which contains 11 items, a 110 patient sample would be required. For the Rasch analysis, a sample of 200-250 patients was sought. This estimate was based on a need for sample selection across the range of measurement. Class interval membership to five class interval groups (see Section 6.5.3 for description) of around 50 patients in each group is suggested (273, 298).

For the sub-study, it was anticipated that up to 100 patients would be required to meet the data requirement for the DIF analysis and to account for the likelihood of missing data from the self-complete group. Rasch measurement methods (260) are able to provide useful exploratory data in small samples (n>30) (261). Small sample has an effect on significance testing and further implications for the interpretation of results, however one of the unique features of the Rasch model includes parameter separatability; relatively robust estimates are possible with small sample sizes and the

stability of fit statistics in different sample sizes has previously been empirically proven (261, 299, 300, 301).

6.4.4 Eligibility

Patients from acute and community NHS Trusts around England and Scotland, with existing PUs were included in the field test and sub-study if they were hospital, intermediate care, nursing home or community patients, and: aged ≥18 years; with an existing PU of any category, location or duration; and able to provide informed consent to participate.

Patients were excluded if they had only moisture lesions, were unconscious, confused, cognitively impaired or deemed ethically inappropriate to approach (e.g. death was imminent), did not speak or understand English or unable to provide informed consent.

Sub-study

To ensure an equivalent or representative sample in both administration groups (groups need to have the same clinical presentation to perform a DIF analysis; Section 6.6.2), the eligibility criteria was adapted from the main study to include only patients able to read and write in English (e.g. patients able to self-complete a questionnaire).

6.4.5 Recruitment and consent procedures

Consecutive patients were identified and approached to participate by members of TVTs at participating trusts. A record of those identified as eligible, approached to participate, refusals, consents and questionnaire returns was made.

Informed consent was obtained prior to baseline data collection and questionnaire completion. Informed consent and data collection was undertaken by the TVT member or researcher (CG).

A verbal explanation of the study and Patient Information Leaflet was provided by the TVT member or researcher[6] (CG) for the patient to consider. These included detailed information about the rationale, design and personal implications of the study. Following information provision, patients had as much time as they needed to consider participation and were given the opportunity to discuss the study with their family and

[6] Where the researcher was involved in the recruitment and consent process, the patient was asked to give verbal permission to be approached by the researcher.

healthcare professional before being asked to take part. The right of the patient to refuse consent without giving reasons was respected.

Assenting patients were invited to provide informed, written consent to collect baseline assessment data and to complete PU-QOLIs. Formal eligibility assessment and informed consent was undertaken by TVT members. All patients were free to withdraw from the study at any time without giving reasons and without prejudicing any further treatment.

6.4.6 Registration and Randomisation

Screened patients eligible for sub-study participation were registered and randomised on a 2:1 basis to receive either self-complete or interview-administered mode of administration. The 2:1 ratio was used to account for the likelihood of increased missing data from self-completed questionnaires (see Section 6.4.3). Randomisation was stratified by: age ($\leq$ 70, $>$70 years) and PU severity (superficial vs. severe PU).

Those assessed 'not eligible' for sub-study participation but who met criteria for field test 1, were enrolled, and registered to the main study. Following registration, baseline data were collected and PU-QOLIs completed.

6.4.7 Data collection/assessment

Study data (registration, baseline clinical data and PU-QOLI) were recorded on case record forms (CRFs) by members of TVTs, the researcher or by patients who were included in the sub-study and randomised to the self-complete group.

Preliminary PU-QOLI

The preliminary PU-QOLI version (Appendix 5.1) consisted of 13 scales (87 items): pain (11 items); exudate (8 items); odour (6 items); sleep (6 items); malaise (3 items); mobility (11 items); activities of daily living (ADL) (9 items); mood (7 items); anxiety (3 items); self-consciousness and appearance (7 items); autonomy (3 items); isolation (4 items); and participation (9 items). Scales represent unique outcomes of the impact of PUs (key domains important to patients with PUs are presented in the conceptual framework, Chapter 3) and responses are in terms of amount of bother attributed (e.g. "During the past week, how much have you been bothered by…?") on a 4-point response scale (e.g. 0=not at all – 3=a lot). A recall period of the past-week was chosen on clinical grounds, as changes in PU severity and symptomology often occur

over days and thus a longer recall period would risk not capturing relevant impact on HRQL (see Chapter 5 for PU-QOLI development).

Self-completed mode

Patients enrolled to the sub-study and randomised to the self-completed group were provided with the PU-QOLI and instructed to complete the instrument on their own. A TVT member returned to collect completed PU-QOLIs

Interview-administered mode

A questionnaire pack was administered to all participants by TVT members or the researcher, following an interview manual. Training in administering the PU-QOLI was provided by CG.

6.4.8 Data monitoring

Data was monitored for quality and completeness by CG upon receipt to the CTRU. Attempts were made to chase any missing clinical data (e.g. baseline data). However, it was not possible to implement a thorough data chasing process as it was not possible to locate some patients after data had been collected and returned due to data being linked anonymised data. Principal investigators at each participating site were responsible for site monitoring of consent forms for all patients enrolled into the study.

6.4.9 Data entry, transfer and cleaning

Data was entered onto a MACRO database with 100% of data cross-checked. In addition, a random sample of records for five patients from each CRF was checked against the data as seen on the database to ensure that the database was set up correctly and that the data entry had been successful. Data were downloaded from MACRO and read into excel spreadsheets for validation by CG. Data was checked for errors to ensure data transfer had been accurate and set-up to be entered into SPSS. Identified errors were corrected in the data file (e.g. entering 45 when 4 was intended). In SPSS, data were checked for outliers and variables defined (302). No unusual or outlying data were observed. Data were transferred from SPSS to RUMM20/30 software to enable Rasch analyses.

6.5 Analysis Plan for Field Test 1

6.5.1 Baseline characteristics of sample

Baseline analysis took the form of data summaries (no formal statistical testing was conducted). Assessment of patients through the study, including those screened,

assessed as eligible, recruited and data returned was summarised. Population screening data and main study baseline characteristics were performed using SPSS and tabulated using frequencies and summary statistics (see Table 6.5). Percentages were calculated using the total number of patients from the relevant population as the denominator (i.e. including all patients with missing data for that variable) and rounded to 1 decimal place. Means, medians, standard deviations and ranges were also summarised to one decimal place.

6.5.2 Sub-study analyses

Response rate and data quality

The proportion of completed and returned PU-QOLIs (response rate) and percentage of missing data (data quality) per PU-QOLI and per item by administration mode group was calculated.

Differential item functioning (DIF)

A Rasch analysis was performed to examine for item bias or DIF within items (260, 273). DIF provides a method of exploring conditional relationships between item response and group membership by examining the significance of differences observed between different levels (class intervals) of a Person Factor (e.g. administration mode group) (303). Groups to be studied are selected based on theoretical considerations about whether or not the construct studied is hypothesised to have the same conceptual meaning across groups. DIF occurs when people from different groups (e.g. self-completed and interview-administered groups) with the same latent trait (e.g. pain) have a different probability of giving a certain response to an item. A DIF analysis can provide an indication of unexpected behaviour by an item(s) on a test (297); respondents with similar ability/disability should respond in similar ways to individual items regardless of their sex, ethnicity or administration mode. Here, the construct studied was HRQL specific to PUs, measured by disease-specific scales and by subgroup mode of administration. PU-QOLs' HRQL subscales should measure the same unidimensional constructs across specified mode groups.

To ensure that any detected bias was a valid interpretation of group differences dependent on administration mode and not an artefact of differences within groups; differences that could present if for example younger, healthier patients were assigned to the self-complete group and older, more frail patients were assigned to the interview-administered group, only patients who met the inclusion criteria (Section 6.4.4) were included in the sub-study. This ensured that group participants were matched on

clinical presentation and relevant underlying ability before determining whether the two administration mode groups differ in their probability for success (297).

A DIF analysis was undertaken to establish measurement equivalence across the two administration mode groups by investigating the equivalence of responses to PU-QOLI items (i.e. whether responses to items are directly comparable between both groups). A between groups analysis (an analysis of variance for each item, comparing across levels of participant characteristics and levels of latent trait (304)) was performed to indicate any patterns of responses (i.e. differences dependent on administration mode) to determine whether it matters how the PU-QOLI is administered; similar responses between groups would support the development of one version suitable for both administration modes, divergent responses would require two mode-specific versions.

DIF occurs when different groups within a sample (e.g. males and females), despite equal levels of the underlying trait being measured, respond in a different manner to an individual item. For example, given the same level of pain, the expected score on any item within the pain scale should be the same irrespective of gender. There are two types of DIF: uniform DIF is where a consistent systematic difference occurs in a subgroup's responses to an item across the whole range of the trait being measured (same amount of item DIF regardless of person ability/disability level) and non-uniform DIF is where the magnitude of DIF varies according to ability/disability (non-uniformity amongst differences between subgroup, for example, responses vary across levels of the trait).

When uniform DIF is detected, the problem can be remedied by splitting the item(s) by subgroup (e.g. administration mode) and separately calibrating the item for each subgroup (298). This effectively allows the item to be specific to the subgroup in question (e.g. in the case of administration mode, Item # for self-completed version and Item # for administered version). Alternately, items with DIF can be grouped into a subtest to determine whether the DIF identified cancels out at scale level (e.g. an item within a scale may be biased towards self-completed mode but another item biased towards administered mode, then if all persons respond to all items, the apparent DIF would cancel out at scale level and the person ability estimates would not be adversely affected) (298). However, little can be done to correct the problem of non-uniform DIF, and it is often necessary to remove the item from the scale (305).

In RUMM the presence of DIF can be detected both statistically and graphically. Analysis of variance is conducted for each item comparing scores across each level of

the selected person factor (e.g. gender) and across different levels of the underlying trait (the class intervals). Uniform DIF is indicated by a significant main effect for the person factor, while the presence of non-uniform DIF is indicated by a significant interaction effect between the person factor and the class intervals (298).

6.5.3 Item and scale analysis using Rasch

Rasch analysis is increasingly used in the development of rating scales for the health and medical sciences (87, 290). Rasch measurement methods add value in the development of new PRO instruments by allowing the evaluation of scale functioning (i.e. response format, item content, response bias, dimensionality, precision) and the transformation of ordinal level scale scores[7] into linear, interval scale measurement[8] (Rasch analysis allows the estimation of the intervals between ordinal numbers), an attribute important when measuring change over time (85, 87, 290, 292, 308). Traditional methods of item reduction that rely on item-total correlations and/or indices of internal consistency can affect the sensitivity of measures and their ability to produce valid scores at the extremes of the construct range as items at the extreme of the measurement range are generally discarded because too many/few respondents affirm them (290). In reality, 'extreme' items are important in a scale for extending the constructs' range of coverage (85).

A Rasch analysis, using the Andrich Rating Scale Model (276), was performed in RUMM2030 (309) to construct PU-QOLI scales that contained the best possible set of items for measurement of each conceptual domain (scale). PU-QOLI data was tested against model expectations. Any deviations from model expectations were examined to determine whether scale attributes could be improved. Final decisions on item inclusion/exclusion were made according to appraisals of the analyses of the observed data against measurement criteria as described below, and clinical relevance (the extent to which items within proposed scales are clinically cohesive), as opposed to examinations carried out singularly or sequentially.

---

[7] Ordinal measurements describe order (numbers are assigned to objects/events in rank order, for example, from strongly agree to strongly disagree) but not the relative size or degree of difference between measurements (306). Numerical values are assigned to item responses but the difference between, say 'strongly agree' and 'agree' is not necessarily the same as the difference between 'agree' and 'disagree' (the interval between values is not interpretable in ordinal measurement).

[8] In interval measurement, the traits measured have more or less equal intervals, or the distance between traits has meaning (307). For example, when temperature is measured, the distance from 10-20 is the same as the distance from 30-40.

Targeting of persons and items

To assess the ability of the PU-QOLI scales to appropriately target the population being measured, person-item threshold distribution maps were inspected. A well targeted scale should include a set of items that span the full range of person estimates (person locations should be covered by items and item locations covered by persons). A well targeted sample is one in which the person distribution closely matches the item distribution when they are both calibrated on the same metric scale (85). Comparison of the mean location score obtained for the persons with that of the value of zero set for the items (e.g. zero logits representing the item average difficulty for a scale, range -5 to +5), provides an indication of how well-targeted the items are for people in the sample (85). For a well-targeted measure (not too easy, not too hard) the mean location for the persons would also be around the value of zero. A positive mean value for persons would indicate that the sample as a whole was located at a higher level of the trait (e.g. pain) than the average of the scale, while a negative value would suggest the opposite. Arguably, if many patients are at the margins, then the scale is not properly targeted (262).

Threshold ordering - are response categories appropriate?

A common source of item misfit occurs due to respondents' inconsistent use of response options; known as disordered thresholds - the failure of respondents to use the response options in a manner consistent with the level of the trait being measured (278). Disordered thresholds occur when respondents have difficulty consistently discriminating between response options either because there are too many options or the labelling is potentially confusing or open to misinterpretation (e.g. *sometimes, often, frequently* terms used) (305). The expectation is that respondents with high levels of the trait (e.g. pain) being measured would endorse high scoring responses (e.g. a lot of bother), while individuals with low levels of the trait would consistently endorse low scoring responses. This would be indicated by an ordered set of response thresholds for each scale item. The term threshold refers to the point between two response options where either response is equally probable. For example, for each PU-QOLI item, the use of response options scored with successive integer scores (e.g. 0 = ‘no bother’ to 3 = ‘a lot of bother’) implies a continuum of increasing impact (bother), from less (no bother) to more (a lot of bother). This assumption was tested by examining the ordering of thresholds (or points of crossover between two adjacent response options; the point where the probability of scoring a 0 or a 1 on an item is 50/50). Category probability curves were inspected to determine appropriate response category ordering (i.e. whether the ordering of polytomous response options was working as expected).

Overall model fit (summary statistics)

The objective of a Rasch analysis was to test how well the observed data fit the expectations of the measurement model. The Summary Statistics screen in RUMM provides a general view of how the data is behaving. Three summary fit statistics are reported: two are item-person interaction statistics, transformed to approximate a *z* score representing a standardized normal distribution (if the items and persons fit the model then a fit residual mean of approximately zero and a standard deviation of 1 is expected) and the third is an item-trait interaction statistic reported as a chi square; a formal test of invariance to the scale (indicates whether or not data fit the model for discrete groups (class intervals) along the scale or across the trait – in HRQL terminology, traits are often called dimensions, subscales or constructs. A significant chi square ($p = <0.05$) indicates that there is no significant deviation between the observed data and what is expected from the model (298). The 'power' of the tests of fit are presented, indicating the power in detecting the extent to which the data do not fit the model, not that data-fit to the model is good or poor (fit statistics are interpreted in light of the power) (298). The power of tests-of-fit is intimately related to the person separation index (PSI), which is also presented (see *Reliability* below).

Individual model fit analyses

In addition to summary fit statistics, individual item-fit statistics were determined. Item fit tests provide information about how well each item in a scale contributes to defining the construct measured by the scale. Misfit implies that an item is not working as intended in a scale and may be regarded as not measuring the scale's intended construct. As there are no absolute criteria for interpreting fit statistics, to ensure meaningful interpretation, findings were considered together and in the context of their clinical usefulness as an item set.

- *Item fit statistics*

The following indicators were examined to determine the extent to which observed data (patient's responses to items) accord with (fit) the responses expected for groups of responders across the trait (class intervals):

- Fit residuals (item-person interaction) - summation of individual person and item deviations.
- Chi-square statistics (item-trait interaction) – for each item, several chi-squares are computed (dependant on sample size) and summed to give an overall chi-square value for items with degrees of freedom being the number of groups minus 1. If chi-square values are less than 0.05, then the item is deemed to

misfit model expectations. To take account of multiple testing, Bonferroni corrections[9] are applied to adjust the chi squared *p* value (312).

- Item characteristic curves (ICC) (graphical indicator of model fit or misfit) (279) - data are plotted against the expected model curve (ICC). Items with good fit show group plots lying on the curve; those with plots steeper than the curve are considered to be over-discriminating; those flatter than the curve, under-discriminating.

- *Item locations*

The aim of scale items is to mark out the trait or construct as a continuum on which people can be located or measured. Measurement continuum implies that individual scale items are located across a continuum in the same way that the location of individual people are spread out across the continuum (ability/disability) (87). Items with similar locations on the continuum may indicate that one of them may be redundant. The spread of item locations was examined for evidence that items spread out to define a measurement continuum (279, 280).

Tests of local dependence and dimensionality

A problem in local dependence of items can be found by response dependency and multidimensionality. The assumption of local independence implies that once the Rasch factors have been extracted (final scales) no leftover patterns in the residuals should be present. An absence of any meaningful pattern in the residuals is considered support for the assumption of local independence and consequently the unidimensionality of the scale (313). Response dependency is where items are linked in some way, such that the response to one item will determine the response to another. Response dependency was investigated by inspecting the residual correlations (279, 280) for pairs of items with correlations exceeding 0.3.

Reliability

In RUMM, the internal consistency reliability of a scale is measured by a person separation index (PSI). A PSI reliability statistic quantifies how reliably the measurements of patients in the sample are separated (the estimates on the logit scale for each person are used to calculate reliability); it is comparable to Cronbach's alpha (314, 315). Higher values indicate greater reliability; minimum values of 0.7 for group use and 0.85 for individual use are recommended (308).

[9] The Bonferroni adjustment is a method for adjusting the significance levels of individual tests when multiple tests are performed on the same data (the test-wise significance levels are divided by the number of tests) (310, 311). An exact probability value using Bonferroni adjustment is calculated in RUMM20/30.

Like alpha, the PSI is sample dependent because it is computed from the person locations. The PSI focuses on the separation of persons therefore it is a property of the scale in relation to the specific sample of persons measured, indicating the ability of scale items to separate the study sample (85). Thus, the PSI is a function of the data, not an independent function of the scale. In contrast, alpha has the same formulaic structure as the PSI but relates the variance among persons to the variance of the items. Fundamental differences between the PSI and alpha is that the PSI: is expressed entirely in terms of person locations and so meets the true definition of reliability; is computed from linear measurements rather than raw scores; and can be computed when data has missing item responses because in Rasch analysis, missing responses affect the standard error of a person location not the ability to generate an estimate, whereas alpha requires complete data to enable computation (85, 316).

6.5.3 Missing data from Rasch measurement perspective

The RUMM software accounts for missing data through the 'Use class intervals complied on individual item basis' function. This function adjusts the calculation of interval class distributions to be estimated on an item-by-item basis rather than by persons to avoid any skew in distributions (e.g. underrepresentation in specific class intervals due to some persons not responding to every item).

6.5.4 Traditional Psychometric Analyses

The Rasch measurement paradigm views the Rasch model as a formulation that represents the structure which data should exhibit in order to obtain meaningful measurements from data (317). To determine whether the Rasch model fulfilled fundamental prerequisites for rigorous measurement as defined by traditional psychometric criteria and FDA guidance (86), the Rasch developed PU-QOLI scales underwent a preliminary psychometric evaluation using standard psychometric tests (85, 86, 102, 132, 227, 281), examining for: acceptability and data quality, scaling assumptions, targeting, reliability and construct validity against pre-specified criteria (see Table 6.1). Psychometric tests were performed using SPSS 15.0 software.

Table 6.1 Psychometric Tests and Criteria (Adapted from (85, 86, 281, 318)

Psychometric Property	Definition/Test	Criteria (Traditional methods)	Criteria (Rasch methods)
Data quality - Acceptability/Data completeness	The extent to which scale items are scored and total scores can be computed; quality of data is assessed by data completeness and score distributions (computing the percentage of missing data for each item, and the percent of people for whom a scale score is computed (319))	• Item-level missing data <10% (320) • Computable scale scores >50% completed items (100) • A total score can only be computed if all items are scored as they have substantially different ranges. • Items in scales rated as 'not relevant' <35%	• Even distribution of endorsement frequencies across response categories (>80%) • Low number of persons at extreme (i.e. floor/ceiling) ends of the measurement continuum
Scaling assumptions	The extent to which it is legitimate to sum a set of item scores, without weighting or standardisation, to produce a total score. Summing item scores is considered legitimate when items: • are approximately parallel (i.e. they measure at the same point on the scale (redundancy)) • contribute similarly to the variation of the total score (i.e. they have similar variances), otherwise these should be standardised • measure a common underlying construct • contain a similar proportion of information concerning the construct being measured, otherwise items should be given different weights (321)	• Similar mean scores (321) and standard deviations (319) • Items have adequate corrected item-total correlation (ITC ≥0.3) (259) • Items have similar ITCs (259)	• Positive correlations above 0.3 suggest local dependency • High negative correlations (>0.6) suggest redundancy • Items share common variance (unidimensionality) • Determine if items have similar locations on the measurement continuum; evenly spaced items spanning wide range of the continuum
Item Response Categories	The extent to which item response categories work in a logical hierarchy reflecting the measurement continuum within the frame of reference of the scale	• NA	• Ordered item response categories
Targeting	The extent to which the range of the variable measured by the scale matches the range of that variable in the study sample; examination of score distributions at both item and scale level in whole sample and defined disease severity subgroups (evidence of matched scale to sample targeting focussed	• Scale scores should span the entire range • Floor (proportion of sample at maximum scale score) and ceiling (proportion of sample at minimum	• Person-item threshold distribution: the extent to which the range of the variable (scale item locations) match the range of that variable in the study sample (person locations) – well

Psychometric Property	Definition/Test	Criteria (Traditional methods)	Criteria (Rasch methods)
	around the scale best point of measurement)	scale score) effects should be low (<15%) (322) • Skewness statistics should range from -1 to +1 (323) • There is no published criteria for item level targeting, therefore scale level criteria will be used	targeted is demonstrated by the mean location of items and persons around the value zero • Items span full range of person estimates
Reliability			
Internal consistency	The extent to which items comprising a scale measure the same construct (e.g. homogeneity of the scale)	• Cronbach's alphas for summary scores (adequate scale internal consistency is ≥ 0.7 (286) • Item-total correlations ≥ 0.4	• High person separation index >0.7 (324) • Power-of-tests indicate the power in detecting the extent to which the data do not fit the model (fit statistics are interpreted in light of the power) (298) • Items with ordered thresholds
*Test-retest reliability	The stability of a measuring instrument; assessed by administering the instrument to respondents on two different occasions and examining the correlation between test and retest scores	• Scale –level Intra-class correlations (ICCs) coefficient >0.7 (102) between test and retest scores	• Statistical stability across time points (no uniform or non-uniform item DIF (p=>0.05 or Bonferroni adjusted value))
Validity	The extent to which a scale measures what it intends to measure (325)		
Content validity	The extent to which the content of a scale is representative of the conceptual domain it is intended to cover	• Qualitative evidence from patients, expert opinion and literature review that items in the scale are representative of the construct being measured	• Clearly defined construct • Validity comes from careful item construction and consideration of what each item is meant to measure, and then testing against model expectations
Construct validity			
i) Within-scale analyses	Evidence that a single entity (construct) is being measured and that items can be combined to form a summary score; Factor analysis (using principal axis factoring, varimax rotation, with criteria for elimination applied to 2-factor model).	• Cronbach alpha for scale scores >0.7 • Item-total correlations >0.3 • Scaling success	• The extent items work together to represent the construct (item fit statistics: fit residuals within given range +/-2.5; non-significant chi square values; no under– or over-

Psychometric Property	Definition/Test	Criteria (Traditional methods)	Criteria (Rasch methods)
			discriminating item characteristic curves; mean fit residual close to 0.0 and SD approaching 1.0 (usually <1.4) for summary statistics) (308) • Person fit residuals within given range +/-2.5
ii) Between scale analysis – analyses against external criteria			
*Convergent validity	Evidence that the scale is correlated with other measures of the same or similar constructs; assessed on the basis of correlations between the measure and other similar measures	• Correlations are expected to vary according to the degree of similarity between the constructs being measured by each instrument. Specific hypotheses are formulated and predictions tested on the basis of correlations.	• NA
*Discriminant validity	Evidence that the scale is not correlated with measures of different constructs; assessed on the basis of correlations with measures of different constructs (e.g. age, gender)	• Low correlations between scale scores and measures of different constructs	• NA
*Known groups differences	The ability of a scale to differentiate known groups; assessed by comparing scores for subgroups who are expected to differ on the construct being measured (significant differences between known groups or difference of expected magnitude)	• Generate hypotheses and compare changes in mean scores (e.g. predict a stepwise change in PU-QOLI scale scores across PU severity groups, and mean scores would be significantly different)	• Hypothesis testing (e.g. clinical questions are formulated as external and the empirical testing comes from data fit to the Rasch model)
Responsiveness	The ability of a scale to detect clinically significant change following treatment of know efficacy; assessed by examining within person change scores before and after treatment and calculating an effect size statistic (mean change score divided by standard deviation of pre-treatment scores).	• Moderate to large effect sizes (small 0.2, moderate 0.5 or large 0.8 or higher) (326)	• Racking and stacking data for analysis (stablility of scale over time) and DIF (item stability over time) (85)

*Additional tests performed for final field test (evaluation of short PU-QOLI)

NA not assessed

Acceptability and Data quality

Acceptability was determined by data quality; assessed by completeness of item- and scale-level data and score distributions (floor/ceiling effects and skew of scale scores). Data completeness concerns the extent to which scale items are completed in the target sample (percent of missing data for items) and the percent of people for whom it is possible to compute scale scores (319). The criterion for acceptable item-level missing data was <10% (320) and for computable scale scores >50% (100). The criterion for maximum endorsement frequencies is <80% (floor/ceiling effects <80%).

Scaling Assumptions

Tests of scaling assumptions examine whether it is appropriate to sum a group of items to generate a scale score. Scaling assumptions are satisfied with similar item means and variances, and when items have adequate corrected item-total correlations (ITC; criterion for corrected ITC was ≥0.3) (259, 319, 321).

Targeting

Targeting assesses the match between the range of each variable (or trait) as measured by each PU-QOLI scale and the range of PU severity in the sample. Scale-to-sample targeting was determined by investigating whether: scale scores span the entire scale range; floor (proportion of the sample at the maximum scale range) and ceiling (proportion of the sample at the minimum scale range) effects are low (<15%) (322); and skewness statistics range from -1 to +1 (323).

Reliability

Reliability refers to the consistency of a measure yielding the same score at each administration, assuming all things being equal (i.e. true change has not occurred in the variable being measured), and the extent to which scale scores are free from random error. A way of estimating reliability is to determine the consistency of results across items on the same measure (i.e. compare scale items that measure the same construct to determine a scales internal consistency).

- *Internal consistency reliability*

Internal consistency, reported as Cronbach's alpha coefficients (314), was determined for all PU-QOLI scales. Alpha provides an indication of the degree of convergence between items hypothesised to represent the same variable. Adequate scale internal consistency is indicated by Cronbach's alpha coefficients ≥0.8 (327), however, internal consistency estimates of ≥0.7 are considered acceptable for group comparisons (286).

- *Item total correlation*

The relationship between an item and the total scale score was assessed using a Pearson correlation coefficient, expressed as a number between -1.0 to +1.0. A negative correlation value indicates that an item is actually lowering an individual's score rather than raising it. Item-total correlations (ITC) in the range of +0.4 to around +0.6 indicate items are moderately correlated with scale scores and higher values indicate well correlated items with scale scores (286).

Validity

Validity refers to the extent to which a scale measures what it intends to measure. Although a scale may be reliable, it may be consistently measuring the wrong thing. For example, demonstrating that a set of items intended to measure pain has good reliability merely indicates that the items are getting at the same true score, but not necessarily tapping into the pain true score (287). Evaluating the validity of a measure involves accumulating evidence from different forms to indicate the degree to which the measure denotes what it is intended to represent.

- *Content*

Content validity is the extent to which a measure samples a representative range of content (items) for the construct measured. Consideration of item sufficiency and the target population is essential and ideally should include systematic comparison with existing standards, well-accepted theoretical definitions, expert opinions and interviews with individuals for whom the measure is targeted (86). These methods are described in detail in Chapters 3 (Development of a conceptual framework) and 5 (PU-QOLI development and pretesting).

- *Construct*

Construct validity evidence indicates the degree to which a measure represents what it is intended to represent. A within-scale construct validity analysis was undertaken to determine the extent to which PU-QOLI scales measure a single entity (are PU-QOLI scales distinct constructs?) and therefore whether items can be combined to form scale scores; assessed on the basis of ITC >0.3.

6.5.5 Missing data from the traditional psychometric perspective

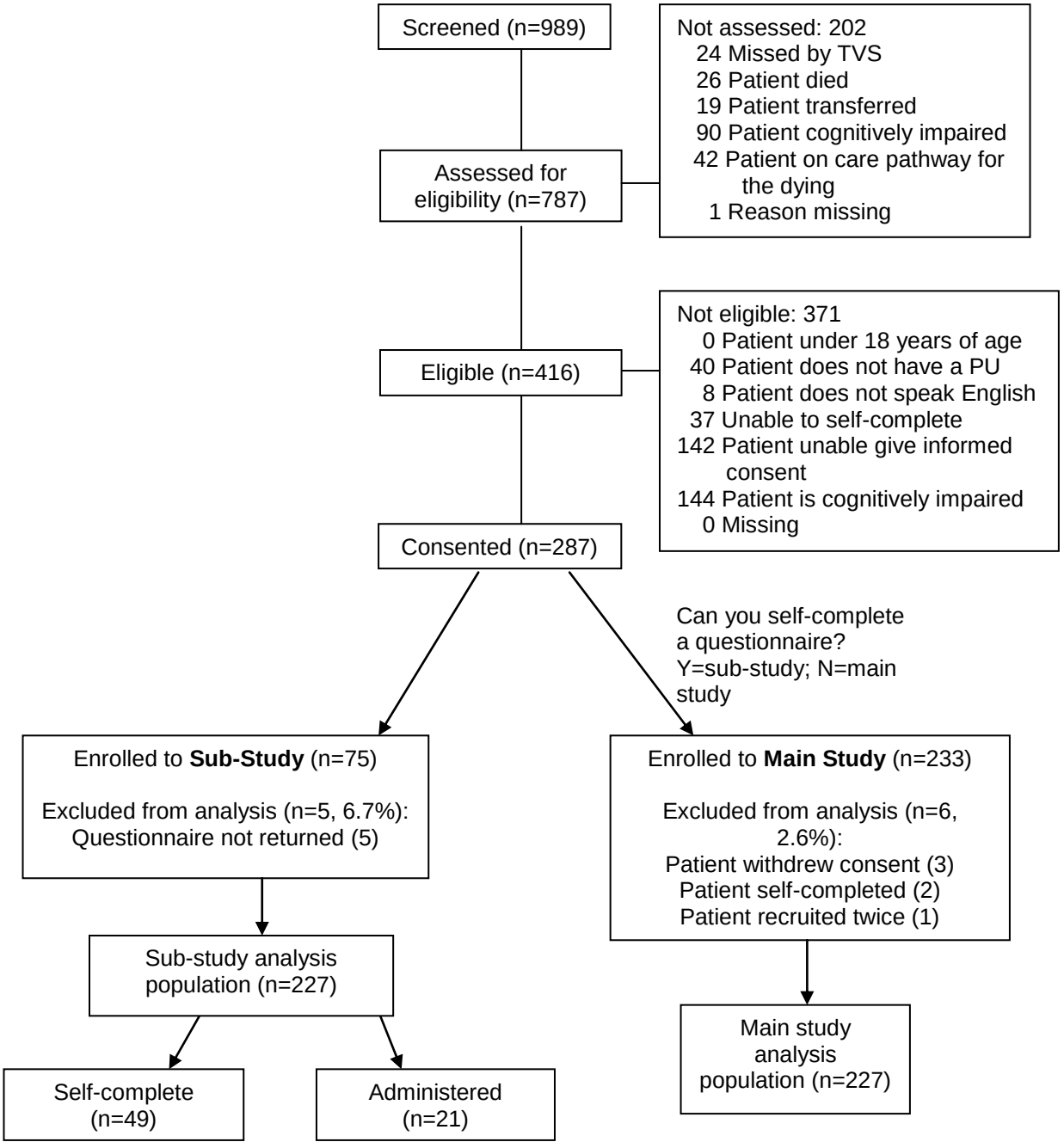
For the traditional psychometric evaluation, missing data was not imputed. The frequency of missing data was determined and items with a response rate of <90% were investigated.

6.6 Results

6.6.1 Sample

A total of 989 patients were screened for study participation from 21 hospitals, 10 community services and one hospice. Of those screened, eligibility was assessed for 787 (79.6%); 416 were considered eligible (52.9%); and of those eligible, 287 (69.0%) consented to participate (Figure 6.1).

Figure 6.1 Assessment flow chart for field test 1 and sub-study



Cognitive impairment was the main reason for ineligibility (38.8%). Those able to self-complete were included in the sub-study (n=75) while those unable to self-complete were enrolled onto the main study.

6.6.2 Sub-Study results: Response rates and data quality

Respondents (n=75) in the sub-study sample represented a wide range of age groups, there was a higher proportion of men, just over half of the sample were hospitalized, and over half of PUs were superficial. Table 6.2 presents the sub-study sample characteristics. All 75 patients completed PU-QOLIs indicating a 100% response rate; no difference in response rate was observed by administration mode group. Table 6.3 indicates the percentage of missing data by sub-groups: mode (self-complete and administered), age (under 70 years and 70 years or over) and healthcare setting (hospital and community). For the administered group (n=21), the possible range of missed items was 0-1827 (i.e. 87 items per PU-QOLI x 21 administrations = 1827 total items); a total of three PU-QOLIs were returned with 29 items missed (1.6%). For the self-completed group (n=49), the possible range of missed items was 0-4263; 19 PU-QOLIs were returned with 619 missed items (14.5%).

Table 6.2 Patient characteristics (sub-study population)

	Self-completed (n=49)	**Administered (n=21)**	**Total (n=70)**
Patient age (years)			
Mean (SD)	65 (15)	62 (16)	64 (15)
Median (range)	68 (21-85)	65 (27-93)	66 (21-93)
Under 70 years of age	25	14	39
70 years or older	24	7	31
Missing	0	0	0
Gender			
Male	33 (67.3%)	14 (66.7%)	47 (67.1%
Female	16 (32.7%)	7 (33.3%)	23 (32.9%)
Type of healthcare setting			
Acute	26 (53.0%)	12 (57.1%)	38 (54.3%)
Community	23 (47.0%)	9 (42.9%)	32 (45.7%)
Pressure ulcer severity			
Superficial grades 1/2	28 (57.1%)	12 (57.1%)	40 (57.1%)
Severe grades 3/4	21 (42.9%)	9 (42.9%)	30 (42.9%)

Participants under 70 years of age who self-completed returned 12 PU-QOLIs with 336 items missed (15.5%). Those 70 years or older who self-completed returned seven PU-QOLIs with 283 items missed (13.6%). Of the administered group, two PU-QOLIs

had three items missed from those under 70 years (0.3%) and one PU-QOLI with 26 items missed from those 70 years or older (4.3%).

Participants hospitalised who self-completed returned 16 PU-QOLIs with 604 items missed (26.7%). Those living in the community who self-completed returned three PU-QOLIs with 15 items missed (0.8%). Of those hospitalised who were administered PU-QOLIs, two were returned with 28 items missed (2.7%). Only one PU-QOLI with one item missed was returned from those living in the community who had PU-QOLIs administered (0.1%) (Table 6.3). Overall, a larger proportion of self-completed PU-QOLIs were returned with missing data compared to those who had them administered. No difference in missing data was observed by age. A difference was observed by healthcare setting; hospitalised patients that self-completed returned PU-QOLIs with the largest amount of missing data.

Table 6.3 Data Quality – Missing data (sub-study population)

	Self-completed (n=49)	**Administered (n=21)**	**Total (n=70)**
PU-QOLIs with missing data	**19** (38.8%)	**3** (14.3%)	**22** (31.4%)
Total number of PU-QOLI items missed (range 1-87 items per PU-QOLI)	619 (14.5%)	29 (1.6%)	648 (10.6%
Age			
Number Under 70 years	(n=12/25)	(n=2/14)	(n=14/39)
Number items missed	336 (15.5%)	3 (0.3%)	345 (10.2%)
Number 70 years or older	(n=7/24)	(n=1/7)	(n=8/31)
Number items missed	283 (13.6%)	26 (4.3%)	309 (11.5%)
Type of healthcare setting			
Number Acute	(n=16/26)	(n=2/12)	(n=18/38)
Number items missed	604 (26.7%)	28 (2.7%)	632 (19.1%)
Number Community	(n=3/23)	(n=1/9)	(n=4/32)
Number items missed	15 (0.8%)	1 (0.1%)	16 (0.6%)

Qualitative observations

PU-QOLIs returned with missed items were examined to investigate patterns in missing responses. The following observations were noted. Of the 19 self-completed PU-QOLIs with missing data, four respondents wrote ‘n/a’ next to items missed, suggesting that the response option ‘My pressure ulcer did not give me this problem’ was not used as intended. Six respondents completed only one item per scale; five respondents missed items at random; two missed a page; one missed items from only the daily

activities scale; and one mostly missed items at the beginning. For the three administered PU-QOLIs with missing data, one had one item missed; one had two items missed; and one hospital patient requested to cease completion due to feeling ill, resulting in a large amount of items (n=26) missed towards the end. No obvious patterns in responses emerged.

6.6.3 Differential item functioning

A summary of the DIF results are presented in Table 6.4. In the context of the mode subgroups, DIF was demonstrated in 9/13 scales (seven scales with one item and two scales with 2 or 3 items); however, the DIF observed for most items was marginal. This finding provides preliminary evidence for stable item performance across mode subgroups; suggesting PU-QOLI subscales could be measured on a common metric (i.e. administration mode did not impact on the way patients responded to PU-QOLI items), as supported by equivalence of self-completed and interview-administered versions. Statistically there were no items with significant DIF by mode for any PU-QOLI scales at the 1% confidence level. Due to the small sample size, it is unlikely that we would find significant DIF below 0.001 Bonferroni adjusted significance level, therefore the 95% confidence level was observed.

As DIF is a product of the sample and not the scale (e.g. chi-squared values and probabilities are sample size dependent), additional exploration of DIF was undertaken with two hypothetical samples (n=200 and n=300) adjusted in RUMM from the original analysis sample (n=70). Amending the sample size provides a better feel for the behaviour of the data and gives the DIF statistics a chance to reveal any DIF (the larger the sample size, the greater the values and the greater the apparent DIF (85)). This estimate was only a guide to give a feel for the behaviour of the data and did not affect other aspects of the analysis. In both adjusted samples, 11/13 scales emerged with items with significant DIF. The anxiety and autonomy scales remained without new items with significant DIF while all other scales had a significant proportion of items with both uniform and non-uniform DIF (Table 6.4).

6.6.4 Main Study respondent characteristics

Respondents in the main study sample represented a wide range of age groups. There was a slightly higher proportion of women, and a small percentage of respondents from minority ethnic groups. Just over half of the sample was hospitalized patients and at on-going long-term risk of PUs. Over two thirds of the sample was

married or cohabiting and just under half the sample had some educational qualification. Table 6.5 presents patient characteristics for the main study.

Table 6.4 Summary of DIF by administration mode for each PU-QOLI subscale

	Uniform		Non-uniform	
Scale (sample size)	Item number (p < 0.001)	Item number (p < 0.05)	Item number (p < 0.001)	Item number (p < 0.05)
Pain				
(n=70)	x	x	x	10 (0.043)
(Adjusted n=200)	10 (0.001)	2 (0.008) 5 (0.002) 8 (0.005) 10 (0.001)	10 (0.000)	10 (0.000)
(Adjusted n=300)	2 (0.000) 5 (0.000) 8 (0.001) 10 (0.000)	1 (0.033) 2 (0.001) 5 (0.000) 8 (0.001) 10 (0.000) 11 (0.019)	10 (0.000)	2 (0.026) 5 (0.031) 8 (0.019) 10 (0.000)
Exudate				
(n=70)	x	x	x	x
(Adjusted n=200)	3 (0.000) 5 (0.000)	2 (0.048) 3 (0.000) 5 (0.000)	2 (0.001)	2 (0.001) 8 (0.037)
(Adjusted n=300)	3 (0.000) 5 (0.000)	1 (0.029) 2 (0.015) 3 (0.000) 5 (0.000)	2 (0.000)	2 (0.000) 8 (0.010)
Odour				
(n=70)	x	3 (0.006)	x	x
(Adjusted n=200)	2 (0.000) 3 (0.000) 6 (0.000)	2 (0.000) 3 (0.000) 6 (0.000)	1 (0.000) 3 (0.000) 6 (0.000)	1 (0.000) 3 (0.000) 6 (0.000)
(Adjusted n=300)	2 (0.000) 3 (0.000) 6 (0.000)	2 (0.000) 3 (0.000) 4 (0.020) 6 (0.000)	1 (0.000) 3 (0.000) 6 (0.000)	1 (0.000) 3 (0.000) 5 (0.021) 6 (0.000)
Sleep				
(n=70)	x	x	x	5 (0.026)
(Adjusted n=200)	5 (0.000) 6 (0.001)	4 (0.006) 5 (0.000) 6 (0.001)	1 (0.002) 5 (0.000) 6 (0.000)	1 (0.002) 5 (0.000) 6 (0.000)
(Adjusted n=300)	4 (0.001) 5 (0.000) 6 (0.000)	1 (0.041) 4 (0.001) 5 (0.000) 6 (0.000)	1 (0.000) 5 (0.000) 6 (0.000)	1 (0.000) 5 (0.000) 6 (0.000)
Malaise				
(n=70)	x	2 (0.026)	x	x
(Adjusted n=200)	2 (0.000) 3 (0.000)	1 (0.031) 2 (0.000) 3 (0.000)	3 (0.003)	3 (0.003)
(Adjusted n=300)	2 (0.000) 3 (0.000)	1 (0.008) 2 (0.000) 3 (0.000)	3 (0.000)	1 (0.021) 3 (0.000)
Mobility				
(n=70)	x	7 (0.038)	x	1 (0.024) 4 (0.023) 7 (0.042)
(Adjusted n=200)	5 (0.000)	5 (0.000)	1 (0.000)	1 (0.000)

	Uniform		Non-uniform	
Scale (sample size)	Item number (p < 0.001)	Item number (p < 0.05)	Item number (p < 0.001)	Item number (p < 0.05)
	7 (0.000)	6 (0.048) 7 (0.000)	2 (0.000) 3 (0.000) 4 (0.000) 6 (0.000) 7 (0.000) 11 (0.000)	2 (0.000) 3 (0.000) 4 (0.000) 6 (0.000) 7 (0.000) 11 (0.000)
(Adjusted n=300)	5 (0.000) 7 (0.000)	4 (0.050) 5 (0.000) 6 (0.015) 7 (0.000)	1 (0.000) 2 (0.000) 3 (0.000) 4 (0.000) 6 (0.000) 7 (0.000) 11 (0.000)	1 (0.000) 2 (0.000) 3 (0.000) 4 (0.000) 6 (0.000) 7 (0.000) 9 (0.031) 10 (0.022) 11 (0.000)
Daily Activities				
(n=70)	x	2 (0.024)	x	2 (0.018)
(Adjusted n=200)	2 (0.000) 4 (0.001)	2 (0.000) 4 (0.001) 7 (0.008)	1 (0.000) 2 (0.000)	1 (0.000) 2 (0.000) 7 (0.004) 8 (0.026) 9 (0.005)
(Adjusted n=300)	2 (0.000) 4 (0.000) 7 (0.001)	2 (0.000) 4 (0.000) 6 (0.017) 7 (0.001)	1 (0.000) 2 (0.000) 7 (0.000) 9 (0.001)	1 (0.000) 2 (0.000) 3 (0.027) 7 (0.000) 8 (0.006) 9 (0.000)
Mood				
(n=70)	x	x	x	4 (0.026)
(Adjusted n=200)	x	2 (0.047) 4 (0.017)	4 (0.000) 6 (0.000)	2 (0.48) 4 (0.000) 6 (0.000) 7 (0.005)
(Adjusted n=300)	x	2 (0.014) 4 (0.003) 5 (0.035) 6 (0.049)	4 (0.000) 6 (0.000) 7 (0.001)	2 (0.015) 4 (0.000) 5 (0.028) 6 (0.000) 7 (0.001)
Anxiety				
(n=70)	x	x	x	x
(Adjusted n=200)	x	x	x	x
(Adjusted n=300)	x	x	x	x
Self-consciousness				
(n=70)	x	x	x	x
(Adjusted n=200)	5 (0.000)	3 (0.012) 4 (0.004) 5 (0.000) 6 (0.025)	5 (0.000)	4 (0.037) 5 (0.000)
(Adjusted n=300)	3 (0.002) 4 (0.000) 5 (0.000)	3 (0.002) 4 (0.001) 5 (0.000) 6 (0.006)	5 (0.000)	3 (0.037) 4 (0.010) 5 (0.000)
Autonomy				
(n=70)	x	x	x	x
(Adjusted n=200)	x	x	1 (0.001) 2 (0.000)	1 (0.001) 2 (0.000)
(Adjusted n=300)	x	x	1 (0.000)	1 (0.000)

	Uniform		Non-uniform	
Scale (sample size)	Item number (p < 0.001)	Item number (p < 0.05)	Item number (p < 0.001)	Item number (p < 0.05)
			2 (0.000)	2 (0.000)
Isolation				
(n=70)	x	x	x	4 (0.042)
(Adjusted n=200)	4 (0.000)	1 (0.012)	3 (0.000)	2 (0.015)
		4 (0.000)	4 (0.000)	3 (0.000)
				4 (0.000)
(Adjusted n=300)	1 (0.000)	1 (0.002)	2 (0.003)	2 (0.003)
	4 (0.002)	3 (0.036)	3 (0.000)	3 (0.000)
		4 (0.000)	4 (0.000)	4 (0.000)
Participation				
(n=70)	x	x	x	3 (0.005)
				8 (0.014)
(Adjusted n=200)	1 (0.001)	1 (0.000)	1 (0.000)	1 (0.000)
	3 (0.001)	2 (0.001)	3 (0.000)	3 (0.000)
	4 (0.000)	3 (0.000)	5 (0.000)	5 (0.000)
	7 (0.001)	4 (0.000)	6 (0.000)	6 (0.000)
		5 (0.015)	7 (0.000)	7 (0.000)
		6 (0.003)	8 (0.000)	8 (0.000)
		7 (0.000)	9 (0.001)	9 (0.001)
		9 (0.000)		
(Adjusted n=300)	1 (0.000)	1 (0.000)	1 (0.000)	1 (0.000)
	2 (0.001)	2 (0.001)	3 (0.000)	3 (0.000)
	3 (0.000)	3 (0.000)	5 (0.000)	5 (0.000)
	4 (0.000)	4 (0.000)	6 (0.000)	6 (0.000)
	7 (0.000)	5 (0.015)	7 (0.000)	7 (0.000)
	9 (0.000)	6 (0.003)	8 (0.000)	8 (0.000)
		7 (0.000)	9 (0.000)	9 (0.000)
		9 (0.000)		

x indicates no items detected with significant DIF within the scale

6.6.5 Rasch analyses of Pain scale and any item reduction

Rasch analyses of PU-QOLI scale data were undertaken for all scales independently. The results and interpretations for each psychometric property described in the methods Section 6.5.3 are reported in detail for the pain scale. For the remaining scales, decisions made to any scale modifications and a summary of findings are presented in Tables 6.11 and 6.12, and results are interpreted and discussed interactively in Sections 6.6.9 to 6.6.26.

It is important to mention that two Rasch analyses were performed. The first analysis included a data set that combined response categories ‘not because of PUs’ and ‘no bother’, both scored as 0. However, merging these two responses was considered an inappropriate method of analysis as ‘not because of PUs’ category was essentially descriptive data about patient comorbidity and not part of the scale data. Therefore, analysis two was performed using a data set where the ‘not because of PUs’ response was treated as missing.

Table 6.5 Respondent Characteristics (n=227)

Characteristics	***Range (Mean, SD)***
Age	*24 to 98 years (72, 13.5)*
	Total n (%)
Gender	
Male	90 (39.6)
Female	137 (60.4)
Ethnicity	
White	223 (98.2)
Asian	1 (0.4)
Black/African	2 (0.4)
Chinese	0
Not stated	1 (0.4)
Setting	
Hospital (surgery)	99 (43.6)
Hospital (medicine)	21 (9.3)
Community	107 (47.1)
PU severity	
Category 1	38 (10.6%)
Category 2	144 (40.2%)
Category 3/4	175 (48.9%)
Missing	1 (0.3%)
PU risk classification	
Short-term	39 (17.2)
New medium to long-term	71 (31.3)
On-going long-term	116 (51.1)
Missing	1 (0.4)
Marital status	
Single (includes divorced, separated, widowed)	59 (26.0)
Married	85 (37.5)
Cohabiting	81 (35.7)
Missing	2 (0.8)
Living arrangements	
Live alone	84 (37.0)
Cohabit with identified carer	63 (27.8)
Cohabit with other	61 (26.9)
Missing	19 (8.4)
Education	
No formal education	129 (56.8)
GCSE or equivalent	39 (17.2)
A-Level or equivalent	25 (11.0)
Degree or higher	15 (6.6)
Missing	19 (8.4)

Prior to formal Rasch analysis on PU-QOLI scales, both analyses were performed on a random selection of five scales and findings compared for consistency. Treating 'not because of PUs' as missing did not significantly change the Rasch or Classical test results; only marginal differences were observed, thus supporting the use of analysis two for the remaining scales. For analysis purposes, 'not because of PUs' was treated as missing for both Rasch and traditional analyses, except for computable scale scores (classical tests); treating 'not because of PUs' as not missing enabled data completeness evaluation.

This methodological evaluation resulted in a distinction being made between descriptive data and scale data. Descriptive data equates to data collected for those patients that have the trait (construct property as measured by a particular item) but not because of PUs and helps formulate known groups hypotheses. Scale data is essentially scale items scored as 0, 1 and 2.

6.6.5.1 Scale-to-sample targeting

Is the pain scale-to-sample targeting adequate for making judgements about the performance of the scale and the measurement of people?

Figure 6.2 illustrates the targeting of the patient sample to the location's 11 items in the pain scale. Scale range set from -4 to +4 units (logits) for symmetry (mean always set at 0). The histogram bars represent the relative location of the item(s) and persons on the same variable (pain). The curve represents where on the continuum the scale performs best. Item locations range from -1.0 to +0.8 (bottom histogram) and person locations range from -4.0 to +3.2 (top histogram) (Figure 6.2).

Item locations are covered by the people, but the person location are not covered by the items; reasonable sample to examine the scale but a suboptimal scale for measuring the sample. There are no items representing people at the extremes of the sample distribution. The scale performs best in the centre (range -1 to +1 logits) and worse at the extremes; about 72 (34.1%) of people in the sample are located outside the best functioning of the scale. Figure 6.3 illustrates the targeting of the patient sample to the thresholds of the 11 items of the pain scale. Item locations now spread from about -2.8 to +1.4 logits but still do not cover the range of person locations in the sample.

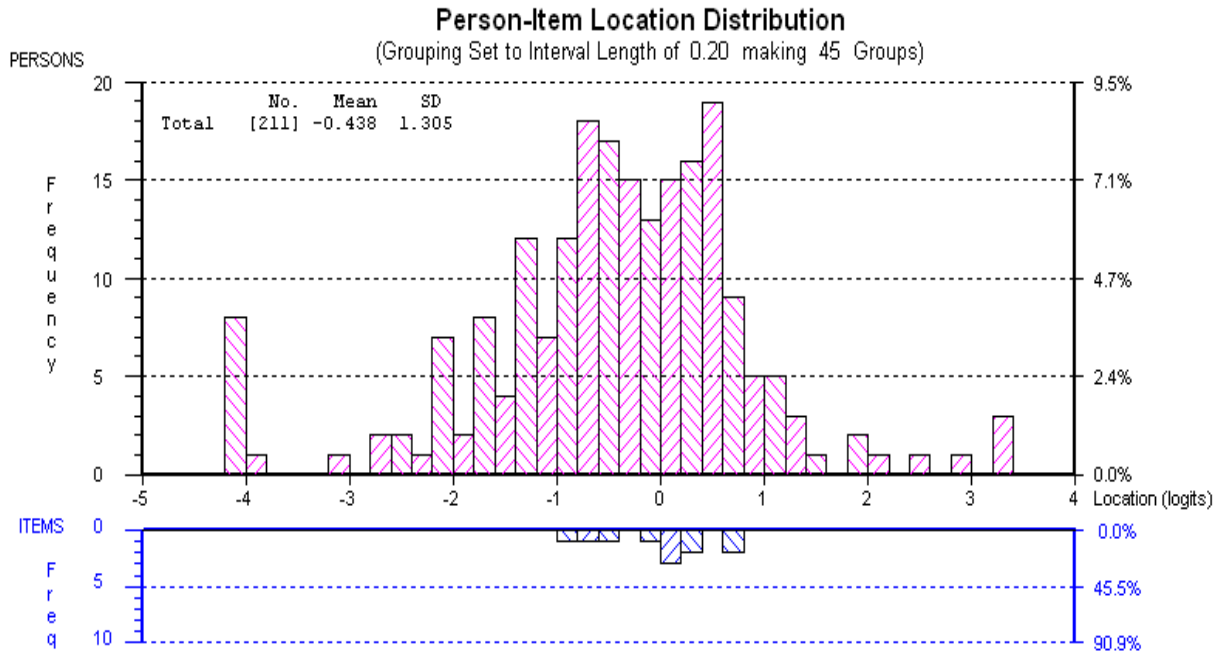


Figure 6.2 Pain scale – targeting of sample to item locations. Person-item location distribution (grouping set to interval length of 0.2, making 45 groups).

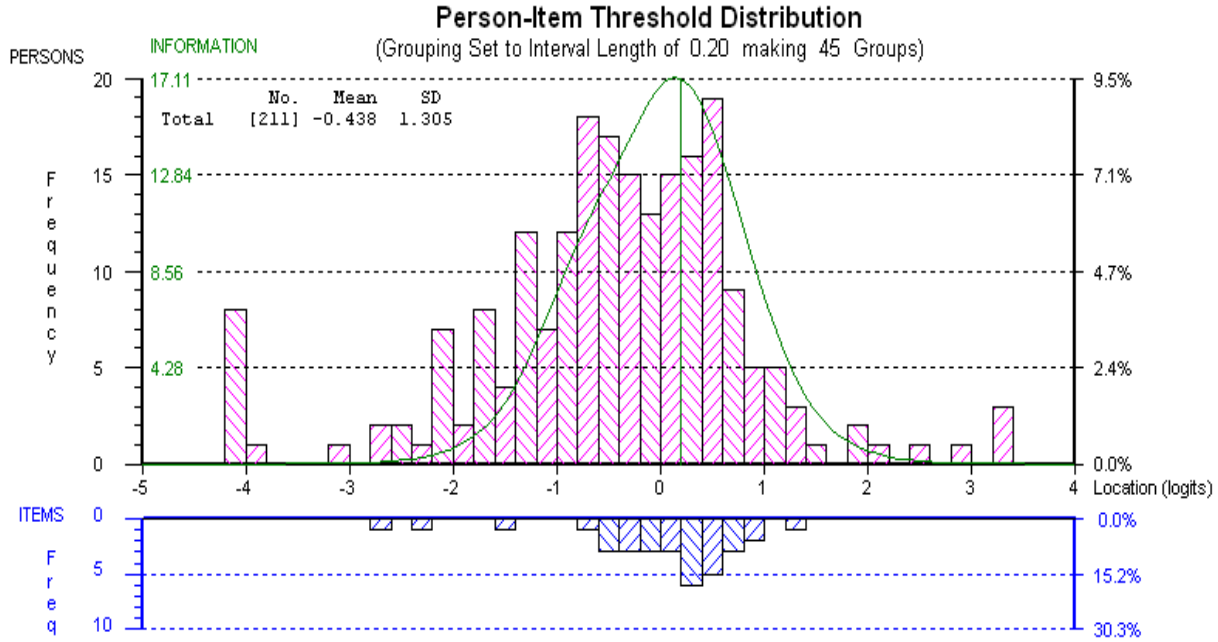
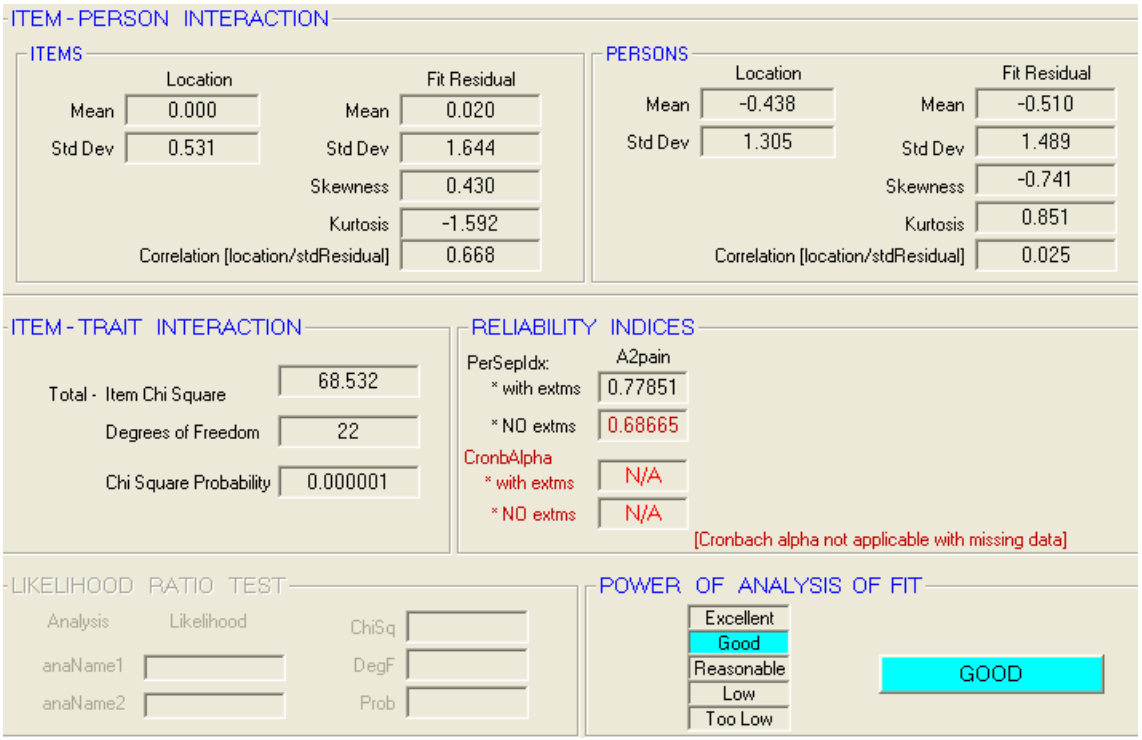


Figure 6.3 Pain scale – targeting of sample to item thresholds. Person-item threshold distribution (grouping set to interval length of 0.2, making 45 groups).

The 'power' of the tests of fit for the pain scale are good (Table 6.6); this indicates that there is good power in detecting the extent to which the data do not fit the model, not that data-fit to the model is good (fit statistics are interpreted in light of the power)

(298). The power of tests-of-fit is intimately related to the person separation index (PSI) (see Section 6.5.3 *Reliability*). If the PSI is low (i.e. persons have similar locations and not spread across the continuum), the power of the tests of fit is low. This is because a lack of variability in person locations makes it impossible to determine whether people with higher locations tend to get higher scores on items (e.g. unable to accurately estimate person ability). The pain scale chi-square probability is 0.000, indicating some items may misfit from model expectations.

Table 6.6 Pain scale summary statistics



6.6.5.2 Thresholds

Do the item response categories work as intended?

Each item instrument has four response categories, ordered to imply a continuum of increasing impact from less ('no bother at all') to more ('a lot of bother'). This continuum of increasing impact is further implied by assigning sequential integers to the response categories (0=no bother to 3= a lot of bother). It is assumed that the response categories work as intended (ordered sequentially) but this needs to be checked empirically. With four response categories, there are three thresholds (the Greek symbol 'τ' (tau) is used to signify threshold):

τ1 – where the probability of scoring either 0 ('no bother') and 1 ('a little bother') is the same

τ2 – where the probability of scoring either 1 (‘a little bother’) and 2 (‘quite a bit of bother) is the same

τ3 - where the probability of scoring either 2 (‘quite a bit of bother) and 3 (‘a lot of bother) is the same

Disordered thresholds, the finding that threshold locations are not ordered sequentially, implies that the item scoring functions are not working as intended (276). There are three main reasons for disordered thresholds: responders cannot use the response categories consistently (impacts item reliability); categories do not characterise the intended meaning of what it takes to reflect more of the property within an item (impacts item validity); and the item does not measure the same underlying trait as the other items in the scale (e.g. pain) (85).

Table 6.7 indicates 3/11 pain scale items have disordered thresholds: items 4 (‘itchiness’); 9 (‘stabbing’); 11 (‘burning’). For example, for item 4, the estimate for τ3 (-0.074) is less than that for τ2 (0.525), meaning that the estimated point on the continuum at which the probability is the same for scoring either 2 (‘quite a bit of bother’) or 3 (‘a lot of bother’) is lower than the point on the continuum at which the probability of scoring either 1 (‘a little bother’) or 2 (‘quite a bit of bother’) is the same; this does not make sense and indicates the response categories are not working as intended.

Table 6.7 Pain scale: item thresholds and location estimates

	Item	**Location**	**τ1**	**τ2**	**τ3**
1	Uncomfortable	-0.828	-1.812	0.448	1.364
2	Dull ache	0.183	-0.474	0.095	0.378
3	Annoying	-0.562	-0.946	-0.041	0.987
4	*Itchiness	0.764	-0.451	0.525	-0.074
5	Tender	-0.793	-1.532	0.528	1.004
6	Tingling	0.625	-0.383	0.165	0.219
7	Throbbing	0.210	-0.712	-0.019	0.732
8	Stinging	0.058	-0.570	-0.089	0.659
9	*Stabbing	0.349	0.117	-0.342	0.225
10	Red raw	-0.882	-0.388	-0.091	0.480
11	*Burning	0.084	-0.283	0.199	0.084

* Indicates disordered threshold

τ signifies threshold

Figure 6.4 illustrates the region of the continuum represented by each item response option for items with ordered thresholds. This figure demonstrates that the area on the continuum represented by response options 1 (a little bother) and 2 (quite a bit of bother) is often very small. For the other items, the response categories are working as intended.

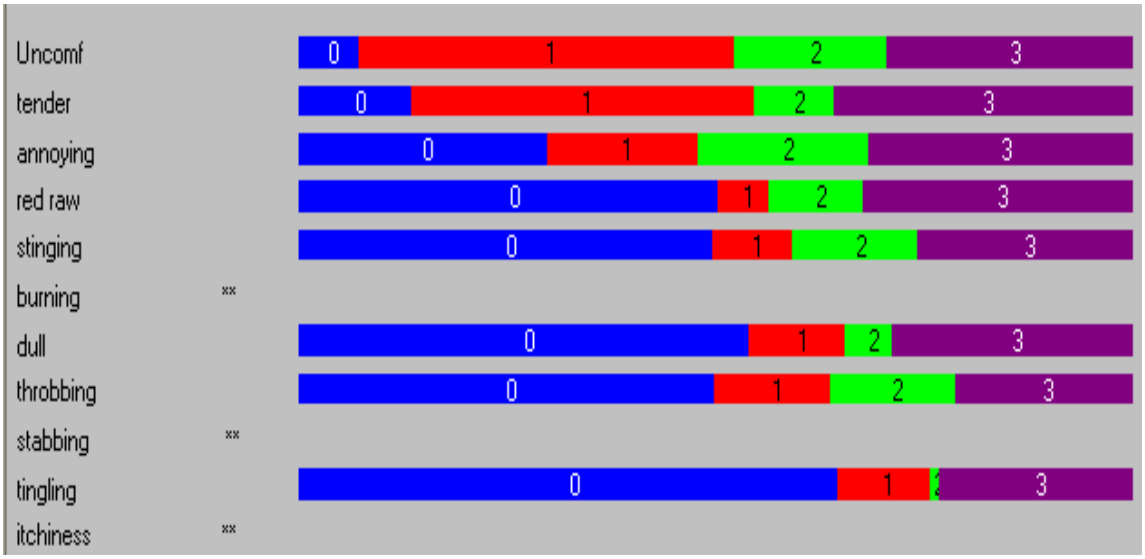


Figure 6.4 Pain scale – threshold maps. xxReversed thresholds. Key: 0, response category labelled 0; 1, response category labelled 1; 2, response category labelled 2; 3, response category labelled 3

Inspecting category probability curves (CPC) indicates that items 1, 2, 3, 5, 6, 7, 8 and 10 (Figure 6.5), despite ordered thresholds, thresholds (τ1 and τ2) are very close to being disordered. Possible explanations for this finding may be that there are too many response categories or the wording attached to response categories is difficult for people to relate to in practice. CPC plot the probability of a response (y-axis) against the person's location on the pain bother continuum mapped out by 11 pain scale items. High scores indicate more bother. Thus, as one moves from left to right on the x-axis, people have more pain bother. The expectation is that as a person's pain bother increases, the probability of that person responding 'no bother' falls and the probability of responding 'a little bother' increases. Figure 6.6 illustrates the CPC for items with reversed thresholds. For example, the graph for item 4 illustrates that there is no point on the continuum at which categories 2 and 3 have the highest probability of being chosen. Thus the intersection of response category 1 and 2 (τ2) is above the intersection of response category 2 and 3 (τ3). Category 2 is problematic in that it appears that people are having difficulty distinguishing between category 1 and 2.

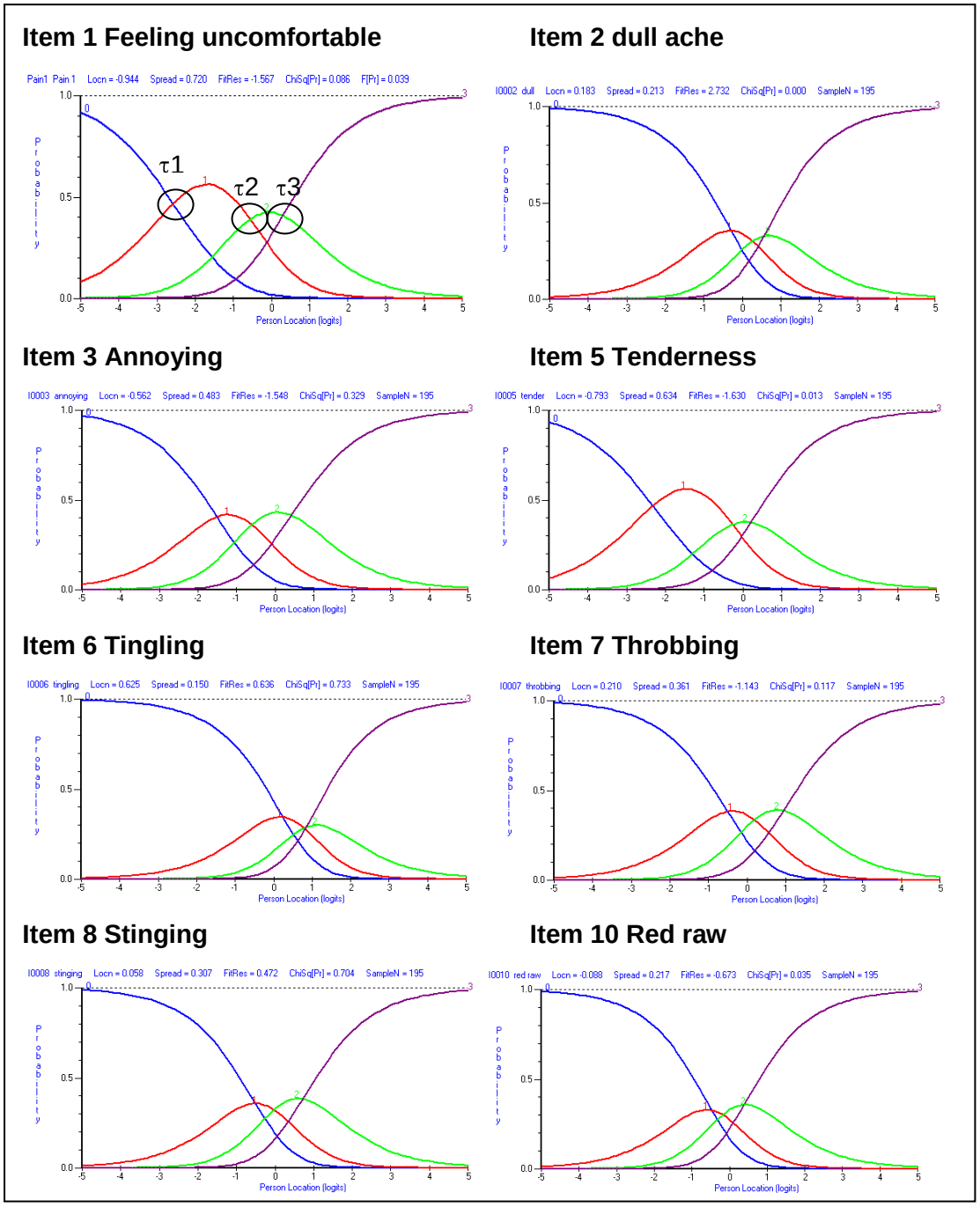
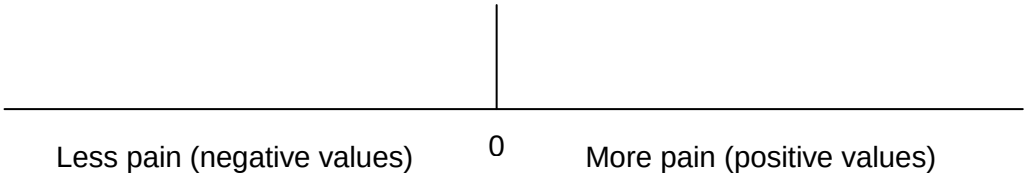


Figure 6.5 Items with ordered thresholds

6.6.5.3 Measurement continuum

Do the items map out a discernable line of increasing intensity?

Item locations, their range, how they are spread, their proximity to each other and the precision of these estimates' standard error were examined.



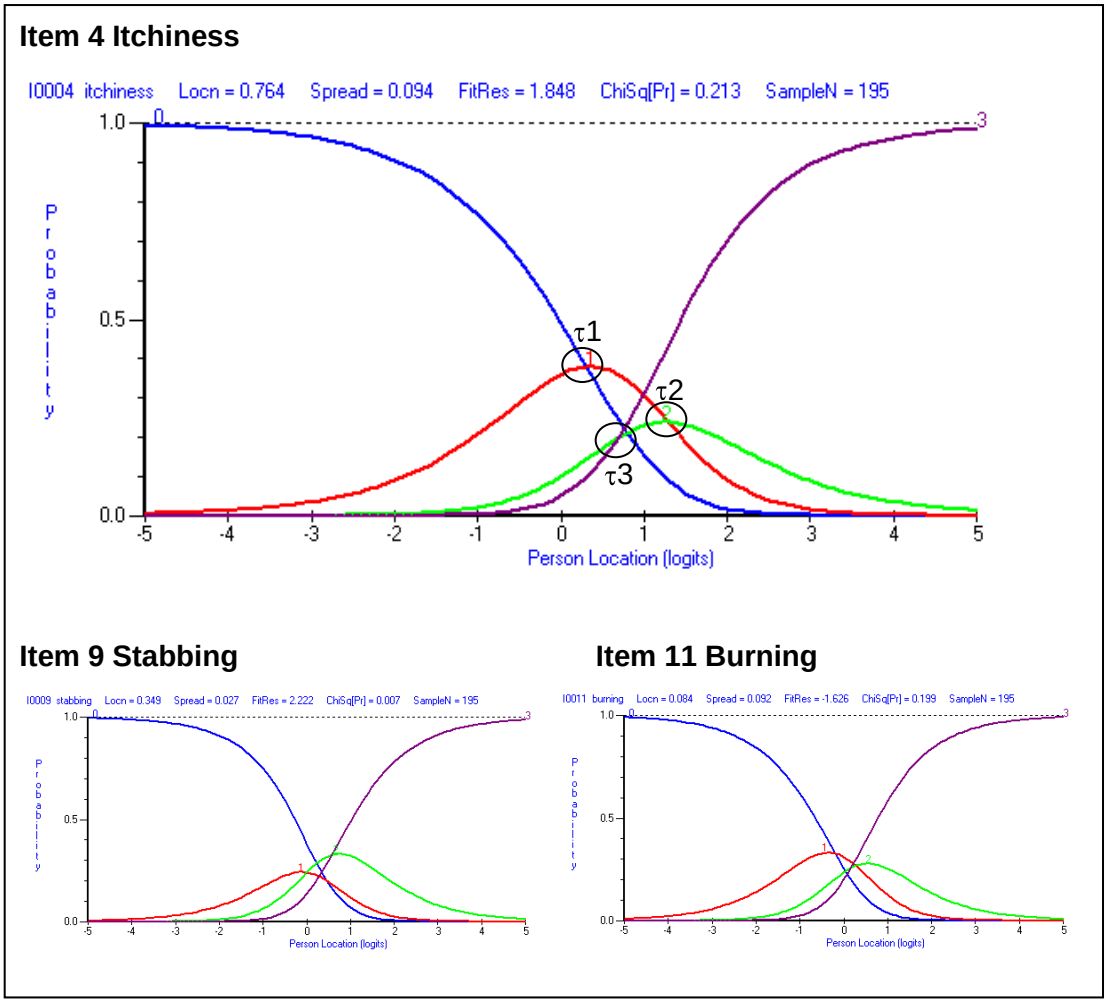


Figure 6.6 Items with disordered thresholds

Table 6.8 illustrates the 11 pain items ordered by their location (also called calibration), in ascending order (from most negative to most positive), and presents the mean item locations. Item locations range from about -1 to +1 logits (a 2 logit spread); the items define a line of increasing intensity, a continuum, rather than just a point but not a wide range. A two logit spread is acceptable because this scale is attempting to tap into a very specific issue, pain. Figure 6.7 illustrates that the scale is within the range where most of the people lie. The mean of item locations is set at 0 and estimates of the locations of items and persons is relative to each other, not their absolute locations, therefore having values ranging from negative to positive.

Figure 6.8 presents the measurement ruler mapped out by the 11 pain items, making its adequacy and limitations explicit and acting as an evidence base for improving the pain scale (e.g. adding items to capture measurement at the floor/ceiling range of the scale). Figure 6.7 complements Table 6.8 by illustrating item locations graphically. The 11 pain items are not evenly spread but bunched in two ways: 1) multiple items

have similar locations on the continuum (bunched around -0.8 to +1.0), indicated by thin arrow (Figure 6.7).

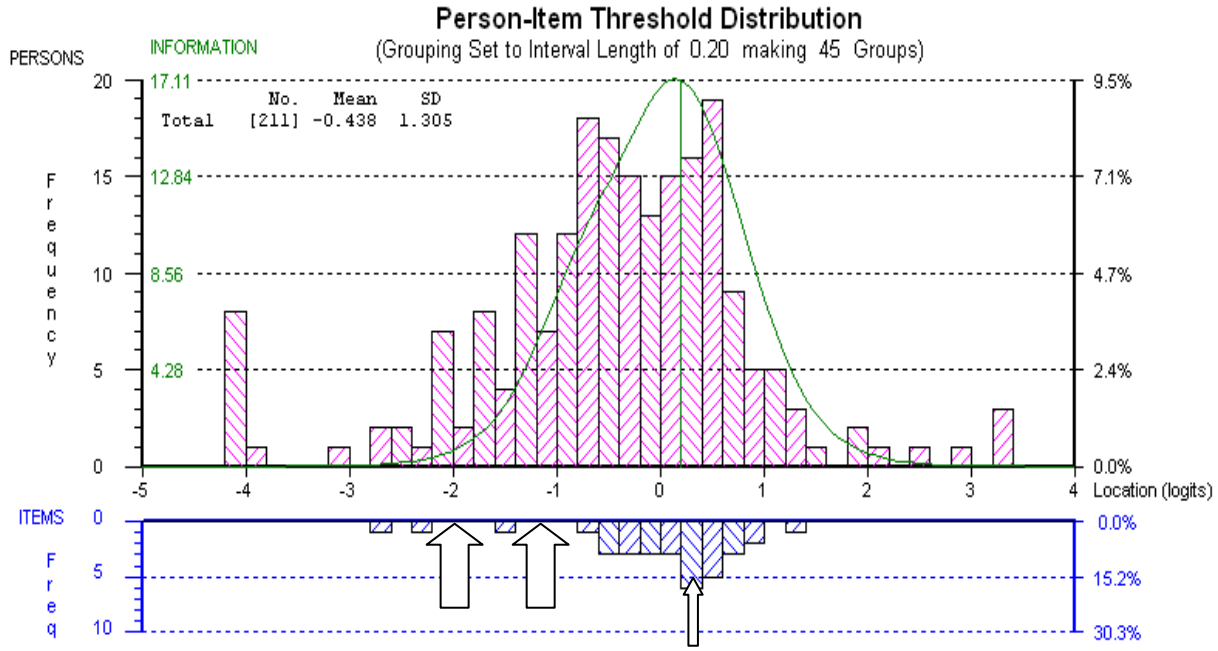


Figure 6.7 Targeting of patient sample to the pain items

Table 6.8 confirms items with similar locations (circled items): Items 'feeling uncomfortable' (-0.828) and 'tenderness' (-0.793); items 'stinging' (0.058) and 'burning' (0.084); and items 'dull ache' (0.183) and 'throbbing' (0.210). Items with similar locations raise the possibility of one of the items being redundant; and 2) there are notable gaps in the continuum mapped out by items between -0.8 and -1.4, and -1.6 and -2.2 units, indicated by thick arrow. Gaps imply limited measurement at those areas on the continuum that they attempt to map out.

6.6.5.4 Item locations

Is the location of items along the line of increasing intensity reasonable?

The ordering of items was examined to determine the extent to which it was consistent with clinical expectations and interpretability. If ordering is consistent with clinical expectation, it provides evidence towards the construct validity of the variable. Departures from expectation require investigation and explanation, and can occur when items are ambiguous, misleading or poorly worded (85). Item ordering for the pain scale is presented in Table 6.8. 'Feeling uncomfortable' (item 1) was predicted to be the least bothersome while 'itchiness' (item 4) was predicted to be the most

bothersome. The ordering of the 11 pain items were considered for clinical sensibility by a group of eight clinical specialists (i.e. TVN specialists). The ordering of items along the variable was clinically sensible, except for the location of the 'dull ache' item. This appeared to be more difficult than predicted. The specialists indicated that 'itchiness' may not be a property of pain but rather a separate symptom.

Table 6.8 Pain item locations in ascending order (n=195; 227 (32 extremes excluded))

Sequence	Item	Location	SE
1	Uncomfortable	-0.828	0.099
5	Tenderness	-0.793	0.096
3	Annoying	-0.562	0.095
10	Red raw	-0.088	0.095
8	Stinging	0.058	0.097
11	Burning	0.084	0.094
2	Dull ache	0.183	0.096
7	Throbbing	0.210	0.102
9	Stabbing	0.349	0.104
6	Tingling	0.625	0.011
4	Itchiness	0.764	0.116

Circled items indicate similar item locations
SE standard error

6.6.5.5 Item fit

Do the items work together to define a single variable?

Three fit statistics were examined: the item-person fit residuals; item-trait chi-squared values and probabilities; and ICC. There is no one indicator of fit of the observed data to the mathematical model, but rather multiple methods that each addresses different aspects of fit. These three indicators were interpreted separately and then interactively to make decisions about item fit.

- *Fit residual*

The fit residual evaluates the fit of the observed data to the Rasch model from the perspective of the items (298). It summarises the interaction between each scale item and all the persons for whom there is a response to that item, therefore the fit residual is a summary of the difference between observed and expected scores from every item response (item-person interaction). The Rasch model derives an observed item score (0, 1, 2 or 3 for items on the pain scale) and an expected score for every item-person interaction (value between 0 and 1). For each pain item, the residuals from the

interactions with each person in the sample (n=195; excluding 32 extremes) were squared, summed and transformed to give a summary value (the fit residual).

Fit residuals are expected to be normally distributed with mean = 0 and SD = 1. If data fit the model, the deviations (residuals) between observed responses and model-derived expected values should be no more than random errors. Therefore, if data fit the Rasch model, the mean fit residual across all items should be close to 0, the SD close to 1, and individual item values distributed in the approximate range +/-2.5 (more specifically, 99.5% of values in the range +/-2.5, and 99.9% in the range +/-3.0) (85). For individual items, observed fit residuals of 0 indicate no difference between observed and predicted scores (perfect fit); greater departure from 0 indicates greater discrepancy between observed and predicted responses, and thus greater misfit of the observed data to the model. The +/- sign associated with fit residual values indicates the type of misfit; negative values indicate over discriminating items while positive values indicate under discriminating items. Items with a fit residual of greater than +/- a certain level (generally set at 2.5) are cause for concern as they misfit to the model. Items with high negative residuals are normally interpreted to indicate item redundancy (i.e. information provided by the item is not adding any new information to the scale).

Table 6.9 presents fit residuals for the 11 pain items. Mean fit residual is 0.020, SD 1.644 (a SD <1.4 would suggest no problems with misfitting items) (Table 6.6). 'A dull ache' (item 2) is outside the +/-2.5 fit residual level (Table 6.9). No items have high negative residuals, suggesting no redundant items.

- *Chi-squared value and its probability*

The chi-squared value is a summary indicator of the interaction between individual items and the variable measured by the set of items (298). It is computed by summing the chi-squared values for a series of class intervals. Class intervals are achieved by dividing a given sample into a number of similarly sized groups based on their level of ability/disability (approximately 50 persons per class interval but it is more important to have equal distribution). Here, the disability is pain, measured by 11 items. For each class interval (in this case four), the mean location of people (n=195) and their mean score on each item (n=11) are computed; then, for each item, the mean observed scores for the class intervals are compared with the scores for those items predicted by the Rasch model at the mean location of the class interval (85). Thus, the chi-squared value for each item is the sum of the chi-squared values computed for each of the four class intervals, indicating the probability that the discrepancy between the observed mean and the expected value is large relative to chance. However, because chi-

squared values increase with sample size; are affected by the number of class intervals chosen; and they only approximate a chi-squared statistic (the value is inflated when the estimated probabilities are close to 0 or 1), they should be best used as an order statistic (i.e. order of degree of misfit) to indicate items showing greater values than others (items with significant chi-squared values (p < 0.05 or 0.01) should be examined), and to examine the ICC (298).

Table 6.9 Pain items in chi-squared probability order (n=195; 227 with 32 extremes excluded; 4 class intervals)

Item	Fit Residual	Chi-square	p
Tingling	0.636	0.622	0.733
Stinging	0.472	0.702	0.704
Annoying	-1.548	2.225	0.329
Itchiness	1.848	3.089	0.213
Burning	-1.626	3.229	0.199
Uncomfortable	-1.065	4.079	0.130
Throbbing	-1.143	4.289	0.117
Red raw	-0.673	6.683	0.035
Tenderness	-1.630	8.663	0.013
Stabbing	2.222	9.858	0.007
Dull ache	2.732*	25.092	0.000*

*indicates values outside recommended Bonferroni adjusted range (i.e. misfit)

Table 6.9 presents the 11 pain items ordered by increasing chi-squared value. Chi-square values range from 0.622 (tingling) to 25.092 (dull ache). Chi-Square values and how they change sequentially across items should be examined (298). All pain items have similar-level chi-square values (gradual increase in chi-square values), but then there is a notable step increase in value for the last item (dull ache; Table 6.9). Fit statistics indicate that one item (dull ache) has a significant chi-square probability; this item is not fitting the model at the Bonferroni adjusted significance level. As chi-square statistics are sample size dependent, as the sample size and number of items increase, the more the likelihood that items will have significance, therefore values are Bonferroni adjusted to account for this.

- *Item characteristic curve (ICC)*

The third indicator of observed data-to-Rasch model fit is the ICC. This graphical indicator of fit aids in the interpretation of the two fit statistics detailed above. The ICC,

a graph (S-shaped curve), plots the expected individual item response (predicted from the Rasch model) at every level of the measurement continuum (85).

Figure 6.8 illustrates the ICC for the item that 'failed' the item-trait chi-square value tests of fit. Item 2 is significantly under-discriminating, indicating that this item is not doing a good job at discriminating those persons with more pain. The observed responses (black dots) are close to representing a horizontal line which indicates that the item does not discriminate difficulty.

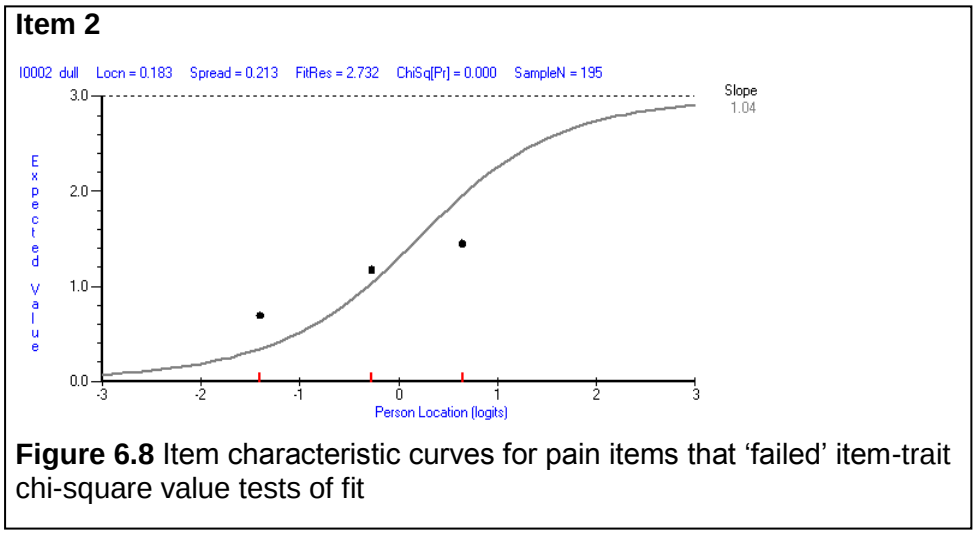


Figure 6.8 Item characteristic curves for pain items that 'failed' item-trait chi-square value tests of fit

Figure 6.9 presents the ICC for all pain items, apart from item 2. ICCs indicate the *expected* score (*y*-axis) (observed responses can be 0, 1, 2 or 3 for pain bother but as responses are continuous, values can vary from 0 to 3) on each item for each possible location on the pain continuum (*x*-axis). The four small vertical marks on the *x*-axis indicate the mean person locations for each of the four class intervals. The four black dots indicate the *observed* mean score on each item for each of these four class intervals.

The chi-squared values in Table 6.9 summarise the consistency between the observed responses (black dots) and expected responses (ICC) at the four points on the continuum; computed by summing the item chi-squared values for each class interval. Observed responses should closely match expected responses.

Ten of 11 pain items pass all three fit criteria. Item 2 has notable criterion failures: fit residual outside recommended range; high chi-squared value with significant p-value and adherence to the ICC (significantly under-discriminating). No other items have departures from expectation.

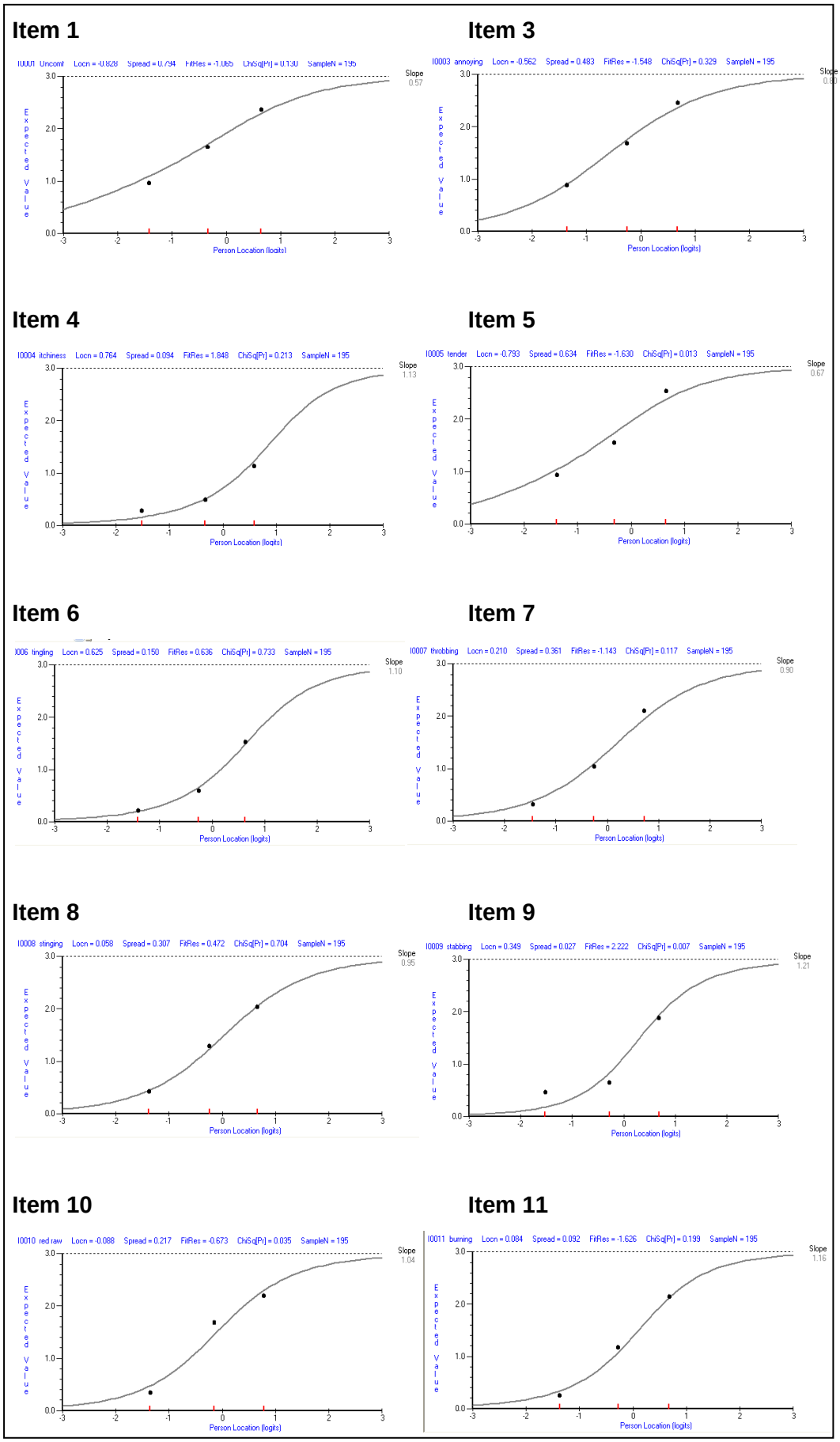


Figure 6.9 Item characteristic curves for 'fitting' items including class intervals

6.6.5.6 Local dependency

Does the response to one item directly influence the response to another?

The local dependency of items was examined by looking at correlations between residuals. The Rasch model requires that items in a scale are locally independent (i.e. response to one item is independent of the response to another item). The assumption of local independence implies that once the Rasch factors have been extracted (the main scales), no leftover patterns in the residuals should be present; the residuals (observed – expected = residual value) should represent random error. An absence of any meaningful pattern in the residuals (i.e. item residuals not correlated), is considered support for the assumption of local independence. Correlated items imply that the answer to one item is dependent on another. Thus, correlations among residuals should be low (the criterion of <0.30118 represents 10% shared variance but a value <0.4 is acceptable) (298).

For the pain scale, in this sample, all residual correlations were below the threshold. This implies that the responses of the items are independent of each other and that the items are locally independent (Table 6.10).

Table 6.10 Person-Item Residual correlation matrix

Item	I0001	I0002	I0003	I0004	I0005	I0006	I0007	I0008	I0009	I0010	I0011
I0001	1.000										
I0002	-0.070	1.000									
I0003	0.262	-0.165	1.000								
I0004	-0.271	-0.035	-0.131	1.000							
I0005	0.171	-0.123	0.003	-0.207	1.000						
I0006	-0.288	-0.003	-0.314	-0.088	-0.369	1.000					
I0007	-0.225	-0.268	-0.103	-0.039	-0.139	0.239	1.000				
I0008	-0.308	-0.267	-0.002	-0.133	-0.074	-0.124	-0.003	1.000			
I0009	-0.288	-0.227	-0.234	-0.140	-0.206	0.019	0.048	0.082	1.000		
I0010	0.032	-0.175	-0.149	-0.152	0.028	-0.200	-0.131	-0.206	-0.182	1.000	
I0011	-0.091	-0.245	-0.135	0.009	-0.184	0.036	-0.234	-0.067	-0.182	0.124	1.000

6.6.5.7 Reliability

Have the people in the sample been measured successfully? (Are the persons in the sample separated along the line defined by the items?).

Figure 6.2 illustrates the distribution of person measurements (locations) relative to the item locations. The sample is well spread with values ranging from around -4.0 to +3.2

logits. The mean is -0.438 (SD 1.305), indicating the sample is off-centre of the items (mean of item locations is always 0).

Figure 6.2 is a graphical indicator that the pain scale items have been successful in separating this sample of people with PUs. A numerical indicator of the degree of separation is the PSI, computed from the person location estimates as the variation among person locations relative to the error of estimate for each person (316). The PSI, comparable to Cronbach's alpha, is consistent with the traditional definition of reliability of a scale. It provides information about how reliably a scale distinguishes between responders and how much of the variation of person estimates can be attributed to error variances (i.e. the extent to which scores are associated with random error); reliability indicators range from 0 (all error) to 1 (no error). A low PSI indicates that the fit statistics obtained may not be reliable as there will be a certain amount of error surrounding them while a high PSI indicates more reliable fit statistics.

For the pain scale, the PSI is 0.810 with extremes and 0.777 with no extremes, indicating good reliability (Table 6.6). A low PSI is likely when there is large mistargeting or the construct is too big (see Section 6.5.3, *Reliability*). In this instance, the construct is not very wide (specific pain issue) so targeting is within the expected range; the distribution curve indicates that the scale is within the range where most of the sample lie (Figure 6.3).

6.6.6 Modification and reanalysis of Pain scale

The sample was adequate for examining the scale but the scale was suboptimal for measuring the sample, suggesting that the scale might provide limited information about people at the extremes of the sample distribution. The pain scale items mapped out a variable of increasing intensity, but not very wide, and the item locations indicated areas on the continuum where measurement could be improved. There was some bunching of thresholds at the centre of the scale, however if persons at the floor were removed then the targeting would be very good.

The ordering of items along the variable was clinically sensible, except for the location of the 'dull ache' and 'itchiness' items; 'dull ache' appeared to be more difficult than predicted while 'itchiness' was considered not a property of pain but a separate symptom. The four-category scoring function for the items did not work as intended for 5/11 items. For the other items, thresholds are very close to being disordered, suggesting that people had difficulty distinguishing between 'a little bother' and 'quite a bit of bother' categories. The fit statistics highlighted that the 'dull ache' item did not fit

the expectations of the measurement model. No dependence among the pain scale items was observed in terms of residual correlations.

Based on the Rasch findings, all response categories were rescored by combining response categories ‘a little bother’ and ‘quite a bit of bother’, resulting in a three-point response scale, and items ‘dull ache’ and ‘itchiness’ were removed from the scale and the Rasch analysis reran.

6.6.7 Summary of findings from final Rasch analysis of the pain scale

Rescoring the response categories and removing items ‘dull ache’ and ‘itchiness’ resulted in a good brief PU-specific pain scale. The PSI was slightly lower (Table 6.11) but the three-category scoring function did not produce any disordered thresholds and the ordering of items along the variable was clinically sensible. The fit statistics indicated no misfitting items and there was no dependence among the items in terms of residual correlations. Item locations ranged from -1.0 to +1.0 and person locations range from -4.665 to 4.033, both ranges slightly wider than in the earlier analysis. However, similarly, the sample is reasonable for examining the scale but a suboptimal scale for measuring the sample, with no items representing people at the extremes of the sample distribution. About 88 (45.8%) of people in the sample are located outside the best functioning of the scale. Table 6.12 provides a summary of the Rasch statistics in item location order (indicating the line of increasing intensity for the pain scale) for the final PU-QOLI items.

The analyses described in Section 6.5.2 were performed on all remaining PU-QOLI scales. A summary of the Rasch results will be presented for each scale, followed by decisions made to any scale modifications, and finally presentation of the psychometric properties in summary tables.

6.6.8 Summary of Rasch findings and recommendations for Exudate scale

Item locations for the exudate scale range from -0.6 to +0.6 and person locations range from -3.2 to +3.2 logits. There are no items representing people at the extremes of the sample distribution. About 53.8% of people in the sample are located outside the best functioning of the scale (i.e. range -1 to +1 logits). Item threshold locations spread from -1.6 to +1.2 logits but still do not cover the range of person locations in the sample. The sample was adequate for examining the scale (item locations are covered by the people) but the scale was suboptimal for measuring the sample (person locations not covered by the items), suggesting that the scale might provide limited

information about people at the extremes of the sample distribution. The exudate scale items mapped out a variable of increasing intensity, but not very wide, and the item locations indicated areas on the continuum where measurement could be improved (i.e. at the extreme ends of the scale).

Item 1 ('Weeping') was predicted to be the least bothersome while item 7 ('Pus') was predicted to be the most bothersome. The ordering of the 8 exudate items along the variable was clinically sensible. 4/8 exudate items had disordered thresholds; the point where the probability for scoring either 2 ('quite a bit of bother') or 3 ('a lot of bother') was lower than the point on the continuum where the probability of scoring either 1 ('a little bother') or 2 ('quite a bit of bother') occurred. For the other items, the response categories were working as intended, however inspecting CPCs indicated thresholds (τ1 and τ2) were close to being disordered, suggesting that people had difficulty distinguishing between 'a little bother' and 'quite a bit of bother' categories.

The exudate scale mean fit residual was -0.224, SD 1.440. Item fit residuals ranged from -1.975 to 2.259. Chi-square values ranged from 0.066 to 6.261. All exudate items had similar-level chi-square values (gradual increase in chi-square values), no significant chi-square probabilities at the Bonferroni adjusted significance level, and no notable step increase in values. The fit statistics indicate that all exudate items fit the expectations of the measurement model. No dependence among the exudate scale items was observed in terms of residual correlations. In this sample, 'weeping' (item 1) and 'running' (item 2) had residual correlations exceeding 0.30 (0.320), however only slightly outside the recommended criterion. No other residual correlations exceeded 0.30; implying that the responses to items are independent of each other and items are locally independent. No items had high negative residuals (all below -0.50), suggesting no redundant items.

The sample is well spread with values ranging -3.2 to +3.2 logits. The mean is -1.144 (SD 1.579), indicating the sample is off-centre of the items (mean of item locations is always 0). The PSI is 0.715 with extremes and 0.585 with no extremes, indicating reasonable reliability. In this instance, the construct is not very wide (specific exudate issue) and not all people (only those with severe PUs) have problems with exudate; therefore the targeting is within the expected context. Based on the Rasch findings, all response categories were rescored by combining response categories 'a little bother' and 'quite a bit of bother', resulting in a three-point response scale and the Rasch analysis reran.

6.6.9 Summary of findings from final Rasch analysis of the exudate scale

Rescoring the response categories resulted in a good brief exudate scale. The PSI was slightly lower (Table 6.11) but still reasonable reliability. The three-category scoring function did not produce any disordered thresholds and the ordering of items along the variable was clinically sensible. The fit statistics indicated no misfitting items and there was no dependence among the items in terms of residual correlations exceeding 0.3. Item locations ranged from -0.709 to +0.697 and person locations range from -3.813 to 3.645, both ranges are slightly wider than in the earlier analysis. However, similarly, the sample is reasonable for examining the scale but a suboptimal scale for measuring the sample, with no items representing people at the extremes of the sample distribution. About 61.4% of people in the sample are located outside the best functioning of the scale. Table 6.12 provides a summary of the Rasch statistics in item location order (indicating the line of increasing intensity for the exudate scale).

6.6.10 Summary of Rasch findings and recommendations for Odour scale

Item locations for the odour scale range from -1.6 to +0.6 and person locations range from -5.6 to +3.6 logits. There are no items representing people at the extremes of the sample distribution. About 75% of people in the sample are located outside the best functioning of the scale (i.e. range -1 to +1 logits). Item threshold locations spread from -4.6 to +2.0 logits but still do not cover the full range of person locations in the sample. The sample was adequate for examining the scale but the scale was suboptimal for measuring the sample (the scale might provide limited information about people at the extremes of the sample distribution). The odour scale items mapped out a variable of increasing intensity, but not very wide, and the item locations indicated areas on the continuum where measurement could be improved (i.e. at the extreme ends of the scale and between -4.2 and -1.2 logits).

Item 1 (‘Unpleasant smell’) was predicted to be the least bothersome while item 6 (‘putrid smell’) was predicted to be the most bothersome. The ordering of the 6 odour items along the variable was clinically sensible. 2/6 odour items had disordered thresholds; the remaining items were close to being disordered, suggesting that people had difficulty distinguishing between ‘a little bother’ and ‘quite a bit of bother’ categories.

The odour scale mean fit residual was -0.141, SD 0.702. Item fit residuals ranged from -0.962 to 0.897. Chi-square values ranged from 0.812 to 3.286. All odour items had similar-level chi-square values (gradual increase in chi-square values), no significant chi-square probabilities at the Bonferroni adjusted significance level, and no notable

step increase in values. The fit statistics indicate that all odour items fit the expectations of the measurement model. No dependence among the odour scale items was observed in terms of residual correlations exceeding 0.3, implying that the responses to items are independent of each other and items are locally independent. No items had high negative residuals (all below -0.55), suggesting no redundant items.

The sample is well spread with values ranging -5.402 to +3.474 logits. The mean is -2.936 (SD 2.910), indicating the sample is off-centre of the items (mean of item locations is always 0). The PSI is 0.550 with extremes and 0.622 with no extremes, indicating reasonable reliability. In this instance, like the exudate scale, the construct is not very wide and not all people (only those with severe PUs) have problems with odour, therefore the targeting is within the expected context. Based on the Rasch findings, all response categories were rescored by combining response categories ‘a little bother’ and ‘quite a bit of bother’, resulting in a three-point response scale and the Rasch analysis reran.

6.6.11 Summary of findings from final Rasch analysis of the Odour scale

Rescoring the response categories resulted in a good brief odour scale. The PSI was slightly lower (Table 6.11) but still reasonable reliability. The three-category scoring function did not produce any disordered thresholds and the ordering of items along the variable was clinically sensible. The fit statistics indicated no misfitting items and there was no dependence among the items in terms of residual correlations exceeding +0.3 or items with high negative residuals (all below -0.59), suggesting no redundant items. Item locations ranged from -2.063 to +0.801 and person locations range from -6.376 to 3.925, both ranges are slightly wider than in the earlier analysis. However, similarly, the sample is reasonable for examining the scale but a suboptimal scale for measuring the sample, with no items representing people at the extremes of the sample distribution, with a large proportion of the sample (about 83%) located outside the best functioning of the scale. Table 6.12 provides a summary of the Rasch statistics in item location order (indicating the line of increasing intensity for the odour scale).

6.6.12 Summary of Rasch findings and recommendations for Sleep scale

Item locations for the sleep scale ranged from -0.5 to +0.3 and person locations range from -3.2 to +3.6 logits. There are no items representing people at the extremes of the sample distribution. About 49.7% of people in the sample are located outside the best functioning of the scale (i.e. range -1 to +1 logits). Item threshold locations spread from -1.6 to +2.0 logits but still do not cover the full range of person locations in the

sample. The sample was adequate for examining the scale but the scale was suboptimal for measuring the sample (the scale might provide limited information about people at the extremes of the sample distribution). The sleep scale items mapped out a variable of increasing intensity, but not very wide, and the item locations indicated areas on the continuum where measurement could be improved (i.e. at the extreme ends of the scale).

Item 6 ('trouble finding comfortable position') was predicted to be the least bothersome while item 1 ('trouble falling asleep') was predicted to be the most bothersome. The ordering of the 6 sleep items along the variable was clinically sensible. 3/6 sleep items had disordered thresholds; the remaining items were close to being disordered, suggesting that people had difficulty distinguishing between 'a little bother' and 'quite a bit of bother' categories.

The sleep scale mean fit residual was 0.148, SD 1.166. Item fit residuals ranged from -1.123 to 1.852. Chi-square values ranged from 1.129 to 10.068. All sleep items had similar-level chi-square values (gradual increase in chi-square values), with a notable step increase in values from 1.868 (item 6, 'trouble finding comfortable position') to 9.657 (item 4, 'not getting amount of sleep needed') and no significant chi-square probabilities at the Bonferroni adjusted significance level. Fit statistics indicated all sleep items fit the expectations of the measurement model. No dependence among the sleep scale items was observed in terms of residual correlations exceeding 0.3, implying that the responses to items are independent of each other and items are locally independent. No items had high negative residuals (all below -0.58), suggesting no redundant items.

The sample is well spread with values ranging -3.031 to +3.574 logits. The mean is -0.466 (SD 1.766), indicating the sample is off-centre of the items (mean of item locations is always 0). The PSI is 0.763 with extremes and 0.624 with no extremes, indicating reasonable reliability. In this instance, the construct is not very wide and not all people have problems with sleep, hence targeting is within the expected context. Based on the Rasch findings, all response categories were rescored by combining response categories 'a little bother' and 'quite a bit of bother', resulting in a three-point response scale and the Rasch analysis reran.

6.6.13 Summary of findings from final Rasch analysis of the Sleep scale

Rescoring the response categories resulted in a good brief sleep scale. The PSI was slightly lower (Table 6.11) but still reasonable reliability. The three-category scoring

function did not produce any disordered thresholds and the ordering of items along the variable was clinically sensible. The fit statistics indicated no misfitting items and there was no dependence among the items in terms of residual correlations exceeding +0.3 or items with high negative residuals (all below -0.55), suggesting no redundant items. Item locations ranged from -0.751 to +0.349 and person locations range from -3.761 to 3.729, both ranges are slightly wider than in the earlier analysis. However, similarly, the sample is reasonable for examining the scale but a suboptimal scale for measuring the sample, with no items representing people at the extremes of the sample distribution, with a large proportion of the sample (about 69.6%) located outside the best functioning of the scale. Table 6.12 provides a summary of the Rasch statistics in item location order (indicating the line of increasing intensity for the sleep scale).

6.6.14 Summary of Rasch findings and recommendations for Mobility scale

Item locations for the mobility scale range from -0.4 to +0.6 and person locations range from -3.8 to +3.8 logits. There are no items representing people at the extremes of the sample distribution. About 39.8% of people in the sample are located outside the best functioning of the scale (i.e. range -1 to +1 logits). Item threshold locations spread from -1.8 to +1.8 logits but still do not cover the range of person locations in the sample. The sample was adequate for examining the scale (item locations are covered by the people) but the scale was suboptimal for measuring the sample (person locations not covered by the items), suggesting that the scale might provide limited information about people at the extremes of the sample distribution. The mobility scale items mapped out a variable of increasing intensity, but not very wide, and the item locations indicated areas on the continuum where measurement could be improved (i.e. at the extreme ends of the scale).

Item 11 (‘walking slowed’) was predicted to be the least bothersome while item 6 (‘difficulty transferring’) was predicted to be the most bothersome. The ordering of the 8 exudate items along the variable was clinically sensible. 4/11 mobility items had disordered thresholds; the point where the probability for scoring either 2 (‘quite a bit of bother’) or 3 (‘a lot of bother’) was lower than the point on the continuum where the probability of scoring either 1 (‘a little bother’) or 2 (‘quite a bit of bother’) occurred. For the other items, the response categories were working as intended, however inspecting CPCs indicated thresholds (τ1 and τ2) were close to being disordered, suggesting that people had difficulty distinguishing between ‘a little bother’ and ‘quite a bit of bother’ categories.

The mobility scale mean fit residual was 0.114, SD 1.241. Item fit residuals ranged from -1.694 to 2.332. Chi-square values ranged from 0.313 to 4.887. All mobility items had similar-level chi-square values (gradual increase in chi-square values), no significant chi-square probabilities at the Bonferroni adjusted significance level, and no notable step increase in values. The fit statistics indicate that all mobility items fit the expectations of the measurement model. Dependence among some mobility items was observed in terms of residual correlations. Positive correlations ranged from 0.03 to 0.86 for item 9 ('ability to walk') and item 10 ('how far able to walk'), suggesting local dependence. Negative correlations ranged from -0.14 to -0.81 for item 3 ('difficulty turning or moving in bed') and item 9 ('ability to walk') suggesting one of the items was redundant.

The sample is well spread with values ranging -3.701 to +3.738 logits. The location mean is -0.067 (SD 1.582), indicating the sample is slightly off-centre of the items (mean of item locations is always 0). The PSI is 0.753 with extremes and 0.582 with no extremes, indicating reasonable reliability at distinguishing between responders on mobility impairment. Based on the Rasch findings, all response categories were rescored by combining response categories 'a little bother' and 'quite a bit of bother', resulting in a three-point response scale and the Rasch analysis reran.

6.6.15 Summary of findings from final Rasch analysis of the mobility scale

The three-category scoring function did not produce any disordered thresholds and the ordering of items along the variable was clinically sensible. However, fit statistics indicated two items failed tests of fit; 'pushing up to sitting' (item 4) and 'sitting for long periods' (item 5) had fit residual values outside +/-2.5 logits but not exceeding +/-3.0, indicating a small departure from expectation). Item 9 ('ability to walk') and item 10 ('how far able to walk') had 0.890 residual correlation value, suggesting local dependence and 'sitting for long periods' (item 5) had high negative correlations (>0.7) with two items. Based on these findings, items 5 and 10 were removed and Rasch reran.

Removing these two items resulted in slightly improved item spread; all items met fit expectations of the measurement model; and residual correlations reduced (<0.52). However, the reliability of the scale to distinguish between responders on mobility impairment also reduced (PSI 0.663 with extremes and 0.323 without extremes). Table 6.12 provides a summary of the Rasch statistics in item location order (indicating the line of increasing intensity for the mobility scale).

6.6.16 Summary of Rasch findings and recommendations for Activity scale

Item locations for the activity scale ranged from -0.6 to +0.6 and person locations range from -2.6 to +2.6 logits. There are no items representing people at the extremes of the sample distribution. About 41.3% of people in the sample are located outside the best functioning of the scale (i.e. range -1 to +1 logits). Item threshold locations spread from -2.4 to +1.8 logits but still do not cover the full range of person locations in the sample. The sample was adequate for examining the scale but the scale was suboptimal for measuring the sample (the scale might provide limited information about people at the extremes of the sample distribution). The activity scale items mapped out a variable of increasing intensity, but not very wide, and the item locations indicated areas on the continuum where measurement could be improved (i.e. at the extreme ends of the scale and between -2.2 and -1.2 logits).

Item 1 ('washing') was predicted to be the least bothersome while item 8 ('being emotionally close') was predicted to be the most bothersome. The ordering of the 9 activity items along the variable was clinically sensible. 8/9 activity items had disordered thresholds; people had difficulty distinguishing between 'a little bother' and 'quite a bit of bother' categories.

The activity scale mean fit residual was -0.057, SD 0.733. Item fit residuals ranged from -0.762 to 1.182. Chi-square values ranged from 0.158 to 3.775. All activity items had similar-level chi-square values (gradual increase in chi-square values), no significant chi-square probabilities at the Bonferroni adjusted significance level, and no notable step increase in values. The fit statistics indicate that all activity items fit the expectations of the measurement model. Dependence among some activity items was observed in terms of residual correlations. Items 5 ('gardening') and 6 ('shopping') had 0.62 residual correlations, suggesting local dependence. No items had high negative residuals (all below -0.65), suggesting no redundant items.

The sample is well spread with values ranging -2.580 to +2.520 logits. The mean is -0.461 (SD 1.125), indicating the sample is off-centre of the items (mean of item locations is always 0). The PSI is 0.448 with extremes and 0.290 with no extremes, indicating poor reliability; the scale is not reliably distinguishing between responders on activity impairment. Based on the Rasch findings, all response categories were rescored by combining response categories 'a little bother' and 'quite a bit of bother', resulting in a three-point response scale, item 5 ('gardening') was removed as it was considered only relevant to a small proportion of people with PUs compared to item 6 ('shopping'), and the Rasch analysis reran.

6.6.17 Summary of findings from final Rasch analysis of the activities scale

The three-category scoring function did not produce any disordered thresholds and the ordering of items along the variable was clinically sensible. The fit statistics indicated no misfitting items and there was no dependence among the items in terms of residual correlations exceeding +0.3 or items with high negative residuals (all below -0.57), suggesting no redundant items. However, the PSI was slightly lower (Table 6.11), indicating poor reliability of the scale to distinguish between responders on activity impairment. Table 6.12 provides a summary of the Rasch statistics in item location order (indicating the line of increasing intensity for the activity scale).

6.6.18 Summary of Rasch findings and recommendations for the vitality scale

The decision was made to add items to the vitality scale as the three items did not produce a scale that met requirements for reliable and valid measurement. Patient transcripts were revisited (Chapter 3) and an additional three items were added to the scale. This new scale was empirically tested in the subsequent field test (Chapter 7).

6.6.19 Summary of Rasch findings and recommendations for emotional well-being scale

The decision was made to combine mood, anxiety, autonomy and isolation scale items into an emotional well-being scale. Apart from the mood scale, both the anxiety and isolation scale did not produce scales that met requirements for reliable and valid measurement. Both traits were however considered properties important to people with PUs and therefore it was important to retain these in a PU-specific PRO instrument. There was some uncertainty pertaining to whether the isolation items might fit better with the participation scale but as the isolation items were about feelings and not actually being physically isolated, conceptually they fitted better with emotional well-being; theoretically it made sense to combine all these items into an emotional well-being scale. This would be tested empirically.

Item locations for the emotional well-being scale ranged from -1.2 to +1.5 and person locations range from -4.4 to +4.0 logits. There are no items representing people at the extremes of the sample distribution. About 39.3% of people in the sample are located outside the best functioning of the scale (i.e. range -1 to +1 logits). Item threshold locations spread from -2.4 to +2.2 logits, a much improved coverage of persons but still do not cover the full range of person locations in the sample. The sample was adequate for examining the scale but the scale was suboptimal for measuring the sample (the scale might provide limited information about people at the extremes of the

sample distribution). The emotional well-being scale items mapped out a variable of increasing intensity (> 4 logit spread) but the item locations indicated areas on the continuum where measurement could be improved (i.e. extreme ends of the scale).

Item 2 ('fed-up') was predicted to be the least bothersome while item 17 ('people avoided or treated you differently') was predicted to be the most bothersome. The ordering of the 17 emotional well-being items along the variable was clinically sensible. 6/17 emotional well-being items had disordered thresholds; other items were close to being disordered, suggesting that people had difficulty distinguishing between 'a little bother' and 'quite a bit of bother' categories.

The emotional well-being scale mean fit residual was -0.154, SD 1.379. Item fit residuals ranged from -3.147 to 2.274. Chi-square values ranged from 0.166 to 9.735. All emotional well-being items had similar-level chi-square values (gradual increase in chi-square values), no notable step increase in values and no significant chi-square probabilities at the Bonferroni adjusted significance level. However, item 8 ('feeling upset') had a fit residual statistic outside the recommended range (-3.147). Dependence was observed, in terms of high positive residual correlations (0.56), for items 6 ('feeling down') and 7 ('feeling depressed'), suggesting local dependence. No items had high negative residuals (all below -0.44), suggesting no redundant items.

The sample is well spread with values ranging -3.980 to +3.629 logits. The mean is -0.488 (SD 1.535), indicating the sample is off-centre of the items (mean of item locations is always 0). The PSI is 0.840 with extremes and 0.817 with no extremes, indicating good reliability. Based on the Rasch findings, all response categories were rescored by combining response categories 'a little bother' and 'quite a bit of bother', resulting in a three-point response scale. Items feeling down' and 'feeling upset' were removed and the Rasch analysis reran.

6.6.20 Summary of findings from final Rasch analysis of the emotional well-being scale

Rescoring the response categories resulted in a good emotional well-being scale. The PSI was slightly lower (Table 6.11) but still good reliability. The three-category scoring function did not produce any disordered thresholds and the ordering of items along the variable was clinically sensible. The fit statistics indicated no misfitting items and there was no dependence among the items in terms of residual correlations exceeding +0.41 or items with high negative residuals (all below -0.41), suggesting no redundant items. Item locations ranged from -1.539 to +2.081 and person locations range from -4.866 to

4.506, both ranges are slightly wider than in the earlier analysis. However, similarly, the sample is reasonable for examining the scale but a suboptimal scale for measuring the sample, with no items representing people at the extremes of the sample distribution, with a large proportion of the sample (about 55.0%) located outside the best functioning of the scale. Table 6.12 provides a summary of the Rasch statistics in item location order (indicating the line of increasing intensity for the emotional well-being scale).

6.6.21 Summary of Rasch findings and recommendations for Self-consciousness scale

Item locations for the self-consciousness scale ranged from -1.0 to +0.8 and person locations range from -3.2 to +3.0 logits. There are no items representing people at the extremes of the sample distribution. About 54.3% of people in the sample are located outside the best functioning of the scale (i.e. range -1 to +1 logits). Item threshold locations spread from -1.8 to +1.6 logits but still do not cover the full range of person locations in the sample. The sample was adequate for examining the scale but the scale was suboptimal for measuring the sample (the scale might provide limited information about people at the extremes of the sample distribution). The self-consciousness scale items mapped out a variable of increasing intensity, but not very wide, and the item locations indicated areas on the continuum where measurement could be improved (i.e. at the extreme ends of the scale).

Item 1 (‘feeling helpless’) was predicted to be the least bothersome while item 7 (‘feeling lack of understanding’) was predicted to be the most bothersome. The ordering of the 7 self-consciousness items along the variable was clinically sensible. 84/7 self-consciousness items had disordered thresholds; people had difficulty distinguishing between ‘a little bother’ and ‘quite a bit of bother’ categories.

The self-consciousness scale mean fit residual was -0.356, SD 0.892. Item fit residuals ranged from -1.650 to 0.981. Chi-square values ranged from 1.050 to 3.849. All self-consciousness items had similar-level chi-square values (gradual increase in chi-square values), no significant chi-square probabilities at the Bonferroni adjusted significance level, and no notable step increase in values. The fit statistics indicate that all self-consciousness items fit the expectations of the measurement model. No dependence among the self-consciousness items was observed in terms of residual correlations exceeding 0.3, implying that the responses to items are independent of each other and items are locally independent. No items had high negative residuals (all below -0.46), suggesting no redundant items.

The sample is well spread with values ranging -3.010 to +2.878 logits. The mean is -0.945 (SD 1.427), indicating the sample is off-centre of the items (mean of item locations is always 0). The PSI is 0.658 with extremes and 0.562 with no extremes, indicating reasonable reliability. Based on the Rasch findings, all response categories were rescored by combining response categories 'a little bother' and 'quite a bit of bother', resulting in a three-point response scale, and the Rasch analysis reran.

6.6.22 Summary of findings from final Rasch analysis of the self-consciousness scale

The three-category scoring function did not produce any disordered thresholds and the ordering of items along the variable was clinically sensible. The fit statistics indicated no misfitting items and there was no dependence among the items in terms of residual correlations exceeding +0.3 or items with high negative residuals (all below -0.43), suggesting no redundant items. The PSI improved slightly (Table 6.11), indicating reasonable reliability of the scale to distinguish between responders on self-consciousness impairment. Table 6.12 provides a summary of the Rasch statistics in item location order (indicating a line of increasing intensity for the self-consciousness scale).

6.6.23 Summary of Rasch findings and recommendations for Participation scale

Item locations for the participation scale ranged from -0.6 to +0.6 and person locations range from -3.2 to +3.6 logits. There are no items representing people at the extremes of the sample distribution. About 58.0% of people in the sample are located outside the best functioning of the scale (i.e. range -1 to +1 logits). Item threshold locations spread from -2.2 to +1.8 logits but still do not cover the full range of person locations in the sample. The sample was adequate for examining the scale but the scale was suboptimal for measuring the sample (the scale might provide limited information about people at the extremes of the sample distribution). The participation scale items mapped out a variable of increasing intensity, but not very wide, and the item locations indicated areas on the continuum where measurement could be improved (i.e. at the extreme ends of the scale).

Item 6 ('restricted where go out') was predicted to be the least bothersome while item 2 ('meeting/seeing family/friends') was predicted to be the most bothersome. The ordering of the 9 participation items along the variable was clinically sensible. 3/9 participation items had disordered thresholds; people had difficulty distinguishing between 'a little bother' and 'quite a bit of bother' categories.

The participation scale mean fit residual was 0.202, SD 1.002. Item fit residuals ranged from -1.673 to 1.441. Chi-square values ranged from 0.289 to 6.556. All activity items had similar-level chi-square values (gradual increase in chi-square values), no significant chi-square probabilities at the Bonferroni adjusted significance level, and no notable step increase in values. The fit statistics indicate that all participation items fit the expectations of the measurement model. No dependence among the participation items was observed in terms of residual correlations exceeding 0.3, implying that the responses to items are independent of each other and items are locally independent. Items 2 ('seeing/meeting family/friends') and 9 ('time spent caring for PU') had high negative residuals (-0.62), suggesting one might be redundant items. However, at this stage, both items were retained as they were considered to measure important properties of participation impairment.

The sample is well spread with values ranging -3.346 to +3.590 logits. The mean is -0.380 (SD 1.727), indicating the sample is off-centre of the items (mean of item locations is always 0). The PSI is 0.705 with extremes and 0.647 with no extremes, indicating reasonable reliability. Based on the Rasch findings, all response categories were rescored by combining response categories 'a little bother' and 'quite a bit of bother', resulting in a three-point response scale, and the Rasch analysis reran.

6.6.24 Summary of findings from final Rasch analysis of the participation scale

The three-category scoring function did not produce any disordered thresholds and the ordering of items along the variable was clinically sensible. The fit statistics indicated no misfitting items and there was no dependence among the items in terms of residual correlations exceeding +0.3 or additional items with high negative residuals (all below -0.63). However, the PSI was slightly lower (Table 6.11), indicating reasonable reliability of the scale to distinguish between responders on participation impairment. Table 6.12 provides a summary of the Rasch statistics in item location order (indicating the line of increasing intensity for the participation scale).

6.6.25 Summary of Rasch results and interpretations

The targeting between the distribution of person measurements and the distribution of item locations indicated that the samples were adequate for examining the scales but the scales were suboptimal for measuring the sample. There was a significant ceiling effect where item coverage was poorest above -2.0 logits for most scales, indicating that the scales might provide limited information about people at the extremes of the sample distribution (those with the least disability or impairment). However, the

ordering of scale items along each variable was clinically sensible, providing evidence towards the construct validity of each scale variable.

Some items had notable criterion failures: outside the +/-2.5 fit residual level; high chi-squared values with significant p-value and adherence to the ICC (significantly under-discriminating or over-discriminating). Few items exceeded residual correlations of +0.3, implying that the responses to items are independent of each other and locally independent, and -0.3, suggesting no redundant items. Departures from item fit expectation were relatively small but when considered in combination, resulted in some item removal. PSI values indicated good to reasonable reliability for scales distinguishing between responders on each scale variable.

Items considered clinically important but that did not fit into existing scales were retained as single items (e.g. itchiness). Scales that did not meet requirements for reliable and valid measurement were either conceptually combined (e.g. items representing mood, anxiety, autonomy and isolation were combined into an emotional well-being scale) or items added (e.g. three items were added to the vitality and malaise scale to produce a six item scale).

The Rasch analysis detected important limitations of the PU-QOLI scales. It detected that the four-category item scoring function did not work as intended for multiple items within each PU-QOLI scale. For the other items where the response categories were working as intended, inspecting CPCs indicated thresholds (τ1 and τ2) were close to being disordered, suggesting that people had difficulty distinguishing between ‘a little bother’ and ‘quite a bit of bother’ categories. This provided good evidence that all items would benefit from fewer response categories. Consequently, all scale items were subjected to a post hoc rescoring by collapsing adjacent categories (so that all items had three response categories). Re-analysis of the data demonstrated that all thresholds were now correctly ordered, producing scales with new categories (‘not b/c of PU’ = treated as missing (descriptive); scale responses are: 0 = no bother, 1 = little bother, 2 = lot bother).

6.6.26 Traditional psychometric evaluation of Rasch developed PU-QOLI scales

All Rasch produced scales (Table 6.12) underwent a preliminary psychometric evaluation using traditional psychometric tests (see Table 6.1) to evaluate the newly formed scales for data completeness, targeting, scaling assumptions, reliability and validity. The preliminary results of the traditional analysis supported the PU-QOLI scales as reliable and valid measures of PU-symptoms, physical and social functioning,

and psychological well-being. The criteria were satisfied for most psychometric properties evaluated. Briefly, data quality was high (scale scores were computable for 93-99.6% of respondents; Table 6.13) and scaling assumptions were satisfied (mostly similar mean item scores, corrected item-total correlations range = 0.525 - 0.920; Table 6.14). Scale-to-sample targeting was good (scale scores spanned the scale range but were notably skewed for three scales (value outside +/-1.0), mean scores were near the scale mid-point for 6/9 scales, and ceiling effects were negligible, however floor effects exceeded the 15% criterion for 2 of 9 scales; Table 6.13). Internal consistency reliability was high as demonstrated by Cronbach's alpha values (range 0.893 – 0.962; Table 6.14). The item-total correlations, alpha coefficient and homogeneity coefficient (inter-item correlation mean and range; Table 6.14) provide evidence towards the internal construct validity of the PU-QOLI scales.

6.7 Summary of Chapter 6

The sub-study identified a large proportion of self-completed PU-QOLs returned with missing data, resulting with the continuation on only evaluating an administered version. Although preliminary, DIF analysis indicated that administration mode did not impact on the way community patients responded to PU-QOLI items, supporting the equivalence of self-completed and interview-administered versions in community populations. The first field test performed a psychometric evaluation of the preliminary PU-QOLI, producing a 10 scale instrument.

Scale development and item reduction were primarily guided by Rasch measurement methods, which provided a vehicle for the detection of items deviating from model expectations with the intention of improving scale attributes. Final decisions on item inclusion were made according to appraisals of the analyses of the observed data against measurement criteria and clinical relevance, as opposed to examinations carried out singularly or sequentially. The use of mixed-methods was beneficial for providing evidence in selecting the optimal mode of administration for the PU-QOLI. The DIF analysis was an important methodological step for highlighting areas warranting further investigation if pursuing a self-completed version in the future. A preliminary evaluation of the Rasch produced scales using traditional methods supported the PU-QOLI as reliable and valid. The next stage required was quantitative confirmation and psychometric support for the modified final PU-QOLI scales (Appendix 6.1) which was undertaken in a second field test with patients with PUs.

Table 6.11 Rasch summary statistics for the final PU-QOLI scales (following field test 1 analysis)

Scale	Person-Item threshold distribution (Scale-to-sample targeting)		Item fit		Person fit			Chi Square interaction		PSI	
	Sample location range	Item location range	Mean	SD	FitRes range	Mean	SD	Value (df)	p	With ext	No ext
Pain	-4.665 - 4.033	-1.054 - 1.087^	-0.267	1.213	-5.968* - 2.377	-0.707	1.517	(18) 18.414	0.429	0.734	0.605
Exudate	-3.813 - 3.645	-0.709 - 0.697^	-0.281	1.273	-8.722* - 2.168	-0.981	2.070	(8) 6.739	0.565	0.712	0.522
Odour	-6.376* - 3.925	-2.063 - 0.801^	-0.365	1.001	-5.625* - 1.813	-0.686	1.238	(6) 8.435	0.208	0.579	0.560
Sleep	-3.761 - 3.729	-0.751 - 0.349^	-0.431	1.239	-5.353* - 2.470	-1.072	1.858	(12) 17.056	0.148	0.709	0.482
Mobility & Movement	-4.049 - 4.031	-0.502 - 0.532^	-0.435	1.129	-6.362* - 2.315	-1.635	2.358	(18) 7.178	0.988	0.663	0.323
Daily Activities	-3.342 – 3.312	-0.793 – 0.813	0.089	0.910	-13.339* - 1.636	-2.001	4.053	(8) 9.917	0.271	0.438	0.220
Vitality & Malaise	The 3 items did not produce a measurement scale; retain as 2 descriptive items or add items										
Mood	Items combined to produce an emotional well-being scale										
Anxiety	The 3 items did not produce a measurement scale; add items or combine with mood items into emotional well-being scale										
Autonomy	The 3 items did not produce a measurement scale; add items or combine with mood items into emotional well-being scale										
Emotional well-being	-4.866 – 4.506	-1.593 – 2.081^	-0.281	0.803	-7.591* - 2.373	-0.824	1.763	(30) 47.003	0.025	0.797	0.743
Self-consciousness	-3.741 – 3.740	-1.163 – 0.949^	-0.368	0.942	-10.506* - 2.011	-1.015	2.121	(14) 17.124	0.250	0.672	0.556
Isolation	The 4 items did not produce a measurement scale; add items or combine with emotional well-being scale										
Participation	-4.018 – 3.979	-0.750 – 0.677^	-0.054	0.918	-4.889* - 2.089	-0.866	1.809	(9) 4.740	0.856	0.648	0.532

*indicates values outside recommended range (i.e. misfit)
^indicates suboptimal targeting (items do not spread across persons)
SD standard deviation; FitRes fit residual; df degrees of freedom; p probability value; ext extremes cases; PSI person separation index

Table 6.12 Rasch summary statistics for final PU-QOLI scale items

Item Location order	Thresholds	Location	Standard error	FitRes	ChiSqu	Prob	Residual correlation
Pain Scale							
1 Uncomfortable	Ordered	-1.054	0.156	-0.617	2.701	0.259	<0.3
5 Tenderness	Ordered	-1.044	0.151	-1.227	3.668	0.160	<0.3
3 Annoying	Ordered	-0.639	0.145	-1.104	2.515	0.284	<0.3
10 Red raw	Ordered	0.002	0.146	-0.364	1.714	0.424	<0.3
8 Stinging	Ordered	0.240	0.150	-0.381	1.249	0.536	<0.3
11 Burning	Ordered	0.278	0.143	-1.436	3.467	0.177	<0.3
7 Throbbing	Ordered	0.743	0.157	-0.817	0.926	0.629	<0.3
9 Stabbing	Ordered	0.698	0.162	1.904	1.159	0.560	<0.3
6 Tingling	Ordered	1.087	0.165	1.635	1.015	0.602	<0.3
Exudate scale							
1 Weeping	Ordered	-0.709	0.197	-1.484	0.521	0.470	<0.3
5 Staining	Ordered	-0.436	0.178	-0.368	0.687	0.407	<0.3
4 Messy	Ordered	-0.321	0.197	-1.870	3.232	0.072	<0.3
6 Dressing off	Ordered	0.043	0.204	1.963	0.316	0.574	<0.3
2 Running	Ordered	0.060	0.191	-1.248	0.652	0.419	<0.3
3 Sticky	Ordered	0.195	0.199	0.055	0.013	0.910	<0.3
8 Bleeding	Ordered	0.472	0.206	-0.135	0.647	0.421	<0.3
7 Pus	Ordered	0.697	0.229	0.843	0.671	0.413	<0.3
Odour scale							
1 Unpleasant	Ordered	-2.063	0.380	-0.922	0.843	0.359	<0.3
3 Stench or stink	Ordered	-0.066	0.296	-1.673	0.763	0.382	<0.3
4 Pungent	Ordered	0.325	0.328	-0.735	3.861	0.049	<0.3
2 Lingering	Ordered	0.394	0.338	-0.345	1.774	0.183	<0.3
5 Sickening	Ordered	0.608	0.322	0.294	0.704	0.401	<0.3
6 Putrid	Ordered	0.801	0.335	1.190	0.490	0.484	<0.3
Sleep scale							
6 Trouble finding comfortable position	Ordered	-0.751	0.174	0.066	0.456	0.796	<0.3
5 sleep in one position	Ordered	-0.218	0.169	1.096	0.374	0.829	<0.3
4 Not getting amount of sleep needed	Ordered	0.053	0.184	-1.454	5.435	0.066	<0.3
2 Interrupted sleep	Ordered	0.283	0.184	-1.191	0.997	0.608	<0.3
3 Being kept awake	Ordered	0.284	0.185	-1.862	3.688	0.158	<0.3
1 Trouble falling asleep	Ordered	0.349	0.182	0.762	6.106	0.047	<0.3
Mobility scale							
9 walking slowed	Ordered	-0.502	0.244	-0.032	0.494	0.781	<0.38
8 limited ability to walk	Ordered	-0.191	0.227	-0.600	0.975	0.614	<0.45
2 adjusting in bed	Ordered	-0.127	0.191	-1.759	0.431	0.806	<0.31
3Turning/moving in bed	Ordered	-0.115	0.185	-1.420	0.502	0.778	<0.30
4 pushing up to a sitting position	Ordered	-0.055	0.174	-1.757	0.348	0.840	<0.30

Item Location order	Thresholds	Location	Standard error	FitRes	ChiSqu	Prob	Residual correlation
7 limited ability to go up and down stairs	Ordered	0.063	0.275	0.754	1.718	0.424	<0.39
6 standing long periods	Ordered	0.086	0.195	0.488	0.686	0.710	<0.52
1 sitting up in bed	Ordered	0.308	0.190	-0.905	0.780	0.677	<0.32
5 transferring	Ordered	0.532	0.194	1.312	1.245	0.536	<0.39
Daily activities scale							
1 Washing	Ordered	-0.793	0.203	-0.415	0.250	0.617	<0.30
5 Shopping	Ordered	-0.516	0.313	-0.645	0.185	0.668	<0.30
8 Regular activities	Ordered	-0.334	0.257	0.041	1.211	0.271	<0.30
3 Toileting	Ordered	-0.153	0.214	0.303	0.222	0.637	<0.30
4 Jobs around house	Ordered	-0.023	0.290	-0.718	3.274	0.070	<0.30
2 Dressing	Ordered	0.372	0.237	-0.725	2.395	0.122	<0.30
6 Doing things enjoy	Ordered	0.633	0.235	1.329	0.156	0.693	<0.30
7 Emotionally close with loved ones	Ordered	0.813	0.268	1.541	2.224	0.136	<0.30
Emotional well-being scale							
2 annoyed	Ordered	-1.593	0.149	-1.309	7.302	0.026	<0.30
1 frustrated	Ordered	-1.196	0.148	-1.159	3.706	0.157	<0.31
3 angry	Ordered	-0.792	0.142	-0.143	7.735	0.021	<0.33
5 depressed	Ordered	-0.602	0.149	-1.806	1.854	0.396	<0.30
11 physically dependent	Ordered	-0.432	0.171	0.553	2.912	0.233	<0.30
7 concerned/worried	Ordered	-0.269	0.149	0.390	3.848	0.146	<0.30
8 anxious	Ordered	-0.170	0.146	0.140	2.334	0.311	<0.36
6 upset	Ordered	0.004	0.150	-0.002	1.432	0.489	<0.30
10 no control	Ordered	0.030	0.163	-0.742	0.164	0.921	<0.30
9 burden/nuisance	Ordered	0.031	0.173	-0.792	0.751	0.687	<0.30
4 miserable	Ordered	0.262	0.148	-0.827	1.613	0.447	<0.30
14 missing out	Ordered	0.661	0.177	-0.077	3.614	0.164	<0.41
13 lonely	Ordered	0.838	0.187	1.107	3.881	0.144	<0.30
12 cut off/isolated	Ordered	1.147	0.184	0.536	3.691	0.158	<0.30
15 others avoided/ treated you differently	Ordered	2.081	0.251	-0.087	2.166	0.339	<0.30
Self-consciousness scale							
1 helpless	Ordered	-1.163	0.170	0.677	2.740	0.254	<0.30
2 lacking confidence	Ordered	-0.519	0.171	-0.387	0.734	0.693	<0.30
3 self-conscious	Ordered	-0.300	0.176	-0.108	3.646	0.162	<0.30
4 embarrassed	Ordered	-0.125	0.178	-0.807	2.348	0.309	<0.30
5 physically unattractive	Ordered	0.425	0.200	-1.932	4.621	0.099	<0.30
6 uneasy being close to others	Ordered	0.734	0.216	-0.807	1.782	0.410	<0.30
7 lack understanding from others	Ordered	0.949	0.226	0.784	1.254	0.534	<0.30

Item Location order	Thresholds	Location	Standard error	FitRes	ChiSqu	Prob	Residual correlation
Participation scale							
6 restricted where you could go out	Ordered	-0.750	0.242	-0.989	3.041	0.219	<0.30
8 unable to get away for holiday or make a trip at the weekend	Ordered	-0.554	0.224	-0.726	5.303	0.071	<0.30
5 give up on hobbies or leisure activities	Ordered	-0.343	0.243	-1.502	2.644	0.267	<0.30
1 Difficulty going out	Ordered	-0.307	0.240	0.221	0.441	0.802	<0.30
7 Being restricted to how long you could stay out	Ordered	0.144	0.234	-0.011	0.919	0.632	<0.30
4 Having to plan going out around pressure sore care	Ordered	0.262	0.242	1.267	1.240	0.538	<0.30
9 time involved in caring for PU	Ordered	0.379	0.226	0.447	0.295	0.863	<0.30
3 unable to participate in family gatherings	Ordered	0.492	0.248	-0.240	0.846	0.655	<0.30
2 meeting/seeing family and/or friends	Ordered	0.677	0.237	1.046	4.757	0.093	<0.30

*indicates values outside recommended range (i.e. misfit)

^indicates suboptimal targeting (items do not spread across persons)

+no DIF detected at Bonferroni adjusted level

SD standard deviation; FitRes fit residual; ChiSqu chi square; df degrees of freedom; p probability value; ext extremes cases

Table 6.13 PU-QOLI Scale level analyses – Data completeness and Targeting (n=227)

		Targeting						
Scale	**Data completeness -** Computable scale score (%)	**Possible score range∗**	**Range mid-point**	**Observed score range**	**Mean score**	**SD**	**F/C effect (%)**	**Skewness**
Pain	96.5	0 – 18	9	0 – 18	7.39	4.732	3.5/1.8	0.342
Exudate	96.5	0 – 16	8	0 – 16	3.43	4.290	17.6/1.3	1.224
Odour	99.1	0 – 12	6	0 – 12	2.27	3.760	26.9/2.6	1.560
Sleep	99.6	0 – 12	6	0 – 12	4.68	3.816	9.7/4.4	0.409
Malaise	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Mobility	96.9	0 – 18	9	0 – 18	8.39	5.659	1.3/1.8	0.265
ADL	97.4	0 – 16	8	0 – 16	7.70	6.131	1.8/1.3	-0.032
Emotional Well-being	93.0	0 - 30	15	0 - 30	11.57	8.007	2.6/0.9	0.450
Mood	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Anxiety	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Appearance & self-consciousness	95.2	0 - 14	7	0 - 14	3.45	3.777	11.5/1.8	1.301
Autonomy & independence	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Isolation	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Participation	96.0	0 - 18	9	0 - 18	8.93	6.341	2.6/3.1	0.018

∗High scores indicate great bother/impact

SD standard deviation; F/C floor/ceiling – floor effect =% scoring 100 (greatest bother/impact); ceiling effect = % scoring 0 (least bother/impact)

Table 6.14 PU-QOLI Scale level analyses - Reliability and scaling assumptions: Validity within-scale analysis (n=227)

Scale	**Internal consistency** - Cronbach's alpha	**SEM**	**95% CI**	**Mean IIC (n=227)**	**Range IIC**	**Scaling Assumptions -** Corrected ITC
Pain	0.893	0.453	6.50, 8.29	0.482	0.235 – 0.663	0.525 – 0.703^
Exudate	0.919	0.445	2.55, 4.31	0.582	0.397 – 0.857	0.563 – 0.836^
Odour	0.962	0.378	1.52, 3.02	0.808	0.735 – 0.908	0.825 – 0.920^
Sleep	0.920	0.361	3.96, 5.39	0.655	0.480 – 0.844	0.667 – 0.858^
Vitality	n/a	n/a	n/a	n/a	n/a	n/a
Mobility	0.927	1.069	6.20, 10.59	0.586	0.226 – 0.912	0.666 – 0.799^
ADL	0.952	1.3741	4.83, 10.57	0.710	0.407 – 0.904	0.583 – 0.899^
Emotional Well-being	0.934	0.931	9.71, 13.42	0.486	0.242 – 0.789	0.537 – 0.761^
Mood	n/a	n/a	n/a	n/a	n/a	n/a
Anxiety	n/a	n/a	n/a	n/a	n/a	n/a
Appearance & self-consciousness	0.901	0.372	2.71, 4.18	0.569	0.409 – 0.750	0.598 – 0.792^
Autonomy & independence	n/a	n/a	n/a	n/a	n/a	n/a
Isolation	n/a	n/a	n/a	n/a	n/a	n/a
Participation	0.957	0.979	6.95, 10.90	0.710	0.526 – 0.887	0.733 – 0.900^

SEM standard error mean; CI confidence interval; IIC inter-item correlation
^Range item-total correlation (ITC)

Chapter 7

FIELD TEST 2 - PSYCHOMETRIC EVALUATION OF THE FINAL PU-QOLIV2

7.1 Overview

Chapter 7 presents the methods and results of the second of two quantitative field tests undertaken to evaluate the measurement properties of the final PU-QOLI scales developed in Chapter 6. Consistent with methods undertaken in the previous Chapter, a Rasch analysis was performed first on all PU-QOLI scales, followed by traditional psychometrics.

7.2 Objectives

The purpose of the final stage of psychometric analysis was to carry out a comprehensive psychometric evaluation of the final (10 scale/83-item) PU-QOLIv2, using Rasch analysis and traditional psychometric methods, in a large independent sample of patients with PUs.

7.3 Methods

7.3.1 Participants

Field test 1 sample

A sample of 200-250 patients with PUs were purposively sampled following the methods described in section 6.4.2.

7.3.2 Sample size

To ensure an appropriate degree of precision, a sample of 200-250 patients was sought (see section 6.4.3). A sample size of around 250 would provide sufficient subjects for analyses of test-retest reliability; correlations at the levels expected in test-retest situations (e.g. r >= 0.80) can be estimated with reasonable precision (95% confidence intervals of ±0.1) with relatively few subjects (328, 329).

7.3.3 Eligibility

The eligibility criteria for field test 2 were consistent with the criteria detailed in section 6.4.4.

7.3.4 Recruitment and Consent

The recruitment and consent procedure was consistent with methods detailed in section 6.4.5.

7.3.5 Registration

Screened patients eligible for field test 2 participation were registered to the study via an automated telephone registration system. Upon registration, a study number was issued. Following registration, baseline data was collected and PU-QOLIs completed.

7.3.6 Data collection/assessment

Study data (e.g. registration, baseline clinical data, PU-QOLI) were recorded on CRFs by members of TVTs or CG. A questionnaire pack was administered to all participants and training in administering the PU-QOLI was provided by CG. A subsample of participants who completed a PU-QOLI at baseline also completed a second PU-QOLI 2-7 days after the first adminstration to evaluate test-retest reliability. The length of the test-retest interval had to be short enough to ensure that clinical change in the PU was unlikely to occur, but sufficiently long to ensure that respondents did not recall their responses during the first administration; a short test-retest interval is necessary to ensure that stability per se is evaluated, rather than clinical change in the PU.

7.3.7 Monitoring and data cleaning

The methods for data monitoring and data cleaning are consistent with the methods described in sections 6.4.8 and 6.4.9, respectively.

7.4 Analysis Plan for Field Test 2

7.4.1 Baseline characteristics of sample

Baseline analysis took the form of methods described in section 6.5.1.

7.4.2 Rasch analyses

Methods for data analysis were consistent with those undertaken during the first field test. However, in addition to Rasch statistics described in sections 6.5.3, additional questions were considered during the final psychometric evaluation, as follows:

Person fit

How valid is each person's measurement? As with item fit residuals, person fit residuals summarise the extent to which individual responses are consistent with those

expected by the Rasch model, and can be used to identify misfitting individuals. Person residuals are produced by subtracting expected scores from observed scores; predicted scores are calculated from the Rasch model using the estimates of the person and item locations. The residual is standardised by dividing it by the square root of the variance (computed from the expected value (EV) using the formula: variance = EV – EV2 (330)), producing a standardised fit residual for each person's response to each item. These are then transformed and summarised to form the person fit residual which approximates a standard normal deviate. Residuals between ±2.5 indicate adequate person fit to the model (85).

Person-fit was examined to identify any persons with high positive residuals. Persons who deviate from model expectations may seriously affect fit at item level. In terms of scale validation, this runs the risk of discarding the scale/item when it would be more appropriate to consider why those persons may have responded differently to everyone else (e.g. because of unrecorded comorbidity or cognitive impairment). Removal of these persons from the analysis may make a significant difference to the scales internal construct validity, while at the same time raising questions about the external construct validity of the scale with the particular patient group.

DIF

Are the locations of the items stable across clinically important groups? The extent to which item locations were stable across multiple groups was examined by a DIF analysis (see section 6.5.2, *DIF*) in relation to four clinical subgroupings: age (under 70 years and 70 years over), gender (male and female), PU location (torso, limb, both), and healthcare setting (hospital and community). Both uniform and non-uniform DIF were examined.

Are PU-QOLI scale scores stable over a given period of time in which the respondent's condition is assumed to have remained unchanged? A DIF analysis was undertaken to examine agreement between total scores at two time points (test-retest reproducibility).

Validity

Unlike traditional methods, a Rasch analysis does not examine correlations with other measures or hypothesis testing. With a Rasch analysis, the clinical validating comes from conceptualisation of the instruments' scales. Specifically, careful item construction and consideration of what each item is meant to measure, and then test them against model expectations. Then, clinical questions are formulated as external validations (see Section 7.4.3, *Differentiate known groups*) and the empirical testing

comes from the Rasch model. For example, people who have severe PUs might be expected to score higher on the symptom scales compared to those with superficial PUs. The placements (responses) of individuals on each symptom scale continuum mapped out by symptom items should be largely consistent with expectation.

7.4.3 Traditional psychometric analyses

Consistent with methods undertaken during the first field test, final Rasch scales underwent a psychometric evaluation using the standard psychometric tests detailed in sections 6.5.4, as well as additional psychometric tests described below, for: acceptability (floor/ceiling effects and skew of scale scores), reliability (internal consistency and item-total correlations; test re-test) and validity (convergent, discriminant validity and known groups validity).

Reliability

In addition to reliability statistics described in section 6.5.4, test retest (TRT) reproducibility was determined. A Pearson correlation was performed to investigate the strength of the relationship between scores at time 1 (baseline) and time 2 (retest 2-7 days post baseline administration). This type of test indicates whether scores at the two time point are statistically significantly correlated (criterion $\geq$0.7). High correlations indicate a more reliable scale.

Validity

Validity testing for the final field test includes both within- and between-scales testing.

- *Construct*

Construct validity is assessed, instead of criterion-related validity, when no gold standard exists with which to compare a measure. In the absence of such a validating measure, a within-scale construct validity analysis was undertaken to determine the extent to which PU-QOLI scales measure a single entity and whether items can be combined to form scale scores. Other types of construct validity were evaluated by examining ITC (>0.3), alpha coefficient (>0.7), homogeneity coefficient (inter-item correlation mean and range; >0.3); indicators that a single construct is being measured, and that items can be combined into a scale, and correlations with external criteria (e.g. convergent validity), providing evidence towards the internal construct validity of the PU-QOLI scales.

- *Criterion*

Criterion validity is a special type of construct validity in which stronger hypotheses are made possible by the availability of a criterion or 'gold standard' measure. There are no true gold standard HRQL measures (245) and no PU-specific or chronic wound–specific measures available (see Chapter 4).

- *Convergent validity*

Convergent validity - the degree to which constructs (or scores on a measure) expected to be related are, in fact, related (scores correlate with scores on other measures designed to assess the same construct) (331) - was examined by computing correlations between the SF12 (332) and PU-QOLI scales as well as overall QoL and pain questions. The SF-12v2 Acute, English (UK) version was used (333) to minimise respondent burden. It is a generic measure that asks respondents to rate their health and functioning during the past week on eight domains: physical functioning, role physical, bodily pain, general health, energy/fatigue, social functioning, role emotional and mental health. Two summary scores can be produced: physical (PCS) and mental (MCS) component scores. Most are composite scores consisting of two questions with the exception of four single item indicators for social function, energy/fatigue, pain and general health. Higher scores indicate good HRQL.

A Spearman rank-order correlation coefficient (used with linear or ordinal level data) was calculated between PU-QOLI scales and the Physical and Mental Health Composite Scale Scores of the SF-12 to determine how closely PU-QOLI scales were related to another instrument of similar constructs. For exploratory purposes, the following hypotheses were proposed based on the proximity of the constructs; criteria were used as guides to the magnitude of correlations, as opposed to pass/fail benchmarks (high correlation $r > 0.7$; moderate correlation $r = 0.3 - 0.7$; low correlation < 0.3) (193, 334):

- The pain scale would positively correlate with the SF12 bodily pain item
- The pain scale would positively correlate ($r = > 0.7$) with the PU-QOLI overall pain item
- The vitality scale would correlate positively with the SF12 energy/fatigue item
- The movement and mobility scale would correlate negatively with the SF12 physical functioning scale
- The emotional well-being scale would correlate negatively with the SF12 mental health scale
- The participation scale would correlate negatively with the SF12 social function item

- All PU-QOLI scales would correlate positively with the PU-QOLI overall QoL question

Moderate to high correlations (r= > 0.3) were predicted. The direction of the relationship was also indicted as high SF12 scale scores indicate better outcome, whereas high PU-QOLI scale scores indicate worse outcome. The SF12 pain and fatigue items are negatively scored therefore scoring is consistent with PU-QOLI scores.

- *Discriminant validity*

Discriminant validity (or divergent validity) – the degree to which constructs expected to not be related (have no relationship) are, in fact, not related (or the ability to discriminate between known groups (331)) - was examined by computing Pearson correlation coefficients (used with interval level data) between PU-QOLI scales and age and gender to determine the extent to which responses were biased by these variables. Low (<0.3) correlations were predicted for gender and age.

- *Differentiate known groups hypothesis testing*

Known-group comparisons are used to evaluate the clinical utility of PRO instruments. This method assesses the extent to which scales are able to discriminate between subgroups of patients known to differ in terms of clinical presentations (335). The HRQL and PU literature is not well established and therefore limited for identifying clinical parameters to formulate known groups. A one-way analysis of variance (AVOVA) (used where there are three or more groups) was used to test for statistically significant differences in mean scores between PU severity subgroups and PU-QOLI exudate and odour scales. PU severity groups were classified as category 1, category 2 and category 3 and 4 combined due to the small sample sizes of severe PUs. A stepwise change in PU-QOLI scores across PU severity groups was predicted, and that mean scores would be significantly different (this is based on clinical expectation; people with superficial PUs do not have problems with exudate or odour). For the remainder of scales, known-group difference was investigated on an exploratory basis, predicting a stepwise change in PU-QOLI scores across PU severity groups.

7.5 Results

7.5.1 Study sample

A total 979 patients were screened for study participation of whom 391 were assessed as eligible and 231 eligible patients consented and were registered to the study (see Figure 7.1 for the flow of participants). Respondents represented a wide range of age

groups. There was a slightly higher proportion of men, and a small percentage of respondents from minority ethnic groups. Approximately 60% of the sample were hospitalised patients and 45% at on-going long-term risk of PUs. Over two third of the sample was married or cohabiting and just under half the sample had some educational qualification. Table 7.1 shows patient characteristics for the final field test.

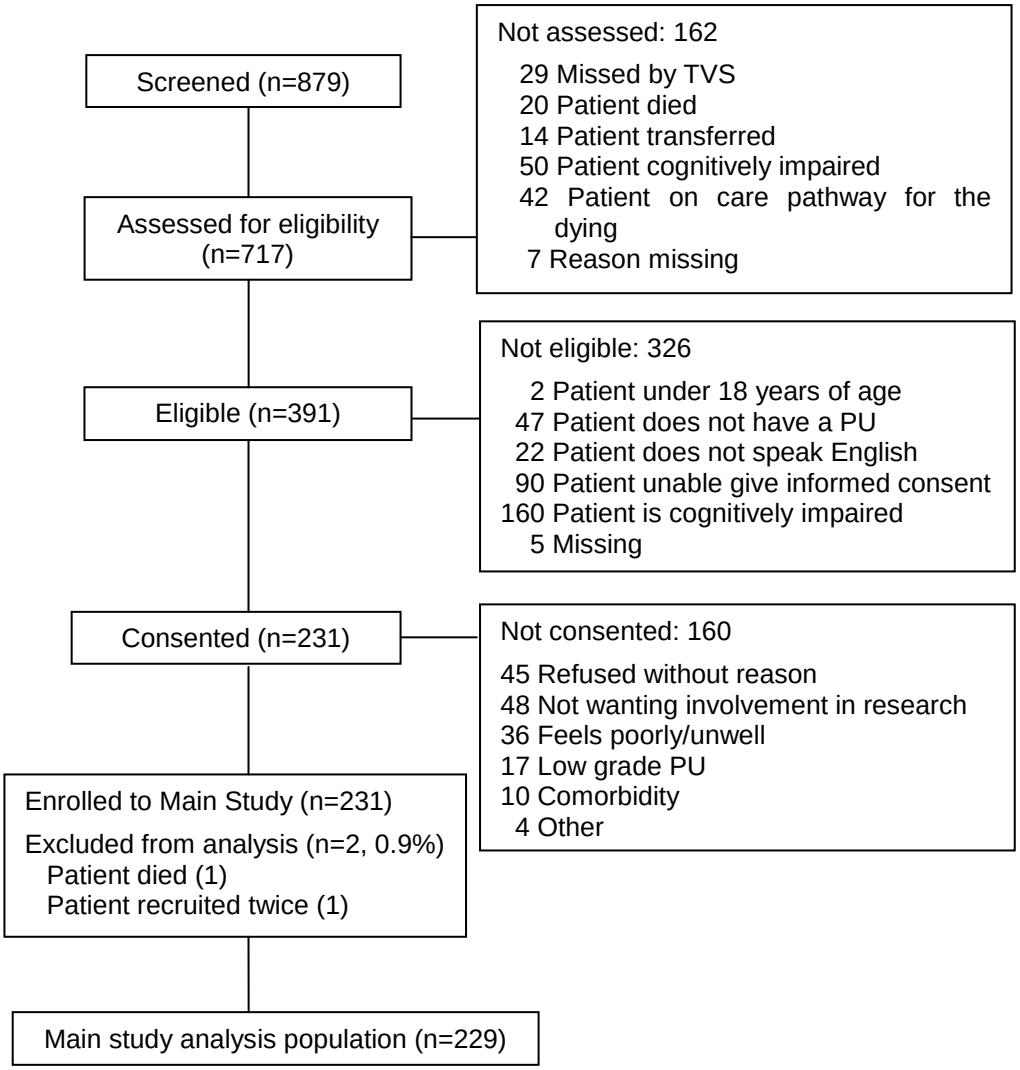


Figure 7.1 Assessment flow chart to field test 2

7.5.2 Rasch analyses results

The analyses described in section 6.5.3 were performed on all PU-QOLI scales independently. A summary of the Rasch results will be presented for each scale, followed by decisions made to any scale modifications, and finally presentation of the final psychometric properties in summary tables.

Table 7.1 Respondent Characteristics (n=229)

Characteristics	Range years (Mean, SD)
Age	20 to 103 (71.3, 16.5)
	Total n (%)
Gender	
Male	119 (52.0)
Female	110 (48.0)
Ethnicity	
White	227 (99.1)
Asian	2 (0.9)
Black/African	0
Chinese	0
Setting	
Hospital (surgery)	62 (27.1)
Hospital (medicine)	74 32.3)
Community	88 (38.4)
PU severity	
Category 1	76 (18.1%)
Category 2	170 (40.5%)
Category 3/4	170 (40.5%)
Missing	4 (0.9%)
PU risk classification	
Short-term	36 (15.7)
New medium to long-term	87 (38.0)
On-going long-term	103 (45.0)
Missing	3 (1.3)
Marital status	
Single (includes divorced, separated, widowed)	71 (31.0)
Married	77 (33.6)
Cohabiting	75 (32.8)
Missing	6 (2.6)
Living arrangements	
Live alone	86 (37.6)
Cohabit with identified carer	51 (22.3)
Cohabit with other	48 (20.9)
Missing	44 (19.2)
Education	
No formal education	125 (54.6)
GCSE or equivalent	40 (17.5)
A-Level or equivalent	16 (6.9)
Degree or higher	21 (9.2)
Missing	27 (11.8)

7.5.3 Summary of Rasch findings for the pain scale

The pain scale items mostly met Rasch measurement criteria. The item 'tingling' had a high chi square value (19.327) and a significant p value (0.000). As the scale had good reliability (PSI =0.823), suggesting confidence in the accuracy of the fit statistics, the item 'tingling' was removed post hoc. Removing the item resulted in a good brief descriptive pain scale. The PSI was slightly lower (Table 7.2) but still represented good reliability. The ordering of items along the variable was clinically sensible; and the fit statistics indicated that all pain items fit the expectations of the measurement model. There was no evidence to suggest that responses to pain items were biased by responses to another; and no evidence of significant DIF across clinically important groups.

Item and person location ranges (Table 7.2) indicate that the sample is good for examining the psychometric properties of the 8-item pain scale but a suboptimal scale for measuring the sample, with no items representing people at the extremes of the sample distribution. A total of 229 people were measured in the sample. Of these, 44 scored at the floor (n = 35 lowest score) or ceiling (n = 9 high scores) and did not have person–item fit residual estimates. Of the 185 people who had estimates of their person–item fit residuals, 9 people (4.9%) had values that were outside the range +/-2.5. These findings indicate that 95.1% of people gave responses that were consistent with expectation.

7.5.4 Summary of Rasch findings for the exudate scale

The sample for examining the psychometric properties of the 8-item exudate scale was small (n=95) due to many extreme persons (n=133); the largest frequency of patients was at the ceiling of the range, scoring the least bother with exudate. In this instance, the construct is not very wide and people with only superficial PUs do not have problems with exudate (exudate is associated with skin breakdown which is consistent with severe PUs) therefore targeting is within the expected context. Of the sample, 59 people had person–item fit residual estimates and all values inside the range +/-2.5.

Patients with only superficial PUs were removed from the analysis, the Rasch analysis reran, and findings compared with the findings from the complete sample. Item and person location ranges (Table 7.2) indicated that the sample was good for examining the psychometric properties of the 8-item exudate scale but a suboptimal scale for measuring the sample, with no items representing people at the extremes of the sample distribution. Removing superficial patients resulted in item 'dressing coming off' (item 4) with reversed thresholds. It would make clinical sense to dichotomise this

item (the dressing is either coming off or it is not). However, due to the very small sample (n=59), additional empirical evidence is needed to be confident about any changes made to the exudate scale. The final exudate scale had reasonable reliability (PSI =0.598); the ordering of items along the variable was clinically sensible; the fit statistics indicated that all exudate items fit expectations of the measurement model; there was no evidence to suggest that item responses were biased by responses to another; and no evidence of significant DIF across clinically important groups.

7.5.5 Summary of Rasch findings for the odour scale

Similar to the exudate scale, the sample for examining the psychometric properties of the 6-item odour scale was very small (n=28) due to many extreme persons (n=201); the largest frequency of patients was at the ceiling of the range, scoring the least bother with odour. In this instance, the construct is not very wide and people with only superficial PUs do not have problems with odour (odour is associated with skin breakdown which is consistent with severe PUs); therefore the targeting is within the expected context. Of the sample, 28 people had estimates of their person–item fit residuals; all people had values inside the range +/-2.5).

Patients with only superficial PUs were removed from the analysis, the Rasch analyses reran, and findings compared with the findings from the complete sample. Item and person location ranges (Table 7.2) indicated that the sample was good for examining the psychometric properties of the 6-item odour scale but a suboptimal scale for measuring the sample, with no items representing people at the extremes of the sample distribution. Removing superficial patients resulted in a 21 patient sample, a sample considered too small to make reliable inferences about the statistics however, removing superficial people improved the PSI from 0.321 to 0.486.

The final odour scale had reasonable reliability; the ordering of items along the variable was clinically sensible; the fit statistics indicated that all odour items fit the expectations of the measurement model; and there was no evidence of significant DIF across clinically important groups. However, ‘stench’ and ‘sickening’ (items 2 and 5) had a high negative correlation (-0.65), indicating that one of the items might be redundant but further empirical evidence is needed due to the very small sample size.

7.5.6 Summary of Rasch findings for the sleep scale

The sleep scale items mostly met Rasch measurement criteria, therefore no changes were made to the items or response categories. Two items failed on numerical tests of

fit; item ‘sleep in one position’ (item 2) had a fit residual of 2.639, but this value is only marginally outside the recommended range, and item ‘not getting amount of sleep needed’ (item 3) had a significant chi square value (0.0013) at the Bonferroni level (set at 0.001667), but again this is only marginally below the significant p value.

The sleep scale had good reliability (PSI =0.719); the ordering of items along the variable was clinically sensible; there was no evidence to suggest that responses to any sleep items were biased by responses to another; and no evidence of significant DIF across clinically important groups. Item and person location ranges (Table 7.2) indicate that the sample is good for examining the psychometric properties of the sleep scale but a suboptimal scale for measuring the sample, with no items representing people at the extremes of the sample distribution. A total of 229 people were measured in the sample, of which 90 scored at the floor (lowest score) and did not have person–item fit residual estimates. Of the 139 people who had person–item fit residual estimates, 7 people (5.0%) had values outside the range +/-2.5, indicating that 95.0% of people gave responses that were consistent with expectation.

7.5.7 Summary of Rasch findings for the movement and mobility scale

Item and person location ranges (Table 7.2) indicate that the sample is good for examining the psychometric properties of the 9-item movement and mobility scale but a suboptimal scale for measuring the sample, with no items representing people at the extremes of the sample distribution. The three-category item scoring function for the scale was inadequate for two items, however the frequency of category endorsements was <10 for both items. The ordering of the nine items along the variable was clinically sensible and the items mapped out a variable of increasing intensity, but only covering a 2 logit spread. The range of item thresholds and locations was narrow, and there were gaps in the continuum.

The fit statistics indicated that all mobility items fit the expectations of the measurement model. However, there was evidence to suggest that responses to some mobility items were biased by responses to another; and there was evidence of significant DIF for two items by healthcare setting (Table 7.3). The PSI was low (Table 7.2) and the sample was small (n=130), with many persons with extreme scores; scoring at the floor or ceiling and did not have person–item fit residual estimates. As such, there was little confidence in making significant changes to the mobility scale, without additional empirical evidence. There may be some benefit in revisiting transcripts to possibly add items to cover the full measurement range.

7.5.8 Summary of Rasch findings for the activity scale

Item and person location ranges (Table 7.2) indicate that the sample is good for examining the psychometric properties of the 8-item daily activities scale but a suboptimal scale for measuring the sample, with no items representing people at the extremes of the sample distribution. The three-category item scoring function for the scale was inadequate for four items, however the frequency of category endorsements was <10 for these items. The ordering of the eight items along the variable was clinically sensible but the range of item thresholds and locations was narrow, with items with similar locations and gaps in the continuum (Table 7.3).

The fit statistics indicated that all daily activity items fit the expectations of the measurement model; there was no evidence to suggest that responses to any items were biased by responses to another; and there was no evidence of significant DIF across clinically important groups (Table 7.3). However, the PSI was low (Table 7.2) and the sample small (n=117), with many persons scoring at the floor or ceiling and therefore did not have person–item fit residual estimates. As such, there was little confidence in making significant changes to the daily activities scale, without additional empirical evidence. There may be some benefit in revisiting transcripts to possibly add items to cover the full measurement range.

7.5.9 Summary of Rasch findings for the vitality scale

Item and person location ranges (Table 7.2) indicate that the sample is good for examining the psychometric properties of the 5-item vitality scale but a suboptimal scale for measuring the sample, with no items representing people at the extremes of the sample distribution. The ordering of the five items along the variable was clinically sensible and the items mapped out a variable of increasing intensity, but only covering a 1.3 logit spread. The range of item thresholds and locations was narrow, and there were gaps in the continuum.

The fit statistics indicate that all vitality items fit the expectations of the measurement model and the three-category item scoring function was adequate for all items (Table 7.3). However, items ‘tired’ and ‘fatigued’ had similar locations and both correlated at 0.49, suggesting responses to these items were biased by responses to each other, and there was evidence of significant DIF for item ‘feeling unwell’ by age (Table 7.3).

The PSI was low (Table 7.2) and the sample was small (n=98), with many persons with extreme scores; scoring at the floor or ceiling and did not have person–item fit residual estimates. As such, there was little confidence in making significant changes to the

vitality scale, without additional empirical evidence. There may be some benefit in revisiting transcripts to possibly add items to cover the full measurement range.

7.5.10 Summary of Rasch findings for the emotional well-being scale

The emotional well-being scale items mostly met Rasch measurement criteria however the three-category item scoring function for the scale was inadequate for two items. These two items were dichotomised and the Rasch analyses reran. This resulted in a good emotional well-being scale, with good reliability (PSI =0.846) (Table 7.2). The ordering of the emotional well-being items along the variable was clinically sensible; the fit statistics indicated that all items fit the expectations of the measurement model; and there was no evidence of significant DIF across clinically important groups. However, there was some evidence to suggest that responses to some emotional well-being items might be biased by responses to another (Table 7.3).

Item and person location ranges (Table 7.2) indicate that the sample is good for examining the psychometric properties of the 15-item emotional well-being scale but a suboptimal scale for measuring the sample, with no items representing people at the extremes of the sample distribution. A total of 229 people were measured in the sample. Of these, 43 scored at the floor (lowest score) and did not have person–item fit residual estimates. Of the 186 people who had person–item fit residual estimates, 64 people (34.4%) had values outside the range +/-2.5, and 49 (26.3%) outside the range +/-3.0, indicating that 65.6% of people gave responses consistent with expectation.

7.5.11 Summary of Rasch findings for the self-consciousness scale

The ordering of the self-consciousness items along the variable was clinically sensible; the fit statistics indicated that all items fit the expectations of the measurement model; the three-category item scoring function was adequate for all items; and there was no evidence of significant DIF across clinically important groups. However, there was some evidence to suggest that responses to items ‘self-conscious’ and ‘embarrassed’ were biased by each other (Table 7.3) and the PSI was low (Table 7.2).

Item and person location ranges (Table 7.2) indicate that the sample is good for examining the psychometric properties of the 7-item self-consciousness scale but a suboptimal scale for measuring the sample, with no items representing people at the extremes of the sample distribution. A total of 229 people were measured in the sample. Of these, 89 scored at the floor (lowest score) and did not have person–item fit residual estimates. Of the 140 people who had estimates of their person–item fit

residuals, 18 people (12.9%) had values that were outside the range +/-2.5, indicating that 87.1% of people gave responses that were consistent with expectation.

7.5.12 Summary of Rasch findings for the participation scale

Item and person location ranges (Table 7.2) indicate that the sample is good for examining the psychometric properties of the 9-item participation scale but a suboptimal scale for measuring the sample (no items representing people at the extremes of the sample distribution). Item ordering along the variable was clinically sensible but item threshold and location ranges were narrow, with items at similar locations and gaps in the continuum (Table 7.3). Fit statistics indicate all participation items fit expectations of the measurement model, and there was no evidence of significant DIF across groups. However, some items had high residual correlations (Table 7.3); suggesting responses to the items were biased by responses to another.

The three-category item scoring function for the scale was inadequate for seven items however the frequency of category endorsements was low for these items. The PSI was low (Table 7.2) and the sample small (n=82), with many persons scored at the floor or ceiling and therefore did not have person–item fit residual estimates. As such, there was little confidence in making significant changes to the participation scale, without additional empirical evidence.

A total of 229 people were measured in the sample. Of these, 117 scored at the floor or ceiling and did not have person–item fit residual estimates. Of the 112 people who had estimates of their person–item fit residuals, 36 people (32.1%) had values that were outside the range +/-2.5, indicating that 67.9% of people gave responses that were consistent with expectation. There may be some benefit in revisiting transcripts to possibly add items to cover the full measurement range.

Table 7.2 provides a summary of the Rasch statistics for the final PU-QOLI scales. Table 7.3 provides a summary of the Rasch statistics for all items in item location order, indicating the line of increasing intensity for the final PU-QOLI scales.

7.5.13 Summary of Rasch findings

The measurement properties of the PU-QOLIs' scales were largely supported as demonstrated through items that mapped out continua of increasing intensity and are located along those continua in a clinically sensible order. Scale items work well together to define single variables, albeit, some item misfit, local dependence and

items exhibiting DIF detected. For example, DIF was demonstrated in three items (Table 7.3), however the deviations from model expectations were marginal, suggesting item performance across the four clinical subgroups is stable and that these groups can be measured on a common ruler.

The Rasch analysis detected important limitations of some PU-QOLI scales. It detected that the three-category item scoring function did not work as intended for some scale items. This would imply that these items had too many response options and required a post hoc rescoring of items by collapsing adjacent categories. However the frequency of category endorsements was <10 for these items (336) therefore further empiric evidence might be required. Some item locations indicated areas on the continuum within the scale range measured where the measurement could be improved (i.e. at extreme ends of the scale range). As the sample sizes for these scales were quite small, major modifications to items and the scoring function were demeaned premature without additional empirical evidence.

Another limitation pertains to the sample distribution. For most scales, the sample was not normally distributed (normal distribution is neither expected nor wanted as sample distribution is an empirical finding rather than a requirement, but it does suggest that assumptions about the distribution of people and the variables measured in populations should not be made (85)). The largest frequency of patients was often at the floor of scale ranges (least bother), suggesting suboptimal targeting of the PU-QOLI scales to the study sample. Ideally there should be a good match between the scale range and sample range, with people falling within the range of the items. For the symptom scales, the targeting can be justified as not all patients with PUs are expected to have problems with symptoms so it is clinically reasonable that these people would fall outside the scale range. Importantly, where people have symptom bother, there needs to be items within the scales that will discriminate symptom bother, and in this instance, the symptom scales perform this function. However, other scales were unable to separate the sample reliably, and people's patterns of responses were inconsistent with expectation as indicated by many misfitting persons in some instances.

7.6 Traditional psychometric evaluation of final PU-QOLI scales

The final PU-QOLI scales underwent a preliminary psychometric evaluation using traditional psychometric tests (see Table 6.1). Results of the preliminary analysis for: data completeness, scaling assumptions, targeting, internal consistency reliability and construct validity are presented in Tables 7.4 to 7.6 and briefly described below.

Table 7.2 Field Test 2 Rasch analysis summary statistics for final PU-QOLI scales

Scale	Person-Item threshold distribution (Scale-to-sample targeting)		Item fit residual		Person fit residual			Chi Square interaction		PSI	
	Sample location range	Item location range	Mean	SD	FitRes range	Mean	SD	Value (df)	p	With ext	No ext
Pain	-4.017 - 3.232	-1.104 - 1.024^	-0.235	0.812	-2.722 - 2.339	-0.216	1.049	(24) 33.613	0.092	0.814	0.721
Exudate	-3.156 - 1.740	-0.751 - 0.843^	0.020	0.908	-3.869 - 1.740	-0.156	1.100	(8)10.503	0.231	0.598	0.688
Odour	-5.070 - 2.869	-1.303 - 0.906^	-0.169	0.587	-1.742 - 1.809	-0.171	0.778	(6) 7.988	0.239	0.486	0.655
Sleep	-3.076 - 2.860	-0.907 - 0.451^	0.071	1.805	-4.096 – 2.424	-0.317	1.277	(12) 39.710	0.000	0.719	0.616
Movement & Mobility	-3.011 - 2.686	-0.457 - 0.572^	0.196	0.934	-16.062 – 1.970	-1.295	3.539	(18) 27.855	0.064	0.505	0.422
Activities	-2.418 - 2.150	-0.299 - 0.561^	0.264	1.028	-8.209 – 1.790	-0.261	1.381	(8) 13.604	0.093	0.102	0.268
Vitality	-3.014 - 2.672	-0.500 - 0.804^	0.072	1.479	-17.571 - 2.118	-1.263	3.900	(5) 11.381	0.044	0.557	0.375
Emotional well-being	-4.005 - 3.896	-1.478 - 2.442^	-0.391	1.248	-10.177 - 2.128	-0.333	1.192	(45) 75.307	0.003	0.846	0.863
Self-consciousness	-3.084 - 3.145	-1.268 - 1.022^	-0.090	0.693	-2.149 – 1.795	-0.241	0.864	(7) 11.340	0.125	0.529	0.579
Participation	-2.610 - 2.650	-0.912 - 0.995^	-0.107	0.517	-17.236 – 2.137	-1.128	3.847	(9) 6.998	0.637	0.435	0.571

^ indicates suboptimal targeting (items do not spread across persons)

SD standard deviation; FitRes fit residual; ChiSqu chi square; df degrees of freedom; p probability value; ext extremes cases

Table 7.3 Field Test 2 Rasch analysis summary statistics for PU-QOLI scale items

	Thresholds	Location	FitRes	ChiSqu	Prob	Person-item correlations	DIF age		DIF gender		DIF HC setting	
							Uni	Non	Uni	Non	Uni	Non
Pain scale items (n=180; 4 CI)												
Uncomfortable	Ordered	-1.104	-0.771	6.527	0.089	<0.3	+	+	+	+	+	+
Tenderness	Ordered	-1.069	-0.192	3.708	0.295	<0.3	+	+	+	+	+	+
Annoying	Ordered	-0.670	-1.817	7.508	0.057	<0.3	+	+	+	+	+	+
Red raw	Ordered	0.219	-0.145	2.065	0.559	<0.3	+	+	+	+	+	+
Stinging	Ordered	0.388	-0.496	2.817	0.421	<0.3	+	+	+	+	+	+
Burning	Ordered	0.482	0.328	0.314	0.957	<0.3	+	+	+	+	+	+
Throbbing	Ordered	0.729	0.692	4.205	0.240	<0.3	+	+	+	+	+	+
Stabbing	Ordered	1.024	0.519	6.468	0.091	<0.3	+	+	+	+	+	+
Exudate scale items (n=59; 2 CI – people with only superficial PUs removed, if retained n=95)												
Dressing off	Disordered Dichotomised	-0.751	0.804	0.020	0.887	<0.3	+	+	+	+	+	+
Staining	Ordered	-0.398	-0.758	0.367	0.544	<0.3	+	+	+	+	+	+
Weeping	Ordered	-0.356	-0.342	0.670	0.413	<0.3	+	+	+	+	+	+
Sticky	Ordered	-0.266	-0.129	0492	0.483	<0.3	+	+	+	+	+	+
Messy	Ordered	-0.006	-1.264	4.607	0.032	<0.3	+	+	+	+	+	+
Running	Ordered	0.252	-0.269	0.255	0.613	<0.3	+	+	+	+	+	+
Bleeding	Ordered	0.683	0.552	1.493	0.222	<0.3	+	+	+	+	+	+
Pus	Ordered	0.843	1.566	2.599	0.107	<0.3	+	+	+	+	+	+
Odour scale items (n=21; 2 CI – people with only												

	Thresholds	Location	FitRes	ChiSqu	Prob	Person-item correlations	DIF age		DIF gender		DIF HC setting	
							Uni	Non	Uni	Non	Uni	Non
superficial PUs removed, if retained n=27)												
Unpleasant	Ordered	-1.303	-0.039	1.340	0.247	<0.3	+	+	+	+	+	+
Lingering	Ordered	-0.207	-0.915	2.186	0.139	<0.3	+	+	+	+	+	+
Pungent	Ordered	-0.187	0.330	1.195	0.274	<0.3	+	+	+	+	+	+
Stench	Ordered	0.047	-0.566	0.273	0.602	<0.3	+	+	+	+	+	+
Putrid	Ordered	0.745	-0.465	0.404	0.525	<0.3	+	+	+	+	+	+
Sickening	Ordered	0.906	0.640	2.591	0.108	<0.3	+	+	+	+	+	+
Sleep scale items (n=133; 3 CI)												
Comfortable position	Ordered	-0.907	0.777	0.468	0.792	<0.3	+	+	+	+	+	+
Sleep in one position	Ordered	-0.058	2.639*	3.239	0.198	<0.3	+	+	+	+	+	+
Interrupted sleep	Ordered	0.027	-1.144	9.530	0.009	<0.3	+	+	+	+	+	+
Not getting amount of sleep needed	Ordered	0.065	-1.789	13.303	0.001*							
Kept awake	Ordered	0.422	-1.485	6.333	0.042	<0.3	+	+	+	+	+	+
Trouble falling asleep	Ordered	0.451	1.427	6.838	0.033	<0.3	+	+	+	+	+	+
Mobility & Movement scale items (n=130; 3CI)												
Pushing up to sitting	Ordered	-0.457	-0.123	3.303	0.192	<0.3	+	+	+	+	+	+
Adjusting in bed	Ordered	-0.349	-0.832	6.928	0.031	0.498	+	+	+	+	+	+

	Thresholds	Location	FitRes	ChiSqu	Prob	Person-item correlations	DIF age		DIF gender		DIF HC setting	
							Uni	Non	Uni	Non	Uni	Non
Difficulty sitting	Ordered	-0.155	2.310	0.990	0.610	<0.3	+	+	+	+	+	+
Difficulty turning/ moving in bed	Ordered	-0.138	-0.454	3.079	0.214	0.498	+	+	+	+	+	+
Walking slowed	Ordered	-0.006	-0.501	6.426	0.040	0.701	+	+	+	+	0.000	+
Difficulty standing long periods	Disordered	0.165	-0.060	1.008	0.604	<0.3	+	+	+	+	+	+
Limited in ability to walk	Ordered	0.168	0.198	4.790	0.091	0.701	+	+	+	+	0.001	+
Difficulty transferring	Ordered	0.201	0.747	0.378	0.828	<0.3	+	+	+	+	+	+
Limited in ability to go up/down stairs	Disordered	0.572	0.475	0.954	0.621	<0.3	+	+	+	+	+	+
Activity scale items (n=95; 2 CI)												
Regular activities	Disordered	-0.299	0.956	0.652	0.419	<0.3						
Washing	Ordered	-0.298	1.564	0.097	0.756	<0.3	+	+	+	+	+	+
Shopping	Disordered	-0.230	-1.446	1.825	0.177	<0.3	+	+	+	+	+	+
Toileting	Ordered	-0.125	0.962	0.055	0.815	<0.3	+	+	+	+	+	+
Dressing	Ordered	-0.003	0.281	5.084	0.024	<0.3	+	+	+	+	+	+
Jobs around house	Disordered	0.059	-0.814	2.247	0.134	<0.3	+	+	+	+	+	+
Doing things enjoy	Ordered	0.334	0.872	2.002	0.157	<0.3	+	+	+	+	+	+
Being	Disordered	0.561	-0.263	1.642	0.200	<0.3	+	+	+	+	+	+

	Thresholds	Location	FitRes	ChiSqu	Prob	Person-item correlations	DIF age		DIF gender		DIF HC setting	
							Uni	Non	Uni	Non	Uni	Non
emotionally close												
Vitality scale items (n=98; 2 CI)												
Tired	Ordered	-0.500	-0.327	0.992	0.319	0.49	+	+	+	+	+	+
Fatigued	Ordered	-0.493	-2.177	7.987	0.005	0.49	+	+	+	+	+	+
Energy reduced	Ordered	-0.148	0.104	0.824	0.364	<0.3	+	+	+	+	+	+
Unwell/poorly	Ordered	0.338	1.624	0.415	0.521	<0.3	0.00	+	+	+	+	+
Appetite reduced	Ordered	0.804	1.133	1.164	0.281	<0.3	+	+	+	+	+	+
Emotional well-being scale items (n=181; 4 CI)												
Fed-up	Ordered	-1.478	1.109	2.428	0.489	<0.3	+	+	+	+	+	+
Frustrated	Ordered	-1.055	-1.298	9.373	0.025	<0.3	+	+	+	+	+	+
Annoyed/irritated	Ordered	-0.673	1.542	4.816	0.186	<0.3	+	+	+	+	+	+
Physically dependent	Ordered	-0.598	0.558	2.208	0.530	<0.3	+	+	+	+	+	+
Miserable	Ordered	-0.441	-1.073	7.850	0.049	<0.3	+	+	+	+	+	+
Anxious	Ordered	-0.298	1.223	7.749	0.052	0.560	+	+	+	+	+	+
No control	Ordered	-0.120	-2.261	7.078	0.069	<0.3	+	+	+	+	+	+
Burden/nuisance	Ordered	-0.113	-0.096	3.332	0.343	<0.3	+	+	+	+	+	+
Concerned/ worried	Ordered	-0.104	0.795	0.719	0.867	0.560	+	+	+	+	+	+
Angry	Ordered - dichot	0.164	-0.735	4.103	0.250	<0.3	+	+	+	+	+	+
Missing out	Ordered	0.223	-1.209	5.481	0.140	<0.3	+	+	+	+	+	+

	Thresholds	Location	FitRes	ChiSqu	Prob	Person-item correlations	DIF age		DIF gender		DIF HC setting	
							Uni	Non	Uni	Non	Uni	Non
Depressed	Ordered	0.235	-2.361	9.588	0.022	<0.3	+	+	+	+	+	+
Lonely	Ordered	0.891	-0.832	1.770	0.621	0.519	+	+	+	+	+	+
Cut off/isolated	Ordered	0.926	-1.345	3.520	0.318	0.519	+	+	+	+	+	+
Others avoided	Ordered - dichot	2.442	0.117	5.294	0.152	<0.3	+	+	+	+	+	+
Self-consciousness scale items (n=100; 2 CI)												
Helpless	Ordered	-1.268	-0.517	0.784	0.376	<0.3	+	+	+	+	+	+
Lacking confidence	Ordered	-0.654	-0.025	0.143	0.705	<0.3	+	+	+	+	+	+
Self-conscious	Ordered	-0.465	0.114	1.388	0.239	0.415	+	+	+	+	+	+
Embarrassed	Ordered	-0.290	0.077	0.731	0.393	0.415	+	+	+	+	+	+
Feeling physically unattractive	Ordered	0.727	-1.131	3.061	0.080	<0.3	+	+	+	+	+	+
Lack understanding from others	Ordered	0.928	1.137	4.227	0.040	<0.3	+	+	+	+	+	+
Uneasy being close to others	Ordered	1.022	-0.283	1.007	0.315	<0.3	+	+	+	+	+	+
Participation scale items (n=82; 2 CI												
Restricted where go out	Disordered	-0.912	-0.962	2.095	0.148	<0.3	+	+	+	+	+	+
Difficulty going out	Ordered	-0.877	-0.801	0.991	0.319	<0.3	+	+	+	+	+	+
Restricted how	Disordered	-0.664	-0.227	0.424	0.515	<0.3	+	+	+	+	+	+

	Thresholds	Location	FitRes	ChiSqu	Prob	Person-item correlations	DIF age		DIF gender		DIF HC setting	
							Uni	Non	Uni	Non	Uni	Non
long stay out												
Holiday/weekend	Disordered	-0.016	0.403	0.690	0.406	<0.3	+	+	+	+	+	+
Give up hobbies/ leisure	Disordered	0.188	-0.387	1.193	0.275	<0.3	+	+	+	+	+	+
Participate family gatherings	Disordered	0.356	0.342	0.689	0.407	0.694	+	+	+	+	+	+
Meeting family/ friends	Disordered	0.428	0.116	0.156	0.693	0.694	+	+	+	+	+	+
Plan going out around PU care	Disordered	0.501	0.163	0.151	0.698	<0.3	+	+	+	+	+	+
Time involved caring for PU	Ordered	0.995	0.387	0.608	0.436	<0.3	+	+	+	+	+	+

* indicates values outside recommended range (i.e. misfit)

+ no DIF detected

CI class interval; SD standard deviation; FitRes fit residual; ChiSqu chi square; DIF differential item functioning; df degrees of freedom; p probability value; ext extremes cases; HC healthcare; Uni Uniform; Non non-uniform

7.6.1 Data quality

Total scores could be computed for most people for all scales (computable scale score range 95.6 – 99.6%; Table 7.4), implying good data quality.

7.6.2 Scaling assumptions

Scaling assumptions were mainly satisfied. Mean scale scores and standard deviations were mostly similar to scale mid-points (Table 7.4). All item–own-scale correlations were high (corrected item-total correlations range = 0.525 - 0.920; Table 7.5), satisfying the recommended criteria (> 0.3), thus providing support that items within scales measured a common underlying construct. Further, corrected item-total correlation > 0.3 indicate that items within each scale contain a similar proportion of information.

7.6.3 Targeting

Scale-to-sample targeting was reasonable. Scale scores spanned the scale range but were notably skewed for four scales (value outside +/-1.0); mean scores were near the scale mid-point for only 3/10 scales, however, due to many people responding at the floor (lowest score), this finding is expected; and ceiling effects were negligible, however floor effects exceeded the 15% criterion for 4/10 scales (Table 7.4).

7.6.4 Reliability

Internal consistency reliability was high as demonstrated by Cronbach's alpha values for all PU-QOLI scales exceeding the standard criterion of 0.7 (range 0.893 – 0.969; Table 7.5). Item–total correlations ranged from 0.525 - 0.920, fulfilling the recommended criteria of >0.3. Finally, test-retest correlations for 8/10 scales exceeded 0.7 and 6/10 scale correlated over 0.8; two scales had correlations below the recommended criteria, but marginally (Table 7.4), thus mostly fulfilling the recommended minimum criteria and indicating good scale stability.

7.6.5 Validity

Within-Scale Analyses

Evidence towards the internal construct validity of the PU-QOLI scales is supported by moderate to high item-total correlations; high Cronbach's coefficient alphas; and moderate to high inter-item correlations (means > 0.48 and ranges between 0.226 – 0.934 indicate that PU-QOLI scale items were mostly correlated with scale scores; Table 7.5), indicating that each scale measures a single construct.

Convergent validity

Correlations between PU-QOLI scales and hypothesised related scales of the SF-12 and the PU-QOLI overall QoL and pain items were consistent with most predictions (Table 7.6), providing support that PU-QOLI scales measure what they intend to measure; moderate to high correlations (r >0.30) were predicted.

Discriminant validity

Correlations between PU-QOLI scales and sociodemographic variables (age, gender) were consistent with predictions (r <0.30; Table 7.6), thus suggesting that responses to PU-QOLI scales are not biased by age or gender.

Known Groups validity

The PU-QOLI exudate and odour scales confirm hypothesised group differences as predicted; there was a significant step increase in mean scores by PU severity groups (Table 7.6). Tests of group difference also found a significant step increase in scores for scales vitality, daily activities, emotional well-being, and self-consciousness by PU severity. In contrast, no step increase in scores was observed for scales pain, sleep, mobility and movement, and participation. These tests were considered exploratory as samples sizes used for scale analysis ranged between 4 and 82 patients. For all scales apart from the sleep scale, the mean score on HRQL outcomes for category 1 PU severity was lower than category 3/4 PU severity, suggesting that HRQL outcomes are worse for people with severe PUs compared to those with superficial category 1 PUs. It is important to note that category 1 PUs had small samples (range 4-14 patients) therefore the known groups validity results are preliminary and require further empirical evidence to provide confidence in whether PU-QOLI scales can detect small differences in the constructs being measured.

7.6.6 Summary of traditional psychometric analyses

The traditional analysis supported the PU-QOLI scales as reliable and valid measures of PU-symptoms, physical and social functioning, and psychological well-being. The criteria were satisfied for most psychometric properties evaluated. Briefly, data quality was high (scale scores were computable for 95.2-99.6% of respondents) and scaling assumptions were satisfied (mostly similar mean item scores, corrected item-total correlations range = 0.511 - 0.940). Scale-to-sample targeting was good, apart from those scales where skew was clinically reasonable (scale scores spanned the scale range but were notably skewed for four scales (value outside +/-1.0), mean scores were near the scale mid-point for 6/10 scales, and ceiling effects were negligible, however floor effects exceeded the 15% criterion for 4/10 scales).

Table 7.4 PU-QOLI scale level analyses – Data completeness and Targeting (n=229)

Scale	**Data completeness** - Computable scale score (%)	**Targeting**						
		Possible score range*	**Range mid-point**	**Observed score range**	**Mean score**	**SD**	**F/C effect (%)**	**Skewness**
Pain	95.6	0 – 16	8	0 – 16	6.14	4.586	15.2/3.9	0.396
Exudate	98.3	0 – 15	7.5	0 – 15	2.09	3.494	57.0/0.9	1.898
Odour	99.6	0 – 12	6	0 – 12	0.97	2.850	83.0/4.3	3.144
Sleep	99.6	0 – 12	6	0 – 12	4.66	4.302	10.7/4.1	0.434
Vitality	98.3	0 - 10	5	0 - 10	2.72	3.217	27.0/2.2	0.896
Mobility	97.8	0 – 18	9	0 – 17	7.077	5.377	1.5/0.4	0.362
ADL	95.6	0 – 16	8	0 – 14	3.67	4.389	3.9/0.4	1.058
Emotional Well-being	95.2	0 - 30	15	0 - 28	10.15	9.190	8.3/1.3	0.673
Appearance & self-consciousness	96.5	0 - 14	7	0 - 14	2.53	3.632	38.7/2.2	1.566
Participation	95.6	0 - 18	9	0 - 18	5.658	6.264	6.2/0.4	0.587

*High scores indicate great bother/impact
SD standard deviation; F/C floor/ceiling – floor effect =% scoring 100 (greatest bother/impact); ceiling effect = % scoring 0 (least bother/impact)

Table 7.5 PU-QOLI scale level analyses - Reliability and scaling assumptions: Validity within-scale analysis (n=227)

Scale	**Internal consistency** Cronbach's alpha	**SEM**	**95% CI**	**Mean IIC**	**IIC**	**Scaling Assumptions -**Corrected ITC	**Test Retest Reproducibility**		
							ICC consistency	**ICC absolute**	**Correlation**
Pain	0.893	0.453	6.50, 8.29	0.482	0.235 – 0.663^	0.525 – 0.703^	0.803	0.805	0.804
Exudate	0.907	0.233	1.63, 2.55	0.544	0.316 – 0.715^	0.511 – 0.752^	0.622	0.625	0.622
Odour	0.969	0.187	0.60, 1.35	0.841	0.716 – 0.934^	0.794 – 0.940^	0.681	0.680	0.700
Sleep	0.920	0.327	4.01, 5.30	0.657	0.491 – 0.805^	0.681 – 0.846^	0.822	0.816	0.824
Vitality	0.900	0.275	2.18, 3.27	0.638	0.488 – 0.902^	0.628 – 0.898	0.735	0.738	0.736
Mobility	0.927	1.069	6.20, 10.59	0.586	0.226 – 0.912^	0.666 – 0.799^	0.873	0.864	0.879
ADL	0.952	1.3741	4.83, 10.57	0.710	0.407 – 0.904^	0.583 – 0.899^	0.866	0.872	0.870
Emotional Well-being	0.934	0.931	9.71, 13.42	0.486	0.242 – 0.789^	0.537 – 0.761^	0.829	0.820	0.832
Appearance & self-consciousness	0.894	0.271	1.99, 3.07	0.557	0.371 – 0.789^	0.617 – 0.755^	0.812	0.814	0.814
Participation	0.932	0.719	4.23, 7.09	0.601	0.359 – 0.877^	0.599 – 0.861^	0.627	0.639	0.634

SEM standard error mean; CI confidence interval; IIC inter-item correlation
^Range item-total correlation (ITC)

Table 7.6 PU-QOLI scale level analyses – Validity (n=229)

	Convergent Validity								Discriminant Validity		Known Groups		
Scale	**SF12 Physical Function Scale r[1]**	**SF12 Social Function Scale r[1]**	**SF12 Role Physical Scale r[1]**	**SF12 Mental Health Scale r[1]**	**SF12 Pain item r[1]**	**SF12 Fatigue item r[1]**	**PU-QOL pain item r[1]**	**PU-QOL QOL item r[1] (n)**	**Gender R^2 (n)**	**Age r^2 (n)**	**PU severity^ R^3**	**Mean score (n)**	**P value (CI)**
Pain	-	-	-	-	0.48[b]	-	0.79[b]	0.38[b] (206)	0.13[b] (214)	0.11[b] (214)	0.895		0.895
Category 1												5.36 (14)	(2.85, 7.86)
Category 2												5.81 (77)	(4.78, 6.83)
Category 3/4												5.51 (68)	(4.49, 6.54)
Exudate	-	-	-	-	-	-	-	0.25[a] (216)	0.08[b] (225)	-0.14[b] (224)	0.000*		0.000*
Category 1												0.64 (14)	(-0.43, 1.72)
Category 2												1.07 (81)	(0.55, 1.60)
Category 3/4												3.26 (72)	(2.31, 4.21)
Odour	-	-	-	-	-	-	-	0.20[a] (217)	0.05[b] (228)	-0.14[b] (227)	0.004*		0.004*
Category 1												0.07 (14)	(-0.08, 0.23)
Category 2												0.28 (82)	(-0.05, 0.61)
Category 3/4												1.60 (72)	(0.77, 2.43)
Sleep	-	-	-	-	-	-	-	0.32[b] (171)	0.21[b] (178)	0.10[b] (178)	0.774		0.774
Category 1												4.89 (9)	(1.36, 8.42)
Category 2												4.49 (65)	(3.41, 5.57)
Category 3/4												4.02 (54)	(2.84, 5.20)
Vitality	-	-	-	-	-	0.36[b]	-	0.52[b] (135)	0.03[b] (137)	-0.16[b] (137)	0.036*		0.036*
Category 1												1.22 (9)	(-0.40, 2.84)
Category 2												1.82 (50)	(1.02, 2.62)
Category 3/4												3.25 (48)	(2.26, 4.24)

Mobility	-0.50[b]	-	-	-	-	-	-	0.39[b] (37)	0.04[b] (39)	0.22[b] (39)	0.137		0.137
Category 1												5.00 (4)	(-1.62, 11.62)
Category 2												4.36 (11)	(1.94, 6.79)
Category 3/4												8.31 (13)	(4.82, 11.80)
ADL	-	-	-0.389[b]	-	-	-	-	0.35[b] (48)	-0.05[b] (49)	-0.19[b] (49)	0.094		0.094
Category 1												1.60 (5)	(-0.66, 3.86)
Category 2												1.73 (11)	(-1.06, 4.51)
Category 3/4												4.63 (24)	(2.81, 6.44)
Emotional Well-being	-	-	-	-0.44[b]	-	-	-	0.58[b] (133)	0.16[b] (135)	-0.15[b] (135)	0.001*		0.001*
Category 1												4.13 (8)	(1.39, 6.86)
Category 2												7.41 (46)	(4.98, 9.84)
Category 3/4												13.28 (47)	(10.39,16.16)
Appearance and self-consciousness	-	-	-	-0.40[b]	-	-	-	0.50[b] (176)	0.23[b] (179)	-0.03[b] (178)	0.014*		0.014*
Category 1												0.92 (12)	(-0.42, 2.26)
Category 2												1.85 (62)	(1.02, 2.68)
Category 3/4												2.52 (58)	(2.43, 4.71)
Participation	-	-0.523[b]	-	-	-	-	-	0.51[b] (75)	0.01[b] (76)	-0.29[b] (76)	0.018*		0.018*
Category 1												3.67 (6)	(-1.55, 8.88)
Category 2												2.55 (22)	(0.43, 4.66)
Category 3/4												7.35 (31)	(4.84, 9.87)

[1] Spearman correlation
[2] Pearson correlation
[3] ANOVA
^PU severity categorised into 3 PU groups: category 1, category 2, and category 3 and 4 combined
*Correlation significant at p=0.05
[a]Correlations falling outside of the predicted range; [b]Correlations consistent with predictions
CI confidence interval

Findings from the reliability analyses supported PU-QOLI scales as reliable measures, with acceptable internal consistency (i.e. supporting the scales are measuring single constructs), and the ability to produce highly reproducible scores on repeated applications. The high item-total correlations and alpha coefficients provide evidence towards the internal construct validity of the PU-QOLI scales. There was good evidence for convergent and discriminant construct validity; moderate to strong correlations were demonstrated between PU-QOLI scales with related constructs and low correlations between PU-QOLI scale scores and external unrelated variables. Scales exudate and odour were able to differentiate known groups as predicted. All other tests of known group difference tests were considered exploratory.

7.7 Discussion

The final psychometric evaluation demonstrated that PU-QOLI scales mostly satisfy criteria for rigorous measurement, confirming the acceptability, reliability and validity of PU-QOLI scales. Some problems were detected that suggest improvements could be made (i.e. revisit qualitative work to add items to extend the measurement at the floor/ceiling scale range) and some scales would benefit from additional empirical evidence to support any changes or modifications required to improve the scales.

The final version of the PU-QOLI is a 10 scale PRO instrument for measuring symptoms, physical functioning, psychological well-being and social participation specific to PUs (Appendix 7.1). There are three symptom scales measuring: pain (8 items), exudate (8 items), and odour (6 items); four physical functioning scales measuring: sleep (6 items), movement and mobility (9 items), daily activities (8 items) and vitality (5 items); two psychological well-being scales measuring: emotional well-being (15 items) and self-consciousness and appearance (7 items); and one social participation scale (8 items). Scale scores are generated by summing items and then transforming to a 0-100 scale. High scores indicate greater patient bother.

7.8 Summary of Chapter 7

The psychometric evaluation of the PU-QOLI was undertaken in two field tests, applying both Rasch measurement and tradition psychometric methods. The final psychometric evaluation confirmed the acceptability, reliability and validity of the PU-QOLI scales but also demonstrated important limitations in some of the scales that require further testing. The next Chapter presents a discussion of the findings of the whole thesis, and considers study limitations and future directions.

CHAPTER 8

GENERAL DISCUSSION

8.1 Overview

This thesis aimed to establish the impact of PUs on HRQL and determine the need for a PU-specific PRO instrument. The research methods undertaken, important findings and the methodological developments woven into the research have been discussed thoroughly in their respective chapters. Thus, this chapter aims to summarise the main findings from the thesis, discuss study limitations, and explore the implications for PU management and future research.

8.2 Summary of findings from phase one

Phase one of this research consisted of two systematic reviews of the literature, a qualitative study, and development of a conceptual framework. The existing research in PUs was reviewed to establish the impact of PUs on HRQL and the potential need for a PU-specific PRO instrument (Chapter 2). Despite a sparse PU and HRQL literature, some important findings emerged. PUs and PU interventions cause patients significant burden, specifically severe pain and increased discomfort due to treatment. PU symptoms, use of various PU treatments and interventions, and the healthcare environment have detrimental effects on patients, contributing to physical, social and psychological impairment. People with PUs have mixed beliefs about PU causes; the most common being a direct result of the healthcare received, and unless a patient had a history of PUs or belonged to groups at high-risk of PU development, there was lack of knowledge about PUs, suggesting the need for patient information and education about PU risks and treatments. Other important issues identified related to patients' views and concerns being ignored by healthcare professionals, early warning signs (e.g. pain) did not always prompt action, and the physical, social and psychological aspects of patient need were not routinely met.

The HRQL literature in the PU field is mainly qualitative, with emphasis on pain and physical functioning impairment rather than a comprehensive exploration of issues important to patients that may be broader than the commonly investigated HRQL domains. Potential sources of bias in the qualitative work are due to low sample sizes (n=≤10) and underrepresentation of people with superficial PUs, the elderly (>70 years), and those acutely ill or with various comorbidity. PUs are common in patients with serious acute and chronic illness and the current evidence does not include the

views of all types of people with PUs. Further, there is a lack of clear definitions and consistency in conceptualisation of HRQL outcomes, making it difficult to formulate precise operational constructs to guide the development of new specific PRO instruments.

The review highlighted HRQL outcomes that are unique to PUs although there is a lack of quantitative work and these outcomes are currently not systematically included as outcomes in clinical trials. Therefore the PU literature is unconvincing in terms of robust evaluation of the impact of PUs treatments on HRQL. The few quantitative studies designed to explore HRQL in PUs had used measures not developed or validated for use with patients with PUs. For RCTs, some included *ad hoc* questions (i.e. a questionnaire designed for a specific study, but not psychometrically evaluated) to measure sleep, comfort and pain rather than utilising existing validated instruments to assess HRQL outcomes. Consequently, the limitations associated with the conduct and reporting of quantitative studies limited the furthering of our understanding of the impact of PUs and PU interventions on HRQL.

The intention of the review was to identify all possible research on PU-specific HRQL. Interventional studies are intended to evaluate an intervention and not necessarily to illuminate the nature of HRQL, however interventional studies were reviewed as PROs associated with the impact of PU interventions (e.g. asking patients about their preference, comfort or pain) or HRQL outcomes using existing PRO instruments may have been assessed. Quality assessment was applied, based on standards set by others (39, 337), for inclusion of RCTs. Reasoning for applying the criteria was for consideration of study quality and consequently confidence in study results. The loss to follow-up criteria was imposed on the study primary outcome rather than PRO data as failure to complete questionnaires or return them is likely in research. Compliance with PU treatment was a problem for all RCTs, with many patients lost through the course of the study, resulting in RCTs being excluded. If this had been a review of intervention effectiveness then a strict quality criteria would have been required. However, as the intention was to generate content (e.g. qualitative elements from quantitative research), such a strict criteria was not needed. Although all RCTs were excluded at this stage, those that included PROs were considered for content and relevance to people with PUs in subsequent work (Chapter 4).

Despite limited evidence, phase one of this research produced a working conceptual framework of HRQL domains specific to PUs and highlighted the potential need for outcome measures that can accurately depict the impact of PUs on HRQL. The next

stage of the first phase was to confirm the relative importance of findings and formulate a final conceptual framework.

New instrument construction needs to be underpinned with a strong conceptual base to ensure valid measurement; one that adequately defines the variables and relationships conceptually and gives operational meaning that guides the development (or selection) of PRO instruments (or scales) (86). Development of a PU-specific conceptual framework was hampered by poor quality and quantity of literature, and as a result the framework development relied heavily on qualitative interviews with patients, endorsed by expert opinion.

A PU-specific conceptual framework of HRQL was developed (Chapter 3) that includes four constructs: PU-specific symptoms, physical functioning, psychological well-being and social functioning. These constructs are similar to those in generic HRQL models (78, 89, 96) however, this framework also incorporates components specific to PUs. There is no universally accepted definition of HRQL; however, central to most definitions is that HRQL is reflected by an individual's subjective perceptions, suggesting that the patients' unique perspective is vital and necessary when developing new PRO instruments. On the basis of the Food and Drug Administration recommendations (86), the perspective of people with PUs was kept central at all stages of the development and validation of the PU-QOLI.

The researcher, having reviewed the literature, may be criticised during data collection and/or analysis as being constrained by preconceived ideas of the relationships between some HRQL constructs and their properties. However, as this research was applied in nature, some focus as to the constructs to be measured (i.e. HRQL) was considered acceptable. The intention was to provide support for the work that had been done before and formulate a final conceptual framework based on previous research and new patient-reported data. To minimise the risk of bias, an interview guide was developed to probe patient responses and ensure comprehensiveness rather than force comments or use leading questions. Other precautions were used to ensure that data was analysed in a systematic manner, and that data interpretations were grounded in actual patient data to avoid bias that could be introduced with familiarisation with previous literature.

Informal discussions with tissue viability nurse specialists/consultants allowed greater understanding of the experience of PUs and opinions regarding patient needs. However, the clinical experts consulted were part of the wider project team and

therefore, the opinions expressed may not have been generalisable to all tissue viability professionals at other clinical sites. Consultation with clinical experts at this stage of the research was intended to help contextualise the impact of PUs, identify any important missed information, and provide clinical support for the findings (outcomes of interest) derived qualitatively.

The framework includes conceptual domains for: symptoms, including pain, exudate and odour; sleep impairment; difficulty with range of movement and mobility; limitations in daily activities, including aspects of self-care, home life, and doing things one enjoys; psychological functioning, including emotional well-being (i.e. mood, anxiety, isolation, dependence) and self-consciousness about ones physical appearance; and ability to participate socially. Importantly, in the PU literature, the studies that developed and used *ad hoc* measures, despite seldom assessing symptoms, sleep, comfort and physical functioning outcomes, suggested that these outcomes are important in wound care. The reason for them not being systematically included in PU research may therefore be due to a lack of appropriate (reliable and valid) measures to assess these outcomes.

This research was the first to conceptually map the range and nature of the impact of PUs on HRQL and investigate patterns of association unique to specific patient factors in patients across all adult age-groups and gender, PU severity and skin site, clinical specialities and healthcare settings. This was to ensure that the PU-specific HRQL conceptual model captured the views of all types of people with PUs and allowed a between case analysis comparison, where arising themes were coded and analysed so that commonalities between patients were identified. Inclusion of patients with varying comorbidities assisted in separating out the effect of existing conditions from the effect of PUs to ensure that the conceptual framework was representative of PU-specific impact on HRQL and relevant to all types of patients with PUs. However, the sample was limited to those who are English-speaking British nationals without cognitive impairment. Differences in outcomes were explored by age, gender, PU severity and location. PU symptoms were problematic for all patients, however older patients were less likely to report pain and for those with impaired sensitivity it was assumed pain was not a problem. This suggests that special attention to PU management procedures that could inflict pain, particularly in older patients who may be less likely to report pain, is needed. It also warrants further PU pain assessment and pain management research. Further cross-case analysis, investigating patterns of association by other PU subgroups and investigation of mediating factors that affect

HRQL in PUs will add to the growing body of evidence about the impact of PUs on HRQL.

The review of the PU and HRQL literature (Chapter 2) failed to identify PRO instruments specific for PUs. The few studies that measured HRQL outcomes related to PU and PU treatment impact assessed outcomes with generic or chronic-wound specific PRO instruments. This suggested that measures for other similar chronic wounds were available but suitability for use in PUs had to be determined. Therefore, before developing a new instrument, existing instruments were examined. The final conceptual framework provided a structured and formal method to assess the content of any available measures against and potentially provided the basis for the development of a new PU-specific measure of HRQL if required.

The review of HRQL measures used in PU and other chronic wound research (Chapter 4) revealed that HRQL outcomes important in PUs are inadequately covered by generic and chronic wound-specific instruments. Despite similarities between conceptual models, the content differed at sub-domain and item level. For example, the VLU measure had 75% of relevant HRQL domains, but items contained in relevant domains were not PU-specific. The VLU physical functioning scale contains the item “difficult to move” but misses important PU-specific issues such as difficulty with transferring, sitting for long periods, and moving around on pressure-relieving mattresses. Due to issues of content validity and specificity, the appropriateness of using existing PRO measures in PU research was questioned. The 75% cut-off (included 10 of 13 PU-specific domains) was selected to determine whether existing PROs represented PU-specific issues (domains). If a lower threshold had been used then additional PROs would have been considered. The items from existing PROs used in chronic wounds research were inspected to include in the item pool but items were found to not be PU-specific.

Underlying the limitations of using these instruments in PUs is a lack of understanding of what patients perceive to be important in terms of HRQL outcome. Consequently none of the existing PRO instruments used can be considered valid measures of PU patient perceptions of outcome, as there has been no research to comprehensively uncover these perceptions. In addition, measurement of general health status and HRQL in relation to outcome of PU treatment has been largely ignored to date (i.e. lack of data concerning benefits, cost effectiveness and long term outcomes of many common PU interventions). When selecting a PRO instrument, careful consideration of the conceptualisation of PRO instruments content as well as determining fit-for-purpose

is needed, prior to evidence for validity, reliability and responsiveness of outcome measures. As no PU-specific PRO instruments exist, and to address some of the limitations of using available instruments in PU research, patient-assessed HRQL measures specific for patients with PUs are needed for use in clinical practice and future research.

It was considered that a new measure, using stringent international criteria from the health measurement literature, would provide a vehicle by which data on PUs could be collected. Measures to address issues concerning outcomes of PUs and the impact of treatment was pertinent as to date there was no easy, standardised method for assessing outcomes important to people with PUs due to a lack of a suitable instrument (i.e. need precise (required for individual clinical decision making) and less burdensome PRO measures). The empirically derived PU-specific conceptual framework informed the development of a new HRQL instrument for PUs.

8.3 Summary of findings from phase two

Thorough and systematic steps were taken to develop and pretest (Chaoter 5) the PU-specific PRO instrument of HRQL (the PU-QOLI).

8.3.1 Development and Pre-Testing the PU-QOLI scales

The objective of this phase of the research was to develop a PRO instrument for PUs that assessed the conceptual domains generated from the review of the literature, patient interviews and expert opinion. Properties of the conceptual model formed the content (items) of the new instruments scales. The patient experience needed to be captured and therefore patient statements were included as items to ensure that the PU-QOLI was from the perspective of those with PUs using their words. However, some words included may not be meaningful to all people in the UK (e.g. “foisty”). At this stage the inclusive approach was used to ensure that all issues important to patients with PUs were captured and then tested and refined in subsequent work (i.e. pre-test and first field test). In order to encourage optimal instrument completion (the best responses and completion with no missing data), the way the PU-QOLI was designed was tailored to the characteristics of patients with PUs (i.e. understood by and relevant to the intended population). Careful consideration of the layout and instructions, framing of questions, response format and recall period was taken to reduce potential biases and cognitive and respondent burden.

Criteria designed to guide questionnaire design and item construction where followed (132, 211, 244). Wherever possible, the exact wording used by patients during the qualitative work was retained for items. Jargon (words not part of everyday vocabulary), ambiguity (vague terms) and double-barrelled questions (a questions asks about more than one issue) were avoided to reduce confusion. The appearance of a PRO instrument needs to be designed so that variability between interviewers is reduced, response rates enhanced and non-response bias and bias arising from respondent error (e.g. ticking wrong box) reduced. Instructions contained only essential information needed for PU-QOLI completion. Important instructions were highlighted and each section (question) defined in order to ensure that meaning was consistent across respondents.

Deciding on the number of response options to include is a balance between maximising precision and minimising bias and respondent burden (85, 132). The response categories were chosen so as not to have more than five responses in order to maximise the extent to which patients were distributed across response categories (132). Wording used was clear and appropriate to the question stem and intended population and a clear distinction was made between choices; discrete response option category scored with successive integer scores that imply a continuum of increasing impact, from less to more.

Important disease changes/progression and memory error (recall bias) need to be considered when choosing a time-frame. A recall period of the past-week was chosen on clinical grounds, as changes in PU severity and symptomology often occur over days and thus a longer recall period would risk not capturing relevant changes to HRQL outcomes. Events that occurred over a month ago may no longer be relevant or have since been resolved or treated.

Obtaining the best possible health outcomes data requires the use of appropriate methods to ensure high quality data with limited bias. A potential problem for data quality is interviewer bias; bias introduced due to the social interaction between the respondent and interviewer (e.g. social acquiescence or characteristics of the interviewer may affect both response rates and the nature of responses given). It was envisaged that the PU-QOLI would be used in future PU treatment effectiveness research; therefore the mode of administration selected for the new instrument was self-completed. In research, self-completed versions are often preferred as they are associated with lower costs, can cover largely dispersed populations (if postal), require less time to collect data, and avoid potential interviewer bias (244). However,

importantly, as many patients had difficulty with self-completion during pretesting, the mode of administration was changed to interview-administered (expanded in section 7.4). If mode was not changed then a large proportion of eligible patients would be excluded in future research as they would be unable to self-complete the PU-QOLI scales. The equivalence of self-completed and interview-administered versions of the PU-QOLI was investigated during preliminary field testing (expanded in section 7.3.2). In hindsight, the advantages of administration (response rates are typically higher, reducing non-response bias) make this method preferable for elderly or chronically ill populations, and accordingly suitable for people with PUs.

Pre-testing was carried out in order to ensure that the steps taken to reduce sources of potential bias had been effective and that the preliminary PU-QOLI was clear, understood and relevant to the intended population. This process was effective for identifying problems with PU-QOLI items and design early in the development process, and for guiding changes to layout, content and mode of administration. Using mixed methods provided useful information needed for item selection and deletion, particularly for ensuring clinical meaningfulness, importance to patients and good measurement properties (expanded in section 8.5).

Pretesting was particularly important in this population as patients with PUs often have underlying comorbidities. Therefore, to ensure that reported issues were PU-specific, outcomes associated with other comorbidities were isolated from PU-specific outcomes. Interview probes were carefully developed to ensure unbiased phrasing (i.e. ensure probes were not leading respondents) and further clarifying questions were asked (e.g. Do you think [that] is only because of your PU or resulting from a combination of things?), to ensure that the PU-QOLI detected PU-impact rather than impact on HRQL due to other medical conditions.

Caution should be used when making significant changes to newly developed instruments on the basis of small samples. Modifications made should improve the content and quality of data collected. For the PU-QOLI, a crude estimate (>40%) was used to provide a systematic approach based on standard psychometric criteria (37, 38) for the removal or merging of items. During development, an inclusive approach was taken, where all patient verbatim were retained as items, including verbatim/items with very similar wording. During cognitive interviews, patients were specifically asked to comment on similarities with the intention of retaining those items most representative of the patient perspective. Decisions to remove or merge items were made after extensive discussion with both clinical and health outcome methodology

experts to ensure that specific components of each construct/scale were not removed but rather only items pertaining to the same component but that were worded slightly different. Expert consultation also resulted in the addition of items representing PU-symptoms. Counting frequency or assessing intensity of symptoms may not be an adequate measure of HRQL (39), but the impact of these symptoms and the meaning for individuals is an important aspect of HRQL assessment and therefore considered important outcomes for inclusion in a PU-specific PRO instrument.

Minor changes made following pre-testing resulted in a 13 scale instrument intended for interview-administration. The modified PU-QOLI was tested to ensure that no subsequent problems were introduced after modification, and confirmed acceptability of the final preliminary version, which took around 30 minutes to complete. The completion time may be considered a burden to patients, however, part of the rationale for the next stage of research (item-reduction) was to reduce patient burden through the generation of scales that included items with the best measurement properties.

Pretesting was useful for identifying problems with PU-QOLI items and design. The methods guided changes to layout, content, administration mode, and item selection to reduce respondent burden, decrease data errors and non-response, and provide further validity and clinical utility of each PU-QOLI scale as reflected by clinically meaningful hierarchical scales, prior to formal psychometric evaluation. Patient input proved to be the most important element of the development process. The Rasch analysis was not used to make changes to the PU-QOLI scales. However, it was a useful demonstration of how Rasch measurement methods can provide a complementary method alongside standard qualitative pretesting, for evaluating the strengths and weaknesses of PROs during early testing and illustrated the changes that could have been made if both methods were used simultaneously. This was a novel methodological development, as Rasch measurement methods have not been previously used during early PRO development therefore the ideas presented here are new. Historically, the Rasch paradigm has not been widely adopted in health measurement research, but more recently the benefits of using Rasch methods are demonstrated (87, 262, 284, 305). This work highlights that iterative, small tests, were PRO development is based on qualitative and empirical data is beneficial, with the Rasch evidence collected to highlight anomalies in the data and interpreted interactively, rather than used as strict cut-off criteria to throw out items.

8.4 Summary of findings from phase three

Phase three of this research involved psychometrically evaluating the PU-QOLI in two field tests (Chapters 6 and 7) and determining optimal mode of administration in a sub-study.

8.4.1 Preliminary psychometric evaluation of PU-QOLI scales

The preliminary PU-QOLI was field tested in 236 patients with PUs. The feasibility and acceptability was established as demonstrated by acceptable data quality, indicating the suitability of interview administration. The sample included mostly white British nationals. It could be argued that there were a disproportionate number of people from different ethnicities represented in this sample. However, this sample reflected the patients with PUs under the care of tissue viability teams in the 34 hospitals and community services around the UK participating in this research, and therefore, as one of the goals of the instrument was to be used in PU research, it was important for the PU-QOLI to be based upon representative samples of patients.

This research took an approach to scale development that is strongly recommended (87, 284) but differs somewhat from approaches adopted by others. Specifically, PU-QOLI was developed on the basis of a conceptual model that defined the areas for scale development and then used an explicit mathematical model (Rasch) to guide the development of each PU-QOLI scale. More typically health rating scale development uses statistical techniques, such as factor analysis of an item pool, to define the content of the constructs (scales) to be measured, and then traditional psychometric methods to test the reliability and validity of produced scales and for further refinement (item reduce) of scales based on whether individual items meet certain psychometric criteria. With this approach, the content (items) of a scale, rather than the construct intended for measurement, define what the scale measures (87). Grouping items statistically or thematically can be misleading, as it assumes, based on correlations between items, that they measure the same thing, however this does not ensure that items in a group measure the same construct.

Most outcome measures used in healthcare are ordinal in nature, making them valid for group-level-based research. Ordinal measurements describe order but not the relative size or degree of the difference between measurements, which means that equal meaning of change scores for an underlying latent trait cannot be assumed. Put another way, a change from ‘no bother’ to ‘a little bother’ for say ‘difficulty transferring from a bed to chair’ may not have the same significance as change from ‘a little bother’

to ‘lot of bother’ on the same item. Rasch methods offer the ability to construct linear, interval-level measurements from ordinal-level rating scale data (291, 338), thereby addressing a major concern of using rating scales as outcome measures (339). Linear transformation produces a hierarchical ordering of items along the underlying measurement (construct) continuum, thus examination of changes in scores are comparable as the distance between items has meaning (290, 340).

The Rasch model also provides a powerful framework to guide scale construction, enable sophisticated checks of internal validity and consistency of scores, and a means for determining whether fundamental measurement properties are achieved, properties required for valid transformation to interval scaling. These properties include: numerical order (one mark on the ruler (the measurement continuum or item set) represents more or less of the construct than another); addition (points on rulers may be added); and specific objectivity (the calibration of the ruler is independent of the sample used to calibrate it and vice versa) (273, 290). To determine whether these properties are confirmed, testing: the scale for unidimensionality, a requirement for construct validity (132); the invariance of items (the ratio of difficulty between any pair of items remains constant across the ability levels of respondents), required for interval-level scaling; appropriate category ordering (whether the items response categories are working as expected); and DIF (whether responses to items are biased by subgroups in the sample) (85, 262, 308), is undertaken. When these tests are confirmed, or when data fit the Rasch model, the PRO instruments developed with Rasch methods are useful clinical measurement tools for individual patients, making them well-suited for use in both research and routine clinical practice (90, 341). However, for some clinicians and researchers, Rasch methods may appear complicated, requiring both specialist knowledge and training in undertaking the analyses, use of the software, and interpretation of the results. The clinically meaningful scientific advantages of using Rasch potentially outweigh these concerns.

Despite Rasch methods providing a unified approach to evaluating PRO instruments, there are no strict criteria for guiding item reduction; rather, decisions for the inclusion/exclusion of items were made according to appraisals of multiple analyses of the observed data against measurement criteria, as opposed to examinations carried out singularly or sequentially, with consideration of clinical meaningfulness. As such, some decisions about item reductions were based on conservative estimates. Therefore, it must be considered that if more stringent criteria were used for item selection then this would have resulted in different items being retained for the final scales. For example, some scale items had residual correlations above 0.30, suggesting local

dependency. However, these items were not necessarily removed based on this criteria alone; consideration of multiple criterions allowed for a fairer assessment of content, in which different item properties were assessed in combination and considered for clinical meaningfulness were retained. This is a problem faced by all developers of new rating scales. In order to justify these decisions, importantly, the items retained were needed for the breadth, range and measurement precision of each construct they measure. This was considered particularly important for clinical care and research where precise estimates of people's locations on a measurement continuum are needed for detecting clinically significant change; items are needed to cover the full measurement range of each construct measured. Thus, at this stage of PU-QOLI development, although issues relating to psychometric cut-off points are pertinent, there was reluctance to reduce the number of items any further so as not to compromise content and clinical validity or risk reducing the measurement range but still ensuring that only the best indicators of outcome were retained.

In an attempt to create scales that captured the entirety of the patient experience as represented by the *a priori* conceptual model, for each conceptual domain, items were selected to produce a scale representing that domain. The psychometric evaluation revealed that not all scales met requirements for reliable and valid measurement. Therefore some exploratory post hoc analyses were undertaken, where items were added to the vitality scale and conceptually related constructs were combined, and the psychometric properties of the new scales explored. A criticism of this method might be that the content of the new scales was thematically driven rather than by the explicit definitions of the constructs (domains) contained in the conceptual model. Patient transcripts were revisited to identify additional items for the vitality scale. The vitality scale was selected in the first instance as a means of testing the methodology in one scale, with the potential for using the same methods to improve the other scales (i.e. revisit transcripts and/or targeted qualitative work to improve the scales' measurement range). The measurement properties of the modified vitality scale were tested during the first field test. Only psychological constructs mood, anxiety, autonomy and isolation were combined into a higher order construct - emotional well-being, as items for these constructs all pertained to feelings (not items covering a wide range of constructs). The rationale for retaining these items was so that content capturing important data related to the PU-experience of individuals was not excluded. The process of modifying a newly developed instrument is part of an evolving measurement process intended to strengthen the hypothesised conceptual relationships with empiric evidence (82). Importantly, the new emotional well-being scale demonstrated good reliability and within scale validity. Finally, the item 'itchiness' was considered a separate

consequence of PUs that's has clinical value, especially in terms of patient bother and symptom management, and therefore retained as a single descriptive item.

The first field test was carried out in order to retain scale items with the best psychometric measurement properties (items that perform best against established psychometric criteria) and investigate the item response format selected for the PU-QOLI against a strict paradigm, the Rasch measurement model (276). The empirical evidence suggested that in an attempt to get good precision (greater number of response categories), people were unable to discriminate between the 4-point response options as demonstrated by disordered thresholds. Disordered thresholds threaten validity. A post hoc reduction of response options from 4 to 3-points indicated that the modification was justified as demonstrated by no subsequent disordering of thresholds.

In addition, an empirical investigation of optimal mode of administration for the PU-QOLI scales revealed that self-completion was not suitable for patients with PUs. Consequently, mode was changed to interview-administered to ensure that the PU-QOLI would be applicable for a wider range of people with PUs and potentially yield higher quality data. Administering a PRO instrument to a patient rather than them self-completing it on their own may potentially affect the responses given. The suitability of self-completion was investigated in this patient group and found not to be acceptable (pre-test and sub-study results, Chapter 5 and 6). A patient-self-completed PU-QOLI was preferred but practical constraints meant that this method was not appropriate. A user manual has been developed to guide PU-QOLI administration to ensure that the interviewer acts only as an intermediary for recording patients own responses on the instrument and to minimise the interviewer's potential for interpreting responses.

Due to the requirement for equivalent groups for a DIF analysis, only those able to self-complete the PUQOLI were included in the sub-study. The sub-study findings do not tell us anything about those people unable to self-complete. The intention was to provide preliminary empirical evidence for whether two mode versions could be developed. In the process, it set the ground work for hypotheses that can be tested in subsequent research (e.g. determine whether both mode versions are suitable for community patients).

A preliminary psychometric evaluation of the Rasch produced PU-QOLI scales supported the acceptability (demonstrated through a high percentage of computable scale scores) and reliability (demonstrated through internal consistency and item-total

correlations exceeding psychometric criteria). Construct validity was limited to within scales analysis. The final PU-QOLI scales underwent a full psychometric evaluation in a large independent sample of patients with PUs.

8.4.2 Final psychometric evaluation of PU-QOLI scales

The purpose of the final field test was to assess data quality, scaling assumptions, targeting, reliability and validity to ensure that the final PU-QOLI performed appropriately in an independent sample. Final field testing was carried in 231 patients with PUs. The empirical evidence supported the underlying conceptual model; producing 10 factors, including scales for PU symptoms, limitations of physical functioning, psychological functioning, and social participation. The final PU-QOLI contains 81 items which might be considered too long for routine clinical use or research. Importantly, the instrument consists of 10 independently evaluated scales. Each scale is a stand-alone measurement instrument that contains items needed for measurement precision of each construct they measure. However, where evaluation of PROs is often undertaken in busy multidisciplinary healthcare services, development of a short-form version of the PU-QOLI, where the most clinically appropriate items from each scale are selected (342), may be beneficial (expanded Section 8.8.3, *Clinical application*).

Overall support was established for the psychometric properties of the PU-QOLI scales demonstrated by adequate fit to the Rasch model, marginal DIF bias, good internal consistency and support for unidimensionality of the scales. Two items presented with item fit statistics below the recommended criteria, however the discrepancies were small and therefore unlikely to impact on measurement. However, important issues emerged in relation to the item response categories and sample-to-scale targeting. Inspection of the item response format revealed issues regarding the ordering of response categories for items in 5/10 scales. A reduction in the number of response categories, from three to two, might resolve any disordering; however no rescoring was undertaken at this stage. The problem with the response format may be influenced by the relatively small sample used in this study. As such, verification of findings is needed in a larger sample before specific recommendations are made. It does however suggest that dichotomizing items may be warranted in future studies to reduce potential confusion for respondents as the findings suggest that respondents are unable to reliably distinguish between what is meant by 'a little' and 'a lot'. It might be that for people with PUs, they either have the problem or they do not, in which case, a yes/no response option would be justified.

Another important finding was suboptimal scale-to-sample targeting. It is expected that in a sample of people with mixed severity PUs, people would have a range of difficulty or impairment due to PUs that would consequently affect aspects of their HRQL. However, 94% of PU-QOLI items had 'no bother' as the highest category frequency, with almost half (48.8%) of the items having a substantial percentage of people (>50%) scoring 0 (no bother). This finding implies that these patients had no bother with PUs impacting on HRQL outcomes, which is inconsistent with expectation. In practice, there is likely to be more clinical heterogeneity in patients with PUs than is implied by this finding, which indicates a targeting problem (possible mismatch between item difficulties and patients' abilities in this sample). As discussed by others (343), this is important because the limited item level targeting will impact on the overall ability of the PU-QOLI scales to detect HRQL differences between people and groups and potentially be less sensitive to the effects of interventions. Ideally, the range of outcome measured by the PU-QOLI scales should be well-matched to the range of outcome present in the study sample so that the scales have the ability to detect variability among and within individuals. Poorly targeted scales are likely to underestimate changes over time and differences between groups (274), which is relevant for future PU clinical trials that tend to recruit people with superficial PUs (123).

The issue of targeting could be improved with the addition of items to cover the construct measurement range at the extreme ends. Further work is required to determine whether some items would benefit from different response categories (i.e. dichotomising responses) or additional items, to make the scales psychometrically stronger, yet retaining their clinical meaningfulness.

The final PU-QOLI scales were also analysed using traditional methods to provide traditional psychometric evidence for the final scales, in line with current FDA guidelines (86). The methods and criteria used were the same as those used in the preliminary psychometric evaluation (i.e. post Rasch analysis and item reduction). The PU-QOLI was acceptable to patients, as the proportion of missing data was low. Tests of scaling assumptions mostly confirmed that items were grouped correctly. For example, corrected item-total correlations exceeded recommended criteria for all scales, however, item mean scores and variances were not especially similar. Whilst this implies that some criteria for scaling assumptions were not satisfied, it is important to note that the samples were not evenly distributed, with many people scoring at the lowest range of the scale, which may be the reason for this finding; notably, all item mean scores are below the mid-point range for each scale, supporting this explanation. Scale scores spanned the entire range of the scales however there were significant

floor effects for four scales and four were notably skewed. These findings indicate adequate scale-to-sample targeting but potentially poor item-to-sample targeting; the range mapped out by items in these scales was poorly matched to the range of variable in this sample. Standard criteria were effectively satisfied for internal consistency reliability as demonstrated with high alpha coefficients and item-total correlations. Test-retest ICCs were high for 8/10 scales. Two scales were below the recommended minimum (>0.70), however, only marginally (>0.60).

In addition, comparison with external measures were employed for assessing construct validity including the SF12 (333) and PU-QOLI overal QoL and overal pain items. The order of the battery of instruments administered to patients was the PU-QOLI adminsitered first, followed by the SF12, consistently throughout. No alternative order of administrations were tested, as this was beyond the remit of this thesis. It could be argued that presenting the questionnaire battery in this fashion may have resulted in order effects. Mixed findings have been reported when the effects of changing questionnaire order have been investigated; some have found significant order effects (344) while others found no differences in scores on any measure (345). Hence, the potential order effects on the PU-QOLI would need investigation.

Construct validity revealed satisfactory levels regarding within scales analysis and against external measures; the generated hypotheses were supported, thus providing evidence for convergent and discriminant validity. Ideally for the results of correlations between PU-QOLI scales and other scales to be fully interpretable the external measures should be reliable and valid. The psychometric properties of the SF12 have not previously been examined in PUs. Reasoning for selecting the SF12 was on the basis of its wide-spread use in chronic wounds and other dermatological conditions. However, it should be noted that conflicting results have been found when changes in HRQL due to PUs have been assessed using the SF36 and other generic instruments (109, 112). Generic instruments are intended for population comparisons on broad domain level, for a wide range of health conditions, and therefore may not be sensitive enough to detect the true impact of PUs on HRQL and provide information about the real changes within patients with PUs or differences among them. As such, additional evidence for convergent validity would provide further support for the PU-QOLI scales.

In general, the Classical psychometric evaluation provided the first evidence towards support for the acceptability, reliability and validity of PU-QOLI scales. The PU-QOLI scales meet FDA criteria for use in clinical trials (86) with patients with PUs at group level. However, the Rasch analysis demonstrated that some of the scales would benefit

from further work to improve their content and measurement properties and suitability for individual patient assessment. Consequently, further work is proposed (see Section 8.8) and any future use of the PU-QOLI scales should include further psychometric evaluation.

8.5 Further methodological outcomes of this research

8.5.1 Optimal retrieval methods for HRQL qualitative data

Part of this research included an evaluation of the effectiveness of qualitative methodology search strategies with subject-specific (HRQL) search strategies in the retrieval of qualitative PRO data of the impact of PUs on HRQL (Appendix 2.2). The subject-specific (HRQL) search strategy developed for this research identified all studies reporting qualitative HRQL data, whereas, research methodology-based (qualitative) strategies did not identify qualitative data reported in mixed method studies. Thus use of the HRQL search strategy maximised yield of relevant studies, making subject-based strategies more effective in retrieving the data of interest and the preferred method for comprehensive literature searching. However, for time efficiency, a broad-based qualitative-methodology search strategy is suitable for identifying most qualitative patient-reported HRQL research. Importantly it was found that in the HRQL field, qualitative data are reported in both qualitative and mixed methodology research and searching for this type of data involves trade-offs between yield, sensitivity and specificity. Accurate indexing of subject-specific outcomes and methodology used in electronic databases and publications is needed.

8.5.2 Mixed methods

This research was the first study to use Rasch measurement methods during the early stage of PRO instrument development. Rasch analyses provided useful information for scale development, specifically support for item -fit, ordering and validity by supporting patients' views that items made contextual sense and were relevant to them. This was further demonstrated by determining that items work together to define single variables and map out a discernable line of increasing intensity. This important contribution ensures that scale construction is underpinned with a strong conceptual base, a process central to valid measurement (3, 25). This work adopted an experimental approach that suits the Rasch paradigm, providing cross validation, for example, where patients reported problems with the PU-QOLI, the quantitative preliminary analysis verified qualitative comments made. One of the unique features of the Rasch model includes parameter separatability; where relatively robust estimates are possible with small sample sizes (13, 14, 15, 16), justifying inclusion of Rasch methods for pretesting

where small samples are often used. The ultimate goal was to provide an early quantitative view of the extent of potential problems/issues with PU-QOLI scales, however all findings needed to be verified and empirically strengthened in a larger sample as small sample sizes have implications for the interpretation of results.

Mixed methods, with the inclusion of a DIF analysis, were also used in a sub-study intended to determine optimal mode-of-administration for the PU-QOLI, providing both qualitative and empirical evidence for the selection of the most appropriate administration mode for the target population. This methodology allowed for consideration of characteristics response rate, data quality and potential for bias from sources other than non-response such as measurement equivalence between self-completed and interview-administered modes. Qualitative findings highlighted difficulty with patient self-completion. Investigation of response rate and data quality indicated no difference in response rate between mode groups but a difference was observed in data quality; a large proportion of PU-QOLIs were returned with missing data in the self-completed group, supporting the qualitative findings. Sub-group analysis indicated that the missing data was mainly observed in acute hospital patients who self-completed, suggesting a self-completed version might be inappropriate for acute patients but potentially feasible for community patients; there was no difference in response rate and data quality by mode for the community setting group.

The DIF observed was marginal thus providing preliminary evidence of stable item performance across mode groups. However, given the small sample size; DIF is a product of the sample and not the scale (85), the sample size was amended (inflated) and DIF re-examined. Increasing the sample size identified additional items with DIF, highlighting items to be cognisant of if pursuing a self-completed version in the future. Additional empirical evidence is needed for support of measurement equivalence between mode groups. As a significant proportion of missing data was found in the self-completed group, it was considered that a self-completed PU-QOLI would limit the people that could be assessed. In longitudinal research, this can be problematic as the progress of PUs and the impact they have on patients may not be accurately assessed due to high levels of missing responses on repeated measurement, hence the need for interview-administered methods. The Rasch/DIF analyses provided a complementary method alongside standard testing for examining optimal administration methods, with the intention of flagging issues with DIF for future examination.

8.5.3 Comparison of psychometric methods modern and classical

It is important to highlight that the traditional psychometric methods implied that the PU-QOLI scales are reliable and valid indicators of PU-impact and patient bother. However, the more sophisticated techniques employed in the Rasch analysis uncovered a number of inherent problems with some of the final PU-QOLI scales, specifically with the scoring function (e.g. participation scale) and scale-to-sample targeting that indicated limited measurement range at the extreme ends of scales (e.g. movement and mobility scale). Classical and Rasch methods are related in that they are both based on mathematical principles and used to evaluate rating scale's psychometric properties (e.g. the equation 0=T+E). There are equivalences such as the PSI and Cronbach's alpha for reliability estimates. However, Rasch extends on Classical methods by providing sophisticated tests intended for disclosing anomalies that may threaten valid measurement. Traditional methods are limited in the information they provide at item level, particularly about the adequacy of the response options and fail to provide specific guidance on how items might be improved. Rasch methods overcome these limitations as they are able to better diagnose specific issues surrounding the performance of rating scales (284, 346).

Statistical tests of fit provide stringent checks of the extent to which observed data satisfy measurement model requirements. Some misfit is to be expected, however the key focus of the analysis is not to only identify misfit but seek to explain why items initially hypothesised to belong to a common variable do not support that prediction (291). As was demonstrated here, reversed response categories and item misfit were detected, despite meeting traditional criteria for reliable and valid measurement. Reversed response categories indicate that the scoring function does not work as intended which is problematic as ordered thresholds are a fundamental necessity for establishing scale validity (85, 276, 305). Reversed thresholds can also contribute to item misfit. Some of the scales had small sample sizes and low frequency of response category endorsement, indicating further testing is required in a larger sample.

PRO instruments developed from a traditional perspective often include an item reduction stage, where items that do not meet psychometric criteria are removed from the scale (182, 294, 318, 330, 347). As already discussed, evidence from a Rasch analysis is used to understand why scale items might not be working, rather than used as a method to remove items from a scale. The evidence helps to establish which items are working/not working, and pin points where improvements could be made. For example, if the ordering of polytomous response options is not working as expected, categories with disordered thresholds could be collapsed. In practice it

might be preferred to have instruments with fewer items however, item reduction using classical methods is often detrimental to the measurement range and construct validity, as items at the floor and/or ceiling are often removed due to infrequent endorsement (85).

8.6 Methodological issues and study limitations

In the qualitative phase of PU-QOLI development, in-depth interviews were used to develop and refine the content of scales. Additional qualitative methods such as focus groups and interviews with the carers of people with PUs may have provided a further opportunity to marry findings. However, the qualitative interviews were continued to the point at which no new information emerged, ensuring that the conceptual framework adequately covered outcomes important to patients with PUs. Subsequent pretesting with patients confirmed the framework that emerged from the qualitative work, providing support for the adequacy of the qualitative method used in this thesis.

Traditionally, clinical intuition and experience have often been regarded as sufficient prerequisites for the construction of PRO instruments. Yet, PRO measurement in patients with PUs is complex, and there are number of measurement issues in studying this patient group. For example, the validity evidence for the PU-QOLI scales must be interpreted with some caution. This is because of the inherent difficulty in differentiating the impact of PUs on HRQL from patient comorbidity. Efforts were made during the qualitative work to elicit outcomes specific to PUs, but it may be difficult for some patients to think only of PU impact when they have a multitude of medical problems. Clarifying questions were used to try and extrapolate PU specific issues from patients' reports.

The qualitative work focused on PU bother and impact. Questioning was intended to elicit the worst aspect of having a PU and consequently assumptions were made about how PUs can impact all people with PUs. This line of questioning may have also missed valuable information about the patient's experience of when PU symptoms and other aspects are managed well. Better use of qualitative questioning would have included patients with healed PUs or severe PUs close to healed. These patients could have been asked about the entirety of their experience, with more thought given to covering the full spectrum of the PU experience. For example, questions pertaining to the patients' experience of when treatment was effective or the PU impact was milder/not at its worst, and words to describe the benefit/PU improvement. This may have helped with improving the measurement range by inclusion of items that

represented milder PU impact/bother. The patient transcripts could be revisited with this in mind or new qualitative work undertaken for further conceptualisation.

Each item within a scale is a component of the construct being measured and each construct (scale) is operationalised by the items (content) within it. To enable measurement of clinical variables, each scale requires a set of items that map out the construct they purport to measure. The aim of items is to mark out a measurement continuum for each construct on which people can be located or measured (87). However, consideration is needed of whether quantification of PU symptoms into multiple item scales is necessary. For scales where a measurement hierarchy is required to capture less of or more of a construct (e.g. difficulty with mobility), multiple descriptors are needed, but for exudate and odour, where a patient either has smell or leaking and is bothered by it, one item rather than multiple descriptors may be appropriate. For PU pain, the impact of pain is not necessarily related to the intensity therefore multiple descriptors may be required to differentiate between the types of pain a patient is experiencing (e.g. inflammatory pain versus neuropathic pain). Each PU-QOLI scale is intended to measure one construct and therefore multiple items represent a measurement continuum upon which patients can be located and monitored for changes. The legitimacy of including particular items in a scale was investigated using Rasch methods; principles that are based on measurement theory. Using Rasch methods allowed the assessment of whether each item within a scale contributed towards defining the construct measured. The analyses supported the items contained within the three symptom scales; however subsequent work could consider the selection of the most clinically appropriate items from each scale (see Section 8.8.3).

Key indicators of the quality of a PRO instrument are the reliability and validity of the measurement. A limitation of this study was that the psychometric properties of the final version of the PU-QOLI were estimated using data from a single study, where, due to many people scoring at the floor of the measurement range, small samples were often used for scale analysis (expanded Section 8.8.3, *Reanalysis*). Since the psychometric estimates are subject to sampling variation, it is possible that different items might have been selected if more data had been available. The samples were small for some scale test-retest which can influence the generalisability of the results. Evidence indicates that useful reliability estimates can be obtained from small samples (348, 349) but the scales would benefit from further examination of the measurement properties to ensure reliability. Further, as due to small sample sizes, the severe category 3 and 4 PUs were combined for the known groups tests, reanalysis is

required in larger samples to allow a true picture of how well PU-QOLI scales are able to discriminate between subgroups of patients known to differ in terms of clinical presentation. In addition, responsiveness - the ability of an instrument to accurately detect true change in the trait measured when true change has occurred - is of interest in healthcare applications where improvement in outcomes as a result of treatment is often a primary goal of research (350). Examination of responsiveness was not undertaken as it requires longitudinal research for testing (e.g. RCTs).

The validity testing of PU-QOLI scales was limited, in part, by a lack of appropriate validating measures and the ability to formulate hypothesises to enable known group difference testing. The literature is limited about the role PU severity, duration and location play in affecting HRQL outcomes. Such gaps in knowledge limit the ability to develop strong hypotheses to evaluate known group validity. Scale development is an on-going process. As the PU-QOLI scales are implemented and used in different samples, the estimates of validity, reliability and responsiveness may change as subsequent data is collected, and consequently various modifications to the scales may be needed. The usefulness of new measures is therefore demonstrated by multiple applications in different studies (accumulative body of evidence to support the measurement properties of new scales).

The PU-QOLI scales have the potential to be scientifically strong measurement tools however they are currently limited in their ability to detect differences in HRQL outcomes between people with different PU severity. Thus, in order for the scales to be valuable measurement tools for individual patients, these limitations need to be addressed. The targeting issues already discussed and the findings from the known group analysis suggest that some PU-QOLI scales may be limited in their ability to detect clinical change when it occurs and some items may underestimate differences in PU impact in people with superficial PUs. This has important implications for the inferences of future research using the PU-QOLI. Further evaluations are warranted to maximise the PU-QOLI as a measure of PU impact on HRQL in people with PUs.

Another consideration is that during field testing, as is standard practice, many patients would have received some form of treatment for their PU. This information was not collected (e.g. amount of analgesia). As such, the true impact of PUs may not have been captured (lower severity represented in the sample due to treatment effect) and be the reason for, at least in part, mistargeting and misrepresentation of known groups testing. In actual fact, PUs appear to cause patients more bother than was

represented in the sample (as indicated from the qualitative work) but because of good care received, lower PU impact was demonstrated in the sample.

Given the heterogeneity of the population with PUs, further work is required to ensure that the PU-QOLI scales fit the needs of all people with PUs including patients with superficial PUs. One tool is intended for effectiveness intervention research where patients with all PU severity might be included in research (e.g. a mattress trial where both improvement and deterioration in HRQL needs to be assessed) and differences in HRQL are explored by severity. In research and clinical practice, information pertaining to treatments received can be collected and accounted for when analysing PU-QOLI data for changes within patients. The intention is to flag issues, promote patient-clinician communication, and help guide decisions on patient treatment and PU management. The next step of this research will investigate the sensitivity of the PU-QOLI scales to change, accounting for any treatment received.

The PU-QOLI was developed in adult patients with PUs of any location, duration or severity, receiving any PU-specific intervention from primary and secondary care. Until further validation, PU-QOLI scales are not considered valid for patients that were not represented in the development process. Despite being a multi-centre study including 34 centres around the UK from various settings and specialities, there were a disproportionate number of different ethnicities represented in this sample. Convenience sampling resulted in recruitment of predominantly British-white nationals. Therefore the results of this study may not be generalisable to all patients with PUs in the UK. More research is needed to determine suitability for use of PU-QOLI scales with non-white ethnic groups. Further, it was not possible to include patients with cognitive impairment due to the unavoidable difficulty of obtaining informed consent from these patients.

Another limitation of this study was that items with DIF were not accounted for. A post-hoc qualitative evaluation of the possible reasons for DIF are recommended (303, 351), however were beyond the scope of this research. At this stage the intention was to develop good scales (item level statistics) and examine some key clinically reasonable variables with the intention of flagging any potential issues with DIF for generating hypotheses/experimental tests for future investigation. DIF was marginal in this study and therefore not worrying at this stage (scale formation stage of the research). However, as DIF is a product of the sample and not the scale, the potential issues with DIF might need to be accounted for psychometrically in subsequent evaluations.

On a similar note, misfitting persons were not accounted for. To ensure good quality data is collected when obtained using PU-QOLI scales, person fit and item responses should be considered. As the intention was to develop good scales, only item level statistics were considered at this stage. Once scales demonstrate valid and reliable measurement, then problematic person data can be considered and interpreted. The Rasch model can be used to identify person misfit by identifying individual persons' response patterns across scale items (i.e. responses to each item for misfitting persons are reviewed in an attempt to highlight any unexpected responses), and determining whether responses are inconsistent with expectation, hence misfitting. The pattern of responses invokes enquiry as to why such an unlikely pattern of responses occurred. The cause of the misfit requires further qualitative exploration of the individuals (85).

Mixed methods were used throughout this work but as new ideas were explored in this thesis, the work was often undertaken with little guidance. Consequently, a range of exploratory work was carried out including use of mixed methods during pretesting and investigation of optimal administration methods. Mixed methods research was challenging in that each has established methods that often constrained the work in this thesis. For example, current guidance recommends a three step process to instrument development including early qualitative work to generate constructs and content, pretesting and psychometric evaluation. This model was followed but proved suboptimal for PU-QOLI development. An interactive, iterative approach, with better use of qualitative methods would have been beneficial. Both qualitative and empirical findings should be used to inform subsequent work and to make improvements to new scales. The findings suggest that current scale development guidelines may need to be revised.

8.7 Implications of findings

This research stemmed from a need for precise and comprehensive PRO measures for research and clinical practice to assess important outcomes for patients with PUs. The use of *ad hoc* measures without a theoretical basis or psychometric testing in much PU research motivated the work in this thesis. In the first instance, appropriate assessment of outcomes in healthcare requires inclusion of PRO assessment. In developing the PU-QOLI, HRQL conceptual domains important to patients with PUs were identified. Elucidation of conceptual domains important in PUs provides a useful framework for designing future research, and consequently improving the quality of research in the PU field by inclusion of PU-specific PROs. Implications for clinical practice and future PU research are considered below.

8.7.1 Clinical practice

There are many potential benefits for using PRO measures in clinical practice, including: facilitation of clinician-patient communication and shared decision making; identifying and prioritising patient problems; screening for hidden problems; identifying patient preferences; monitoring changes or outcomes of treatment; training new staff; measurement of the performance of healthcare providers; and clinical audit (113, 352, 353, 354, 355, 356). The PU-QOLI potentially provides a means for comprehensive assessment of the impact of PUs, and a way of quantifying the benefits of PU interventions from the patient's perspective; a recognised essential part of healthcare evaluation (357, 358). Thus, subject to further work, PU-QOLI data could provide an important source of information for supporting patient-focused decision making, provide a PRO measure for intervention and evaluation research, increase understanding of the impact of PUs on individuals, and ultimately result in adjustments in care delivery to meet patient requirements. This is particularly important for changing practice through mandated NICE guidance, where the perceived value of PU interventions and evaluating PROs associated with treatment and relative PU burden must have a robust evidence base (86, 120, 359) and can help inform decisions about the most appropriate management of PUs.

Importantly, trials of PU intervention clinical effectiveness have not used the highest quality PROs (122, 123, 360). The selection of appropriate outcome measures underpins the meaningful interpretation of study results (i.e. data can only be meaningful if the instruments used to collect the data are valid and reliable and appropriate to address the research question) (87, 361). Further, the assessment of treatment benefits as perceived by patients is particularly important as they may differ from clinically important outcomes. Hence, a range of carefully selected and complementary outcomes may be needed for clinical practice for the measurement of benefits from health expenditure and in the assessment of the management and delivery of healthcare in the PU field.

It may be premature to expect that the PU-QOLI will rapidly gain acceptance for clinical practice. In practice, the focus is on PU prevention or treatment that ultimately results in healing, rather than assessment of PROs. PU management should involve multidisciplinary care plans with consideration of all factors contributing to and affecting the wound as well outcomes important to patient. However, concerns about PU-QOLIs relevance to patients with superficial PUs and its ability to accurately detect differences in people known to differ on clinical presentation, suggests that PU-QOLI scales are currently not suitable for clinical use. Further work is needed to improve the

measurement precision and psychometric properties before it is used for the assessment and management of individual patients in clinical practice.

8.7.2 Research

Quality outcome instruments are the cornerstone of clinical research. The findings support the PU-QOLI as valid and reliable according to FDA criteria for use in clinical trials (86), however the Rasch analysis findings highlighted that further work is required before the PU-QOLI scales can be used as the main PRO measure in future clinical trials or other research. PU-QOLI scales could be included as one outcome measure amongst others in future PU research (e.g. clinical trials, observational or epidemiological studies, audit and service evaluation) on the proviso that studies have built in a parallel psychometric analysis to indicate the performance (psychometric evaluation) of the scales in future samples. Currently, the PUQOLI scales are most appropriate for patients with severe PUs, as demonstrated by a lack of items to represent people with little or no bother due to PUs. The exudate and odour scales are not intended for people with superficial PUs (<category 2 PU). Electronically defined skip questions would assist in the selection of scales and items that are relevant to each patient's circumstance. Appropriateness for PU-QOLIs use in individual decision-making requires further investigation.

The PU-QOLI and practical information necessary for administration and scoring in a user's manual will be made freely available via the University of Leeds, CTRU website. The PU-QOLI should be administered using the manual. Due to limitations to PU-QOLIs measurement properties, the manual will also include information to potential users about the purposes for which the PU-QOLI should and should not be used.

Individual scales can be selected depending on the nature of the research and the research question of interest (individual scale scores rather than an overall HRQL score)Imputation of missing data, based on methods undertaken for scoring the SF36 (95), can be undertaken provided at least 50% of items are complete for an individual. For these cases missing data can be imputed using the person mean substitution (the mean scale score divided by the completed items) (100, 362). Cases with less than 50% data complete are excluded. There are some limitations to imputation (i.e. making assumptions about unknown item values, bias in estimating values (363)) but these can be overcome by providing a scoring algorithm for all PU-QOLI scales; this method would not require imputation of missing data (expanded in Section 8.8.3, *Clinical application*). Questions 3 (overall pain), 4 (itchiness) and 13 (overall QoL) (Appendix 6.1) are not included in scale scores but provide descriptive information for clinical

decision making. All scale items are scored 0 (no bother) to 2 (a lot of bother). Scale items are summed to produce a total scale score. A lower score indicates better outcome.

It is increasingly important to demonstrate the value of healthcare interventions within the cost effectiveness framework used by NICE. To date, PU interventions have struggled to demonstrate their value within this framework due to the lack of appropriate health outcome measures. NICE guidelines (54, 55, 364) highlight that in the PU field resource availability is not based upon health economic evaluation and there is no systematic way of considering patients priorities for interventions. So whilst efforts have been made to develop evidence-based practice, a major limitation of national guidelines is that they are largely based upon consensus and not an evidence-base. In addition, PUs usually develop as a secondary medical problem in a wide case-mix of patients. Due to the multitude of medical problems these patients often face, issues surrounding resource allocation and priority setting arise. Data from the PU-QOLI could be used to improve knowledge of different PU treatments (i.e. comparison of different wound management approaches and 'end results' such as patients' perceived benefit of treatment) and of the health problems faced by patients with PUs. This information could then be used (in part) as a basis for determining efficient allocation of resources for healthcare. Further, a mapping of PU-specific constructs to a utility measure would allow cost-effectiveness analysis (expanded Section 8.8.1, *Use in economic evaluation*). The use of rigorous outcome measures will provide evidence-based information to allow health authorities to select the most effective healthcare for patients and to audit and monitor the quality of care given.

During the last decade, the NHS has become more interested in audit, and have started the routine collection of PRO data for all patients before and after receiving NHS-funded care via its PRO measures initiative (365). Routine audit in PUs using the PU-QOLI could assist clinicians in monitoring the outcomes of care. This data could be part of a dataset collected in addition to other clinical data, as has already proven possible by the NHS PRO measures initiative. Such data can enhance evidence-based policy and help to inform treatment guidelines in the PU field, which are of interest to government bodies such as NICE. The use of such data in conjunction with other treatment outcome and cost information will help determine the cost effectiveness of different methods of PU care and help prioritise patient care according to the impact on patient's HRQL. This is pertinent as it is increasingly necessary to carry out multi-centre clinical trials to demonstrate benefits gained from various treatments and interventions.

Finally, using a PU-specific instrument may provide a more appropriate comprehensive assessment of important outcomes that compliment the information generated by generic measures. Importantly, generic measures do not include, for example, the assessment of PU symptoms or limitations of movement such as transferring that are specific consequences of PUs; thus useful clinical indicators for tissue viability specialists. In addition, the methods used to develop and evaluate the PU-QOLI could be applied to other PRO instruments. Uniformity of research approaches used to develop PRO measures could lead to consistency in health measurement and the inclusion of Rasch measurement methods in accepted international guidelines.

8.8 Future research

8.8.1 Further development and evaluation of PU-QOLI

The PU-QOLI was found to be acceptable, reliable and valid in line with FDA criteria (86), supporting the applicability for use with patients with PUs in future research. However, issues with item response categories and sample-to-scale targeting emerged which suggest that some of the scales require further work to improve the top end of the measurement range and subsequent empirical demonstrations in well targeted samples. Inspection of threshold distributions demonstrated sub-optimal targeting for most scales (items did not span the full range of the patient sample). However, scale scores for >65% of the samples were within the best performing part of the scales. For example, the pain scale items spread 2-logits compared to a person spread of 7 logits, indicating suboptimal targeting. But for the vast majority of people in the sample, the pain scale performed well as the measurement range distribution was within the range where most people lay. This work has been the first step to producing PU-specific scales. The final Rasch analyses provide an initial evidence-base for future testing to improve the PU-QOLI scales and to establish the extent that psychometrically sound scales have been developed. Small samples are considered adequate for a Rasch analysis if there is good targeting (261), however future studies utilising larger samples should be undertaken using Rasch analysis to confirm this study's findings. Despite attempts to sample the broad PU population, including a wide variety of patients with PUs drawn from different settings, a problem with targeting emerged.

The problem of targeting could be improved by developing PU-QOLI scale items that span a wider range of measurement. Although item-level floor and ceiling effects are likely to exist to some extent, attempts should be made to minimise them in order to maximise the potential of the PU-QOLI to detect change. Three scales had large floor effects, implying that it may be beneficial to extend their measurement range in the

future. This can be achieved without affecting the scales as they stand, because the item locations are calibrated relative to each other. This work was the first step towards establishing item estimates for each PU-QOLI scale. Future scale developments can be empirically driven as the distribution of item locations highlight where 'gaps' in the measurement continuum are (notable distances in item locations could be filled with items, particularly those representing superficial PU impact), and the distribution of person measurements indicate that it may be valuable to extend the measurement range at the extreme ends of the continuum. Subsequent work would extend this work and move towards establishing better precision and construct definition (cumulative data will allow for establishing better precision of item estimates or finding empirical support for the ones established in this study). Qualitative work would be needed to explore the addition of items to extend the measurement range or more PU-QOLI data for item calibration to confirm the fixed item estimates.

Another consideration is that the scales produced are hypotheses of how each variable could be measured in PUs. So despite the finding of narrow measurement ranges for some scales, in reality this may be true in PUs. Put another way, the true measurement range for each variable is actually captured and represented by each PU-QOLI scale but in this instance the sample was mistargeted. The finding is that the targeting was not optimal, not necessarily the item estimates; in this sample this might be the reality. Noteworthy, during development, scale construction was carefully thought-out (content of each scale) and underpinned by sound qualitative work that included patients at every stage of the development process, which led to the produced hypothesised scales. Finally, further evaluation of the response format of the scales should be undertaken as some items indicated problems with the scoring function. The scoring function in this study was assessed in small samples, therefore requires further examination to guide decisions about whether rescoring is appropriate.

In the short-term, to begin exploring some of the measurement issues outlined, it is proposed that data from both field tests will be merged and the psychometric properties reanalysed. Data from both field tests will increase the sample size and provide accumulating evidence for the appropriateness of PU-QOLI scales and items. Further psychometric evaluation and replication is a basic requirement for any new instrument, and although all the results require replication, test-retest reliability in particular should be evaluated in a larger sample, and further consideration should be given to the targeting and response option categories.

Evaluation of responsiveness

Responsiveness of the PU-QOLI needs to be evaluated in future research. Determining PU-QOLIs ability to accurately detect true change is essential if it is to be used to evaluate PU interventions. PRO instruments need to be able to detect the effects of treatment on outcomes and changes in outcome over time. Further, the effect of PU interventions on HRQL is largely unknown. It is intended that responsiveness will be evaluated in a future PU mattress clinical trial (PRESSURE II).

Proxy and cross-cultural PU-QOLI

PU-QOLI scales are not considered valid for patients that were not represented in the development process. To address this limitation, development of condition-specific modules (e.g. SCI-specific) and proxy measures (to enable patients with cognitive impairment to be included in research) are recommended. The PU-QOLI is not for use with children (under 18 years) or those with cognitive impairment. People with cognitive impairment are often excluded from PU research studies, despite the considerable number of people with cognitive impairment that develop PUs. This presents a major challenge to health and social services. During field testing, approximately 40% of patients screened for study participation were assessed as cognitively impaired. Given the high percentage of patients with cognitive impairment, outcome measures are needed to enable inclusion of these patients in future research as well as for evaluations of PU interventions to widen the potential use of the PU-QOLI. Such work would require the development of a proxy measure intended for patients who are cognitively impaired or who lack capacity. The FDA guidance identifies that some patients over the course of a clinical trial may become too ill to complete a questionnaire or respond to an interview, suggesting that proxy reporting may help to prevent missing data. Qualitative work with carers of people with PUs is needed to determine whether the perspectives of carers are consistent with the perspective of people with PUs. If so, the current PU-specific conceptual framework provides a useful basis for the development of PU-QOLI-proxy. Work is needed to assess the relationship between patient and proxy reports. An alternative to a proxy measure may be validating the PU-QOLI for carer-assisted mode of administration

The PU-QOLI was developed in a UK population. Relatively few participants in this study were from minority ethnic groups. Given the cultural diversity in the UK there is a need to be able to judge the effectiveness of interventions in people with PUs from minority ethnic groups. Moreover, there is need for PRO instruments that can be used in cross-national studies such as multi-country and multi-centre RCTs. This requires linguistic validation studies on language translations and cross-cultural adaptation of

the PU-QOLI for use in other groups and cultures, particularly as patient perceptions are not independent of their cultural environment.

Use in economic evaluation

There is need for a PRO instrument for use in economic evaluations of interventions for people with PUs. Work is planned, as part of the larger NIHR PU Programme of Research, to use the PU-QOLI to generate health scenarios for the development of a preference-based measure specific to PUs for use in economic evaluation.

8.8.2 Methodological research

Feasibility of self-completed PU-QOLI for community patients

The mixed methods work undertaken in this thesis provided evidence for the feasibility of a self-completed version for community patients. However, as the study was not powered accordingly (n=33 community patients), more work is needed to confirm appropriateness. Nonetheless, the results provide positive preliminary evidence for a community self-completed version and items where DIF is reaching significance were identified. Parallel use of qualitative methods to determine reasons for DIF may allow these issues to be accounted for (i.e. adapt/improve items) if pursuing a self-completed version in the future.

Prognostic modelling

The PU-QOLI can be included in clinical research evaluating PU interventions, wound care management and healthcare services; subject to further replication and demonstration of responsiveness. In addition, there has been little study of mediating factors associated with HRQL in PUs. The PU-QOLI will allow for the investigation of how HRQL is affected by other factors related to the person with PUs (e.g. severity, location and duration of PUs, motivation and other psychosocial factors), their family and/or carer (e.g. nature and quality of relationship), and factors associated with experience of care (e.g. satisfaction with treatment, perceived professional competence). There is also a lack of prospective studies measuring changes in HRQL over time. Such studies would allow for the identification of factors predicting higher and lower HRQL. At group level, longitudinal studies are needed to enable the interpretation of scores, in particular the meaning of change scores. Further work may be needed to elaborate on particular aspects of wound care management and skincare following healing in order to provide a richer picture of the impact of PUs on HRQL. Finally, unlike other chronic wounds that may be the principal health problem (i.e. leg ulcers), PUs are usually a consequence of other comorbidity associated with mobility

impairment (e.g. stroke, multiple sclerosis, dementia). Little work has been done to evaluate HRQL in these sub-groups.

Computer adaptive testing

There is growing interest in, and use of, computer technology in the collection of PRO data. Computer adaptive testing (CAT) can help minimise routine errors and omissions that can occur when asking questions, recording responses and following complex skip instructions (253). The concept of CAT is based on presenting the respondent with only the most relevant items to his/her ability or disability level (item presentation is guided by the respondent's response to previous items) (366), thus minimising the number of items required to obtain a reliable measure on a particular trait or variable (367). Despite benefits of using CAT in healthcare (e.g. reducing respondent burden), set-up costs associated with purchasing hardware and programming computers may be prohibitive for use by many researchers and application in clinical settings.

8.8.3 Next steps

Reanalysis

For strengthening the psychometrics and decisions made to PU-QOLI items and scales, reanalysis of PU-QOLI scales from data combined from field tests one and two is proposed. In a Rasch analysis, people at the extremes of the scale range (those at the floor/ceiling) are excluded from the estimation of item statistics as they offer no comparison across the items to facilitate the examination of relative item difficulty (85). This is because people at the extremes of the scale range achieve the same score on all the items of the scale. In the PU-QOLI analysis, people at the extremes were excluded thus the sample for evaluating item statistics were computed from only n = 27 to 184 range. Combining the samples would increase the sample size.

The presence of many extreme scores can influence the variability across the range of the scale and influence the targeting. For example, for exudate and odour symptoms, patients with superficial PUs are not expected to have problems with these symptoms; only severe PUs where skin breakdown occurs is attributed with exudate and odour. As the sample consisted of patients with both superficial and severe PUs, approximately half the sample would not be expected to have problems with these symptoms. As these patients were included in the analysis, the samples produced poor targeting with many patients considered extremes. For the next PU-QOLI version, consideration of the addition of screening questions such as "have you had any problems with smell or odour because of your PU in the past week?" (If yes, complete the questions below; if no, move onto the next question) will be made.

Clinical application

A pertinent question for tissue viability specialists is the appropriateness of using the PU-QOLI for individual clinical decision-making. The intention is to develop the PU-QOLI as a clinical intervention tool intended for targeted treatment. This may require the development of a short-form version of the PU-QOLI, where the most clinically appropriate items from each scale are selected (342). Like in other clinical areas (113, 368), it would be used to facilitate patient-clinician communication and assist with priority setting. Importantly, completion of PU-QOLI prior to meeting with a tissue viability specialist would allow for issues most important (or bothersome) to patients to be discussed during the consultation and direct treatment decisions that consequently improve HRQL outcomes for patients. It is advised that the development work would include the construction and testing of a scoring algorithm programme to enable wide use of the PU-QOLI to inform clinical management of individual patients across different clinical settings and ease the interpretation of scale scores (368).

Translation and validation

Interest has been expressed from members of the EPUAP regarding PU-QOLI translation into various European languages however uncertainty about which methods to use to undertake this work were stressed. It is proposed that as part of the PU-QOLI user manual, guidelines for language translation will be included based on internationally accepted methods (86, 102, 369).

8.9 Summary and Conclusions

The thesis aimed to assess the applicability of developing a new PRO instrument for people with PUs. A review of the literature revealed that few studies measured HRQL outcomes related to PUs and treatment. If assessed, generic or chronic-wound specific PROs were used but the outcomes important in PUs were inadequately covered despite similarities between conceptual models. This suggested the need for an easy, standardised method for assessing health outcomes important to patients with PUs. Given the limitations of the PU research literature, this work was the first to conceptualise the impact of PUs on HRQL from the patients perspective, and develop a PRO instrument that captures issues important to patients using development and evaluation methods accepted and applied in the wider health measurement field. In addition, various methodological developments were woven into the research. They highlighted that the traditional three stage approach may not be appropriate for producing scientifically sound scales in all clinical areas. What was demonstrated was that small iterative steps, using mixed methods in an interactive way, particularly at early content and scale format/design development, is needed.

A PRO instrument for people with PUs (the PU-QOLI) was developed and validated. It includes 10 outcomes of PU symptoms, limitations of physical functioning, psychological functioning, and social participation. This thesis demonstrates that PUs impact on HRQL and the PU-QOLI provides a method for evaluating HRQL in PUs. The PU-QOLI is intended for interview-administration, following a user manual, and is currently appropriate for use in adults with severe PUs of any location or duration, and suitable for use in all UK healthcare settings. Individual scales can be selected for use, depending on the nature of the research and scale items can be summed to produce scores, without weighting or standardisation.

Longitudinal studies should be undertaken to assess the responsiveness of the PU-QOLI over time, as clinical studies evaluating the efficacy of various PU treatments and interventions require accurate detection of true change. Further research is also needed to: confirm the study findings in an independent sample and the generalisability of the findings from this study; investigate the feasibility of use in specific subgroups, economic evaluation; and the development of proxy measures and language translations given the prevalence of cognitively impaired patients with PUs and the high PU prevalence worldwide. Long-term goals include developing the PU-QOLI as a clinical tool intended for individual-patient decision making and applying the methodology used in this thesis to other medical areas. This could lead to the development of guidelines that include mixed methods as well as the more sophisticated psychometric methods, such as Rasch measurement. This work was the first step towards establishing measurement precision but additional research is needed to improve some of the measurement properties of PU-QOLI scales.

This study makes important contributions to the PU and wider health measurement fields. The findings demonstrate that mixed methods, including Rasch measurement methods were a suitable approach to developing a new PRO instrument specific for PUs (the PU-QOLI); a methodology that can be applied for further development of the PU-QOLI as well as PROs in other health areas. The PU-QOLI provides a means for the comprehensive assessment of PU impact and for quantifying the benefits of PU interventions from the patient's perspective, thus far lacking in the area. And finally, PRO measurement needs to become more common place in the PU field so that the goal of PU management can be to enhance and maintain the HRQL of people with PUs. Subject to further development as detailed above, PU-QOLI is a tool with which to evaluate whether PU treatments and the healthcare given achieve this; outcomes that are ultimately best judged by patients themselves.

References

1. National Pressure Ulcer Advisory Panel and the European Pressure Ulcer Advisory Panel (NPUAP/EPUAP). (2009). Prevention and Treatment of Pressure Ulcers: Clinical Practice Guideline. 16-20

2. Bridel, J. (1993). The epidemiology of pressure sores. *Nursing Standard*, 42, 25-30

3. Kaltenthaler, E,. Whitfield, MD,. Walters, SJ,. Akehurst, RL. & Paisley, S. (2001). UK, USA and Canada: how do their pressure ulcer prevalence and incidence data compare? *Journal of Wound Care*, 1, 530-535

4. Thomas, DR., Goode, PS., Tarquine, PH & Allman, RM. (1996). Hospital-acquired pressure ulcers and risk of death. *Journal of the American Geriatrics Society*, 12, 1435-1440

5. Allman, RM., Goode, PS., Patrick, MM., Burst, N & Bartolucci, AA. (1995). Pressure ulcer risk factors among hospitalized patients with activity limitation. *JAMA*, 11, 865-870

6. Bennett, G., Dealey, C & Posnett, J. (2004). The cost of pressure ulcers in the UK. *Age & Ageing*, 3, 230-5

7. DOH. (2001). National service framework for older people. Department of Health document, London

8. DOH. (2001). Essence of care: Patient-focused benchmarking for health care practitioners. Department of Health document, London

9. Spilsbury, K., Nelson, A., Cullum, N., Iglesias, C., Nixon, J & Mason, S. (2007). Pressure ulcers and their treatment and effects on quality of life: hospital inpatient perspectives. *Journal of Advanced Nursing*, 5, 494-504

10. Gebhardt, K. (1995). Tissue viability. What causes pressure sores? *Nursing Standard*, 31, 48-51

11. Mustoe, T. (2004). Understanding chronic wounds: a unifying hypothesis on their pathogenesis and implications for therapy. *The American Journal of Surgery*, Suppl May, S65-S70

12. Halfens, RJ., Bours, GJ & Van, Ast W. (2001). Relevance of the diagnosis 'stage 1 pressure ulcer': an empirical study of the clinical course of stage 1 ulcers in acute care and long-term care hospital populations. *Journal of Clinical Nursing,* 10(6):748-57

13. Nixon, J., Brown, J., McElvenny, D., Mason, S & Bond, S. (2000). Prognostic factors associated with pressure sore development in the immediate post-operative period. *International Journal of Nursing Studies*, 4, 279-289

14. Nixon, J., Cranny, G & Bond, S. (2007). Skin alterations of intact skin and risk factors associated with pressure ulcer development in surgical patients: a cohort study. *International Journal of Nursing Studies,* 44(5):655-63

15. Safeguading Vulnerable Groups Bill, HL bill 79 (2006)

16. Byers, PH., Carta, SG & Mayrovitz, HN. (2000). Pressure ulcer research issues in surgical patients. *Advances in Skin & Wound Care*, 3 Pt 1, 115-21

17. Nixon, J., Thorpe, H., Barrow, H., Phillips, A., Nelson, EA., Mason, SA & Cullum, N. (2005). Reliability of pressure ulcer classification and diagnosis. *Journal of Advanced Nursing,* 50(6):613-23

18. Moore, Z. (2005). Pressure ulcer grading. *Nursing Standard*, 52:56-64

19. NICE. (2003). *CG7 Pressure ulcer prevention.* Retrieved from the NICE, UK website: www.nice.org.uk

20. Brandeis, GH., Wee, LO,. Hossain, M., Morris, JN & Lipsitz, LA. (1994). A longitudinal study of risk factors associated with the formation of pressure ulcers in nursing homes. *Journal American Geriatrics Society,* 42:388-393

21. Defloor, T. & Grypdonck, MF. (2005). Pressure ulcers: validation of two risk assessment scales. *Journal of Clinical Nursing*, 3, 373-382

22. NICE. (2001). *Pressure ulcer risk assessment and prevention.* Retrieved from the NICE, UK website: www.nice.org.uk

23. Nixon, J. & McGough, A. (2001). Principles of Patient Assessment: Screening for Pressure Ulcers and Potential Risk. *In The Prevention and Treatment of Pressure Ulcers*, 2 ed. M. Morison, ed., Edinburgh: Mosby, 55-74.

24. Nixon, J., Cranny, G., Iglesias, C., Nelson, EA., et al. (2006). Randomised, controlled trial of alternating pressure mattresses compared with alternating pressure overlays for the prevention of pressure ulcers: PRESSURE (pressure relieving support surfaces) trial. *BMJ*, 7555, 1413-1416

25. Perneger, TV., Rae, AC., Gaspoz, JM., Borst, F., Vitek, O & Heliot, C. (2002). Screening for pressure ulcer risk in an acute care hospital: Development of a brief bedside scale. *Journal of Clinical Epidemiology,* 55(5), 498-504

26. Reed, RL., Hepburn, K., Adelson, R., Center, B & McKnight, P. (2003). Low serum albumin levels, confusion, and fecal incontinence: Are these risk factors for pressure ulcers in mobility-impaired hospitalized adults? *Gerontology*, 4, 255-259

27. Schultz, A., Bien, M., Dumond, K., Brown, K & Myers, A. (1999). Etiology and incidence of pressure ulcers in surgical patients. *AORN Journal 70(3):434, 437-40, 443-9,* 3, 437-440

28. Allman, RM., Laprade, C. A & Noel, LB. (1986). Pressure sores among hospitalized patients. *Annals of Internal Medicine*, 3, 337-342

29. Clough, NP. (1994). The cost of pressure area management in an intensive care unit. *Journal of Wound Care*, 1, 33-35

30. Bale, S., Finlay, I & Harding, KG. (1995). Pressure sore prevention in a hospice. *Journal of Wound Care*, 10, 465-8

31. Bergstrom, N., Braden, BJ., Laguzza, A & Holman, V. (1987). The Braden scale for predicting pressure sore risk. *Nursing Research*, 4, 205-210

32. Gosnell, DJ. (1973). An assessment tool to identify pressure sores. *Nursing Research*, 1, 55-59

33. Lowthian, P. (1989). Identifying and protecting patients who may get pressure sores. *Nursing Standard*, 4, 26-29

34. Norton, D., McLaren, R & Exton-Smith, AN. (1962). *An investigation of geriatric nursing problems in hospital.* Edinburgh:Churchill Livingstone

35. Pritchard, V. (1986). Calculating the risk. *Nursing Times*, 8, 59-61

36. Towey, AP. & Erland, SM. (1988). Validity and reliability of an assessment tool for pressure ulcer risk. *Decubitus*, 2, 40-48

37. Waterlow, J. (1988). The Waterlow card for the prevention and management of pressure sores: towards a pocket policy. *Care Science and Practice*, 1, 8-12

38. NHS Centre for Reviews and Dissemination. (1995). *The prevention and treatment of pressure sores: How useful are the measures for scoring people's risk of developing a pressure sore?* Effective Health Care Bulletin, 2(1): University of York

39. Altman, DG. (2001). Systematic reviews in health care: Systematic reviews of evaluations of prognostic variables. *British Medical Journal*, 7306, 224-228

40. Cullum, N., Deeks, JJ., Fletcher, AW., Sheldon, TA & Song, F. (1995). Preventing and treating pressure sores. *Quality in Health Care*, 4, 289-97

41. Defloor, T. & Grypdonck, MF. (2004). Validation of pressure ulcer risk assessment scales: a critique. *Journal of Advanced Nursing 48(6):613-21,*

42. Margolis, DJ,. Bilker, W,. Knauss, J,. Baumgarten, M. & Strom, BL. (2002). The incidence and prevalence of pressure ulcers among elderly patients in general medical practice. *Annals of Epidemiology*, 321-325

43. Beckrich, K & Aronovitch, SA. (1999). Hospital-acquired pressure ulcers: a comparison of costs in medical vs. surgical patients. *Nursing Economics*, 263-71

44. European Pressure Ulcer Advisory Panel (EPUAP). (2005). *Scope of pressure ulcer prevention guidelines.* Retrieved http://www.epuap.org/glprevention.html

45. National Pressure Ulcer Advisory Panel (NPUAP). (2001). Pressure ulcers in America: Prevalence, incidence, and implication for the future. Retrieved http://www.epuap.org website

46. Lahmann, NA,. Halfens, RJ. & Dassen, T. (2006). Pressure ulcers in German nursing homes and acute care hospitals: prevalence, frequency, and ulcer characteristics. *Ostomy Wound Management*, 52:20-33

47. Meehan, M. (1990). Multisite pressure ulcer prevalence survey. *Decubitus*, 4: 14-7

48. Whittington, KT. & Briones, R. (2004). National Prevalence and Incidence Study: 6-year sequential acute care data. *Advances in Skin & Wound Care*, 9: 490-494

49. Brandeis, GH,. Morris, JN,. Nash, DJ. & Lipsitz, LA. (1990). The epidemiology and natural history of pressure ulcers in elderly nursing home residents. *JAMA*, 22:2905-9

50. Santamaria, N., Carville, K., Prentice, J., Ellis, I., Ellis, T., Lewin, G & Newall, N. (2005). Pressure ulcer prevalence and its relationship to comorbidity in nursing home residents: results from phase 1 of the PRIME Trial. *Primary Intention*, 13(3):107

51. Bergquist, S. (2003). Pressure ulcer prediction in older adults receiving home health care: implications for use with the OASIS. *Advances in Skin & Wound Care*, 3:132-9

52. Clark, M. & Watts, S. (1994). The incidence of pressure sores within a National Health Service Trust hospital during 1991. *Journal of Advanced Nursing*, 1:33-36

53. Graves, N., Birrell, F. A & Whitby, M. (2005). Modeling the economic losses from pressure ulcers among hospitalized patients in Australia. *Wound Repair & Regeneration,* 13(5):462-7

54. DOH. (1993). Pressure sores - a key quality indicator: a guide for NHS purchasers and providers. Department of Health document, London

55. NICE. (2005). *CG29 Pressure ulcer management.* Retrieved from the NICE, UK website: www.nice.org.uk

56. McInnes, E,. Jammali-Blasi, A,. Bell-Syer, SEM,. Dumville, JC. & Cullum, N. (2011). Support surfaces for pressure ulcer prevention. *Cochrane Database of Systematic Reviews*, 4

57. Dealey, C. (2007). Review: support surfaces, nutritional supplements, and topical agents help prevent pressure ulcers. *Evidence-Based Nursing,* 10(2): 54

58. European Pressure Ulcer Advisory Panel (EPUAP). (2006). *Pressure ulcer treatment guidelines.* Retrieved http://www.epuap.org/glprevention.html

59. Whitney, J., Phillips, L., Aslam, R., Barbul, A., Gottrup, F., et al. (2006). Guidelines for the treatment of pressure ulcers. *Wound Repair & Regeneration,* 14(6):663-79

60. Cullum, N., Nelson, EA., Flemming, K & Sheldon, T. (2001). Systematic reviews of wound care management: (5) beds; (6) compression; (7) laser therapy, therapeutic ultrasound, electrotherapy and electromagnetic therapy. *Health Technology Assessment*, 9

61. Bradley, M., Cullum, N., Nelson, EA., Petticrew, M,. Sheldon, T & Torgerson, D. (1999). Systematic reviews of wound care management: (2) Dressings and topical agents used in the healing of chronic wounds. *Health Technology Assessment*, 17

62. The Cochrane Collaboration. (2009). *The Cochrane Library.* Retrieved http://www.thecochranelibrary.com/view/0/index.html

63. Padula, CA,. Osborne, E. & Williams, J. (2008). Prevention and early detection of pressure ulcers in hospitalized patients. *Journal of WOCN*, 1, 65-75

64. Eisenberger, A. & Zeleznik, J. (2003). Pressure ulcer prevention and treatment in hospices: a qualitative analysis. *Journal of Palliative Care*, 1, 9-14

65. Girouard, K,. Harrison, MB. & VanDenKerkof, E. (2008). The symptom of pain with pressure ulcers: a review of the literature. *Ostomy Wound Management*, 5, 30-42

66. Hack, A. (2003). Malodorous wounds--taking the patient's perspective into account. *Journal of Wound Care*, 8, 319-21

67. Keast, DH., Bowering, CK,. Evans, AW,. Mackean, GL., Burrows, C. & D'Souza, L. (2004). MEASURE: A proposed assessment framework for developing best practice recommendations for wound assessment. *Wound Repair and Regeneration*, 17(6):797-805

68. Stewart, A,. Ware, JE. & editors. (1992). Measuring functioning and well-being: The Medical Outcomes Study approach. *Medical Care*, 724-35

69. WHO. (1947). *Constitution*. Geneva:World Health Organisation

70. Engel, GL. (1977). The need for a new medical model: a challenge for biomedicine. *Science*, 129-136

71. Greenfield, S & Nelson, EC. (1992). Recent developments and future issues in the use of health status assessment measures in clinical settings. *Medical Care*, Suppl, S23-41

72. McHorney, CA. (1999). Health status assessment methods for adults: past accomplishments and future challenges. *Annual Review Public Health*, 309-335

73. Patrick, DL. & Bergner, M. (1990). Measurement of health status in the 1990s. *Annual Review of Public Health*, 165-83

74. WHO. (2003). Learning to live with health ecconomics. Copenhagen: World Health Organisation Regional office for Europe

75. WHO. (1980). International classification of impairments, disabilities and handicaps. Generva: World Health Organisation

76. Gray, DB & Hendershot, GE. (2000). The ICIDH-2: developments for a new era of outcomes research. *Archives of Physical Medicine & Rehabilitation*, Suppl 2, S10-14

77. WHO. (1999). ICIDH-2; International classification of functioning and disability. Geneva: World Health Organisation

78. Wilson, IB. & Cleary, PD. (1995). Linking clinical variables with health-related quality of life: A conceptual model of patient outcomes. *Journal of the American Medical Association*, 1, 59-65

79. Bennett, SJ,. Perkins, SM,. Lane, KA. & et al. (2001). Social support and health-related quality of life in chronic heart failure patients. *Quality of Life Research*, 8, 671-682

80. Cunningham, WE,. Crystal, S,. Bozzette, S. & et al. (2005). The association of health-related quality of life with survival among persons with HIV infection in the United States. *Journal of General Internal Medicine*, 1, 21-27

81. Wettergren, L,. Bjorkhom, M,. Axdorph, U. & Languis-Eklof, A. (2004). Determinants of health-related quality of ife in long-term survivors of Hodgkin's Lymphoma. *Quality of Life Research*, 8, 1369-79

82. Rothman, ML., Beltran, P,. Cappelleri, JC., Lipscomb, J. & Teschendorf, B. (2007). Patient-Reported Outcomes: Conceptual Issues. *Value in Health*, s2, S66-S75

83. Fitzpatrick, R., Davey, C., Buxton, MJ & Jones, DR. (1998). Evaluating patient-based outcome measures for use in clinical trials. *Health Technology Assessment*, 14, 1-74

84. Garratt, A,. Schmidt, L,. Mackintosh, A. & Fitzpatrick, R. (2002). Quality of life measurement: bibliographic study of patient assessed health outcome measures. *BMJ*, 7351, 1417-22

85. Hobart, JC & Cano, S. (2009). Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. *Health Technology Assessment*, 12

86. US Department of Health & Human Services FDA. (2009). Patient reported outcome measures: Use in medical product development to support labelling claims. *Federal Register*, 74:1-43

87. Hobart, J,. Cano, S,. Zajicek, J. & Thompson, A. (2007). Rating scales as outcome measures for clinical trials in neurology: problems, solutions, and recommendations. *Lancet Neurology*, 1094-95

88. Fallowfield, L. (1990). *The Quality of Life: Missing Measurement in Health care,* London:Souvenir Press

89. Ware, JE., Jr. (1995). The status of health assessment 1994. *Annual Review of Public Health*, 327-54

90. Fayers, PM & Machin, D. (2000). *Quality of life assessment, analysis and interpretation.* Chichester:John Wiley & Sons Ltd

91. Patrick, DL. & Erickson, P. (1993). Concepts of health-related quality of life. 76-112

92. Sullivan, M. (1992). Quality of life assessments in medicine. *Nordic Journal of Psychology*, 79-83

93. Furlong, WJ,. Feeny, DH,. Torrance, GW. & Barr, RD. (2001). The Health Utilities Index (HUI) system for assessing health-related quality of life in clinical studies. *Annual of Medicine*, 375-384

94. Patrick, DL. & Chiang, YP. (2000). Measurement of health outcomes in treatment effectiveness evaluations: conceptual and methodological challenges. *Medical Care*, 38(9 Suppl):14-25.

95. Ware, JE,. Kosinski, MA. & Keller, S. D. (1994). SF-36 Physical and metal component summary measures: A users' manual.

96. Bowling, A. (2005). *Measuring Health: A review of quality of life measurement scales*. Berkshire: Open University Press

97. Carlson, LE & Bultz, BD. (2003). Benefits of psychosocial oncology care: Improved quality of life and medical cost offset. *Health and Quality of Life Outcomes*, 8, 1-8

98. Guyatt, GH,. Feeny, DH. & Patrick, DL. (1993). Measuring health-related quality of life. *Annals of Internal Medicine*, 8, 622-629

99. Patrick, DL. & Deyo, RA. (1989). Generic and disease-specific measures in assessing health status and quality of life. *Medical Care*, 3 Suppl, 217-32

100. Ware, JE,. Snow, KK,. Kosinski, M. & Gandek, B. (1997). SF-36 Health survey: Manual and interpretation guide.

101. Drummond, MF,. Sculpher,. O'Brien, BL,. Stoddart, GL. & Torrance, GW. (2005) *Methods for economic evaluation of health care programmes*. Oxford: Oxford University Press

102. Scientific Advisory Committee of the Medical Outcomes Trust. (2002). Assessing health status and quality-of-life instruments: attributes and review criteria. *Quality of Life Research*, 3, 193-205

103. McClimans, LH & Browne, J. (2011). Choosing a patient-reported outcome measure. *Theoretical Medicine and Bioethics*, 47-60

104. Fox, C. (2002). Living with a pressure ulcer: a descriptive study of patients' experiences. *British Journal of Community Nursing*, 6 Suppl, 10, 12, 14, 16-

105. Hopkins, A., Dealey, C & Bale, S. (2006). Patient stories of living with a pressure ulcer. *J Advanced Nursing*, 4, 345-353

106. Langemo, DK,. Melland, H,. Hanson, D,. Olson, B. & Hunter, S. (2000). The lived experience of having a pressure ulcer: a qualitative analysis. *Advances in Skin & Wound Care*, 5, 225-235

107. Rastinehad, D. (2006). Pressure ulcer pain. *Journal of WOCN*, 3, 252-257

108. Franks, P & Collier, M. (2001). Quality of life: the cost to the individual. 4:37-45

109. Lyons, S. & Sorenson, M. (2007). Quality of life in spinal cord injury patients with pressure ulcers. *SCI Nursing: Electronic Journal*, 3

110. Thein, HH,. Gomes, T,. Krahn, MD. & Wodchis, WP. (2009). Health status utilities and the impact of pressure ulcers in long-term care residents in Ontario. *Quality of Life Research*, 19:81-89

111. Essex, HN., Clark, M., Sims, J., Warriner, A & Cullum, N. (2009). Health-related quality of life in hospital inpatients with pressure ulceration: Assessment using generic health-related quality of life measures. *Wound Repair and Regeneration*, 6, 797-805

112. Franks, PJ., Winterberg, H & Moffatt, C. J. (2002). Health-related quality of life and pressure ulceration assessment in patients treated in the community. *Wound Repair & Regeneration*, 3, 133-140

113. Velikova, G,. Booth, L,. Smith, AB,. Brown, P,. Lynch P. & Brown, J. (2004). Measuring quality of life in routine oncology practice improves communication and patient well-being: A randomised controlled trial. *Journal of Clinical Oncology*, 4, 714-724

114. Hibbard, JH. (2003). Engaging health care consumers to improve the quality of care. *Medical Care*, 41:61-70

115. Detmar, SB, Muller, MJ, Schornagal, JH, Wever, LD & Aaronson, NK. (2002). Health-related quality -of-life assessments and patient-physician communication: A randomised controlled trial. *Journal of the American Medical Association*, 288(23):3027-37

116. Calkins, DR,. Rubenstein, LV,. Cleary, PD. & et al. (1991). Failure of physicians to recognise functional disability in ambulatory patients. *Annals of Internal Medicine*, 6, 451-454

117. Chauhan, Cynthia. (2007). Denouement: A Patient-Reported Observation. *Value in Health*, s2, S146-S147

118. Revicki, DA, Osoba, D, Fairclough, et al. (2000). Recommendations on health-related quality of life research to support labeling and promotional claims in the United States. *Quality of Life Research*, 9:887-900

119. O'Boyle, CA,. Hofer, S. & Ring, L. (2005). Individualized quality of life. 4.1, 225-242

120. Rothman, M,. Burke, L,. Erickson, P,. Kline Leidy, N,. Patrick, DL. & Petrie, CD. (2009). Use of existing patient-reported outcome (PRO) instruments and their modification: The ISPOR good research practices for evaluating and documenting content validity for the use of existing instruments and their modification PRO tasck force report. *Value in Health*, 8, 1075-83

121. Marlow, S. (1999). System 4: the four-layer bandage system from SSL International. *British Journal of Nursing*, 16, 1104-7

122. Cullum, N., Deeks, J., Fletcher, A., Long, A., Mouneimne, H., Sheldon, T & Song, F. (1995). The prevention and treatment of pressure sores: How effective are pressure-relieving interventions and risk assessment for the prevention and treatment of pressure sores? *Effective Health Care Bulletin*, 1, 1-16

123. Nixon, J,. Nelson, EA,. Cranny, G,. Iglesias, C. P,. Hawkins, K,. Cullum, NA,. Phillips, A,. Spilsbury, K,. Torgerson, D. J. & Mason, S. (2006). Pressure relieving support surfaces: A randomised evaluation. *Health Technology Assessment*, 22

124. Hudson, M, Impens, A, Baron, M, Seibold, JR, Thombs, BD, Walker, JG, the Canadian Sc, leroderma Research Group, and Steele, R. (2010). Discordance between patient and physician assessments of disease severity in Systemic Sclerosis. The Journal of Rheumatology, 37(11):2307-12

125. Yen, JC, Neville, C & Fortin, PR. (1999). Discordance between patients and their physicians in the assessment of lupus disease activity: relevance for clinical trials. *Lupus*, 8:660-70

126. Campbell, R., Quilty, B & Dieppe, P. (2003). Discrepancies between patients' assessments of outcome: qualitative study nested within a randomised controlled trial. *BMJ*, 7383, 252-3

127. McWhinney, IR., Bass, MJ & Donner, A. (1994). Evaluation of a palliative care service: Problems and pitfalls. *British Medical Journal*, 6965, 1340-1342

128. Greenhalgh, J. & Meadows, K. (1999). The effectiveness of the use of patient-based measures of health in routine practice in improving the process and outcomes of patient care: a literature review. *Journal of Evaluation in Clinical Practice*, 4, 401-416

129. Marshall, S,. Haywood, K. & Fitzpatrick, R. (2006). Impact of patient-reported outcome measures on routine practice: a structured review. *Journal of Evaluation in Clinical Practice*, 5, 559-568

130. Valderas, JM., Kotzeva, A., Espallargues, M., Guyatt, G., Ferrans, CE., Halyard, MY., Revicki, DA., Symonds, T., Parada, A & Alonso, J. (2008). The impact of measuring patient-reported outcomes in clinical practice: a systematic review of the literature. *Quality of Life Research*, 2, 179-193

131. Blazeby, J,. Sprangers, M,. Cull, A,. Groenvold, M. & Bottomley, A. (2002). EORTC Quality of Life Group: Guidelines for Developing Questionnaire Modules, 3rd ed. Brussels:EORTC Quality of Life Group

132. Streiner, DL & Norman, GR. (2003). Health Measurement Scales: A Practical Guide to Their Development and Use, 3rd ed. London:Oxford University Press

133. Bondas, T. & Hall, EOC. (2007). Challenges in approaching metasynthesis research. *Qualitative Health Research,* 17(1):113-21

134. Pearson, A. (2004). Balancing the evidence: incorporating the synthesis of qualitative data into systematic reviews. *JBI Reports,* 2(2):45-64

135. Thomas, R. (2003). Examining quality of life issues in relation to endocrine therapy for breast cancer. *American Journal of Clinical Oncology*, 4, 40-4

136. Clarke, L. (1995). Nursing research: science, visions and telling stories. *Journal of Advanced Nursing*, 3, 584-93

137. Shaughnessy, J & Zechmeister, E. (1997). Observation. *In Research methods in psychology*, 4th ed. Singapore:McGraw-Hill, 80-117

138. Thomas, J & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology*, 45,

139. MRC. (2003). Clinical trials for tomorrow: An MRC review of randomised control trials. *Clinical Trials Series*, Medcial Research Council, UK

140. NHS Centre for Reviews and Dissemination. (2001). *Undertaking systematic reviews of research on effectivness.* Report 4. Retrieved University of York: http://www.york.ac.uk/inst/crd/report4.htm

141. Gorecki, C,. Brown, JM,. Briggs, M. & Nixon, J. (2010). Evaluation of five search strategies in retrieving qualitative patient-reported electronic data on the impact of pressure ulcers on quality of life. *Journal of Advanced Nursing*, 3, 645-652

142. InterTASC. (2006). *Search filter resource:Patient views.* Retrieved http://www.york.ac.uk/inst/crd/intertasc/public.htm

143. Centre for Evidence-Based Medicine. (2006). *Searching for the best evidence in clinical journals.* Retrieved http://www.cebm.net/searching.asp#aetiology

144. Shaw, RL,. Booth, A,. Sutton, AJ,. Miller, T,. Smith, JA,. Young, B,. Jones, DR. & Dixon-Woods, M. (2004). Finding qualitative research: an evaluation of search strategies. *BMC Medical Research Methodology*, 5

145. de Laat, EH., Scholte op Reimer, WJ & van, Achterberg T. (2005). Pressure ulcers: diagnostics and interventions aimed at wound-related complaints: a review of the literature. *Journal of Clinical Nursing*, 4, 464-72

146. Fleurence, RL. (2005). Measuring quality of life in patients with pressure ulcers to include in economic evaluations. *Journal of Wound Care*, 3, 129-131

147. Goode, ML. (2004). Psychological needs of patients when dressing a fungating wound: a literature review. *Journal of Wound Care,* 13(9):380-2

148. EPUAP. (1999). *Pressure Ulcer Treatment Guidelines.* Retrieved: http://www.epuap.org/gltreatment.html

149. Joanna Briggs Institute. (2005). *Qualitative Assessment and Review Instrument*. Adelaide:Joanna Briggs Institute

150. CONSORT. (2001). *CONSORT Statement.* Retrieved from the CONSORT website: http://www.consort-statement.org/

151. STROBE. (2005). *STROBE statement: Strengthening the reporting of observational studies in epidemiology.* Retrieved July 12th, 2006 from the STROBE statement website: http://www.strobe-statement.org/PDF/STROBE-Checklist-Version3.pdf

152. Briggs, M. & Flemming, K. (2007). Living with leg ulceration: a synthesis of qualitative research. *Journal of Advanced Nursing*, 4, 319-328

153. O'Cathain, A,. Murphy E. & Nicholl, J. (2010). Three techniques for integrating data in mixed method studies. *BMJ*, 1147-1150

154. Pearson, A. (2004). Balancing the evidence: incorporating the synthesis of qualitative data into systematic reviews. *JBI Reports*, 2, 45-64

155. Kitching, M. (2004). Patients' perceptions and experiences of larval therapy. *Journal of Wound Care*, 1, 25-29

156. Wellard, S. & Rushton, C. (2002). Influences of spatial practices on pressure ulcer management in the context of spinal cord injury. *International Journal of Nursing Practice*, 4, 221-227

157. Krause, J. S. (1998). Intercorrelations between secondary conditions and life adjustment in people with spinal cord injury. *SCI Psychosocial Process*, 1, 3-7

158. Sanchez, HM., Munoz Villalba, JC., Roldan, MB., Cardo, JE., Rivilla, FD., de la Calle, OA & Cunarro Alonso, JM. (2008). Treatment using a hydrocolloid dressing and its influence on the life quality a patient has. *Revista De Enfermeria*, 1, 32-38

159. Sapountzi-Krepia, D., Soumilas, A., Papadakis, N., Sapkas, G., Nomicos, J., Theodossopoulou, E & Dimitriadou, A. (1998). Post traumatic paraplegics living in Athens: the impact of pressure sores and UTIs on everyday life activities. *Spinal Cord*, 6, 432-437

160. Dewey, A., Rice-Oxley, M & Dean, T. (2004). A qualitative study comparing the experiences of tilt-in-space wheelchair use and conventional wheelchair use by clients severely disabled with multiple sclerosis. *British Journal of Occupational Therapy*, 2, 65-74

161. Gunes, U. Y. (2008). A descriptive study of pressure ulcer pain. *Ostomy Wound Management*, 2, 56-61

162. Hatcliffe, S. & Dawe, R. (1996). Monitoring pressure sores in a palliative care setting. *International Journal of Palliative Nursing*, 4, 182-187

163. Manfredi, PL., Breuer, B., Meier, DE & Libow, L. (2003). Pain assessment in elderly patients with severe dementia. *Journal of Pain & Symptom Management*, 1, 48-52

164. Quirino, J., Santos, VLC., Quednau, TJP., Martins, APF., Lima, P & Almeida, M. R. M. (2003). Pain in pressure ulcers. *Wounds: A Compendium of Clinical Research and Practice*, 12, 381-389

165. Roth, RS., Lowery, JC & Hamill, JB. (2004). Assessing persistent pain and its relation to affective distress, depressive symptoms, and pain catastrophising in patients with chronic wounds: a pilot study. *American Journal of Physical Medicine & Rehabilitation*, 11, 827-834

166. Szor, JK. & Bourguignon, C. (1999). Description of pressure ulcer pain at rest and at dressing change. *Journal of Wound, Ostomy, & Continence Nursing*, 3, 115-120

167. Feng, DY. & Qin, LL. (2004). Relativity of psychological status and coping style with the occurrence of pressure sore in senile cancer patients after surgical stress. *Zhongguo Linchuang Kangfu*, 35, 7888-7890

168. Harding-Okimoto, MB. (1997). Pressure ulcers, self-concept and body image in spinal cord injury patients. *SCI Nursing*, 4, 111-117

169. Krause, JS. (1998). Skin sores after spinal cord injury: relationship to life adjustment. *Spinal Cord*, 1, 51-56

170. Widerstrom-Noga, EG., Felipe-Cuervo, E., Broton, JG., Duncan, RC & Yezierski, RP. (1999). Perceived difficulty in dealing with consequences of spinal cord injury. *Archives of Physical Medicine & Rehabilitation*, 5, 580-586

171. Baltzi, E. & Dafogianni, C. (2005). Quality of life and bedsores. *ICUs and Nursing Web Journal*, 4, 1-10

172. Forbes, A., While, A., Mathes, L & Griffiths, P. (2006). Health problems and health-related quality of life in people with multiple sclerosis. *Clinical Rehabilitation*, 1, 67-78

173. Fuhrer, MJ., Rintala, DH., Hart, KA., Clearman, R & Young, ME. (1993). Depressive symptomatology in persons with spinal cord injury who reside in the community. *Archives of Physical Medicine & Rehabilitation*, 3, 255-260

174. Krause, JS., Coker, JL., Charlifue, S & Whiteneck, GG. (2000). Health outcomes among American Indians with spinal cord injury. *Archives of Physical Medicine & Rehabilitation*, 7, 924-931

175. Rintala, DH., Young, ME., Hart, KA., Clearman, RR & Fuhrer, MJ. (1992). Social support and the well-being of persons with spinal cord injury living in the community. *Rehabilitation Psychology*, 3, 155-163

176. Wellard, S. (2001). An Australian experience of managing pressure ulcers in persons with SCI. *SCI Nursing*, 1, 11-17

177. Barr JE, Day AL Weaver VA Taler GM. (1995). Assessing clinical efficacy of a hydrocolloid/alginate dressing on full-thickness pressure ulcers. *Ostomy Wound Management*, 3, 28-30, 32, 34-36, 38, 40

178. Miller, GM,. Clinton, M,. Thomsen, D,. Miller, GM,. Clinton, M, et al. (2002). Pressure sores: Part II: an incurable malady? *Paraplegia News*, 1, 35-37

179. Price, P & Harding, K (2004). Cardiff wound impact schedule: the development of a condition-specific questionnaire to assess health-related quality of life in patients with chronic wounds of the lower limb. International Wound Journal, 1:10-17.

180. Herber, OR., Schnepp, W & Rieger, MA. (2007). A systematic review on the impact of leg ulceration on patients' quality of life. *Health & Quality of Life Outcomes*, 5(44)

181. Wall, J. (2002). Preventing pressure sores among wheelchair users. *Professional Nurse*, 5, 321-324

182. Smith, SC,. Lamping, DL,. Banerjee, S,. Harwood, R,. Foley, B, et al. (2005). Measurement of health-related quality of life for people with dementia: development of a new instrument (DEMQOL) and an evaluation of current methodology. *Health Technology Assessment*, 10

183. Guyatt, GH & Cook, DJ. (1994). Health status, quality of life, and the individual. *JAMA*, 8, 630-631

184. Hobart, JC., Riazi, A., Lamping, DL., Fitzpatrick, R & Thompson, AJ. (2005). How responsive is the Multiple Sclerosis Impact Scale (MSIS-29)? A comparison with some other self report scales. *Journal of Neurology, Neurosurgery & Psychiatry,* 76(11):1539-43

185. Evans, D. (2002). Database searches for qualitative research. *Journal of the Medical Library Association*, 3, 290-293

186. Braun, V & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 77-101

187. Strauss, AL & Corbin, J. (1998). *Basics of qualitative research: techniques and procedures for developing grounded theory*. Thousand Oaks:Sage

188. Morse, JM. (1995). The significance of saturation. *Qualitative Health Research*, 2, 147-149

189. Hsieh, HF. & Shannon, SE. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 9, 1277-88

190. Krippendorf, K. (1980). *Content Analysis: An Introduction to its Methodology*. Beverly Hills:Sage

191. Graneheim, UH. & Lundman, B. (2004). Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. *Nurse Education Today*, 2, 105-112

192. Thompson, C,. McCaughan, D,. Cullum, N,. Sheldon, TA & Raynor, P. (2004). Increasing the visibility of coding decisions in team-based qualitative research in nursing. *International Journal of Nursing Studies*, 1, 15-20

193. Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 1, 37-46

194. DeCoster, J. (2004). *Data analysis in SPSS*. Retrieved: http://www.stat-help.com/notes.html

195. Landis, JR & Koch, GG. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 1, 159-174

196. Ritchie, J & Spencer, L. (1994). Qualitative data analysis for applied policy research. *In Analyzing qualitative data*, A. Bryman & R. G. Burgess, eds., London: Routledge

197. Ritchie, J & Lewis, J. (2003). *Qualitative research practice*. London:Sage, 219-262

198. Patton, MQ. (2002). *Qualitative research and evaluation methods*, 3rd ed. Thousand Oaks:Sage

199. Gorecki, C,. Brown, JM,. Nelson, EA,. Briggs, M,. Schoonhoven, L,. Dealey, C,. Defloor, T,. Nixon, J. & and on behalf of the European Quality of Life Pressure Ulcer Project. (2009). Impact of pressure ulcers on quality of life in older patients: A systematic review. *Journal of the American Geriatrics Society*, 1175-1183

200. Anand, SC., Dean, C., Nettleton, R & Praburaj, DV. (2003). Health-related quality of life tools for venous-ulcerated patients. *British Journal of Nursing*, 1:48-49

201. Brown, A. (2008). Does social support impact on venous ulcer healing or recurrence? *Br J Community Nursing*, 3, 14

202. de Laat, EH., Scholte op Reimer, WJ & van, Achterberg T. (2005). Pressure ulcers: diagnostics and interventions aimed at wound-related complaints: a review of the literature. *Journal of Clinical Nursing*, 4, 464-472

203. Heinen, MM., van, Achterberg T., op Reimer, WS., van de Kerkhof, PCM & de, Laat E. (2004). Venous leg ulcer patients: a review of the literature on lifestyle and pain-related interventions. *Journal of Clinical Nursing,* 13(3):355-66

204. Howard, A. & Davies, AH. (2001). Health-related quality of life in patients with venous ulceration. *Phlebology,* 16(1):12-16*)*

205. McMullen, M. (2004). The relationship between pain and leg ulcers: a critical review. *British Journal of Nursing,* 13(19):S30, S32, S34

206. Persoon, A., Heinen, MM., Van Der, V,. de Rooij, MJ., van de Kerkhof, PCM & van, Achterberg T. (2004). Leg ulcers: a review of their impact on daily life. *Journal of Clinical Nursing*, 13(3):341-54

207. Price, P. (2004). The diabetic foot: Quality of life. *Clinical Infectious Diseases,* 39(SUPPL 2):S129-S1312

208. Sutton, MR., McGrath, C., Brady, L & Ward, J. (2000). Diabetic foot care: Assessing the impact of care on the whole patient. *Practical Diabetes International,* 17(5):147-151

209. Vileikyte, L. (2001). Diabetic foot ulcers: A quality of life issue. *Diabetes/Metabolism Research and Reviews,* 17(4):246-249

210. Young, C. V. (2005). The effects of malodorous fungating malignant wounds on body image and quality of life. *Journal of Wound Care 14(8):359-62*

211. McDowell, I & Newell, C. (1996). *Measuring Health: a guide to rating scales and questionnaires*. New York:Oxford University Press

212. Haywood, KL,. Garratt, A., Schmidt, L., Mackintosh, A & Fitzpatrick, R. (2004). *Health Status and Quality of Life in Older People: a structured review of patient-assessed health instruments*. Report from the Patient-assessed Health Instruments Group to the Department of health

213. Pusic, AL., Chen, CM., Cano, S, et al. (2007). Measuring quality of life in cosmetic and reconstructive breast surgery: a systematic review of patient-reported outcomes instruments. *Plastic & Reconstructive Surgery*, 4, 823-837

214. Palfreyman, S., Michaels, J & Brazier, J. (2007). Development of a tool to examine the effect of venous ulcers on patients' quality of life. *Nursing Standard*, 45, 57-58

215. Naylor, W. (2002). Part 2: Symptom self-assessment in the management of fungating wounds. *World Wide Wounds*

216. Ferrans, CE. & Powers, MJR. (1985). Quality of life index: development and psychometric properties. *Advances in Nursing Science*, 1, 15-24

217. Clark, M. (2002). Pressure ulcers and quality of life. *Nursing Standard*, 22, 74-78, 80

218. Vogel, LC., Krajci, KA & Anderson, CJ. (2002). Adults with pediatric-onset spinal cord injuries: part 3: impact of medical complications. *Journal of Spinal Cord Medicine*, 4, 297-305

219. Hyland, ME., Ley, A & Thomson, B. (1994). Quality of life of leg ulcer patients: questionnaire and preliminary findings. *Journal of Wound Care*, 6, 294-298

220. Hareendran, A., Doll, H., Wild, DJ., Moffatt, CJ., Musgrove, E., Wheatley, C & Franks, PJ. (2007). The venous leg ulcer quality of life (VLU-QoL) questionnaire: Development and psychometric validation. *Wound Repair and Regeneration*, 4, 465-473

221. Smith, JJ., Guest, MG., Greenhalgh, RM & Davies, AH. (2000). Measuring quality of life in patients with venous ulcers. *Journal of Vascular Surgery*, 4, 642-649

222. Launois, R,. Reboul-Marty, J. & Henry, B. (1996). Construction and validation of a quality of life questionnaire in chronic lower limb venous insufficiency (CIVIQ). *Quality of Life Research*, 5:539-554

223. Augustin, M,. Dieterle, W,. Zschocke, I,. Brill, C,. Trefzer, D,. Peschen, M. & Vanscheidt, W. (1997). Development and validation of a disease-specific questionnaire on the quality of life of patients with chronic venous insufficiency. *VASA*, 26:291-301

224. Venkatraman, PD., Anand, SC., Dean, C., Nettleton, R., El, SA & Afify, S. (2005). Pilot study investigating the feasibility of an ulcer-specific quality of life questionnaire. *Phlebology,* 20(1):14-27

225. Chren, M,. Lasek, RJ,. Quinn, LM,. Mostow, EN. & Zyzanski, SJ. (1996). Skindex, a quality of life measure for patients with skin disease: reliability, validity, and responsiveness. *Journal of Investigative Dermatology*, 707-713

226. Chren, M,. Lasek, RJ,. Flocke, SA. & Zyzanski, SJ. (1997). Improved discriminative and evaluative capability of a refined version of Spindex, a quality of life instrument for patients with skin diseases. *Archives of Dermatology*, 1433-1440

227. Lamping, DL,. Schroter, S,. Kurz, X,. Kahn, S. R. & Abenhaim, L. (2003). Evaluation of outcomes in chronic venous disorders of the leg: development of a scientifically rigorous, patient-reported measure of symptoms and quality of life. *Journal of Vascular Surgery*, 2, 410-9

228. Abetz, L., Sutton, M., Brady, L., McNulty, P & Gagnon, DD. (2002). The Diabetic Foot Ulcer Scale (DFS): A quality of life instrument for use in clinical trials. *Practical Diabetes International,* 19(6):167-175

229. Bann, CM., Fehnel, SE & Gagnon, DD. (2003). Development and validation of the Diabetic Foot Ulcer Scale-short form (DFS-SF). *PharmacoEconomics,* 12(1):1277-1290

230. Johanson, NA,. Liang, MH,. Daltroy, L,. Rudicel, S. & Richmond, J. (2004). American academy of Orthopaedic surgeons lower limb outcomes assessment

instruments: reliability, validity, and sensitivity to change. *The Journal of Bone and Joint Surgery*, 5, 902-909

231. Vileikyte, L., Peyrot, M., Bundy, C., Rubin, RR., Leventhal, H., Mora, P., Shaw, JE., Baker, P & Boulton, AJM. (2003). The development and validation of a neuropathy- and foot ulcer-specific quality of life instrument. *Diabetes Care,* 26(9):2549-55

232. Mendonca, DA., Drew, PJ., Harding, KG & Price, PE. (2007). A pilot study on the effect of topical negative pressure on quality of life. *Journal of Wound Care*, 2, 49-53

233. Smith, SC,. Lamping, DL,. Banerjee, S,. Harwood, R,. Foley, B,. Smith, P,. Cook, JC,. Murray, J,. Prince, M,. Levin, E,. Mann, A. & Knapp, M. (2005). Measurement of health-related quality of life for people with dementia: development of a new instrument (DEMQOL) and an evaluation of current methodology. *Health Technology Assessment*, 10

234. Laupacis, A. & The Canadian Erythropoietin Study Group. (1991). The use of generic and specific quality-of-life measures in hemodialysis patients treated with erythropoietin. *Controlled Clinical Trials*, 4 Suppl, 168-179

235. Franks, PJ. & Moffatt, CJ. (2001). Health related quality of life in patients with venous ulceration: use of the Nottingham health profile. *Quality of Life Research*, 8, 693-700

236. Walters SJ, Morrell CJ Dixon. (1999). Measuring health-related quality of life in patients with venous leg ulcers. *Quality of Life Research,* 8(4):327-336

237. Ettema, TP., Droes, RM., de, Lange J., Mellenbergh, GJ & Ribbe, MW. (2005). A review of quality of life instruments used in dementia. *Quality of Life Research*, 3, 675-686

238. Haywood, KL., Garratt, AM & Fitzpatrick, R. (2005). Quality of life in older people: a structured review of generic self-assessed health instruments. *Quality of Life Research*, 7, 1651-1668

239. Schmitt, J., Langan, S., Williams, HC & Network, European Dermato-Epidemiology. (2007). What are the best outcome measurements for atopic eczema? A systematic review. *Journal of Allergy & Clinical Immunology*, 6, 1389-1398

240. Mokkink, LB., Terwee, CB., Patrick, DL, et al. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*, 7, 737-745

241. Valderas, JM., Ferrer, M., Mendivil, J, et al. & Scientific Committee. (2008). Development of EMPRO: a tool for the standardized assessment of patient-reported outcome measures. *Value in Health*, 4, 700-708

242. Skevington, SM., Sartorius, N., Amir, M., Sartorius, N., Orley, et al. (2004). Developing methods for assessing quality of life in different cultural settings - The history of the WHOQOL instruments. *Social Psychiatry & Psychiatric Epidemiology*, 1, 1-8

243. Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist*, 1012-27

244. McColl, E., Jacoby, A., Thomas, L., Soutter, L., Bamford, C., et al. (2001). Design and use of questionnaires: a review of best practice applicable to surveys of health service staff and patients. *Health Technology Assessment*, 31:1-256.

245. Fayers, P. & Hays, R. (2005). *Assessing quality of life in clinical trials: Methods and practice*. 2nd ed, Oxford:Oxford University Press

246. Jenkins, CR & Dillman, DA. (1997). Towards a theory of self-administered questionnaire design. *In Survey measurement and process quality*, L. Lyberg et al., eds., New York: Wiley, 165-96

247. Sudman, S & Bradburn, N. (1982). *Asking questions: a practicle guide to questionnaire design*. San Francisco:Mossey-Bass

248. Foddy, W. (1993). *Constructing questions for interviews and questionnaires: Theory and practice in social research*. Cambridge:Cambridge University Press

249. Roberson, MT & Sundstrom, E. (1990). Questionnaire design, return rates, and response favourableness in an employee attitude questionnaire. *Journal of Applied Psychology*, 75:354-7

250. Moser, CA & Kalton, G. (21971). *Survey methods in social investigation*. Gower:Aldershot

251. Tourangeau, R. (1984). Cognitive sciences and survey methods. I*n Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines*, T. Jabine et al., eds., Washington DC: National Academy Press., 73-100

252. Oppenheim, AN. (1992). *Questionnaire design, interviewing and attitude measurement*. London:Pinter

253. McColl, E. (2005). Developing questionnaires. *In Assessing quality of life in clinical trials*, 2nd ed. P. Fayers & R. Hays, eds., New York: Oxford University Press, 9-23

254. Jobe, J. B. (2003). Cognitive psychology and self-reports: Models and methods. *Quality of Life Research*, 3, 219-227

255. Jobe, JB & Mingay, DJ. (1989). Cognitive research improves questionnaires. *American Journal of Public Health*, 19:1053-1055

256. Boynton, PM. & Greenhalgh, T. (2004). Hands-on guide to questionnaire research: Selecting, designing, and developing your questionnaire. *British Medical Journal,* 328(7451):1312-1315

257. Mullin, PA., Lohr, KN., Bresnahan, BW & McNulty, P. (2000). Applying cognitive design principles to formatting HRQOL instruments. *Quality of Life Research*, 1, 13-27

258. Stewart, AL,. Hays, RD. & Ware, JE. (1988). The MOS Short-Form General Health Survey: reliability and validity in a patient population. *Medical Care*, 724-35

259. Ware, JE,. Harris, WJ,. Gandek, B,. Rogers, B. & Reese, PR. (1997). MAP-R for windows: multitrait/multi-item analysis program - revised user's guide.

260. Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Chicago:University of Chicago

261. Linacre, JM. (1994). Sample size and item calibration stability. *Rasch Measurement Transactions*, 4, 328

262. Hagquist, C,. Bruce, M. & Gustavsson, JP. (2009). Using the Rasch model in nursing research: an introduction and illustrative example. *International Journal of Nursing Studies*, 380-393

263. Willis, GB. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. Thousand Oaks:Sage

264. Ericsson, KA & Simon, HA. (1980). Verbal reports as data. *Psychological Review*, 87:215-250

265. Willis, GB., Royston, P & Bercini, D. (1991). The use of verbal report methods in the development and testing of survey questionnaires. *Applied Cognitive Psychology*, 3, 267

266. Watt, T,. Rasmussen, AK,. Groevold, M,. Bjorner, JB,. Watt, AH,. Bonnema, SJ,. Hegedus, L. & Rasmussen, UF. (2008). Improving a newly developed patient-reported outcome for thyroid patients, using cognitive interviewing. *Quality of Life Research*, 17:1009-1017

267. Collins, D. (2003). Pretesting survey instruments: an overview of cognitive methods. *Quality of Life Research*, 12:229-238

268. Willis, GB. (1999). *Cognitive interviewing: A "how to" guide.* Retrieved from http://appliedresearch.cancer.gov/areas/cognitive/interview.pdf

269. Willis, GB & Lessler, JT. (1999). *Question appraisal system: QAS-99.* Retrieved http://appliedresearch.cancer.gov/areas/cognitive/qas99.pdf

270. Christodoulou, C., Junghaenel, DU., DeWalt, DA., Rothrock, N & Stone, AA. (2008). Cognitive interviewing in the evaluation of fatigue items: Results from the patient-reported outcomes measurement information system (PROMIS). *Quality of Life Research*, 10, 1239-1246

271. Housen, P., Shannon, GR., Simon, B., Edelen, MO., Cadogan, MP., Sohn, L., Jones, M., Buchanan, JL & Saliba, D. (2008). What the resident meant to say: use of cognitive interviewing techniques to develop questionnaires for nursing home residents. *Gerontologist*, 2, 158-169

272. Andrich, D. (1989). Distinctions between assumptions and requirements in measurements in the social sciences. *Proceedings of the XXIVth International Congress of Psychology*, 7-16

273. Andrich, D. (1988). *Rasch models for measurement*. Beverly Hills:Sage

274. Cano, SJ,. Barrett, LE,. Zajicek, JP. & Hobart, JC. (2011). Beyond the reach of traditional analyses: using Rasch to evaluate the DASH in people with multiple sclerosis. *Multiple Sclerosis Journal*, 2, 214-222

275. Andrich, D. (1985). An elaboration of Guttman scaling with Rasch models for measurement. *In Sociological Methodology*, N. Brandon-Tuma, ed., San Fransisco: Jossey-Bass, pp. 33-80

276. Andrich, D. (1978). Rating formulation for ordered response categories. *Psychometrika*, 561-573

277. Andrich, D., Luo, G., & Sheridan, B. (2004). *Interpreting RUMM2020.* Perth:RUMM Laboratory

278. Hagquist, C & Andrich, D. (2004). Is the Sense of Coherence instrument applicable on adolescents? A latent trait analysis using Rasch modelling. *Personality Individual Differences*, 36:955-968

279. Wright, BD & Masters, G. (1982). *Rating scale analysis: Rasch measurement*. Chicago:MESA

280. Hobart, JC,. Riazi, A,. Thompson, AJ, et al. (2006). Getting the measure of spasticity in multiple sclerosis: the Multiple Sclerosis Spasticity Scale (MSSS-88). *Brain*, 224-234

281. Lamping, DL,. Schroter, S,. Marquis, P,. Marrel, A,. Duprat-Lomon, I. & Sagnier, PP. (2002). The community-acquired pneumonia symptom questionnaire: a new, patient-based outcome measure to evaluate symptoms in patients with community-acquired pneumonia. *Chest*, 3, 920-9

282. Schroter, S. & Lamping, DL. (2004). Coronary revascularisation outcome questionnaire (CROQ): development and validation of a new, patient based measure of outcome in coronary bypass surgery and angioplasty. *Heart 90(12):1460-6*

283. Turner, RR., Quittner, AL., Parasuraman, BM,. Kallich, JD. & Cleeland, CS. (2007). Patient-Reported Outcomes: Instrument Development and Selection Issues. *Value in Health*, s2, S86-S93

284. Andrich, D. (2002). Implication and applications of modern test theory in the context of outcomes based research. *Studies In Educational Evaluation*, 2, 103-21

285. Allen, MJ & Yen, WM. (1979). *Introduction to measurement theory*. Montery:Boorks/Cole

286. Nunnally, JC & Bernstein, IH. (1994). Psychometric Theory, 3rd ed. New York:McGraw-Hill

287. DeVellis, RF. (2006). Classical test theory. *Medical Care*, 44(11):S50-S59

288. Lipscomb, J,. Donaldson, MS. & Hiatt, RA. (2004). Cancer outcomes research and the arenas of application. *JNCI Monographs*, 1-7

289. Reeve, BB,. Hays, RD,. Chang, CH. & Perfetto, EM. (2007). Applying item response theory to enhance health outcomes assessment. *Quality of Life Research*, 16:1-3

290. Tennant, A,. McKenna, SP. & Hagell, P. (2004). Application of Rasch Analysis in the Development and Application of Quality of Life Instruments. *Value in Health*, 7(Suppl 1):S22-S26

291. Andrich, D. (2004). Controversy and the Rasch model: a characteristic of incompatible paradigms? *Medical Care*, 1 Suppl 1, S7-S16

292. Bond, TG & Fox CM. (2001). *Applying the Rasch model: fundamental measurement in the human sciences*. New Jersey: Lawrence Erlbaum Associates

293. Smith, EV. (2001). Evidence for the reliabilityof measures and validity of measure interpretation: a Rasch measurement perspective. *Journal of Applied Measurement*, 2:281-311

294. Hilari, K,. Byng, S,. Lamping, DL. & Smith, SC. (2003). Stroke and Aphasia Quality of Life Scale-39 (SAQOL-39): evaluation of acceptability, reliability, and validity. *Stroke*, 8, 1944-50

295. Stead, ML,. Brown, JM,. Velikova, G, et al. (1999). Development of an EORTC questionnaire module to be used in health-related quality-of-life assessment for patients with multiple myeloma. European Organization for Research and Treatment of Cancer Study Group on Quality of Life. *British Journal of Haematology*, 3, 605-11

296. Pusic, A,. Klassen, A,. Scott, A,. Klok, J,. Cordeiro, P. & Cano, S. (2009). Development of a new patient reported outcome measure for breast surgery: the BREAST-Q. *Plastic & Reconstructive Surgery*, 124:345-353

297. Zumbo, BD. (1999). A Handbook on the Theory and Methods of Differential Item Functioning (DIF): Logistic Regression Modeling as a Unitary Framework for Binary and Likert-Type (Ordinal) Item Scores. Ottawa: ON: Directorate of Human Resources Research and Evaluation, Department of National Defence.

298. Andrich, D., Luo, G & Sheridan, BE. (2004) Interpreting RUMM2020. Perth:RUMM Laboratory

299. Fisher, WP,. Elbaum, B. & Coulter, A. (2010). Reliability, precision, and measurement in the context of data from ability tests, surveys, and assessments. *Journal of Physics, Conference Series*, 1, http://iopscience.iop.org/1742-6596/238/1/012036/pdf/1742-6596_238_1012036.pdf-

300. Green, KE & Smith, RM. (1987). A comparison of two methods of decomposing item difficulties. *Journal of Educational Statistics*, 4, 369-381

301. Lord, FM. (1983). Small N justifies Rasch model. *In New horizons in testing: Latent trait test theory and computerized adaptive testing*, D. Weiss, ed., New York: Academic Press, Inc, 51-61

302. Pallant, J. (2007). SPSS survival manual, 3rd ed. Maidenhead:Oxford University Press

303. Teresi, JA,. Ramirez, M,. Lai, JS. & Silver, S. (2008). Occurences and sources of Differential Item Fuctioning (DIF) in patient-reported outcome measures: Description of DIF methods, and review of measures of depression, quality of life and general health. *Psychol Sci Q*, 4, 538

304. van der Velde, G,. Beaton, D,. Hogg-Johnston, S,. Hurwitz, E. & Tennant, A. (2009). Rasch analysis provides new insights into the measurement properties of the Neck Disability Index. *Arthritis & Rheumatism*, 4, 544-51

305. Pallant, J & Tennant, A. (2007). An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). *British Journal of Clinical Psychology*, 46:1-18

306. Tabachnick, BG & Fidell, LS. (2001). Using multivariate statistics. MA: Allyn & Bacon

307. Stevens, S. (1951). *Handbook of Experimental Psychology*. New York:Wiley

308. Tennant, A & Conaghan, PG. (2007). The measurement model model in rheumatology: What is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis & Rheumatism*, 57(8):1358-1362

309. Andrich, D,. Sheridan, B. & Luo, G. (2010). RUMM 2030. Perth: RUMM Laboratory PTY LTD

310. Bonferroni, CE. (2011). Teoria statistica delle classi e calcolo delle probabilit`a. *Pubblicazionidel R Istituto Superiore di Scienze Economichee Commerciali di Firenze*, 8:3-62

311. Miller, RG. (1981). Simultaneous statistical inference, 2nd ed. Verlag:Springer, 6-8

312. Bland, JM and Altman, DG. (1995). Multiple significance tests: the Bonferroni method. *BMJ*, 310:170

313. Smith, EV. (2002). Detecting and evaluating the impact of multidimensionality using item fit statistics and principal components analysis of residuals. *Journal of Applied Measurement*, 3(2):205-231

314. Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16:297-334

315. Cronbach, L. & Meehl, P. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 4, 302

316. Andrich, D. (1982) An index of person separation in latent trait theory, the traditional KR20 index and the Guttman scale response. *Education Research and Perspectives*, 9:95-104

317. Cano, SJ, Warner, TT, Thompson, AJ, Bhatia, KP, Fitzpatrick, R, and Hobart, JC. (2008). The cervical dystonia impact profile (CDIP-58): Can a Rasch developed patient reported outcome measure satisy traditional psychometric criteria? *Health and Quality of Life Outcomes*, 6:58

318. Cano, SJ., Browne, JP., Lamping, DL., Roberts, AH., McGrouther, DA & Black, NA. (2006). The Patient Outcomes of Surgery-Head/Neck (POS-head/neck): a

new patient-based outcome measure. *Journal of Plastic, Reconstructive & Aesthetic Surgery*, 1, 65-73

319. McHorney, CA., Ware, JEJ,. Lu, JFR. & Sherbourne, CD. (1994). The MOS 36-Item Short-Form Health Survey (SF-36): III.Tests of data quality, scaling assumptions and reliability accross diverse patient groups. *Medical Care*, 1, 40-66

320. WHOQOL Group. (1998). The World Health Organization Quality of Life Assessment (WHOQOL): development and general psychometric properties. *Social Science & Medicine,* 46(12):1569-85

321. Likert, RA. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 5-55

322. McHorney, CA. & Tarlov, AR. (1995). Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Quality of Life Research*, 4:293-307

323. Hays, RD,. Anderson, R. & Revicki, DA. (1993). Psychometric consideration in evaluating health-related quality of life measures. *Quality of Life Research*, 2:441-449

324. Lohr, KN., Aaronson, NK., Alonso, J, et al. (1996). Evaluating quality-of-life and health status instruments: development of scientific review criteria. *Clinical Therapeutics,* 18(5):979-92

325. Kaplan, RM,. Bush, JW. & Berry, CC. (1976). Health status: Types of validity and the Index of Well-being. *Health Services Research*, 11:478-507

326. Kazis, LE,. Anderson, JJ. & Meenan, RP. (1989). Effect sizes for interpreting changes in health status. *Medical Care*, 27:S178-S189

327. McGraw, KO & Wong, SP. (1996). Forming inferences about some intra-class correlation coefficients*. Psychological Methods*, 1:30-46

328. Fleiss, JL. (1986). Reliability of measurements. *In The design and analysis of clinical experiments*, New York: John Wiley & Sons, 2-312-31

329. Guyatt, G,. Walter, S. & Norman, G. (1987). Measuring change over time:assessing the usefullness of evaluative instruments. *Journal of Chronic Disability*, 2, 171-78

330. Hobart, J,. Lamping, DL,. Fitzpatrick, R,. Riazi, A. & Thompson, AJ. (2001). The multiple sclerosis impact scale (MSIS-29): a new patient-based outcome measure. *Brain*, 124:962-73

331. Campbell, DT & Fiske, DW. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 2, 81-105

332. Ware, J., Jr., 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(3):220-33

333. Ware, JE,. Kosinski, M. & Turner-Bowker, DM. (2002). How to score version 2 of the SF-12 Health Survey (with a supplement documenting version 1). Lincolm: Quality Metric Incorporated

334. Burnand, B,. Kernan, WN. & Feinstein, AR. (1990). Indexes and boundaries for "quantitative significance" in statistical decisions. *Journal of Clinical Epidemiology*, 143:1278-84

335. Kerlinger, FN. (1973). Foundations of behavioural research. New York: Holt, Rinehart and Winston

336. Linacre, JM. (1999). Investigating rating scale category utility. *Journal of Outcome Measurement*, 3:103-122

337. CONSORT. (2001) CONSORT Statement. Retrieved August 12, 2006: http://www.consort-statement.org/

338. Wright, BD. (1977). Solving measurement problems with the Rasch model. *Journal of Educational Measurement*, 14:97-116

339. Platz, T,. Eickhof, C,. Nuyens, G. & Vuadens, P. (2005). Clinical scales for the assessment of spasticity, associated phenomena, and function: a systematic review of the literature. *Disability and Rehabilitation*, 27:7-18

340. Stuki, G,. Daltroy, L,. Katz, JN,. Johannesson, M. & Liang, MH. (1996). Interpretation of change scores in ordinal clinical scales and health status measures: the whole may not equal the sum of the parts. *Journal of Clinical Epidemiology*, 7, 711-717

341. Svensson, E. (2001). Guidelines to statistical evaluation of data from rating scales and questionnaires. *Journal of Rehabilitation Medicine*, 33:47-48

342. Cano, S,. Riazi, A,. Schapira, AH,. Cooper, JM. & Hobart, J. (2009). Friedreich's Ataxia Impact Scale@ a new measure striving to provide the flexibility required by today's studies. *Movement Disorders*, 7, 984-992

343. Cano, S,. Posner, HB,. Moline, ML,. Hurt, SW,. Swartz, J,. Hsu, T. & Hobart, J. (2010). The ADAS-cog in Alzheimer's disease clinical trials: psychometric evaluation of the sum and its parts. *Journal of Neurology, Neurosurgery & Psychiatry*, 81:1363-1368

344. Choong-Wan, W,. Seok-Man, K,. Young-Jin, L. & Min-Sup, S. (2010). The Obsessive-Compulsive Inventory-Revised (OCI-R): Psychometric properties of the Korean version and the order, gender, and cultural effects. *Journal of Behavior Therapy AndExperimental Psychiatry*, 41:220-227

345. Barry, M,. Walker-Corkey, E,. Chang, Y,. Tyll, LT,. Cherkin, D. & Fowler, F. (1996). Measurement of overall and disease-specific health status: Does the order of questionnaires make a difference? *Journal of Health Services Research and Policy,* 5(1):20-27

346. Hobart, J. C., Riazi, A., Lamping, D. L., Fitzpatrick, R & Thompson, A. J. (2004). Improving the evaluation of therapeutic interventions in multiple sclerosis: development of a patient-based measure of outcome. *Health Technology Assessment,* 8(9):1-48

347. Kahn, SR., Lamping, DL., Ducruet, T., Arsenault, L., Miron, M. J., Roussin, A., Desmarais, S., Joyal, F., Kassis, J., Solymoss, S., Desjardins, L., Johri, M., Shrier, I & VETO, Study investigators. (2006). VEINES-QOL/Sym questionnaire was a reliable and valid disease-specific quality of life measure for deep venous thrombosis. *Journal of Clinical Epidemiology,* 59(10):1049-56

348. Cano, S,. Warner, T. & Hobart, J. (2003). What sample sizes for reliability and validity studies II: a prospective study. *Quality of Life Research*, 12:771

349. Hobart, J,. Cano, S. & Thompson, A. (2002). What sample sizes for reliability and validity studies? *Quality of Life Research*, 11:636

350. Deyo, RA,. Diehr, P. & Patrick, DL. (1991). Reproducibility and responsiveness of health status measures: Statistics and strategies for evaluation. *Controlled Clinical Trials*, Suppl 4, S142-S158

351. Roussos, LA & Stout, WF. (1996). A multidimensionality-based DIF analysis paradigm. *Applied Psycholgical Measurement*, 20:355-371

352. Fayers, P. (2008). Evaluating the effectiveness of using PROs in clinical practice: a role for cluster-randomised trials. *Quality of Life Research*, 17:1315-1321

353. Higginson, I & Carr, A. (2001). Using quality of life measures in the clinical setting. *British Medical Journal*, 322:1297-1300

354. Browne, J,. Jamieson, L,. Lawsey, J,. van der Meulen, J,. Black, N,. Cairns, J,. Lamping, D,. Smith, S,. Copley, L. & Horrockes, J. (2007). Patient reported outcome measures (PROMs) in elective surgery. Report to the Department of Health. Available from: www.lshtm.ac.uk/hsru/research/PROMs-Report-12-Dec-07.pdf-

355. Devlin, NJ,. Parkin, D. & Browne, J. (2010). Patient-reported outcome measures in the NHS: New methods for analysing and reporting EQ-5D data. *Health Economics*, 19:886-905

356. Greenhalgh, J. (2009). The applications of PROs in clinical practice: what are they, do they work, and why? *Quality of Life Research*, 18:115-123

357. Cleary, PD. (1997). Subjective and objective measures of health: which is better when? *Journal of Health Services & Research Policy*, 2:3-4

358. Hobart, J. (2002). Measuring disease impact in disabling neurological conditions: are patients' perspectives and scientific rigour compatible. *Current Opinions in Neurology*, 15:721-724

359. NICE. (2008). Guide to methods of technology appraisals. Publication reference N1618, London: National Institute for Health and Clinical Excellence

360. Guyatt, G., Veldhuyzen Van Zanten, SJ., Feeny, DH. & Patrick, DL. (1989). Measuring quality of life in clinical trials: a taxonomy and review. Canadian Medical Association Journal, 12:1441-48

361. Donaldson, G. (2008). Patient-reported outcomes and the mandate of measurement. *Quality of Life Research*, 17:1303-1313

362. Huisman, M. (2000). Imputation of missing item responses: some simple techniques. *Quality and Quantity*, 34:331-351

363. Engels, JM & Diehr, P. (2003). Imputation of missing longitudinal data: a comparison of methods. *Journal of Clinical Epidemiology*, 56:968-976

364. NICE. (2003). *CG7 Pressure relieving devices.* Retrieved from the NICE, UK website: www.nice.org.uk

365. DOH. (2008). Guidance of the routine collection of patient reported outcome measures (PROMs). Department of Health document DH_081179[1].pdf-

366. Green, BF,. Bock, RD,. Humphreys, LG,. Linn, RL. & Reckase, MD. (1984). Technichal guidelines for assessing computerised adaptive tests. *Journal of Educational Measurement*, 4, 347-360

367. Hahn, EA,. Cella, D,. Bode, RK,. Gershom, R. & Lai, JS. (2006). Item banks and their potential applications to health status assessment in diverse populations. *Medical Care*, 11, S189-S197

368. Cano, SJ,. Klassen, AF. & Pusic, AL. (2011). From Breast-Q © to Q-score ©: using Rasch measurement to better capture breast surgery outcomes. *Joint International IMEKO Symposium*, urn:nbn:de:gbv:ilml-2011imeko:2,

369. Cull, A,. Sprangers, M,. Bjordal, K,. Aaronson, N,. West, K. & Bottomley, on behalf of the EORTC Quality of Life Group. (2002). EORTC quality of life group. Translation procedure, 2nd ed. Brussels:EORTC

Appendices

Appendix 2.1 Search strategies

Pressure Ulcer filter

1. decubitus.sh
2. skin ulcer.sh
3. exp decubitus ulcer
4. decubitus ulcer$.tw
5. pressure ulcer$.tw.
6. pressure damage$.tw
7. pressure sore$.tw
8. bed sore$.tw
9. skin ulcer$.tw
10. or/1-9

PU Symptom terms

11. exp pain
12. pain.tw
13. comfort$.tw
14. acceptab$.tw
15. discomfort.tw
16. exp quality of sleep
17. sleep.tw
18. exp smell
19. smell$.tw
20. odo?r$.tw
21. exudat$.tw
22. or/11-21

QoL terms

23. (sf36 or sf 36).ti,ab,tw,sh,kw
24. (eq5d or eq 5d or euroqol).ti,ab,tw,sh,kw
25. (short form 36 or shortform 36 or sf thirtysix or sf thirty six or shortform thirtysix or shortform thirty six or short form thirtysix or short form thirty six).ti,ab,tw,sh,kw
26. (hrql or hrqol or qol or hql or hqol).ti,ab,tw,sh,kw
27. (hye or hyes or health$ year$ equivalent$ or health utilit$).ti,ab,tw,sh,kw
28. rosser.ti,ab,tw,sh,kw
29. (quality of wellbeing or quality of well being or quality of wellbeing index or qwb).ti,ab,tw,sh,kw
30. (wellbeing or well being).ti,ab,tw,sh,kw
31. (health utilities index or hui).ti,ab,tw,sh,kw
32. (medical outcomes survey or mos).ti,ab,tw,sh,kw
33. (qaly$ or qualy$ or quality adjusted life or quality adjusted life year or quality of life or life quality).ti,ab,tw,sh,kw
34. exp quality of life
35. quality of living.tw
36. (health status or health state$).ti,ab,tw,sh,kw
37. health status indicators.mp
38. sickness impact profile.ti,ab,tw,sh,kw
39. health measurement$.ti,ab,tw,sh,kw
40. (health survey questionnaire$ or health survey$ or health care survey$).ti,ab,tw,sh,kw
41. (general health questionnaire$ or ghq).ti,ab,tw,sh,kw

42. or/23-41

Patient issue terms

43. patient.tw,mp
44. view$.tw
45. satisfaction$.tw
46. preference$.tw
47. opinion$.tw
48. perspective$.tw
49. concern$.tw
50. issue$.tw
51. experience$.tw,mp
52. journey$.tw,mp
53. (worry or worries).tw,mp
54. (attitude$ or emotion$ or feeling$).tw
55. ((psycho$ or social) adj (adjust$ or adap$)).tw
56. (cope$ or coping).tw
57. or/44-56
58. 43 and 57
59. exp emotion
60. depression.tw,mp
61. exp stress
62. exp stress, psychological
63. exp adaptation, psychological
64. exp acceptance, psychological
65. or/59-64
66. 58 or 65
67. 22 or 42 or 66
68. 10 and 67

RCT filter

1. randomi?ed-controlled-trial.pt
2. meta-analysis.pt
3. controlled-clinical-trial.pt
4. clinical-trial.pt
5. (clin$ trial$).tw
6. control$ and (trial$ or stud$).tw
7. random$.tw
8. (meta-analys?s or metaanalys?s or meta analys?s).tw
9. (singl$ or doubl$ or tripl$ or trebl$) and (blind$ or mask$).tw
10. exp clinical trial.sh
11. research-design.sh
12. comparative study.sh
13. placebo$.tw
14. crossover.ti,ab,sh
15. or/1-14
16. 15 and *pressure ulcer filter* above

Comprehensive search strategies including (a) Thesaurus terms, (b) Free text terms, and (c) Broad-based terms for Qualitative studies

(A) Thesaurus Terms

1. qualitative research
2. nursing methodology research
3. questionnaires
4. exp attitude
5. focus groups
6. discourse analysis.mp
7. content analysis.mp
8. ethnographic research.mp
9. ethnological research.mp
10. ethnonursing research.mp
11. constant comparative method.mp
12. qualitative validity.mp
13. purposive sample.mp
14. observational method$.mp
15. field stud$.mp
16. theoretical sampl$.mp
17. phenomenology
18. phenomenological research.mp

19. life experience$.mp
20. cluster sampl$.mp
21. or/1-20
22. 21 and *pressure ulcer filter* above

(B) Free Text Terms

1. ethnonursing.af
2. ethnograph$.mp
3. phenomenol$.af
4. grounded theory.mp
5. (grounded adj (theor$ or stud$ or research or analys?s)).af
6. ((life stor$) or women$ stor$)).af
7. (emic or etic or hermeneutic$ or heuristic$ or semiotic$).af or (data adj1 saturat$).tw or (participant observ$).tw
8. ((social construct$ or (postmodern$ or post-structural$) or post structural$ or poststructural$) or post modern$) or post-modern$ or feminis$ or interpret$).mp
9. (action research or cooperative inquir$ or (co operative inquir$) or (co-operative inquir$)).mp
10. (humanistic or existential or experiential or paradigm$).mp
11. (field adj (study or studies or research).tw
12. (human science).tw
13. (biographical method).tw
14. (qualitative validity).af
15. (purposive sampl$).af
16. (theoretical sampl$).af
17. ((purpose$ adj4 sampl$) or (focus adj group$)).af
18. (account or accounts or unstructured or open-ended or (open ended) or text$ or narrative$).mp
19. ((life world) or life-world or conversation analys?s or personal experience$ or theoretical saturation).mp
20. (lived experience$).mp
21. (life experience$).mp
22. (cluster sampl$).mp
23. (theme$ or thematic).mp
24. category$.mp
25. observational method$.mp
26. field stud$.mp
27. focus group$.af
28. questionnaire$.mp
29. (content analysis).af
30. (thematic analysis).af
31. (constant comparative).af
32. (discourse analys?s).af
33. ((discourse$ or discurs$ adj3 analys?s).tw
34. (constant adj (comparative or comparison)).af
35. (narrative analys?s).af
36. Heidegger$.tw
37. colaizzi$.tw
38. speigelberg$.tw
39. (van adj manen$).tw
40. (van adj kaam$).tw
41. merleau adj ponty$).tw
42. Husserl$.tw
43. giorgi$.tw
44. Foucault$.tw
45. (corbin$ adj2 strauss$).tw
46. (strauss$ adj2 corbin$).tw
47. (glaser$ adj2 strauss$).tw
48. glaser$.tw
49. or/1-48

50. 49 and *pressure ulcer filter* above

(C) Simple broad-based terms

1. findings.af
2. interview$.af or Interviews
3. qualitative.af
4. or/1-3
5. 4 and *pressure ulcer filter* above

26. stomach ulcer.mp
27. fistula$.mp
28. bite.tw
29. or/15-28
30. (all final searches above) not 29

The following refinement filter was added to all the searches listed above:

Refinement filter

1. historical article.pt.
2. review.pt.
3. (systematic adj review$).ti,ab,pt
4. (meta adj analysis).ti,ab
5. audit.ti,ab,pt
6. case report.tw,sh,mp
7. (case adj stud$).ti,ab,pt
8. exp guidelines
9. letter.pt.
10. comment.pt.
11. editorial.pt.
12. or/1-11
13. (all final searches above) not 12
14. limit 13 to humans
15. leg ulcer.mp
16. varicose ulcer.mp
17. pilonidal.tw
18. surgical flap$.mp
19. skin transplantation$.mp
20. burn$.mp
21. gunshot.mp
22. corneal ulcer.mp
23. exp dentistry
24. peptic ulcer.mp
25. duodenal ulcer.mp

Appendix 2.2 Additional Searches

Theses and dissertations

The following databases were searched for theses and dissertations:

- Proquest
- Networked Digital Library of Theses and Dissertations (NDLTD)
- International Theses in Progress
- Theses Canada Portal
- Australian Digital Theses Program (ADT)
- Russian Academy of Sciences Bibliographies
- Index to Theses

Hand searching

Journals

The following specialist journals were hand searched:

- Journal of Tissue Viability, 1990-present
- Journal of Wound Care, 1991-present
- Wounds Repair and Regeneration, 2000-present
- Review European Pressure Ulcer Advisory Panel, 1999-present
- International Wound Journal, 2004-present
- European Wound Management Association Journal, 2001-May 2007
- Journal of Health and Quality of Life Outcomes, 1999-present
- Journal of the American Medical Association archive collection of 'Quality of Life', 1998-present http://pubs.amaassn.org/cgi/collection/quality_of_life?page=1)

Conference proceedings

The following conference proceedings were hand searched:

- European Conference on Advances in Wound Management, 1991 - 2000
- Conference of the European Wound Management Association, 2001 - 2006
- Proceedings of the European Wound Management Association and Journal of Wound Care, 1997 - 1998
- 2nd World Union of Wound Healing Societies' Meeting, 2004
- Journal of Wound Healing 2nd Conference, 2005
- Wounds UK Conference, 2004
- The European Pressure Ulcer Advisory Panel Open Meeting, 1997 - 2007
- European Tissue Repair Society, Focus Meeting, 2000 - 2005
- Conference of the International Society Of Quality Of Life Research, 1997 - 2007

Experts

The following experts in the field were contacted to enquire about ongoing and recently published research: P Price; D Mendoca; S Bale; M Fox; A Hopkins; D Langemo; D Rastinehad; S Wellard; C Moffatt; J Krause; P Franks; C Dealey

Additional searches

- UK National Research Register
  http://www.dh.gov.uk/en/Aboutus/Researchanddevelopment/AtoZ/DH_4002357
- Internet search of web content relating to PU self-help and focus group:
  - http://www.ukselfhelp.info/
  - http://www.patient.co.uk/selfhelp.asp
  - http://www.dipex.org/DesktopDefault.aspx (website offering personal patient experiences of various health and illnesses)
- Patient-Reported Outcome and Quality of Life Database (PROQOLID, http://www.qolid.org/)

Appendix 3.1Topic Guide: Example questions

I. Intro/Background and General Questions

- Tell me a little about your experience-
- When was the skin problem first noticed? What were you told about the PU?
- Did you tell anyone? What did you do when you first realised you had a PU?

II. Experience and impact

- What is it like to live with a PU?
- How has your life changed since your ulcer developed?-anything else?
- What kinds of things are more difficult for you to do? Any other tasks?

III. Treatment and HC/wound management questions

- Have you received any treatment on that area?
- Can you tell me about your experience of the wound care treatments that you received (i.e. symptoms; acceptability; satisfaction)?
- How has your PU been attended to, what sorts of things have been done?

IV. Patient involvement

- Have the different types of wound treatments been discussed with you?
- Have you sought ways to help yourself rather than tell someone/ask for help?
- Have you considered alternative treatment and actually sought it?

V. Visualisation and PU improvements

- Have you seen it? (If yes, what did you think about it? How did it make you feel? If no, has anyone else seen it, how have they described it to you?
- Would you have liked to have seen it? (pictures, mirror, during dressings)

VI. General health and symptoms

- How would you describe your general health?
- Has developing a PU changed how you perceive your health status?
- What kind of symptoms have you experienced?

VII. Pain & pain assessment

- Can you try and describe the pain you feel? Do you feel pain at other times?
- Tell me how it feels when you move/sit/stand/reposition?

VIII. Final questions

- If someone came to you and said “I’m going to develop a PU what should I expect”, what would you tell them? How would you prepare them?
- Have you experienced anything else that we have not covered today?
- Is there anything else that you want to add about your experience?

Appendix 4.1. Search Strategy for review of existing PROs used in pressure ulcers and similar chronic wounds

Pressure Ulcer terms

1. decubitus.sh
2. skin ulcer.sh
3. exp decubitus ulcer
4. decubitus ulcer$.tw
5. pressure ulcer$.tw.
6. pressure damage$.tw
7. pressure sore$.tw
8. bed sore$.tw
9. skin ulcer$.tw
10. or/1-9

Chronic wound terms

11. chronic wound$.tw
12. leg ulcer$.tw
13. foot ulcer$.tw
14. venous ulcer$.tw
15. necrotic wound$.tw
16. ischaemic ulcer$.tw
17. arterial ulcer$.tw
18. fungating wound$.tw
19. diabetic ulcer$.tw
20. varicose vein$.tw
21. dehisced wound$.tw
22. pilonidal.tw
23. or/11-22
24. 10 or 23

QOL terms (Symptoms, social functioning, mood, life satisfaction)

25. (wellbeing or well being).ti,ab,tw,sh,kw
26. (hrql or hrqol or qol or hql or hqol).ti,ab,tw,sh,kw
27. exp quality of life
28. quality of living.tw
29. (health status or health state$).ti,ab,tw,sh,kw
30. (satisfaction or life satisfaction or satisfaction with life).tw
31. (attitude$ or emotion$ or feeling$ or mood$).tw
32. ((psycho$ or social) adj (adjust$ or adap$ or function$)).tw
33. (cope$ or coping).tw
34. exp emotion
35. exp psychological
36. exp adaptation, psychological
37. exp acceptance, psychological
38. symptom$.tw,ab,sh,kw
39. exp pain
40. pain.tw
41. exp smell
42. smell$.tw
43. odo?r$.tw
44. exudat$.tw
45. or/25-44

Measures terms

46. (instrument$ or questionnaire$ or survey$ or measure$).kw,ab,ti
47. (patient outcome$ or patient reported outcome$ or PRO$).ti,ab,tw,sh,kw
48. (sf36 or sf 36).ti,ab,tw,sh,kw
49. (eq5d or eq 5d or euroqol).ti,ab,tw,sh,kw
50. (short form 36 or shortform 36 or sf thirtysix or sf thirty six or shortform thirtysix or shortform thirty six or

short form thirtysix or short form thirty six).ti,ab,tw,sh,kw
51. sickness impact profile.ti,ab,tw,sh,kw
52. (medical outcomes survey or mos).ti,ab,tw,sh,kw
53. (quality of wellbeing or quality of well being or quality of wellbeing index or qwb).ti,ab,tw,sh,kw
54. (hye or health$ year$ equivalent$ or health utilit$ or health utilities index or utilit$ approach$ or health gain or hui or hui2 or hui 2 or hui3 or hui 3).ti,ab,tw,sh,kw
55. (qaly$ or qualy$ or quality adjusted life or quality adjusted life year or quality of life or life quality).ti,ab,tw,sh,kw
56. health measurement$.ti,ab,tw,sh,kw
57. (health survey questionnaire$ or health survey$ or health care survey$).ti,ab,tw,sh,kw
58. (general health questionnaire$ or ghq).ti,ab,tw,sh,kw
59. health status indicators.mp
60. (disutilities or disutility or daly or disability adjusted life).ti,ab.
61. preference based.ti,ab.
62. (state adj2 valu$).ti,ab.
63. (categor$ scal$ or linear scal$ or linear analog$ scal$ or visual scal$ or magnitude estimat$).ti,ab.
64. (multiattribute$ health or multi attribute$ health).ti,ab.
65. (multiattribute$ theor$ or multi attribute$ theor$ or multiattribute$ analys$ or multi attribute$ analys$).ti,ab.
66. (multiattribute$ utilit$ or multi attribute$ utilit$).ti,ab.
67. or/46-66
68. 24 and 45
69. 24 and 67
70. 68 or 69

Refinement

71. historical article.pt.
72. review.pt.
73. (systematic adj review$).ti,ab,pt
74. (meta adj analysis).ti,ab
75. audit.ti,ab,pt
76. case report.tw,sh,mp,pt
77. (case adj stud$).ti,ab,pt
78. exp guidelines
79. letter.pt.
80. comment.pt.
81. editorial.pt.
82. or/71-81
83. 70 not 82
84. limit 83 to humans, English, adult

Appendix 4.2 Additional searches

Hand Searched Journals

- Journal of Tissue Viability, 1990-present
- Journal of Wound Care, 1991-present
- Wounds Repair and Regeneration, 2000-present
- International Wound Journal, 2004-present
- European Wound Management Association Journal, 2001-May 2007
- Journal of Health and Quality of Life Outcomes, 1999-present
- Journal of the American Medical Association archive collection of 'Quality of Life',1998-present http://pubs.amaassn.org/cgi/collection/quality_of_life?page=1)

Hand search Conference Proceedings

- European Conference on Advances in Wound Management, 1991 - 2000
- Conference of the European Wound Management Association, 1997 - 2006
- 2nd World Union of Wound Healing Societies' Meeting, 2004
- Journal of Wound Healing 2nd Conference, 2005
- Wounds UK Conference, 2004
- The European Pressure Ulcer Advisory Panel Open Meeting, 1997 - 2008
- European Tissue Repair Society, Focus Meeting, 2000 - 2005
- Conference of the International Society Of Quality Of Life Research, 1997 - 2008

Searched Theses and dissertation databases

- Proquest
- Networked Digital Library of Theses and Dissertations
- International Theses in Progress
- Theses Canada Portal
- Australian Digital Theses Program
- Index to Theses

Experts

The following experts in the field were contacted to enquire about ongoing and recently published research: P Price; D Mendoca; S Bale; M Fox; A Hopkins; D Langemo; D Rastinehad; S Wellard; C Moffatt; J Krause; P Franks; C Dealey

Additional searches

- UK National Research Register
  http://www.dh.gov.uk/en/Aboutus/Researchanddevelopment/AtoZ/DH_4002357
- Patient-Reported Outcome and Quality of Life Database http://www.qolid.org/)

Appendix 5.1 Final PU-QOLI item map

Domain (Sub-domain)	Reduced Item List (87) – post pretest	Original Item list (118)
Symptoms	Feeling uncomfortable	Feeling uncomfortable
(Pain & Discomfort)	Annoying pain or discomfort	Annoying pain or discomfort
	Itchiness	Itchiness
	Tenderness	Tenderness
		Niggling
		Soreness
	A dull ache	Aching
		Pins and needles
	Tingling	Tingling
	Throbbing	Throbbing
		Nagging
		Shooting
	Stinging	Stinging
	Stabbing pains	Stabbing
		Electric shocks
	Red raw	Red raw
	Burning	Burning
(Exudate)	Weeping	Weeping
		Oozing
	Running	Running
	Sticky	Sticky
		Slimy
		Wet
	Messy	Messy
	Staining	Staining
	Causing dressing to come off	Causing dressing to come off
		Gungy
	Pus	Pus
	Bleeding	Bleeding
(Odour)	Unpleasant smell	Unpleasant smell
	Lingering smell	Lingering smell
		Dirty smell
		Foisty smell
	Stench or stink	Stench
		Stink
		Stale smell
	Pungent smell	Pungent smell

Domain (Sub-domain)	Reduced Item List (87) – post pretest	Original Item list (118)
	Sickening smell	Sickening smell
	Putrid smell	Putrid smell
Physical functioning	Difficulty sitting up in bed	Difficulty sitting up in bed
(Mobility/movement)	Difficulty adjusting yourself in bed	Difficulty adjusting yourself in bed
	Difficulty turning or moving in bed	Difficulty turning in bed
	Difficulty pushing up to a sitting position	Difficulty pushing up to a sitting position
	Difficulty sitting in one position for long periods	Difficulty sitting in one position for long periods
	Difficulty standing for long periods	Difficulty standing for long periods
	Difficulty transferring (e.g. from bed to a chair or to a car)	Difficulty transferring from bed to a chair
	Feeling limited in your ability to walk	Feeling limited in ability to walk
	Feeling limited in your ability to go up and down stairs	Feeling limited in ability to go up and down stairs
	Feeling limited in how far you were able to walk	Feeling limited in how far you were able to walk
	Feeling that your walking was slowed down	Feeling that your walking was slowed down
(ADL)	Being able to wash yourself in your usual way (e.g. hand wash, bath, shower)	Washing yourself in the bath or shower
	Getting dressed or undressed	Getting dressed or undressed
	Doing jobs around the house (e.g. cooking, housework, DIY)	Doing housework
	Doing gardening	Doing gardening
	Doing shopping	Doing shopping
	Being able to go to the toilet	Going to the toilet
		Being able to travel or drive a car
	Doing things that you enjoy (e.g. reading a book, watching a movie, using a computer)	Doing things that you enjoy
		Getting up and about to do things that you enjoy
	Being emotionally close or affectionate with loved ones	Being intimate with loved ones
	Doing your regular daily activities (e.g. work, volunteering, religious service, clubs, university)	Doing usual work
(General vitality)	Feeling that your appetite has reduced	Feeling that your appetite has reduced
	Feeling unwell or poorly	Feeling unwell or poorly

Domain (Sub-domain)	Reduced Item List (87) – post pretest	Original Item list (118)
	Feeling that your energy levels have been reduced (e.g. feeling tired, fatigued)	Feeling that your energy levels have been reduced
(Sleep)	Trouble falling asleep	Trouble falling asleep
	Interrupted sleep (e.g. restless sleep or being woken up during your sleep)	Restless sleep
	Being kept awake	Being kept awake
		Being woken up during the night
	Not getting the amount of sleep that you needed	Not getting the amount of sleep that you needed
	Having to sleep in one position (e.g. your back or side)	Having to sleep in one position
	Trouble finding a comfortable position	Trouble finding a comfortable position
Psychological well-being (Mood)	Feeling frustrated	Feeling frustrated
	Feeling fed-up	Feeling fed-up
	Feeling annoyed or irritated	Feeling annoyed
		Feeling irritated
		Feeling bad tempered
	Feeling angry	Feeling angry
	Feeling miserable	Feeling miserable
	Feeling down	Feeling down
	Feeling depressed	Feeling depressed
(anxiety & worry)		Feeling fearful
		Feeling afraid
	Feeling upset	Feeling upset
	Feeling concerned or worried	Feeling concerned
		Feeling worried
	Feeling anxious	Feeling anxious
		Feeling surprised
		Feeling shocked
(Self-efficacy & dependence)	Feeling like a burden or nuisance on others	Feeling like a burden or nuisance on others
	Feeling like you have no control over your life because of your sore	Feeling like you have no control over your life
	Feeling physically dependent on others	Feeling physically dependent on others
(Appearance/self-consciousness)	Feeling helpless	Feeling helpless
		Feeling a lack of self-esteem
	Feeling self-conscious	Feeling self-conscious

Domain (Sub-domain)	Reduced Item List (87) – post pretest	Original Item list (118)
	Lacking in confidence	Feeling a lack of self-confidence
	Feeling embarrassed	Feeling embarrassed
	Feeling physically unattractive	Feeling physically unattractive
		Feeling disinterested in socialising
	Feeling uneasy being close to or around other people	Feeling uneasy being close to people
		Feeling worried about how others will react to your ulcer
	Feeling a lack of understanding from those close to you	Feeling a lack of understanding from those close to you
Social functioning		Feeling left out
(Isolation)		Feeling isolated
	Feeling cut off or isolated from others	Feeling cut off
	Feeling lonely	Feeling lonely
	Feeling like you were missing out	Feeling like you were missing out
	Feeling like people avoided you or treated you differently now	Feeling like people avoided you or treated you differently now
(Participation)	Difficulty going out	Difficulty going out
		Being unable to meet up with others
	Difficulty meeting up or seeing family and/or friends	Difficulty seeing family and/or friends
	Being unable to participate in family gatherings or activities	Being unable to participate in family gatherings or activities
	Having to plan going out around ulcer care	Having to plan going out around ulcer care
		Being unable to do things spontaneously
	Having to give up on hobbies or leisure activities	Giving up on hobbies or leisure activities
	Being restricted to where you could go out	Being restricted to where you could go out
	Being restricted to how long you could stay out	Being restricted to how long you could stay out
	Being unable to get away for a holiday or make a trip at the weekend	Being unable to get away for a holiday or make a trip at the weekend
	The amount of time involved in caring for your ulcer	The amount of time involved in caring for your ulcer

Appendix 5.2 - Preliminary PU-QOL version

Pressure Ulcer Quality of Life Questionnaire (PU-QOL v1)

This questionnaire asks your views about the impact that your pressure ulcer(s) has had on your everyday life during the past week.

To start, please tick the box(es) below to indicate the part(s) of your body where you currently have a pressure ulcer(s).

I currently have a pressure ulcer on my: (*tick all that apply*)

☐ Sacrum (the area at the bottom of your spine area)	☐ Ankle/foot
☐ Buttocks	☐ Heel
☐ Back of leg and/or thigh	☐ Elbow
☐ Hip	☐ Head and/or face
☐ Other, please specify:	

This questionnaire asks about your physical and social functioning, mood, and any symptoms that you may have experienced as a result of having a pressure ulcer(s). All of the information you provide is completely confidential.

Please read all the questions that follow, and for each one choose the answer that best describes your experience by ticking ☑ in the box. Only tick one box for each question.

There are no right or wrong answers. If you are not sure how to answer a question, the first response you think of is often the best one. If you have more than one pressure ulcer, please try to think about the overall impact of your pressure ulcers when answering the questions.

People with pressure ulcers may experience some type of pain or discomfort.

1. During the <u>past week</u>, how much were you bothered by these types of pain or discomfort from your pressure ulcer(s)?	Not at all	A little	Quite a bit	A lot
a) Feeling uncomfortable	☐ 0	☐ 1	☐ 2	☐ 3
b) Annoying pain or discomfort	☐ 0	☐ 1	☐ 2	☐ 3
c) Itchiness	☐ 0	☐ 1	☐ 2	☐ 3
d) Tenderness	☐ 0	☐ 1	☐ 2	☐ 3
e) Niggling	☐ 0	☐ 1	☐ 2	☐ 3
f) Soreness	☐ 0	☐ 1	☐ 2	☐ 3
g) Aching	☐ 0	☐ 1	☐ 2	☐ 3
h) Pins and needles	☐ 0	☐ 1	☐ 2	☐ 3
i) Tingling	☐ 0	☐ 1	☐ 2	☐ 3
j) Throbbing	☐ 0	☐ 1	☐ 2	☐ 3
k) Nagging	☐ 0	☐ 1	☐ 2	☐ 3
l) Shooting	☐ 0	☐ 1	☐ 2	☐ 3
m) Stinging	☐ 0	☐ 1	☐ 2	☐ 3
n) Stabbing	☐ 0	☐ 1	☐ 2	☐ 3
o) Electric shocks	☐ 0	☐ 1	☐ 2	☐ 3
p) Red raw	☐ 0	☐ 1	☐ 2	☐ 3
q) Burning	☐ 0	☐ 1	☐ 2	☐ 3

2. During the <u>past week</u>, how would you rate the overall pain or discomfort you experienced from your pressure ulcer(s)?

- ☐ None
- ☐ Very mild
- ☐ Mild
- ☐ Moderate
- ☐ Severe
- ☐ Very severe (excruciating)

3. If you have experienced any pain or discomfort from your pressure ulcer(s) during the <u>past week</u>, how would you describe it?

- ☐ I did not experience any pain/discomfort during the past week
- ☐ Occasional
- ☐ Intermittent
- ☐ Only when I move
- ☐ Only when the ulcer is touched
- ☐ Only when the ulcer dressing is changed
- ☐ Frequent
- ☐ Constant

People with pressure ulcers may experience leaking from their ulcer(s).

4. During the <u>past week</u>, how much were you bothered by these types of leaking from your pressure ulcer(s)?	Not at all	A little	Quite a bit	A lot
a) Weeping	☐ 0	☐ 1	☐ 2	☐ 3
b) Oozing	☐ 0	☐ 1	☐ 2	☐ 3
c) Running	☐ 0	☐ 1	☐ 2	☐ 3
d) Sticky	☐ 0	☐ 1	☐ 2	☐ 3
e) Slimy	☐ 0	☐ 1	☐ 2	☐ 3
f) Wet	☐ 0	☐ 1	☐ 2	☐ 3
g) Messy	☐ 0	☐ 1	☐ 2	☐ 3
h) Staining	☐ 0	☐ 1	☐ 2	☐ 3
i) Causing dressing to come off	☐ 0	☐ 1	☐ 2	☐ 3
j) Gungy	☐ 0	☐ 1	☐ 2	☐ 3
k) Pus	☐ 0	☐ 1	☐ 2	☐ 3
l) Bleeding	☐ 0	☐ 1	☐ 2	☐ 3

People with pressure ulcers may experience a smell or odour from their ulcer.

5. During the <u>past week</u>, how much were you bothered by these types of smell or odour from your pressure ulcer(s)?	Not at all	A little	Quite a bit	A lot
a) An unpleasant smell	☐ 0	☐ 1	☐ 2	☐ 3
b) A lingering smell	☐ 0	☐ 1	☐ 2	☐ 3
c) A dirty smell	☐ 0	☐ 1	☐ 2	☐ 3
d) A foisty smell	☐ 0	☐ 1	☐ 2	☐ 3
e) A stench	☐ 0	☐ 1	☐ 2	☐ 3
f) A stink	☐ 0	☐ 1	☐ 2	☐ 3
g) A stale smell	☐ 0	☐ 1	☐ 2	☐ 3
h) A pungent smell	☐ 0	☐ 1	☐ 2	☐ 3
i) A sickening smell	☐ 0	☐ 1	☐ 2	☐ 3
j) A putrid smell	☐ 0	☐ 1	☐ 2	☐ 3

Pressure ulcers may interfere with a person's quality of sleep.

6. During the <u>past week</u>, how much were you bothered by these sleep problems because of your pressure ulcer(s)?	Not at all	A little	Quite a bit	A lot
a) Trouble falling asleep	☐ 0	☐ 1	☐ 2	☐ 3
b) A restless sleep	☐ 0	☐ 1	☐ 2	☐ 3
c) Being kept awake	☐ 0	☐ 1	☐ 2	☐ 3
d) Being woken up during your sleep	☐ 0	☐ 1	☐ 2	☐ 3
e) Not getting the amount of sleep that you needed	☐ 0	☐ 1	☐ 2	☐ 3
f) Having to sleep in one position (e.g. your back or side)	☐ 0	☐ 1	☐ 2	☐ 3
g) Trouble finding a comfortable position	☐ 0	☐ 1	☐ 2	☐ 3

Now some questions about everyday movements.

7. During the <u>past week</u>, how much were you bothered by these everyday movements because of your pressure ulcer(s)?	Not at all	A little	Quite a bit	A lot
a) Difficulty sitting up in bed	☐ 0	☐ 1	☐ 2	☐ 3
b) Difficulty adjusting yourself in bed	☐ 0	☐ 1	☐ 2	☐ 3
c) Difficulty turning in bed	☐ 0	☐ 1	☐ 2	☐ 3
d) Difficulty pushing up to a sitting position	☐ 0	☐ 1	☐ 2	☐ 3
e) Difficulty sitting up in one position for long periods	☐ 0	☐ 1	☐ 2	☐ 3
f) Difficulty standing for long periods	☐ 0	☐ 1	☐ 2	☐ 3
g) Difficulty transferring from a bed to a chair	☐ 0	☐ 1	☐ 2	☐ 3
h) Feeling limited in your ability to walk	☐ 0	☐ 1	☐ 2	☐ 3
i) Feeling limited in your ability to go up and down stairs	☐ 0	☐ 1	☐ 2	☐ 3
j) Feeling limited in how far you were able to walk	☐ 0	☐ 1	☐ 2	☐ 3
k) Feeling that your walking was slowed down	☐ 0	☐ 1	☐ 2	☐ 3

Now some questions about how you might be feeling within yourself.

8. During the <u>past week</u>, how much were you bothered by these health aspects because of your pressure ulcer(s)?	Not at all	A little	Quite a bit	A lot
a) Feeling that your appetite has reduced	☐ 0	☐ 1	☐ 2	☐ 3
b) Feeling unwell or poorly	☐ 0	☐ 1	☐ 2	☐ 3
c) Feeling that your energy levels have been reduced (i.e. feeling tired or fatigued)	☐ 0	☐ 1	☐ 2	☐ 3

The following questions are about everyday activities.

9. During the <u>past week</u>, how much were you bothered by difficulty doing these everyday activities because of your pressure ulcer(s)?	None at all	A little	Quite a bit	A lot
a) Washing yourself in the bath or shower	☐ 0	☐ 1	☐ 2	☐ 3
b) Getting dressed or undressed	☐ 0	☐ 1	☐ 2	☐ 3
c) Doing housework (e.g. cooking or cleaning)	☐ 0	☐ 1	☐ 2	☐ 3
d) Doing gardening	☐ 0	☐ 1	☐ 2	☐ 3
e) Doing shopping	☐ 0	☐ 1	☐ 2	☐ 3
f) Going to the toilet	☐ 0	☐ 1	☐ 2	☐ 3
g) Being able to travel in or drive a car	☐ 0	☐ 1	☐ 2	☐ 3
h) Doing things that you enjoy (e.g. reading a book, watching a movie)	☐ 0	☐ 1	☐ 2	☐ 3
i) Getting up and about to do things that you enjoy	☐ 0	☐ 1	☐ 2	☐ 3
j) Being intimate with loved ones	☐ 0	☐ 1	☐ 2	☐ 3
k) Doing usual work (e.g. employment, volunteering, university, clubs)	☐ 0	☐ 1	☐ 2	☐ 3

Next are some questions about your feelings.

10. During the <u>past week</u>, how much were you bothered by these feelings because of your pressure ulcer(s)?	Not at all	A little	Quite a bit	A lot
a) Feeling frustrated	☐ 0	☐ 1	☐ 2	☐ 3
b) Feeling fed-up	☐ 0	☐ 1	☐ 2	☐ 3
c) Feeling annoyed	☐ 0	☐ 1	☐ 2	☐ 3
d) Feeling irritated	☐ 0	☐ 1	☐ 2	☐ 3
e) Feeling bad tempered	☐ 0	☐ 1	☐ 2	☐ 3
f) Feeling angry	☐ 0	☐ 1	☐ 2	☐ 3
g) Feeling miserable	☐ 0	☐ 1	☐ 2	☐ 3
h) Feeling down	☐ 0	☐ 1	☐ 2	☐ 3
i) Feeling depressed	☐ 0	☐ 1	☐ 2	☐ 3
j) Feeling fearful	☐ 0	☐ 1	☐ 2	☐ 3
k) Feeling afraid	☐ 0	☐ 1	☐ 2	☐ 3
l) Feeling upset	☐ 0	☐ 1	☐ 2	☐ 3
m) Feeling concerned	☐ 0	☐ 1	☐ 2	☐ 3
n) Feeling worried	☐ 0	☐ 1	☐ 2	☐ 3
o) Feeling anxious	☐ 0	☐ 1	☐ 2	☐ 3
p) Feeling surprised	☐ 0	☐ 1	☐ 2	☐ 3
q) Feeling shocked	☐ 0	☐ 1	☐ 2	☐ 3
r) Feeling helpless	☐ 0	☐ 1	☐ 2	☐ 3
s) Feeling a lack of self-confidence	☐ 0	☐ 1	☐ 2	☐ 3
t) Feeling a lack of self-esteem	☐ 0	☐ 1	☐ 2	☐ 3
u) Feeling self-conscious	☐ 0	☐ 1	☐ 2	☐ 3
v) Feeling embarrassed	☐ 0	☐ 1	☐ 2	☐ 3
w) Feeling physically unattractive	☐ 0	☐ 1	☐ 2	☐ 3
x) Feeling disinterested in socialising	☐ 0	☐ 1	☐ 2	☐ 3
y) Feeling uneasy being close to people	☐ 0	☐ 1	☐ 2	☐ 3
z) Feeling worried about how others will react to your ulcer	☐ 0	☐ 1	☐ 2	☐ 3

Some more questions about your feelings.

11. During the <u>past week</u>, how much were you bothered by these feelings because of your pressure ulcer(s)?	Not at all	A little	Quite a bit	A lot
a) Feeling physically dependent on others	☐ 0	☐ 1	☐ 2	☐ 3
b) Feeling left out	☐ 0	☐ 1	☐ 2	☐ 3
c) Feeling isolated	☐ 0	☐ 1	☐ 2	☐ 3
d) Feeling cut off	☐ 0	☐ 1	☐ 2	☐ 3
e) Feeling lonely	☐ 0	☐ 1	☐ 2	☐ 3
f) Feeling like you were missing out	☐ 0	☐ 1	☐ 2	☐ 3
g) Feeling that people avoided you or treated you differently now	☐ 0	☐ 1	☐ 2	☐ 3
h) Feeling a lack of understanding from those close to you	☐ 0	☐ 1	☐ 2	☐ 3
i) Feeling like a burden or nuisance on others	☐ 0	☐ 1	☐ 2	☐ 3
j) Feeling like you have no control over your life?	☐ 0	☐ 1	☐ 2	☐ 3

Next are some questions about your usual social activities with family, friends, neighbours and groups.

12. During the <u>past week</u>, how much were you bothered by these limitations in your social activities because of your pressure ulcer(s)?	Not at all	A little	Quite a bit	A lot
a) Difficulty going out	☐ 0	☐ 1	☐ 2	☐ 3
b) Being unable to meet up with others	☐ 0	☐ 1	☐ 2	☐ 3
c) Difficulty seeing family and/or friends	☐ 0	☐ 1	☐ 2	☐ 3
d) Being unable to participate in family gatherings or activities	☐ 0	☐ 1	☐ 2	☐ 3
e) Having to plan going out around ulcer care	☐ 0	☐ 1	☐ 2	☐ 3
f) Being able to do things spontaneously	☐ 0	☐ 1	☐ 2	☐ 3
g) Giving up on hobbies or leisure activities	☐ 0	☐ 1	☐ 2	☐ 3
h) Being restricted to where you could go out	☐ 0	☐ 1	☐ 2	☐ 3
i) Being restricted to how long you could stay out	☐ 0	☐ 1	☐ 2	☐ 3
j) Being unable to get away for a holiday or make a trip at the weekend	☐ 0	☐ 1	☐ 2	☐ 3
k) The amount of time involved in caring for your ulcer	☐ 0	☐ 1	☐ 2	☐ 3

13. We have asked about the impact of your pressure ulcer(s) on many aspects of your everyday life. Thinking about all of these aspects of your everyday life, how would you rate your overall quality of life?

☐ Very Good	☐ Good	☐ Fair	☐ Poor	☐ Very Poor

Thank you for completing the questions above. To end, some questions about you

Your initials? _____

Your date of birth? _____(day/month/year)_

Your gender? (please tick) Male ☐ Female ☐

How long you have had your pressure ulcer(s)? _____(months)_

Today's date: ___/___/_____ (day/month/year)

Appendix 5.3 Final pretested PU-QOLI for field testing

PURPOSE

Pressure Sore Quality of Life Questionnaire Test 1 Version

For office use only	Patient Initials	Date of Birth (Day Month Year)	Centre Number	Patient Study ID

Mode of administration: ☐ Self-complete ☐ Administered

Please read these instructions before you begin

This questionnaire asks your views about the impact that your pressure sore(s) has had on your everyday life **during the past week**.

We understand that you may have a combination of medical problems, but please try to think about only your pressure sore(s) when you answer the questions.

Please read and **answer all the questions** that follow by ticking ☑ in the box that best describes your experience.

If you have more than one pressure sore, please try to think about the pressure sore that has caused you the most bother when answering the questions.

PURPOSE

Pressure Sore Quality of Life Questionnaire Test 1 Version

To start, **how many** pressure sore(s) do you have? ☐

(Please write the number of pressure sores that you have in the box)

On **which part of your body** do you currently have pressure sore(s)?
(Please tick all that apply)

- ☐ Area at the bottom of your spine (sacrum)
- ☐ Buttocks
- ☐ Back of leg and/or thigh
- ☐ Hip
- ☐ Other, please specify: ☐
- ☐ Ankle/foot
- ☐ Heel
- ☐ Elbow
- ☐ Head and/or face

If you have experienced pain or discomfort because of your pressure sore(s), how would you describe it? (Please tick all that apply)

- ☐ I did not experience any pain/ discomfort during the past week
- ☐ Comes and goes (intermittent)
- ☐ When I move
- ☐ When I sit, stand or put pressure on my sore
- ☐ When the dressing is changed
- ☐ Frequent
- ☐ Constant

PURPOSE

Page 2 of 9

Pressure Sore Quality of Life Questionnaire Test 1 Version

1. During the **past week**, how much were you **bothered** by pain or discomfort because of your **pressure sore(s)**? (Please tick one box on each line)

	The pressure sore did not give me this problem	I had the problem because of pressure sores and it bothered me…			
		No bother at all	**A little bother**	**Quite a bit of bother**	**A lot of bother**
a) Feeling uncomfortable	☐	☐	☐	☐	☐
b) A dull ache	☐	☐	☐	☐	☐
c) Annoying pain or discomfort	☐	☐	☐	☐	☐
d) Itchiness	☐	☐	☐	☐	☐
e) Tenderness	☐	☐	☐	☐	☐
f) Tingling	☐	☐	☐	☐	☐
g) Throbbing	☐	☐	☐	☐	☐
h) Stinging	☐	☐	☐	☐	☐
i) Stabbing pains	☐	☐	☐	☐	☐
j) Red raw	☐	☐	☐	☐	☐
k) Burning	☐	☐	☐	☐	☐

2. During the **past week**, how would you rate the overall pain or discomfort you experienced because of your pressure sore(s)? (Please tick only one box)

- ☐ None
- ☐ Very mild
- ☐ Mild
- ☐ Moderate
- ☐ Severe
- ☐ Very severe (excruciating)

3. During the **past week**, how much were you **bothered** by leaking from your **pressure sore(s)**? (Please tick one box on each line)

	The pressure sore did not give me this problem	I had the problem because of pressure sores and it bothered me…			
		No bother at all	**A little bother**	**Quite a bit of bother**	**A lot of bother**
a) Weeping	☐	☐	☐	☐	☐
b) Running	☐	☐	☐	☐	☐
c) Sticky	☐	☐	☐	☐	☐
d) Messy	☐	☐	☐	☐	☐
e) Staining	☐	☐	☐	☐	☐
f) Causing dressing to come off	☐	☐	☐	☐	☐
g) Pus	☐	☐	☐	☐	☐
h) Bleeding	☐	☐	☐	☐	☐

4. During the **past week**, how much were you **bothered** by smell or odour from your **pressure sore(s)**? (Please tick one box on each line)

	The pressure sore did not give me this problem	I had the problem because of pressure sores and it bothered me…			
		No bother at all	**A little bother**	**Quite a bit of bother**	**A lot of bother**
a) An unpleasant smell	☐	☐	☐	☐	☐
b) A lingering smell	☐	☐	☐	☐	☐
c) A stench or stink	☐	☐	☐	☐	☐
d) A pungent smell	☐	☐	☐	☐	☐
e) A sickening smell	☐	☐	☐	☐	☐
f) A putrid smell	☐	☐	☐	☐	☐

PURPOSE

5. During the **past week**, how much were you **bothered** by sleep problems because of your **pressure sore(s)**? (Please tick one box on each line)

	The pressure sore did not give me this problem	I had the problem because of pressure sores and it bothered me…			
		No bother at all	**A little bother**	**Quite a bit of bother**	**A lot of bother**
a) Trouble falling asleep	☐	☐	☐	☐	☐
b) Interrupted sleep (e.g. restless sleep or being woken up during your sleep)	☐	☐	☐	☐	☐
c) Being kept awake	☐	☐	☐	☐	☐
d) Not getting the amount of sleep that you needed	☐	☐	☐	☐	☐
e) Having to sleep in one position (e.g. your back or side)	☐	☐	☐	☐	☐
f) Trouble finding a comfortable position	☐	☐	☐	☐	☐

6. During the **past week**, how much were you **bothered** by health aspects because of your **pressure sore(s)**? (Please tick one box on each line)

	The pressure sore did not give me this problem	I had the problem because of pressure sores and it bothered me…			
		No bother at all	**A little bother**	**Quite a bit of bother**	**A lot of bother**
a) Feeling that your appetite has reduced	☐	☐	☐	☐	☐
b) Feeling unwell or poorly	☐	☐	☐	☐	☐
c) Feeling that your energy levels have been reduced (i.e. feeling tired, fatigued)	☐	☐	☐	☐	☐

PURPOSE

Pressure Sore Quality of Life Questionnaire Test 1 Version

7. During the **past week**, how much were you **bothered** by these everyday movements because of your **pressure sore(s)**? (Please tick one box on each line)

	The pressure sore did not give me this problem	I had the problem because of pressure sores and it bothered me…			
		No bother at all	**A little bother**	**Quite a bit of bother**	**A lot of bother**
a) Difficulty sitting up in bed	☐	☐	☐	☐	☐
b) Difficulty adjusting yourself in bed	☐	☐	☐	☐	☐
c) Difficulty turning or moving around in bed	☐	☐	☐	☐	☐
d) Difficulty pushing up to a sitting position	☐	☐	☐	☐	☐
e) Difficulty sitting up in one position for long periods	☐	☐	☐	☐	☐
f) Difficulty transferring (e.g. from a bed to a chair or to a car)	☐	☐	☐	☐	☐
g) Difficulty standing for long periods	☐	☐	☐	☐	☐
h) Feeling limited in your ability to go up and down stairs	☐	☐	☐	☐	☐
i) Feeling limited in your ability to walk	☐	☐	☐	☐	☐
j) Feeling limited in how far you were able to walk	☐	☐	☐	☐	☐
k) Feeling that your walking was slowed down	☐	☐	☐	☐	☐

8. The following questions are about everyday activities. During the **past week**, how much were you **bothered** by difficulty doing everyday activities because of your **pressure sore(s)**? (Please tick one box on each line)

	The pressure sore did not give me this problem	I had the problem because of pressure sores and it bothered me…			
		No bother at all	**A little bother**	**Quite a bit of bother**	**A lot of bother**
a) Being able to wash yourself in your usual way (e.g. hand wash, bath, shower)	☐	☐	☐	☐	☐
b) Getting dressed or undressed	☐	☐	☐	☐	☐
c) Being able to go to the toilet	☐	☐	☐	☐	☐
d) Doing jobs around the house (e.g. cooking, housework, DIY)	☐	☐	☐	☐	☐
e) Doing gardening	☐	☐	☐	☐	☐
f) Doing shopping	☐	☐	☐	☐	☐
g) Doing things that you enjoy (e.g. reading a book, watching a movie, using a computer)	☐	☐	☐	☐	☐
h) Being emotionally close or affectionate with loved ones	☐	☐	☐	☐	☐
i) Doing your regular daily activities (e.g. work, volunteering, religious service, clubs, university,)	☐	☐	☐	☐	☐

PURPOSE

Page 7 of 9

Pressure Sore Quality of Life Questionnaire Test 1 Version

9. During the **past week**, how much were you **bothered** by these feelings because of your **pressure sore(s)**? (Please tick one box on each line)

	The pressure sore did not give me this problem	I had the problem because of pressure sores and it bothered me…			
		No bother at all	**A little bother**	**Quite a bit of bother**	**A lot of bother**
a) Feeling frustrated	☐	☐	☐	☐	☐
b) Feeling fed-up	☐	☐	☐	☐	☐
c) Feeling annoyed or irritated	☐	☐	☐	☐	☐
d) Feeling angry	☐	☐	☐	☐	☐
e) Feeling miserable	☐	☐	☐	☐	☐
f) Feeling down	☐	☐	☐	☐	☐
g) Feeling depressed	☐	☐	☐	☐	☐
h) Feeling upset	☐	☐	☐	☐	☐
i) Feeling concerned or worried	☐	☐	☐	☐	☐
j) Feeling anxious	☐	☐	☐	☐	☐
k) Feeling helpless	☐	☐	☐	☐	☐
l) Lacking in confidence	☐	☐	☐	☐	☐
m) Feeling self-conscious	☐	☐	☐	☐	☐
n) Feeling embarrassed	☐	☐	☐	☐	☐
o) Feeling physically unattractive	☐	☐	☐	☐	☐
p) Feeling uneasy being close to or around other people	☐	☐	☐	☐	☐
q) Feeling a lack of understanding from those close to you	☐	☐	☐	☐	☐
r) Feeling like a burden or nuisance on others	☐	☐	☐	☐	☐
s) Feeling like you have no control over your life because of your sore	☐	☐	☐	☐	☐

PURPOSE

Page 8 of 9

Pressure Sore Quality of Life Questionnaire Test 1 Version

10. During the **past week**, how much were you **bothered** by these feelings because of your **pressure sore(s)**? (Please tick one box on each line)

	The pressure sore did not give me this problem	I had the problem because of pressure sores and it bothered me…			
		No bother at all	**A little bother**	**Quite a bit of bother**	**A lot of bother**
a) Feeling physically dependent on others	☐	☐	☐	☐	☐
b) Feeling cut off or isolated from others	☐	☐	☐	☐	☐
c) Feeling lonely	☐	☐	☐	☐	☐
d) Feeling like you were missing out	☐	☐	☐	☐	☐
e) Feeling that people avoided you or treated you differently now	☐	☐	☐	☐	☐

11. The next questions are about your usual social activities. During the **past week**, how much were you **bothered** by limitations in your social activities because of your **pressure sore(s)**? (Please tick one box on each line)

	The pressure sore did not give me this problem	I had the problem because of pressure sores and it bothered me…			
		No bother at all	**A little bother**	**Quite a bit of bother**	**A lot of bother**
a) Difficulty going out	☐	☐	☐	☐	☐
b) Difficulty meeting up or seeing family and/or friends	☐	☐	☐	☐	☐
c) Being unable to participate in family gatherings or activities	☐	☐	☐	☐	☐
d) Having to plan going out around pressure sore care	☐	☐	☐	☐	☐
e) Having to give up on hobbies or leisure activities	☐	☐	☐	☐	☐
f) Being restricted to where you could go out	☐	☐	☐	☐	☐

PURPOSE

Page 9 of 9

Pressure Sore Quality of Life Questionnaire Test 1 Version

12. During the **past week**, how much were you **bothered** by limitations in your social activities because of your **pressure sore(s)**? (Please tick one box on each line)

	The pressure sore did not give me this problem	I had the problem because of pressure sores and it bothered me…			
		No bother at all	**A little bother**	**Quite a bit of bother**	**A lot of bother**
a) Being restricted to how long you could stay out	☐	☐	☐	☐	☐
b) Being unable to get away for a holiday or make a trip at the weekend	☐	☐	☐	☐	☐
c) The amount of time involved in caring for your sore	☐	☐	☐	☐	☐

13. We have asked about the **impact of your pressure sore(s)** on many aspects of your everyday life. Thinking about all of these aspects, how would you rate your **overall quality of life** because of your pressure sore(s)? (Please tick only one box)

☐ Very Good ☐ Good ☐ Fair ☐ Poor ☐ Very Poor

Finally, please fill in the following details:

Your initials _____

Your date of birth _____/_____/_____ (day/month/year)

Your gender (please tick) ☐ Male ☐ Female

How long you have had your pressure sore(s) _____ (weeks) _____ (months) _____ (years)

Today's date: _____/_____/_____ (day/month/year)

Please check that you have answered all the questions on each page

THANK YOU FOR COMPLETING THESE QUESTIONS

Appendix 6.1 has been removed.
Please refer to the thesis abstract page for more information: http://etheses.whiterose.ac.uk/2422