

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

**WHO Long Form Scoring, Reliability, Validity and Norms for New
Zealand**

**A thesis presented in fulfilment of the requirements for the degree of
Master of Public Health**

at Massey University, Wellington Campus, New Zealand

Karen Sarah Blakey

July 2006

Abstract

Background

Self-reported health measures provide information about a wider range of health outcomes than objective measures of health status, such as mortality and hospitalisation rates. National health surveys play a role in monitoring population health. The New Zealand Health Monitor (NZHM) is the organised, co-ordinated and integrated survey programme of the Ministry of Health in New Zealand. The New Zealand Health Survey (NZHS) is one of the chief surveys of the NZHM. One of the categories of information collected in the NZHM is health outcomes, and within this there is the subcategory of health status. The International Classification of Functioning and Disability (ICF) provides the framework to describe the critical elements of non-fatal health outcomes captured by health status instruments. NZHM is to collect data on most if not all of these 21 ICF dimensions.

The WHO Long Form was developed as the health module in the WHO Multi-country Survey Study. The WHO Long Form is made up of 20 health domains, some overlapping with the eight SF-36 domains. The WHO Long Form did not have a set scoring system for scales, unlike the SF-36 instrument. The SF-36 has been previously tested and validated in New Zealand in the 1996/97 NZHS.

Methods

The 2002/03 NZHS used a complex sample design. A total of 12,929 people responded to the survey, with 12,529 respondents being included in the CURF dataset available for research. The health status section of the 2002/03 NZHS measures health-related quality of life (HRQL) covered 16 health and health-related domains. The questions were derived from the SF-36 and the WHO Long Form questionnaire on health status. The health domains covered in the 2002/03 NZHS were general health, vision, hearing, digestion, breathing, pain, sleep, energy and vitality, understanding, communication, physical functioning, self-care. The health-related domains covered in the 2002/03 NZHS were mental health, role-physical and role-emotional (usual activities), and social functioning.

There were five key aims specific to the current thesis. First, to group the WHO Long Form items in the 2002/03 NZHS into scales for each health domain and develop standard scoring protocols for each scale. Second, to test the reliability of the scales using standard

psychometric tests for the total NZ population and for major population subgroups. Third, to test the validity of the scales using the standard psychometric tests for the total NZ population and for major population subgroups. Fourth, to construct norms for the WHO Long Form scales for the NZ population. And finally, to provide recommendations for the health status component of future NZ health surveys.

Results

In summary, this thesis developed a method for producing scale scores for domains of health not previously measured in New Zealand Health Surveys, providing greater coverage of domains from the ICF. There were virtually no missing data for all items and subgroups within the questions used to develop the scales. The scaling approach was consistent with that for the SF-36, allowing the new scales to be presented alongside the SF-36 scales. All scales for the total population and major population subgroups met the required criterion for satisfactory psychometric properties, with the exception of digestion and bodily excretions scale. For the digestion and bodily excretions scale, the Cronbach's alpha was lower than that required for between group comparisons. The composite physical functioning and social functioning scales performed no better than the existing SF-36 scales and were highly correlated with these scales.

Conclusion

Notwithstanding the limitations of this study, key findings of interest are that the new WHO Long Form questions can be used to form scales that cover physical functioning, social functioning, vision, hearing, digestion and bodily excretions, breathing, self-care, understanding, communication and sleep. The majority of the questions and scales work for the NZ population and subgroups. All but one of the scales, digestion and bodily excretions, have satisfactory psychometric properties for the total population and major subpopulation groups of interest. The respondent burden is an important consideration for the NZHS, thus it cannot be argued that enough is gained from adding questions to the physical functioning and Social Functioning domains, thus it would be recommended that the SF-36 scales are used to measure these two domains of health. The new WHO Long Form scales can now be presented alongside the SF-36 scales and used in future analyses looking at interrelationships between factors such as health risk and health status.

Acknowledgements

Firstly I would like to thank the participants of the 2002/03 NZHS who gave freely of their time to take part in the survey. Without them this research would never have been possible!

A large number of people have provided me with support and encouragement during the course of this research. In particular I am indebted to Professor Neil Pearce, Director Centre for Public Health Research (CPHR) at Massey University in Wellington and Dr Barry Borman, Manager of Public Health Intelligence (PHI) in the Ministry of Health. Neil and Barry were both my two bosses and thesis supervisors during my time as Training Fellow with CPHR and PHI. I thank you both for believing that I would finish the thesis (despite there being times when I thought I would never get to the end of the tunnel!) and providing the environment for me to develop both professionally and personally. I am also especially grateful to Martin Tobias, Principal Technical Specialist (Epidemiology) at PHI for providing the initial motivation for this research. I am indebted to Matthew Cronin, previously employed as a statistician with PHI for his assistance with the statistical analyses. I would also like to thank Maria Turley, Senior Advisor (Nutrition/Epidemiology) for ongoing support as I carried out this research.

Finally I wish to acknowledge my wonderful partner Ralf who cooked me the most delicious meal ever to celebrate the completion of my first full thesis draft. Ralf, you have greatly enriched my life and I thank you for all the love and support that you give me.

I thought I would end with a quote from Albert Einstein, "Learn from yesterday, live for today, hope for tomorrow. The important thing is not to stop questioning."

Table of contents

Abstract	2
Background	2
Methods	2
Results	3
Conclusion	3
Acknowledgements	4
Table of contents	5
List of Tables	9
List of Figures	11
Chapter 1 - Introduction	12
Background	12
Thesis outline	14
Chapter 2: Background - Measuring health-related quality of life	16
Measuring health status	16
<i>Concepts of health</i>	<i>16</i>
<i>International Classification of Functioning, Disability and Health (ICF)</i>	<i>17</i>
<i>Measurement instruments</i>	<i>19</i>
SF-20	21
<i>SF-20 Development</i>	<i>21</i>
<i>SF-20 Content</i>	<i>21</i>
<i>Physical functioning</i>	<i>22</i>
<i>Role functioning</i>	<i>22</i>
<i>Bodily pain</i>	<i>22</i>
<i>Current health</i>	<i>22</i>
<i>Social functioning</i>	<i>22</i>
<i>Mental health</i>	<i>22</i>
SF-36	23
<i>Development of SF-36</i>	<i>23</i>
<i>SF-36 Content</i>	<i>23</i>
<i>Physical functioning</i>	<i>25</i>
<i>Role functioning – Physical and Emotional</i>	<i>26</i>
<i>Bodily pain</i>	<i>26</i>
<i>General health</i>	<i>26</i>
<i>Vitality</i>	<i>26</i>
<i>Social functioning</i>	<i>26</i>
<i>Mental health</i>	<i>26</i>

Comparison of SF-20 and SF-36 scales	27
Physical functioning	27
Role functioning – Physical and Emotional	27
Bodily pain	27
General health	28
Vitality	28
Social functioning	28
Mental health	28
Scoring	28
Standardisation	29
Item recoding	29
Item recalibration	29
Computing scale scores	30
Summary measures	30
Psychometric properties	31
Reliability	31
Validity	31
Data quality	32
Scaling assumptions	32
WHO Long Form	33
Development of the WHO Long Form	33
WHO Long Form Content	33
Application of SF-36 to New Zealand	34
Australian/New Zealand adaptation of SF-36	34
Adaptation	34
Internal consistency	35
Construct validity	35
Items and domains	35
Psychometric properties	37
Data quality	37
Scaling assumptions	37
Reliability	37
Validity	37
Norms	38
Issues for use with Māori and Pacific peoples	38
Problems with principal components	38
Māori/Pacific models of health	39
Application of WHO Long Form to New Zealand	40
Items and domains	40
WHO Long Form Validation	40
Chapter 3: Methods	42
New Zealand Health Monitor	42
New Zealand Health Survey	44
Methodology of the 2002/03 NZHS	44
Aims	44
Sample design	45
Sample selection	46

<i>Data collection</i>	47
<i>Sample allocation</i>	48
<i>Questionnaire</i>	48
<i>Response rate</i>	48
<i>Quality control</i>	49
<i>Weighting estimation</i>	49
<i>CURF</i>	50
<i>Data reliability</i>	50
Health Status module of 2002/03 NZHS	50
<i>Embedding of SF-36 within WHO Long Form</i>	55
Statistical analysis	55
<i>Scoring</i>	55
<i>Subgroups</i>	56
<i>Reliability analysis</i>	56
<i>Items</i>	56
<i>Scales</i>	57
<i>Validity analysis</i>	59
<i>Construct validity</i>	59
<i>Discriminative validity</i>	60
<i>Norms</i>	61
<i>Data</i>	61
<i>Software</i>	61
Chapter 4: Results - Scoring, reliability, validity and norms	62
Sample characteristics	62
<i>Response rate</i>	62
<i>Sex</i>	62
<i>Age Group</i>	63
<i>Ethnic Group</i>	63
<i>Deprivation</i>	64
Scoring	65
<i>SF-36 Domains</i>	65
<i>WHO Long Form Domains</i>	65
<i>Process for summing scales</i>	65
Reliability analysis	66
<i>Items</i>	66
<i>Completeness of data</i>	66
<i>Descriptive statistics</i>	66
<i>Scales</i>	69
<i>Descriptive statistics</i>	69
<i>Summated rating scale assumptions</i>	85
<i>Reliability of scale scores</i>	98
<i>Deprivation Quintile</i>	101
<i>Comparative analysis for physical functioning scales</i>	102
<i>Comparative analysis for social functioning scales</i>	103

Validity analysis	104
<i>Construct validity</i>	<i>104</i>
<i>Scaling success rates</i>	<i>104</i>
<i>Relationships between scales</i>	<i>104</i>
<i>Discriminative validity</i>	<i>114</i>
<i>Summary</i>	<i>115</i>
<i>Item completeness and item internal consistency</i>	<i>116</i>
<i>Reliability</i>	<i>117</i>
<i>Floor and ceiling effects</i>	<i>117</i>
National Norms	117
Chapter 5 - Discussion	124
Introduction	124
Summary of findings	128
<i>Scoring</i>	<i>128</i>
<i>Reliability analysis</i>	<i>128</i>
<i>Validity analysis</i>	<i>129</i>
<i>National norms</i>	<i>129</i>
References	134
Appendix I – 2002/03 NZHS Relevant sections of questionnaire	137

List of Tables

Table 1. Key dimensions of health recognised by the ICF.....	18
Table 2. Number of items and scale levels in the SF-20	21
Table 3. Number of items and scale levels in the SF-36	24
Table 4. SF-36 Domains and abbreviated item content.....	25
Table 5. WHO Long Form Health and Health-Related Domains	34
Table 6. SF-36 scale and abbreviated item content in Health status module of NZHS 1996/97 and 2002/03	36
Table 7. WHO Long Form scales and abbreviated item content in Health status module of 2002/03 NZHS.....	41
Table 8. WHO Long Form and Composite Domains and abbreviated item content.....	52
Table 9. Number of domains covered by SF-36, WHO Long Form (NZ Version) and composite instrument used in Health status module of 2002/03 NZHS.....	54
Table 10. Total sample of 2002/03 NZHS and target population by age group.....	63
Table 11. Total sample of 2002/03 NZHS and target population by ethnic group	63
Table 12. Total sample of 2002/03 NZHS and target population by deprivation quintile	64
Table 13. Item frequency distributions in percent for total population	67
Table 14. Scale descriptive statistics for total population	70
Table 15. Scale descriptive statistics for males	72
Table 16. Scale descriptive statistics for females	73
Table 17. Scale descriptive statistics for 15-24 years.....	74
Table 18. Scale descriptive statistics for 25-44 years.....	75
Table 19. Scale descriptive statistics for 45-64 years.....	75
Table 20. Scale descriptive statistics for 65-74 years.....	76
Table 21. Scale descriptive statistics for 75+ years.....	77
Table 22. Scale descriptive statistics for Māori.....	78
Table 23. Scale descriptive statistics for Pacific peoples	79
Table 24. Scale descriptive statistics for Asian peoples.....	80
Table 25. Scale descriptive statistics for European/Other	81
Table 26. Scale descriptive statistics for Deprivation Quintile 1	82
Table 27. Scale descriptive statistics for Deprivation Quintile 2	83
Table 28. Scale descriptive statistics for Deprivation Quintile 3	83
Table 29. Scale descriptive statistics for Deprivation Quintile 4	84
Table 30. Scale descriptive statistics for Deprivation Quintile 5	84
Table 31. Correlations between items and scales for total population	86

Table 32. Results of item scaling tests for total population.....	87
Table 33. Results of item scaling tests for males	88
Table 34. Results of item scaling tests for females	88
Table 35. Results of item scaling tests for 15-24 years	89
Table 36. Results of item scaling tests for 25-44 years	90
Table 37. Results of item scaling tests for 45-64 years	90
Table 38. Results of item scaling tests for 65-74 years	91
Table 39. Results of item scaling tests for 75+ years	92
Table 40. Results of item scaling tests for Māori	93
Table 41. Results of item scaling tests for Pacific peoples	93
Table 42. Results of item scaling tests for Asian peoples	94
Table 43. Results of item scaling tests for European/Other	94
Table 44. Results of item scaling tests for Deprivation Quintile 1.....	95
Table 45. Results of item scaling tests for Deprivation Quintile 2.....	96
Table 46. Results of item scaling tests for Deprivation Quintile 3.....	96
Table 47. Results of item scaling tests for Deprivation Quintile 4.....	97
Table 48. Results of item scaling tests for Deprivation Quintile 5.....	97
Table 49. Reliability estimates for scales (Cronbach's alpha) for the total population.....	98
Table 50. Reliability estimates for scales (Cronbach's alpha) by sex	99
Table 51. Reliability estimates for scales (Cronbach's alpha) by age group	100
Table 52. Reliability estimates for scales (Cronbach's alpha) by ethnic group	101
Table 53. Reliability estimates for scales (Cronbach's alpha) by deprivation quintile.....	102
Table 54. Comparison of SF-36 and Composite physical functioning scales in 2002/03 NZHS for total population	103
Table 55. Comparison of SF-36 and Composite social functioning scales in 2002/03 NZHS for total population	104
Table 56. Reliability coefficients and inter-scale correlations for total population	105
Table 57. Reliability coefficients and inter-scale correlations for males	106
Table 58. Reliability coefficients and inter-scale correlations for females	106
Table 59. Reliability coefficients and inter-scale correlations for 15-24 years.....	107
Table 60. Reliability coefficients and inter-scale correlations for 25-44 years.....	108
Table 61. Reliability coefficients and inter-scale correlations for 45-64 years.....	108
Table 62. Reliability coefficients and inter-scale correlations for 65-74 years.....	109
Table 63. Reliability coefficients and inter-scale correlations for 75+ years.....	109
Table 64. Reliability coefficients and inter-scale correlations for Māori.....	110
Table 65. Reliability coefficients and inter-scale correlations for Pacific peoples	110

Table 66. Reliability coefficients and inter-scale correlations for Asian peoples	111
Table 67. Reliability coefficients and inter-scale correlations for European/Other	111
Table 68. Reliability coefficients and inter-scale correlations for Quintile 1	112
Table 69. Reliability coefficients and inter-scale correlations for Quintile 2	112
Table 70. Reliability coefficients and inter-scale correlations for Quintile 3	113
Table 71. Reliability coefficients and inter-scale correlations for Quintile 4	113
Table 72. Reliability coefficients and inter-scale correlations for Quintile 5	114
Table 73. Discriminative ability of the WHO Long Form (NZ Version) scales between subgroups with and without chronic diseases reported	115
Table 74. Summary of tests of scales associated with best and worst results	116
Table 75: WHO Long Form (NZ Version) norms for the total population	117
Table 76. WHO Long Form (NZ Version) norms by age group for males (standard errors)	120
Table 77. WHO Long Form (NZ Version) norms by age group for females (standard errors)	120
Table 78. WHO Long Form (NZ Version) norms for by ethnic group (standard errors)	121
Table 79. WHO Long Form (NZ Version) norms for by deprivation (NZDep01) quintile (standard errors).....	122

List of Figures

Figure 1. Evan and Stoddard's model of health	17
Figure 2. Model of health and its causes	19
Figure 3: WHO Long Form (NZ Version) norms by sex	118
Figure 4. WHO Long Form (NZ Version) norms by life cycle age group	119
Figure 5. WHO Long Form (NZ Version) norms by ethnic group	122
Figure 6. WHO Long Form (NZ Version) Profile by deprivation quintile	123

Chapter 1 - Introduction

Background

Due to the rising burden of chronic disease and a decrease in mortality from infectious diseases, the use of measures of health-related quality of life (HRQL) has increased. Self-reported health measures provide information about a wider spectrum of health outcomes than objective measures of health status, such as mortality and hospitalisation rates.(Scott, Tobias, & Sarfati, 1999)

A common way to measure health status or health-related quality of life is through scoring standardised responses to standardised questions (Lohr, 1989). Measures of positive health are important when measuring the health of the general population (Bowling, 1997). Self-reports introduce a consumer perspective to population health monitoring. The emphasis of such measures is on quality of life and wellbeing (Ministry of Health, 1999a, 1999b).

Numerous questionnaires have been developed to measure HRQL, but the Medical Outcomes Study Short Form 36 (SF-36), which was developed in the United States, is the most widely used (Bowden & Fox-Rushby, 2003) & (Scott et al., 1999).

National health surveys play a role in monitoring population health. The New Zealand Health Monitor (NZHM) is the organised, co-ordinated and integrated survey programme of the Ministry of Health in New Zealand (NZ). The aim of the New Zealand Health Monitor (NZHM) is to collect data that cannot be collected more effectively and efficiently through other means (e.g. administrative data collection or epidemiologic studies), are needed to inform decision making of the Ministry of Health or district health boards, and are population based (Ministry of Health, 2005).

One of the information domains of the NZHM is “health outcomes”, and within this there is “health status”. The two portions of health status relevant to this thesis are subjective (self-rated) health and functional limitation. The International Classification of Functioning and Disability (ICF) provides the framework to describe the essential elements of non-fatal health outcomes measured by health status instruments. The ICF was approved by the World Health Organisation (WHO) in May 2001 and identifies 21 key domains of health (World Health Organization, 2001), and the NZHM aims to collect data on most if not all of these dimensions.

The New Zealand Health Survey (NZHS) is one of the main surveys of the NZHM. The 1996/97 NZHS was a nationally representative survey of the general population that measured health status, health-related behaviour and health service utilisation. The major measure of self-reported health status (HRQL) was the SF-36 (Australia/New Zealand adaptation). The SF-36 measures eight health domains using eight scales: physical functioning, role-limitation, bodily pain, general health, vitality, social functioning, role-emotional and mental health. From these eight scales, two summary measures are calculated, the physical component score and the mental component score (Scott et al., 1999)). The SF-36 performed reasonably well nationally meeting standard criteria for psychometric assessment, but not for older Māori or Pacific peoples (Scott, Sarfati, Tobias, & Haslett, 2000).

An alternative survey instrument for measuring health status is the WHO Long Form which was developed as the health module in the WHO Multi-country Survey Study. The aim of the health module was to develop valid, reliable and comparable instruments to describe a core set of health domains (Ustin et al., 2001). An extensive review of existing instruments was carried out using the International Classification of Functioning, Disability and Health (ICF) Framework (Ustin et al., 2001). The WHO Long Form consists of 20 health domains, some overlapping with the eight SF-36 domains (Ustin et al., 2001). The WHO Long Form does not have a standard scoring system for scales, unlike the SF-36 instrument.

One of the aims of the 2002/03 NZHS was “to measure the health status of New Zealand adults, including their self-reported physical and mental health status, and the prevalence of selected health conditions” (Ministry of Health, 2004 p1). The 2002/03 New Zealand Health Survey self-reported health status (HRQL) module included a combination of the SF-36 and the WHO Long Form (NZ Version).

The health status component of the health survey was broadened to cover a greater proportion of the health and health-related domains from the ICF - fifteen health domains were covered in the 2002/03 NZHS, compared to the eight covered by the SF-36 alone in the previous health survey completed in 1996-1997. These domains were the SF-36 domains outlined previously, plus physical functioning, social functioning, vision, hearing, digestion and bodily excretions, breathing, self-care, understanding, communication and sleep.

Questionnaires can have different meanings in different cultures and countries, and also within countries between population subgroups. Also, one aim of the 2002/03 NZHS was to “examine differences between population subgroups (as defined by sex, ethnicity, age and the

New Zealand Deprivation Index 2001 (NZDep01)” (Ministry of Health, 2004 p1) Thus it is important to assess the psychometric performance of instruments for the NZ population and subgroups within it, before national norms can be used in practice (Scott et al., 1999, Gandek, 1998). Thus it is necessary to validate instruments for population subgroups as well as the total NZ population. The SF-36 was validated in the previous 1996/97 health survey, but the WHO Long Form questions had not been previously validated. The focus of this thesis is on the validation of the WHO Long Form questions and scales for the NZ population and important population subgroups.

The primary objectives of this project were:

- i. to group the WHO Long Form items in the 2002/03 NZHS into scales for each health domain and develop standard scoring protocols for each scale
- ii. to test the reliability of the scales using standard psychometric tests for the total NZ population and for major population subgroups
- iii. to test the validity of the scales using the standard psychometric tests for the total NZ population and for major population subgroups
- iv. to construct norms for the WHO Long Form scales for the NZ population
- v. to provide recommendations for the health status component of future NZ health surveys

Thesis outline

The thesis, which describes the rationale, data collection, scoring development and psychometric testing of the WHO Long form in the New Zealand Health Survey 2002/03, of is organised as follows.

Chapter 2 discusses the background to the measurement of health-related quality of life (HRQL), both internationally and in the New Zealand context. The chapter contains an introduction to the measurement of health status and its decomposition into different health and health-related domains. I review the development and content of the major instruments for measuring HRQL, i.e. the SF-20, SF-36 and WHO Long Form. This is followed by a discussion of their adaptation and application in the New Zealand context.

Chapter 3 is divided into two main sections. First the methodology of the 2002/03 New Zealand Health Survey (2002/03 NZHS) is described followed by a discussion of the use of

SF-36 and the WHO Long Form in this survey. The second half of the chapter covers the specific methods used to apply, and to test, the WHO Long Form in the 2002/03 NZHS. The methods for developing scoring guidelines for the scales is outlined. This is followed by a description of the methods for testing the reliability and validity of the scales and for producing population norms.

Chapter 4 has four main sections. First, the process for creating the WHO Long Form scales in the 2002/03 NZHS is described. Second, the results of the reliability tests are discussed for items and scales. Third, the validity analysis for WHO Long Form scales is presented. Finally, population norms for the scales are presented. The analyses in this chapter were performed for the whole population and major subgroups of interest within the population, separating the population by ethnic group, age group and deprivation.

The thesis concludes with a summary in Chapter 5 of the major findings and a discussion of the implications of this work for future health surveys in New Zealand.