



## **Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin**

### **Copyright statement**

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

### **Liability statement**

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

### **Access Agreement**

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.



*Outcome measurement in physiotherapy practice & the  
preliminary development of a new measurement instrument*

Emma K. Stokes  
Dip Stat, BSc(Physio), MSc

A thesis submitted in fulfilment of the requirements for the degree of  
Doctor in Philosophy

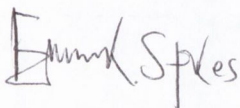
University of Dublin  
Trinity College

January 2005

*Declaration*

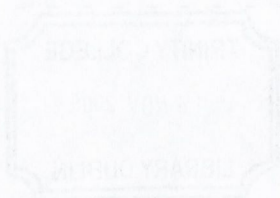
I, Emma K. Stokes, declare that this thesis is entirely my own work has not been submitted as an exercise for a degree at this university or any other university.

I hereby give permission for this thesis to be lent or copied on request, with the consent of the Library, and with due acknowledgement of the author.



Emma K. Stokes

10th November 2005  
Date

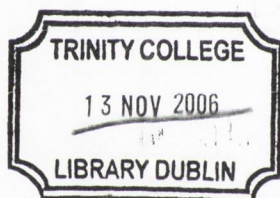


Author's name

I, Emma K. Stokes, declare that this thesis is entirely my own work, has not been submitted as an exercise for a degree at this university or any other university.

I hereby give permission for this thesis to be lent or copied on request, with the consent of the Library, and with due acknowledgement of the author, with the

*[Faint handwritten signatures and text, including the name Emma K. Stokes]*



THESIS  
8037

*When you can measure what you are speaking about and express it in numbers, you know something about it- but when you cannot measure it in numbers your knowledge is of a meagre and unsatisfactory kind- it may be the beginning of knowledge but you have scarcely in your thought, advanced to the stage of science whatever the matter may be*

*Lord Kelvin*

## *SUMMARY*

The 1990's became the era of accountability in health care services throughout the world and across all disciplines. Internationally, the profession of physiotherapy faced challenges in the acceptance and systematic use of standardised ways of measuring intervention. In the early part of this decade, their routine and systematic use is still not widespread across many domains of physiotherapy practice. The reasons for some of these difficulties can be attributed to the development of the profession and the lack of informative models and appropriate instruments.

This project considers the changes in practice in Ireland among physiotherapists working with older people. It measured the change in practice in Ireland between 1998 and 2003; a 5-year period during which the activities of the outcome measure movement within the physiotherapy profession have increased throughout the world.

The results of the two national surveys of physiotherapists working with older people who require rehabilitation, charts the changes in practice-

- The increase use in standardised outcome measures – from 30-50% of participants to 100% in some areas of practice.
- The acknowledgement of the barriers to their complete acceptance into routine practice- lack of time, resources/administrative support.
- A recognition that there is a lack of professional consensus on what instruments to use and for some client groups and the fact that none meeting their clients' needs.
- The degree of confidence of physiotherapists in using the results of their outcome measurements to inform clinical decision-making.

The initial survey of user needs informed the development and evaluation of a valid and reliable outcome measurement. The measurement properties of this new standardised outcome measurement (SOM) were evaluated in a sample of older people. Content validity was investigated in an iterative manner with the participation of specialist physiotherapists. The first version of the new SOM was

trialed with a sample of specialist physiotherapists (n=8), and their feedback was incorporated into a second iteration of the OM. A third version was then prepared following further trialing (n=13).

To evaluate construct validity, a number of constructs were considered

- That the new SOM would have a strong association with two scales of mobility and balance, most commonly used in Ireland at the time of the first survey 1998. The association between the new SOM and the Elderly Mobility Scale (EMS) ( $r=0.80$ ) and with the Performance Oriented Assessment of Mobility – balance subscale (POAM-B) (0.82-0.86) and gait sub-scales (POAM-G) (0.61-0.72) was strong.
- That the new SOM would be able to discriminate between two groups of older people, namely those having rehabilitation and healthy, community dwelling older people, which it does.
- That the scale would contain one factor- that would be explained by the individual items. Factor analysis revealed that one factor existed with all items contained and accounting for 70% of the variance within the data.

Both concurrent and predictive criterion validity have been considered through comparison with the admission and discharge scores of the EMS and POAM-B. The new SOM is more responsive than existing methods in use.



## *ACKNOWLEDGEMENTS*

Sincere thanks are due to the supervisors of this work-Mr. Peter Yung and Prof. Des O'Neill.

I wish to acknowledge the help, advice, assistance and participation of many physiotherapists who work with older people who are in need of rehabilitation.

This work would not have been possible if the older people, for whom part of this work is designed, had not agreed to give freely of their time and enthusiasm.

I would like to thank Dr. Bernie Brady for her help with proof reading the text.

I wish to also thank my friends and my beloved. This work has taken a long time and without their support and encouragement, especially in the final stages, it would not have been completed.

And my parents and sister, who never doubted this thesis would be completed and who have always supported all of my endeavours.

## TABLE OF CONTENTS

Summary		i
Acknowledgements		iii
Table of contents		iv
List of tables		x
List of figures		xii
Preface- contribution overview & dissertation summary		xiv
Chapter 1	Introduction	1
1.0	Background	1
1.1	Standardised outcome measures	2
1.2	Standardised outcome measures in physiotherapy	3
1.3	Measurement properties	4
1.4	Comparisons in practice	6
1.5	Conclusion	6
Chapter 2	Measurement properties	
2.0	Introduction	7
2.1	The description and methodology of validity	7
2.1.1	Face & content validity	7
2.1.2	Sensibility	10
2.1.3	Criterion validity	12
2.1.3.1	Sensitivity and specificity, positive and negative predictive values, likelihood ratios	13
2.1.4	Construct validity	20
2.1.5	Measuring change- the responsiveness or sensitivity of a measurement instrument	21

2.2	Statistical expressions of validity	26
2.2.1	Convergent and criterion validity	27
2.2.2	Factor analysis- content and construct validity	30
2.2.3	Sensitivity and specificity, positive & negative predictive values	32
2.2.4	Group differences, discriminant validity	37
2.2.5	Responsiveness	38
2.2.5.1	Receiver operator characteristic (ROC) curves	38
2.2.5.2	Standardised response mean and the effect size	40
2.3	Reliability	43
2.4	Analysis of reliability studies	45
2.5	Conclusion	50
<b>Chapter 3</b>	<b>Outcomes movement &amp; physiotherapy</b>	<b>51</b>
3.0	Introduction	
3.1	The outcomes movement	51
3.2	Evidence- based practice	52
3.3	The use of outcome measures in practice	56
3.3.1	The use of standardised outcome measures in practice	56
3.3.2	Attitudes and barriers to the use of SOM	61
3.3.3	The role of professional organisations in promoting the systematic use of SOM's	63
3.4.1	Conclusion	65
<b>Chapter 4</b>	<b>Methodology</b>	<b>69</b>
4.0	Introduction	69
4.1	Survey 1998	69
4.1.1	Aim & objectives	69
4.1.2	Methodology	70
4.1.2 (i)	Operational definition of standardised outcome measures	70
4.1.2 (ii)	Participants	70

4.1.2 (iii)	Procedure	70
4.1.3	Analysis of data	71
4.2	Survey 2003	71
4.2.1	Aim	71
4.2.2	Objectives	71
4.2.3	Participants	72
4.2.4	Procedure	72
4.3	Development and evaluation of a new instrument	73
4.3.1	Establishing an operational definition for the new instrument	75
4.3.1 (i)	A practice review to identify the barriers older people experience to achieving functional independence in the context of physiotherapy	76
4.3.1 (ia)	Aim & objectives	76
4.3.1 (ib)	Methodology	76
4.3.1 (ic)	Results	77
4.3.2	Item selection	77
4.3.2.1	Face validity	77
4.3.2.2	Scaling	77
4.3.3	Reliability	78
4.3.3.1	Inter-rater reliability	79
4.3.3.1(i)	Subjects	79
4.3.3.1(ii)	Raters	79
4.3.3.1(iii)	Procedures	79
4.3.3.1(iv)	Training	79
4.3.3.1(v)	Testing	80
4.3.3.2	Intra-rater reliability	80
4.3.3.2(i)	Subjects	80
4.3.3.2(ii)	Raters	80
4.3.3.2(iii)	Procedures	80
4.3.3.2(iv)	Training	80
4.3.3.2(v)	Testing	80
4.3.3.3	Analysis of inter- and intra-rater reliability	81

4.3.4	Validity	81
4.3.4 (i)	Content validity	81
4.3.4 (ii)	Construct and criterion validity	82
4.3.4 (iii)	Responsiveness	83
4.4	Ethical approval	84
<b>Chapter 5</b>	<b>Results</b>	
5.0	Introduction	85
5.1	Survey 1998	85
5.1.1	Profile of participants	85
5.1.2	Use of standardised outcome measures	85
5.1.3	Level of satisfaction	87
5.2	Survey 2003	90
5.2.1	Participants	91
5.2.2	Use of SOM	90
5.2.3	Reported barriers to the use of SOM	92
5.2.4	Confidence-self-efficacy in the use of SOM	93
5.3	Establishing an operational definition for the new instrument	96
5.3.1	The function of an outcome measure	96
5.3.2	A practice review to identify the barriers older people experience to achieving functional independence in the context of physiotherapy	98
5.4	Validity	103
5.4.1	Face validity	103
5.4.2	Content validity	110
5.4.3	Criterion validity	124
5.4.4	Construct validity	127
5.4.4.1	Factor analysis	127
5.4.4.2	Community dwelling older people	128
5.4.5	Responsiveness	131
5.5	Reliability	133
5.5.1	Inter-rater reliability	133

5.5.1.1	Subjects & rater	133
5.5.1.2	Percentage agreement & the kappa co-efficient	133
5.5.1.3	Bland & Altman	135
5.5.2	Intra-rater reliability	139
5.5.2.1	Subjects and rater	139
5.5.2.2	% agreement and weighted kappa coefficients	139
5.5.2.3	Bland & Altman method	140
5.5.2.4	Internal consistency	142
5.6	Summary of results	143
5.6.1	Surveys	143
5.6.2	Requirements of an outcome measure	144
5.6.3	Measurement properties-comparison between NS-total and NS-GP	147
<b>Chapter 6</b>	<b>Discussion</b>	<b>149</b>
6.0	Introduction	149
6.1	Surveys	149
6.1.1	Use of standardises outcome measures	149
6.1.2	Frustrations & barriers	153
6.2	Conclusion	155
6.3	A review of the item selection & evaluation	156
6.4	Conclusion	160
6.5	Measurement properties	160
6.5.1	Criterion validity-predictive and concurrent validity	160
6.5.2	Discriminant validity	161
6.5.3	Responsiveness	162
6.5.4	Construct validity	163
6.5.5	Reliability	165
6.6	Two versions of the new SOM	166
6.7	Conclusion	167
6.8	Limitations	168
6.9	Future work	169

References

170

Appendices

184

## *LIST OF TABLES*

Table 2.1	The concept of sensibility	11
Table 2.2	Examples of concurrent and predictive validity	13
Table 2.3	Likelihood ratios in physiotherapy research	15
Table 2.4	Responsiveness to change	23
Table 2.5	Reported coefficients of validity for the Berg Balance Scale	28
Table 2.6	2 X 2 table	33
Table 2.7	Multi-train-multimethod matrix	37
Table 2.8	SRM for five orthopaedic measurement instruments	41
Table 2.9	Effect sizes reported for PASS	42
Table 2.10	Methods to calculate the SEM	46
Table 2.11	Reliability datasets	48
Table 3.1	The potential relationships between outcomes evaluation and research, clinical effectiveness, clinical guidelines and evidence-based practice; and between the individual practitioner, researchers, developers and the provision of quality health care	55
Table 3.2	SOM's used by Canadian PT's	59
Table 3.3	Confidence in the use of SOM's	60
Table 5.1	Frustrations with current methods of measurement	89
Table 5.2	The specific SOM's used by PT's in 2003	91
Table 5.3	Mean % confidence reported for each statement, along with the standard deviation (SD) and the range	95
Table 5.4	Rank order scores for each measurement function, in the each domain- mean (standard deviation)	97
Table 5.5	Items for inclusion in an assessment of Parkinson's Disease	104
Table 5.6	Items for inclusion in an assessment of balance	106
Table 5.7	Items for inclusion in an assessment of stroke	109



Table 5.8.1	Bed rise (BR)	111
Table 5.8.2	Sit-to-stand (STS)	112
Table 5.8.3	Standing (Stand)	113
Table 5.8.4	Reach & lift (R&L)	114
Table 5.8.5	Bend & reach (B&R)	115
Table 5.8.6	Repeated sit-to-stand (RSTS)	116
Table 5.8.7	Gait	117
Table 5.8.8	Gait speed	118
Table 5.8.9	360 turn	119
Table 5.9	Results from second phase of content validity	120
Table 5.10	Summary statistics, baseline scores- total sample	125
Table 5.11	Correlation matrix- baseline scores, all instruments	125
Table 5.12	Summary statistics, admission and discharge score	126
Table 5.13	Correlation matrix- admission and discharge score	127
Table 5.14	Relative Efficiency (RE) of the new SOM	132
Table 5.15	%agreement & kappa co-efficients for agreement between raters	134
Table 5.12	% agreement & weighted K coefficient between T1 and T2	139
Table 5.13	Inter-item correlation matrix	142
Table 5.14	Survey findings- 1998 & 2003	143
Table 5.15	Outstanding questions from content validity methodology	145
Table 5.16	Measurement properties	147

## *LIST OF FIGURES*

Figure 2.1	A Scree test	32
Figure 2.2	The influence of cut-off scores	34
Figure 2.3	Nomogram for application to likelihood ratios	36
Figure 2.4	Scatterplot of datasets	48
Figure 2.5	Bland & Altman method	49
Figure 4.1	The development of a new instrument and the placement of the surveys to inform this process	73
Figure 5.1	Use of SOM versus non SOM	86
Figure 5.2	Frustrations reported by users of non-SOM	87
Figure 5.3	Frustrations reported by users of SOM	88
Figure 5.4	Use of SOM 1998 VS 2003	91
Figure 5.5	Frequency and pattern use of SOM	92
Figure 5.6	Barriers to the use of outcome measures	93
Figure 5.7	Confidence in the use of SOM	94
Figure 5.8	Measurement functions across domains- rank order	98
Figure 5.9	Mobility as a barrier to achieving functional independence	99
Figure 5.10	Balance problems as a barrier to achieving functional Independence	99
Figure 5.11	Decreased fitness as a barrier to achieving functional Independence	100
Figure 5.12	Impaired cognition as a barrier to achieving functional Independence	100
Figure 5.13	Communication difficulties as a barrier to achieving functional independence	101
Figure 5.14	Mood as a barrier to achieving functional independence	101
Figure 5.15	Pain as a barrier to achieving functional independence	102
Figure 5.16	Comparison of impact barriers	103
Figure 5.17	Suggestions for inclusion in an assessment of mobility	105
Figure 5.18	Items for inclusion in an assessment of gait disorders	107
Figure 5.19	Items for inclusion in an assessment of exercise tolerance	107

Figure 5.20	Scree plot for factor analysis	128
Figure 5.21.1	Community dwelling older people- NS-GP	129
Figure 5.21.2	Community dwelling older people- NS-GP	130
Figure 5.22	Boxplots of admission and discharge scores	131
Figure 5.23	Bland & Altman plot-NS total, R1 &R2	136
Figure 5.24	Bland & Altman plot-NS total, R1 &R3	136
Figure 5.25	Bland & Altman plot-NS total, R2 &R3	137
Figure 5.26	Bland & Altman plot NS- GP, R1 &R2	137
Figure 5.27	Bland & Altman plot NS- GP, R1 &R3	138
Figure 5.28	Bland & Altman plot NS- GP, R2 &R3	138
Figure 5.29	Bland & Altman plot for NS- GP	140
Figure 5.30	Bland & Altman plot NS- total	141

*PREFACE    CONTRIBUTION SUMMARY & DISSERTATION  
OVERVIEW*

The aim of this thesis was to investigate an aspect of physiotherapy practice- the use of standardised outcome measurement by physiotherapists working in the area of rehabilitation of older people. The review of this practice occurred at two points in time, with a view to evaluating any changes that might have occurred during this period of time.

In addition the thesis investigated the measurement properties of a newly developed standardised outcome measurement. The thesis aimed to describe the design and analysis phase of the development of the new SOM and compared its measurement properties to the most commonly used standardised outcome measures reported to be in use by physiotherapists working in Ireland.

This thesis contributes to the field both nationally and internationally in the following ways:-

1. A review of measurement properties and how, during the development of a new measurement instrument, they can be investigated. This review also considers the relevance of the output of any statistical analysis reported when the measurement properties are being evaluated. It utilises, where available, relevant examples from the physiotherapy literature.
2. A review of the literature published internationally that considers the practice of physiotherapy in the context of the use of standardised outcome measures. Such an international review of the published literature has not been completed to date. The activities of professional organisations in supporting the use of standardised outcome measures in practice are also considered.
3. The first and only review of practice in Ireland, which focuses on the area of rehabilitation of older people (published 1999). A repeat review of practice that allows a change in practice in Ireland to be clearly documented and compared to the only other review published to date.
4. The design of a new standardised outcome measurement for use by physiotherapists both in Ireland and other countries. The design phase was informed by the end-users. The evaluation and analysis of the

measurement properties of the new standardised outcome measurement are presented.

The dissertation takes the following format:-

Chapter 1- Introduction. This chapter present the local and international background to this work and places the activities undertaken in the context of developments in both health care and physiotherapy practice.

Chapter 2- Measurement properties. A comprehensive review of the measurement properties of validity and reliability. The methods used to investigate them and the analysis used to evaluate the output of these methodologies.

Chapter 3- Outcomes movement & physiotherapy. A review of the published literature on physiotherapy practice internationally.

Chapter 4- Methodology. This chapter describes the various methodologies employed in this work.

Chapter 5- Results. This chapter describes the results of each of the sections of work.

Chapter 6- Discussion. This chapter discusses the results in light of other published findings and clinical experience. It considers the limitations of the work and areas for future development.

**CHAPTER 1 INTRODUCTION**

The purpose of this chapter is to outline the research presented within this thesis and to place it in the context of the practice of physiotherapy during the course of the research work from 1998-2003.

**1.0 Background**

It is hard to imagine that there was a time when routine standardised outcome measurement within physiotherapy practice was not employed and yet Partridge reports in a paper in the 1980's that this was the case. She states that 'information about specific changes that occur in patients during a course of treatment is not usually recorded, outcome being described in terms of the patient 'getting better', 'improving' or being discharged' (Partridge, 1982). During the 1980's, it was increasingly recognised, in theory, that measurement, within the context of a skilled assessment, was necessary in rehabilitation therapy, nonetheless barriers did exist to the routine incorporation into practice. Measurement is the application of standard scales/instruments to variables, giving a numerical score, which may be combined for each variable to give an overall score, while assessment is the process of understanding a measurement in a specific context (McDowell & Newell, 1996). Tallis (1988), as part of the Marjory Warren Lecture, discussed some of the issues surrounding the reasons why measurement of the effects of therapy was not taking place. He suggested that the reasons could be due to

- Misplaced confidence- the notion that 'I know it works I've seen it work'.
- Misinterpretation- a fear that measured findings could be misinterpreted with a view to justifying rationalisation of services.
- Time consuming- using standardised outcome measurement was time-consuming and the urgency to deliver conventional therapy was more important than considering its effectiveness.
- Ideological hostility- a reluctance to measure might have been expressed or rationalised by suggesting that measurement instruments or methods available would not capture the real needs of the patient. For example, impairment measures bear no relation to the quality of life of the patient.
- Signal to noise- the limited ability of existing measurement instruments to capture change that is attributable to therapy.

- Problems with definition-measurement tended to be 'ill-defined'. For example, if a measurement scale had a score for 'performs task with help'- no specific definition of 'help' was provided. This lack of detail left the application of scores open to individual interpretation

### 1.1 Standardised outcome measures

At a more strategic level, significant challenges faced the physiotherapy profession with the emergence of the outcomes movement, namely the absence of an agreed framework for measurement and the absence of an ethos of using standardised measurement instruments. The 1990's have been described as the 'era of accountability' and in physiotherapy as in many other health care disciplines it was the time when the focus on the application and development of standardised outcome measurement began to be considered. The International Classification of Impairment, Disability and Handicap (ICIDH) (WHO, 1980) was proposed as a framework for informing measurement practice (Wade 1992, Enderby & Kew 1995, Jette 1995, Duckworth 1999). The Royal College of Physicians and the British Geriatric Society published a report suggesting a 'package' of assessments for use by multidisciplinary teams- the Barthel Index, the Abbreviated Mental Test Score, The Geriatric Depression Scale and the Philadelphia Morale Scale (RCP & BGS, 1992). Simpson & Forster (1993) reported that physiotherapists expressed some concern about the *mobility* section of the Barthel Index being used as the sole measure of mobility. In response to this absence of a more acceptable indicator of mobility, the Elderly Mobility Scale was developed (Smith, 1994). By 1997, only three papers had been published, examining the actual use of standardised OM in physiotherapy (Mayo *et al.* 1993, Chesson *et al.* 1996, Turner *et al.* 1996). Nevertheless, the 1990's were also a time when a number of professional organisations began a series of activities to promote and enhance the use of SOM's in physiotherapy practice (Hammond, 1999, Hammond 2000, Kay *et al.* 2001).

### 1.2 Standardised outcome measures in physiotherapy in Ireland

At the start of the current project (1998), in Ireland, the use of SOM was emerging in clinical practice, although no actual review had been performed. The Irish Society of Chartered Physiotherapists (ISCP)- the professional organisation of physiotherapy- had no formal policy or strategy to support such activities. The model

of ICIDH had become accepted, as a paradigm for choosing SOM's and while many measures existed for consideration of impairments and handicaps, few adequately designed and evaluated disability measures were available for physiotherapists working with older people. Within the context of both national and international developments in adopting the systematic use of standardised outcome measures, a review of the practice of physiotherapists (PT's) in Ireland working with older people was considered timely. The area of physiotherapy rehabilitation of older people was of specific interest to this author due to her personal clinical experience. Furthermore, personal interest notwithstanding, people over the age of 65 years represent a growing sector of the population and represent a group that can have significant health care needs. In Ireland, people aged  $\geq 65$  years have the highest age-specific admission rate to hospital. Functional disability increases with age and while the majority of people aged 65-69 years in Ireland report no functional disability, these figure drop to 63% (women) and 71% (men) in people aged 80 years and older (National Council of Ageing & Older People 1998). The purpose of this review was to identify trends in practice and to begin the process of gathering information from physiotherapy experts about what should be included in a new standardised outcome measurement (SOM), should one be developed. The results of this survey suggested that the use of standardised outcome measures was not widespread, that PT's had frustrations with both standardised and non-standardised methods of measurement and that the instruments in common use lacked full evaluation of their measurement properties. The frustration expressed with both the standardised and non-standardised methods used, personal clinical experience of using both the Elderly Mobility Scale and the Performance Oriented Assessment of Mobility (found to be the two most commonly used SOM's) and the limitations that existed in some aspects of their measurement properties led to the final decision to construct a new SOM. One that would meet the needs of Irish PT's working with older people, the components and construction of which would be closely informed by current practice and the needs of the physiotherapy users.

A wide range of measurement scales or instruments exist, both generic and disease or condition specific. Disease or condition specific measures, such as the Motor Assessment Scale for stroke (Carr *et al.* 1985), are designed for a given diagnostic group and are not applicable beyond that group. Examples of generic physiotherapy outcome measurements are the Elderly Mobility Scale (Smith 1993),



the Rivermead Mobility Index (Collen *et al.* 1991) and the Performance Oriented Assessment of Mobility (Tinetti, 1986). In general they can be used across a group of patients regardless of their specific diagnostic group. Rehabilitation of older people, embraces a broad spectrum of potential disease and disability. In choosing what to measure, it is important to capture the characteristics that are related to the intervention provided. Physiotherapists concern themselves with movement- both at a micro and macro level and mobility is of particular relevance to many people with disabilities (Chio & Burnett 1985, DeWeert & Harrison 1985). The *Description of Physical Therapy*, developed during the years 1996-1999 and agreed by all member organisations of the World Confederation for Physical Therapy (1999) formalises the fact that physiotherapy in its broadest sense is concerned with movement -'Physical Therapy is providing services to people and populations to develop, maintain and restore maximum movement and functional ability throughout the lifespan'. When designing the new SOM in this study, it was considered important that any new measurement instrument would operate within accepted frameworks and would be informed by physiotherapy practice. Hence, with a view to considering the most common difficulties encountered by older patients attempting to achieve functional independence, in the context of physiotherapy, a more focussed survey was completed in two rehabilitation units in Dublin in 1999. The members of the physiotherapy staff in these units completed a questionnaire on their current patient list, 130 questionnaires were completed. The results identified that the domains of 'immobility' and 'instability' were the two most common barriers to achieving functional independence in their older patients. Thereafter the development of a measurement instrument that would address aspects of mobility and stability was designed. The results of the 1998 survey, clinical experience and a review of normative data informed the construction of the first iteration of the scale.

### 1.3 Measurement properties

Measurement properties describe the attributes of standardised outcome measurements (SOM). Validity and reliability should be reported prior to the acceptance of a SOM into practice. The method employed and the analysis used informs the way a SOM influences clinical-decision making. This information is available in individual publications outlining the development of instruments. After the initial publication relating to a new instrument, many more may be published

that add to the body of knowledge about its measurement properties. Collecting and synthesizing this information is a time-consuming task and is often presented in databases (Hammond 1999), review articles (Forlander & Bohannon 1999, VanSwearingen & Brach 2001) or text books (Bowling 2005, McDowell & Newell 1996). All of these media presuppose an understanding of the relevance of the reported measurement properties. The volume of such information is growing. It is interesting to note the significant increase in the amount of information provided on measurement properties and their interpretation from 1995 and 2002, in the only physiotherapy specific reference text (Cole *et al.* 1995, Finch *et al.* 2002). Apart from recent graduates, most physiotherapists would not have been exposed to a course on outcome measurement as part of their undergraduate education. The professional body in Ireland, unlike those in the United Kingdom, US and Canada, has not provided outcome measurement courses in a strategic way with a view to changing practice. Chapter two of this thesis considers in detail the methodology, analysis and interpretation of measurement properties, using specific physiotherapy examples where they exist. The measurement properties of the newly developed outcome measure were evaluated as part of this work. This involved in excess of 100 people aged 60 years and older. Content validity was investigated in an iterative manner. The first version of the new OM was trialed with a sample of specialist PT's, and their feedback was incorporated into the second iteration of the OM. A third and final version was prepared following further trialing and comment from physiotherapy specialists. Both concurrent and predictive criterion validity have been considered through comparison with other rehabilitation scales, at admission and discharge. The following constructs have been evaluated

- That there is a strong relationship between the new OM and existing commonly used OMs.
- That the new OM is more responsive than existing OMs. Responsiveness to change is considered by evaluating the effect size and relative efficiency.
- That there is a significant difference in scores in the new OM between community dwelling older people and those in rehabilitation programmes.

#### **1.4 Comparisons in practice**

In 2003, a comparative survey of the practice of senior physiotherapists was completed. The output of this survey enabled a review of changes in practice in Ireland and a consideration of the reported barriers to the use of SOM. If a new outcome measurement is to be incorporated into practice, clearly identifying any barriers that might impact on its use is important. Equally, in presenting the measurement properties of the new SOM to practising PT's should be informed by their confidence in the use and interpretation of SOM's. A comparison with international practice was also made possible by completing this survey. If the systematic use of SOM is to become routine in physiotherapy practice in Ireland and beyond, a full understanding of the extent to which practice is changing, the ways it is adapting and concerns of PT's is required. This work contributes to a small body of international work in this field.

#### **1.5 Conclusion**

In conclusion, the systematic use of standardised outcome measurement to enhance clinical-decision making is not a simple endeavour. If it were, their use would be comprehensive, informative and optimal. This has yet to be achieved in physiotherapy practice in Ireland and in other countries where practice has been investigated. The research work that forms part of this thesis contributes to our knowledge of the changes physiotherapy practice from the late 1990's over a five-year period. It also makes available a new outcome measure designed for and by Irish physiotherapists for use with older people, with preliminary supportive information on its measurement properties.

## CHAPTER 2      MEASUREMENT PROPERTIES

### 2.0 Introduction

Measurement or psychometric properties of an instrument, tool or scale are the properties that describe the 'qualifications' of the instrument. They describe the genesis of the scale or tool, the methods used to assess its functionality and the analysis, both quantitative and qualitative, employed to the methods, to yield some expression of support, or otherwise for its inclusion in practice. This chapter describes the measurement properties of validity (sections 2.1 and 2.2) and reliability (sections 2.3 and 2.4). It considers the description of the concepts, the methods employed to investigate them, the statistical analysis that may be utilised and how the results of such analysis can be used to inform practice. Where available, examples of physiotherapy literature and measurement scales/instruments have been used to illustrate points.

### 2.1 The description and methodology of validity

#### 2.1.1 Face & content validity

Face validity is, as the name suggests, a question of whether on the face or surface of the instrument, it measures what it intends to measure. Face validity increases the acceptance of the instrument to the users. If items seem irrelevant to their practice, it is possible they will simply be omitted (Streiner & Norman, 1989). Streiner & Norman (1989) suggest that face validity and content validity are generally defined within the sphere of subjective professional consensus or require an explicit process which includes establishing the theoretical framework of the measure, research and expert opinion based on clinical observation. These inform which items should be considered for inclusion, before the instrument is submitted for further measurement testing. Feinstein (1987) describes a more detailed exposition of the process of face validity, with a view to dissecting what he considers to be a primarily intuitive process. He suggests four main attributes, namely, 'focus of interpersonal exchange', 'focus of basic evidence', 'biologic coherence of components' and 'attention to personal collaboration'.

- *Focus of interpersonal exchange.* In the case of an instrument that considers personal beliefs and/or attitudes as opposed to general public opinion, Feinstein (1987) suggests that the way in which the questions are posed

should be considered. The construction of questions that are too impersonal will not yield personal attributes or beliefs.

- *Focus of basic evidence.* If an instrument is designed to measure a particular construct, for example lower extremity function, it might include 'basic evidence' such as strength or range of motion (impairments of function/structures), or it might include information about functional activity involving the lower extremity (disabilities). In substituting impairments and disabilities as measures or indicators of one another, the author suggests that the basic evidence becomes displaced. The appropriateness of substitutions and displacements should be considered in the context of face validity.
- *Biologic coherence of components.* When an instrument contains multiple domains, for example the Elderly Mobility Scale (Smith 1994) that contains bed mobility, transfer skill and speed, the 'biologic coherence of the components' refers to the rationality of the combination of cognate domains.
- *Attention to personal collaboration.* If factors such as motivation and effort in performing activities within an instrument are not considered in instructions or scoring, the results of the scale may not be valid. Feinstein (1987) suggests that this 'attention to personal collaboration' within the instrument, is an important factor in the face validity of an instrument, especially if the instrument is designed to measure change as a result of intervention.

If an instrument contains within it a number of domains that represent a sample of the total population of domains (most likely theoretical) that could be included, content validity considers how adequate the sample of domains is, with regard to the expected functionality of the instrument. If an instrument aims to measure functional mobility in older people, the content validity of the instrument is a measure of how well the constituent parts reflects the aim of the instrument, which should be clearly stated in the design phase (McDowell & Newell, 1996). On the basis of the domains included in an instrument, the assessor may wish to infer beyond the output of the instrument. In a unidimensional instrument e.g. Functional Reach (Duncan *et al.*, 1990), it is not possible to infer about the

functional balance performance of the subject but, with a broader measure of balance e.g. the Performance Oriented Assessment of Mobility (POAM) (Tinetti, 1986), it is possible to consider balance performance beyond simply the components of the instrument. Streiner & Norman (1989) suggest that the higher the content validity, the wider the inferences that can be validly applied about the person under other conditions and in alternative situations. Wade (1992) proposes that in the context of a multi-item instrument, all the component items should come from the same level of impairment, disability and handicap. According to Feinstein (1987), content validity requires digging deeper than face validity. It requires consideration not simply the components that are included and their suitability but also important omissions. He also considers the output of the domains that constitute the instrument, scaling and weighting. The above descriptions of face and content validity are supported throughout the literature on measurement properties (Green 1981, Bush 2000, Finch *et al.* 2002).

In the descriptions of the development of instruments, it is often not explicitly stated how face and content validity have been considered (Reuben & Siu 1990, Di Fabio & Seay 1997). Nonetheless in some cases brief descriptions are given, for example in the development of the Elderly Mobility Scale, which employed discussion with physiotherapists, and a literature review (Smith, 1994). Tinetti (1986) describes the process as a review of 'previous work by bioengineers, orthopedists, neurologists, rheumatologists, and physical therapists to identify what observations should be included'. Platt *et al.* (1998) used a panel of physiotherapists in the development of the Physiotherapy Functional Mobility Profile. Berg (1989), in the preliminary development of the eponymous balance scale, considered face and content validity in three phases. Using both professional samples (physiotherapists, nurses, physicians and occupational therapists) and patient samples (older people with an impairment of balance), groups of items were identified as being important for inclusion in the scale. The next phase asked the health care professionals to rank the importance of the items and asked the patients to rate their sense of unsteadiness when performing the activities. The third iteration of the scale was further tested on patients and their responses were compared to the scores they achieved on the scale. The health care professionals were asked to view videotapes of patients performing the items in the scale and asked to identify which components were essential for inclusion and which were

extraneous. The final iteration of the scale was tested for reliability using five physiotherapists. It resulted in the production of a 14-item instrument, which is scored on a 5-point ordinal scale.

### **2.1.2 Sensibility**

It is Feinstein (1987) who takes the concept of face and content validity and extends it to consider the 'sensibility' of a measure or instrument. He describes a checklist for considering the sensibility of any instrument, a list of suggestions for a reviewer of an instrument to consider prior to its employment in the individual's practice. Table 2.1 outlines the five major topics that the author advocates should be considered when evaluating an instrument for sensibility. Each has clearly described subdivisions. Face and content have been discussed above.

Table 2.1 The concept of sensibility

<p><b>Purpose &amp; framework</b></p> <p>Are the attributes listed across clearly stated?</p>	<ul style="list-style-type: none"> <li>▪ Clinical function-prediction of outcome, measure of change, generation of problem list.</li> <li>▪ Justification – what need is addressed? Is the instrument a new or alternative expression?</li> <li>▪ Applicability – Is it clear the clinical situation in which the instrument may be applied? And the subjects most appropriate for its application.</li> </ul>
<p><b>Overt format</b></p> <p>Judgements that could be made following a ‘thoughtful review’, especially by a non-specialist</p>	<ul style="list-style-type: none"> <li>▪ Comprehensibility – the input and output of the instrument.</li> <li>▪ Replicability – are the instructions clear?</li> <li>▪ Output – is the scope acceptable? Are the categories mutually exclusive? Is the scaling/rating logical?</li> </ul>
<p><b>Face validity</b></p> <p>Judgements made from an initial appraisal of the instrument.</p>	<ul style="list-style-type: none"> <li>▪ Discussed in detail above.</li> </ul>
<p><b>Content validity</b></p> <p>i.e. appraisal of the underlying components</p>	<ul style="list-style-type: none"> <li>▪ Quality of basic data</li> </ul>
<p><b>Ease of usage</b></p>	<ul style="list-style-type: none"> <li>▪ Personnel involved or required</li> <li>▪ Risks</li> <li>▪ Amount of time</li> </ul>



### 2.1.3 Criterion validity

Following the consideration of face and content validity, more formal evaluation of validity is conducted. Criterion validity considers the performance or accuracy (Feinstein, 1987) of the new instrument by comparing it to a *gold standard*. The level of agreement between the two sets of measurements is then examined, using an 'appropriate indicator of agreement' (McDowell & Newell, 1996); since this is often a correlational comparison, criterion validity may also be referred to as correlational validity. Regardless of the term utilised, there is a temporal component associated with this type of validity – it can be either concurrent or predictive. The former relates to how the new instrument compares to the *gold standard* at a given point in time and the latter refers to its relationship to an assessment in the future. Messick (1980) proposed using the descriptors *diagnostic suitability* or *substitutability* for concurrent validity and *predictive utility* for predictive validity as a means of clearly describing the rationale of criterion validity. These terms have not been widely adopted. In describing how criterion validity is assessed, Streiner & Norman (1989) pose the following question, if a good criterion measure exists, why create another instrument? That the new instrument is less expensive, less invasive or carries less risk, are the main reasons for creating a new means of measurement. The consideration of criterion validity becomes less absolute when a *gold standard* does not exist. This is often the case for rehabilitation measures. Nevertheless, criterion validity may still be assessed by substituting cognate instruments as a substitute for the *gold standard* or through informal analysis. Table 2.2 illustrates a variety of ways the criterion validity has been established for a sample of physiotherapy scales.

Table 2.2 Examples of concurrent and predictive validity

Author (s)	New scale	Validity	Cognate scale
Smith (1994)	Elderly Mobility Scale (EMS)	Concurrent	Barthel Index (BI) Functional Independence Measure (FIM)
Prosser & Canby (1997)			Barthel Index
Collen <i>et al.</i> (1991)	Rivermead Mobility Index (RMI)	Concurrent	BI, Functional Ambulation Categories (FAC), gait speed, six minute distance test (SMDT)
Benaim <i>et al.</i> (1999)	Postural Assessment Scale for Stroke (PASS)	Concurrent & predictive	FIM- transfers, locomotion, total concurrent measure and at 90 days after stroke (DAS)
Mao <i>et al.</i> (2002)			BI, Berg Balance Scale (BBS)

### 2.1.3.1 Sensitivity and specificity, positive and negative predictive values, likelihood ratios.

Sensitivity (Sn) and specificity (Sp) and their evaluation are embraced within the concept of criterion validity. In this regard the criterion is diagnostic i.e. it considers how accurate the instrument is in identifying correctly the presence of a disease/disorder in an individual (sensitivity) but also its precision in correctly identifying that the disease/disorder is not present (specificity). Similarly, positive and negative predictive values ( PPV & NPV) may also be presented for an instrument – the former being the proportion of subjects with a positive test result who have the condition of interest and the latter being the proportion of subjects with a negative test result who do not have the condition. If a positive predictive value is near 100% it is more likely that the disease or condition being predicted by

the test is actually present when the test is positive. Whereas sensitivity considers the instrument and its ability to identify subjects or patients who have a disorder, positive predictive values (PPV) focus on the number of people correctly identified by the test as having the condition. The relationship between PPV, Se and Sp is outlined below

$$PPV = \frac{Pr * Sn}{Pr * Sn + (1 - Pr) * (1 - Sp)} \quad \text{where Pr= prevalence}$$

Hence, PPV and NPV are influenced by the prevalence of the condition. This information is not often known, and in the past PPV were calculated using estimated prevalence of 50% or the prevalence based on the results of samples in studies (Dujardin *et al.*, 1994). The converse applies to the calculation of specificity and negative predictive values (Riddle *et al.* 1999, Sackett *et al.*, 1991).

Likelihood ratios (LR's) combine sensitivity and specificity and give an indication of how much a particular score (or range of scores) on a measurement instrument or test will raise or lower the pre-test probability of a particular condition being present e.g. falls risk, improvement in pain (Fritz *et al.*, 2000). LR's of greater than 10 or less than 0.1 may generate large changes in pre and post-test probability, LR's of 1-2 and 0.5 -1 add rarely important changes to pre-test probability.

A simple review of physiotherapy literature (using CINAHL and keywords 'likelihood ratios' and 'physical therapy' or 'physiotherapy', 1994-2004) yielded 22 citations, of which 7 citations were research papers related to physiotherapy intervention. The remainder either did not relate to PT interventions (n= 12) or were review papers (n=3). The search results are listed in Annex 1. Table 2.3 outlines the content and results of the papers (discussed below). Each study design reflects the specific aims of the research, nevertheless the essential feature of each of the studies was the need to examine how certain tests or examinations could predict a future endpoint e.g. reduction in pain, return to work. The use of the positive or negative likelihood ratio depends on whether the measuring test or instrument needs to rule 'out' a condition e.g. screening for a behaviour which if not present does not require further investigation, or rule 'in' a condition e.g. a behaviour is present, therefore the person is at risk (a diagnostic test).

Table 2.3 Likelihood ratios in physiotherapy research

Authors	Profile of participants	Predicted end-point	Measurement	Likelihood ratio (LR) (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Pre-test probability	Post-test probability
Werneke & Hart (2004)	Acute work-related low back pain, 1 year after discharge (n=171)	Return to work status	Pain pattern classification, time dependent i.e. after intervention completed. Presence or absence of centralisation of pain.	+LR 3.82 (2.29-6.35) -LR 0.38 (0.2-0.75)	68% (46-85)	82% (74-88)	15%	25%
Fritz & George (2002)	Acute work-related low back pain (n=78)	Persistent work restrictions	Work sub-scale of Fear Avoidance Beliefs (FAB) Questionnaire 0-42, higher=more	>34 +LR 3.33 (1.05-6.77) <30/42 -LR 0.08 (0.01-0.54)	>34/42 = 55% (34-75) <30/42= 95% (87-100)	>34/42= 84% (73-94) <30/42= 58% (45-70)	29%	>34/42= 58% <30/42= 3%

			FAB +LR >34/42 -LR <30/42					
Sutlive <i>et al.</i> (2004)	Patello-femoral pain syndrome (n=45)	Reduction in pain i.e. >50% improvement on VAS	Forefoot valgus (FFV) $\geq 2^\circ$ Passive great toe extension (PGTE) $\leq 78^\circ$  Navicular drop (ND) $\leq 3\text{mm}$	FFV $\geq 2^\circ$ +LR 4.0 (0.7-21.9) PGTE +LR 4.0 (0.7-21.9)  ND +LR 2.3 (1.3-4.3)	FFV= 13% (4-24)  PGTE= 13% (4-24)  ND= 47% (32-61)	FFV= 97% (90-100)  PGTE= 97% (90-100)  ND= 86% (67-93)	60%	Forefoot valgus (FFV) $\geq 2^\circ$ = 86% Passive great toe extension (PGTE) $\leq 78^\circ$ = 86%  Navicular drop (ND) $\leq 3\text{mm}$ = 78%
Fritz <i>et al.</i> (2000)	Return-to-work within 4 weeks, following occupational acute low back pain (n=69)	5 non-organic signs (NOSg) and 7 non-organics symptom descriptors (NOSym) & non-organic	$\geq 2$ NOSg $\geq 3$ NOSymp $\geq 3$ NOI	$\geq 2$ NOSg -LR 0.75 (0.51-1.10)  $\geq 3$ NOSymp -LR 0.62 (0.40-0.96)	$\geq 2$ NOSg= 41% (20-61)  $\geq 3$ NOSymp= 50% (29-71)  $\geq 3$ NOI= 64%	$\geq 2$ NOSg= 79% (67-90)  $\geq 3$ NOSymp= 81% (70-92)  $\geq 3$ NOI= 62%	33%	Negative NOSg= 27%  Negative NOSymp= 24%  Negative NOI= 23%

		index (NOI)		$\geq 3$ NOI	(44-84)	(48-79)		
				-LR 0.59				
				(0.32-1.07)				
Stratford <i>et al.</i> (1998a)	Investigation of test characteristics of Roland- Morris Questionnaire in subjects with low back pain of less than 6 weeks (n=226)	Global rating of change in status, score of change =5	Roland Morris Questionnaire (RMQ) change score of $\geq 5$	4 (2.49-6.37)	72% (65-79)	82% (74-90)	63%	87%
Stratford <i>et al.</i> (1998b)	Investigation of test characteristics of Roland- Morris Questionnaire in subjects	Chance of achieving treatment goals	Roland Morris Questionnaire (RMQ) important change score, for different initial (iRMQ) scores	iRMQ 0-8 +LR 5.33 (1.47-20.4) iRMQ 9-16 +LR 3.7 (1.35- 10.02) iRMQ 17-24	iRMQ 0-8 64% (53-75) iRMQ 9-16 85% (72-98) iRMQ 17-24 100% (78%)	iRMQ 0-8 88% (81-95) iRMQ 9-16 77% (65-89) iRMQ 17-24 78% (65-91%)	iRMQ 0-8 78% iRMQ 9-16 72% iRMQ 17-24 57%	iRMQ 0-8 95% iRMQ 9-16 86% iRMQ 17-24 86%

	with low back pain (n=143)		e.g. 0-8, 9-16, 17-24	+LR 4.55 (1.33-15.28)					
Abbott & Mercer (2003)	Patients with low back pain (new or also within previous 3 months) or those with persistent pain of at least 3 months duration (n=12)	Presence of lumbar segmental hypomobility, defined using radiography	Active range of motion of lumbar spine (AROM) Abnormality of segmental motion during AROM (AbnROM) Passive accessory inter-vertebral movements (PAIVM) Passive physiological inter-vertebral movements (PPIVM)	+LR 1.88 (0.57-6.8) -LR 0.42 (0.07-1.9) AbnROM +LR3.6 (0.84-15.38) -LR 0.65 (0.28- 1.06) PAIVM +LR 1.16 (0.44-2.03) -LR 0.71 (0.12-2.75) PPIVM +LR 3.86 (0.89-16.31) -LR 0.64	NOTE all results 90% CI	75% (36-94)	60% (27-86)	28%	+ AROM signs ≈41% + AbnROM signs =58%* +PAIVM signs ≈ 30% +PPIVM signs ≈60%  - AROM signs = 14% -AbnROM signs ≈20% -PAIVM signs ≈22% -PPIVM signs

(0.28-1.04)

≈20%



#### 2.1.4 Construct validity

A construct is an idea or concept, and in the context of measurement properties, construct validity is a process, 'part science and largely art form' (McDowell & Newell, 1996), which draws the developer and later the reviewer along an evolving set of hypotheses about the new instrument. A series of constructs are considered, for example in the case of a new instrument to measure functional activities in older people, these might be that,

- the new instrument correlates well with the Barthel Index (convergent validity)
- it measures one or more (in the case of multidimensional instruments) identified domains (factor analysis)
- older people living in the community who require support services such as home-help and meals-on-wheels would perform less well than those who do not need help from these services or their family/carers in this regard (discriminant, known group validity).

Analysis of the hypotheses around these constructs will either support or otherwise the construct validity of the new instrument. Finch *et al.* (2002) discuss each of the aspects of construct validity mentioned above (convergent validity etc.) in terms of cross-sectional validity and longitudinal validity. The authors suggest that this taxonomy enables validity associated with a single point in time to be differentiated from that, which is associated with change in scores over time. Kirshner & Guyatt (1985) suggest that if instruments are used to evaluate change over time, it is not simply enough to assume that an instrument with cross-sectional validity demonstrates longitudinal validity; this needs to be explicitly examined during the instrument construction. Construct validity, and all it incorporates, differs from the previous types of validity described because it is an ongoing and evolving process, not solved in one or two experimental designs. The process may address hypotheses that consider an instrument designed to replace a cognate instrument or measurement; or it may be that a series of behaviours, signs and symptoms are combined in an attempt to define a new construct and its measurement.

### 2.1.5 Measuring change- the responsiveness or sensitivity of a measurement instrument.

There is unresolved debate in the literature about where the ability of an instrument to measure change falls in the domains of measurement properties. Some authors suggest that the responsiveness or sensitivity to change of an instrument is a separate measurement property and others believe it is part of the art-science continuum of activities that is incorporated in the investigation and evaluation of construct validity (Stratford *et al.* 1996, Hays & Hadorn 1992, Guyatt *et al.* 1987). Guyatt *et al.* (1987) support the thesis that measurement responsiveness/sensitivity is a property separate to validity and reliability since it is possible to have a measure that is reliable but not responsive, responsive but not valid and unreliable but yet responsive. Hays & Hadorn (1992) respond to the thesis of Guyatt *et al.* (1987) and suggest that while all of these situations are indeed possible, the artificial distinction between responsiveness/sensitivity may arise from the desire to decide whether measurement instruments are either valid or not. The authors suggest that validity is in fact an ongoing iterative process and degrees of validity exist. Stratford (1996), in a paper describing the various methods that may be employed to establish measurement sensitivity/responsiveness and the analytical methods that may be employed, opts for including sensitivity/responsiveness as one component of validity. To a certain extent, where sensitivity/responsiveness is placed may be irrelevant as long as the property is examined. Cole *et al.* (1995) note that for many outcome measures used by physiotherapists, a full examination of measurement properties has not been completed and Liang (1995) note that sensitivity/responsiveness has had the least study.

There is ongoing debate on the subject of sensitivity versus responsiveness and which of the two terms to use. (Fritz 1999, Stratford 1999, Finch *et al.* 2002). Kirshner and Guyatt first referred to 'responsiveness' as the power of the test to detect clinically relevant change (Kirshner and Guyatt, 1985). More recently the two terms have been further defined by Liang (1995) as,

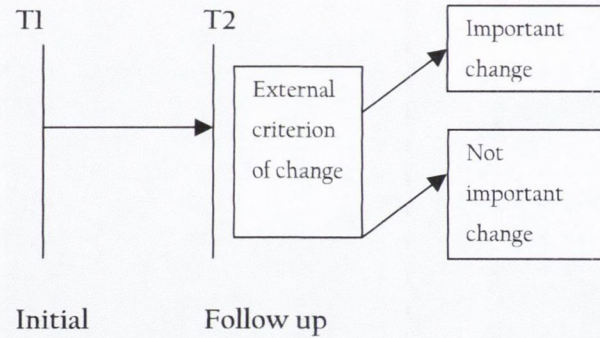
- Sensitivity- the ability to measure any change in a state
- Responsiveness- the ability to measure clinically meaningful or important change.

The author adds that responsiveness implies that the change is noticeable to the patient or physician and may allow for the person to perform a functional task more efficiently or with less pain or difficulty (Liang, 2000). It is the term responsiveness that is widely used in physiotherapy literature (Wright *et al.* 1998, Stratford 1996, Hammond 2000, Finch *et al.* 2002) and as such it will be used throughout this text. Study designs to consider responsiveness have been outlined by Stratford *et al.* (1996) and are represented in Table 2.4.



	will be demonstrated between T2 & T3		measuring this may not be sufficiently long enough.
3	Patients receiving the effective therapy will display more improvement than patients receiving placebo	<p>T1</p> <p>Initial Follow up</p>	T2 Difficulties arise if an established and effective treatment is not available and also with the ethics of not providing such a treatment to a group of patients.
4	Using the natural course of a disease or condition, where one group may be expected to have a more rapid rate of change than another e.g. acute and chronic low back pain	<p>T1</p> <p>Initial Follow-up</p>	T2 It is possible that two groups of patients with a differing temporal component will not actually exist within the group of patients for whom the measurement is designed.

- 5 This design is similar to 1 but it employs an external measure of change. At the follow up point in time the patient is judged as having achieved change or not. The construct applied then is whether the new measurement has identified this change also.



The main limitation that occurs in this study design is identifying another measure, which has a criterion standard for change.

## 2.2 Statistical expressions of validity

As outlined above, analysis of face and content validity are primarily discursive and descriptive, more qualitative than quantitative. When statistical analysis of criterion and construct validity is considered, the type of data - nominal, ordinal, interval-ratio- will inform the process.

### 2.2.1 Convergent and criterion validity

The output of criterion validity and convergent validity (an aspect of construct validity) will be validity coefficients. These are products of correlating the scores obtained on the new instrument with a *gold standard* or with existing measurements of similar domains. Coefficients of validity reported in the context of criterion and convergent validity are (Cohen & Holliday, 1996) –

- *Pearson product-moment correlation (PPMCC)* – consideration of the linear relationship between variables (interval or ratio data). It assumes random sampling, homoscedasticity (equal variance) and normal distributed data. If the data from two sets of measurements are plotted against one another, the PPMCC illustrates the closeness of the points to a straight line.
- *Spearman's rank order correlation*. When x and y are not linearly related, but show a consistently increasing or decreasing trend in rank, a nonparametric correlation such as Spearman's (rho) may be employed. It assumes that the data are ordinal and the sample is randomly generated.
- *Kendall's rank order correlation (tau)*. This coefficient may be applied in cases where Spearman's rho is appropriate. The power efficiency of this test relative to the PPMCC is reported to be 91%.
- *Phi coefficient ( $\phi$ )* – used in the analysis of data which is dichotomous i.e. presence of absence of a disease. It is derived from a 2x2 table (see 2.2.3) and related to chi-squared ( $\chi^2$ ) as follows,  $\phi = \sqrt{\chi^2/N}$  (Streiner & Norman, 1989).

In examining the validity coefficients associated with a new test and a criterion test, it is possible to explore the relevance of the coefficients with respect to how useful the new test will be for prediction of scores on the criterion or *gold standard*. For instance, let us say we want to predict how successful an individual would be in a health education programme, for example, a cardiac rehabilitation programme,

where success has been previously reported as an average score on a test instrument used on completion of the programme- *completion test*. In the absence of any other information, this *completion test* average is the best estimate that can be made about how successful a potential participant might be; and the variance associated with this average score is the estimate of how far away from the true performance score we might be, with no additional information (error without test). If a pre-course test is created that has a relationship with the *completion test* scores, then the variance of the *completion test* scores for those who received a given pre-course score is an estimate of the error associated with using the pre-course test (error with test).

There is a mathematical relationship between these error values ( $\sigma$ ) and the correlation coefficient [ $r$ ] between the pre-course test and the *completion test* scores, namely,

(correlation coefficient)<sup>2</sup> = error without test - error with test / error without the test  
 i.e.  $r^2 = \sigma_y^2 - \sigma_{yx}^2 / \sigma_y^2$

This relationship may be used in a number of ways to explore how relevant a reported validity coefficient might be in terms of informing practice (Helmstadter, 1966).

- The *standard error of the estimate* is the amount of error associated when a given score on the pre-course test is used to estimate a score on the *completion test* scores. It is used to establish confidence intervals around the score. It is expressed in the units of the criterion measure; in the case of the example above, the scores on the *completion test* scores.
- The *coefficient of alienation* is a ratio of the standard error of the estimate to the standard deviation of the criterion scores. The coefficient of alienation is low when the efficiency of predicting an individual score is high.
- The *relative reduction of error*- since  $\sigma_{yx}^2$  represents the error with the use of the pre-course test and  $\sigma_y^2$  represents the error without using the pre-course test, the difference between these two represent the reduction of error that results when the pre-course test is used. The relative reduction in error is calculated by dividing this number by the original amount of error -  $r^2$ . Hence a correlation co-efficient of 0.3 between the pre-course test and the completion test, would



imply that the relative reduction in error is .09, whereas a higher correlation coefficient of 0.7 would yield an almost 50% reduction in error.

Helmstadter (1966) also describes how when the output of two measurements is compared, the maximum possible correlation between them is the square root of the product of their respective reliabilities. McDowell & Newell (1996) set out an example of how this might be used to inform a judgement of convergent validity coefficients. If the coefficient of comparison of two instruments – one new and one extant, with reliability coefficients of 0.7 and 0.75 respectively – is 0.6, this should be considered in light of the maximum possible coefficient of validity, which would be 0.72. By extension, the authors suggest, a *gold standard* criterion validity correlation may be extrapolated – observed correlation/maximum possible correlation, in this example 0.83.

Squaring the coefficient indicates how much variance there is in common with the two tests. This suggests that a coefficient of 0.7 between the new instrument and an existing measure indicates that they share less than half the variance. Kline (1998) notes that in constructing a new instrument, components of the criterion or cognate instrument may be incorporated, and this may enhance the correlation. Table 2.5 illustrates a variety of coefficients of validity reported for the Berg Balance Scale.

Table 2.5 Reported coefficients of validity for the Berg Balance Scale

Authors	Subjects	Validity coefficient	Variable compared to BSS	Coefficient
Stevenson & Garland (1996)	24 subjects 8.4 +/- 6.16 years post stroke. 68.8 +/- 7.72 years*.	Pearson Product Moment Correlation Coefficient (PPMCC)	Mean Centre of Pressure (COP) speed in quiet stance Mean COP speed during arm flexion Peak arm acceleration	-.76 -.65 -.67
Wolf <i>et al.</i>	28 subjects with	PPMCC	Functional Reach	.62

(1999)	stroke, 56.4 +/- 13.8 years*, 162.5-190.5 months post stroke				
				Gait speed	.63
Usuda <i>et al.</i> (1998)	46 subjects with stroke, 69.3 +/- 9.6 years*	Spearman's Rank Order Coefficient (SRCC)		Barthel Index	.84
				Brunnstrom's recovery stage	.56
Shumway-Cook <i>et al.</i> (1997)	44 community dwelling older people - faller 77.6 +/- 5.4 years* & non-fallers, 74.6 +/- 54 years*.	SRCC		Use of assistive device	-.53
				History of imbalance	-.50
				Balance self-perceptions	.76
				Dynamic Gait Index	.67
Kokko <i>et al.</i> (1997)	40 patients with Parkinson's Disease, mean age 67 years, range 35-84 years.	PPMCC		Relative gait velocity (height/speed for 10m)	.56
				Relative stride length	.60
				Hand dexterity	.56
				Functional Status Questionnaire	.67

Piotrowski & Cole (1994)	60 older people volunteers – nursing home residents and community dwelling, 78 +/- 6.88 years*	PPMCC	Falls efficacy scale	.86
			Functional Assessment Inventory	.86
			Self-paced walk test	.73
Hsieh <i>et al.</i> (2000)	38 subjects with stroke	SRCC	Rivermead Mobility Index	.80
Nilsson <i>et al.</i> (1998)	28 subjects, 3-8 weeks post stroke, 55+/-9.3 years*	SRCC	Step frequency over 10m	-.69
			Borg Category Scale	-.73
			Functional Ambulation Category	.63

\* mean age and standard deviation (years)

### 2.2.2 Factor analysis- content and construct validity

Factor analysis may be applied in the consideration of both content and construct validity. In the former, the items within the instrument, or the sub-scales within a multidimensional inventory are examined to identify how they fit into one or more themes. In the latter, factor analysis may contribute to construct validity by indicating the associations between scales measuring similar constructs and lack of associations with scales measuring different concepts (McDowell & Newell, 1996, Streiner & Norman, 1989, Reis *et al.* 1991). The aim of factor analysis is to simplify a correlation matrix (Kline, 1998). In the case of instrument development, this allows the developer to evaluate the themes that emerge from the output of the instrument. For example, Washburn *et al.* (2002) used factor analysis in the development of a scale to measure physical activity for individuals with disabilities.

A 12-item scale was administered to 372 subjects with disabilities, and the results on each item were used to compute a 12-item correlation matrix. Factor analysis was utilised to identify if the responses to items within the scale fell into what could be considered 'reasonable and predictable patterns'. The analysis indicated that the scale assessed five dimensions of physical activity and that these items accounted for 63% of the total item variance. Similarly, Beattie *et al.* (2002) employed factor analysis in the creation of a *patient satisfaction survey*. Their analysis revealed two potential designs for their satisfaction survey instrument – one containing two factors (10 items) and the other containing three factors (12 items). On the basis of further evaluation of the scale (which considered the reliability of the instrument), the 10-item questionnaire was retained.

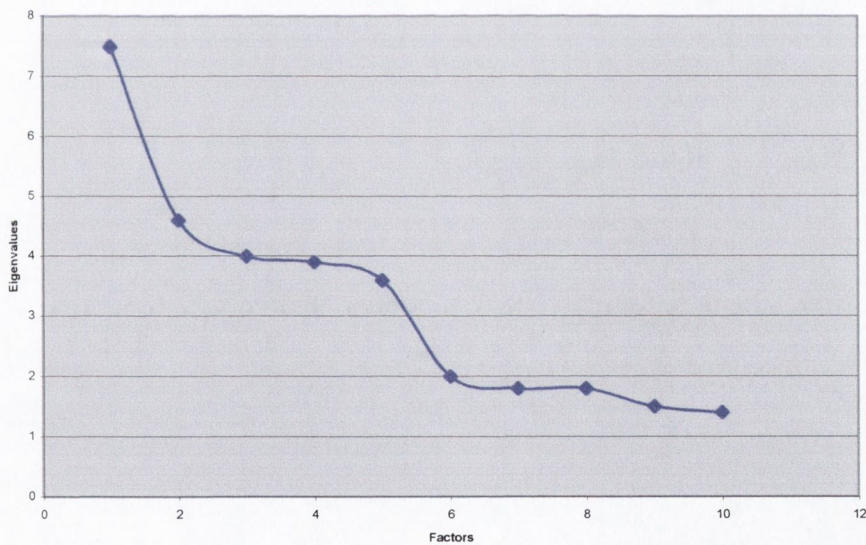
The process of factor analysis is as follows - an inter-item correlation matrix is created, and through principal factor or principle component analysis the number of variables is reduced to a smaller number of components. This is based on similarities between the items, and is believed to be indicative of the constructs that explain the underlying relationships between the original variables (Roush & Sonstroem, 1999). A new correlation matrix is generated, to explore the correlations between the original variables and the new components. This yields a number of outputs (Kline, 1998).

- *Factor loadings* – the correlation between the components and the original variables or items. Items with components loadings of  $> 0.3-0.4$  and with only one component may be retained (Kline 1998, Washburn *et al.* 2002). Items should 'load on' only one component to be included.
- *Eigenvalues* – are calculated by squaring and adding the loadings on each factor. Each item has an eigenvalue of 1, hence the factor must have an eigenvalue of  $> 1$  if it is to be logically retaining.
- *Percentage variance* – the total variance within the correlation matrix that the factor accounts for. It is calculated by dividing the eigenvalue by the number of variables.
- *Cumulative percentage variance* – variance accounted for by all the factors. The larger this is the more variance that is accounted for by the analysis.

- *Communality* – squaring and adding the loadings for each item or variable calculates this. It is an indicator of the proportion of variance for each item that each factor accounts for.

Once the initial factor analysis has been computed, there are many different sets of factors which could produce the observed matrix, and factors need to be 'rotated'. Rotation of the factors is a procedure used to clarify the relationships with the correlation matrix, to ensure that the simplest structure is obtained. In terms of identifying which or how many factors should be extracted for rotation, the Scree test is one method that has been proposed (Kline, 1998), and is considered the best solution (Kline, 1994). The Scree test (Cattell, 1966) evaluates a plot of eigenvalues and it is possible to identify from the slope where there is a plateau in the decreasing eigenvalues. One potential flaw in this process is its subjectivity. Figure 2.1 is an example of a Scree test. In this example, six factors would be selected for rotation.

Figure 2.1- A Scree test



### 2.2.3 Sensitivity and specificity, positive and negative predictive values.

Sensitivity and specificity are reported in percentages, which are generated as a result of the following type of 2x2 table.

Table 2.6 2 X 2 table

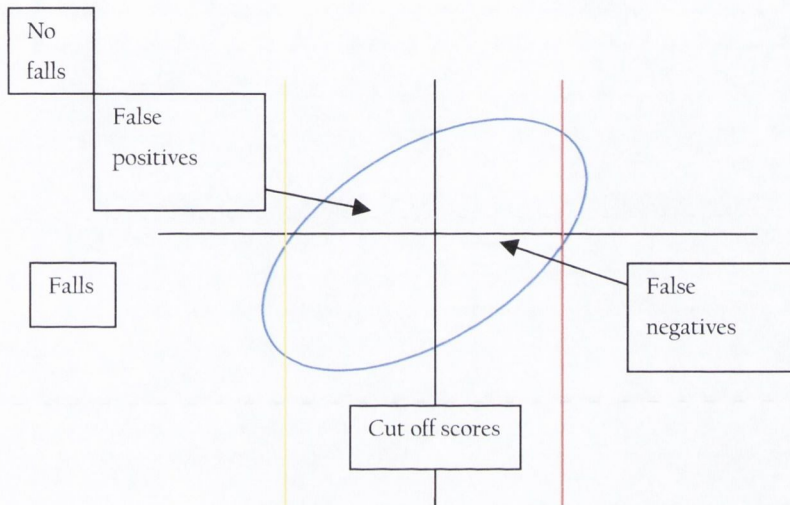
Result of test	Gold standard or criterion		
	Condition present	Condition absent	Total
Condition present	True positive (a)	False positive (b)	a+b
Condition absent	False negative (c)	True negative (d)	c+d
Total	a+c	b+d	

Sensitivity is calculated as  $100\% \times (a / [a + c])$  and specificity as  $100\% \times (d / [d+b])$ . Positive predictive values and negative predictive values may also be calculated from such a table – positive predictive values [ $100\% \times (a / [a + b])$ ], negative predictive value [ $100\% \times (d / ([c + d]))$ ] (Riddle *et al.* 1999). Using data from two studies on the Berg Balance Scale (BBS), Riddle *et al.* (1999) explore the sensitivity and specificity of this scale. The BBS is a scale that measures balance in older people. It is scored out of 56, with higher scores suggesting better balance. A single cut-off point of  $<45/56$  has been utilised to identify people at risk of falling. Combining the data from the two previously published studies (Bogle Thorbahn & Newton 1996, Shumway-Cook *et al.* 1997) and using the cut-off point of  $<45/56$  to identify those at risk of falling yielded a sensitivity of 64% and a specificity of 90%. This suggests that 1/3 of fallers were missed by the BBS and that it is more accurate in identifying non-fallers, since only 10% of them were incorrectly identified. The authors note that since sensitivity and specificity are not dependent on the prevalence of the disorder, the information can be applied to how useful the BBS will be when used with any group of older people with balance disorders.

Positive and negative predictive values are influenced by the prevalence of the disorder and the data on the BBS used by Stratford *et al.* (1996) related to prevalence of 50% (Shumway-Cook *et al.* 1997) and 17% (Bogle Thorbahn & Newton 1996) combined. If the prevalence of the disorder in the studies reporting the development of the instrument does not approximate to the subjects or patients to whom the test is to be applied in practice, these values are of a limited nature. The positive and negative predictive values for the BBS are reported as 72% and 85% respectively when using the cut off point of  $<45/56$ . This suggests, as with specificity and sensitivity, that the BBS is better at identifying correctly non-fallers, since 28% of subjects who had a history of falls were not clearly identified as fallers.

Figure 2.2 demonstrates the effect of moving a cut-off point up and down on the number of false positive and false negative results. The blue ellipse represents a scatterplot of the relationship between, in this example, scores on the Berg Balance Scale and whether a subject falls or not. The black line in the centre represents a cut-off point of 45, which is suggested by Berg (1989) as the score below which people were at risk of falling.

Figure 2.2 The influence of cut-off scores



It can be seen that this yields a set of false positive and false negative results, and corresponds to the positive predictive value of 72%, the negative predictive value of 85%, a sensitivity of 64% and a specificity of 90%. Moving the cut-off point up and down changes the relative values of each of these validity indices. Hence the decision of where to locate the cut off point is informed by the consequences of false positives and false negatives i.e. is it worse to identify a patient as a risk of falling when they are not, than not to identify a patient who is at risk of falls when they are.

Table 2.3 above illustrates the reported sensitivity, specificity and likelihood ratios (LR's) for a number of physiotherapy measurements. The interpretation of reported likelihood ratios for various tests or measurements helps the user to decide if the addition of a given test may enhance clinical decision-making. LR's are used to more accurately establish how performance on a measurement or presence of a series of signs or symptoms will change the probability that a patient will have a condition e.g. be at risk of falls, improve due to a specific intervention. Pre-test

probability is estimated before the test is used. This may be estimated by using individual clinical judgement based on the history of the patient, or in the case of the studies listed in Table 2.3, the actual prevalence in the study sample. Thereafter, the likelihood ratio nomogram (Figure 2.3) may be utilised by

1. marking the pretest probability on the left of the nomogram
2. marking the LR of the test on the middle line
3. drawing a line through these two points will estimate the post-test probability.

The nomogram (Fagan, 1975) provides a user-friendly way of utilising LR's as opposed to the mathematical calculations which would otherwise have to be completed i.e. convert pre-test probability to odds, then multiplying the result by the LR which would give the post-test odds, which then have to be converted into post-test probability (Boyko, 1994). Jaeschke *et al.* (1994) provide the following guide to how to interpret the size of reported LR's i.e. is the reported LR's such that the use of the associated instrument of measurement will better inform practice

- LR's of greater than 10 or less than 0.1 may generate large changes in pre and post-test probability
- LR's of 5-10 or 0.1-0.2 generate moderate changes in pre and post-test probability
- LR's of 2-5 and 0.5-0.2 can generate small changes in probability
- LR's of 1-2 and 0.5 -1 add rarely important changes to pre-test probability.

Riddle *et al.* (1999) present the following positive likelihood ratios (PLR) for different intervals of the BBS,

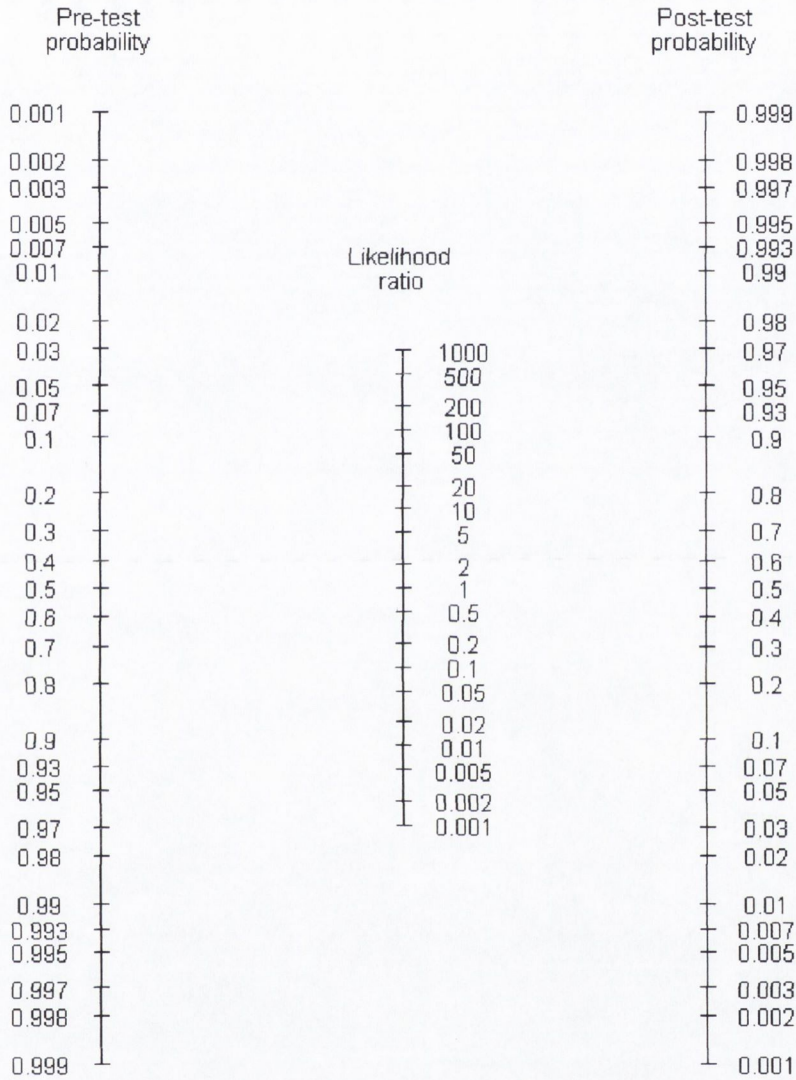
BBS <40/56	PLR 11.7
BBS 40-44/56	PLR 2.8
BBS 45-49/56	PLR 1.3
BBS >54/56	PLR 0.1

A pre-test probability of risk of falls of 50% estimated by the PT based on the history and presenting signs, and a BBS score of 38/56 which would yield a PLR of 11.7, applied to the nomogram results in a post-test probability of 92%, suggesting that in using the BBS test score, the PT has increased the certainty about the patient's risk of falling from that identified as a result of an educated guess. Of note in the studies listed in Table 2.3 is that the majority of the PLR's are not greater



than 4 and have wide 95% confidence intervals (CI's), suggesting a limited value for the use of the instruments or measurements, to date.

Figure 2.3 Nomogram for application to likelihood ratios (Fagan, 1975).



### 2.2.4 Group differences, discriminant validity

In considering the construct that scores on a new instrument will be different in one group of subjects versus another, often two extreme groups, the application of statistical tests for two unrelated groups is reported. These may be unpaired t-tests, with confidence intervals, or the Wilcoxin rank sum (two-sample) or Mann-Whitney U tests (Petrie & Sabin, 2000); the choice depending on the type of data and the assumption of normal distribution. If an instrument is designed to be a diagnostic tool, it is not possible to infer from this type of analysis about finer discriminations. Alternative methods and analysis should be employed (see 2.2.3). The multi-train-multimethod matrix is one method of considering both convergent and discriminant validity and is explained with the help of Table 2.7, using the example of comparing a new mobility scale with an existing one e.g. the Elderly Mobility Scale (EMS), in two groups of older people - one community dwelling and the other in hospital in receipt of rehabilitation.

Table 2.7 Multi-train-multimethod matrix

		Community dwelling		Hospital	
		EMS	New scale	EMS	New scale
Community dwelling	EMS	•			
	New scale	⊗	•		
Hospital	EMS	□	◆	•	
	New scale	◆	□	⊗	•

• - level of agreement between the same method

⊗ - homotrait-heteromethod i.e. same people, different method

□ - heterotrait-homomethod i.e. different people, same method

◆ - heterotrait-heteromethod i.e. different people, different method

The correlation that would be expected to be highest would be the correlation between the same measures, a method often used to report on the reliability of the instrument. The lowest correlation would be expected between different people and different groups i.e. the scores of the new scale in hospital patients and the scores of EMS in community dwelling people. For discriminant validity to be supported, it is also to be expected that there would be a low correlation between scores on the new scale in both groups. But for convergent validity, it would be

expected that a relationship does exist between scores for people in hospital on the EMS and the new scale and for people living in the community and the two scales i.e. convergent validity and correlational validity. Despite a search of AMED, Medline and CINAHL, no physiotherapy studies were found employing such a methodology.

## 2.2.5 Responsiveness

There are a variety of co-efficients that are reported when the responsiveness of a measurement is being analysed (Deyo & Centor 1986, Guyatt 1987, Liang *et al.* 1990, Liang 1995) and these include

- Receiver operator characteristic curve
- Relative efficiency
- Standardised response mean
- Effect size

### *2.2.5.1 Receiver operator characteristic (ROC) curves*

Receiver operator characteristic (ROC) curves, used in the context of considering the responsiveness of a change scores in an instrument or scale between two time points, investigate the sensitivity and specificity of the change scores in comparison to an external 'gold standard' or criterion of important change (Deyo & Centor 1986). In this instance, sensitivity is defined as the number of subjects correctly identified as undergoing important change by a particular scale or instrument and specificity is the number of subjects correctly identified as not having important change. The area under the curve represents the probability of correctly identifying the improved subject from a randomly selected group of improved and unimproved subjects (Stratford *et al.* 1996a). A ROC curve can be presented for varying cut-off points. The area under the curve ranges from 0.5 (no accuracy in detecting improved from unimproved) to 1.0 (perfect accuracy) (Deyo & Centor 1986). Stratford *et al.* (1998) reviewed the responsiveness of the Roland-Morris Questionnaire in 226 subjects who presented with a current episode of LBP of less than 6 weeks duration. The Roland-Morris Questionnaire was measured at baseline and following 3-6 weeks of physiotherapy. Clinician and patient rating of global rating of important change ranged from -7 (deterioration) to +7 (improvement) and a mean clinician and patient global rating score of 5 was considered to represent an

important amount of change. This represented the external 'gold standard' and 63% of the sample was classified as having changed an important amount. The area under the ROC curve result for the RMQ was 0.84. Riddle *et al.* (1998) in an accompanying paper also presents the responsiveness of the RMQ, but this time using a different 'gold standard' of change i.e. a clinician's judgement of whether the participant had achieved their treatment goals. The area under the ROC curve in this study of 143 people with low back pain was 0.68. The authors of both papers, used ROC curves to identify the RMQ change score that maximised the overall accuracy of classifying patients as those who had undergone their respective 'gold standards of change' and those who had not. The authors noted that this was a change in RMQ score of 5 points.

Davidson & Keating (2002) compared the responsiveness of the

- Modified Oswestry Disability Questionnaire (mODC)
- Quebec Back Pain Disability Scale (QBPD)
- Roland-Morris Disability Questionnaire
- Waddell Disability Index
- Physical health scale of the SF-36

in patients undergoing physiotherapy(PT) for low back pain. Measurements were taken at initial consultation and after 6 weeks. External criterion for improvement was if the participant rated their back pain as 'completely gone', 'much better', or 'better' at follow-up. Area under the ROC ranged from 0.73-0.78 for the scales, no significant difference was noted between the ROC curve results for each scale. This is consistent with the results of a similar study comparing the mODC and the QBPD) scale in a group (n=67) of subjects with acute, work-related low back pain, also having PT intervention (Fritz & Irrgang 2001). The authors found not difference between the area under the curve (AUC) for the two scales (mODC-0.94, QBPD-0.87). Wagner *et al.* (1993) also employed this method when comparing health status measures in older adults. Self-reported health status measures such as number of restricted activity days, number of days spent in bed due to illness or injury, a physical limitations scale, self-evaluated health and a scale indicating psychological well-being were employed and measured at baseline and one year later. The external criterion was an episode of hospitalisation or major illness. The results are presented by comparing the area under the ROC curve obtained for each measure to a hypothetical result of 0.5 since this would represent a non-responsive

measure. The most responsive health status measure were days spent in bed and restricted activity days and the physical activity measure, all of whose area under the ROC curve was significantly different to 0.5.

### 2.2.5.2 Standardised response mean and the effect size

The standardised response mean (SRM) is the mean change in scores taken at two time points, divided by the standard deviation of the changes (Liang 1995). The effect size (ES) provides a ratio of the 'signal to noise' within measurement scores on a particular instrument, where the 'signal' is the change that occurs due to intervention and the 'noise' is the level of inherent variability within the instrument. The ES has been used in conjunction with design 1 in Table 2.2. Using the standard deviation of baseline scores as a measure of the variance within the data i.e. noise and the change in scores from admission to discharge as the 'signal', this ratio may be calculated. Cohen (1977) notes that effect sizes that are larger e.g. 0.8 or greater suggest that the measurement instrument is responsive to clinical change. Effect sizes of less than 0.2 should be interpreted as small and 0.5 as moderate. Wright *et al.* (1998) measured effect size for both the Rivermead Mobility Index with the Barthel Index and noted effects sizes of 1 and 0.87 respectively. Stokes *et al.* (2003) noted effect sizes of 0.9 for the Elderly Mobility Scale and 0.8 for the Barthel Index. Liang *et al.* (1990) compared the responsiveness of five health status instruments for orthopaedic evaluation- Arthritis Impact Scale (AIMS), Functional Status Index (FSI), Health Assessment Questionnaire (HAQ), Index of Well Being (IWB), and the Sickness Impact Profile (SIP). Thirty-eight participants two weeks prior to hip (55%) or knee (45%) arthroplasty, and at 3, 12 and 15 months after surgery were included in the study. Responses were divided into 'early' response (baseline-three months postoperatively) and 'net' response (baseline to 12 and 15 months postoperatively). The five questionnaires were gathered at each time point. Rather than evaluate the total score from each instrument, four domains were identified- pain, global health (original total scores), mobility and social functioning and compared across the five instruments, except pain which is only considered by three of the five instruments, namely AIMS, FSI and HAQ. Cohen's (1977) guide to interpreting effect sizes (see above) was applied by the authors to the SRM. Table 2.8 illustrates the range of SRM reported by the authors and their associated 95% CI.

Table 2.8 SRM for five orthopaedic measurement instruments (adapted from Liang *et al.* 1990)

	Pain	Global health	Mobility	Social
AIMS early	1.11 (0.68-1.54)	0.88 (0.38-1.38)	1.01 (0.42-1.60)	0.17 (-0.18-0.52)
AIMS net	1.15(0.67-1.63)	1.36 (0.87-1.85)	1.41 (1.08-1.74)	<b>0.39 (0.09-0.69)</b>
FSI early	1.08 (0.51-1.65)	<b>0.40(-0.02-0.82)</b>	0.78(0.33-1.23)	<b>0.54 (0.20-0.88)</b>
FSI net	1.00 (0.51-1.49)	<b>0.61 (0.15-1.07)</b>	0.84 (0.45-1.23)	<b>0.51 (0.03-0.99)</b>
HAQ early	0.95 (0.43-1.47)	<b>0.33 (-0.10-0.76)</b>	0.57 (0.08-1.06)	<b>0.35 (-0.09-0.79)</b>
HAQ net	0.89 (0.46-1.32)	1.00 (0.53-1.47)	0.84 (0.37-1.31)	0.98 (0.70-1.29)
IWB early	N/A	1.13 (0.80-1.46)	<b>0.68 (0.17-1.19)</b>	<b>0.34 (-0.07-0.75)</b>
IWB net	N/A	0.88 (0.56-1.20)	0.91 (0.45-1.37)	0.82 (0.48-1.16)
SIP early	N/A	0.71 (0.32-1.10)	0.94 (0.54-1.34)	0.51 (0.14-0.88)
SIP net	N/A	1.11 (0.74-1.48)	1.10 (0.79-1.41)	0.93 ((0.62-1.24)

If one evaluates the above SRM in the light of Cohen's guideline - 0.8 or greater suggest that the measurement instrument is responsive to clinical change, effect sizes of less than 0.2 should be interpreted as small and 0.5 as moderate- the sections highlighted in bold are those where either the SRM or the associated 95% CI may be considered to be less than a 'moderate' effect, suggesting limited responsiveness. It is interesting to note the influence of timeframe on the responsiveness of the various instruments i.e. 0-3months(early) versus 0-12/15 months (net), suggesting that some of the instruments would be more useful in capturing change over a short period of time. This is consistent with the findings of Benaim *et al.* (1999) who investigated the responsiveness of the Postural Assessment Scale for Stroke (PASS) in a group of participants with acute stroke. They divided their sample into mild, moderate and severe stroke and reported the effect size of the PASS for different numbers of days post stroke (DAS). Table 2.9 summarises their findings.

Table 2.9 Effect sizes reported for PASS (adapted from Benaim *et al.* 2000)

Days after stroke	Total sample	Mild	Moderate	Severe
14-30	0.89	0.60	1.14	1.01
30-90	0.64	<b>0.29</b>	0.5	1.02
90-180	<b>0.31</b>	<b>0.34</b>	<b>0.29</b>	<b>0.36</b>
14-90	1.07	0.76	1.18	1.51
14-180	1.12	0.85	1.20	1.54

These results would suggest that the PASS is best utilised for all people with stroke in the time period up to three months after stroke and thereafter it is possibly not responsiveness in those people with mild disability after stroke.

Relative efficiency has also been used to consider a number of rehabilitation measurement instruments (Wright *et al.* 1998, Stokes *et al.* 2003, Liang *et al.* 1985). Wright *et al.* (1998) and Stokes *et al.* (2003) both evaluated the mobility scales and compared them to the Barthel Index, considered a 'gold standard' in the measurement of functional status. Relative efficiency (Liang *et al.* 1985) is a method of comparing a number of measurement instruments with a view to identifying the relative responsiveness. Depending on the output of the instrument, for example, whether the data is ordinal or interval a *t* or *z* score is obtained by comparing admission and discharge scores, or indeed scores from any other two designated time points. Relative efficiency is calculated using the following formula

$$\text{Relative efficient (X versus Y)} = (\frac{X}{Y})^2$$

A relative efficiency score of 1 would indicate that X is equally as responsive as Y, a score of >1 would suggest that X was more responsive than Y and a score of <1 that Y was more responsive than X. Using this analysis, it has been demonstrated that both the Rivermead Mobility Index and the Elderly Mobility Scale are more responsive than the Barthel Index in older people participating in rehabilitation programmes (Wright *et al.* 1998, Stokes *et al.* 2003). Liang *et al.* (1985) compared the relative efficiency of the AIMS, HAQ, IWB and SIP to the FSI (see discussion above in 2.2.5.1, same sample). The relative efficiencies reported for each of the four domains were as follows

- Mobility- AIMS (0.85), HAQ (0.48), IWB (0.45), SIP(1.11)
- Pain- AIMS (0.79), HAQ (0.57)
- Social- AIMS (0.18), HAQ (0.62), IWB (0.54), SIP (0.74)
- Global- AIMS (4.12), HAQ (1.15), IWB (7.50), SIP (3.51)

Suggesting that for measuring change in global health, each of the other four instruments is more responsive than the FSI.

The choice of which analysis to use is dependent on the design of the study and the type of data gathered i.e. ordinal or interval/ratio. Liang (1995) suggests that to date, no one method has become standard.

### 2.3 Reliability

The measurement property of reliability refers to the degree to which the measurement instrument is free of random or variable errors (McDowell & Newell, 1996).<sup>1</sup> Each observed score on a measurement instrument is the composite of the true score and error. Random errors may occur during any part of the measuring process and may be a product of inattention, fatigue or inaccuracy. Random errors display certain characteristics (McDowell & Newell, 1996)-

- Their size does not relate to the size of the true score
- They cancel one another out, if sufficient measures are taken
- They are as likely to increase as decrease the true score.

Rothstein (2001) suggests that reliability 'is not an all-or-none phenomenon, but rather it lies along a continuum'. Interpreting the ways in which reliability has been tested and considering the degree of reported error will inform if the instrument is used in practice. Reliability has been described as relative and absolute (Finch 2002). Absolute reliability is expressed as the standard error of the measurement and is expressed in terms of the actual unit of the original instrument (Stratford, 2004). This is probably of more use to the clinician on a day-to-day basis. The SEM can be used to generate the minimal detectable change (MDC), which is the

---

<sup>1</sup> In measurement, there is the potential for a variety of errors. These have been classified as namely random and systematic (McDowell & Newell, 1986), or variable, personal, constant and interpretative. The degree to which a measurement instrument is free of random or variable errors is considered in the context of reliability (Helmstadter 1966). The other types of error are normally considered during the construction of the instrument.



minimal amount of change in the score of an instrument that must occur in an individual in order to be sure that the change in score is not simply attributable to measurement error (Stratford *et al.* 1996, Stratford & Goldsmith, 1997). MDC values are reported at 90% and 95% confidence levels. The reliability of an instrument is often reported in three ways, inter-rater reliability, intra-rater reliability and internal consistency. Inter-rater reliability requires the same group of subjects to be measured at the same time by different observers. Intra-rater reliability considers the same subjects, the same rater and measurements taken at different time points. Internal consistency studies investigate the measurements of subjects in parallel (Finch *et al.* 2002). In evaluating the reliability of the Elderly Mobility Scale (EMS), Smith (1994) and another physiotherapist measured the performance of 15 older people. The two raters performed the measurements independently. Neither internal consistency or intra-rater reliability was reported in the paper. A later study on the EMS considered inter-rater reliability using a similar method (Prosser & Canby, 1997). Berg (1989) reported the inter- and intra-rater reliabilities in addition to the internal consistency of the Berg Balance Scale in the first paper published on the instrument. Five raters reviewed videotaped performance of 14 subjects and the results of their scoring were compared. One week later four of the raters reviewed the videotapes and re-scored the subjects, with a view to considering the intra-rater reliability. The Postural Assessment Scale for Stroke Patients (PASS) is a scale for measuring the postural abilities in people with stroke. Similar methods were employed by the authors (Benaim *et al.*, 1999) to establish inter-rater reliability i.e. two different raters reviewing 12 subjects with stroke and intra-rater reliability i.e. one rater reviewed the same subjects three days later. It can be seen that while these methods seem similar, there are different potential sources of error within each of the designs. In the review of the BBS, all of the raters viewed the same subject, at the same time on videotape. For the PASS study, the raters viewed the subjects separately but on the same day and this method was employed in the EMS studies. In the BBS study, the only source of error is the raters but in the other studies, there is an additional potential source of error and that is the client error.

## 2.4 Analysis of reliability studies

The output of the analysis of reliability studies may be termed the reliability coefficient and as with the co-efficients described for criterion and convergent validity, the type of data- nominal, ordinal or continuous- will inform selection. The use of the kappa and weighted kappa coefficients are used for reporting on ratings that are dichotomous. A 2x2 table identifying the level of agreement between raters can be used as a simple and effective way of illustrating the level of agreement. It does not take into consideration the proportion of agreement that is attributable simply to chance (Streiner & Norman, 1989). The K coefficient corrects for chance agreement by calculating the extent of agreement that could exist between raters by chance. The weighted K (Cohen, 1968) extends this concept and allows for partial agreement between raters. For measurements that generate continuous data, analysis of variance (ANOVA) is employed and a resulting intraclass correlation is reported (Streiner & Norman 1989, Deyo *et al.* 1991, McDowell & Newell 1996, Finch *et al.* 2002). The Intra-class correlation co-efficient (ICC) is an indication of the variance due to error, this is calculated using different contributory sources of variance. Six forms of ICC are described by Shrout and Fleiss (1979) for inter-rater reliability and each depends on the experimental design and the later use of the instrument whose reliability is being tested. In a study where two observers rate a number of subjects, there will be variance due to the subjects, the observers and random error. The ICC is calculated from the results of an ANOVA for repeated measures. This provides an indication of the relative reliability of the measurement instrument. The ICC is not related to the measurement units of a given instrument. The standard error of the measurement (SEM) can be calculated in a number of ways (Stratford, 2004) and by providing this additional information in reliability studies may provide more meaningful data to physiotherapists. Table 2.10 illustrates the methods that may be used to calculate the SEM (Stratford, 2004).

Table 2.10 Methods to calculate the SEM ( $\sqrt{\text{square root}}$ )

Method 1	Design	Output	SEM=
Method 1	Testing at two time points, with both sets of data having equal variance (sd)	Standard deviation (sd) and Pearson's correlation coefficient (r)	$s\sqrt{1-r}$
Method 2A	Testing at two time points.	One-way ANOVA to calculate relative reliability. Use Mean Squares for within subjects from ANOVA table	$\sqrt{\text{MSW}}$
Method 2B	Testing at two time points but with a randomised block format	One-way ANOVA to calculate relative reliability. Use Mean Squares error term from ANOVA table	$\sqrt{\text{MSE}}$
Method 3	Testing at two time points.	Estimate the sd of the difference scores	$\text{sd}/\sqrt{2}$

Because of the methods employed to calculate the SEM, it demonstrates population invariance (Wywich *et al.*, 1999) i.e. it is sample independent. Its interpretation is that there is a 68% chance that a true score will fall within  $\pm 1$  SEM of the observed score. To ascertain 95% confidence levels the SEM is multiplied by 1.96. In order to calculate the MDC at this level the calculation is as follows,

$$\text{SEM} * 1.96 * \sqrt{2}$$

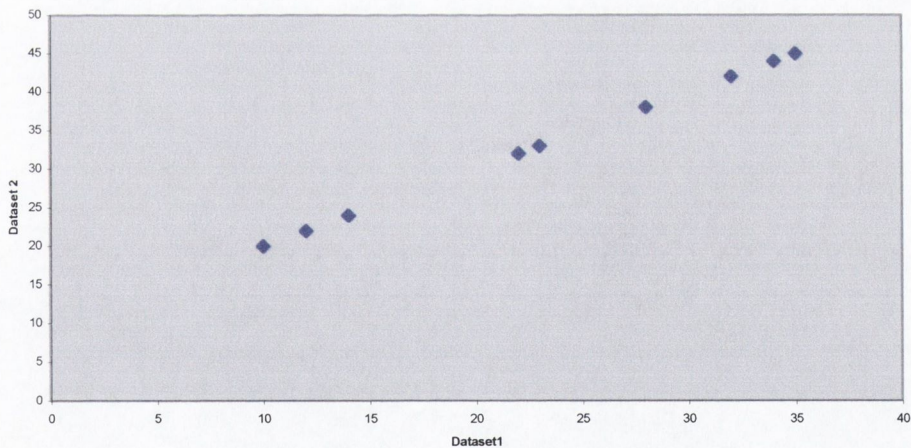
The incorporation of the  $\sqrt{2}$  allows for the fact that error may also occur as a result of two measuring episodes. Although the SEM is reported to be relatively constant across all levels of a population's ability, except extremes, SEM's are reported for a variety of different initial scores (Stratford 1996b, 1998). The SEM for the Fugl-Meyer assessment of sensori-motor recovery after stroke is 9.4 for the total instrument (maximum score 226) (Sanford *et al.* 1993) although to date the MDC has not been reported. The MDC<sub>95</sub> and MDC<sub>90</sub> for the Berg Balance Scale are reported as 6.9 and 5.8 points respectively (Stevenson, 2001). In this study, Stevenson demonstrates that in a group of subjects with stroke who were still receiving in-patient rehabilitation, the MDC<sub>95</sub> may vary depending on functional status from 6.0 points in subjects who require standby assistance to 8.1 in those who require assistance. Conditional MDC<sub>90</sub> of 4-5 points is reported for the Roland Morris Questionnaire, since the authors (Stratford *et al.* 1996) note that this applies only to subjects with initial scores of between 4-20 on the scale.

The Pearson-product moment correlation coefficient (PPMCC) is frequently cited in support of the reliability of an instrument. Nonetheless this may not be an accurate representation of the level of agreement between raters or within a rater (Bland & Altman 1986). The PPMCC reflects the degree to which two sets of continuous data have a linear relationship. Thus, the two sets of data in Table 2.11 yield the scatterplot in Figure 2.4 and a correlation co-efficient of 1.0. If this represents the same rater measuring the same subjects at different times (but with no anticipated changes in the subjects' scores), or indeed, if it reflects two raters measuring the same subject at the same time, despite reporting a very high PPMCC, there is no agreement between the raters. Hence, when interpreting the results of reliability studies on measurement instruments the PPMCC should be interpreted with caution.

Table 2.11 Reliability datasets

Subject	Dataset 1	Dataset 2
1	10	20
2	12	22
3	14	24
4	22	32
5	32	42
6	34	44
7	35	45
8	28	38
9	23	33
10	14	24

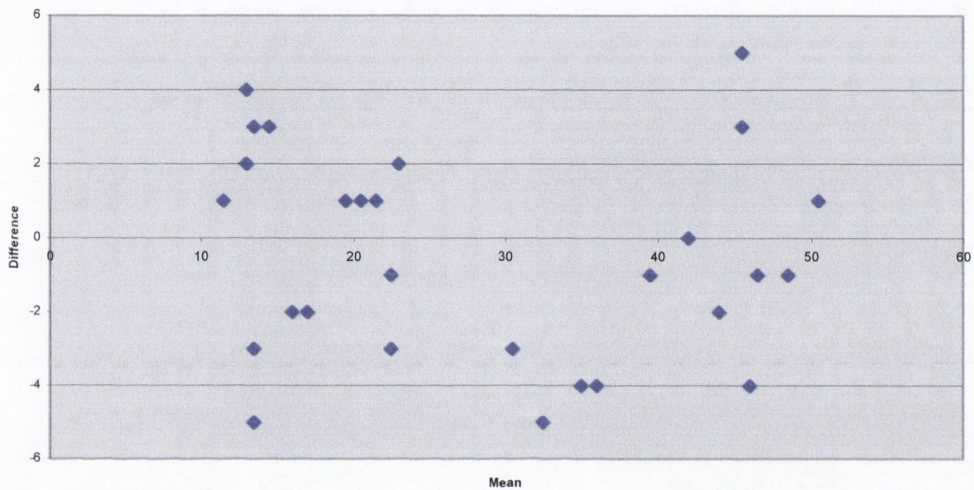
Figure 2.4 Scatterplot of datasets



Bland and Altman (1995) report a more accurate method of assessing the degree of error between sets of data, over the full range of reported scores. By plotting the mean of the sets of scores against the differences of the two scores for each person, it is possible to consider if error changes over the range of the scale. Figure 2.5

illustrates the mean versus the difference for a synthetic dataset. The 95% limits of agreement can then be calculated- the mean of the differences  $\pm 1.96^*$  standard deviation of the differences (-6 to +5 in this sample). In this figure, it can be seen that almost all the points within the limits of agreement, suggesting a high level of agreement. This method is reported by Benaim *et al.* (1999) and Stokes *et al.* (1998) when analysing the inter- and intra-rater reliabilities of rehabilitation outcome measures.

Figure 2.5 Bland &amp; Altman method



A further method of analysing a measurement instrument for reliability is termed internal consistency. This is also a measure of the homogeneity of the instrument. It is often employed for evaluating questionnaires; nonetheless it is useful also in the case of multi-item instruments where each of the internal items or domains is summed to form a composite score (Finch *et al.* 2002). In essence the relationship between the items and their relationship with the overall score is considered. The higher the relationship between the individual items, the easier it would be to create two similar versions that are reliable, thus improving test-retest reliability. Nevertheless, inter-item correlation should not be too strong because it may negate the need to retain strongly related items. All the items should relate to the total score because this supports the construct that one theme or domain is being measured, albeit through the contributory components. Internal consistency may be reported through the expression of inter-item correlations, using PPMCC to

illustrate the level of association between the items (Streiner & Norman, 1989). When comparing individual items to the overall score, the latter should exclude the individual item to which it is being compared. The most common statistical expression of internal consistency is Cronbach's alpha (Finch *et al.* 2002, McDowell & Newell 1996, Streiner & Norman 1989). An  $\alpha$  of  $> 0.8$  is considered has been reported as high but McDowell and Newell (1996) caution that this does not always indicate a moderate -high inter-item correlation. Berg (1989) reported a Cronbach's  $\alpha$  of 0.96 for the BBS with inter-item correlations ranging from 0.71-0.89. Benaim *et al.* (1999) report a Cronbach's  $\alpha$  of 0.95 for PASS, the authors did not report inter-item correlation coefficients.

## 2.5 Conclusion

The relevance of reported measurement properties is important when the instrument is being chosen for either research or clinical practice. The methods employed to investigate the properties and the analysis used should be subject to critical appraisal and in this light of this review, the relevance of the studies to clinical decision-making can be ascertained. This does not routinely happen in clinical practice. Difficulties arise in the diversity of publications about a given instrument or scale and there are few published reviews of the measurement properties of individual instruments or scales. A small number of professional organisations have considered ways of making these findings available to physiotherapists in a meaningful way and this will be discussed in more detail in the next chapter.

This chapter has reviewed the descriptions of the two principle measurement properties, and how methods to evaluate them have been employed in some rehabilitation measures. The relevance of the expression of various forms of analysis has been reported. To date many instruments being used by physiotherapists lack comprehensive reporting of measurements properties, and often even when this information is available, it is not used to maximise the potential of the measurement instrument. This is possible because information on measurement properties such as that described in this chapter is often not available in a format that is user friendly or when it is available it is used in a superficial manner.

CHAPTER 3 THE OUTCOMES MOVEMENT & PHYSIOTHERAPY

3.0 Introduction

In the past two decades, the focus of many health care policies and initiatives has been associated with the increased desire for both accountability and quality in health care (Kane 1997). The delivery of quality health care requires information on both the appropriateness of the intervention or management and its effectiveness. Kane (1994) suggests that appropriateness is informed by 'clear evidence of efficacy... under a specified situation' and effectiveness requires the measurement of outcome, the two terms, he believes are not synonymous. Liebenson & Yeomans (1997) suggest that quality is demonstrated by improved outcomes and the utilisation of evidence-based intervention. Bury & Mead (1998) employ the premise that clinical effectiveness encompasses evidence-based practice (EBP), and suggest that clinical effectiveness requires the consideration of evidence in the context of external environmental and organisational influences. These represent just a fraction of the definitions used to illustrate a variety of terms employed in the language of health care evaluation. By defining each of these terms individually, the relationship between them may be fractured and it does not reflect the place of each and all in day-to-day clinical practice. Table 3.1 illustrates the dynamic relationship that may exist between clinical effectiveness, evidence-based practice, clinical guidelines, outcomes research and the systematic use of standardised outcome measurement.

This chapter briefly considers the outcomes movement, evidence-based practice and clinical effectiveness in physiotherapy as a means of providing a context for the main review of outcome measurement in physiotherapy practice. This review considers the practice of systematic outcome measurement, the barriers reported to the systematic use of standardised outcome measures (sSOM) and the role of professional organisations in promoting and supporting sSOM.

3.1 The outcomes movement

In the late 1980's Relman described the outcomes movement as the 'third revolution in medical care' (Relman 1988). In his opinion, the *Era of Expansion* came after the Second World War and continued until the late 1960's. This was followed by the *Era of Cost Containment*. He suggested that the third stage was the *Era of Assessment and Accountability*. The origins of this era are not completely clear (Kane 1997) but it is



likely that a number of factors contributed. Large variations in the delivery of service, coupled with increasing expenditure, prompted questions about whether differences in outcome existed. With increasing amounts of national budgets being spent on health care, decisions about health care expenditure required information about the relative effectiveness of interventions and services, with a view to minimising unnecessary expenditure- cost containment within the context of preserving quality of care. Managed care, with industrial accountability and productivity models, generated a revised way of thinking. Globally, the health care system had become a competitive market place. Individuals, insurance companies, health maintenance organisations, national health services and individual governments are all purchasers of health care. Market decisions are informed by outcomes of care (Epstein 1990, Jette 1995, Enderby 1995, Kane 1997, Hammond 2000, Beattie 2001).

Epstein (1990) described the three-fold effect of the outcomes movement on assessing outcomes. The emergence of a value placed on outcomes information had a subsequent impact on the way information is collected and stored. In some health care systems, large computerised databases are used to inform billing and reimbursement; while the presence of desktop computers in many departments and services has led to local analysis of data. This large-scale collection of data can be used to inform outcomes research and thus expand the existing knowledge base, working in tandem with the results of randomised controlled trials. The range of outcomes measured is broader. If health as defined by the World Health Organisation<sup>1</sup> (1948), has become more measurable because of the development of well-constructed and evaluated measurement instruments (McDowell and Newell 1996), the outcomes movement has influenced the way that these outcome measures are used to inform practice and decision-making (Epstein 1990, Kane 1997).

### 3.2 Evidence-based practice

Evidence-based practice became the buzzword of the 1990's (Bury & Mead 1998), albeit that the origin of the concept is at least 150 years old (Sackett 1996); its emergence fuelled by increasing research activity and the desire to bring findings into practice and supported by published formats that made the information more accessible and advances in information technology. In physiotherapy, a long tradition

---

<sup>1</sup> 'Health is a state of complete physical, mental and social well-being and not merely the absence of disease and infirmity.'

of research does not exist, and a critical mass is still only emerging in some areas of practice and not present at all in others (Twomey 1996, Parry 1997). A number of factors have been cited as potential contributors to explain why some areas of healthcare do not have supporting evidence derived from research (Appleby *et al.* 1995, cited in Bury & Mead, 1998):

- Difficulties in designing studies
- Difficulty in removing an intervention, which through custom and practice is now accepted, to perform a clinical trial
- Existing research that is of poor quality
- Randomised controlled trials not always appropriate to specific areas of healthcare
- Inadequate attention paid to the cost-effectiveness of interventions
- Failure to disseminate research findings.

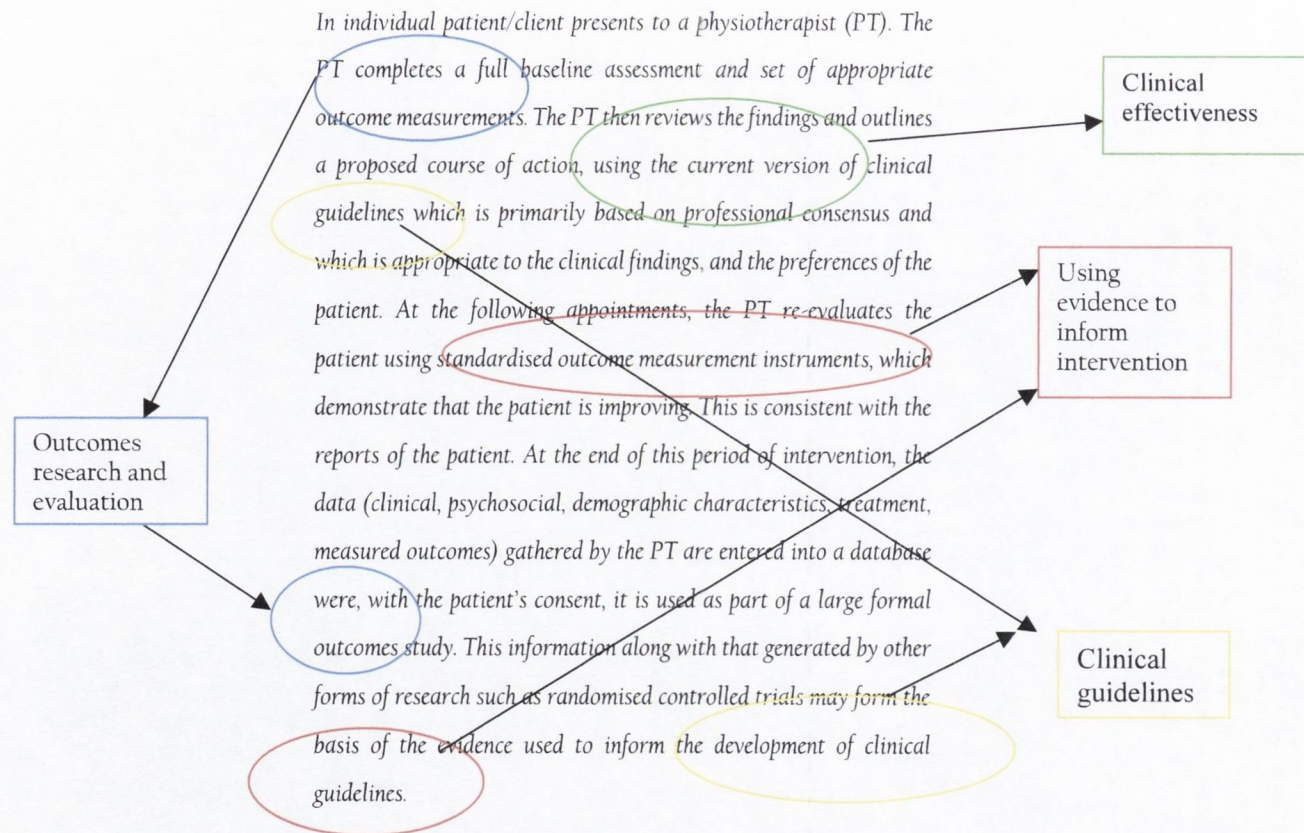
In addition, specific factors in physiotherapy relate to the development of the profession, its historical placement under the auspices and protection of the medical profession (Roberts 1994, Parry 1995) and its lack of professional autonomy. The profession grew out of the establishment of the Incorporated Society of Trained Masseuses in 1894 (Barclay 1994), which was later to become the Chartered Society of Physiotherapy. The practice of its members could only occur under doctor's orders. It was only in 1977 in the United Kingdom<sup>2</sup> that physiotherapists finally gained professional autonomy and became first contact practitioners. This occurred in the 1980's in other countries (Turner 2001). Prior to this there was a requirement for a referral from a medical practitioner, who also prescribed the treatment. The development of educational programmes for physiotherapy occurred in a non-uniform manner internationally. The United States had the first graduate programme (Moore 1995) in the 1920's, South Africa, Canada and Australia in the 1940's, 50's and 70's respectively (Turner 2001). The first honours degrees programmes commenced in Ireland in the 1980's and the United Kingdom in the 1990's (Palastanga 1990). To date, the proportion of members of the Irish and UK professional organisations that hold research doctoral degrees represent approximately 0.005% of the total membership (Moore 2004). Despite the changes in the professional role, the emergence of a body of graduates with research skills and equipped for the

---

<sup>2</sup> The professional organisation of physiotherapists in Ireland was established in 1983. Prior to this it was branch of the CSP in the UK. Education and practice in Ireland was the same as in the UK until that time.

responsibility of autonomous practice and the drive for EBP in healthcare services, the use of research findings to inform the choice of treatment techniques is limited in physiotherapy (Turner 2001, Pomeroy & Tallis 2000). The factors that physiotherapists/physical therapists (PT's) use to inform treatment choice have been cited as original professional education, attendance at continuing professional development (CPD) courses, previous experience with a client or peer suggestion (Turner *et al.* 1997, 1999, Carr *et al.* 1994, Nilsson & Nordholm 1992). There is no mandate that requires the profession of physiotherapy to demonstrate efficacy of an intervention prior to its inclusion in an undergraduate course of study, or indeed in the area of continuing professional development (Stratford 1999). Often, in practice decisions are made on the basis of 'personal observations, precedence and consensus' (Parry 1997).

Across all aspects of healthcare, EBP initiatives, driven by the desire for creating greater consistency in the provision of services, have resulted in demands for and the development of Clinical Guidelines. This has presented and continues to present a challenge to many areas in health care provision where a large body of evidence does not exist to support interventions and where professional consensus is hard to reach (Kane 1997, Stratford 1999). This is especially the case in physiotherapy because our research tradition is short but also because many 'higher' forms of research suggest that some physiotherapy interventions are ineffective or selectively effective (Stratford 1999, Pomeroy & Tallis 2001). Nevertheless, clinical guidelines exist in areas such as the management of osteoporosis, stress incontinence and soft tissue injury (CSP 1998, 1999, 2001).



**Table 3.1 The potential relationships between outcomes evaluation and research, clinical effectiveness, clinical guidelines and evidence-based practice; and between the individual practitioner, researchers, developers and the provision of quality health care.**

### 3.3 The use of outcome measures in practice

Significant challenges facing the physiotherapy profession with the emergence of the outcomes movement were the absence of an agreed framework for measurement and the absence of an ethos of using standardised measurement instruments. With respect to the latter, this chapter reviews the use of systematic outcomes measurement (SOM) in physiotherapy practice under the following headings

- The extent to which standardised outcome measures (SOM) are employed in physiotherapy and related rehabilitation practice and the profile of this practice.
- The attitudes towards use and the barriers identified by PT's in hindering the use of SOM's.
- The role of professional organisations policy in promoting the use of SOM.

The information for this review was obtained from a number of sources including a review of the published literature (Medline, CIHAHL, and AMED), use of the World Confederation for Physical Therapy (WCPT) EBP advisory group and correspondence with the individual PT associations.

#### ***3.3.1 The use of standardised outcome measures in practice***

Two methods are reported in the literature for the investigation of the extent to which standardised outcome measures are used in physiotherapy practice and rehabilitation- survey instruments using self-report and retrospective chart retrieval or chart audit. The first such reported work in physiotherapy was undertaken in 1992; a task force commissioned by the Canadian Physiotherapy Association and Health Canada completed a national survey, using a stratified random sample that examined the use of client outcome measures by PT's. The definition of an outcome measure was a 'published measurement scale'. The sample included individual practitioners and PT managers and was a random sample (n=309) from a list of licensed therapists, with a response rate of approximately 80% (Mayo 1993, Cole 1994). The findings suggested that use of standardised outcome measures was limited- 50% reported using the measures, but only 20% were able to identify one 'published measurement scale'. At the same time (the study was published 3 years after its completion), Chesson *et al.* (1996) conducted a similar survey of PT's and occupational therapists (OT's). The survey participants were PT and OT managers in hospital and community based departments. There was a 79% PT response, with

only 44% of PT departments reporting to be using at least one standardised outcome measure, many of whom had only introduced the practice in the 1990's. Low rates of usage were reported in some regions suggesting the influence of local policy and SOM's more commonly used in the speciality of rehabilitation of older people than other specialities. Both studies suggested that the use of SOM's was emerging in physiotherapy practice. A pan-European review of the use of outcome measures was completed in 1998 (Torenbeek *et al.* 2001). Using a postal questionnaire, 581 rehabilitation facilities in Germany, Ireland, Italy, Austria and the Netherlands were surveyed about a number of aspects of outcome measurement. The overall response rate was only 17.5% but the results are consistent the findings of the previous two surveys; the authors concluded that, in the area of rehabilitation post stroke and for low back pain, systematic use of standardised outcome measurement is not yet common practice. The results identified that many of the measures used were 'neither published nor validated'. In a review of 182 rehabilitation centres in the UK, Turner-Stokes and Turner-Stokes (1997) also noted that 77% of the centres represented by the respondents used at least one SOM.

Retrospective chart retrieval and chart audit were the methods employed by Turner *et al.* (1996, 1999) and Kirkness & Korner-Bitensky (2002). In 1993 in the UK, Turner *et al.* (1996) screened a sample of case notes and included the case notes if pain was on the problem list, being treated or was noted in the initial assessment. 1010 case notes were selected for audit. On initial assessment 90% of cases presented with pain as a problem, 64% were treated for pain but only 21% actually quantified pain in any way. Reassessment occurred in 73% of cases and in those cases 94% of cases had pain as a noted problem on reassessment. 63% noted treatment for pain but only 2.5% quantified the pain on reassessment. Similar findings were reported in a later study by Turner *et al.* (1999), in 1994-1996, 1254 patient records were reviewed from 5 hospitals to consider the measurement of muscle strength and range of motion in the management of low back pain and after knee replacement surgery. 810 charts met the criteria i.e. the relevant parameter listed as a problem, in treatment plan, or treated by PT. In 95% of cases treatment for increasing muscle strength (MS) and range of motion (ROM) was documented. Both were initially quantified in approximately 64% of cases, but reassessment only took place in 10% of cases for MS and 30% for ROM. In both of these papers, charts from paediatric wards, patients <14 years, day hospitals for older people and patients aged > 85 years were excluded. Nevertheless,

the authors still noted that patients aged > 55 years were less likely to have pain assessed than younger patients. Pain was measured using a body chart, and the end of range pain. MS was measured using manual muscle testing, limb girth and dynamometry. ROM was measured using goniometry. The attrition in the use of SOM across the period of intervention was also noted by Kirkness & Korner-Bitensky (2002) in their investigation of the prevalence of OM use by PT's in the management of LBP. They reviewed 256 randomly selected charts from 40 PT practice settings in Canada. The prevalence of PT's consistently using standardised outcome measures was low- 34%. All but one of the respondents (53 PT's and 265 charts) employed measures of impairment of structures such as pain and range of motion, while one PT used a measure of activity limitation (disability). The authors divided their respondents into 'consistent users' and 'inconsistent users'. Clients of the former received more treatment sessions and were treated over a longer period of time; payment for their services was more likely to come from a hospital source as opposed to a private source or third party source i.e. insurance company, workers' compensation scheme. The numeric pain rating score was the most commonly noted standardised measure employed, 27% used it at initial assessment, this level of usage dropped over the duration of intervention, with only 4% using it again at discharge. This attrition in the use of standardised outcome measures was consistent across all the instruments employed, suggesting that the information obtained was not used to measure change as a result of intervention.

In 2001 the results of a further survey on practice in Canada (completed in 1998) were published (Kay *et al.* 2001). This survey focused on a review of practice following a number of strategic interventions by the Canadian PT Association (detailed below). In addition to general and specific questions on the use of outcome measurement, respondents were asked questions about their sources of information about outcome measurements and their confidence in a variety of situations relating to outcome measurement. In the second survey, respondents were identified as being either staff PT's (n=69) (the equivalent of rotational or basic grade PT's in Ireland and the United Kingdom) or professional practice leaders (PPLs) (n=20) i.e. clinical specialists/senior PT's. Direct comparison could be made between the 1992 and 1998 surveys. 41% (n=58) of staff PT's reported that published measurement scales were used in their department in 1992, this increased by only 2% (to 43%, n=26) in 1998. A more focused question asked respondents to consider a list of published outcome

measures; check if they were familiar with the measure, and if they currently used it. Almost the entire sample reported that they currently used at least one of the outcome measures. The discrepancy between the two answers leading the authors to observe that 'it is difficult to conclude whether or not overall use of client outcome measures has increased since the 1990's'. It is interesting to note that two of the three stroke specific outcome measures reported as being used in 1992, were no longer reported as in use in the 1998 survey, although approximately one in three staff PT's and PPLs reported using the Chedoke McMaster Stroke Assessment Impairment Inventory Scale- an increase of 25%. The Berg Balance Scale was used by 45% of respondents (n=40), an increase of 28% from 1992.

The five most frequently cited measures and the pattern of their use are outlined in Table 3.2. This question was not asked in the 1992 survey; hence no comparison of practice is available. The degree of attrition of use by OM's such as ROM, MS and pain is significantly less than that reported by Turner *et al.* (1996, 1999) and Kirkness & Korner-Bitensky (2002), nevertheless one third of PT's would still not measure balance, for example at both admission and discharge.

**Table 3.2** SOM's used by Canadian PT's

Outcome measurement	Current use % total	At admission % total	At admission & discharge %total	More often % total
Range of motion	90	90	85	61
Manual muscle testing	88	92	85	68
Goal setting	73	95	85	57
Visual analogue scale pain	57	86	67	61
Berg Balance Scale	45	83	63	38

These results indicate that optimum use of outcome measurements may not be occurring in clinical practice. This is consistent with the results of the section of the survey on self-efficacy. The ability to carry out an activity with confidence may be quantified by quantifying self-efficacy. Respondents were asked to consider 12



statements and rate their confidence from 0% (not confident) to 100% (completely confident). Table 3.3 is a summary of the results reported.

Table 3.3 Confidence in the use of SOM's

Item	Staff PT Mean (sd)	Staff PT Range	PPL Mean (sd)	PPL Range
Knowing enough about test construction to develop own measure	27.8(22.2)	0-80	29.5(21.4)	0-80
Knowing how to link information to other information	45(24.3)	0-90	52(22.4)	10-90
Knowing how to compare scores to baseline levels across client groups	51.1(25)	0-100	54.8(22.3)	10-90
Knowing enough about measurement properties to choose	58.7(21)	10-90	64.5(20.4)	20-90
Knowing whether suitable measures are available	62.5(22.7)	10-100	73(21.7)	40-100
Knowing what to do with scores	62.9(19.7)	10-100	70(22.9)	10-90
Overall, knowing what to do with the information obtained	64(21.6)	10-100	68.5(23.9)	10-90
Knowing why to measure	71.5(20)	10-100	77.5(23.4)	10-100
Knowing how to track clients' progress with outcome measures	73(16.2)	20-100	73(22)	10-100
Knowing how to score measures	73.7(16.2)	20-100	75.5(18.8)	30-90
Knowing what to measure for client groups	73.8(15.6)	20-100	77.5(23.1)	0-100
Knowing how to administer OM in standardised manner	74.1(16.3)	20-100	75(18.2)	30-100
Total Confidence Score (higher score, greater confidence)	65.2(14.2)	10-100	68.5(25.9)	10-90

No statistically significant difference was noted between the groups for self-efficacy scores. The least amount of confidence was reported in the areas of measurement

properties, linking information, comparing scores across groups and overall, what to do with the results. It would appear that outcome measures are used with confidence, but PT's are less confident or familiar with their broader use in the context of overall evaluation.

### 3.3.2 Attitudes and barriers to the use of SOM

Similar barriers to using outcome measures were reported in both Canadian surveys and were

- Lack of time- reported by 52% staff PT's & 55% PPLs in 1998
- Lack of knowledge about measures- reported by 82% staff PT's & 75% PPLs in 1998
- Limited availability of measures- reported by 51% staff PT's & 50% PPLs in 1998
- Not meeting needs of clients- reported by 33% staff PT's & 60% PPLs in 1998
- Lack of professional consensus on what to use- reported by 27% staff PT's & 15% PPLs in 1998.

In addition to the quantitative survey results, the CPA undertook a series of focus groups to further examine the themes that had emerged following analysis of the survey (Huijbregts *et al.* 2002). The results of this qualitative research methodology were supported by the quantitative survey results. It was observed that while it was accepted that the use of client outcome measures had become intrinsic in physiotherapy practice, consistent application was not uniform; and the utilisation of the information gathered in a meaningful way lagged behind the collection of data. In terms of how practice is influenced, the authors noted that it is a combination of an organisational mandate to use standardised measures and the availability of an expert to advise on the outcomes evaluation that was required. To enhance practice in this area, the participants advocated that both employers and the professional organisation needed to be proactively supportive. Enhanced communication with the developers of measurement instruments was repeatedly mentioned as a way of optimising the subsequent use of newly developed outcome measures. Consistent with the survey results, limiting factors reported were insufficient knowledge, time constraints, lack of equipment, accessibility of forms and space e.g. for timed walks. Kirkness & Korner-Bitensky (2002), although their methodology differed from that of the previous authors, also noted that barriers such as limited knowledge of

instruments and their development; time; failure of the instruments to meet client needs; and lack of consensus on what to use were all reported by respondents. Russek *et al.* (1997) investigated the attitudes of PT's and OT's to standardised data collection and the characteristics of the therapists that inform these attitudes. Standardised evaluation forms designed for an orthopaedic patient population were provided to all clinics that agreed to participate. These included three principle forms (i) a data collection tool (DCT) for initial and subsequent re-evaluation, (ii) daily treatment forms and (iii) discharge status forms. Both videotapes and written instructions were provided for the measurements that formed part of the data collection tool. A number of measurements were included- for shoulder, elbow, knee, foot and ankle- but only those appropriate to each specific client needed to be used. Completed data sets were sent to a central site where the information could be entered on a computer. Information about attitudes and personal demographic data was gathered through a 67-item questionnaire. Principle-axis factor analysis (on 33 items) yielded 5 factors that accounted for 42.5% of the total variability, namely

1. Inconvenience associated with the completion of the data collection tool i.e. *initial and re-evaluation*
2. Acceptance of operational definitions- the instructions given for the various measurements that were performed as part of the evaluation
3. Automation- the usefulness of computers with respect to gathering and processing this type of information
4. Daily treatment forms- issues specifically surrounding the use and impact of the daily treatment forms
5. Training.

Two thirds of respondents to the questionnaire reported that they had 'studied thoroughly' the operational definitions for at least one body part included in the study. Nonetheless, reported participation rate for completing and submitting one complete data set was only 50%. The extent to which the latter activity occurred was significantly associated with those respondents who identified research as a professional goal. This was also significantly associated with factor 2 (acceptance of operational definitions). Increased numbers of data sets completed and submitted were both significantly associated with factors 2 and 5- comfort with operational definitions and adequate training. The inconvenience associated with the DCT's was significantly associated with greater client numbers per day. The authors conclude

that adequate training in operational definitions of measures and data collection will enhance participation in systematic use of measures but that this needs to occur in the context of optimising the method of data gathering, perhaps through the provision of automated patient reports. In the European survey, it was interesting to note that the degree to which collected data was processed statistically later varied from country to country with Ireland the lowest at 17% and Germany the highest at 90%. This may reflect the reason why the outcomes data is gathered; statistical processing of the data occurs most often in countries where the information is gathered for the purposes of quality management and informing providers and policy makers. Greater skill and knowledge of how gathered SOM data may be used at a later stage to inform practice would perhaps influence more consistent measurement practice.

### *3.3.3 The role of professional organisations in promoting the systematic use of SOM's*

Table 3.4 outlines the activities of a variety of international professional physiotherapy organisations in the context of encouraging and supporting the use of SOM. Three questions were asked of representatives of the organisations-

1. Does the professional organisation have a policy on the use of standardised and systematic outcome measurement and if so, is it contained in a larger strategy e.g. research and development, evidence-based practice?
2. What are the activities/services that the professional organisation has implemented/provided to enhance the use of OM in physiotherapy practice for individual members?
3. To what extent are OM's used in practice?

This information was obtained through direct correspondence and where appropriate journal publications, although for the most part, the latter applies to the Chartered Society of Physiotherapy (CSP) which is the representative body for physiotherapists within the United Kingdom of Great Britain and Northern Ireland and the Canadian Physiotherapy Association (CPA). In September 1999, the Council of the CSP adopted a strategy (CSP, 1999) with its focus on research and clinical effectiveness (CSP 1999). Prior to 1999, two separate strategy documents had been developed to focus on the areas- the Research Strategy (1995) and the Clinical Effectiveness Strategy (1997). As part of the achievements of the work incorporated in these

strategies, the one that focused specifically on SOM was the development of an outcome measures database (Hammond 1999). This provides information about the measurement properties of a large range of outcome measurements. A number of clinical guidelines have been developed and within those documents are suggestions for which standardised measurements should be used in practice e.g. osteoporosis and incontinence (CSP 1999, 2001). Two documents are designed specifically to provide information on SOM's, one deals with depression (CSP, 2002) and the other, low back pain related functional limitation (CSP, 2001). Additional national networking activities have been created with Clinical Interest Groups and these activities have been ongoing (Hammond 2000). To date no formal evaluation of the impact of these measures has been completed.

The Canadian Physiotherapy Association (CPA) has clearly described the process and progress their organisation has taken to enhance the systematic use of outcome measurements. Their definition of evidence-based practice as 'practice which has a theoretical body of knowledge, uses the best available scientific evidence in clinical decision making and uses standardised outcome measures to evaluate the care provided', is unique in that it includes an emphasis on the use of standardised outcome measures (Parker-Taillon 2002). The CPA EBP initiatives include activities in six key areas, outlined below.

1. Entry-level Curriculum Guidelines will be developed which promote the development of evidence-based practitioners.
2. Access to high quality, recognized continuing professional development that promotes and supports EBP will be available to Canadian PT's.
3. Clinical practice guidelines and strategies for their implementation, which will promote EBP and assist in clinical decision-making will be developed.
4. CPA Accreditation Programme will be revised to promote EBP and support such practice.
5. Outcome Measures. Canadian PT's are expected to consistently use standardised outcome measures in their daily practice.
6. A database will be developed which will examine the effectiveness & efficiency of Canadian PT services.

The European Region of the World Confederation for Physical Therapy adopted a series of Core Standards in May 2002 (ER-WCPT, 2002). Standard 6 reads as follows 'Taking account of the patient's problems, a published, standardised, valid,

reliable and responsive outcome measure is used to evaluate the change in the patient's health status.' The relevant criteria (6.1, 6.2, 6.6, 6.7) are listed below.

- Criteria 6.1 The physiotherapist selects an outcome measure that is relevant to the patient's problem.
- Criteria 6.2 The physiotherapist ensures the outcome measure is acceptable to the patient. The physiotherapist selects an outcome measure that he/she has the necessary skill and experience to use, administer and interpret.
- Criteria 6.6 The result of the measurement is recorded immediately
- Criteria 6.7 The same measure is used at the end episode of care.

The ER-WCPT adopted the Core Standards in 2002 and the expectation is that each of the member organisations will adopt the Standards for their own country.

### 3.4 Conclusion

It is accepted that the systematic use of standardised outcome measures is required in physiotherapy practice. This requires a commitment from individual practitioners, the organisations where they work and the professional physiotherapy organisations. The extent to which SOM's are used may be influenced by an organisational mandate, but there still appears to be a lack of understanding as to why SOM's are useful, this is reflected in the attrition that occurs after the initial measure is taken at baseline. To fully understand the relevance of outcomes data, the individual has to be equipped with the relevant knowledge and ability to interpret measurement properties. This is enhanced by publications such as 'Physical Rehabilitation Outcome Measures' (Finch *et al.* 2002) and databases of outcome measures. It appears from the literature review that the 1990's have been a period of significant change in practice, with increasing numbers of PT's using SOM's. The future of this practice in the coming years will be to optimise this practice with a view to enhancing individual clinical decision-making and broader outcomes research.

Table 3.4 The role of professional organisations in the promotion and use of SOM's.

<b>Does the professional organisation have a policy on the use of standardised and systematic outcome measurement and if so, is it contained in a larger strategy e.g. research and development, evidence-based practice?</b>	
New Zealand	The organisation does not have a policy on standardised outcome measures however a priority in their Strategic Plan for 2004 is the development of educational programmes on the use of clinical outcome measures. The main funder of physiotherapy services for private practitioners is the Accident Assurance Corporation (AAC). At present they do not require the use of standardised outcomes – they are interested in the number of visits and return to work.
Sweden	The association has a policy document concerning research and development and this includes the use of standardised test in clinical practice.
USA	The APTA does not have an explicit policy on the use of a standardised outcome measurement in research or in practice.
Canada	As part of the CPA EBP action list, the area of outcome measures is one of six key objectives, namely 'Canadian physiotherapists will consistently use standardised outcome measures in their daily practice'
UK	In the Research & Effectiveness Strategy (1999) The CSP pledges to 'improve access to and the use of clinical data and validated outcome measures to facilitate their use in practice.' (Objective 10).
Australia	The Australian Physiotherapy Association does not have any policy on the use of standardised and systematic outcome measurements.
Ireland	The ISCP has no formal policy but has adopted the ER-WCPT Core standards.
Germany	The ZVK (German PT Association) has no formal policy.
South Africa	There is no such policy in place in South Africa. At the moment they are trying to foster the use of SOM through courses and information in their newsletters
<b>What are the activities/services that the professional organisation has implemented/provided to enhance the use of OM in physiotherapy practice for individual members?</b>	
New Zealand	The organisation has a database of commonly used clinical outcome measures with their supporting articles. This has been published in their monthly newsletter and is available on the members only section of the website. Copies of the outcome measure and supporting articles are sent out to members when requested. Regular articles are written in the newsletter on the outcome measures for a particular site or condition to try and raise the awareness of their use and the different types of parameters that can be measured

Sweden	A review of available outcome measures was completed, with a view to simplifying the choice and use of measurement. The result is 150 measures presented on the Association's website <a href="http://www.lsr.se">www.lsr.se</a> (in Swedish). It is available to members only. Out of these about 50 have a more thorough description; using a format similar to that of the CPA's book (Finch <i>et al.</i> 2002). These measures should have been investigated for measurement properties. When it is allowed, according to copyright rules, the manual and protocol are provided as pdf-files.
USA	APTA has developed an instrument that assesses function at the level of actions from a patient's perspective. The instrument collects data using a five point Likert-type scale that assesses patients' perceptions of difficulty in and confidence about performing a series of actions describing movement. Dubbed OPTIMAL (Outpatient Physical Therapy Improvement in Movement Assessment Log), the instrument is undergoing analyses defining its psychometric properties.
UK	<ul style="list-style-type: none"> <li>▪ The CSP web site also contains information on the work of the Society with regard to outcome measurement</li> <li>▪ An editorial was written to promote the CSP outcome measures database (Hammond 2000)</li> <li>▪ An information paper on outcome measures has been published: Outcome measures</li> <li>▪ Various specialty specific papers or chapters on outcome measures <ul style="list-style-type: none"> <li>○ Measures for depression</li> <li>○ Low back pain-related functional measures</li> <li>○ Chapter in the Falls audit pack</li> <li>○ Chapter in Osteoporosis audit pack.</li> <li>○ Chapter in Incontinence audit pack.</li> <li>○ A series of workshops was held in 2000 around the UK</li> <li>○ The Outcome measures network, as part of the pilot interactive CSP website.</li> </ul> </li> </ul>
Canada	<ul style="list-style-type: none"> <li>▪ National workshops</li> <li>▪ Teleconferences on outcome measures</li> <li>▪ Two editions of Physical Rehabilitation Outcome Measures <ul style="list-style-type: none"> <li>○ Cole <i>et al.</i> (1994)</li> <li>○ Finch <i>et al.</i> (2002)</li> </ul> </li> </ul>
Australia	No activities described
Ireland	No activities described
Germany	<ul style="list-style-type: none"> <li>▪ Text book on measurement- Qualitätsmanagement in physiotherapeutischen Einrichtungen and Standardisierte</li> </ul>



	<p>Messungen und Tests - DGQ Band 31-01.</p> <ul style="list-style-type: none"> <li>▪ For individual members, the German physiotherapy association offers special courses of training <ul style="list-style-type: none"> <li>○ Standardisierte Messverfahren" through our PhysioAkademie (<a href="http://www.physio-akademie.de">www.physio-akademie.de</a>)</li> <li>○ Qualitätsmanagement in physiotherapeutischen Einrichtungen - one course lasting five weekends</li> </ul> </li> </ul>
South Africa	<ul style="list-style-type: none"> <li>▪ Developing evidence-based clinical guidelines for several conditions by doing literature reviews and combinations of existing guidelines- applying it to South African situations and within these, advocating validated outcome measures.</li> <li>▪ Development of an intervention strategy for chronic low back pain and an outcome measure for the medical insurers to monitor whether physiotherapy is worthwhile paying for before surgery is performed.</li> </ul>
<b>To what extent are OM's used in practice in various countries?</b>	
New Zealand	No information to date
Canada	Mayo <i>et al.</i> (1993), Kay <i>et al.</i> (2001), Huijbregts <i>et al.</i> (2002), Kirkness & Korner-Bitensky (2002)
UK	Chesson <i>et al.</i> (1996) Turner <i>et al.</i> (1996) Turner-Stokes & Turner-Stokes (1997) Turner <i>et al.</i> (1999)
South Africa	No information to date
Australia	No information to date
Germany	No information to date
Sweden	No information to date
USA	No information to date
Ireland	Stokes & O'Neill (1999) Stokes (2004, in preparation)

## CHAPTER 4      METHODOLOGY

### 4.0 Introduction

This chapter describes the methodology employed in this research. The methodology contains two distinct strands, namely,

1. a survey of the use of systematic outcome measurement in physiotherapy practice in Ireland and
2. the development of a new measurement for use by physiotherapists working with older people. The actual methodologies overlap since some of the experimental data gathered in the survey informs part of the methodology for developing a new measurement. Nevertheless, for clarity, the methods employed will be described separately.

### 4.1 Survey 1998

Two surveys were performed to consider the practice of physiotherapists working with older people in Ireland in the context of their use of outcome measurement. The first survey took place in 1998.

#### 4.1.1 Aim & objectives

The aim of the first survey was to examine the extent to which PT's working in the area of rehabilitation of older people personally used standardised outcome measurements in their practice. More specifically, the objectives were-

1. To identify if generic standardised measurement instruments or non-standardised methods were used in their assessment of mobility/transfers, balance, gait disorders, and exercise tolerance. In addition, to identify what disease specific outcome measures, if any were used in their assessment of stroke and Parkinson's disease.
2. To investigate the level of satisfaction with the instruments utilised in their practice.

The survey also contained questions about the ideal content of physiotherapy assessments and the function of measurement instruments. It was expected that the latter two aspects could be used to inform the initial stages of the content validity methodology of the new outcome measurement and they will be discussed in detail below.

## 4.1.2 Methodology

### *4.1.2 (i) Operational definition of standardised outcome measures*

The term standardised assessment or outcome measure suggests that the scale has been fully evaluated and the properties of validity, reliability and sensitivity to change are established. For many measures used by PT's, these measurement or psychometric properties have not been fully established (Cole *et al.* 1995). For the purposes of this survey, the operational definition employed was 'a published measurement tool', designed for the purpose in a given population, with detailed instructions provided as to when and how it is to be administered and scored, interpretation of the scores and results of investigations into reliability and validity (Cole *et al.* 1995). This is the definition used by the Canadian Physiotherapy Association in their initial survey of practice in 1992 and by other authors both at the time of the first survey and in later publications (Chesson 1996, Kay *et al.* 2001, Huijbregts *et al.* 2002, Kirkness & Korner-Bitensky 2002). By employing the same definition, comparison could be made with those studies.

### *4.1.2 (ii) Participants*

As part of their routine work in the health care system in Ireland, many PT's work with older people. This is because people over the age of 65 years, who make up 11% of the population, account for 25% of admissions to hospital and 40% of bed days (Brenner & Shelley, 1998). With a view to obtaining a representative picture of the activities of physiotherapists in Ireland, all Departments of Medicine for the Elderly (n=25), at the time (1998), were approached to participate. The Senior Physiotherapist in each department was invited to participate in the survey and report on her/his individual practice.

### *4.1.2 (iii) Procedure*

The survey took the form of a structured interview, with the author travelling to each participating centre. Each interview took 40-60 minutes to complete, the interview forms were completed by the author. The interview comprised four questions to be answered for each of the following areas,

1. Mobility and transfers
2. Gait disorders
3. Stroke

4. Parkinson's disease/parkinsonism
5. Exercise tolerance
6. Balance disorders.

It was considered that these domains covered some of the specific pathologies encountered in the rehabilitation of older people and also aspects of physiotherapy that crossed a range of pathologies. The data-gathering instrument is contained in Appendix 2. The first question was a closed question and asked whether the participant used standardised or non-standardised measurement. If a standardised measurement was employed, the participant was asked to name the measurement. The participant was asked to outline their frustrations with their chosen mode of measurement, if indeed any existed. The two remaining questions related to the face and content validity data and will be discussed in section 4.3 below.

#### **4.1.3 Analysis of data**

The data from this part of the survey was analysed descriptively.

### **4.2 Survey 2003**

#### **4.2.1 Aim**

The aim of the second survey, which took place in 2003, was to re-examine the degree to which systematic outcome measurement is used in the practice of physiotherapy in rehabilitation of older people in Ireland, with a view to establishing any changes in practice that may have occurred in the 5 years since the 1998 survey. The nature of practice was also considered, as well as barriers to the use of SOM. The level of confidence of physiotherapists was also investigated.

#### **4.2.2 Objectives**

The specific objectives were

1. To consider what outcome measurements were used in the rehabilitation of older people and to compare this to 1998.
2. To examine at what stages during the rehabilitation process outcome measures were employed.
3. To investigate the level of confidence participants display with the use of systematic outcome measurement in practice

It was anticipated that the output of this survey would add to the small body of published international data on the practice of outcome measurement, with a view to identifying how this might be optimised. Secondly, it would generate information about the level of confidence of PT's have in using outcome measurement with a view to informing the future dissemination of the new SOM.

#### 4.2.3 Participants

The participating centres and where relevant individuals in the 1998 survey were invited to participate. Due to changes within the structure of Departments of Medicine for the Elderly, two centres were no longer in a position to participate since due to reorganisation, they did not provide the services. Twenty centres were surveyed as part of this review of practice. Forms were returned anonymously, hence individual data from 1998 and 2003 could not be matched.

#### 4.2.4 Procedure

The survey instrument (Appendix 3) contained a definition of an 'outcome measurement' and, for each of the areas included in the 1998 survey, posed the question 'do you use a published outcome measurement or one developed by your department?' The second question asked the participant to identify what standardised outcome measures they used in practice for each of the areas. Thereafter a question was asked about the frequency of use of the specific standardised outcome measures. The final section considered the confidence of the participants in systematically using outcome measurement. Self-efficacy is a method for assessing the self-reported level of confidence in performing a task or activity. Each participant was asked to rate their confidence from 0% (not at all confident) to 100% (totally confident) for twelve statements, which were also employed by the Canadian Physiotherapy Association (CPA) in their 1998 survey (Kay *et al.* 2001). The statements were as follows

1. Knowing **whether** suitable measures are available
2. Knowing enough about reliability and validity **to choose** the best measure
3. Knowing about **scale construction** to develop my own instrument
4. Knowing **what to measure** for my client group(s)
5. Knowing **how to administer** measures in a standardised manner
6. Knowing **how to score** measures

7. Knowing **what to do** with the scores
8. Knowing how to track clients' progress
9. Knowing **how to link information** from outcome measures with other sources of information e.g. client characteristics
10. Knowing **how to compare scores** to baseline levels across clients
11. Overall, have a clear purpose for measurement- **knowing why** to measure
12. Overall, knowing **what to do** with the information obtained.

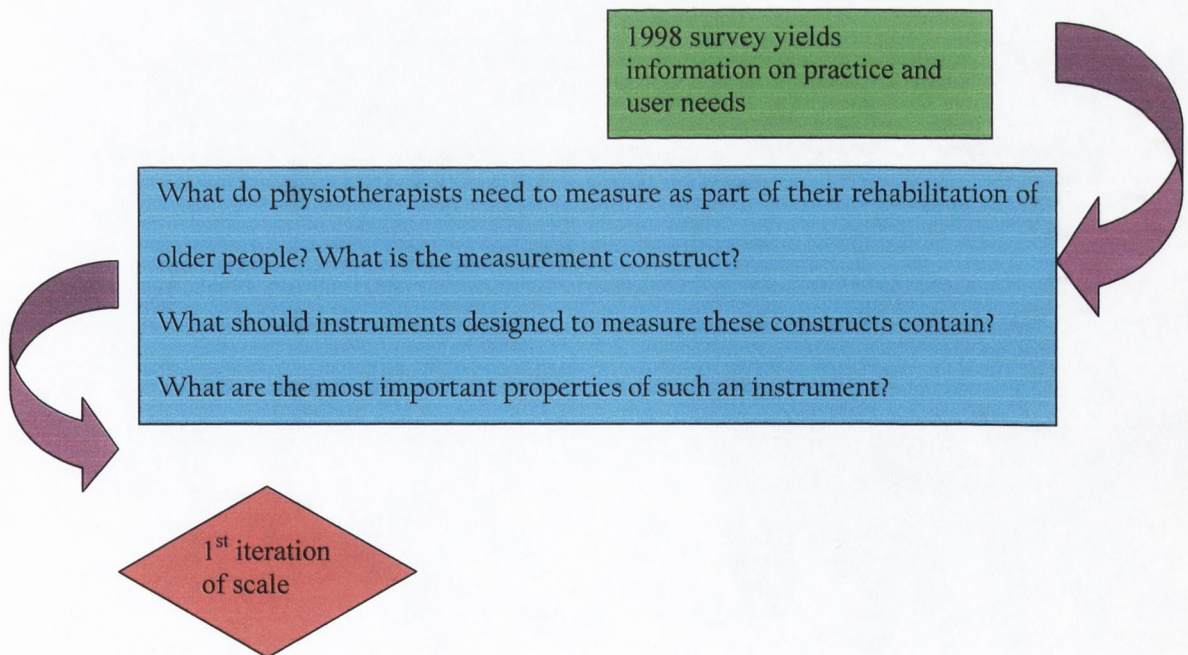
The Canadian study (2001) addressed the confidence of both staff physiotherapists and clinical specialists and reported the results for both separately. By investigating the level of confidence of Irish Senior PT's, using the same questions, direct comparison could be made.

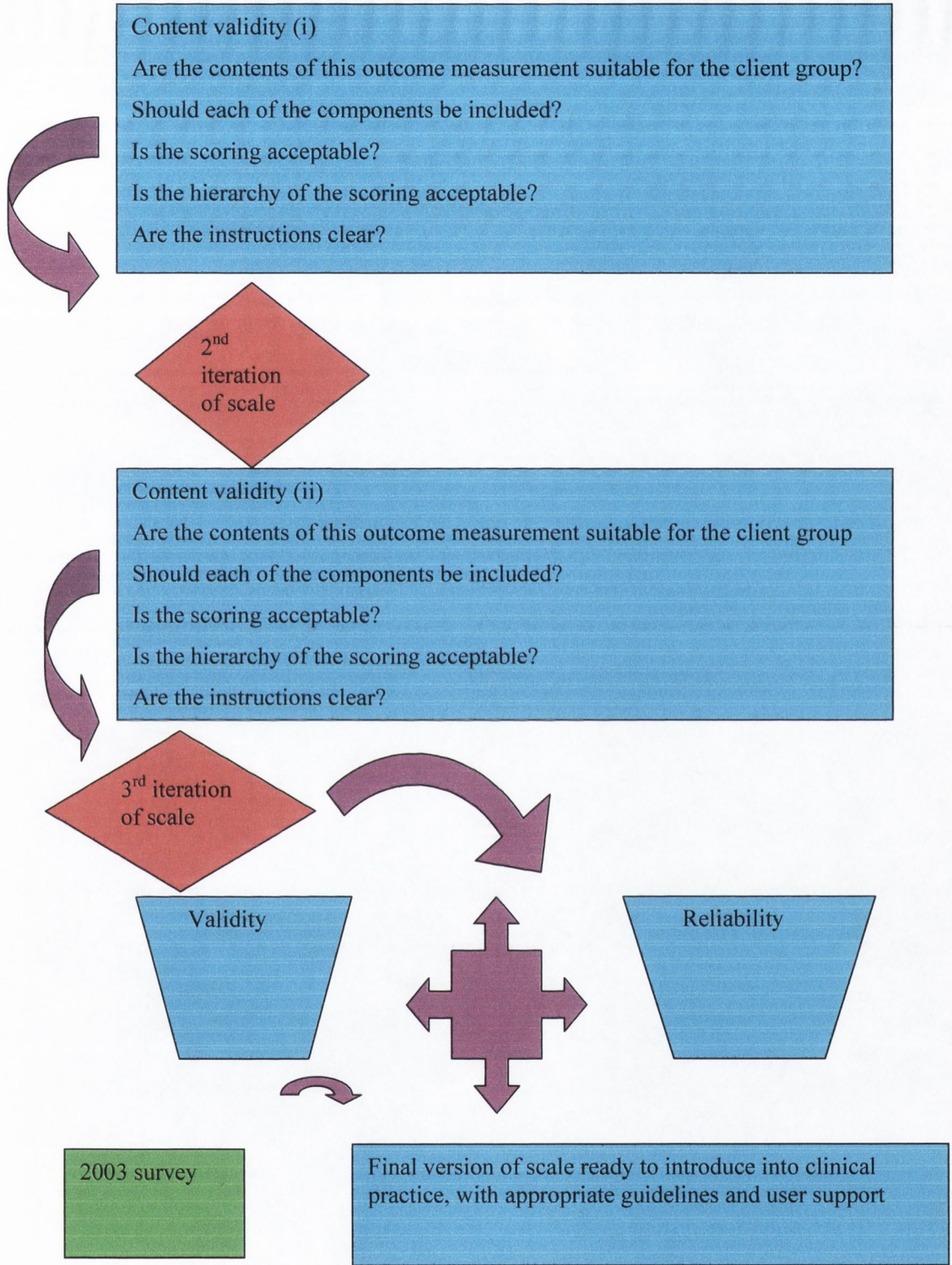
#### 4.3 Development and evaluation of a new instrument

In creating a new measurement instrument, its properties of validity and reliability must be examined. Figure 4.1 outlines the process employed to consider these properties during the course of this work and places the surveys in context.

Figure 4.1

The development of a new instrument and the placement of the surveys to inform this process.





The remainder of the chapter describes the methods utilised to develop and evaluate the new instrument.

#### 4.3.1 Establishing an operational definition for the new instrument

Four sources were used to inform an operational definition for the new instrument.

1. The description of physical therapy adopted at the World Confederation for Physical Therapy (WCPT) General Meeting in 1999 (Appendix 4), which states that physiotherapy is about working with people to maintain and restore maximum movement and functional ability throughout the life span. It is concerned with identifying and maximising movement potential, within the spheres of promotion, prevention, treatment and rehabilitation (WCPT 1999).
2. The International Classification of Impairments, Disability and Handicap (ICIDH) (WHO 1980), now the International Classification of Functioning (ICF) (WHO 2001). Wade (1992) notes that to measure outcome, it should always be considered at the level of disability. The domain of the ICF that corresponds with 'disability' in ICIDH is limitations of activities i.e. the difficulties an individual may experience in executing activities. This instrument considers common activities that contribute to mobility and stability.
3. Information generated as part of question 4, 1998 survey.
4. A practice review to identify the enunciate the barriers to achieving functional independence in the context of physiotherapy.

Question 4 in the 1998 survey asked participants to identify, within each of the domains of mobility/transfer and stability, the **function** of an outcome measure. Feinstein (1987) identified a number of functions for clinical indices, these include, description of a state, measure of change, estimation of prognosis, offering a guideline. The full list offered to the participating PT's included-

- Provide a problem list
- Identify a state
- Denote a change due to intervention
- Predict an outcome
- Offer a guideline



- Teaching tool
- Communication aid with other services.

The participants were asked to rank these in order of importance.

*4.3.1 (i) A practice review to identify the barriers older people experience to achieving functional independence in the context of physiotherapy.*

*4.3.1 (ia) Aim & objectives*

Older people often present for rehabilitation with multiple pathologies, which may in turn act as barriers that hinder the attainment of functional independence during physiotherapy. This practice review was carried out to address the following questions-

1. In general, what is the degree to which each of these variables contributes to difficulties in achieving functional independence? and
2. What is the relative impact of these barriers?

*4.3.1 (ib) Methodology*

This was a cross sectional survey of older patients (aged 65 years or older) attending the physiotherapy departments in St. James's Hospital and the Adelaide & Meath Hospital, in Dublin. Five physiotherapists participated in completing anonymously a questionnaire (Appendix 5) on the patients in their care during a six-week period in 1999. The questionnaire contained eight questions. Seven questions were scored using a Likert Scale, with anchors of 'Not at all' and 'Hinders intervention'. For each of the following potential variables- cognition, pain, mood, communication, problems with balance/stability, problems with mobility and poor fitness/deconditioning- the therapists were asked to score the impact of the variable as a barrier to the patient achieving functional independence. These barriers were chosen because in clinical practice they represent many of the potential contributory factors to poor function but also those which can inhibit participation in physiotherapy intervention and treatment. Space was provided for additional barriers that were not listed in the data-gathering instrument. The final question asked the therapist to rank the relative contribution of each variable that acted as a barrier. Questionnaires were completed on 133 patients.

### 4.3.1 (ic) Results

The results, which will be discussed in Chapter 5, confirmed that problems in mobility and stability are the most common sources of difficulty in achieving functional independence that PT's have to consider for their patients/clients.

Using this information, along with the results from question 4 of the 1998 survey which suggests that outcome measures should be able to capture change and have detailed information that could contribute to the development of a problem list, item selection was then considered.

### 4.3.2 Item selection

#### 4.3.2.1 Face validity

Face validity, as discussed in Chapter 3, considers whether on the face of it, the instrument measures what it sets out to measure. As part of the process of generating the contents of the new instrument, question 3 in the 1998 survey asked 22 Senior Physiotherapists what they considered should be included in the outcome measures for mobility problems, balance problems, gait disorders, stroke, Parkinson's disease/parkinsonism and problems with exercise tolerance. All answers were documented and the themes extrapolated after all the interviews were completed. These answers formed the basis of the items included in the new instrument. Further analysis of the sensibility of the items took place as part of the evaluation of content validity and is discussed below.

#### 4.3.2.2 Scaling

The final version of the scale for analysis in this project has eight subsections, each of which has a variety of descriptors. Numbers were assigned to each descriptor category within the scale. This was done in a hierarchical manner for each of the categories within the scale. This method is useful in practice since it allows the evaluator to discontinue testing once a component cannot be completed. This applies to the following sections-

- Bed rise 0-6
- Sit to stand 0-6
- Standing ability & stability 0-13
- Reach & lift 0-6
- Bend & reach 0-6

- Repeated sit to stand                    0-8
- Gait ability                                    0-12

It was not possible to do this for the section on gait analysis (Gait pattern 0-40).

The scoring is such that a high score indicates better performance. In the first iteration of the instrument, the scoring was reversed but this was changed as a result of feedback from the therapists who reviewed this iteration of the scale. Although the output of this scale could be considered interval data, it is in fact ordinal. It follows the standards for the use of ordinal scales in clinical trials outlined by MacKenzie & Charlson (1986), namely that scales

- Must be composed of individual elements that are clearly defined do not overlap
- Should have ranks that are defined and with a 'reasonable order'
- Have a 'sufficient range to encompass the spectrum of the phenomenon in the population' to be able to measure both improvement and dis-improvement equally.

The upper end of each section employs normative data from a variety of published sources (Reuben & Siu 1990, Alexander *et al* 1992, Bassey *et al* 1992, Bohannon *et al* 1996, Bofelli *et al* 1996, Bohannon 1997). Whittle (2001) was used as a reference for normal and pathological gait parameters.

#### 4.3.3 Reliability

The reliability of an outcome measurement refers to the extent to which a measurement contains random errors (Helmsdater 1967); minimising the potential for error within an instrument will lead to reproducibility, consistency and repeated measures that demonstrate change which is, in fact, true change and not error. **Inter-rater** reliability refers to the level of agreement between two or more raters using the same outcome measure on the same subject at the same time. This considers inter-observer agreement. **Intra-rater** reliability describes the level of agreement between scores obtained on a measurement of a subject, taken on two separate occasions by the same rater, and where, for the purposes of analysis, no actual change is anticipated to have occurred within the subject. It considers the variation that occurs within an observer as a result of multiple exposures to the same subject (Streiner and Norman 1995). A common way of evaluating the

internal consistency of an instrument is to estimate the correlations that exist between all of the possible pairs of items within an instrument and within individual item and total scores. This lends support to the concept of inter-rater reliability (McDowell & Newell 1996).

#### *4.3.3.1 Inter-rater reliability*

##### *4.3.3.1(i) Subjects*

Fifteen older people who were in-patients in St. James's Hospital participated in this study. The inclusion criteria were that the participant was 65 years of age or older and able to give informed consent to take part in the study. If an individual had a cognitive impairment (Mini mental state examination (MMSE) (Folstein *et al.*, 1975) <24/30, Abbreviated mental test score (AMTS) <7/10) or a documented receptive communication disorder, they were excluded from participation.

##### *4.3.3.1(ii) Raters*

Three raters participated in this evaluation. They were chosen to reflect that range of PT's who will in future use the new OM i.e. both novice and experienced practitioners.

##### *4.3.3.1(iii) Procedures*

A PT not involved in the reliability study but familiar with the content of the new outcome measure (OM) was requested to identify a sample of suitable subjects. The nature of the study was explained to the participants and informed consent was obtained.

##### *4.3.3.1(iv) Training*

The PT participants received a short training session in the use of the new OM. Scoring was explained and any queries raised by the raters were clarified. The instructions for the patient outlined in the new OM were considered. The equipment required was checked and one trial of the OM was performed by a rater on another rater.

#### 4.3.3.1(v) *Testing*

The experienced PT performed an assessment using the new OM on each participant. The other raters viewed this concurrently. Each rater was asked to score the participants performance independently from the other raters. This was the case for all 15 older participants.

#### 4.3.3.2 *Intra-rater reliability*

##### 4.3.3.2(i) *Subjects*

Twenty older people who were in-patients in St. James's Hospital participated in this study. The inclusion criteria were that the participant was 65 years of age or older and able to give informed consent to take part in the study. If an individual had a cognitive impairment (MMSE <24/30, AMTS <7/10) or a documented receptive communication disorder, they were excluded from participation.

##### 4.3.3.2(ii) *Raters*

A qualified physiotherapist performed the two measurements on each participant.

##### 4.3.3.2(iii) *Procedures*

A PT not involved in the reliability study but familiar with the content of the new OM was requested to identify a sample of suitable subjects. The nature of the study was explained to the participants and informed consent was obtained.

##### 4.3.3.2(iv) *Training*

The PT participant received a short training session in the use of the new OM. Scoring was explained and any queries raised by the rater were clarified. The instructions for the patient outlined in the new OM were considered.

#### 4.3.4.2(v) *Testing*

The rater performed a measurement using the new OM on each participant on two occasions. The second measurement session took place within 48 hours of the first, this ensured that no actual change had occurred in the status of the participants' status. The rater was blind to the score at the first measurement time, when they repeated the measurement.

#### 4.3.3.3 Analysis of inter- and intra-rater reliability

The percentage agreement between raters was calculated, as well as kappa ( $k$ ) and weighted kappa coefficients (Cohen, 1968). The Bland and Altman (1986) method was also employed to consider at the level of agreement between two raters. Cronbach's alpha was calculated to evaluate the internal consistency of the measurement instrument. Excel, SPSS (Version 12.0.1) and MedCalc (Version 7.6.0.0) were the software packages utilised in the analysis.

#### 4.3.4 Validity

The establishment of the validity of a new SOM is a multifarious process. Validity is not simply a binary state i.e. a measure is valid or not, but a continuum. A number of different methodologies were employed within this study to consider the validity of the new SOM. Some of the methodologies overlap with the survey and with each other but they are listed separately for clarity.

##### 4.3.4 (i) Content validity

Content validity examines the content of the measurement. In the context of this study, it considers whether the sample of all potential items that could be included is an acceptable sample, whether the structure of each domain is sensible, if the scoring is acceptable and if the instructions were useful. After the first iteration of the scale was developed (Appendix 6), it was circulated to eight Senior Physiotherapists working in the rehabilitation of older people. The PT's reflected the areas where the new SOM would be used in the future – in-patient rehabilitation, out-patient rehabilitation and research. The participants were asked to use the new SOM for two weeks with a self-selected sample of patients and thereafter, to answer a series of questions for each item. Three questions were asked, as follows-

- Rate the importance of this item 0-5 (not at all important-very important)
- Is the scoring acceptable?
- Is the hierarchy of the scoring acceptable?

There was an invitation for additional comments on each item (Appendix 7). Each PT was interviewed individually and after the feedback from the first evaluation of the content validity, the second iteration of the scale was developed (Appendix 8). It was circulated to 13 Senior Physiotherapists, who were asked specific questions

about each item (Appendix 8). Feedback from this exercise yielded the third and final iteration of the new SOM and the remainder of the work on validity was performed on this final version (Appendix 9).

#### 4.3.4 (ii) *Construct and criterion validity*

Construct validity has been described by McDowell and Newell (1996) as a series of activities that are 'part science and largely art'. It requires that a series of constructs been established about the new SOM and subsequently examined. Criterion validity relates to how the newly developed instrument relates to a 'gold standard'. In rehabilitation existing rehabilitation measures are employed, since no 'gold standard' measurement exists. Criterion validity has a temporal component - it can be either concurrent or predictive. The former relates to how the new instrument compares to the *gold standard* at a given point in time and the latter refers to its relationship to an assessment in the future.

For the purposes of this study, concurrent criterion validity was examined by comparing scores on the new instrument with scores on two rehabilitation measures- the Elderly Mobility Scale (EMS) (Smith, 1994) and the Performance Oriented Assessment of Mobility (POAM) (Tinetti, 1986) (Appendix 10). These two instruments were chosen because the results of the first survey of practice revealed that these two scales were the most commonly used by physiotherapists in Ireland. The value of comparing the new SOM with these existing measures was to identify that it had a strong relationship with those instruments that formed part of current, widespread practice. Both concurrent and predictive criterion validity was evaluated. The comparison of the new SOM with the EMS and POAM also formed part of the investigation of construct validity. A number of specific constructs were established for consideration:-

- That the new SOM would have a strong association with two scales of mobility and balance, most commonly used in Ireland at the time of the first survey 1998.
- That the new SOM would be able to discriminate between two groups of older people, namely those having rehabilitation and healthy, community dwelling older people.
- That the scale would contain one factor- that would be explained by the individual items.

Seventy-five older people who were undergoing rehabilitation were recruited for inclusion in the validity studies. Eighteen community dwelling older people were included. People were excluded if they would have been unable to give informed consent due to cognitive (MTS <24/30) or communication problems (documented diagnosis of a receptive dysphasia). All of the older people who were hospital patients were referred for physiotherapy. All of the community dwelling older people were recruited from Active Retirement Groups in Dublin or personal contact. Each participant was given an information leaflet and signed a consent form.

Following their agreement to participate, each subject was measured using the new SOM, the EMS and the POAM. This measurement occurred once only for the community dwelling older people and for a sample of the hospital group. Thirty-seven of the hospital group had this measurement repeated at the time of their discharge.

Spearman rank correlation co-efficients were used to investigate the level of association between the new measure and the EMS and POAM. An appropriate non-parametric test - the Mann Whitney U test- was used to examine the difference between the hospital and community groups.

#### 4.3.4 (iii) *Responsiveness*

Thirty-seven of the hospital group, outlined above, were measured using the new SOM, the POAM and the EMS at admission and discharge. The responsiveness of the data was considered by evaluation of the effect size and the relative responsiveness of the three scales. Relative efficiency and effect size have been used to consider a number of rehabilitation measurement instruments (Wright *et al.* 1998, Stokes *et al.* 2003). The effect size provides a ratio of the 'signal to noise' where the 'signal' is the change that occurs due to intervention and the 'noise' is the level of variability within the instrument. It uses the change in scores between admission and discharge and the standard deviation. For the purposes of this study, effect sizes of less than 0.2 would be interpreted as small and 0.5 as moderate. Anything greater than 0.5 would be acceptable.

Relative efficiency (Liang *et al.* 1985) is a method of comparing two scales to ascertain which is more responsive. Relative efficiency is calculated using the following formula-



$$\text{Relative efficient (X versus Y)} = (\hat{X}/\hat{Y})^2$$

Where  $X$  is either the EMS or POAM and  $Y$  is the new SOM and  $z$  is obtained by comparing the admission and discharge scores. A relative efficiency score of 1 would indicate that  $X$  was equally as responsive as  $Y$ , a score of  $>1$  would suggest that  $X$  was more responsive than  $Y$  and  $<1$  that  $Y$  was less responsive than  $X$ . The new scale should be at least as responsive as those used in existing practice.

Excel, SPSS (Version 12.0.1) and MedCalc (Version 7.6.0.0) were the software packages utilised in the analysis.

#### 4.4 Ethical approval

Ethical approval for all aspects of the study involving older people was obtained from the Joint Ethics Committee of St. James's Hospital and the Adelaide & Meath Hospital, incorporating the National Children's Hospital. This letter is copied in Appendix 11.

## CHAPTER 5 RESULTS

### 5.0 Introduction

This chapter describes the results of the various components of this research. It follows the same structure as the methodology i.e. two distinct strands, namely,

1. a survey of the use of systematic outcome measurement in physiotherapy practice in Ireland and
2. the development of a new measurement for use by physiotherapists working with older people.

As described in the previous chapter, the actual methodologies overlap since some of the experimental data gathered in the survey informs part of the methodology for developing a new measurement. Nevertheless for clarity, the results will be described separately.

Two surveys were performed to consider the practice of physiotherapists working with older people in Ireland in the context of their use of outcome measurement.

#### 5.1 Survey 1998

The first survey took place in 1998. Its aims were to investigate the use of outcome measurement

##### 5.1.1 Profile of participants

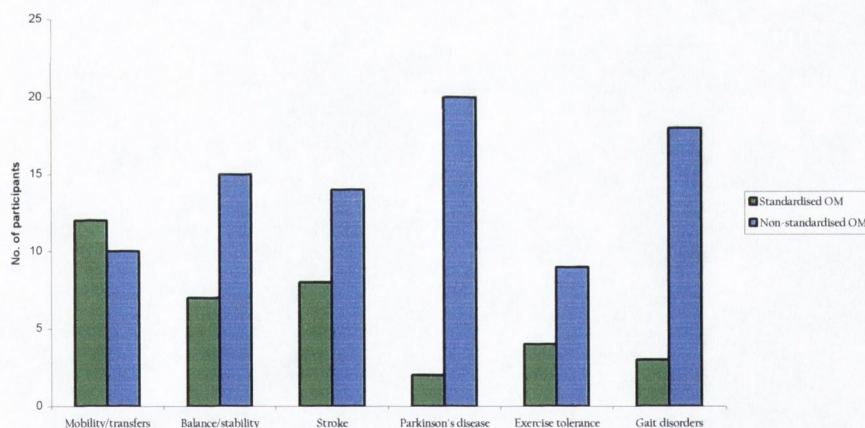
With a view to obtaining a representative picture of the activities of physiotherapists in Ireland, all Departments of Medicine for the Elderly (n=25), at the time (1998), were approached to participate. The Senior Physiotherapist in each department was invited to participate in the survey. 22 agreed to participate, representing a response rate of 88%.

##### 5.1.2 Use of standardised outcome measures

One of the objectives of the first survey was to examine the degree to which PT's working in the area of rehabilitation of older people used standardised outcome measures (SOM) versus non-SOM in their assessment of mobility/transfers, balance, stroke, Parkinson's disease/parkinsonism and gait disorders and used standardised outcome measurements in their practice. Figure 5.1 below illustrates this reported practice. Of the 22 PT's surveyed, 12 used standardised outcome

measures for mobility and transfers. Ten of the twelve used the Elderly Mobility Scale (Smith, 1994) and two used other scales; the Communication Chart (n=1) designed by Finlay (1994) and the 'Get Up and Go' test (n=1) (Mathias et al, 1986). Seven of the 22 PT's surveyed used SOM to measure balance performance. Five used the balance sub-scale of the Performance Oriented Assessment of Mobility (Tinetti, 1986). Two physiotherapists (PT's) used Functional Reach (Duncan et al, 1990). For Parkinson's disease/parkinsonism, only 2 respondents reported using SOM – the Webster Scale (Webster, 1967) and Therapy Outcome Measures (TOM) (Enderby and Kew, 1995). Two other PT's reported using the Hoehn and Yahr Staging to monitor the status of the disease. In the case of patients with stroke, 8/22 PT's used SOM: five used the Motor Assessment Scale (MAS) (Carr et al, 1995), one used TOM and two PT's employed the Motor Club Assessment (MCA) (Ashburn, 1982). For measurement of gait disorders, only 3 of the 21 PT's who reported assessing gait disorders used a standardised instrument- the gait subscale of the Performance Oriented Assessment of Mobility (Tinetti, 1986) and for exercise tolerance of the 13 PT's who assessed it, four used standardised methods.

Figure 5.1 Use of SOM versus non SOM



### 5.1.2 Level of satisfaction

The second objective was to investigate the level of satisfaction with the instruments they utilised in their practice. Figures 5.2 and 5.3 outline the frequency of common themes reported by users of SOM and non-SOM. The question was asked for the 4 domains of physiotherapy- mobility, balance, exercise tolerance and gait as well as the two diagnoses- Parkinson's disease and stroke, hence multiple answers are reported. Table 5.1 reports these results with additional detail.

Figure 5.2 Frequency of common themes reported by users of non-SOM

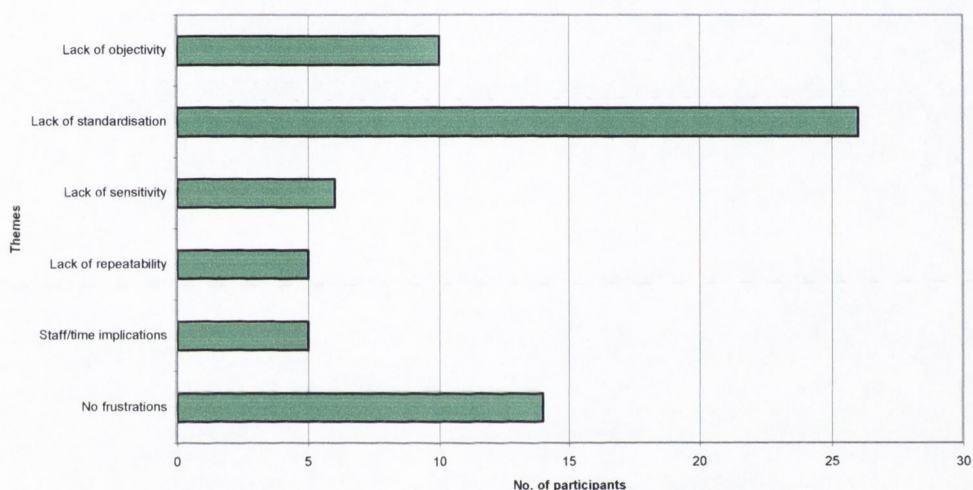


Figure 5.3 Frequency of common themes reported by users of SOM

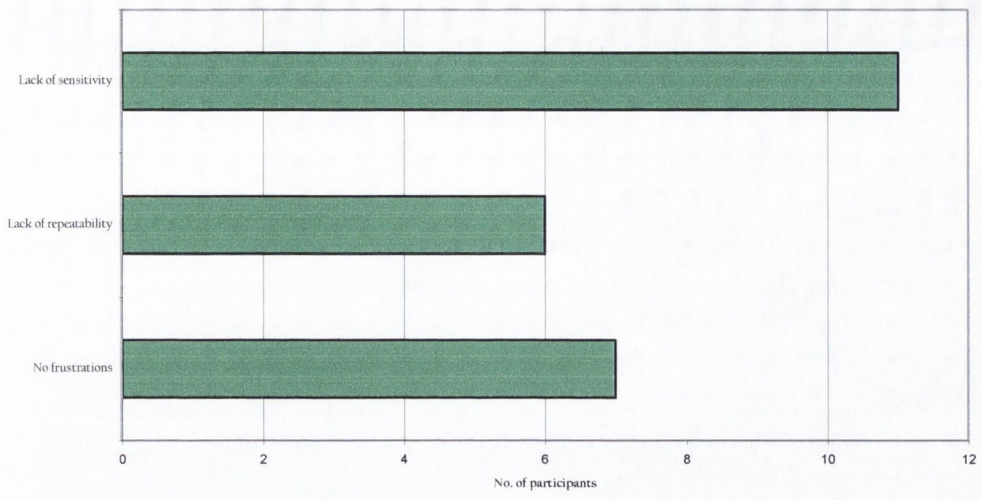


Table 5.1 Frustrations with current methods of measurement

Domain	Non-SOM	SOM
Mobility	Lack of standardisation (n=2)	EMS issues with sensitivity (n=5), floor & ceiling effect (n=2)
	Lack of sensitivity to change (n=2)	EMS lack of detail (n=3)
	Problems with repeatability (n=2)	EMS lack of standardised instructions (n=2)
	Staff and time issues (n=2) No frustrations (n=2)	No frustrations (n=3)
Balance	Lack of objectivity (n=8)	FR lack of sensitivity (n=2)
	Lack of standardisation (n=4)	No frustrations (n=4)
	No frustrations (n=1)	
Stroke	Lack of repeatability (n=2)	MAS upper extremity and mobility components presented some difficulty with scoring & hierarchy (n=4)
	Lack of standardisation (n=2)	MCA insufficiently sensitive (n=1)
	Length of time (n=5) No frustrations (n=3)	
Parkinson's disease/parkinsonism	Lack of standardisation (n=6)	Webster Scale insufficiently sensitive (n=1)
	Lack of repeatability (n=2) No frustrations (n=8)	
Gait disorders	Lack of sensitivity (n=4)	10m walk not sensitive (n=1)
	Lack of objectivity (n=2)	Shuttle test lack details (n=1) and time consuming (n=1)
	Lack of standardisation (n=12)	No frustrations (n=1)

Exercise tolerance	Lack of standardisation (n=4)	
	Time as an issue (n=1)	
	Lack of objectivity (n=1)	
	No reported frustrations (n=5)	

## 5.2 Survey 2003

The aim of the second survey, which took place in 2003, was to re-examine the degree to which systematic outcome measurement is used in the practice of physiotherapy in rehabilitation of older people in Ireland, with a view to establishing any changes in practice that may have occurred in the previous 5 years.

The specific objectives were

1. To consider what outcome measurements were used in the rehabilitation of older people.
2. To examine at what stages during the rehabilitation process outcome measures were employed.

To investigate the level of confidence participants display with the use of systematic outcome measurement in practice

### 5.2.1 Participants

The participants in the 1998 survey were invited to participate. Due to changes within the structure of Departments of Medicine for the Elderly, 2 centres were no longer in a position to participate since they no longer provided the services. 20 centres were surveyed as part of this review of practice, 15 responded representing a 75% response rate.

### 5.2.2 Use of SOM

Figure 5.4 illustrates the difference in the reported use of SOM's by PT's in 1998 and 2003 and Table 5.2 gives details of the SOM's used in various domains. Percentages are used to enable comparison across the years.

Figure 5.4 Use of SOM 1998 vs 2003

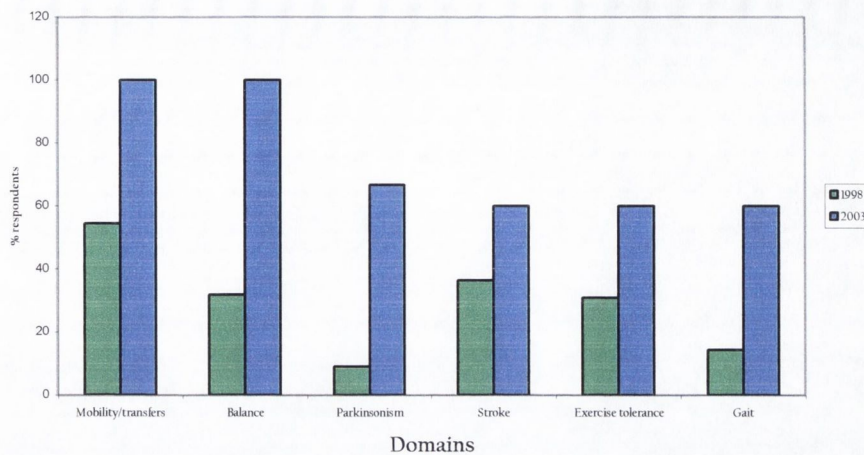


Table 5.2 The specific SOM's used by PT's in 2003

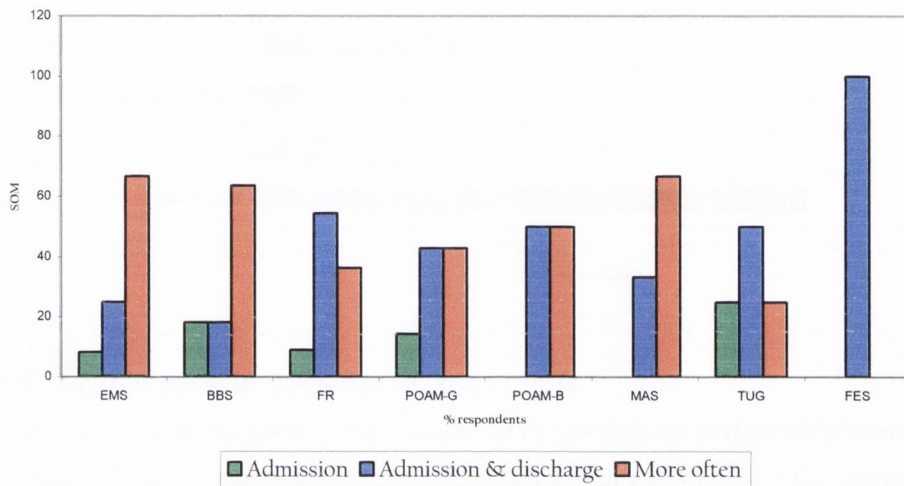
Mobility	Balance	Gait	Stroke	Parkinson's disease/ parkinsonism	Exercise tolerance
EMS 86.7%	BBS 73.3%	POAM-G 60%	MAS 60%	Webster 13.3%	SMDT 33.3%
TUG 40%	FR 80%		MCA 6.7%	Hoehn & Yahr 26.6%	Shuttle test 13.3%
POAM-G 73.3%	POAM-B 66.7%		9HPT 20%	UPDRS 6.7%	RPE 33.3%
SMDT 6.7%	FES 26.7%		ROM to pain 6.7%		VAS 40%
RMI 6.7%			MI 6.7%		
BI 6.7%			RMI 6.7%		
FAC 6.7%					

EMS- Elderly Mobility Scale, TUG- Timed Up & Go, POAM-G- Performance Oriented Assessment of Mobility Gait subscale, SMDT- Six Minute Distance Test, RMI- Rivermead Mobility Index, BI- Barthel Index, FAC- Functional Ambulation Category, BBS- Berg Balance Scale, FR- Functional Reach, POAM-B- Performance Oriented Assessment of Mobility Balance subscale, FES- Falls Efficacy Scale, 9HPT- Nine-hole peg test, ROM to pain- shoulder range of motion to the point of pain, MI- Motricity Index, UPDRS- Unified Parkinson's Disease Rating Scale, RPE- Rate of Perceived Exertion, VAS- Visual analogue scale.



In addition to ascertaining whether and what types of SOM were used, a further question was posed to identify the pattern and frequency of use of the SOM. Figure 5.5 illustrates the results of this question for the most commonly used SOM.

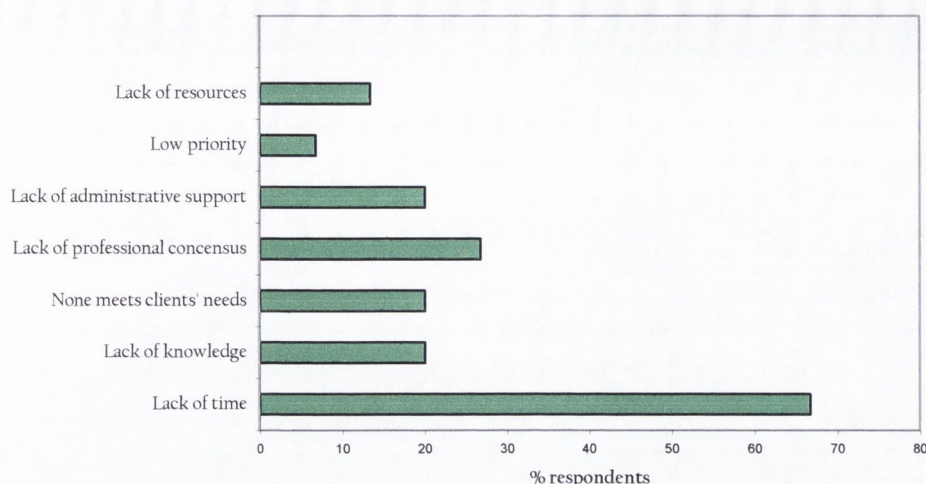
Figure 5.5 Frequency and pattern of use of SOM



### 5.2.3. Reported barriers to the use of SOM

Participants were asked to identify any barriers that existed to their use of SOM. Multiple answers were allowed; hence the percentages in Figure 5.6, which identifies the frequency with which a variety of barriers were reported, exceed a total of 100.

Figure 5.6 Barriers to the use of outcome measures



#### 5.2.4 Confidence-self-efficacy in the use of SOM

The final section of the 2003 survey considered the confidence of the participants in systematically using outcome measurement. Self-efficacy is a method for assessing the self-reported level of confidence in performing a task or activity. Each participant was asked to rate their confidence from 0% (not at all confident) to 100% (totally confident) for twelve statements, which were also employed by the CPA in their 1998 survey (Kay et al 2001). The statements are listed below.

1. Knowing **whether** suitable measures are available.
2. Knowing enough about reliability and validity **to choose** the best measure.
3. Knowing about **scale construction** to develop my own instrument.
4. Knowing **what to measure** for my client group(s).
5. Knowing **how to administer** measures in a standardised manner.
6. Knowing **how to score** measures.
7. Knowing **what to do with** the scores.
8. Knowing how to track clients' progress.
9. Knowing **how to link information** from outcome measures with other sources of information e.g. client characteristics.
10. Knowing **how to compare** scores to baseline levels across clients.
11. Overall, have a clear purpose for measurement- **knowing why** to measure.
12. Overall, knowing **what to do with** the information obtained.

Figure 5.7 Confidence in the use of SOM

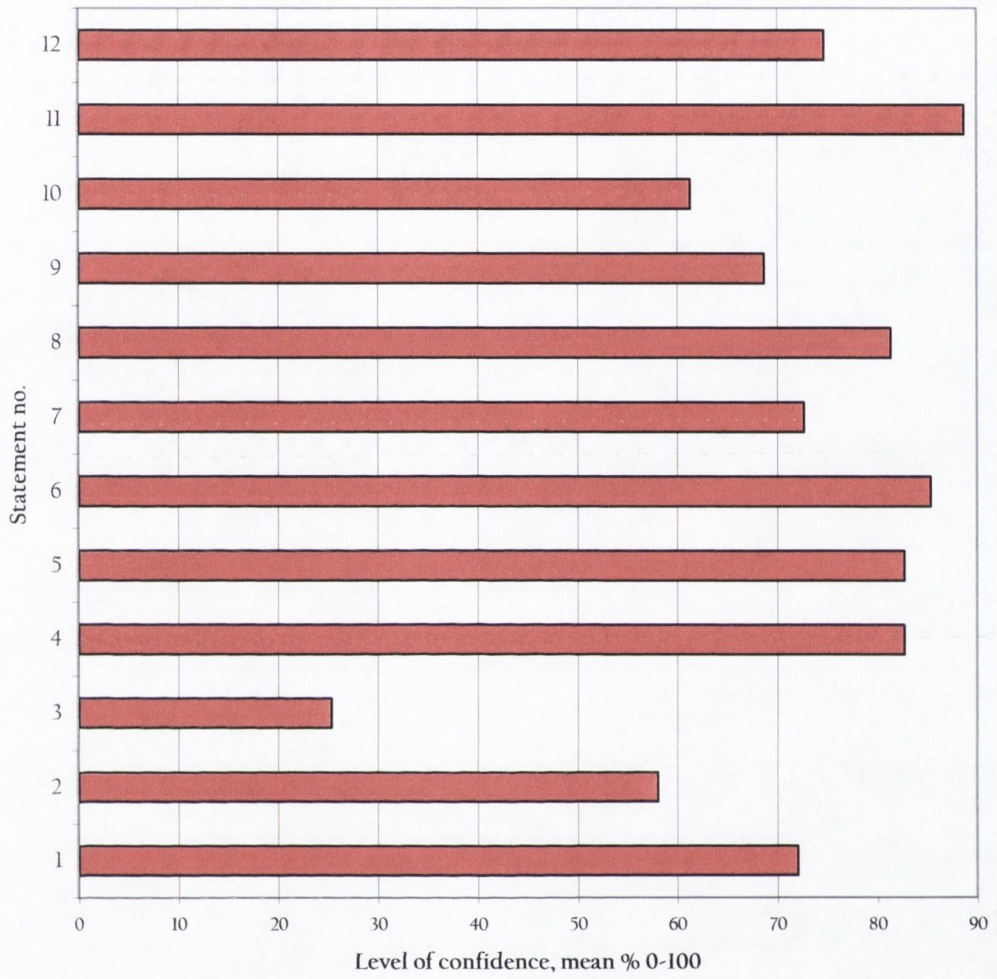


Figure 5.7 (above) illustrates the mean percentage confidence reported for each statement, Table 5.3 outlines the mean % confidence reported for each statement, along with the standard deviation (SD) and the range.

Table 5.3 Levels of confidence (%) in the use of SOM

Statement	Mean	SD	Range
Knowing whether suitable measures are available	72	15.7	40-90
Knowing enough about reliability & validity to choose the best measure	58	23	20
Knowing about scale construction to develop own	25	23	0-70
Knowing what to measure for my clients group (s)	83	11	60-100
Knowing how to administer in a standardised way	83	12	60-100
Knowing how to score measures	85	8	70-100
Knowing what to do with score	73	26	0-100
Knowing how to track clients' progress	81	11	60-100
Knowing how to link information from OM's to other information	69	21	30-90
Knowing how to compare scores to baseline levels across clients	61	28	10-100
Overall having a clear purpose for measurement	89	11	70-100
Overall knowing what to do with information	75	26	0-100

These two surveys of practice, albeit using different methods of data collection allow a comparison to be made between practice in 1998 and 2003. The later survey identifies difficulties and barriers that still prevent the comprehensive use of SOM. These results will be discussed in detail in Chapter 6.

The remainder of this chapter considers the development of a new standardised outcome measure (SOM) for use by physiotherapists working in rehabilitation of older people.

### 5.3 Establishing an operational definition for the new instrument

For the purposes of this work, an operational definition was taken to mean the broad framework for the new SOM. Two significant publications relevant to physiotherapy are the Description of Physical Therapy (WCPT 1999) and the International Classification of Function (WHO 2001). The measure was designed to fall into the category of 'disability' or 'activity' within the ICF framework since it recognised that change at this level is more relevant to individuals that change at the level of impairment of structures and functions. Consideration was given to what functions the new OM should have and also a review of the barriers considered by physiotherapists to be relevant to their patients' achievement of functional independence was also considered.

The results in this section are from two sources- question 4 in the 1998 survey, which asked participants to identify, within each of the domains of mobility/transfer and stability etc., the important functions of an outcome measure and a survey of the barriers to achieving functional independence.

#### 5.3.1 The function of an outcome measure

Feinstein (1987) has identified a number of functions for clinical indices, these include, description of a state, measure of change, estimation of prognosis, offering a guideline. The full list offered to the participating PT's included

- Provide a problem list (problem)
- Identify a state (state)
- Denote a change due to intervention (change)
- Predict an outcome (predict)
- Offer a guideline (guide)
- Teaching tool (Teach)
- Communication aid with other services (Comm)

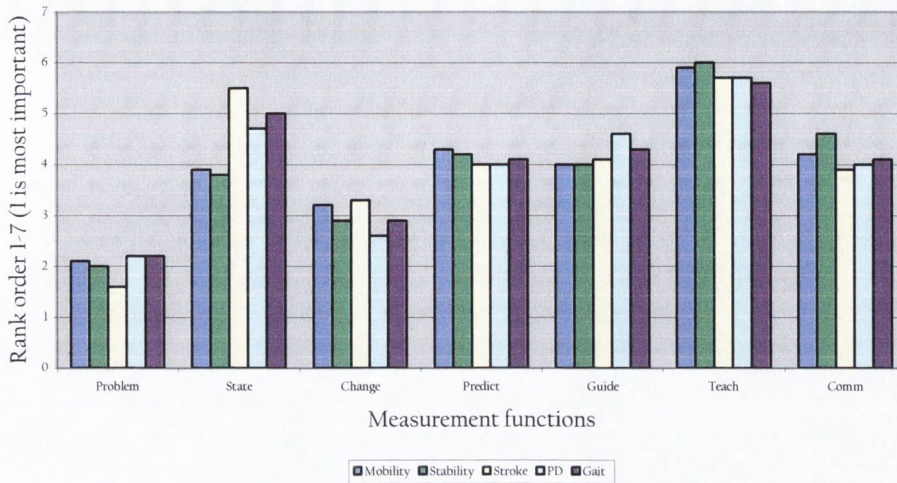
As part of the 1998 survey, the participants were asked to rank these in order of importance for each domain i.e. mobility, balance, gait, stroke and Parkinson's disease. The ranking ranged from 1-7, where one was the most important function.

- Each function was analysed separately, for each domains.
- For each measurement function, the individual rank given for each domain was noted e.g. in the case of 'denote a change' in the context of 'mobility', 22 values were available.
- Absent values were allocated a score of 4, which is the theoretical mean of the range of scores (1-7). If zero was used this would influence the later calculation of means and standard deviations. When, tied scores were presented by some participants, a mean of the two scores was calculated.
- Thereafter, the means, standard deviations and ranges were calculated. These are listed in Table 5.4. The lower the mean rank, the more important the function, since in the ranking 1= most important and 7=least important.
- Taking the mean rank for each domain as an indication of how important the particular function is considered, a lower score indicates higher importance. Figure 5.8 illustrates the relative importance of each measurement function across each domain.

**Table 5.4 Rank order scores for each measurement function, in the each domain- mean (standard deviation)**

	Problem	State	Change	Predict	Guide	Teach	Comm
Mobility	2.1(1.3)	3.9(1.3)	3.2(1.4)	4.3(1.7)	4.0(1.6)	5.9(1.6)	4.2(1.5)
Stability	2.0(1.5)	3.8(1.5)	2.9(1.4)	4.2(1.6)	4.0(1.4)	6(1.6)	4.6(1.3)
Stroke	1.6(1.1)	5.5(1.1)	3.3(1.5)	4.0(1.3)	4.1(1.5)	5.7(1.2)	3.9(1.6)
PD	2.2(1.4)	4.7(1.4)	2.6(1.5)	4.0(1.6)	4.6(1.5)	5.7(1.2)	4(1.7)
Gait	2.2(1.8)	5.0(1.8)	2.9(1.2)	4.1(1.6)	4.3(1.5)	5.6(1)	4.1(1.7)

Figure 5.8 Measurement functions across domains- rank order



The results of this data would suggest that two important features of outcome measures are the capacity to measure change due to intervention and the ability to inform the development of a problem list. Both of these functions suggest that when developing a new SOM, it requires an increased level of detail than in SOM in use at the time.

### 5.3.2 A practice review to identify the barriers older people experience to achieving functional independence in the context of physiotherapy.

This aim of this survey, completed in 1999, was to examine the extent to which a variety of barriers could impact on the achievement of functional independence in older people attending physiotherapy in-patient and out-patient rehabilitation programmes. Seven questions were scored using a Likert Scale, with anchors of 'Not at all' and 'Hinders intervention'. For each of the following potential variables- cognition, pain, mood, communication, problems with balance/stability, problems with mobility, problems with mobility and poor fitness/deconditioning- the therapists were asked to score the impact of the variable as a barrier to the patient achieving functional independence. Questionnaires were completed on 133 patients. Figures 5.9-5.15 represent the results of this survey.

Figure 5.9 Mobility as a barrier to achieving functional independence

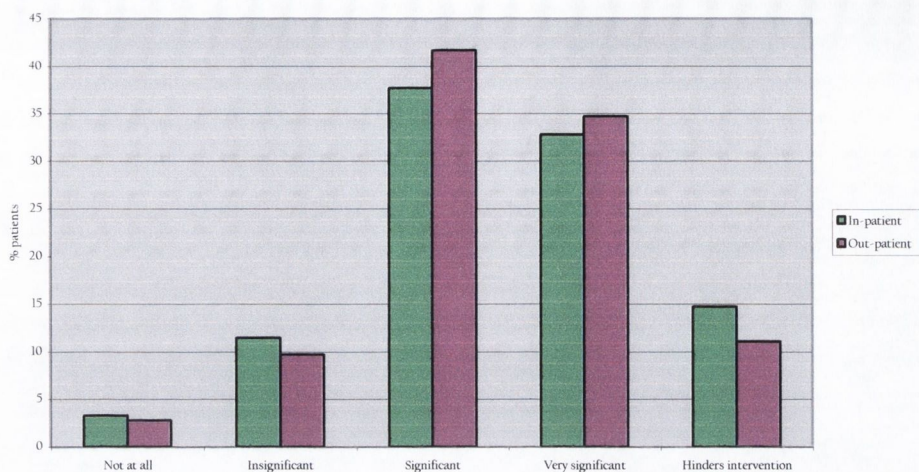


Figure 5.10 Balance problems as a barrier to achieving functional independence

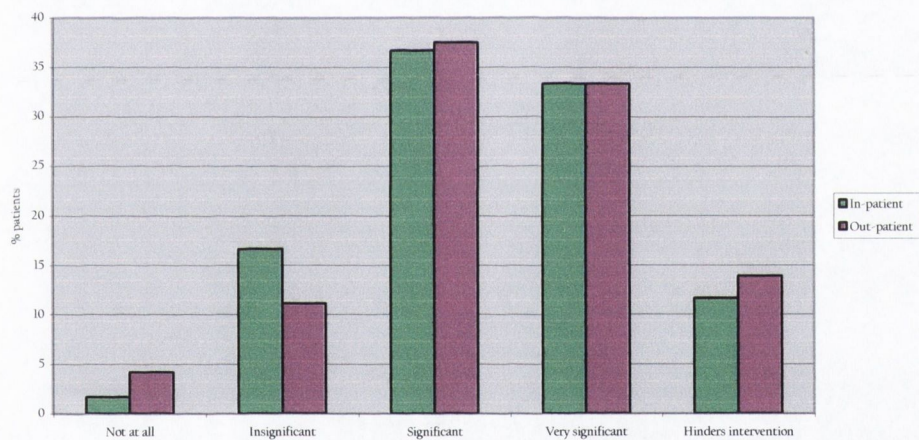




Figure 5.11 Decreased fitness as a barrier to achieving functional independence

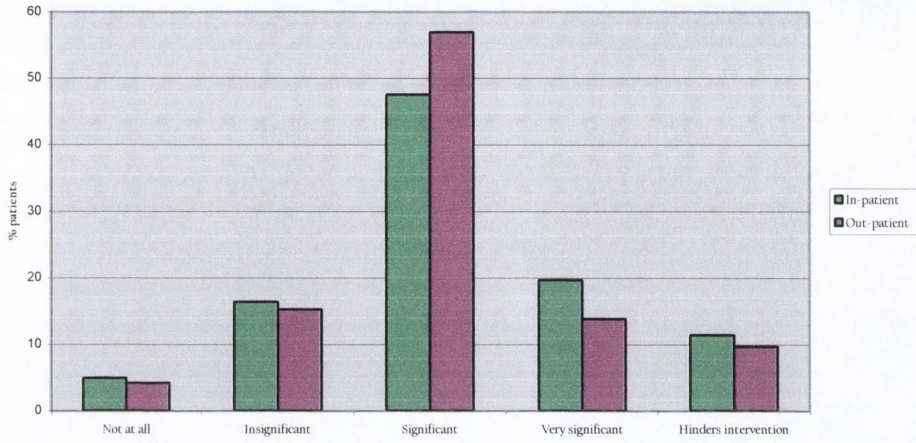


Figure 5.12 Impaired cognition as a barrier to achieving functional independence

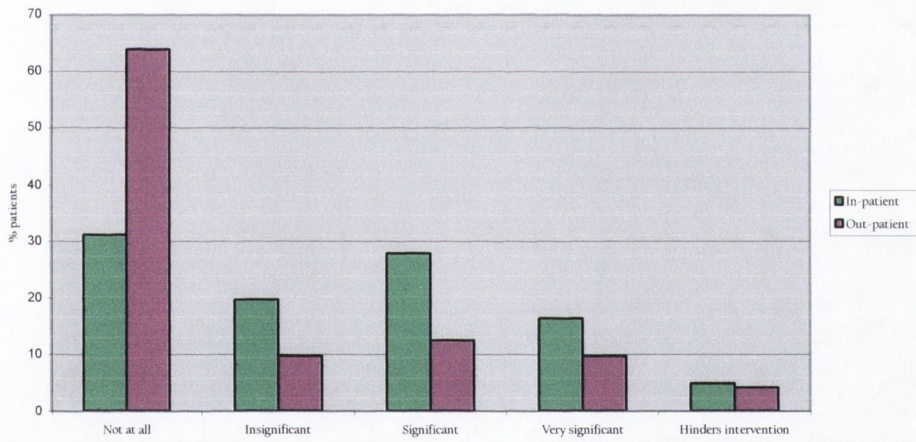


Figure 5.13 Communication difficulties as a barrier to achieving functional independence

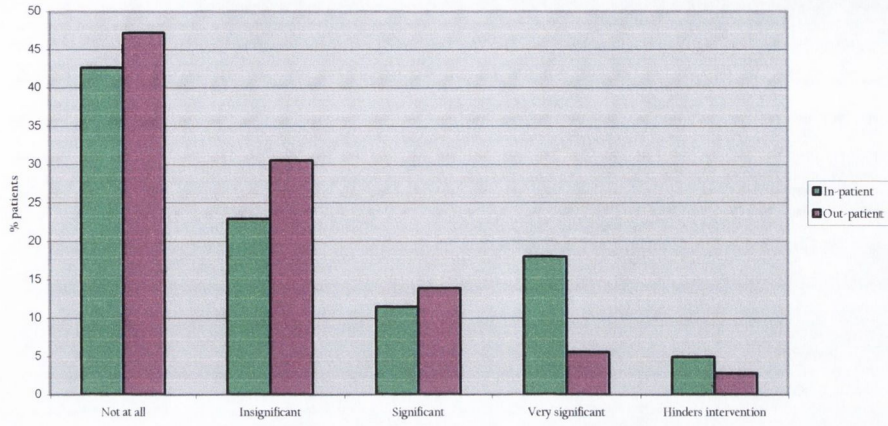


Figure 5.14 Mood as a barrier to achieving functional independence

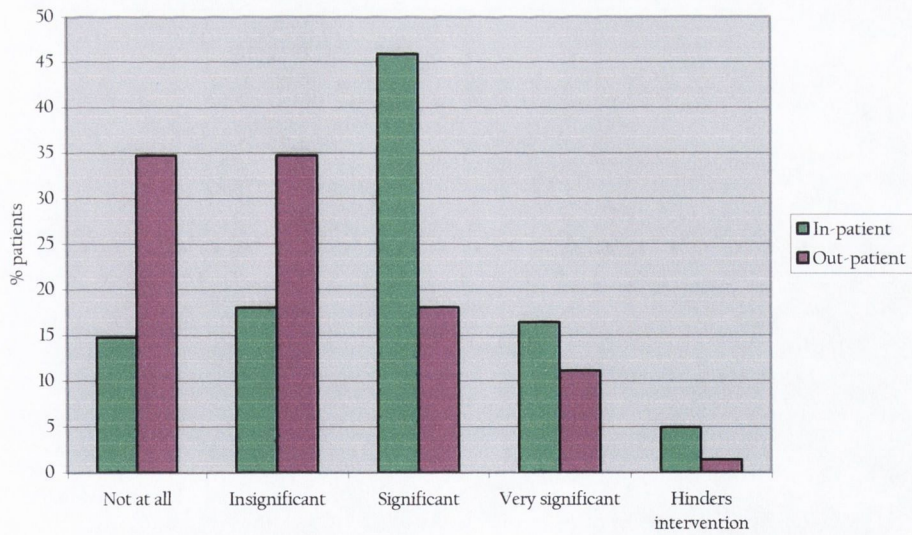


Figure 5.15 Pain as a barrier to achieving functional independence

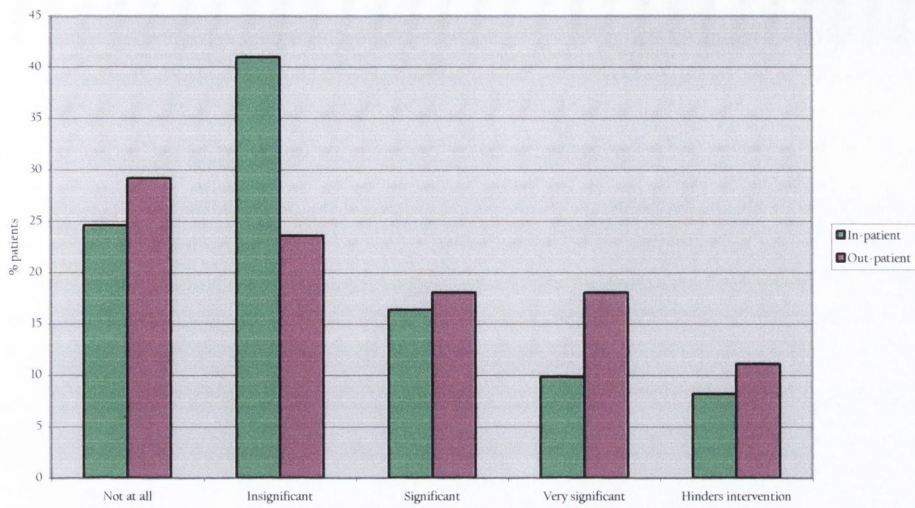
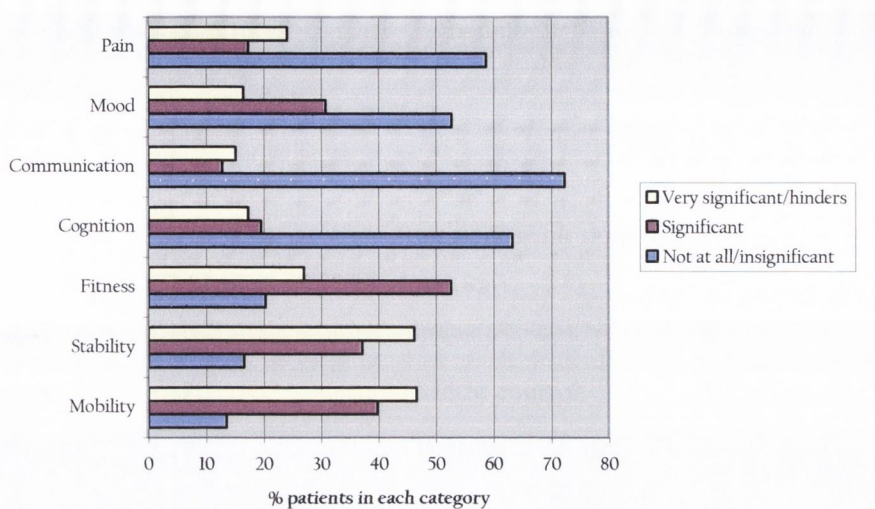


Figure 5.16 illustrates the comparison of all the domains. The data for 'not at all significant' and 'insignificant' and 'very significant' and 'hinders intervention' are combined for clarity in the illustration. It is clear that problems with mobility, stability and fitness/deconditioning are the main barriers associated with achieving functional independence in the context of physiotherapy rehabilitation. Hence, the creation of a new SOM that considered functional mobility and stability was relevant to practice in Ireland.

Figure 5.16 Comparisons of impact of barriers



Sections 5.4 and 5.5 report the results of the specific development of the new SOM

## 5.4 Validity

### 5.4.1 Face validity

Face validity considers whether on the face of it, the instrument measures what it sets out to measure. When the survey of practice was undertaken in 1998, it was considered likely that a new SOM might be later developed and that preliminary data on what PTs considered should be included in SOM's would provide an initial way forward. To this end, question 3 in the 1998 survey asked 22 Senior Physiotherapists what they considered should be included in the an assessment for mobility problems, balance problems, gait disorders, stroke, Parkinson's disease/parkinsonism and problems with exercise tolerance. Tables 5.5-5.7 and Figures 5.17-5.19 outline the items reported for inclusion in an assessment of Parkinson's disease, stroke, mobility, balance, exercise tolerance and gait disorders and the frequency of their report.

The results were very disparate, including both impairments of structure/functions and activity limitations. No items of impairment were included in this functional scale, since it was to be a measure of disability. The decision to include items on the first iteration of the scale was based on a review of the results, the relevance of functional tasks, the availability of normative data and clinical experience of the author.

Table 5.5 Items for inclusion in an assessment of Parkinson's disease

Item- activity, participation	Frequency of report	Item-impairment	Frequency of report
Gait	12	Tone/power/sensation	19
Balance	12	Range of motion/trunk range of motion	10
Transfers	9	Tremor	6
Function	7	Motor control	3
Timed tasks	5	Posture/ Base of support	3
Bed mobility	4	Dexterity	1
Turning	2	Freezing	2
Stairs	1	Micrographia	1
Quality of life	1	Safety	1
Other- Respiratory status, Drug therapy, Cognition	7	Quality of movement	3

Figure 5.17 Suggestions for inclusion in an assessment of mobility

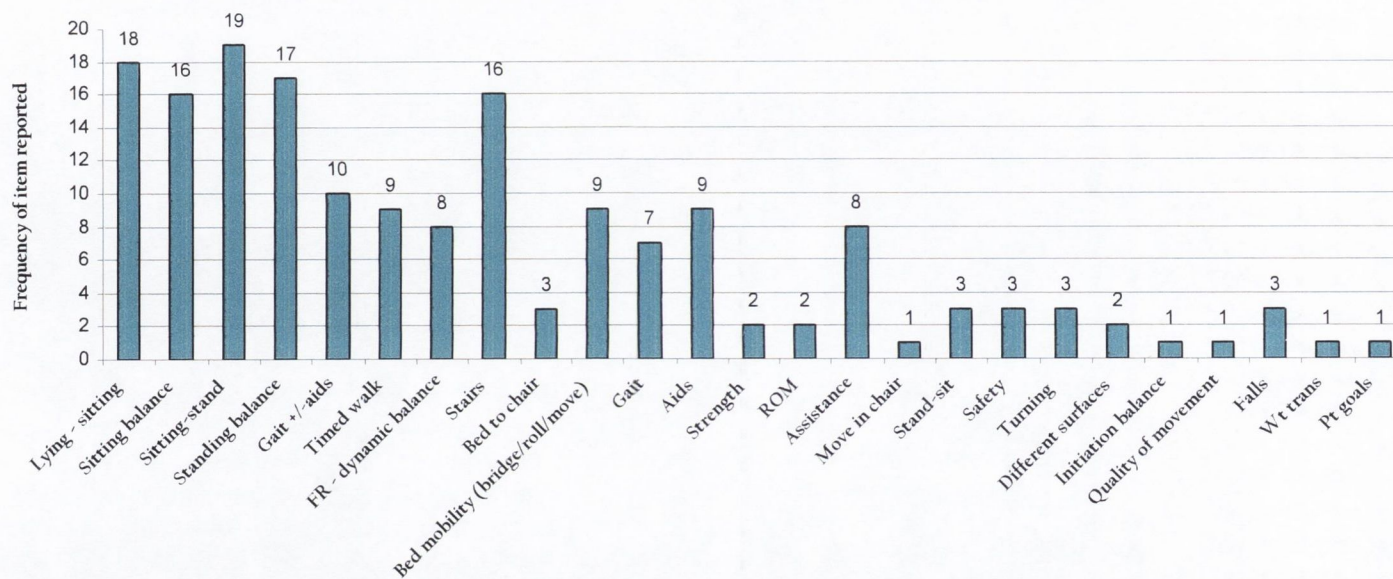
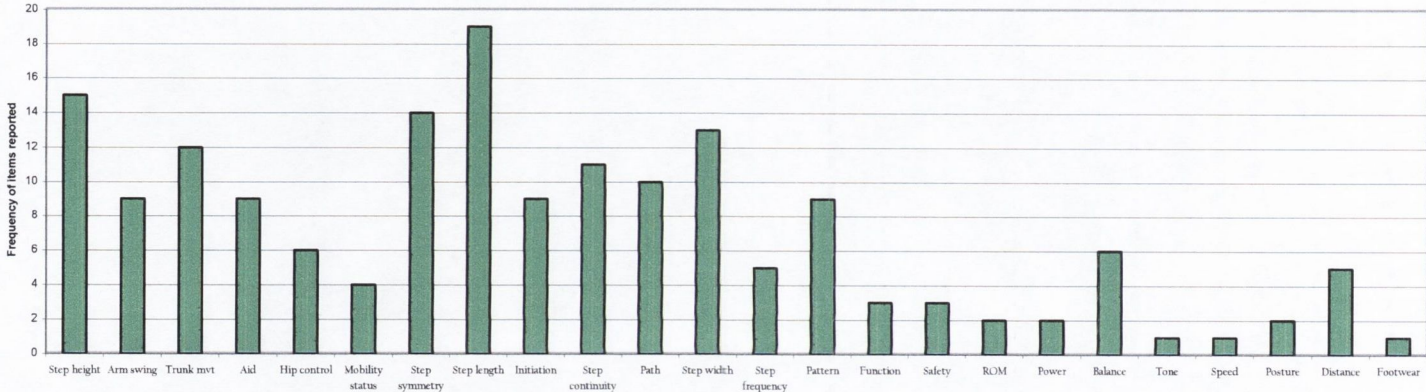


Table 5.6 Items for inclusion in an assessment of balance

Item- activity & participation	Frequency of report	Item	Frequency of report
Sitting balance	20	Vision	12
Walking Uneven surfaces- directions Obstacles	21	Proprioception Co-ordination Range of motion- Cervical spine	14
Reach & bend- displacement response	22	Causes of imbalance dizziness- vestibular-	5
Standing balance- stance positions	25	Associated conditions	2
Turning	8	Power	3
Sit-to-stand	14		
Function-safety	4	Others-	1
Stairs	2		
Lying-to-sit	4		1

Figure 5.18 Items for inclusion in an assessment of gait disorders





5.19 Items for inclusion in an assessment of exercise tolerance

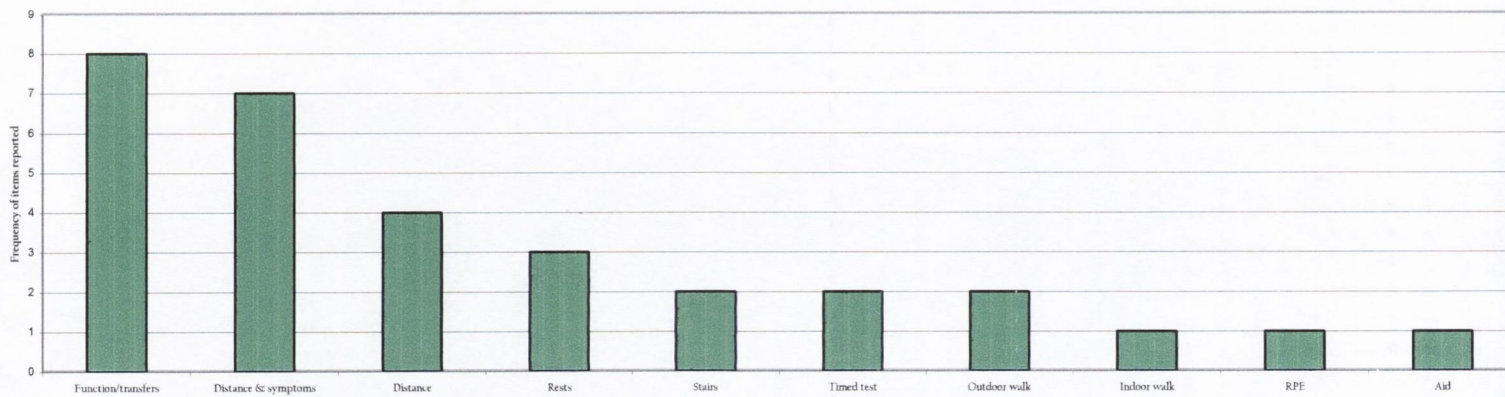


Table 5.7 Items for inclusion in an assessment of stroke

Item-activity, participation	Frequency of report	Item-impairment	Frequency of report
Gait & walking ability	19	Tone	13
Bed mobility	12	Quality of movement/weight transference	11
Sitting balance	12	Strength	8
Sit-to-stand	12	Range of motion	8
Standing balance	12	Sensation	8
Sitting to lying	13	Perception	7
Lying to sitting	13	Cognition/ Comprehension	7
Level of dependence/function	11	Co-ordination	6
Stairs	6		
Floor to stand	3		
Patient goals	4		

#### 5.4.2 Content validity

Content validity examines the content of the measurement, after the first iteration of the scale was developed (Appendix 6), it was circulated to eight Senior Physiotherapists working the rehabilitation of older people. The PT's reflected the areas where the new SOM would be used in the future – in-patient rehabilitation, out-patient rehabilitation and research. The participants were asked to use the new SOM for two weeks of a self-selected sample of the patients and thereafter answer a series of questions for each item. Three questions were asked, as follows,

- Rate the importance of this item 0-5 (not at all important-very important)
- Is the scoring acceptable?
- Is the hierarchy of the scoring acceptable?

There was an invitation for additional comments on each item (Appendix 7). Each PT was interviewed individually. Tables 5.8.1-5.8.9 illustrates the results of this first part of the content validity methodology

Table 5.8.1 Bed rise (BR)

PT code number	Scoring	Instructions-bed or plinth?	Rank*	Other comments
1	Why is 0 the top score?	Use plinth	5	Query re: upper extremity (UE) separate to trunk Consideration of those who need a prompt because of poor cognition Consider safety
P2	OK	Use plinth	5	
3	Why is 0 the top score?	Use plinth	5	Is use of the UE not normal ageing?
4	OK	Use plinth	5	Consider supervision only versus verbal cueing Consider which side the pt gets of out bed, stroke
5	OK	Use plinth	5	
6	OK	Use plinth	5	Define maximum versus minimum assistance Is there a need to differentiate UE and trunk/lower extremity (LE)?
7	Why is 0 the top score?	Use plinth	5	Clarify LE/trunk/UE, maybe use term momentum?
8	What about if requires verbal cueing & compensation?	Bed	4	Consider mechanical aid with 1 or 2.

Review & actions

Clarification of the use of the upper extremities was made in the next iteration. The scoring was reversed as a review of the PT scales physiotherapy used in the 1998 survey demonstrated that a high score indicated good performance. A specific question about combining the use of 'requires assistance for the LE & trunk' and 'requires assistance for UE' was placed in the next review. Note in instructions on side of bed for people with stroke.  
\*Rank order 0=not at all important-5=very important. Majority regarded this as 'very important'. Median 5.Retain item for next phase.

Table 5.8.2 Sit-to-stand (STS)

PT code number	Scoring	Instructions	Rank	Other comments
1	Why is 0 at the top?	OK	5	Define cueing, is it for safety or task description.
2	OK	Type of chair or perhaps use hips & knees at 90 degrees	5	
3	Why is 0 the top score	OK	5	Consider 1 plus aid, no consideration of supervision.
4	OK	Chair +/- arm rests	5	Assistance of 1, consider using minimum and maximum, with FIM definitions.
5	OK	? use of hands, often natural for older people	5	Supervision is omitted.
6	OK	Re: chair height, use average or note height	5	Add space for additional compensatory strategies
7	Why is 0 the top score?	Define standard rehabilitation chair	5	Consider giving example of mechanical aid Physical assistance vs supervision Momentum in compensations Consider chairs at home for community use
8		Define standard rehabilitation chair	5	Mechanical aid – consider +1 and +2

Results & actions

Scoring was changed as noted in BED RISE above. Cueing was defined and supervision was included. Specific question about FIM definitions of assistance prepared for next review. Momentum included in description of compensations.

Rank order 0=not at all important-5=very important. Majority regarded this as 'very important'. Median 5. Retain item for next phase.

Table 5.8.3 Standing (Stand)

PT code number	Scoring	Instructions	Rank	Other comments
1	Why is 0 the top score	With or without aid, define semi tandem.	4	What is the foot position for immediate standing? Arm position - does this matter?
2	OK	OK	5	
3	Why is 0 the top score?	OK	4	
4	OK	With or without physical support/ assistive device	4	
5	OK	OK	4	
6	OK	Does this allow for assistance of 1? Define semi tandem.	5	Is there enough on the dependent end of the scale, it may be too advanced.
7	Why is 0 the top score?	OK	5	What is relevance of immediate standing balance - perhaps define better.
8	Scoring incorrect order.	How far apart should feet be?	5	Is it possible to place feet - in the case of stroke, for example?
Review & actions	Scoring direction changed as noted in BED RISE above. Clarification of arm position was included in instructions and for positioning of feet. The end of the scale to cover more dependent people was expanded. Rank order 0=not at all important-5=very important. Majority regarded this as 'very important'. Median 4.5. Retain item for next phase.			

Table 5.8.4 Reach &amp; lift (R&amp;L)

PT code number	Scoring	Instructions	Rank	Other comments
1	Why is 0 the top score?	Reach from where?	4	Define assistance of 1 – doing what? Minimum & moderate definitions might be useful.
2	OK	OK	5	
3	Why is 0 the top score		5	
4	OK	OK	5	
5	OK	OK	4	
6	OK	OK	4	Reminder that stopwatch has to record to one decimal place.
7	Why is 0 the top score?	Reach from where?	5	Define assistance in terms of requirement for stabilisation.
8	OK		5	
Review & actions	Scoring changed as noted in BED RISE above. Assistance of one defined i.e. for stability. Include question in next review on FIM definitions. Rank order 0=not at all important-5=very important. Majority regarded this as 'very important'. Median 5.Retain item for next phase.			

Table 5.8.5 Bend &amp; reach (B&amp;R)

PT code number	Scoring	Instructions	Rank	Other comments
1	Why is 0 the top score?	Lift from where?	4	Define assistance of 1 – doing what? Minimum & moderate definitions might be useful, holding what – holding book or help to come up? Consider that for many people, bending without knee flexion is normal habitual way to bend.
2	OK	OK	5	
3	Why is 0 the top score?	OK	5	Define assistance in rising.
4	OK	OK	5	
5	OK	OK	4	Clarify difference between 6 and 5.3.
6	OK	OK	4	Reminder that stopwatch has to record to one decimal place. Do the times allow compensatory strategies?
7	Why is 0 the top score?	Reach from where?	5	Is inadequate knee flexion a compensation? Define assistance.
8	OK	How far in front should the book be placed?	5	If part of the book is placed on the floor and the remainder dropped, how does this score?
Review & comments	<p>Assistance defined in the context of stability. Object changed to 'penny' with more detail given on how far away the object should be placed and on what side i.e. 'dominant or unaffected side in the case of upper extremity dysfunction'</p> <p>Rank order 0=not at all important-5=very important. Majority regarded this as 'very important'. Median 5.Retain item for next phase.</p>			



Table 5.8.6 Repeated sit-to-stand (RSTS)

PT code number	Scoring	Instructions	Rank	Other comments
1	Why is 0 the top score?	End at 5 <sup>th</sup> stand or sit?	4	
2	OK	OK	5	
3	Why is 0 the top score?	OK	4	Using UE's or not?
4	OK	Sit to stand from what	4	
5	OK	OK	5	
6	OK	OK	4	Using UE's or not?
7	Why is 0 the top score?	Specify height of starting surface	5	Using UE's or not?
8	OK	+/- UE	5	Completes with verbal cueing? What about if the patient doesn't control sitting down?
Review & actions	Scoring changed as noted in BED RISE above. Instructions clarified. Note on verbal cueing included i.e. 'verbal cueing to count the number of stands is allowed. No encouragement should be given. Rank order 0-not at all important-5-very important. Majority regarded this as 'very important'. Median 4.5. Retain item for next phase.			

Table 5.8.7 Gait

PT code number	Scoring	Instructions	Rank	Other comments
1	Why is 0 the top score?	Does the patient walk 14m	5	Add descriptors Compensatory with frame/2 crutches = 1 Compensatory with cane = 0.5
2	OK	OK	5	Provide descriptors. Use of rollator versus cane
3	Why is 0 the top score	OK	5	Provide descriptors
4	OK	What surface – carpet, lino	5	Provide descriptors
5	OK	OK	5	Provide descriptors Assistance versus quality
6	OK	OK	4	Unable to walk = 0 in sentence format Descriptors for compensation? Does compensation include physical help e.g. to initiate gait?
7	Why is 0 the top score?	OK	5	Might be difficult in community length of walk Descriptors
8	OK	OK	5	Add descriptors

## Review &amp; comments

Scoring was amended as noted in BED RISE above. Descriptors for each aspect of the gait cycle were added. Gait pattern was made into a separate domain. Ability e.g. cane, crutches detail included in new domain 'gait ability'.

Rank order 0=not at all important-5=very important. Majority regarded this as 'very important'. Median 5.Retain item for next phase.

Table 5.8.8 Gait speed

PT code number	Scoring	Instructions	Rank	Other comments
1	Why is 0 the top score	Should this go before 'gait'	5	Assistance of 1? with aid
2	OK	No instructions given	5	What about practice run?
3	Why is 0 the top score	OK	5	Consider Frame +1
4	OK	Describe test	5	With assistive device, is this independent?
5	OK	No instructions	5	What about supervision
6	OK	OK	5	What exactly is being measured? Speed or dependency?
7	Why is 0 the top score	OK	5	Make timed scoring easier Put all gait under one heading
8	Aid +1, Aid +2		5	Supervision Use of Arjoe frame
Review & actions	<p>Scoring changed as noted in BED RISE above. Definition of categories included in instructions e.g. 'a frame includes a rollator or motorised frame or 2 crutches'. Time as a function of dependence included. Supervision included.</p> <p>Rank order 0=not at all important-5=very important. Majority regarded this as 'very important'. Median 5. Retain item for next phase.</p>			

Table 5.8.9 360 turn

PT code number	Scoring	Instructions	Rank	Other comments
1	Why is 0 the top score?	On the spot or around object?	4	Consider verbal cueing for safety or task completion
2	OK	No instructions given	5	Is a demonstration permitted?
3	Why is 0 the top score	Not needed	5	Consider Frame +1, supervision.
4	OK	Is it continuous turning, with no rests	5	
5	OK	No instructions	5	What about supervision?
6	OK		3	What does the ability to complete 360 degrees with one or two people tell us?
7	Why is 0 the top score	OK	3	How relevant is +2 or +1?
8	Add +1, +2, supervision	Use 'full circle' instead of 360 degrees	5	
Review & actions	On the basis of the comments above and the review of the gait section, turning was included as part of gait.			

After the feedback from the first evaluation of the content validity, the second iteration of the scale was developed (Appendix 8). It was circulated to a further 13 Senior Physiotherapists, who were asked specific questions about each item (Appendix 8). Table 5.9 outlines the results from this exercise.

Table 5.9 Results from second phase of content validity

<i>BED RISE (BR)</i>					
Should assistance for the trunk/lower extremity be combined with upper extremity?	Yes=6	No=5	No answer=2		Action-equivocal response. Combine & review after analysis.
Is the 'verbal cueing' score acceptable as it is or should there be different scores for verbal cueing for safety reasons' and verbal cueing for 'prompting to complete the task'?	Yes=8	No=4	NA=1		Action-verbal cueing isq.
Should this item be included?	Yes=13				Action-retain item.
<i>SIT TO STAND (STS)</i>					
Is the description of the chair acceptable or should the chair dimensions be given?	Description acceptable=8	Dimensions=2	Standardise wrt patient=2	No answer=1	Action-retain description.
Is the 'verbal cueing' score acceptable as it is or should there be different scores for verbal cueing and supervision?	Acceptable as it is=10	Not acceptable=1	No answer=2		Action-retain score isq.
Should 'assistance of one' be	Yes=2	No=5	No answer=6		Action-majority of those who

subdivided further using FIM definitions?					answered indicate 'no'. Do not change.
Are the compensations listed acceptable?	Yes=5	No=3	No answer=5		Retain list. Can be amended in future as required since does not affect scoring.
Should this item be included	Yes=13				Action-retain item.

**STANDING ABILITY & STABILITY  
(Stand)**

Should this item be included?	Yes=13				Action-retain item.
-------------------------------	--------	--	--	--	---------------------

**REACH & LIFT  
(R&L)**

Should the FIM definitions be included for assistance	Yes=3	No=6	No answer=4		Action-majority of those who answered indicate no- do not include FIM definitions.
---	-------	------	-------------	--	--

Should this item be included?	Yes=13				Action-retain item.
-------------------------------	--------	--	--	--	---------------------

**BEND & REACH  
(B&R)**

Is there a need for a compensation section?	Yes=6	No=4	No answer=3		Action-majority of those who answered indicate 'yes', retain.
---	-------	------	-------------	--	---

Should the FIM definitions of assistance be included?	Yes=2	No=5	No answer=5		Action-majority of those who answered indicate no- do not include FIM definitions.
---	-------	------	-------------	--	--

Should this item be included?	Yes=13				Action-retain item.
-------------------------------	--------	--	--	--	---------------------

included?

**REPEATED SIT TO STAND  
(RSTS)**

Is the gap between 1 and 2 too much	Yes-5	No-2	No answer-6		Action-while the function is to consider LE strength, additional sections were included with a view to reviewing after further evaluation. This was also in view of the poor response to this question.
This domain is to consider lower extremity endurance hence no upper extremity use is ideal. Is this acceptable?	Yes-6	No-1	No answer-6		
Should this item be included?	Yes-9	No-3	Not sure-1		Action-include item.

**GAIT ABILITY  
(GA)**

Is 4 acceptable? Or should it be subdivided?	Acceptable-5	Sub-divide-5	No answer-3		Action-equivocal response. Retain isq and review after next evaluation.
Should this be included?	Yes-12	No-0	No answer=1		Action-retain item.

**GAIT PATTERN  
(GP)**

On the basis of inclusion of all the previous sections, does this section need to be included or should it be optional?	Include=5	Optional=5	No answer=2	Missing data from instrument=1	Action-equivocal answer. Retain and review in future analysis
If it is removed, do all the sections	Remove sections=2 No detail	Do not disturb=4	NA=1	MD=1	Action- as above.

need to be provided.  
included?

The respondent was asked to indicate the section they considered should be removed, if they answered 'remove sections' to this answer.

Is the information on compensations necessary?	Yes=4	No=3	NA=5	MD=1	
If the information on compensations is retained. Is it acceptable?	Yes=4	No=0			

Feedback from this exercise yielded the third and final iteration of the new SOM and the remainder of the work on validity and reliability was performed on this final version (Appendix 9), with a view to creating the final version of the scale.



### 5.4.3 Criterion validity

For the purposes of this study, concurrent and predictive criterion validity was examined by comparing scores on the new instrument with scores on two rehabilitation measures- the Elderly Mobility Scale (EMS) (Smith, 1994) and the Performance Oriented Assessment of Mobility (POAM) (Tinetti, 1986) (Appendix 10).

Seventy-five older people in receipt of physiotherapy rehabilitation participate, yielding 73 complete data sets (2 sets had missing data and were excluded from the analysis). They aged from 60-95 years, with a mean age of 79 years (standard deviation 7.3 years). Seventy-five percent of the participants were female. Reasons for referral to physiotherapy varied- stroke rehabilitation (18%), falls management (15%), mobility rehabilitation (50%), mobility & balance rehabilitation (12%), gait re-education (2.7%), strength training (1.4%) other reasons (1.4%).

Table 5.10 illustrates the participants' baseline scores for the

- Elderly Mobility Scale (EMS)
- Problem-oriented assessment of mobility-gait subscale (POAM-G) and Problem-oriented assessment of mobility-balance subscale (POAM-B). These are the two sub-scales of the POAM.
- The new SOM with the gait pattern section included (NS-total)
- The new SOM with the gait pattern section excluded (NS-GP)

It was not clear from an earlier part of the content validity methodology that the gait pattern (GP) sub-section should be included; hence analysis was performed with it included in the total score and also when it was excluded.

Table 5.10 Summary statistics, baseline scores- total sample

	Available range	Mean (sd)	Range
EMS	0-20	13.8 (5.3)	0-20
POAM-G	0-12	6.4 (3.8)	0-12
POAM-B	0-16	8.9 (4.4)	0-16
NS-total	0-95	60.2 (22.4)	1-90
NS-GP	0-55	30.6 (13.1)	1-50

Table 5.11 illustrates the correlation matrix for the admission scores on all the instruments used. A Spearman's rank correlation co-efficient are utilised throughout as the appropriate correlation co-efficient for ordinal data.

Table 5.11 Correlation matrix- baseline scores, all instruments.

	EMS	POAM-B	POAM-G	NS-GP
POAM-B	0.79			
POAM-G	0.73	0.65		
NS-GP	0.80	0.86	0.61	
NS-total	0.78	0.82	0.72	0.96

A subset of the total sample (n=37) was measured at both admission and discharge. Length of stay was an average of 24 days (standard deviation 20 days, range 5-76 days). Tables 5.12 and 5.13 illustrate the summary statistics (mean, standard deviation and range) for the admission and discharge scores and a correlation matrix demonstrating the relationship between admission and discharge scores for all instruments used.

Table 5.12 Summary statistics, admission and discharge score (n=37)

	Mean (sd)	Range
Admission EMS	12.4 (5.9)	0-20
Discharge EMS	14.5 (5.6)	1-20
Admission POAM-B	6.5 (4.3)	0-14
Discharge POAM-B	9.1 (4.3)	0-14
Admission POAM-G	5.1 (4.3)	0-12
Discharge POAM-G	6.6 (4.4)	0-12
Admission NS-GP	34 (14.7)	1-49
Discharge NS-GP	33.9 (14,7)	3-51
Admission NS-total	50.9 (24.4)	1-89
Discharge NS-total	64.2 (24.3)	3-88

Table 5.13 Correlation matrix- admission and discharge score

	Discharge EMS	Discharge POAM-B	Discharge POAM-G
Admission NS-total	0.70	0.80	0.80
Admission NSGP	0.65	0.69	0.44

#### 5.4.4 Construct validity

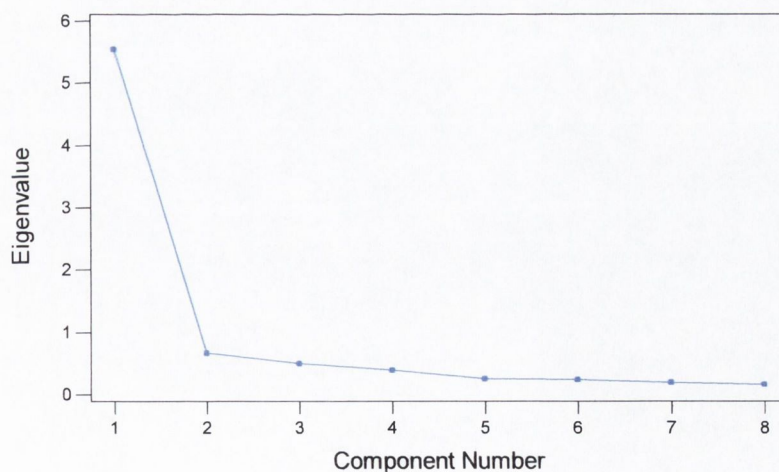
Three constructs were investigated in this study

- That the new SOM would have a strong association with two scales on mobility and balance and that the new SOM would be at least as responsive as the POAM and the EMS- results outlined in 5.4.3.
- That the new SOM would be able to discriminate between two groups of older people, namely those having rehabilitation and healthy, community dwelling older people
- That the new SOM would contain not more than one construct or factor.

##### 5.4.4.1 Factor analysis

A principal components factor analysis was performed on the baseline data (n=78), which identified 1 eigenvalue of greater than 1- Figure 5.20. This contained all eight components of the new SOM and explained 70% of the variance within the data. This suggests that the components of the new SOM can be considered to measure one single domain.

Figure 5.20 Scree plot for factor analysis



#### 5.4.4.2 Community dwelling older people

Sixteen older people, who were self-reported healthy and not in attendance at physiotherapy at the time of the measurement participated in the study. Their mean age was 70 years (standard deviation 6.4 years, range 61-85). Figure 5.21.1 illustrates the difference in score for the NS-GP in the community group and the group of older people who were participating in rehabilitation.

Figure 5.21.1 Score for NS-GP- Community versus rehabilitation groups

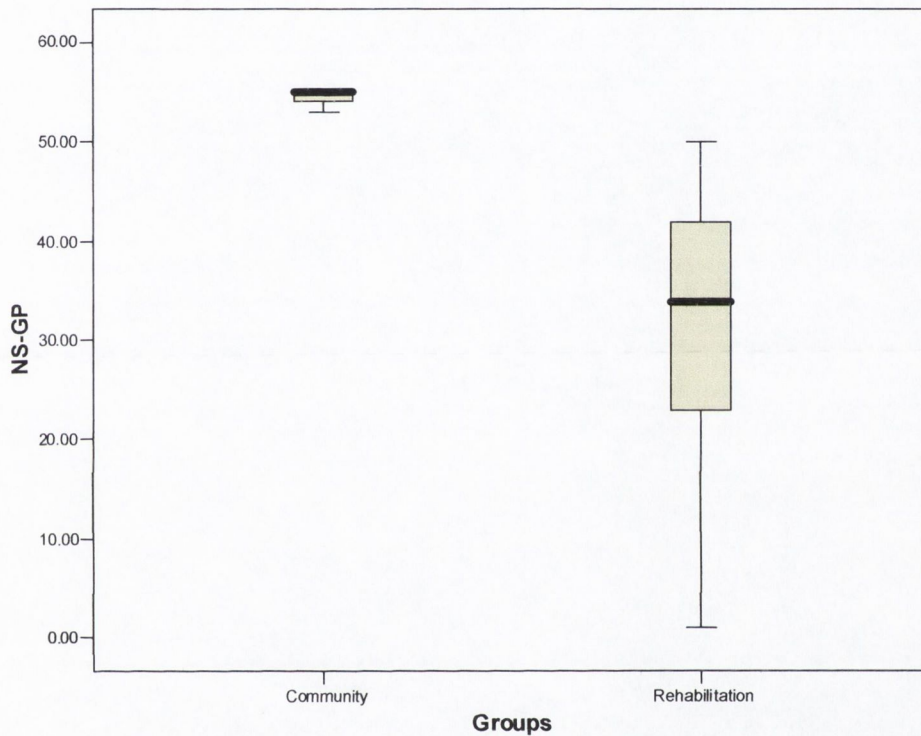
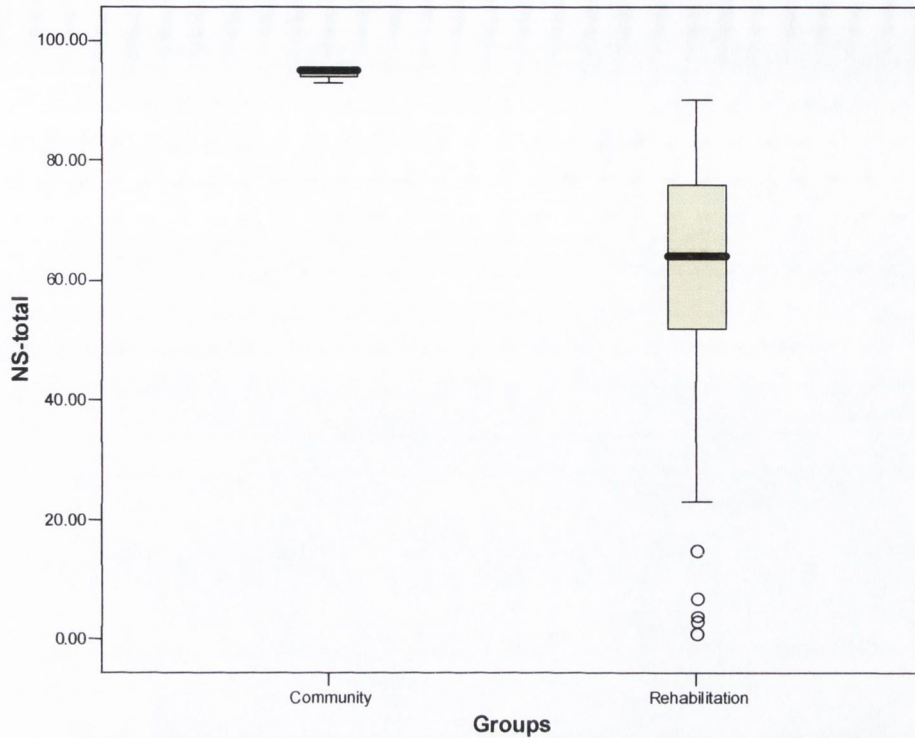


Figure 5.21.2 Score for NS-total- Community versus rehabilitation groups



To compare the medians of the two groups, the analysis used was a Mann-Whitney test. For NS-GP the median for the community group was 55 (95% confidence interval (CI) 54-55) and for the rehabilitation group 34 (95% CI 29-38). The null and alternate hypothesis is as follows,

$H_0$ : that there is a significant difference between the community and rehabilitation groups.

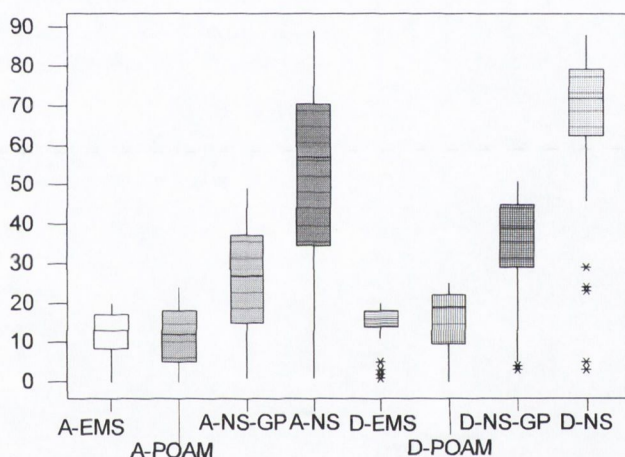
$H_1$ : that there is no difference between the medians of the two groups.

Alpha is set at 5%. The test statistics (z) is 6.54,  $p < 0.0001$ , hence  $H_0$  can be rejected and it can be inferred that there is a significant difference between the two groups. Similarly, for NS-total, the median for the community group was 95 (95% CI 94-95) and for the rehabilitation group 64 (95% CI 59-73). The hypotheses are the same; alpha is also set at 5%. The test statistic (z) is 6.54,  $p < 0.0001$ , hence  $H_0$  can be rejected and it can be inferred that there is a significant difference between the two groups.

### 5.4.5 Responsiveness

A sub-set of the larger sample of 75 older people was measured using the new SOM, the POAM and the EMS at admission and discharge. The average length of stay was 24 days (standard deviation 24, range 5-76). Figure 5.22 illustrates the range of admission and discharge scores for each of the scales. As mentioned previously, the POAM contains two subscales – POAM-B (a balance sub-scale) and POAM-G (a gait subscale). For the evaluation of responsiveness, the total POAM score is evaluated. The boxplots identify the median (solid horizontal line), inter-quartile range (open box) and the range (end point of vertical lines)

Figure 5.22 Boxplots of admission and discharge scores



The mean scores on admission for the EMS, POAM, NS-GP and NS-total were above their respective midpoints. The responsiveness of the data was considered by evaluation the effect size and the relative responsiveness of the three scales.

The effect size provides a ratio of the 'signal to noise' where the 'signal' is the change that occurs due to intervention and the 'noise' is the level of variability within the instrument. It uses the change in scores between admission and discharge and the standard deviation. Since each of the measurement instruments had differing total scores, all the scores were transformed to give a score out of one hundred using the formula:



(Patient score/Total possible score) \*100 (Wright *et al.* 1998)

The general convention for effect sizes is effect sizes of less than 0.2 would be interpreted as small and 0.5 to 0.79 as moderate.

Effect sizes for the scales were as follows-

- EMS 0.36
- POAM 0.52
- NS-GP 0.64
- NS-total 0.51

Relative efficiency (Liang *et al.* 1985) is a method of comparing two scales to ascertain which is more responsive. Relative efficiency is calculated using the following formula

$$\text{Relative efficient (X versus Y)} = (z_X / z_Y)^2$$

Where  $X$  is either the EMS or POAM and  $Y$  is the new SOM and  $z$  is obtained by comparing the admission and discharge scores using a Wilcoxin Signed Rank test. A relative efficiency score of 1 would indicate that  $X$  was equally as responsive as  $Y$ , a score of  $>1$  would suggest that  $X$  was more responsive than  $Y$  and  $<1$  that  $Y$  was less responsive than  $X$ . The new scale should be at least as responsive as those used in existing practice.

Relative efficiency was tested for both the new SOM with the gait pattern sub-section (NS-GP) and without this sub-section (NS-total). Table 5.14 indicates the results.

**Table 5.14 Relative Efficiency (RE) of the new SOM**

Scales	RE
NS-GP versus EMS	1.75
NS-GP versus POAM	1.50
NS-GP versus NS-total	0.96
NS-total versus EMS	1.82
NS-total versus POAM	1.56

## 5.5 Reliability

As discussed in Chapter 2 and 4, **inter-rater** reliability refers to the level of agreement between two or more raters using the same outcome measure on the same subject at the same time. This considers inter-observer agreement. **Intra-rater** reliability describes the level of agreement between scores obtained on a measurement of a subject, taken on two separate occasions, and where, for the purposes of analysis, no actual change is anticipated to have occurred within the subject. It considers the variation that occurs within an observer as a result of multiple exposures to the same subject (Streiner and Norman 1995).

### 5.5.1 Inter-rater reliability

#### 5.5.1.1 Subjects & rater

Fifteen older people who were in-patients in St. James's Hospital participated in this study. Three raters participated in this evaluation and they (R1, R2, and R3) concurrently measured the 15 subjects. A number of techniques exist for the analysis of such ordinal data. The kappa co-efficient is an index of agreement designed specifically for nominal and ordinal scales. While the output of the new measurement in this study appears to be continuous, it is, as in the case of many rehabilitation scales actually ordinal, since the output is neither interval or ratio data, hence methods such as the intra-class correlation coefficient are not appropriate in the analysis of this data. As previously, the total scale (NS-total) and the scale without the *Gait pattern* section (NS-GP) are analysed and presented.

#### 5.5.1.2 Percentage agreement & the kappa co-efficient

The kappa co-efficient takes into consideration the extent to which a proportion of the amount of agreement between raters may be simply due to chance i.e. the proportion of agreement that exists between two sets of measurement when the proportion due to chance has been removed from both the numerator and denominator.

$$k = \frac{P_o - P_c}{1 - P_c}$$

Where  $P_o$  is the observed proportion of agreement,  $P_c$  is the proportion of agreement due to chance.

Cohen (1968) states that the above formula 'quite reasonably yields negative values when there is less observed agreement than is expected by chance, zero when observed agreement can be (exactly) accounted for by chance and unity where there is complete agreement.' The weighted kappa co-efficient allows for partial agreement e.g. a difference of 1 or 2 points is acceptable. A system is employed where weights are assigned to cells within the contingency table. The most commonly used weighting system is quadratic weights where differences in one category are given a weight of 1 and differences in 2 categories a weight of 2<sup>2</sup> (Streiner & Norman 1995, Cohen 1968). Table 5.15 outlines the percentage of agreement for each sub-section of the new instrument and the associated weighted *k* co-efficient.

**Table 5.15 %agreement & kappa co-efficients for agreement between raters**

	Rater 1 and 2		Rater 2 and 3		Rater 1 and 3	
	% agreement mean (sd) range	<i>Weighted</i> <i>k</i> -statistic	% agreement mean (sd) range	<i>Weighted</i> <i>k</i> -statistic	% agreement mean (sd) range	<i>Weighted</i> <i>k</i> -statistic
Bed rise	98.8(4.5) 83.3-100	0.97	98.6 (5.4) 80-100	0.96	97.4 (6.7) 80-100	0.93
Sit-to-stand	98.8 (4.5) 83-100	0.96	94.8 (10.8) 67-100	0.69	93.6 (11.1) 67-100	0.75
Stand	100	0.99	100	0.98	100	0.98
Reach & lift	92 (13.5) 67-100	0.95	97.6 (8.9) 67-100	0.80	94.4 (11.4) 67-100	0.93
Bend & reach	100	1	97.6 (8.9) 67-100	0.98	97.6 (8.9) 67-100	0.98
Repeated STS	100	1	100	0.99	100	0.99
Gait ability	100	1	100	1	100	1
Gait pattern	94.5 (4.8) 86-100	0.77	90.7(9.2) 72-100	0.59	92 (5.8) 83-100	0.89

Total-GP	98.4(2.3)	0.99	97.9 (3.7)	0.98	97.1 (4.3)	0.98
	95.5-100		88-100		88-100	
Total	96.6(3.4)	0.97	95.4 (5.7)	0.94	95.4 (3.6)	0.97
	90-100		83-100		87-100	

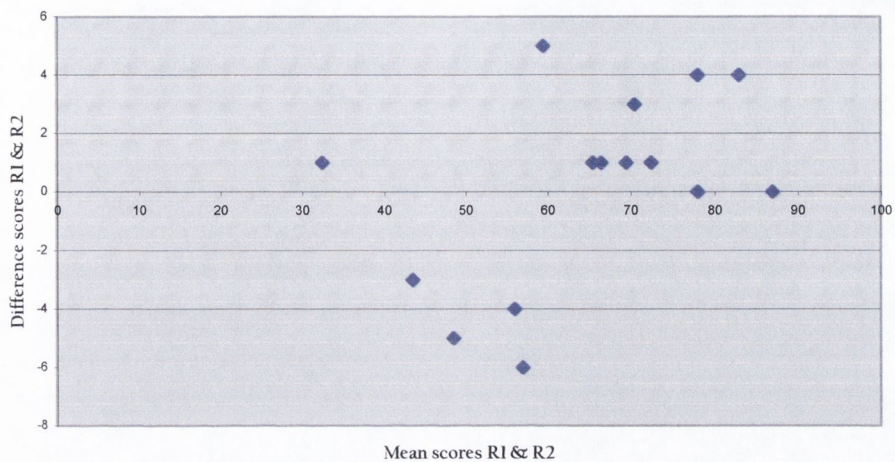
### 5.5.1.3 Bland & Altman

A further method of analysis has been described by Bland & Altman and discussed elsewhere (Chapter 2, Section 2.4). For both NS-GP and NS-total, the method was applied as follows-the differences between the measurements of R1 and R2, R2 and R3 and R1 and R3 were calculated, as was the average of the each of the two sets of measurement. These two variables were plotted against one another. This gives a visual representation of the size of each difference and the distribution around zero – Figures 5.23-5.28. If no difference existed between the two sets of measurements between raters, all the points would lie along 0. These figures also allows the observer to note how the differences may relate to the mean i.e. is more error likely to occur at a given point with the measurement range. The 95% limits of agreement can be calculated for each set of raters, as follows,

$$95\% \text{ limits of agreement} = \text{mean of the differences} \pm 1.96 \text{ SD}_{\text{diff}}$$

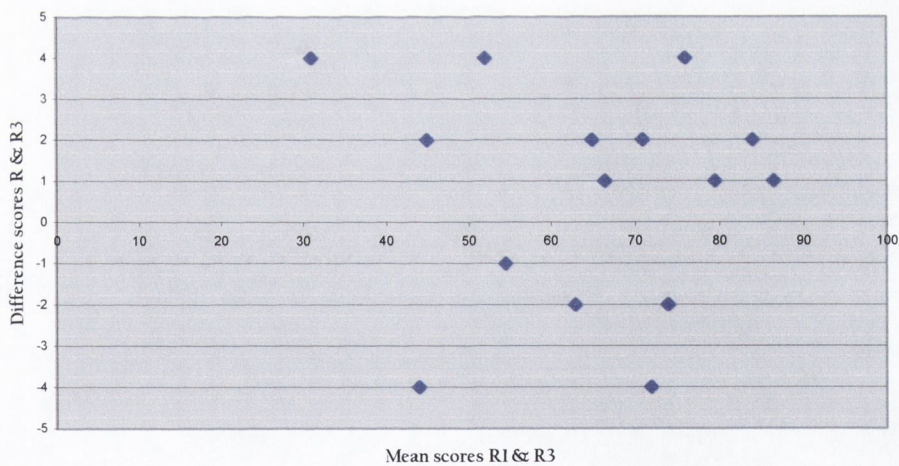
They are listed below each of the figures.

Figure 5.23 Bland & Altman plot- NS-total, R1 & R2



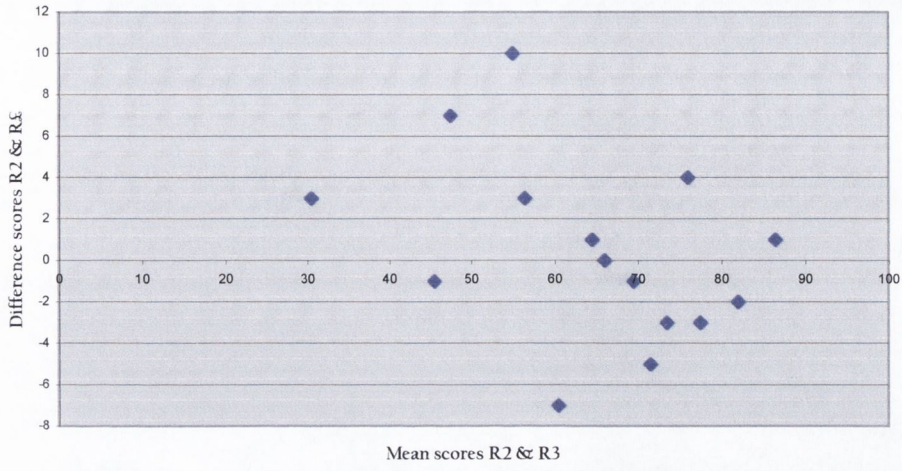
R1 & R2 - 95% limits of agreement are -6.4 to +6.8.

Figure 5.24 Bland & Altman plot NS-total, R1 & R3



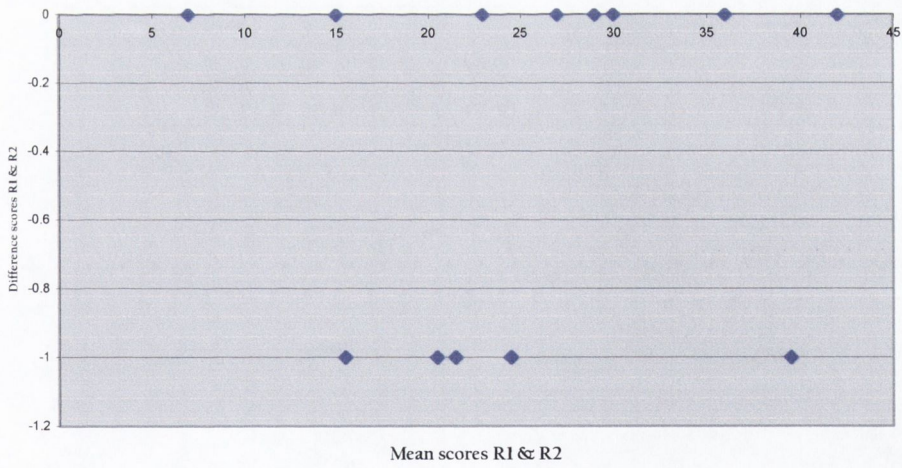
R1& R3 - 95% limits of agreement are -4.6 to +5.9.

Figure 5.25 Bland & Altman plot NS-total, R2 & R3



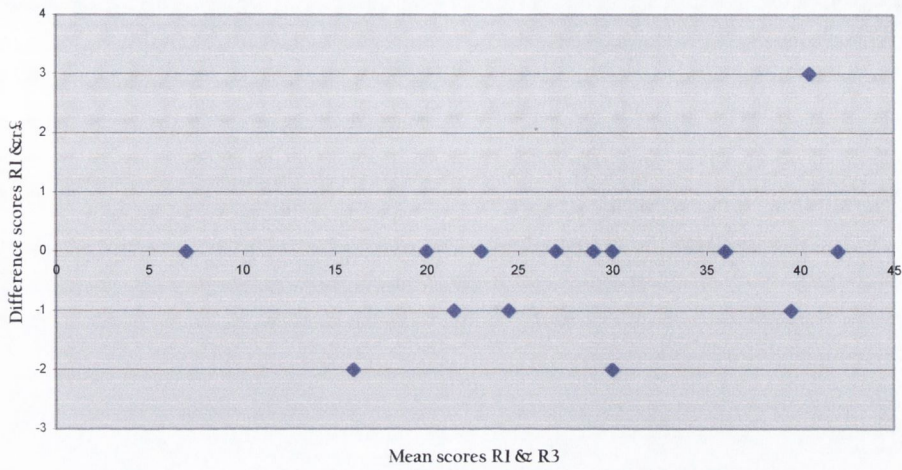
R2& R3 - 95% limits of agreement are -8.3 to 9.2.

Figure 5.26 Bland & Altman plot NS-GP, R1 & R2



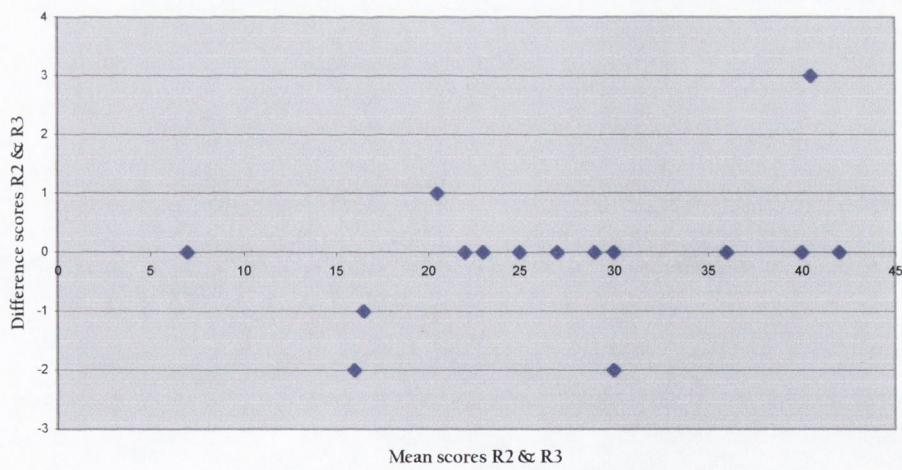
R1& R2 - 95% limits of agreement are -1.3 to +0.6.

Figure 5.27 Bland & Altman plot NS-GP, R1 & R3



R1 & R3 - 95% limits of agreement are  $-2.8$  to  $+2.0$ .

Figure 5.28 Bland & Altman plot NS-GP, R2 & R3



R2 & R3 - 95% limits of agreement are  $-2.3$  to  $+2.2$ .

## 5.5.2 Intra-rater reliability

### 5.5.2.1 Subjects and rater

Twenty older people who were in-patients in St. James's Hospital participated in this study. One rater measured the participants at time 1 (T1) and within 24-48 hours thereafter (T2).

### 5.5.2.2 % agreement and weighted kappa coefficients.

Table 5.12 illustrates the percentage agreement for each section of the scale.

Table 5.12 % agreement & weighted K coefficient between T1 and T2

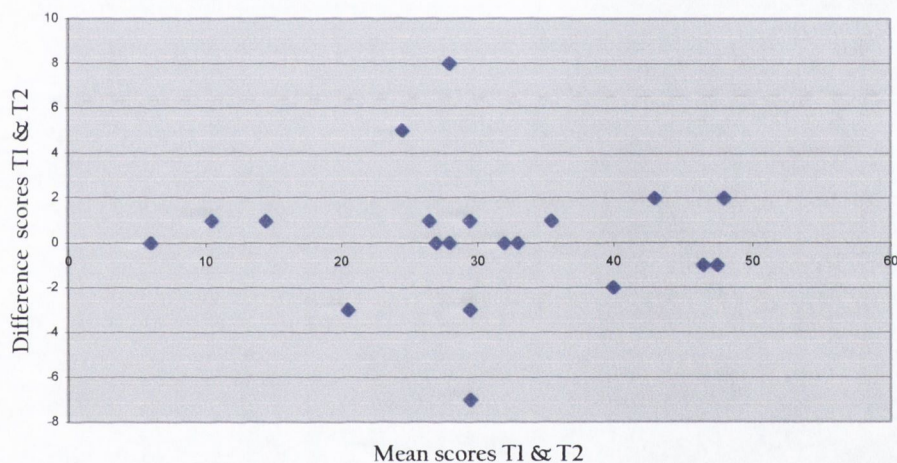
	% agreement mean (sd) range	Weighted k-statistic
Bed rise	87 (21) 40-100	0.73
Sit-to-stand	96 (12) 57-100	0.88
Stand	91 (17) 33-100	0.86
Reach & lift	87 (22) 40-100	0.86
Bend & reach	97 (13) 40-100	0.97
Repeated STS	89 (20) 40-100	0.90
Gait ability	96 (11) 57-100	0.87
Gait pattern	92 (9) 77-100	0.56
NS-GP	93 (7.5) 75-100	0.93
NS-total	97 (5.1) 87.5-100	0.95



### 5.5.2.3 Bland & Altman method

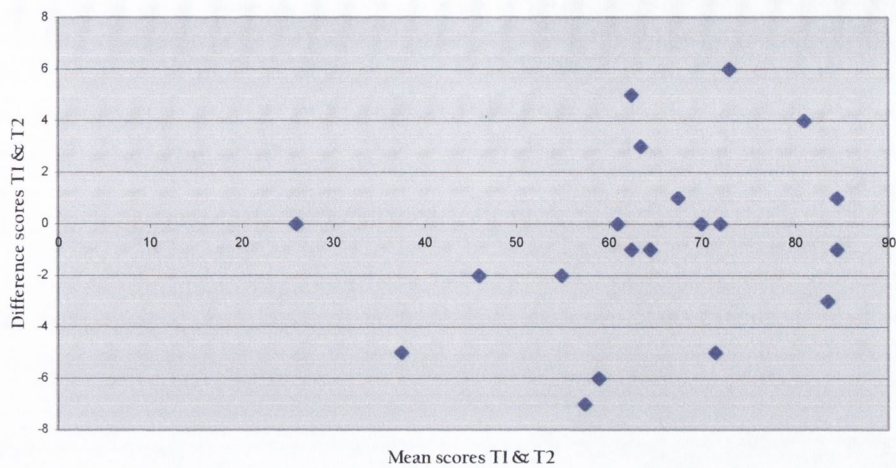
For the total score of the new SOM- NS-total and the total score without the gait pattern sub-section- NS-GP, the differences between the measurements taken at time 1 (T1) and time 2 (T2) were calculated, as was the average of the two sets of measurement. These two variables were plotted against one another. This gives a visual representation of the size of each difference and the distribution around zero – Figures 5.29 and 5.30. If no difference existed between the two sets of measurements, all the points would lie along 0. These figures also allows the observer to note how the differences may relate to the mean i.e. is more error likely to occur at a given point with the measurement range.

Figure 5.29 Bland & Altman plot for NS-GP



95% limits of agreement for T1 & T2 are -5.7 to +6.2

Figure 5.30 Bland &amp; Altman plot NS-total



95% limits of agreement for T1 & T2 are -7.5 to +6.2

The standard error of the measurement was calculated for both NS-GP and NS-total, thereafter the minimal detectable change at the 95% confidence level was calculated as described by Stratford (2004)-  $MDC_{95} = SEM * 1.96 * \sqrt{2}$ . Hence for the NS-GP and the NS-total the change required between repeated measures to ensure that the change can be attributed to more than simply measurement error is 5.5 and 7.3 respectively.

#### 5.5.2.4 Internal consistency

Table 5.13 is an inter-item correlation matrix, from this information, Cronbach's alpha was measured to investigate the degree of internal consistency within the scale -  $\alpha = 0.783$ .

Table 5.13 Inter-item correlation matrix

	BR	STS	Stand	R&L	B&R	RSTS	GA	GP
BR	1.00	.377	.223	.298	.332	.227	.248	.334
STS		1.00	.556	.613	.665	.811	.541	.548
Stand			1.00	.643	.513	.589	.580	.705
R&L				1.00	.743	.589	.591	.669
B&R					1.00	.660	.597	.619
RSTS						1.00	.507	.509
GA							1.00	.818
GP								1.00

## 5.6 Summary of results

This section comprises a brief summary of the results of this work. Where appropriate comparisons are made between the data sets.

### 5.6.1 Surveys

Table 5.14 Survey findings- 1998 & 2003

Observation	1998	2003
Use of SOM	Mobility- 54.5%	Mobility- 100%
	Balance- 31.8%	Balance- 100%
	Parkinson's-9.1%	Parkinson's-66.7%
	Stroke-36.4%	Stroke-60%
	Exercise tolerance- 30.8%	Exercise tolerance-60%
	Gait disorders-14.3%	Gait disorders-14.3%
Scales/instruments reported as utilised	Elderly Mobility Scale	Elderly Mobility Scale
	POAM- gait & balance	POAM- gait & balance
	Communication chart	Berg Balance Scale
	Get Up & Go Test	Timed Up & Go Test
	Functional Reach	Functional Reach
	Therapy outcome measures	Rivermead Mobility Index
	Webster Scale	Webster Scale
	Hoehn & Yahr	Hoehn & Yahr
	Motor Assessment Scale	Motor Assessment Scale
	Motor Club Assessment	Motor Club Assessment
		Barthel Index
		Functional Ambulation Category
		Falls Efficacy Scale
		9 hole peg test
	Motricity Index	
	Range of motion to pain	
	Unified PD Rating Scale	
	Six minute distance test	

		Shuttle test Rate of perceived exertion Visual analogue scale
Barriers-frustrations	Lack of time Issues with measurement properties e.g. repeatability, standardisation, sensitivity etc.	Lack of time, knowledge, professional consensus, administrative support, resources. Low priority None meets clients needs
Confidence in the use of SOM		Lowest levels (25%) about creating own scale, knowing measurement properties to chose scale, comparing scores across baselines and linking OM data with other characteristics. Highest levels in having a clear purpose for measurement, operational issues such as scoring and tracking progress.

### 5.6.2 Requirements of an outcome measure

A number of activities were completed to ascertain information that would inform the development of the new SOM. The output of the first survey suggested that common themes existed within the answers to questions that asked PT's what should be included within an assessment of various domains. The detail included in a new SOM should help in the creation of problem lists and should be responsive- hence the inclusion of sections dealing with compensatory strategies and higher level timed aspects of the majority of sub-sections. A number of questions remained unanswered after the final content validity exercise, to be revisited after the review of reliability.

Table 5.15 Outstanding questions from content validity methodology

<i>BED RISE (BR)</i>		
Should assistance for the trunk/lower extremity be combined with upper extremity?	Action-equivocal response. Combine & review after analysis.	A review of the BR scores does not suggest any pooling of data points around the score of '2'. This would suggest that this section should remain combined.
<i>REPEATED SIT TO STAND (RSTS)</i>		
Is the gap between 1 and 2 too much  This domain is to consider lower extremity endurance hence no upper extremity use is ideal. Is this acceptable?	Action-while the function is to consider LE strength, additional sections were included with a view to reviewing after further evaluation. This was also in view of the poor response to this question Yes-5	Additional items were placed in final version, 1-4 allowed the participant to complete the task with use of the arms, in different times. All potential scores from 0-4 were applicable to a proportion of the sample, suggesting that the use of UE's should be retained. The inter-item correlation matrix indicates that RSTS does not have strong relationships (r=0.23-0.66) with the other items in the scale, apart from STS and that its relationship to this item is 0.88. RSTS requires repeated activity as opposed to one single activity and may mimic other functional tasks requiring repeated sub-maximal muscle contractions and as such should be retained in its final format.
<i>GAIT ABILITY (GA)</i>		
Is 4 acceptable? Or should it be subdivided?	Action-equivocal response. Retain isq and review after next evaluation.	As with Bed Rise, there was no clustering of the participants' scores at this item. Contact guard assistance, supervision and verbal cueing all require a physical presence of another person and as such may not represent a relevant difference to the person who requires this help and the person providing it. Thus, it would appear to make sense to retain the item in status quo.

*GAIT PATTERN  
(GP)*

<p>On the basis of inclusion of all the previous sections, does this section need to be included or should it be optional? If it is removed, do all the sections need to be included? The respondent was asked to indicate the section they considered should be removed, if they answered 'remove sections' to this answer.</p>	<p>Action-equivocal answer. Retain and review in future analysis</p>	<p>See section 5.6.3 below.</p>
--	--	---------------------------------

## 5.6.3 Measurement properties-comparison between NS-total and NS-GP

Table 5.16 lists the majority of the measurement properties evaluated in this study.

Table 5.16 Measurement properties

Property	Analysis	NS-total	NS- without GP
Concurrent validity	Relationship with	EMS- 0.78	EMS-0.80
	EMS, POAM-G,	POAM-B- 0.72	POAM-B-0.86
	POAM-B using Spearman's rank correlation coefficient	POAM-G-0.72	POAM-G-0.61
Predictive validity	Relationship with Discharge EMS,	Admission scores EMS-0.70	Admission scores EMS-0.65
	POAM-G, POAM-B using Spearman's rank correlation coefficient	POAM-B- 0.80	POAM-B-0.69
		POAM-G-0.80	POAM-G-0.44
Discriminant	Difference between community & active rehabilitation group. Mann-Whitney.	Able to discriminate	Able to discriminate
Responsiveness	Effect size EMS=0.36 POAM=0.52	0.51 > EMS 0.51 < POAM	0.64 > EMS and POAM
	Relative efficiency compared to EMS and POAM	EMS- 1.82 POAM- 1.56	EMS- 1.75 POAM- 1.50
Inter-rater	% agreement &	95.4-96.6%	97.1-98.4%



reliability	weighted k coefficient	0.94-0.97	0.98-0.99
	Bland & Altman plots	Majority of points within limits of agreement	Majority of points within limits of agreement
Intra-rater reliability	% agreement & weighted k coefficient	97% 0.97	93% 0.93
	Bland & Altman plots	Majority of points within limits of agreement	Majority of points within limits of agreement
Internal consistency	Inter-item correlation coefficient	GP correlates with the other items in the scale, ranging from 0.3-0.8.	

Sections highlighted in red suggest a more successful outcome, albeit in some cases the difference is marginal. Physiotherapists who participated in the face and content validity suggested that the two most important functions of an outcome measure were that it would measure change (responsiveness) and that it would inform the development of a problem list. If the results of Table 5.16 are considered, then the new SOM may include the item on gait pattern, but could be optional for the individual PT, department or service, since the measurement properties without it are also robust.

## CHAPTER 6      DISCUSSION

### 6.0 Introduction

This chapter contains three sections, 6.1 a discussion of the surveys of practice, 6.3 a review of the item selection and evaluation as part of the development of a new standardised outcome measure (SOM) and 6.5 a consideration of validity and reliability. After each section, a conclusion is presented.

### 6.1 Surveys

Two surveys were performed to consider the practice of physiotherapists working with older people in Ireland, specifically, in the context of their use of outcome measurements. Two different methods of data gathering were employed- a structured interview and later a postal questionnaire. The former was considered more appropriate in 1998 since, it was still the early stages of the use of standardised outcome measurement (SOM) in physiotherapy (PT) practice and this method of retrieval of information enabled any confusion about terms to be clarified. This also attempted to optimise the participation rate. The second survey of practice took place in 2003; on the same sample of same physiotherapy departments and this time took the form of a postal survey. This allowed comparison with the only other published comparative review of practice (Kay *et al.* 2001). Senior physiotherapists were invited to participate in the surveys. These PT's represent the specialists within the field of physiotherapy rehabilitation of older people and as such are the professional practice leaders. Their individual practice frequently reflects that of the members of physiotherapy staff in their departments and so the results of these surveys can be considered representative of practice in Ireland in dedicated departments of rehabilitation for older people. The participation rates of 88% and 75% are good although it is possible that those PT's who did not respond were not using SOM and that some bias may exist within the results.

#### 6.1.1 Use of standardised outcome measures

The extent to which SOM are used by physiotherapists (PT's) working with older people in Ireland has increased in the five-year period from 1998 to 2003, as has the range of SOM employed. 100% of the participants now use a SOM in dealing with

mobility and balance problems, compared to approximately 30-50% in 1998. To date only one other comparative survey of practice has been reported in the literature (Kay *et al.* 2001). Due to the contradictory nature of some responses in that study, the authors conclude that it is difficult to identify 'whether or not overall use of client outcome measures in clinical practice has increased since the early 1990's'. Nevertheless, the same authors also employed a focus group methodology in five Toronto centres and the results of this suggest that the idea of using outcome measures as an intrinsic part of PT practice is now widely accepted (Huijbregts *et al.* 2002). This increase in the use of SOM's in Canada was not considered to be unique to physiotherapy and reflects the broad restructuring of the health care system in the 1990's. This is similar to the two health strategies which have been introduced in Ireland since the 1990's – *Shaping a healthier future: a strategy for effective healthcare in the 1990's* (1994) and *Quality and Fairness- a health system for you* (2001). Both strategies include, among other principles, those of accountability and quality of service.

The results of the 1998 survey in Ireland were consistent with other such surveys in the 1990's. Cole *et al.* (1995) reported on practice in Canada and noted that although 50% of PT's reported using a SOM, 20% identified only one instrument. Chesson *et al.* (1996) complete a survey of practice in Scotland in 1992 among OT's and PT's and, while the methodology was slightly different to the Canadian survey, 44% of PT's reported using SOM. In this survey the author noted that SOM were more commonly used in 'care of the elderly' than any other speciality. Although the use of SOM's for mobility was 54.5%, this was lower in other areas. Approximately 1 in 3 of the participants used a standardised method of measuring balance performance and a disease specific measure for stroke. In 2003, the use of SOM in Ireland had certainly increased but with regard to disease specific measures, for example, stroke and Parkinson's disease/parkinsonim one third of PT's do not use a standardised method of measurement. While the prevalence of Parkinson's disease in Ireland is not known, the European prevalence is estimated to be 1.6 per 1000 (National Council for Ageing & Older People, 1998). Stroke is reported to be the leading cause of acquired disability in adults and its prevalence increases with age (Irish Heart Foundation, 2000). Torenbeek *et al.* (2001) report that 75% of respondents to their survey, who worked in stroke

rehabilitation, used SOM's. Of the measures reported to be in use in stroke rehabilitation, only one was a stroke-specific measure, suggesting that change in this area is gradual.

Similar numbers of Pt's do not employ a standardised instrument in the area of gait disorders and exercise tolerance. This is notable since, one of the barriers to achieving functional independence noted by PT's who participated in another part of this research work, was the area of fitness/deconditioning. However, the SOM's available to measure exercise tolerance, which are appropriate for use with older people with functional limitations, are limited and this may explain why 1/3 of participants do not use them. Nevertheless, it is likely that there are a proportion of older people who are being treated by PT's in Ireland for whom no standardised method of measuring their response to intervention is employed.

The range of instrument employed has increased since 2003. In the case of mobility, the most common instrument remains the Elderly Mobility Scale (Smith, 1994). Two other instruments that specifically report on mobility status are the Functional Ambulation Category and the Rivermead Mobility Index. Only one respondent used the Performance Oriented Assessment of Mobility -Gait subscale (POAM-G) as the sole standardised method of measuring mobility; all others report using a number of different SOM's. With respect to balance, the two most commonly used methods are now the Performance Oriented Assessment of Mobility-Balance subscale (POAM-B) and the Berg Balance Scale (Berg, 1989). Falls efficacy, the self-reported confidence in carrying out activities without falling, is used by 27% of participants, reflecting the aspiration to capture the impact of falling which is beyond physical capabilities. One unidimensional instrument is employed - Functional Reach. The POAM-G is the only standardised method reported which is used for gait analysis. In stroke rehabilitation, three stroke-specific instruments are used- the Motor Assessment Scale, the Motor Club Assessment and the Motricity Index. Range of motion (ROM) to the point of reported pain was reported by 7% of respondents (n=1) and is not actually a SOM, since while ROM has reported measurement properties, its use in the context of pain in the shoulder post stroke has not been researched for validity or reliability. The Hoehn & Yahr staging reported by 26.6% of respondents is designed for use in Parkinson's disease but it not designed to measure change in response to physiotherapy treatment. Of the three SOM's reported for use in Parkinson's

disease, the Unified Parkinson's Disease rating scale (UPDRS) is the only one which has a structure that would be suitable for physiotherapy intervention.

In 1998, when the first survey of practice was initiated, the questions sought to identify if SOM's were being utilised among physiotherapists. At this time, both from within the profession and as a result of external trends within the health service, the incorporation of SOM into practice was desirable. This survey simply asked the question 'are SOM being used?' and it did not consider the nature of how SOM were being incorporated into practice. In 2003, the results of the survey demonstrate that for the majority of SOM employed, they are used at admission and discharge or indeed more frequently. Ten to twenty percent of respondents reported using SOM at admission only. The degree of attrition in the use of SOM - from admission to discharge - is less than that reported in the literature, albeit that the studies reported by Turner *et al.* (1996, 1999) employ a different method for data gathering. The authors employed retrospective chart retrieval and the measures they investigated relate specifically to measures of impairment - pain, muscle strength and joint range of motion. The drop-off in the use of SOM's for these aspects of intervention was from 21% to 2.5% for pain and from 64% to 10% and 30% for muscle strength and range of motion respectively. This 'attrition' in the use of SOM's through the course of a treatment cycle is also reported by Kirkness & Korner-Bitensky (2002) and Kay *et al.* (2001) and suggests that while SOM's are used as part of PT management, the information they yield is yet to be fully incorporated into practice and decision-making.

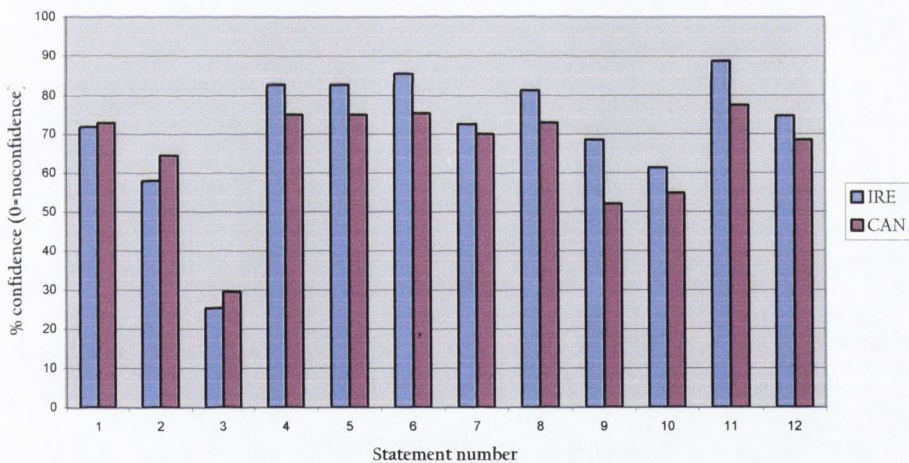
A review of self-efficacy<sup>1</sup> in the use of SOM's reported by Irish PT's suggests that overall, they have a clear purpose for measurement - mean confidence 89% with a standard deviation (sd) of 11% and a range of scores from 70-100%. They are less confident about what to do with the information (mean 75%, sd 26%, range 0-100%). When it comes to the operational issues, they know what to measure for their client groups, how to administer and score the relevant SOM's and how to track progress. For these activities, the mean confidence reported was 70-85%, with sd's of 8-12%. Confidence is less absolute when it comes to using the generated SOM scores in a meaningful way (statements 9 and 10) and in the areas of knowledge about measurement properties (statements 2 and 3). This may reflect the fact that there are few continuing professional development courses in Ireland

---

<sup>1</sup> Self-efficacy measured as 0% no confidence, 100% total confidence.

available on the use of SOM's in practice and during undergraduate training that no such modules were included in the curriculum. Furthermore, the Irish Society of Chartered Physiotherapists has not engaged in any strategic exercise to optimise the systematic use of SOM's in physiotherapy practice. It may also reflect the lack of an outcomes database to gather and use the output of SOM's both at local and national level. Torenbeek *et al.* (2001) notes that of the five European countries included in their study, statistical use of outcomes data after they are gathered (in stroke and low back pain) occurred in 17-90% of cases, with Ireland reporting the lowest level. Figure 6.1 illustrates the level of confidence between the results of a Canadian sample (Kay *et al.* 2001) and the Irish survey – the statement number on the X-axis relates the statements previously discussed in Section 5.2.4. Given the difference in the attributes of the samples (rehabilitation of older people (Ireland) and all aspects of PT (Canada)), there appears to be little difference of note between the levels of confidence reported.

Figure 6.1 Confidence levels



### 6.1.2 Frustrations and barriers

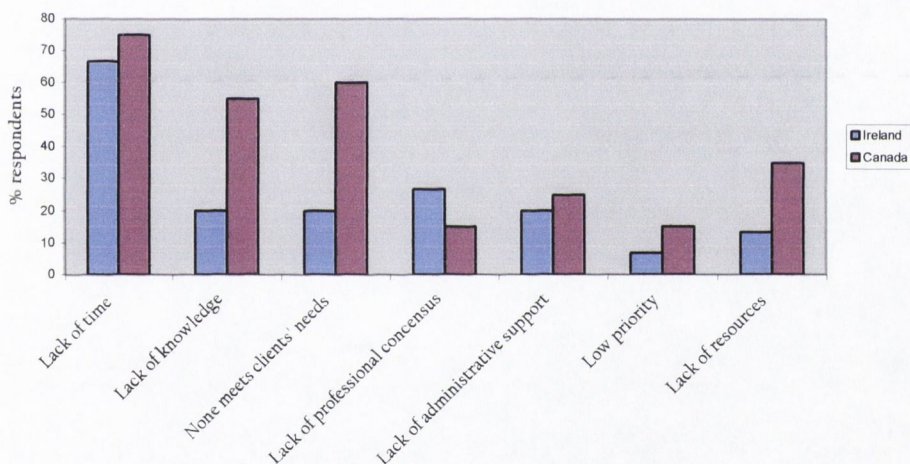
In the first survey of practice, participants were asked to identify any frustration they experienced with their current method of measurement- standardised or not. The responses can be allocated to two main themes-

- Issues with the measurement instruments/methods

- Operational issues

The former represents the responses that describe frustrations with objectivity, standardisation, sensitivity and repeatability. The latter clearly suggesting that time and staffing were implicated in barriers to the use of SOM. The later survey, in 2003, explored in more detail barriers that existed to the use of SOM's- time remains a significant issue (65%), with problems that could be described as 'local' being administrative support (20%) and resources (20%). 20% of respondents noted that no outcome measures met their clients' needs while more strategic issues which should be of interest to clinical interest groups and the ISCP is the issue of the lack of professional consensus in the choice of SOM and lack of knowledge. Figure 6.2 compares the responses of the Irish survey to a similar question asked by the Canadian Physiotherapy Association (Kay *et al.* 2001)

Figure 6.2 Barriers to use of SOM



These results compare the professional practice leaders (PPL's) in Canada and their equivalent in Ireland but the Canadian sample includes all areas of PT practice. Nevertheless, time appears to be a universal obstacle to the use of SOM's. Huijbregts *et al.* (2002) report that when discussing the issue of time, the Canadian PT's in their focus groups, clearly identified time as relating to the identification of a suitable measure, its administration and scoring, and thereafter the interpretation of the results and not simply the use of SOM's with patients or clients. The PT's in this Canadian study also note that where an organisational mandate exists to use

SOM's, their utilisation is enhanced when there is expertise and support available at departmental level. These results identify the need for the investigation of methods that will minimise the time issues associated with the use of SOM. Novel software solutions, which aid in the interpretation of individual patient scores and are linked to the specific measurement properties of the SOM being employed, could be developed. Clinical guidelines, based on professional consensus, and with appropriate information about measurement properties, which suggest SOM's for use in specific client groups could be developed in areas of common practice.

## 6.2 Conclusion

The surveys carried out in Ireland suggest that change has been occurring in the practice of physiotherapy measurement. This has occurred against a backdrop of reviews of practice in physiotherapy in Canada, the UK, and the US; the increase in the activities of some professional organisations throughout the world to optimise the systematic use of SOM's in routine practice; and within Europe a recognition by the professional organisations that *best practice* requires the use of 'a published, standardised, valid, reliable and responsive outcome measure' (European Region WCPT, 2002). The review in Ireland focussed on one specific area of physiotherapy practice, namely, that relating to services for people over the age of 65 years, and the results of this study cannot be applied beyond this area. Nevertheless, it is likely, given the national demand within health care services for increased accountability and the inherent recognition within the profession that outcome has to be quantified, that the systematic use of SOM's is increasing in practice in general.

The results of the survey of Irish PT's, in combination with findings of other studies suggest that a number of approaches may need to be taken to increase the meaningful use of SOM's in physiotherapy practice. These include:-

- Outcome measurement courses at undergraduate level to inculcate knowledge about interpreting reported measurement properties with a view to choosing an appropriate measure and using its output for clinical decision making.
- Similar courses at professional level.
- Developing professional consensus about what SOM's to use. This requires panels of experts to identify a review methodology



for instruments relevant to the clinical field or client in question. Following a review of the relevant instruments, a number of SOM's may be recommended for use within the area of practice. This review could form part of a set of clinical guidelines or could exist as a stand-alone document.

These suggested actions might relieve some of time pressures reported by PT's as being a barrier to their use. They also may improve the confidence of PT's in their choice and use of SOM's in practice.

- Technological solutions using artificial intelligence paradigms and case-based reasoning may also be a method of optimising the use of the output of SOM's. In the future it is likely that personal digital assistants (PDA's) may become part of e-health solutions to health care planning and decision. Software programmes which support interpretation of the results on SOM's may enable PT's to use the information generated by SOM's to plan interventions and predict outcomes, based on their measurements.

### 6.3 A review of the item selection and evaluation as part of the development of a new standardised outcome measures (SOM)

The methods used to consider face and content validity in this study were two-fold. In the initial stages, information was sought, in a broad sense, from PT's about what should be included in an assessment of various domains in physiotherapy and two specific diagnostic groups- stroke and Parkinson's disease/parkinsonism. The information could be categorised as impairments of structures or functions and disabilities/limitations of activities. The extent to which functional activities such as transfers, sitting to standing, walking ability, stability in the context of movement, were represented across the physiotherapy domains<sup>2</sup> and diagnostic groups<sup>3</sup> supported their inclusion in the first version of the new scale. While aspects of fitness/deconditioning/exercise tolerance were recognised as a barrier to achieving functional independence, it was considered beyond the scope of the development of this new SOM.

---

<sup>2</sup> Mobility, balance, gait, exercise tolerance

Bed rise or lying to sitting was included since difficulty with this task is a problem that can affect a proportion of older people and may limit independence. It has been reported to be a problem for 6-8% of community-dwelling older people and at least 63% of older people in residential care (Alexander *et al.* 2000). Normal bed rise (flat bed with 1-inch firm padding) is characterised by transient use of the upper extremities, almost simultaneous trunk flexion and lower limb movement off the bed, a pivot on the hip and the feet clearing the surface. Normal rise times have been reported for community dwelling older people (Alexander *et al.* 1992) and these were incorporated into the higher end of the scoring scheme. This task requires trunk strength and co-ordination of movement; and performance (movement and time) on this task has been shown to differ between community-dwelling older people and those living in residential care (Alexander *et al.* 2000). Bofelli *et al.* (1996) observed that performance on a scale to measure bed rise difficulty could be poor despite high scores on other scales of functional ability suggesting that performance on this task might be able to detect early and/or mild disability. During the development stage of the new SOM, there were two opportunities to identify whether items should be retained. When the first iteration of the new SOM was being considered, the participating PT's were asked to rank the importance of the inclusion of each item, with 0 being 'not at all important' and 5 being very important. The median score for this item was 5 indicating the PT's considered it an important item to retain. On reviewing the second iteration of the new SOM, 9 of the 13 participants felt it should be included. Consistent with Bofelli *et al.* (1996), there was not a strong association with bed rise performance and the other activities in the new SOM. This item may enable subtle changes in performance to be observed and it provides information deemed to be important by PT's for inclusion in a scale of functional abilities.

**Sit-to-stand and repeated sit-to-stand** were both included in this scale, since the ability to rise independently from a chair is fundamental to functional independence. Analysis of the performance of this task, as well as the measurement of ability, can indicate the source of factors that may be limiting independence (Schenkman *et al.* 1996). The time taken to perform a single chair rise correlates with the power in the extensor muscles of the lower limb and is also reported to correlate with walking speed in older men and women (Bassey *et al.* 1992).

Although balance could be considered to have a role in chair rise, Schenkman *et al.* (1996) investigated the relative importance of both strength and balance in chair rise and reported that lower extremity strength is the stronger predictor of success. Repeated chair rise was included as a measure of power in the lower extremities. Muscle power can be measured using an isokinetic dynamometer and these are increasingly available in physiotherapy departments, however the use of functional activities to estimate this parameter may be more informative for functional performance (Schenkman *et al.* 2002). On both occasions, when asked, PT's considered this to be a very important item for inclusion (rank of 5, 13/13 suggest inclusion).

**Bend & reach and lift & bend, standing ability & stability-** In this study, stability was reported as being one of two main barriers to the achievement of functional independence. 'Balance' was reported as being important to include in assessment of mobility, balance, gait, stroke and Parkinson's disease/parkinsonism. Thus, some method of quantifying stability needed to be included in the new SOM. These components represent two aspects of the control necessary for functional balance- balance in a steady state and anticipatory postural control (Shumway-Cook & Wollacott, 2001). The third aspect of postural control is ability to withstand perturbation. Asking a patient to withstand a 'nudge' traditionally tests it. The difficulty in this regard is the replication of the extent of the 'nudge'. The three aspects of balance activity included in the new SOM received median scores of 4.5-5 in the first evaluative exercise. Thereafter, the agreement to include all three was also unanimous. They all correlate with one another but not to the extent that one item should be excluded.

**Gait ability** aims to capture not simply information about ability but also speed. Reference values exist for both men and women who are independent and ambulatory (Bohannon *et al.* 1996). Comfortable walking speed may be influenced by variables such as calf strength and health problems (for men) and calf strength, leg pain and customary activity (for women) (Bendall *et al.* (1989). Use of a walking aid such as a cane has been associated with gait speed that is significantly faster than walking with a frame (Bassey *et al.* 1992). In this study, those older people using walking frames walked 3.6 times slower than those who were independent or used a cane. Unsurprisingly, gait ability correlated highly with gait scores on the gait pattern component of the new SOM ( $\rho$  0.82) and also strongly with other

functional aspects of the new SOM. Its inclusion was considered 'very important' by all PT's in the first evaluation of the SOM and all who answered (12/13) suggested it be retained in the second evaluation.

The retention of the **gait pattern** subsection of the new SOM was clearly identified as 'very important' by all who participated in the first evaluation of the new SOM. For the second evaluation, the respondents were asked to state whether this item should be optional. 10 respondents answered the question, 5 in favour of it being optional and five against. The item was retained and its final inclusion is discussed below.

As discussed in Chapter 2, no definitively agreed methods exist to consider the issues of item selection and evaluation in the literature. For the most part, the methods used employ groups of users, generally expert users, to enable professional consensus. The new SOM was compared to the most commonly used SOM's at the time of completion of the survey in 1998- the Elderly Mobility Scale (EMS) and the Performance Oriented Assessment of Mobility (POAM). The methods used for the generation of these scales were reviews of the published literature. No professional judgement of the items or scoring was sought. This study used a combination approach to the development of the new SOM -review of literature, clinical expertise, and requirements expressed by potential users, a review of included items by experienced users and finally a review of measurement properties. Increasingly this type of methodology is reported in the literature (Powell & Myers 1995, Platt *et al.* 1998, McIntosh *et al.* 1999) and the greater the involvement of the end-users in any design procedure, the more likely their satisfaction with the final product.

The items in this new SOM were designed for use in the clinical setting, with a view to combining the scores on the individual components, to yield a composite score. The World Health Organisation International Classification of Functioning, Disability and Health (2001) aims to provide a unified and standard language and framework for the description of health and health-related states. It notes that for the quantification to be used in a universal manner, assessment procedures need to be developed through research (WHO, 2001 p22). The output of the individual components of the new SOM could be used as a means of applying the WHO performance qualifiers. Broad ranges of percentages are provided for those cases where standards are available to quantify the limitation. 'Moderate

problem' is defined as up to half the scale of total difficulty. The items in this new SOM could be used to inform the following activities described in the ICF chapter on mobility:-

- d410 Changing and maintaining body position
- d430 Carrying, moving and handling objects
- d450 Walking and moving.

## 6.4 Conclusion

Item selection for the new SOM was informed by a number of different sources. The aim of using a variety of sources and a number of iterations was to optimise the design of the version of the scale for final evaluation in the context of this study. This scale was designed primarily for use by physiotherapists in Ireland but there is nothing to prevent its use in other English-speaking countries or indeed for translations to be employed for its use in other countries. There are no explicit cultural indicators within the new SOM that would prevent this from occurring. One potential international application for this instrument is how performance of the various items or subsections within the scale, could be used in the application of the World Health Organisation ICF indicators of activity limitation. Further work would be required to ensure the reliability of the application method both between and within raters.

## 6.5 Measurement properties of validity and reliability

### 6.5.1 Criterion validity-predictive and concurrent validity

As discussed in Chapter 2 (2.1.3 p12), traditionally criterion validity considers the performance or accuracy (Feinstein, 1987) of the new instrument by comparing it to a *gold standard*. The level of agreement between the two sets of measurements is then examined, using an 'appropriate indicator of agreement' (McDowell & Newell, 1996). It can be either concurrent or predictive. In rehabilitation 'gold standards' rarely exist, and accepted practice is to compare the new SOM to existing SOM's. In the case of the POAM, it has been used as the 'gold standard' in the evaluation of concurrent validity of the FEMBAF, Bed Rise Difficulty (BRD), SturdyGrip and a number of measures of postural sway (Lichtenstein *et al.* 1990, Thapa *et al.* 1994, Bofelli *et al.* 1996, DiFabio & Seay 1997, Cooper & Stewart 1998, Stokes *et al.* 1998). It is widely used as an outcome measure in rehabilitation research. The Elderly

Mobility scale was also utilised since it was the only relevant standardised outcome measure used in practice in Ireland at the time. The relationship between NS-total and NS-GP scores on admission were strongly related to discharge scores as measured by the EMS (0.70 and 0.65), suggesting that the new SOM can be used to predict mobility status. Equally, the relationship was strong between the admission scores on the new SOM and the balance subscale of the POAM (0.80 and 0.69), suggesting that admission scores are predictive of balance performance. Unsurprisingly, the association between NS-total on admission and the performance at discharge on the POAM-G is stronger (0.80) than the relationship between the NS-GP and POAM-G (0.44), suggesting that without the gait pattern subsection, the new SOM is not predictive of gait performance on discharge. Obviously this provides no information about gait ability, which is a separate section in the new SOM and not featured in POAM-G. Concurrent validity explores the relationship between measures on the NS-Total and NS-GP with the EMS and POAM-B and POAM-G and has been discussed previously.

#### 6.5.2 Discriminant validity

Among the healthy community dwelling older people, all participants attained the total scores in bed rise, sit-to-stand, bend and reach, lift and reach. 17 attained the total score in gait pattern, one participant scored 39/40. 13/18 attained the top score in standing ability and stability. The task that proved most difficult was tandem stance with the eyes closed. It is possible that performance on this task would improve with practice but this was not allowed in this trial. One participant (aged 78 years) scored two points below the total score, with slowness in repeated sit-to-stand and gait speed being the areas where he did not attain a high score. It is unclear why these scores were lower than those of his counter-parts. Using timed data allows fine discrimination to occur in high functioning older people. Small changes in performance in older people who themselves report no difficulty have been shown to be predictive of changes in mobility status in a two-year follow-up period (Pine *et al.* 2002). It is possible that scores at the high end of this scale could be used to screen for slight changes in performance that require further evaluation to initiate a programme of exercise which may prevent later mobility problems.

### 6.5.3 Responsiveness

Many of the frustrations noted by participants in the first survey of practice in 1998 related to the sensitivity of the methods employed- both standardised and non-standardised. When asked what they thought the attributes of an outcome measure should be, it clearly emerged that an outcome measure should measure change. For a SOM to capture change, it should not be disposed to a 'floor' or 'ceiling' effect. The developer of the POAM suggested that the POAM and other tests of functional balance 'are not sensitive enough to pick up the changes' due to interventions (Tinetti, 1993) and this may be due to the range of scores available within the scale. Despite this the POAM is widely used in rehabilitation research as an outcome measure, lending support to its responsiveness, although no formal evaluation of this has been reported. In reviewing the EMS, no report has been made of its responsiveness, and within its construction, there are perhaps some reasons why this property might be limited. For example, scoring within the section which evaluates *sit-to-stand* moves from a score of 1 (needs help of one person), which could range from supervision to physical assistance to a score of 2, which is independent in greater than 3 seconds. This scoring system may not be able to reflect relevant clinical change in performance of the function of sit-to-stand. The methods used to evaluate responsiveness of the new SOM are effect size and relative efficiency. Effect sizes for the new SOM were greater for the version without the gait pattern item. The effect size of 0.64 suggests that the new SOM will be moderately sensitive to change within a group of older people participating in rehabilitation. In this preliminary review of the measurement properties of the new instrument, it was more responsive to change than the EMS (effect size 0.36) and the POAM (effect size 0.52). Including the gait pattern subsection will lessen the new SOM's responsiveness and this is explained by the relatively smaller range of scoring options in this item, which is essentially dichotomous i.e. normal/abnormal. The second method of analysis performed was an investigation of the relative efficiency of the new SOM compared to the EMS and POAM. One of the constructs (construct validity 6.5.4) established prior to the evaluation of the new SOM was that it would be at least as responsive, if not more so, than the commonly used SOM's. Both versions of the new SOM were more sensitive than the EMS and POAM.

Responsiveness to change is a property of outcome measures that was often neglected in the development and evaluation stage of a new method of measurement. Many instruments entered practice without any clear indication of the important property (Cole *et al.* 1995, Liang *et al.* 1995). More recently, the description of the development of new instruments often includes a report of this property and SOM's with moderate to high levels of responsiveness are used in physiotherapy practice (Mao *et al.* 2002, Trombly *et al.* 1998). On a day-to-day basis what may be of more relevance to individual practitioners is the  $MDC_{95}$  – the minimal detectable change and this is discussed below (6.4.6 reliability). The ability to measure change due to intervention is important to ensure effectiveness, quality and accountability. Measures that are insufficiently responsive may not capture change in functional status regardless of how effective the intervention may be. Equally, changes may be reported in the status of individuals or groups of patients that are nothing more than measurement error. The ability to appropriately interpret change scores must be informed by the measurement properties of the SOM.

#### 6.5.4 Construct validity

One of the constructs investigated in this study of the measurement properties was that the new SOM would display a strong association with two commonly used extant SOM's- the EMS and the POAM-B and POAM-G. McDowell & Newell (1996) report that a correlation of 0.6 between two measures represents an 'extremely strong association'. They and Helmstadter (1966) also describe how when the output of two measurements is compared, the maximum possible correlation between them is the square root of the product of their respective reliabilities. Hence, the results of the comparison of the NS-GP and the NS-total with the EMS and the POAM-G and POAM-B, should be considered in light of the above. Reliability coefficients are not reported for the EMS in the first publication on the instrument, however in a paper by Prosser & Canby (1997) the reliability coefficient for inter-rater reliability is reported as 0.88. A similar coefficient was generated, for the purposes of this analysis, for NS-GP (0.99) and NS-Total (0.98). Hence, the maximum correlation possible between the EMS and NS-GP and NS-total is 0.93. The results of this study illustrate that there is a strong relationship between NS-GP and the EMS (Spearman's rho 0.80) and between NS-Total and



EMS (rho 0.78). In the initial work on the POAM (Tinetti 1986), the measures of reliability used were percentage agreement and kappa coefficients, and in the few publications on its reliability (Cipriany-Dacko *et al.* 1997, McGinty *et al.* 1999, Roller *et al.* 2001) none report a Spearman's rank correlation coefficient, therefore it is not possible to calculate the maximum correlation possible between the new SOM and POAM-B and POAM-G. Nonetheless, the relationship between NS-GP and NS-Total and the POAM-B is strong, rho-0.86 and 0.82. The relationship between the two versions of the new SOM and POAM-G is less strong 0.61-0.72 but still in the region described by McDowell & Newell as a strong association.

This method of relating the performance of a new SOM against another set of measures, during the development phase of a new SOM, is widely reported in the physiotherapy literature. There is no consensus about which measures the new one should be compared. Podsiadlo & Richardson (1991), in the development of the timed Up & Go, compared it to the Berg Balance scale (-0.81), gait speed (0.61) and the Barthel Index (-0.78). Collen *et al.* (1991) compared the new Rivermead Mobility Index with the Barthel Index (0.91), the Functional Ambulation Category (0.89), and the six-minute distance test (0.63). In this study, the choice on what to compare the new measure against was the most commonly used SOM's in Ireland at the time, namely the EMS and the POAM.

A second construct was that the new SOM would be equally as responsive as the standardised outcome measures in common use. This has been demonstrated. The third construct suggested that the items in the new SOM would be explained by one factor or domain. Using principal components factor analysis, one factor emerged containing all the items with the scale. The factor analysis identified only one item with an eigenvalue of greater than one, illustrated in the scree plot (5.4.4.1 p128). This illustrates the unidimensionality of the scale and as a method to investigate the number of domains or as an item reduction exercise has been reported in the development of the Hierarchical Assessment of Balance and Mobility (MacKnight & Rockwood 2000), the Physical Therapy Outpatient Satisfaction Survey (Roush & Sonstroem 1999).

### 6.5.5 Reliability- inter-rater, intra-rater and internal consistency

If an instrument is not shown to be reliable, it is of limited use in clinical practice. Both versions of the new SOM were evaluated for reliability. The methods employed have been reported for the same purposes in the physiotherapy literature. Percentage agreement and kappa or weighted kappa coefficients have been used to evaluate the degree of reliability for many outcome measures with similar ordinal outputs- POAM (Tinetti, 1986), the Berg Balance Scale (Berg, 1989), the Postural Assessment Scale for Stroke (Mao et al, 2002), Ashworth Scale for tone (Pomeroy *et al.* 2000, Gregson *et al.* 2000). Other methods are also reported such as correlation coefficients (Prosser & Canby, 1997), and the Mann Whitney test (Smith, 1994) but these have limitations and were discussed in detail in Chapter 2. In addition the intraclass correlation coefficient is often reported (Lim *et al.* 2005, Platt *et al.* 1998) but this type of analysis is appropriate for data that is continuous and not ordinal (Shrout & Fleiss 1979, Eliasziw *et al.* 1994, Rankin & Stokes 1998).

The percentage agreement between three raters for both versions of the new SOM was high, 97-98% for the NS-GP and 95-96% for the NS-total. Kappa coefficients ranged from 0.94-0.99. These weighted kappa coefficients are very good (Brennan & Silman, 1992). Similar results were demonstrated for intra-rater reliability. The Bland and Altman method (Bland & Altman, 1986) for reviewing the agreement between two sets of measurements plots the mean difference between the two sets of measurements (either two different times or two different raters) against the difference of the two sets of measurements. The authors suggest that the mean difference represents the estimated bias and the standard deviations (sd) of the differences would measure the 'random fluctuations' around the mean (Bland & Altman, 1995). They recommend 95% limits of agreement should be calculated to inform the reviewer of how far apart the two sets measurements were likely to be for most individuals. A review of all the Bland and Altman plots generated for both inter and intra-rater reliability indicates that the majority of observations fall within the acceptable limits of agreement. This supports the reliability of the two versions of the new SOM.

In addition the  $MDC_{95}$  was calculated. It is this figure that is will probably be more meaningful on a day-to-day basis for physiotherapists using this scale, since it is the figure that will inform them if 'true' change has occurred rather than just measurement error. The  $MDC_{95}$  was calculated as 5.5 for NS-GP and 7.3 for NS-

total. This method of calculating the MDC has been reported for the Fugl-Meyer assessment of sensorimotor recovery after stroke (Sanford *et al.* 1993), the Berg Balance Scale (Stevenson, 2001) and the Roland Morris Questionnaire (Stratford 1996b, 1998), all of which are SOM's with ordinal data. However, this method of calculation should be considered with caution since the statistical approach employed to generate the MDC subjects ordinal data to an arithmetic function that may not be appropriate.

It is useful to compare the reliability of the new SOM with the two most commonly used SOM's in 1998. The intra-rater reliability of the EMS has never been reported. The inter-rater reliability has been reported, but both methods used are not optimal for this analysis. Smith (1994) compared the means of the two sets of measurements generated by two rater (inter-rater reliability) and noted no significant difference between their scores. Prosser & Canby (1997) reported the Spearman's rank correlation coefficient as 0.88. In the first publication describing the POAM, Tinetti (1986) reports that agreement existed for 85% of items between a nurse and a physician and where differences existed the total score never differed by greater than 10%. In order to compare the reliability of the POAM to the new SOM, reference is made to the paper published by Cipriany-Dacko *et al.* (1997), which is similar to the methodology used in this study. The authors report that the percentage agreement between their physiotherapy raters as ranging from 73-90% with kappa coefficients ranging from 0.57-0.82. It would appear that the percentage agreements for the new SOM were marginally greater than those reported for the POAM. Hence the new SOM is at least equally as reliable as one of the commonly used SOM's. There have been no publications that describe effect sizes or the MDC for either the EMS or the POAM.

## 6.6 Two versions of the new SOM

During the evaluation of content validity, it emerged that there was no consensus about whether the section that described gait pattern should be included or optional. It was decided to evaluate the measurement properties of the two versions of the new SOM with a view to making a final decision thereafter. Table 5.16 in Chapter 5 outlines a comparison between the two versions of the new SOM and it would appear that there is little significant difference between the two versions of the new SOM. As a result, it is recommended that the new SOM be presented to

physiotherapists as a 7-item outcome measurement, with an optional 8th item that can be incorporated when required. Since further review of the measurement properties is likely to be ongoing in the coming years. This attribute could be re-analysed in the future.

### **6.7 Conclusion**

The measurement properties of the new SOM are consistent with those reported in the literature for other outcome measurements used in physiotherapy practice. The analysis of the reliability of the new SOM is more informative than that published about the EMS. Both inter and intra rater reliability is evaluated. The new SOM is as reliable as the POAM. It is more responsive than both and has information about the MDC; this may be helpful in decision-making about individual patients. It was developed with the advice of and feedback from clinical experts in Ireland and as such responds to the needs they expressed. It represents a new standardised outcome measurement, the preliminary evaluation of which suggests that it will be a useful measure for use on physiotherapy practice in Ireland. In order to optimise its use, the measurement properties of each version will be presented with the new SOM to aid interpretation and use.

### 6.8.1 Survey

- In 1998, all Senior PT's in the area of Medicine for the Elderly were invited to participate in the review of practice. Their departments were invited to participate again in 2003. This sample is not a random sample of PT's working with older people in Ireland and although is likely to be representative of practice, if it were possible to identify a random sample of PT's this would have ensured it was representative.
- Postal surveys like the one used in 2003 and structured interviews that use closed questions can limit the richness of the information obtained. Focus groups enable the expression of themes not considered in structure questionnaires. The qualitative nature of the responses can provide a broader canvas upon which to review practice, barriers, confidence and many of the issues explored in the survey.
- It is possible that non-response to surveys such as those undertaken in this study occurred because PT's were not using SOM's and were not happy to participate or return a survey confirming their practice. Hence, the figures of use of SOM's may have been overestimated.

### 6.8.2 New SOM

- The sample sizes in this study are similar to those reported for preliminary development of instruments but may limit the application of the scale, since no older people in institutional care were evaluated and this is an area where rehabilitation occurs albeit that the goals may be different than an active rehabilitation unit.
- The new SOM was compared to the two most commonly used SOM's at the time of its development. Other measures are now in use since practice changed during the time of the design and evaluation study and comparison with these measures would be informative.
- In the preliminary development of an instrument, there are basic measurement properties that need to be addressed before it can be widely disseminated. Wide dissemination enables evaluation to be considered in different ways. Prediction of long term function, of the likelihood of falls, and many other functions can be examined once the new SOM is in regular use.

- In examining the extent to which change was captured by the new SOM was limited because no external measure of whether change had actually occurred was employed.

### 6.9 Future work

- The survey focussed on one area of physiotherapy practice. For a more representative reflection of the use of SOM's in physiotherapy practice in Ireland, a sample could be generated from the membership list of the Irish Society of Chartered Physiotherapists.
- This survey could take place in parallel with focus groups to explore how the use of SOM's in practice can be optimised.
- Review of some solutions to enhance clinical decision making through the use of outcome measurements for example CPD courses, web-based learning, health informatics solutions, and professional consensus panels.
- Once the new SM is disseminated, a later review of the opinions of users would allow further review of its structure, scoring and instructions for use.
- Ongoing evaluation of its measurement properties such as
  - Predictive ability of falls, change in mobility status
  - How responsiveness compares with global measures of change wither reported by the PT or the patient
  - Use in extended care and community settings.

## REFERENCES

- Alexander NB, Fry-Welch DK, Ward ME, Folkmier LC (1992) Quantitative assessment of Bed Rise Difficulty in young and elderly women. *J Am Geriatr Soc* 40: 685-691.
- Alexander NB, Grunawalt JC, Carlos S, Augustine J (2000) Bed mobility task performance in older adults. *Journal of Rehabilitation Research and Development* 37(5): 633-638.
- Appleby Beattie P (2001) Measurement of health outcomes in the clinical setting: applications to physiotherapy. *Physiotherapy Theory and Practice* 17: 173-185.
- Abbott JH, Mercer SR (2003) Lumbar segmental hypomobility: criterion-related validity of clinical examination items (a pilot study). *NZ J Physiotherapy* 31(1): 3-9.
- Barclay J (1994) *In Good Hands. The history of the Chartered Society of Physiotherapy 1894-1994*. Butterworth Heinemann, Oxford.
- Bassey EJ, Fiatarone M, O'Neill EF, Kelly M, Evans WJ, Lipsitz LA. Leg extensor power and functional performance in very old men and women. *Clinical Science* 82: 321-327.
- Beattie PF, Pinto MB, Nelson MK, Nelson R (2002) Patient satisfaction with outpatient physical therapy: instrument validation. *Phys Ther* 82(6): 557-565.
- Benaim C, Perennou DA, Villy J, Rousseaux M, Pelissier JY (1999) Validation of a standardised assessment of postural control in stroke patients. *Stroke* 30:1862-1868.
- Bendall MJ, Bassey JM, Pearson MB (1989) Factors affecting walking speed of elderly people. *Age Ageing* 18:327-332.
- Berg KO, Wood-Dauphinee SL, Williams JI, Gayton D (1989) Measuring balance in the elderly: preliminary development of an instrument. *Physio Canada* 41:304-311.
- Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1: 07-310.
- Bland JM, Altman DG (1995) Comparing methods of measurement: why plotting difference against standard method is misleading. *Lancet* 346:1085-87.

- Bogle Thorbahn LD, Newton RA (1996) Use of the Berg Balance Test to predict falls in elderly persons. *Phys Ther* 76:576-583.
- Bohannon RW, Andrews AW, Thomas MW (1996) Walking speed: reference values and correlates for older adults. *JOSPT* 24(2): 86-90.
- Bohannon RW (1997) Comfortable and maximum walking speed of adults aged 20-79 years: reference values and determinants. *Age Ageing* 26:15-19.
- Boffelli S, Franzoni S, Rozzini R, Barbisoni P, Bertozzi B, Trabucchi M (1996) Assessment of functional ability with the Bed Rise Difficulty scale in a group of elderly patients. *Gerontology* 42: 294-300.
- Bowling A (2005) *Measuring health- a review of quality of life measurements*. Maidenhead, Open University Press.
- Brennan P, Silman A (1992) Statistical methods for assessing observer variability in clinical measures. *Br Med J* 304:1941-1944.
- Brenner H, Shelley E (1998) *Adding years to life and life to years- a health promotion strategy for older people*. Dublin: National Council for Ageing and Older People.
- Bush KW (2000) The Statistics of Outcome Measurement. *Orth Phys Ther Clin North Am* 9(1): 55-68.
- Carr JH, Mungovan SF, Shepherd RB, Dean CM, Nordholm LA (1994) Physiotherapy in stroke rehabilitation: basis for Australian physiotherapists choice of treatment. *Physiotherapy Theory & Practice* 10: 201-209.
- Carr JH, Shepherd RB, Nordholm L, Lynne D (1985) Investigation of a new motor scale for stroke patients. *Phys Ther* 65:175-80.
- Catell RB (1966) The Scree test for number of factors. *Multivariate Behavioural Research* 1:140-161.
- Chartered Society of Physiotherapy (2000) *Research and Clinical Effectiveness Strategy*. London, CSP.
- Chartered Society of Physiotherapy (1999) *Clinical guidelines for the management of osteoporosis*.  
<http://www.csp.org.uk/effectivepractice/clinicalguidelines/physiotherapyguidelines.cfm#4> (Accessed 22.11.2004)
- Chartered Society of Physiotherapy (1998) *Clinical guidelines for the management of soft tissue injury with PRICE*.



- <http://www.csp.org.uk/effectivepractice/clinicalguidelines/physiotherapyguidelines.cfm#4> (Accessed 22.11.2004)
- Chartered Society of Physiotherapy (2001a) Low back pain-related functional measures. CLEF4.  
<http://www.csp.org.uk/effectivepractice/outcomemeasures/publications.cfm?id> (Accessed 25.06.2004)
- Chartered Society of Physiotherapy (2001b) Clinical guidelines for the management of stress urinary incontinence.  
<http://www.csp.org.uk/effectivepractice/clinicalguidelines/physiotherapyguidelines.cfm#4> (Accessed 22.11.2004)
- Chartered Society of Physiotherapy (2002) Outcome measures for people with depression (a working document) CLEF5.  
<http://www.csp.org.uk/effectivepractice/outcomemeasures/publications.cfm?id> (Accessed 25.06.2004)
- Chesson R, MacLeod M, Massie S (1996) Outcome measures used in Therapy Departments in Scotland. *Physiotherapy* 82(12): 673-679.
- Chio IL, Burnett CN (1985) Values of activities of daily living: a survey of stroke patients and their home therapists. *Phys Ther* 65:901-906.
- Cipriany-Dacko LM, Innerst D, Johannsen J, Rude V (1997) Interrater reliability of the Tinetti Balance Scores in novice and experienced physical therapy clinicians. *Arch Phys Med Rehabil* 78:1160-1164.
- Cohen J (1968) Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psych Bull* 70:213-220.
- Cohen J (1977) *Statistical power analysis for the behavioural sciences*. Academic Press, New York.
- Cohen L, Holliday M (1996) *Practical statistics for students- an introductory text*. Paul Chapman Publishing Ltd, London. P76-87.
- Cole B, Finch E, Gowland C, Mayo N (1995) In: Basmajian J (Editor). *Physical rehabilitation outcome measures*. Toronto (ON) Health and Welfare Canada and Canadian Physiotherapy Association.
- Collen FM, Wade DT, Robb GF, Bradshaw CM (1991) The Rivermead Mobility Index: a further development of the Rivermead Motor Assessment. *Int Disabil Studies* 13:50-54.

- Cooper BA, Stewart D (1997) The effect of a transfer device in the homes of elderly women- the STURDYGRIP vertical pole. *Phys Occup Ther Geriatr* 15(2): 67-77.
- Davidson M, Keating JL (2002) A comparison of five low back pain disability questionnaires: reliability and responsiveness. *Phys Ther* 82(1): 8-24.
- Davidson M (2003) The interpretation of diagnostic tests: a primer for physiotherapists. *Australian Journal of Physiotherapy* 48:227-233.
- Department of Health & Children (2001) Quality and Fairness- A health system for you. Health Strategy. Dublin: Stationary Office.
- Department of Health & Children (1994) Shaping a Healthier Future: a strategy for effective healthcare in the 1990's. Dublin: Stationary Office.
- DeWeert WJG, Harrison MA (1985) Care for stroke patients- against whose yardstick should we measure. *Physiotherapy* 71:298-300.
- Deyo RA, Centor RM (1986) Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. *J Chron Dis* 39:897-906.
- DiFabio RP, Seay R (1997) Use of the 'Fast Evaluation of Mobility, Balance and Fear' in elderly community dwellers: validity & reliability. *Phys Ther* 77(9): 904-917.
- Duckworth M (1999) Outcome measurement selection and typology. *Physiotherapy* 85(1): 21-27.
- Dujardin B, Van den Ende J, Van Gompel A, Unger J, Van der Stuyft P (1994) Likelihood ratios: a real improvement for clinical decision making. *Eur J Epidemiol* 10:29-36.
- Duncan PW, Weiner DK, Chandler J, Studenski S (1990) Functional Reach: a new clinical measure of balance. *J Gerontol* 45(6): M192-197.
- Eliasziw M, Young LS, Woodbury MG, Fryday-Field K (1994) Statistical methodology for the concurrent assessment of interrater and intrarater reliability: using goniometric measurements as an example. *Phys Ther* 74(8): 777-788.
- Enderby P, Kew E (1995) Outcome measurement in physiotherapy using the WHO's Classification of Impairment, Disability and Handicap 81(4): 177-180.
- Epstein A (1990) The outcomes movement- will it get us where we want to go? *The New Eng J Med* 323:266-270

- Fagan TJ (1975) Nomogram for Bayes theorem (letter). *N Engl J Med* 293:257.
- Feinstein A (1987) *Clinimetrics*. Yale University Press. New Haven. 141-211.
- Finch E, Brooks D, Stratford P, Mayo E (2002) *Physical Rehabilitation Outcome Measures- A guide to enhanced clinical decision making* (2<sup>nd</sup> Edition). BC Decker, Ontario.
- Folstein MF, Folstein SE, McHugh PR (1975) 'Mini-mental state'- a practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 12:189-198.
- Forlander DA, Bohannon RW (1999) Rivermead Mobility Index: a brief review of research to date. *Clin Rehabil* 13:97-100.
- Fritz JM (1999) Sensitivity to change (letter). *Phys Ther* 79(4): 420-421.
- Fritz JM, Wainner RS, Hicks GE (2000) The use of nonorganic signs and symptoms as a screening tool for return-to-work in patients with acute low back pain. *Spine* 25(15): 1925-1931.
- Fritz JM, Irrgang JJ (2001) A comparison of a modified Oswestry Low Back Pain Disability Questionnaire and the Quebec Back Pain Disability Scale. *Phys Ther* 81(2): 776-788.
- Fritz JM, George SZ (2002) Identifying psychosocial variables in patients with acute work-related low back pain: The importance of fear-avoidance beliefs. *Physical Therapy* 82 (10): 973-84.
- Green BF (1981) A Primer of Testing. *American Psychologist*; 36(10): 1001-1011.
- Gregson JM, Leathley MJ, Moore AP, Smith TL, Sharma AK, Watkins CL (2000) Reliability of measurements of muscle time and muscle power in stroke patients. *Age Ageing* 29:223-228.
- Guyatt G, Walter S, Norman G (1987) Measuring changes over time: the usefulness of evaluative instruments. *J Chron Dis* 40:171-8.
- Hammond R (1999) Why an outcome measures database? *Physiotherapy* 85(5): 234-235.
- Hammond R (2000) Evaluation of physiotherapy by measuring the outcome. *Physiotherapy* 86(4): 170-172.
- Hays RD, Hadorn D (1992) Responsiveness to change: an aspect of validity, not a separate dimension. *Quality of Life Research* 1:73-75.
- Helmstadter GC (1966) *Principles of psychological measurement*. Methuen & Co Ltd, London. P112-133.

- Hsieh CL, Hsueh IP, Mao HF (2000) Validity and responsiveness of the Rivermead Mobility Index in stroke patients. *Scand J Rehab Med* 32:140-2.
- Huijbregts MPJ, Myers AM, Kay TM, Gavin TS (2002) Systematic outcome measurement in clinical practice: challenges experiences by physiotherapists. *Physiotherapy Canada* 54:25-31,36.
- Irish Heart Foundation, Council on Stroke (2000) Towards excellence in stroke care in Ireland. Dublin: Eireann Health Publications.
- Jaeschke R, Singer J, Guyatt GH (1989) Measurement of health status: ascertaining the minimal clinically important difference. *Control Clin Trials* 10:407-415.
- Jette AM (1995) Outcomes research: shifting the dominant research paradigm in physical therapy 75: 965-970.
- Kane R (1994) Looking for physical therapy outcomes. *Phys Ther* 74:425-429.
- Kane R (1997) Understanding health care outcomes. Aspen, Maryland p1-15.
- Kay TM, Myers AM, Huijbregts MPJ (2001) How far have we come since 1992? A comparative survey of physiotherapists' use of outcome measures. *Physiotherapy Canada* 53:268-275,281.
- Kirkness C, Korner-Bitensky N (2002) Prevalence of outcome measure use by physiotherapists in the management of low back pain. *Physiotherapy Canada* Fall; 54(4); 249-57.
- Kirshner & Guyatt (1985) A methodological framework for assessing health indices. *J Chron Dis* 38(1): 27-36.
- Kline P (1994) An easy guide to factor analysis. Routledge, London. P1-13.
- Kline P (1998) The new psychometrics- science, psychology and measurement. Routledge, London. P51-69.
- Kokko SM, Paltamaa J, Ahola E, Malkia E (1997). The assessment of functional ability in patients with Parkinson's disease: The PLM-tests and three clinical tests. *Physio Res Int* 2: 29-45.
- Liang MH, Cullen KE, Schwartz JA (1985) Comparative measurement efficiency and sensitivity of five health status instruments for arthritis research. *Arthritis Rheum* 28:542-547.
- Liang MH, Fossel AH, Larson MG (1990) Comparison of five health status instruments for orthopaedic evaluation. *Med Care* 28:632-642.

- Liang MH (1995) Evaluating measurement responsiveness. *J Rheumatol* 22:1191-2.
- Liang MH (2000) Longitudinal construct validity: establishment of clinical meaning in patient evaluation instruments. *Med Care* 38 Suppl11:II-89-90.
- Lichenstein ML, Burger C, Shields SL, Shiavi RG (1990) Comparison of biomechanics platform measures of balance and videotaped measures of gait with a clinical mobility scale in elderly women. *J Gerontol* 45(2) M49-54.
- Liebenson C, Yeomans S (1997) Outcomes assessment in musculoskeletal medicine. *Manual Therapy* 2(2) : 67-74.
- Lim LHK, van Wegen EEH, deGoede CJT, Jones D, Rochester L, Hetherington V, Nieuwboer A, Willems AM, Kwakkel G (2005) Measuring gait and gait-related activities in Parkinsons' patients own home environment: a reliability, responsiveness and feasibility study. *Parkinsonism and related disorders* 11:19-24.
- MacKnight C, Rockwood K (2000) Rasch analysis of the hierarchical assessment of balance and mobility (HABAM). *J Clin Epidemiol* 53:1242-1247.
- Mao HF, Hseuh IP, Tang PF, Shue CF, Hsieh CL (2002) Analysis and comparison of the psychometric properties of three balance measures for stroke patients. *Stroke* 33:1022-1027.
- Mayo N, Cole B, Dowler J, Gowland C, Finch E (1993). Use of outcome measurement in physiotherapy: Survey of current practice. *Can J Rehab* 7: 81-83.
- Mayo N (1994). Outcome measures or measuring outcome. *Physiother Can* 46:145-146, 148.
- Messick S (1980) Test validity and the ethics of assessment. *American Psychologist*. 35(11): 1012-1027.
- McDowell I, Newell C (1996) *Measuring Health. A guide to Rating Scales and Questionnaires* (2<sup>nd</sup> Edition) Oxford University Press, New York. P31.
- McEvoy (1991) The use of ADL indices by occupational therapists. *Br J Occ Ther* 47: 225-30.
- McGinty SM, Masters LD, Till DB (1999) Inter-tester reliability using the Tinetti Gait and Balance assessment scale. *Issues on Aging* 22(1): 3-5.
- McIntosh L, Woronuk J, Whitridge M, Elton-Smith G (1999) Development of the Glenrose Ambulation Rating: Validity and inter-rater reliability. *J Rehabil Outcomes* 3(3): 1-11.

- Moore W (1995) *Healing the generations: a history of physical therapy and the American Physical Therapy Association*. Greenwich Publishing Co., Connecticut.
- Moore A (2004) Personal communication.
- National Council on Ageing & Older People. *Adding years to life and life to years. A health promotion strategy for older people*. Dublin: National Council on Ageing & Older People.
- Nilsson LM, Nordholm LA (1992) Physical therapy in stroke rehabilitation: Basis for Swedish therapists choice of treatment. *Physiotherapy Theory & Practice* 8: 49-55.
- Nilsson LM, Carlsson JY, Brimby G, Nordholm LA (1998) Assessment of walking, balance and sensorimotor performance of hemiparetic patients in the acute stage after stroke. *Physiotherapy Theory and Practice* 14: 149-57.
- Palastanga N (1990) *Physiotherapy* 76: 124-126.
- Parker-Taillon (2002) CPA initiatives put the spotlight on evidence-based practice in physiotherapy. *Physiotherapy Canada Winter*: 12-15,24.
- Parry A (1995) Ginger Rogers did everything that Fred Astaire did backwards and in high heels. *Physiotherapy* 81(6): 310-319.
- Parry A (1997) New paradigms for old: musings on the shape of clouds. *Physiotherapy* 83(8): 423-433.
- Partridge CJ (1982) The outcome of physiotherapy and its measurement. *Physiotherapy* 68(11): 362-363.
- Petrie A, Sabin C (2000) *Medical statistics at a glance*. Blackwell Science, Oxford, p49.
- Podsiadlo D, Richardson S (1991) The Timed 'Up & Go'@ a test of basic functional mobility for frail elderly. *J Am Geriatric Soc* 39:142-148.
- Pine ZM, Gurland B, Chren MM (2002) Use of a cane for ambulation: marker and mitigator of impairment in older people who report no difficulty walking. *J Am Geriatric Soc* 50:263-268.
- Piotrowski A, Cole J (1994). Clinical measures of balance and functional assessments in elderly persons. *Aust J Physio* 40: 183-88.
- Platt W, Bell B, Kozak J (1998) *Physiotherapy Functional Mobility Profile*. A tool for measuring functional outcome in chronic care clients. *Physiotherapy Canada Winter*: 47-74.

- Pomeroy VM, Dean D, Sykes L, Farragher EB, Yates M, Tyrrell PJ, Moss S, Tallis RC (2000) The unreliability of clinical measures of muscle tone: implications for stroke therapy. *Age Ageing* 29:229-233.
- Powell LE, Myers AM (1995) The Activities-specific Balance Confidence (ABC) Scale. *J Gerontol Biol Sci Med Sci* 50:M28-34.
- Prosser L, Canby A (1997) Further validation of the Elderly Mobility Scale for measurement of mobility of hospitalised elderly. *Clin Rehabil* 11:338-343.
- Pomeroy V, Tallis R (2000) Physical therapy to improve movement performance and functional ability post stroke. Part 1: Existing evidence. *Reviews in Clinical Gerontology* 10: 261-290.
- Rankin G, Stokes M (1998) Reliability of assessment tools in rehabilitation: an illustration of appropriate statistical analyses. *Clin Rehabil* 12: 187-199.
- Reis AL, Kaplan RM, Blumberg E (1991) Use of factor analysis to consolidate multiple outcome measures in chronic obstructive disease. I *Clin Epidemiol* 44(6): 497-503.
- Relman AS (1988) Assessment and accountability: the third revolution in medical care. *N Engl J Med* 319:1220-1222.
- Reuben DB, Siu AL (1990) An objective measure of physical function of elderly outpatients. The Physical Performance Test. *J Am Geriatr Soc* 38:1105-1112.
- Riddle DL, Stratford PW (1999) Interpreting validity indexes for diagnostic tests: an illustration using the Berg Balance Test. *Phys Ther* 79(10): 939-948.
- Roberts P (1994) Theoretical models of physiotherapy. *Physiotherapy* 80(6): 361-366.
- Roller ML, Avila NR, Bell RG, Spencer N (2001) The interrater reliability of the Tinetti Performance-Oriented Mobility Assessment. *Journal of Geriatric Physical Therapy* 24(3): 27.
- Rothstein J (2001) Sick and tired of reliability. Editor's note. *Phys Ther* 81(2): 774-776.
- Roush SE, Sonstroem RJ (1999) Development of the physical therapy outpatient satisfaction survey (PTOPS). *Phys Ther* 79(2):159-170.
- Royal College of Physicians and The British Geriatrics Society (1992) Standardised assessment scales for elderly people. RCP and BGS, London.
- Russek L, Wooden M, Ekedahl S, Bush A (1997) Attitudes toward standardised data collection. *Phys Ther* 77(7): 714-728.

- Sackett DL, Haynes RB, Guyatt GH, Tugwell P (1991) *Clinical epidemiology. A basic science for clinical medicine*. 2ed ed. Boston, mass: Little, Brown and Co Inc. P69-86.
- Sackett et al (1996) Evidence based medicine: what it is and what it isn't. *BMJ* 312: 71-72.
- Sanford J, Moreland J, Swanson LR, Stratford PW, Gowland C (1993) Reliability of the Fugl-Meyer assessment for testing motor performance in patients following stroke. *Physical Therapy* 73(7): 447-54.
- Schenkman M, Hughes MA, Samsa G, Studenski S (1996) The relative importance of strength and balance in chair rise by functionally impaired older people. *J Am Geriatr Soc* 44:1441-1446.
- Schenkman M, Scherer S, Riegger-Krugh C, Cutson TM (2002) Measurement of impairments and functional limitations in older adults: conceptual considerations and practical applications. *Critical Reviews in Physical and Rehabilitation Medicine*. 14(2): 83-161.
- Shrout PE, Fleiss JL (1979) Intraclass correlations: uses in assessing rater reliability 86(2): 420-428.
- Shumway-Cook A, Baldwin M, Polissar NL, Gruber W (1997). Predicting the probability for falls in community-dwelling older adults. *Phys Ther* 71:812-819.
- Shumway-Cook A, Woollacott M (2001) *Motor Control- theory and practical applications*. Baltimore, Lippincott Williams & Wilkins.
- Simpson J, Forster A (1992) Assessing elderly people. Should we all use the same scales? *Physiotherapy* 79(12): 836-838.
- Smith R (1994) Validation and reliability of the Elderly Mobility Scale. *Physiotherapy* 80(11): 744-747
- Stevenson TJ, Garland SJ (1996) Standing balance during internally produced perturbations in subjects with hemiplegia: validation of the balance scale. *Arch Phys Med Rehabil* 77: 656-61.
- Stevenson TJ (2001) Detecting change in patients with stroke using the Berg Balance Scale. *Australian Journal of Physiotherapy* 47:29-38.
- Stokes EK, Finn AM, Walsh JB, Coakley D, Kirkham D (1998). The 'Balance-Meter': investigation of an apparatus to measure postural sway. *Health Care in Later Life* 3(3): 215-225.



- Stokes EK, O'Neill D (1999) The use of standardised assessments by physiotherapists. *Br J Ther Rehab* 6(11): 560-565.
- Stokes EK, Jennings A, Lyons M, Murphy F, O'Hara G, Walsh C, Wright G, Noel S, Murphy B (2003). Measuring change in the physiotherapy rehabilitation of older people – Barthel Index and Elderly Mobility Scale. *Proceedings of the 14<sup>th</sup> World Confederation for Physical Therapy*: p1944.
- Stratford PW, Binkley JM, Riddle DL (1996a) Health status measures and analytic methods for assessing change scores. *Phys Ther* 76(10): 1109-1123.
- Stratford PW, Binkley J, Soloman P, Finch E, Gill C, Moreland J (1996b) Defining the minimal detectable change for the Roland-Morris Questionnaire. *76(4)*: 350-368.
- Stratford PW, Goldsmith CH (1997) Use of the standard error as a reliability index of interest: an applied example using elbow flexor strength data. *Phys Ther* 77(7): 745-750.
- Stratford PW, Binkley J, Riddle D, Guyatt GH (1998a). Sensitivity to change of the Roland-Morris Back Pain Questionnaire: Part 1 *Phys Ther* 78(11); 1186-95.
- Riddle D, Stratford PW, Binkley J (1998b). Sensitivity to change of the Roland-Morris Back Pain Questionnaire: Part 2. *Phys Ther* 78(11): 1197-1207.
- Stratford P (1999) Who will decide the efficacy of physiotherapeutic interventions? *Physiotherapy Canada* Fall 235-236,238.
- Stratford PW, Binkley JM, Riddle DL (1999) Sensitivity to change (letter). *Phys Ther* 79(4): 421-422.
- Stratford PW (2004) Getting more from the literature: estimating the standard error of the measurement from reliability studies. *Physiotherapy Canada* 56(1): 27-30.
- Streiner DL, Norman GR (1989) *Health Measurement Scales. A Practical Guide to their Development and Use*. Oxford University Press. Oxford. P4-5, 17, 108-109.
- Tallis R (1988) *Measurement in rehabilitation*. Majory Warren Lecture.
- Thapa PB, Gideon P, Fought RL, Kormicki M, Ray WA (1994) Comparison of clinical and biomechanical measures of balance and mobility in elderly nursing home residents. *J Am Geriatr Soc* 42:493-500.
- Tinetti ME (1986) Performance-oriented assessment of mobility problems in elderly patients. *J Am Geriatr Soc* 34:119-126.

- Tinetti ME (1993) Is there one simple measure for balance? *PT Magazine* January: 74-80.
- Torenbeek M, Caulfield B, Garrett M, Van Harten W (2001) Current use of outcome measures for stroke and low back pain rehabilitation in five European countries: first results of the ACROSS project. *Int J Rehabil Res* 24:95-101.
- Trombly CA, Randomsky MV, Davis ES (1998) Achievement of self-identified goals by adults with traumatic brain injury: phase I. *Am J Occup Ther* 52:810-818.
- Turner PA, Whitfield A, Brewster S, Halligan M, Kennedy J (1996) The assessment of pain: an audit of physiotherapy practice. *Australian Journal of Physiotherapy*. 42(1): 55-62.
- Turner PA, Harby-Owen H, Shackelford F, So A, Fosse T, Whitfield TWA (1999) Audits of physiotherapy practice. *Physiotherapy Theory and Practice* 15: 261-74.
- Turner P (2001) Evidence-based practice and physiotherapy in the 1990's. *Physiotherapy Theory and Practice* 17: 107-121.
- Turner-Stokes L, Turner-Stokes T (1997) The use of standardised outcome measures in rehabilitation centres in the UK. *Clin Rehabil* 11:306-313.
- Twomey L (1996) Editorial- research, more essential than ever. *Physiotherapy Research International* 1(2): iii-iv.
- Sutlive TG, Mitchell SD, Maxfield SN, McLean CL et al. (2004) Identification of individuals with patellofemoral pain whose symptoms improved after a combined program of foot orthosis use and modified activity: a preliminary investigation. *Phys Ther* 84(1): 49-61.
- Usuda S, Araya K, Umehara K, Endo M, Shimizu T, Endo F (1998). Construct validity of functional balance scale in stroke patients. *J Phys Ther Sci*; 10: 53-6.
- VanSwearingen JM, Brach JS (2001) Making geriatric assessment work: selecting useful measures. *Phys Ther* 81(6): 1233-1252.
- Wade D. (1992) *Measurement in Neurological Rehabilitation*. Oxford University Press. Oxford.
- Wagner EH, LaCroix AZ, Grothaus LC, Hecht JA (1993) Responsiveness of health status measures to change among older adults. *Jour Am Geriatr Soc* 41:241-248.

- Washburn RA, Weimo Z, McAuley E, Frongley M, Figoni SF (2002) The Physical Activity Scale for individuals with physical disabilities: development and evaluation. *Arch Phys Med Rehabil* 83:193-200.
- Werneke MW, Hart DL (2004) Categorizing Patients With Occupational Low Back Pain by Use of the Quebec Task Force Classification System Versus Pain Pattern Classification Procedures: Discriminant and Predictive Validity. *Physical Therapy* 84 (3): 243-55.
- Whittle M (2001) *Gait analysis: an introduction*. Oxford, Butterworth-Heinemann.
- Wolf SL, Catlin PA, Gage K, Gurucharri K, Robertson R, Stephen K (1999). Establishing the reliability and validity of measurements of walking time using the Emory Functional Ambulation Profile. *Phys Ther* 79: 1122-33.
- World Confederation for Physical Therapy (1999) *Description of Physical Therapy*. London: WCPT.
- World Health Organisation (1948) *Constitution*. In: *Basic documents*. Geneva, Switzerland.
- World Health Organisation (1980) *International Classification of impairments, disabilities and handicaps*. Geneva, Switzerland.
- World Health Organisation (2001) *International Classification of Functioning, Disability and Health*, Geneva: WHO.
- Wright J, Cross J, Lamb S (1998) Physiotherapy outcome measures for rehabilitation of elderly people. Responsiveness to change of the Rivermead Mobility Index and Barthel Index. *Physiotherapy* 84(5): 216-221.
- Wyrwich KW, Tierney WM, Wolinsky FD (1999) Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. *52(9): 861-873*.

## APPENDIX I      SEARCH RESULTS- LIKELIHOOD RATIOS

1. Phillips BA (1998) Clinical diagnosis of sleep apnea... proceedings of a conference held in Cancun, Mexico, October 17-19, 1997. *Respiratory Care* Apr; 43(4); p. 288-95.
2. Paul W Stratford, Jill M Binkley, Daniel L Riddle, Gordon H Guyatt (1998). Sensitivity to change of the Roland-Morris Back Pain Questionnaire: Part 1 *Physical Therapy* 78( 11); 1186-95.
3. Paul W Stratford, Jill M Binkley, Daniel L Riddle, Gordon H Guyatt (1998). Sensitivity to change of the Roland-Morris Back Pain Questionnaire: Part 2 *Physical Therapy* 78( 11): 1197-1208.
4. Wainner RS, Boninger ML, Balu G, Burdett R, Helkowski W (2000) Durkan gauge and carpal compression test: accuracy and diagnostic test properties. *Journal of Orthopaedic and Sports Physical Therapy* 30(11): 676-82.
5. Kliethermes PA; Cross ML; Lanese MG; Johnson KM; Simon SD (1999) Transitioning preterm infants with nasogastric tube supplementation: increased likelihood of breastfeeding. *Journal of Obstetric, Gynecologic, and Neonatal Nursing* 28(3) : 264-73.
6. Riddle DL, Stratford PW (1999) Interpreting validity indexes for diagnostic tests: An illustration using the Berg Balance Test. *79 (10) : 939-49.*
7. Wyer PC, Cordell WH (1999) Evidence-based emergency medicine: updates, feedback, and links. *Annals of Emergency Medicine* 34(5); 657-63.
8. Fritz JM, Winner RS, Hicks GE (2000) The use of nonorganic signs and symptoms as a screening tool for return-to-work in patients with acute low back pain. *Spine* 25(15); 1925-31.
9. Mahadevan M, Graff L (2000) Prospective randomized study of analgesic use for ED patients with right lower quadrant abdominal pain. *American Journal of Emergency Medicine* 18(7); 753-6.
10. Firtz JM, Irrgang JJ (2001) A comparison of a modified Oswestry low back pain disability questionnaire and the Quebec back pain disability scale. *Physical Therapy* 81 (2) : 776-89.
11. Gurrbach RT, Flynn TW, Browder DA, Guffie KL et al (2001) Flexural wave propagation velocity and bone mineral density in females with and without tibial bone stress injuries... including commentary by Levangie PK and Roach KE with author response. *Journal of Orthopaedic and Sports Physical Therapy* 31(2); 54-69.
12. Fritz JM, Wainner RS (2001) Examining diagnostic tests: An evidence-based perspective. *Physical Therapy* 81(9): 1546-65.

13. Vollset SE, Refsum H, Tverdal A, Nygård O et al (2001) Plasma total homocysteine and cardiovascular and noncardiovascular mortality: the Hordaland Homocysteine Study. *American Journal of Clinical Nutrition* 74(1): 130-6.
14. Heitman SJ; Flemons WW (2001) Evidence-based medicine and sleep apnea. *Respiratory Care* 46(12): 1418-34.
15. Blow FC (1998) The spectrum of alcohol interventions for older adults; Alcohol problems and aging, NIAAA research monograph 33: 73-96.
16. Dunn A, McGinn T (2002) An evidence-based approach to the diagnosis of deep vein thrombosis: beyond the ultrasound report: does this patient have deep vein thrombosis? *Journal of the American Geriatrics Society* 50(3): 577-80.
17. Davidson M (2002) The interpretation of diagnostic tests: a primer for physiotherapists. *Australian Journal of Physiotherapy* 48(3): 227-32
18. Fritz JM, George SZ (2002) Identifying psychosocial variables in patients with acute work-related low back pain: The importance of fear-avoidance beliefs. *Physical Therapy* 82 (10): 973-84.
19. Koziatek SM, Powell NJ (2002) A validity study of the evaluation tool of children's handwriting. *American Journal of Occupational Therapy* 56(4) 446-53.
20. Abbott JH, Mercer SR (2003) Lumbar segmental hypomobility: criterion-related validity of clinical examination items (a pilot study). *New Zealand Journal of Physiotherapy* 31(1): 3-9.
21. Sutlive TG, Mitchell SD, Maxfield SN, McLean CL et al. (2004) Identification of individuals with patellofemoral pain whose symptoms improved after a combined program of foot orthosis use and modified activity: a preliminary investigation. *Physical Therapy* 84(1): 49-61.
22. Werneke MW, Hart DL (2004) Categorizing Patients With Occupational Low Back Pain by Use of the Quebec Task Force Classification System Versus Pain Pattern Classification Procedures: Discriminant and Predictive Validity. *Physical Therapy* 84 (3): 243-55.

## APPENDIX 2 1998 SURVEY INSTRUMENT

The following four questions were asked regarding each of the domains (generic \* disease specific) listed below:-

- Mobility/transfers
- Balance
- Gait disorders
- Exercise tolerance/fitness
- Stroke
- Parkinson's disease/parkinsonsim

The definition of standardised assessment employed is 'a published measurement tool', designed for the purpose in a given population, with detailed instructions provided as to when and how it is to be administered and scored, interpretation of the scores and results of investigations into reliability and validity (Cole et al 1995).

For clarity, the example of mobility/transfers will be used in this appendix.

*Question 1-* What type of assessment/measurement is used to measure mobility/transfers?

Standardised

Ad hoc

i.e. a non-standardised assessment or one designed by the department.

*Question 2-* What components should be included in a standardised assessment/measurement of mobility/transfers?

*Question 3-* Do you have any frustrations with the assessment/measurement method you currently use?

Question 4- Please rank, in order of important (1 being most important) the functions that a standardised assessment/measurement of mobility/transfers should demonstrate.

- Provide a problem list
- Identify a state of immobility
- Denote a change due to intervention
- Predict an outcome
- Offer a guideline
- Teaching tool
- Communication aid with other services

## APPENDIX 3 2003 SURVEY INSTRUMENT

Thank you for taking the time to answer this survey, which is a follow-up survey of one that was completed in 1998.

For the purposes of the questions asked, an outcome measure is a scale or instrument that has been published in a journal or textbook.

**1. To measure mobility, do you use**

A published outcome measurement?

One developed by your department?

**2. To measure balance, do you use**

a published outcome measurement?

one developed by the department?

**3. To measure stroke, do you use**

a published outcome measurement?

one developed by the department?

**4. To measure Parkinson's disease/parkinsonism, do you use**

a published outcome measurement?

one developed by the department?

**5. To measure exercise tolerance, do you use**

a published outcome measurement?

one developed by the department?

**6. To measure gait disorders, do you use**

a published outcome measurement?

one developed by the department?



**7. Which published outcome measurement tools do you use (tick as many as appropriate)?**

M O B I L I T Y	Elderly Mobility Scale (EMS)	<input type="checkbox"/>
	Timed Up & Go (TUG)	<input type="checkbox"/>
	Performance oriented assessment of mobility Gait sub-scale (Tinetti - gait)	<input type="checkbox"/>
	Other (please specify)	<input type="checkbox"/>
	Other (please specify)	<input type="checkbox"/>

B A L A N C E & F A L L S	Berg balance scale (BBS)	<input type="checkbox"/>
	Functional Reach (FR)	<input type="checkbox"/>
	Performance oriented assessment of mobility Balance sub-scale (Tinetti - balance)	<input type="checkbox"/>
	Balance-master (BM)	<input type="checkbox"/>
	Falls efficacy scale (FES)	<input type="checkbox"/>
	Other (please specify)	<input type="checkbox"/>
	Other (please specify)	<input type="checkbox"/>

G A I T	Performance oriented assessment of mobility Gait sub-scale (Tinetti - gait)	<input type="checkbox"/>
	Other (please specify)	<input type="checkbox"/>
	Other (please specify)	<input type="checkbox"/>

S T R O K E	Fugl-Meyer Assessment (FM)	<input type="checkbox"/>
	Motor assessment scale (MAS)	<input type="checkbox"/>
	Motor Club Assessment (MCA)	<input type="checkbox"/>
	Other (please specify)	<input type="checkbox"/>
	Other (please specify)	<input type="checkbox"/>

P A R K I N S O N S	Webster (Parkinson's)	<input type="checkbox"/>
	Hoehn & Yahr Staging - Parkinson's	<input type="checkbox"/>
	Therapy Outcome Measure	<input type="checkbox"/>
	Other (please specify)	<input type="checkbox"/>
	Other (please specify)	<input type="checkbox"/>

E X E R C I S E	Six-minute distance test	<input type="checkbox"/>
	Shuttle test	<input type="checkbox"/>
	Borg Scale - rate of perceived exertion	<input type="checkbox"/>
	Visual analogue scale - exertion/fatigue	<input type="checkbox"/>
	Other (please specify)	<input type="checkbox"/>
	Other (please specify)	<input type="checkbox"/>

## 8. At what times do you use the specific outcome measures?

	Admission	Discharge	Admission & discharge	More often
Elderly Mobility Scale	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Timed Up & Go	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Performance oriented assessment of mobility -Gait sub-scale	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Berg balance scale	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Functional Reach	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Performance oriented assessment of mobility-Balance sub-scale	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Falls efficacy scale	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Fugl-Meyer Assessment	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Motor assessment scale	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Motor Club Assessment	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Webster (Parkinson's)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Hoehn & Yahr Staging	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Therapy Outcome Measure	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Six-minute distance test	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Shuttle test	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Borg Scale	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Visual analogue scale - exertion	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Other _____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(please specify)				
Other _____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(please specify)				

9. Do you experience any barriers to using outcome measures in your practice?

Yes  No

If yes, please tick if the following apply.

Lack of time  Lack of knowledge  Availability of measures

None meet clients' needs

Lack of professional consensus on what to use

Lack of administrative support  Lack of equipment

Low priority  Lack of personal interest  Lack of resources

Other (please specify) \_\_\_\_\_

**10. Please rate your confidence [0-100%] in the scenarios described in the following 12 statements.**

0% not at all comfortable or confident  
100% completely comfortable or confident

How confident are you...

1. Knowing whether suitable measures are available

0%  10%  20%  30%  40%  50%  60%  70%  80%  90%  100%

2. Knowing enough about reliability and validity to choose the best measure

0%  10%  20%  30%  40%  50%  60%  70%  80%  90%  100%

3. Knowing about scale construction to develop my own instrument

0%  10%  20%  30%  40%  50%  60%  70%  80%  90%  100%

4. Knowing what to measure for my client group(s)

0%  10%  20%  30%  40%  50%  60%  70%  80%  90%  100%

5. Knowing how to administer measures in a standardised manner

0%  10%  20%  30%  40%  50%  60%  70%  80%  90%  100%

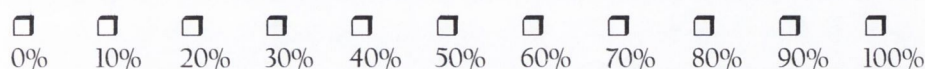
6. Knowing **how** to score measures



7. Knowing **what** to do with the scores



8. Knowing how to track clients' progress



9. Knowing **how** to link information from outcome measures with other sources of information e.g. client characteristics



10. Knowing **how** to compare scores to baseline levels across clients

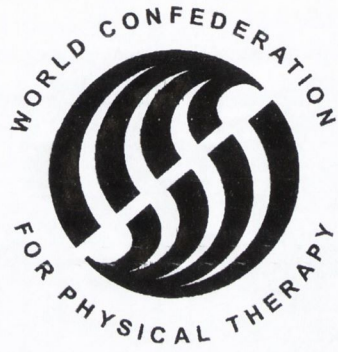


11. Overall, have a clear purpose for measurement- **knowing why** to measure



12. Overall, knowing **what** to do with the information obtained





DESCRIPTION OF PHYSICAL THERAPY

### Why a Description?

The existence of WCPT demonstrates the international dimension of Physical Therapy practice. The variety of its Member Organisations illustrates the diversity of needs and contexts of health care delivery throughout the Confederation.

In response to a motion at the 13<sup>th</sup> General Meeting to develop a description of physical therapy, WCPT initiated a consultative exercise with the intention of providing a foundation on which Member Organisations in different parts of the world could build a description of physical therapy relevant to their needs.

WCPT is committed to supporting Member Organisations - not stereotyping them. It is in this spirit that this description of physical therapy has been drawn up in response to the expressed need of members. It is intended as a Position Statement rather than a Declaration of Principle and is therefore open to be adopted fully, in part or developed to meet the evolving needs of the profession. New research is proving further evidence upon which future practice will build. Nowhere is this more apparent than in our understanding of human movement which is central to the skills and knowledge of the physical therapist. Clearly the uniqueness of the contribution which physical therapy can make to health care in the next millennium remains to be fully defined. This statement is presented as the basis upon which subsequent reviews of the description will continue to be conducted in response to the development of knowledge in physical therapy and the profession's response to changing health needs of society.

## Table of Contents

- 1 **What is Physical Therapy?**
  - 1.1 The nature of physical therapy
  - 1.2 The nature of the physical therapy process
  
- 2 **Where is Physical Therapy Practised?**
  - 2.1 The scope of physical therapy services
  - 2.2 Settings in which physical therapy is practised
  
- 3 **What Characterises Physical Therapy?**
  - 3.1 Assumptions underlying the knowledge and practice of physical therapy
  
- 4 **Where are we now?**
  - 4.1 Principles supporting the description of physical therapy

### Note:

The professional title and term used to describe the profession's practice vary and depend largely on the historical roots of the profession in the country of the WCPT Member Organisation.

The most generally used titles and terms are 'physical therapist' or 'physiotherapist' and 'physical therapy' or 'physiotherapy'. Physical therapist and physical therapy are used in this document but may be replaced by WCPT Member Organisations in favour of those terms officially used by them and their members without any change in the meaning of the document.



## 1 What is Physical Therapy?

### 1.1 The nature of Physical Therapy

Physical Therapy is providing services to people and populations to develop, maintain and restore maximum movement and functional ability throughout the lifespan. Physical therapy includes the provision of services in circumstances where movement and function are threatened by the process of ageing or that of injury or disease. Full and functional movement are at the heart of what it means to be healthy.

Physical therapy is concerned with identifying and maximising movement potential, within the spheres of promotion, prevention, treatment and rehabilitation. Physical therapy involves the interaction between physical therapist, patients or clients, families and care givers, in a process of assessing movement potential and in establishing agreed upon goals and objectives using knowledge and skills unique to physical therapists.

The physical therapists' distinctive view of the body and its movement needs and potential is central to determining a diagnosis and an intervention strategy and is consistent whatever the setting in which practice is undertaken. These settings will vary in relation to whether physical therapy is concerned with health promotion, prevention, treatment or rehabilitation.

### 1.2 The nature of the physical therapy process

Physical therapy is the service only provided by, or under the direction and supervision of a physical therapist and includes assessment, diagnosis, planning, intervention and evaluation.

**Assessment** includes both the **examination** of individuals or groups with actual or potential impairments, functional limitations, disabilities, or other conditions of health by history taking, screening and the use of specific tests and measures and **evaluation** of the results of the examination through analysis and synthesis within a process of clinical reasoning.

**Diagnosis** arises from the examination and evaluation and represents the outcome of the process of clinical reasoning. This may be expressed in terms of movement dysfunction or may encompass categories of impairments, functional limitations, abilities/ disabilities or syndromes.

**Planning** begins with determination of the need for intervention and normally leads to the development of a plan of intervention, including measurable outcome goals negotiated in collaboration with the patient/ client, family or care giver. Alternatively it may lead to referral to another agency in cases which are inappropriate for physical therapy.

**Intervention** is implemented and modified in order to reach agreed goals and may include manual handling; movement enhancement; physical, electro-therapeutic and mechanical agents; functional training; provision of aids and appliances; patient related instruction and counselling; documentation and co-ordination, and communication. Intervention may also be aimed at **prevention** of impairments, functional limitations, disability and injury including the **promotion** and maintenance of health, quality of life, and fitness in all ages and populations.

Evaluation necessitates re-examination for the purpose of evaluating outcomes.

## 2. Where is physical therapy practised?

### 2.1 The scope of physical therapy services

Physical therapy is an essential part of the health services delivery system. Physical therapists practice independently of other health care providers and also within interdisciplinary rehabilitation/habilitation programs for the restoration of optimal function and quality of life in individuals with loss and disorders of movement. Physical therapists are guided by their own code of ethical principles. Thus, they may be concerned with one of the following purposes:

- Promoting the health and well being of the individual and the general public/society.
- Preventing impairments, functional limitations, and disabilities in individuals at risk of altered movement behaviours due to health or medically related factors, socio-economic stressors, and lifestyle factors.
- Providing interventions to restore integrity of body systems essential to movement, maximise function and recuperation, minimise incapacity, and enhance the quality of life in individuals and groups of individuals with altered movement behaviours resulting from impairments, functional limitations, disabilities.

### 2.2 Settings in which physical therapy is practised

Physical therapy is delivered in a variety of settings which allow for it to achieve its purpose.

**Treatment and Rehabilitation** usually occur in community and acute care settings which may include but are not confined to the following:

- Hospices
- Hospitals
- Nursing Homes
- Rehabilitation Centres/Residential Homes
- Physical Therapist Private Office/Practice/Clinic
- Out-Patient Clinics

- Community Settings: Primary Health Care Centres: Individual Homes: Field Settings
- Education and Research Centres

Prevention and Health Promotion are more likely to occur in the following settings although they often form an integral part of treatment and rehabilitation offered within other care settings.

- Fitness Centres/Health Clubs/Spas
- Occupational Health Centres
- Schools
- Senior Citizen Centres
- Sports Centres
- Workplace/Companies
- Public settings (i.e. Shopping Malls) for health promotion

### 3 What Characterises Physical Therapy?

#### 3.1 Assumptions underlying the knowledge and practice of physical therapy

The following assumptions are embedded in this description and reflect the central issues of physical therapy.

##### Movement

The capacity to move is an essential element of health and well-being. Movement is dependent upon the integrated, co-ordinated function of the human body at a number of different levels.

*Movement is purposeful and is affected by internal and external factors.*

Physical therapy is directed towards the movement needs and potential of the individual.

##### Individuals

Individuals have the capacity to change as a result of their responses to physical, psychological, social and environmental factors.

Body, mind and spirit contribute to individuals' views of themselves and enable them to develop an awareness of their own movement needs and goals.

Ethical principles require the physical therapist to recognise the autonomy of the patient or legal guardian in seeking his or her services.

##### Interaction

Interaction aims to achieve a mutual understanding between the physical therapist and the patient/client/family or care giver and forms an integral part of physical therapy.

Interaction is a pre-requisite for a positive change in body awareness and movement behaviours that may promote health and well-being.

Interaction often involves partnership within inter-disciplinary teams, in determining the needs and formulating goals for physical therapy intervention and recognises the patient/client/family and care givers as being active participants in this process.

### Professional Autonomy

Professional education prepares physical therapists to be autonomous practitioners.

Professional autonomy is possible for individual physical therapists as they practice with patients/clients/family and care givers to reach a diagnosis which will direct their physical therapy interventions.

### Diagnosis

Diagnosis within physical therapy is the result of a process of clinical reasoning which results in the identification of existing or potential impairments, functional limitations and abilities/disabilities.

The purpose of the diagnosis is to guide physical therapists in determining the prognosis and identifying the most appropriate intervention strategies for patients/clients and in sharing information with them.

In carrying out the diagnostic process, physical therapists may need to obtain additional information from other professionals.

If the diagnostic process reveals findings that are not within the scope of the physical therapist's knowledge, experience or expertise, the physical therapist will refer the patient/client to another appropriate practitioner.

## 4 Where are we now?

### 4.1 Principles supporting the description of physical therapy

In order to make explicit the underlying values upon which this international description of physical therapy is based there follows a list of principles which are recognised as important by WCPT.

WCPT believes a description must:

- respect and recognise the history and roots of the profession;
- build on the reality of contemporary practice and the growing body of research;
- allow for variation in: cultures, values and beliefs; health needs of people and societies; and structure of health systems around the world;
- use terminology that is widely understood and adequately defined;
- recognise internationally accepted models and definitions (e.g. World Health Organisation definition of health);
- provide for the ongoing growth and development of the profession and for the identification of the unique contribution of physical therapy;
- acknowledge the importance of the movement sciences within physical therapy curricula at all levels;

- emphasise the need for practice to be evidence based whenever possible;
- appreciate the inter-dependence of practice, research and education within the profession:
- recognise the need to continuously review the description as the profession changes in response to the health needs of society and the development of knowledge in physical therapy;
- anticipate that work will flow from this description through utilisation of the document to assist in the development of curricula and identification of areas for research.

#### REFERENCES:

- American Physical Therapy Association: Guide to Physical Therapist Practice, *Physical Therapy*. 77, November 1997
- Cott CA, Finch E, Gasner D, Yoshida K, Thomas SG, Verrier MC(1995): The Movement Continuum of Physical Therapy, *Physiotherapy Canada* 47:87-95
- Higgs J, Jones M (eds)(1995): *Clinical Reasoning in the Health Professions*, Butterworth-Heinemann, Oxford
- Standing Liaison Committee of Physiotherapists in the European Union: *The Practice of Physiotherapy in the European Union*, 1996
- Tornebohm H(1991): What is worth knowing in Occupational therapy? *American Journal of Occupational Therapy*, Vol.45 No.5 p.451 -454
- World Health Organisation (WHO): *International Classification of Impairments, Disabilities and Handicap*, 1980

**APPENDIX 5      BARRIERS SURVEY**

*Below you will find a list of potential barriers to the attainment of functional independence. Following each variable are five boxes labelled as follows:-*

- |                          |                          |                          |                          |                          |
|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Not at all               | Insignificant            | Significant              | Very significant         | Hinders Intervention     |

Please tick the level at which each variable acts as a barrier in preventing the patient from achieving functional independence.

**Cognition**

- |                          |                          |                          |                          |                          |
|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Not at all               | Insignificant            | Significant              | Very significant         | Hinders Intervention     |

**Pain**

- |                          |                          |                          |                          |                          |
|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Not at all               | Insignificant            | Significant              | Very significant         | Hinders Intervention     |

**Mood**

- |                          |                          |                          |                          |                          |
|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Not at all               | Insignificant            | Significant              | Very significant         | Hinders Intervention     |

**Communication**

- |                          |                          |                          |                          |                          |
|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Not at all               | Insignificant            | Significant              | Very significant         | Hinders Intervention     |

**Problems with balance/stability**

- |                          |                          |                          |                          |                          |
|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Not at all               | Insignificant            | Significant              | Very significant         | Hinders Intervention     |

**Immobility**

- |                          |                          |                          |                          |                          |
|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Not at all               | Insignificant            | Significant              | Very significant         | Hinders Intervention     |

**Poor fitness/deconditioning/tolerance for exercise**

- |                          |                          |                          |                          |                          |
|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Not at all               | Insignificant            | Significant              | Very significant         | Hinders Intervention     |

Other (please specify)

- 
- |                          |                          |                          |                          |                          |
|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Not at all               | Insignificant            | Significant              | Very significant         | Hinders Intervention     |

APPENDIX 6 1<sup>ST</sup> ITERATION OF SCALE**Bed rise (0-7)**

Ability to rise from lying on a plinth, bring legs over the edge of the bed and sit up.

- 7 Requires hoist/mechanical aid
- 6 Requires assistance of 2
- 5 Requires assistance of 1 for the lower extremity & trunk
- 4 Requires assistance for upper extremity
- 3 Requires verbal cueing
- 2 Uses compensatory strategies
  - Pulls off bed rails
  - Discontinuity of trunk elevation & leg motion off bed
  - LE use – multiple motions, poor clearance of heels, pulls with leg to aid motion
  - Hands grasp thigh or buttocks to aid trunk movement
  - Rolls onto one side and uses contra lateral UE for push off
- 1 Independent, no compensatory strategies
- 0 Independent in 3 seconds or less

**Sit to stand (0-6)**

Ability to stand from a standard rehabilitation chair.

- 6 Requires mechanical aid
- 5 Requires assistance of 2
- 4 Requires assistance of 1
- 3 Requires verbal cueing
- 2 Uses compensatory strategies
  - Pushing down excessively on upper extremities
  - Repeated forward flexion of trunk to initiate the movement i.e. use of momentum
- 1 Independent in > 2 seconds
- 0 Independent in 2 seconds or less

**Standing (0-8)**

Ability to hold each position for 10 seconds

- 8 Immediate standing with no physical support
- 7 Feet apart- eyes open
- 6 Feet apart- eyes closed
- 5 Feet together- eyes open
- 4 Feet together- eyes closed
- 3 Semi tandem- eyes open
- 2 Semi tandem- eyes closed
- 1 Tandem- eyes open
- 0 Tandem- eyes closed

**Reach & Lift (0-6)**

The patient is asked to lift a textbook and place on a shelf above shoulder height.  
Timed task.

- 6 Unable
- 5 Able with assistance of one, for the purposes of stability
- 4 Requires use of compensatory strategies
  - Needs to hold object e.g. aid or table, on reaching up
- 3 Independent in > 6 seconds
- 2 Independent in 4.6-6 seconds
- 1 Independent in 2.6-4.5 seconds
- 0 Independent in 2.5 or fewer seconds

**Bend & Reach (0-6)**

The patient is asked to place the same object on the floor.  
Timed task.

- 6 Unable
- 5 Able with assistance of one, for the purposes of stability
- 4 Requires use of compensatory strategies
  - Needs to lean on object to lean down or come up
  - Inadequate knee flexion – reaches from trunk
  - Needs assistance on rising
- 3 Independent in > 6 seconds
- 2 Independent in 4.6-6 seconds
- 1 Independent in 2.6-4.5 seconds
- 0 Independent in 2.5 or fewer seconds

**Repeated Sit to Stand (0-54)**

Patient is asked to stand up and sit down 5 times. Time starts at initial sitting position to the final standing position at the fifth stand.

- 4 Unable to complete
- 3 Completed in greater than or equal to 16.7 seconds
- 2 Completed in 13.7-16.6 seconds
- 1 Completed in 11.2-13.6 seconds
- 0 Completed in a time less than or equal to 11.1



Gait

10m test the subject is asked to walk at their normal comfortable speed.

Provide 2m for acceleration and deceleration before and after the test distance.

Gait patterns (0-8)

Gait initiation-	normal	0	Step length-	normal	0
	compensatory	1		compensatory	1
	unable	2		unable	2
Step continuity-	normal	0	Step symmetry-	normal	0
	compensatory	1		compensatory	1
	unable	2		unable	2

Gait speed (0-6)

- 5 Unable to complete
- 4 Able to complete with assistance of 2
- 3 Able to complete with assistance of 1
- 2 Able to complete with frame
- 1 Able to complete with cane
- 0 Able to complete independently in <0.808cm/sec (female), <1.11 cm/second (male)

Turning 360 degrees (0-6)

- 6 Unable to complete
- 5 Able to complete with assistance of 2
- 4 Able to complete with assistance of 1
- 3 Able to complete with frame
- 2 Able to complete with cane
- 1 Able to complete independently
- 0 Able to complete independently in fewer than 5 steps

**APPENDIX 7**      **1<sup>ST</sup> ITERATION- STRUCTURE OF REVIEW FOR  
CONTENT VALIDITY**

**Name**

**Centre**

Item 1- bed rise

Rate the importance of each item 0-5 (not at all important-very important)

Hierarchy presentation of scoring

Instructions

Other comments

Item 2 – sit to stand

Rate the importance of each item 0-5 (not at all important-very important)

Hierarchy presentation of scoring

Instructions

Other comments

Item 3 - standing

Rate the importance of each item 0-5 (not at all important-very important)

Hierarchy presentation of scoring

Instructions

Other comments

Item 4 – reach & lift

Rate the importance of each item 0-5 (not at all important-very important)

Hierarchy presentation of scoring

Instructions

Other comments

Item 5 – bend & reach

Rate the importance of each item 0-5 (not at all important-very important)

Hierarchy presentation of scoring

Instructions

Other comments

Item 6 – repeated sit to stand

Rate the importance of each item 0-5 (not at all important-very important)

Hierarchy presentation of scoring

Instructions

Other comments

Item 7 - gait

Rate the importance of each item 0-5 (not at all important-very important)

Hierarchy presentation of scoring

Instructions

Other comments

Item 8 – gait speed

Rate the importance of each item 0-5 (not at all important-very important)

Hierarchy presentation of scoring

Instructions

Other comments

Item 9 – turning 360 degrees

Rate the importance of each item 0-5 (not at all important-very important)

Hierarchy presentation of scoring

Instructions

Other comments

## APPENDIX 8      2<sup>ND</sup> ITERATION OF SCALE AND STRUCTURE FOR REVIEW OF CONTENT VALIDITY

### Bed rise (0-7)

Ability to rise from lying on a plinth, bring legs over the edge of the bed and sit up. In the case of patients with stroke, orthopaedic conditions, use the side of the plinth that will optimise performance and repeat assessments consistently on this side. The patient should use the number of pillows normally used when sleeping.

- 0      Requires hoist/mechanical aid
- 1      Requires assistance of 2
- 2      Requires assistance of 1 for the leg(s) and/or trunk
- 3      Requires assistance for upper extremity (UE)
- 4      Requires verbal cueing/prompting to complete task or because of safety issues
- 5      Uses compensatory strategies - below are some examples of compensatory strategies that may be employed, the list is not meant to be definitive
  - Pulls off bed rails/monkey pole
  - Takes a long duration or demonstrates repeated use of upper extremities to push off bed
  - Discontinuity of trunk elevation & leg motion off bed
  - LE use – multiple motions, poor clearance of heels, pulls with leg to aid motion
  - Hands grasp thigh or buttocks to aid trunk movement
  - Rolls onto one side and uses *contra lateral* UE for push off
- 6      Independent, no compensatory strategies, transient use of UE's is acceptable
- 7      Independent in 3 seconds or less

#### COMMENTS & QUESTIONS

##### My questions

Should assistance for the trunk/LE be combined with UE?

Is verbal cueing acceptable as it is or should safety vs prompting be different scores?

Should this item be included    Yes                   No

Your comments

### Sit to stand (0-6)

Ability to stand from a chair that allows the patient to sit with their knees and hips at approx 90 degrees and their feet flat on the floor. The patient may use one or both upper extremities.

- 0      Requires mechanical aid
- 1      Requires assistance of 2
- 2      Requires assistance of 1
- 3      Requires verbal cueing for safety and/or task description - supervision
- 4      Uses compensatory strategies - some examples are listed you may wish to add others
  - Use of UE's

- Repeated forward flexion of trunk to initiate the movement i.e. use of momentum
- Cushions on chair enable independence by reducing the height from which the patient has to rise

- \_\_\_\_\_
- \_\_\_\_\_

- 5 Independent in > 2 seconds  
6 Independent in 2 seconds or less

#### COMMENTS & QUESTIONS

##### My questions

Is the chair description acceptable? Or should I include the dimensions of a rehabilitation chair?

Is the verbal cueing score acceptable to included supervision?

Should assistance of one be subdivided further e.g. moderate, minimum assistance?

Are the compensations listed acceptable?

Should this item be included    Yes                       No

Your comments

#### Standing ability & stability (0-13)

The ability to stand in the most upright posture for that patient. With the therapist sitting/standing in front of the patient, ask the patient to stand up. One arm may be supported to allow the patient to position their feet, if the patient needs help positioning their feet; the therapist carries this out. The patient is asked if they are ready, support is removed and timing is begun. The patient's arms should remain close their body.

- 0 Unable to stand unsupported  
1 Able to stand with assistance of one  
2 Able to stand with assistance of an aid  
3 Immediate standing with no physical support  
4 Feet apart, with broad base (> 20cm between medial aspect heels) – eyes open (EO) - maintains for 10 seconds  
5 Feet apart, with broad base (> 20cm between medial aspect heels) – eyes closed (EC) - maintains for 10 seconds  
6 Feet apart, with normal base – EO - maintains for 10 seconds  
7 Feet apart, with normal base – EC - maintains for 10 seconds  
8 Feet together – EO - maintains for 10 seconds  
9 Feet together – EC - maintains for 10 seconds  
10 Semi tandem, the heel of one foot is placed beside the base of the first toe of the opposite foot – EO - maintains for 10 seconds  
11 Semi tandem, the heel of one foot is placed beside the base of the first toe of the opposite foot – EC - maintains for 10 seconds  
12 Tandem, the heel of one foot is placed in front of the toes of the other foot – EO - maintains for 10 seconds  
13 Tandem, the heel of one foot is placed in front of the toes of the other foot – EC - maintains for 10 seconds

**COMMENTS & QUESTIONS****Your comments**

Should this item be included    Yes                       No

**Reach & Lift (0-6)**

The patient is asked to lift a textbook (physiotherapy equipment catalogue) from a standard gym table, hand to the physiotherapist who places his/her receiving hand at a height that requires the patients to achieve 90 degrees flexion at the shoulder or in the case of limited shoulder movement, the greatest range available. In the case of one non-functioning upper extremity, one arm may be used.

This is a timed task. Instructions 'when I say go, I want you to take the book from the table and hand it to me'.

- 0        Unable
- 1        Able with assistance of one, for the purposes of stability
- 2        Requires use of compensatory strategies
  - Needs to hold object e.g. aid or table, on reaching up
- 3        Independent in > 6 seconds
- 4        Independent in 4.6-6 seconds
- 5        Independent in 2.6-4.5 seconds
- 6        Independent in 2.5 or fewer seconds

**COMMENTS & QUESTIONS****My questions**

Should FIM definitions of levels of assistance be included?

Should this item be included    Yes                       No

**Your comments****Bend & Reach (0-6)**

Place a penny on the floor approximately one metre away from the patient on their dominant or unaffected side in the case of upper extremity dysfunction. Ask them to pick it up from the floor and stand up.

This is a timed task. Instructions 'when I say go, I want you to pick the cone up from the floor'

- 0        Unable
- 1        Able with assistance of one, for the purposes of stability, throughout task
- 2        Requires use of compensatory strategies
  - Needs to lean on object to lean down or come up
  - Inadequate knee flexion – reaches from trunk
  - Needs assistance on rising
- 3        Independent in > 6 seconds
- 4        Independent in 4.6-6 seconds
- 5        Independent in 2.6-4.5 seconds
- 6        Independent in 2.5 or fewer seconds

**COMMENTS & QUESTIONS**  
**My questions**  
 Is there a need for the compensation section?

Should the FIM definitions of levels of assistance be included?

Should this item be included Yes  No

**Your comments**

**Repeated Sit to Stand (0-5)**

Patient is asked to stand up and sit down 5 times from a standard chair. Time starts at initial sitting position to the final standing position at the fifth stand.

Instructions 'when I say go, I want you to stand up and sit down as quickly as you can, 5 times'. Please try not to use your arms to help you push up.

Verbal cueing to count the number of stands is allowed. No encouragement should be given.

- 0 Unable to complete
- 1 Able to complete using the upper extremities
- 2 Completed in greater than or equal to 16.7 seconds, without UE use
- 3 Completed in 13.7-16.6 seconds, without UE use
- 4 Completed in 11.2-13.6 seconds, without UE use
- 5 Completed in a time less than or equal to 11.1 seconds, without UE use

**COMMENTS & QUESTIONS**  
**My questions**  
 Is the gap between 1 and 2 too much?

This is to test aspects of lower extremity strength/endurance hence no UE use allowed. Is this acceptable?

Should this item be included Yes  No

**Your comments**

**Gait ability (0-12)**

This test assesses gait ability and speed, at the higher levels of performance. The patient is asked to walk 'at your normal comfortable speed' along a stretch marked out on the floor in the gym or with cones, if necessary. Approximately 2m acceleration and deceleration are allowed. The patient uses their assistive device. A frame includes a rollator or motorised frame or 2 crutches. A cane also incorporates the use of one crutch or one elbow crutch.

Timing occurs over 10m of the course, following the length of acceleration.

- 0      unable to mobilise or complete 10m
- 1      requires assistance of 2 people to mobilise
- 2      requires maximum assistance
- 3      requires moderate assistance
- 4      requires contact guard assistance/verbal cueing/supervision
- 5      completes with frame independently in > 30 seconds
- 6      completes independently with a frame in <30 seconds
- 7      completes independently with a cane in > 10 seconds (male)/>11 seconds (female)
- 8      completes independently with a cane in 7-10 seconds (male)/7-11 seconds (female)
- 9      completes independently with a cane in < 7 seconds (male)/<7.5 seconds (female)
- 10     completes independently with no aid > 10 seconds (male)/>11 seconds (female)
- 11     completes independently with no aid in 7-10 seconds (male)/7-11 seconds (female)
- 12     completes independently with no aid in <7 seconds (male)/<7.5 seconds (female)

**COMMENTS & QUESTIONS****My questions**

Is 4 acceptable or should it be subdivided?

Should this item be included    Yes                       No

Your comments



**Gait pattern**

The patient is asked to walk along a line on the floor in the gym, turning at each end and repeating as requested. The patient is asked to walk at a comfortable speed and may use the assistive device they normally require.

**GAIT INITIATION** - A smooth start is absent, the patient presents with start hesitation or freezing

Normal 2      Abnormal 1

**SWING PHASE** – Initial swing → mid swing → terminal swing

Right

Abnormalities in clearance

Normal 2      Abnormal 1

Step length

Normal 2      Abnormal 1

Left

Abnormalities in clearance

Normal 2      Abnormal 1

Step length

Normal 2      Abnormal 1

**STANCE PHASE** –

**Initial contact** → Foot flat → mid stance → terminal stance → Toe off  
Single and double leg stance

Right

Initial contact

Heel strike      Normal 2      Abnormal 1

Foot flat      Normal 2      Abnormal\* 1

\* not controlled

During stance

Hip control      Normal 2      Abnormal 1

- May be noted as abnormal foot placement i.e. toe in, toe out
- Pain on loading

Knee control      Normal 2      Abnormal 1

- Potentially hyperextension at the knee joint
- Inability to extend knee appropriately

Trunk control      Normal 2      Abnormal 1

- Support required at the lower trunk and pelvis to maintain level of pelvis during single leg stance

Toe off      Normal 2      Abnormal (not forefoot) 1

Left

Initial contact

Heel strike      Normal 2      Abnormal 1

Foot flat      Normal 2      Abnormal\* 1

\*not controlled

During stance				
Hip control	Normal	2	Abnormal	1
	<ul style="list-style-type: none"> <li>• May be noted as abnormal foot placement i.e. toe in, toe out</li> <li>• Pain on loading</li> </ul>			
Knee control	Normal	2	Abnormal	1
	<ul style="list-style-type: none"> <li>• Potentially hyperextension at the knee joint</li> <li>• Inability to extend knee appropriately</li> </ul>			
Trunk control	Normal	2	Abnormal	1
	<ul style="list-style-type: none"> <li>• Support required at the lower trunk and pelvis to maintain level of pelvis during single leg stance</li> </ul>			
Toe off	Normal	2	Abnormal (not forefoot)	1
Turning – en bloc, turn hesitation				
	Normal	2	Abnormal	1
Step continuity – look for toe off corresponding to heel strike on opposite side				
	Normal	2	Abnormal	1
Arm swing (not present if using frame)				
	Normal	2	Abnormal	1

**COMMENTS & QUESTIONS**

**My questions**

On the basis of inclusion of all the previous sections, does this section need to be included?

Or should it be an optional module?

If yes to either of the above options, does it all need to be included?

If no, please mark which sections could be excluded, perhaps with a highlight pen.

Is the information on compensations necessary?

If yes, is this information acceptable?

Should this item be included    Yes                       No

Your comments

*APPENDIX 9 FINAL VERSION OF THE SCALE, FOR EVALUATION*

**Bed rise (0-6)**

Ability to rise from lying on a plinth, bring legs over the edge of the bed and sit up. In the case of patients with stroke or orthopaedic conditions, use the side of the plinth that will optimise performance and repeat assessments consistently on this side. The patient should use the number of pillows normally used when sleeping.

**Instruction:** I'd like you to get up from lying and sit over the edge of the bed, as quickly as you can.

- 0 Requires hoist/mechanical aid
- 1 Requires assistance of 2
- 2 Requires assistance of 1 for the leg(s) and/or trunk and/or arms
- 3 Requires verbal cueing/prompting to complete task or because of safety issues
- 4 Uses compensatory strategies- below are some examples of compensatory strategies that may be employed, the list is not meant to be definitive
  - Pulls off bed rails/monkey pole
  - Takes a long duration or demonstrates repeated use of upper extremities to push off bed
  - Discontinuity of trunk elevation & leg motion off bed
  - LE use – multiple motions, poor clearance of heels, pulls with leg to aid motion
  - Hands grasp thigh or buttocks to aid trunk movement
  - Rolls onto one side and uses contra lateral UE for push off
- 5 Independent, no compensatory strategies, transient use of upper extremities is acceptable
- 6 Independent in 3 seconds or less

Start this section by asking the patient to see if they are able to stand up from the chair without using their arms. If they can, then ask them to repeat the activity using the instructions below.

**Sit to stand (0-6)**

Ability to stand from a chair that allows the patient to sit with their knees and hips at approx 90 degrees and their feet flat on the floor. The patient may use one or both upper extremities, if necessary, but only after trying without using the upper extremities first.

**Instruction:** I'd like you to get up out of the chair and stand up straight, as quickly as you can. I'd like you to try this without using your hands.

- 0 Requires mechanical aid
- 1 Requires assistance of 2
- 2 Requires assistance of 1
- 3 Requires verbal cueing for safety and/or task description - supervision
- 4 Uses compensatory strategies- some examples are listed you may wish to add others
  - Use of UE's
  - Repeated forward flexion of trunk to initiate the movement i.e. use of momentum
  - Cushions on chair enable independence by reducing the height from which the patient has to rise
  - \_\_\_\_\_
  - \_\_\_\_\_
- 5 Independent in > 2 seconds
- 6 Independent in 2 seconds or less

**Standing ability & stability (0-13)**

The ability to stand in the most upright posture for a given patient. With the therapist sitting/standing in front of the patient, ask the patient to stand up. One arm may be supported to allow the patient to position their feet, if the patient needs help positioning their feet; the therapist carries this out. The patient is asked if they are ready, support is removed and timing is begun (relevant for 3 onwards). The patient's arms should remain close their body.

- 0 Unable to stand unsupported
- 1 Able to stand with assistance of one
- 2 Able to stand with assistance of an aid
- 3 Immediate standing with no physical support
- 4 Feet apart, with broad base (> 20cm between medial aspect heels) – eyes open (EO) - maintains for 10 seconds
- 5 Feet apart, with broad base (> 20cm between medial aspect heels) – eyes closed (EC) - maintains for 10 seconds
- 6 Feet apart, with normal base – EO - maintains for 10 seconds
- 7 Feet apart, with normal base – EC - maintains for 10 seconds
- 8 Feet together – EO - maintains for 10 seconds
- 9 Feet together – EC - maintains for 10 seconds
- 10 Semi tandem, the heel of one foot is placed beside the base of the first toe of the opposite foot – EO - maintains for 10 seconds
- 11 Semi tandem, the heel of one foot is placed beside the base of the first toe of the opposite foot – EC - maintains for 10 seconds
- 12 Tandem, the heel of one foot is placed in front of the toes of the other foot – EO - maintains for 10 seconds
- 13 Tandem, the heel of one foot is placed in front of the toes of the other foot – EC - maintains for 10 seconds

**Reach & Lift (0-6)**

The patient is asked to lift a textbook or equipment catalogue from a standard gym table, hand to the physiotherapist who places his/her receiving hand at a height that requires the patients to achieve 90 degrees flexion at the shoulder or in the case of limited shoulder movement, the greatest range available. In the case of one non-functioning upper extremity, one arm may be used.

This is a timed task. It commences with the patient in standing.

**Instructions** 'when I say go, I want you to take the book from the table and hand it to me'.

- 0 Unable
- 1 Able with assistance of one, for the purposes of stability
- 2 Requires use of compensatory strategies
  - Needs to hold object e.g. aid or table, on reaching up
- 3 Independent in > 6 seconds
- 4 Independent in 4.6-6 seconds
- 5 Independent in 2.6-4.5 seconds
- 6 Independent in 2.5 or fewer seconds

**Bend & Reach (0-6)**

Place a cone on the floor approximately one metre away from the patient on their dominant or unaffected side in the case of upper extremity dysfunction. Ask them to pick it up from the floor and stand up.

This is a timed task. The patient is in standing to start the test.

**Instructions** 'when I say go, I want you to pick the cone up from the floor'

- 0 Unable
- 1 Able with assistance of one, for the purposes of stability, throughout task
- 2 Requires use of compensatory strategies
  - Needs to lean on object to lean down or come up
  - Inadequate knee flexion – reaches from trunk
  - Needs assistance on rising
- 3 Independent in > 6 seconds
- 4 Independent in 4.6-6 seconds
- 5 Independent in 2.6-4.5 seconds
- 6 Independent in 2.5 or fewer seconds

Repeated Sit to Stand (0-8)

You will know already if the patient can go from sit to stand without arms, from the earlier test.

Patient is asked to stand up and sit down 5 times from a standard chair.

Time starts at initial sitting position to the final standing position at the fifth stand.

Instructions 'when I say go, I want you to stand up and sit down as quickly as you can, 5 times'. Please try not to use your arms to help you push up.

Verbal cueing to count the number of stands is allowed. No encouragement should be given.

- 0 Unable to complete
- 1 Completed in greater than or equal to 16.7 seconds, with UE use
- 2 Completed in 13.7-16.6 seconds, with UE use
- 3 Completed in 11.2-13.6 seconds, with UE use
- 4 Completed in a time less than or equal to 11.1 seconds, with UE use
- 5 Completed in greater than or equal to 16.7 seconds, without UE use
- 6 Completed in 13.7-16.6 seconds, without UE use
- 7 Completed in 11.2-13.6 seconds, without UE use
- 8 Completed in a time less than or equal to 11.1 seconds, without UE use

Gait ability (0-12)

This test assesses gait ability and speed, at the higher levels of performance. The patient is asked to walk 'at your normal comfortable speed' along a stretch marked out on the floor in the gym or with cones, if necessary. Approximately 2m acceleration and deceleration are allowed. The patient uses their assistive device. A frame includes a rollator or motorised frame or 2 crutches. A cane also incorporates the use of one crutch or one elbow crutch.

Timing occurs over 10m of the course, following the length of acceleration.

- 0 unable to mobilise or complete 10m
- 1 requires assistance of 2 people to mobilise
- 2 requires maximum assistance
- 3 requires moderate assistance
- 4 requires contact guard assistance/verbal cueing/supervision
- 5 completes with frame independently in > 30 seconds
- 6 completes independently with a frame in <30 seconds
- 7 completes independently with a cane in > 10 seconds (male)/>11 seconds (female)
- 8 completes independently with a cane in 7-10 seconds (male)/7-11 seconds (female)
- 9 completes independently with a cane in < 7 seconds (male)/<7.5 seconds (female)
- 10 completes independently with no aid > 10 seconds (male)/>11 seconds (female)
- 11 completes independently with no aid in 7-10 seconds (male)/7-11 seconds (female)
- 12 completes independently with no aid in <7 seconds (male)/<7.5 seconds (female)



**Gait pattern**

The patient is asked to walk along a line on the floor in the gym, turning at each end and repeating as requested. The patient is asked to walk at a comfortable speed and may use the assistive device they normally require.

0 is scored if the patient is unable to ambulate at all.

**GAIT INITIATION** - A smooth start is absent, the patient presents with start hesitation or freezing

Normal	2	Abnormal	1
--------	---	----------	---

**SWING PHASE** – Initial swing → mid swing → terminal swing

**Right**

Abnormalities in clearance

Normal	2	Abnormal	1
--------	---	----------	---

Step length

Normal	2	Abnormal	1
--------	---	----------	---

**Left**

Abnormalities in clearance

Normal	2	Abnormal	1
--------	---	----------	---

Step length

Normal	2	Abnormal	1
--------	---	----------	---

**STANCE PHASE** –

Initial contact → Foot flat → mid stance → terminal stance → Toe off  
Single and double leg stance

**Right**

Initial contact

Heel strike	Normal	2	Abnormal	1
-------------	--------	---	----------	---

Foot flat	Normal	2	Abnormal*	1
-----------	--------	---	-----------	---

\* not controlled

During stance

Hip control	Normal	2	Abnormal	1
-------------	--------	---	----------	---

- May be noted as abnormal foot placement i.e. toe in, toe out

- Pain on loading

Knee control	Normal	2	Abnormal	1
--------------	--------	---	----------	---

- Potentially hyperextension at the knee joint

- Inability to extend knee appropriately

Trunk control	Normal	2	Abnormal	1
---------------	--------	---	----------	---

- Support required at the lower trunk and pelvis to maintain level of pelvis during single leg stance

Toe off	Normal	2	Abnormal (not forefoot)	1
---------	--------	---	-------------------------	---

**Left**

Initial contact				
Heel strike	Normal	2	Abnormal	1
Foot flat	Normal	2	Abnormal*	1

\*not controlled

**During stance**

Hip control	Normal	2	Abnormal	1

- May be noted as abnormal foot placement i.e. toe in, toe out
- Pain on loading

Knee control	Normal	2	Abnormal	1

- Potentially hyperextension at the knee joint
- Inability to extend knee appropriately

Trunk control	Normal	2	Abnormal	1

- Support required at the lower trunk and pelvis to maintain level of pelvis during single leg stance

Toe off	Normal	2	Abnormal (not forefoot)	1
---------	--------	---	-------------------------	---

**Turning – en bloc, turn hesitation**

	Normal	2	Abnormal	1
--	--------	---	----------	---

**Step continuity – look for toe off corresponding to heel strike on opposite side**

	Normal	2	Abnormal	1
--	--------	---	----------	---

**Arm swing (not present if using frame)**

	Normal	2	Abnormal	1
--	--------	---	----------	---

*APPENDIX 10 ELDERLY MOBILITY SCALE AND PERFORMANCE  
ORIENTED ASSESSMENT OF MOBILITY*

*Appendix 10A Elderly Mobility Scale (Smith, 1994)*

Lying to sitting	Independent (2)	Help +1 (1)	Help +2 (0)
Sitting to lying	Independent (2)	Help +1 (1)	Help +2 (0)
Sit to stand	Independent <3 sec (3)	Independent >3 sec (2)	
	Help +1 (1)	Help +2 (0)	
Stand	Without support, able to reach (3)		
	Without support, not able to reach (2)		
	Stands with support i.e. use of UE's (1)		
	Stands only with physical support (0)		
Gait	Independent (incl. cane) (3)		
	Independent with frame (2)		
	Mobile with aid, needs occasional supervision (1)		
	Needs physical help or constant supervision (0)		
Timed 6m walk	< 15 seconds (3)	15-30 seconds (2)	> 30 seconds (1)
	Unable to cover 6m (0)		
Functional Reach	>16 cm (4)	8-16 cm (2)	Under 8cm or unable (0)

*Appendix 10B Performance oriented assessment of mobility (Tinetti, 1986)*

## Balance Tests

Initial instructions: Subject is seated in a hard, armless chair. The following manoeuvres are tested.

1. Sitting balance  
Leans or slides in chair = 0  
Steady, safe = 1
2. Arises  
Unable without help = 0  
Able, uses arms to help = 1  
Able without using arms = 2
3. Attempts to rise  
Unable without help = 0  
Able, requires more than one attempt = 1  
Able to rise, 1 attempt = 2
4. Immediate standing balance (first 5 seconds)  
Unsteady (staggers, moves feet, trunk sway) = 0  
Steady, but uses walker or other support = 1  
Steady without walker or other support = 2
5. Standing balance  
Unsteady = 0  
Steady but wide stance (medial heels more than 4 inches apart) and uses cane  
or other support = 1  
Narrow stance without support = 2
6. Nudged (sternum at maximum position with feet as close together as possible, three times)  
Begins to fall = 0  
Staggers, grabs, catches self = 1  
Steady = 2
7. Eyes closed  
Unsteady = 0  
Steady = 1
8. Turning 360 degree  
Discontinuous steps = 0  
Unsteady steps (grabs, staggers) = 1  
Continuous steps = 2
9. Sitting down  
Unsafe (misjudges distance, falls into chair) = 0  
Uses arms or not smooth motion = 1  
Safe, smooth motion = 2

Balance score            /16

## Gait Tests

Initial instructions: Subject stands with examiner, walks down hallway or across room, first at usual pace, then back at rapid but safe pace (usual walking aids)

10. Initiation of gait (immediately after told to go)  
Any hesitancy or multiple attempts to start = 0  
No hesitancy = 1
11. Step length and height
  - (a) Right swing foot  
Does not pass left stance foot with step = 0  
Passes left stance foot = 1  
Right foot does not clear floor completely with step = 0  
Right foot completely clears floor = 1
  - (b) Left swing foot  
Does not pass right stance foot with step = 0  
Passes right stance foot = 1  
Left foot does not clear floor completely with step = 0  
Left foot completely clears floor = 1
12. Step symmetry  
Right and left step not equal (estimate) = 0  
Right and left step appear equal = 1
13. Step continuity  
Stopping or discontinuity between steps = 0  
Steps appear continuous = 1
14. Path (estimated in relation to floor tiles, 12 inch diameter, observe excursion of 1 foot over about 10 feet of the course)  
Marked deviation = 0  
Mild/moderate deviation or uses walking aid = 1  
Straight without walking aid = 2
15. Trunk  
Marked sway or uses walking aid = 0  
No sway, but flexion of knees or back pain or spreads arms while walking = 1  
No sway, no flexion, no use of arms, and no use of walking aid = 2
16. Step width  
Heels apart = 0  
Heels almost touching while walking = 1

Gait score /12

Total score /28

*APPENDIX II      ETHICAL APPROVAL- CORRESPONDENCE*



Dan Lynch, Joint Research Ethics Committee Secretariat.  
Telephone : 4142860. Fax : 4142371. Email: [dan.lynch@amnch.ie](mailto:dan.lynch@amnch.ie)

THE ADELAIDE & MEATH  
HOSPITAL, DUBLIN  
INCORPORATING  
THE NATIONAL CHILDREN'S HOSPITAL

TALLAGHT, DUBLIN 24, IRELAND  
TELEPHONE +353 1 4142000

Ms. Emma Stokes,  
Lecturer,  
School of Physiotherapy  
Trinity Centre for Health Sciences,  
St. James's Hospital,  
James's Street, Dublin 8.

6<sup>th</sup> July, 2001.

RE : Outcome measurement in physiotherapy – evaluation of a new scale.

Please quote this reference in all  
communications regarding this study : 010603 / 9301

Dear Ms. Stokes,

The Joint Research Ethics Committee at its meeting on 19<sup>th</sup> June 2001 agreed to give ethical approval to the above study subject to the following condition :

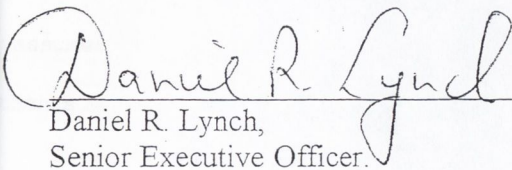
The first paragraph on the second page of the Patient Information Leaflet reads :

“.....It will take 20 minutes and will interfere with your normal treatment.....”

The Committee assumes that this should read “.....will *not* interfere.....”

The Investigator is asked to confirm that this assumption is correct.

Yours sincerely,

  
Daniel R. Lynch,  
Senior Executive Officer.

cc Dr. Desmond O'Neill,  
Dr Bernard Walsh,  
Prof. Davis Coakley,  
Dr. Conal Cunningham.

From: Dan Lynch <Dan.Lynch@amnc.ie>  
To: "'estokes@tcd.ie'" <estokes@tcd.ie>  
Cc: Des O'Neill <Des.O'Neill@amnc.ie>,  
"jbwalsh@stjames.ie"  
<jbwalsh@stjames.ie>,  
'dcoakley@stjames.ie' <dcoakley@stjames.ie>,  
'ccunningham@stjames.ie' <ccunningham@stjames.ie>

Subject: RE: Ethics query  
Date: Wed, 11 Jul 2001 21:22:21 +0100  
Importance: low  
X-Priority: 5  
MIME-Version: 1.0  
Status: RO  
X-Status:  
X-Keywords:  
X-UID: 1

Yes Emma, You may take it that ethical approval for this study is now complete and unconditional.

Kind regards,

Dan.

-----Original Message-----

From: estokes@tcd.ie [mailto:estokes@tcd.ie]  
Sent: 11 July 2001 10:25  
To: dan.lynch@amnc.ie  
Cc: des.oneill@amnc.ie; jbwalsh@stjames.ie; dcoakley@stjames.ie;  
ccunningham@stjames.ie  
Subject: Ethics query

Re: 010603/9301

Dear Dan,

Thanks for your letter regarding the application for Ethics approval for the outcome measure study.

The first paragraph on the second page of the Patient Information Leaflet should it fact read 'will not interfere with your normal treatment'.

Apologies for the typo on my part.

Can I assume that this e-mail serves to answer the query and proceed.

Many thanks.

Emma

Emma K. Stokes  
School of Physiotherapy