The reliability of calf and thigh measurements for the detection of deep vein thrombosis in patients with spinal cord injuries

Ellen Eugarde Bachelor of Nursing, Bachelor of Nursing Informatics

30/05/2013

School of Nursing and Midwifery
Faculty of Health
Griffith University

Submitted in fulfilment of the requirements of the degree of

Master of Philosophy
Abstract

Venous thrombo-embolism (VTE) has been recognised as a leading cause of morbidity and mortality in the acute spinal cord injury (SCI) population. It is thought that deep vein thrombosis (DVT), the most common form of VTE will occur in the majority of patients in the absence of adequate prophylaxis in the SCI group. Pulmonary embolism, which is often preceded by a DVT, is a leading cause of death in the first year after injury for this group of patients.

The reliability of signs and symptoms in the diagnosis of thrombo-embolism has been debated in medicine since the 1960s; however diagnosis of DVT remains problematic. The most commonly used of these signs and symptoms in the SCI patient group is swelling of the calves and/or thighs as indicated by an increase in a daily circumference measurement. After a review of the literature it was identified that there were several aspects of reliability of the current practice of lower limb circumference measurements that had been poorly examined including the location the measurements are performed, the landmarks used to determine the location to be measured, the type of tape measure being used and the inter-rater reliability of the procedure.

The aim of this study was to test the reliability of calf and thigh circumference measurements by testing the stability and equivalence of the equipment, the procedure and the measurements. This three phase prospective study was designed to determine how much the Birch tape measure stretches over time, if the standard (Birch) tape measure can be used instead of the Gulick spring tape for circumference measures, at which measurement location could the most consistent measurements be produced and whether it was possible for staff to consistently reproduce measurements over time.
In Phase I 20 Birch tape measures currently in clinical use were examined for stretch over time. Tape measures were found to have stretched between 1 and 7mm but this was not dependent on time in use. In Phase II the limits of agreement method was used to assess the equivalency of the Birch tape with the Gulick spring tape at the calf and thigh location. A total of 402 measurements at each location were collected. The limits of agreement for the measurement differences between the two tapes were wider at both locations (calf: -3.66 to 1.66 cm and thigh: -6.19 to 2.79 cm) than the +/- 1.5 cm set a priori. Some of this lack of agreement was attributed to poor reproducibility with both tape measures, with reproducibility coefficients wider than the limits of agreement at both the calf (Birch +/- 2.64cm and Gulick +/- 2.84 cm) and thigh (Birch +/- 4.15 cm and for the Gulick +/- 4.65 cm) locations.

In Phase III the consistency of measurements performed at four separate locations, the ankle, mid-calf, calf and thigh across three time points was assessed. Relative standard deviations calculated from the first time point (baseline) indicated that calf measurements were the most consistently performed followed by the ankle measurements. Mid-calf and thigh measurements were found to be the least consistent.

Two way repeated measures analysis of variance were performed for all four locations to assess inter-rater reliability. No significant variation in measurement was found between examiners at any of the locations. The effect of time by group membership was found to be approaching significance at the mid-calf level only (p=0.05). A significant effect of time on measurements was found at the mid-calf level (p=0.002) and the calf level (p=<0.001) indicating poor repeatability of measurements at these locations.

Recommendations for practice based on the results of the study include the continued use of the Birch tape measure over the Gulick spring tape as measurement reproducibility was not improved with the Gulick spring tape and the stretching of the Birch tape does not
appear to be dependent on time in use. The continued use of the calf location is supported by low variation in measurements when based on a single measurement time point. Importantly, the thigh location varied significantly in measurements by examiners indicating that this location should be used with caution. The finding of good inter-rater reliability and low variation in measurements of the ankle location indicates the need for further investigation into the specificity of ankle measurements in DVT diagnosis.

The practice of routine lower limb circumference measurements is a relatively simple, cheap and non-invasive method of DVT surveillance in the spinal cord injured patient. However poor reliability was found in many aspects of the practice of circumference measurements in this study. Without a reliable measure, the everyday practice of calf and thigh measurements is meaningless, with the potential that the diagnosis of DVT in SCI patients is delayed or undetected. There is a need for further examination of the reliability of this fundamental surveillance tool as evidenced by the results of this study and the paucity of prior research in this area.
Statement of originality

This work has not previously been submitted for a degree or diploma in any university. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.

Ellen Eugarde
Acknowledgements

I would like to acknowledge Professor Leanne Aitken and Associate Professor Marion Mitchell, for their ongoing guidance and support throughout this process. Thankyou for talking me off the ledge and making me finish this even when I drove you insane. It has been a privilege to work with you both.

Thankyou to Covidien Pty Ltd for providing me with a research grant which was invaluable in the undertaking of the research.

I would like to thank Ms Liz Burmeister, Ms Bonnie MacFarlane and Ms Helen Skerman for their assistance and support with the data analysis. Many thanks also to Ms Margaret Lavery for her assistance with the editorial process.

I would like to offer my most sincere thanks to my colleagues, the nursing staff of the SIU, who willingly gave me their time and their legs despite their aversion to research.

Thank you to my long suffering family and friends, without you none of this would be possible, especially my husband; my IT support and my confidante, who as always, held my hand, nodded and offered advice even when you had no clue what I was talking about.
## Contents

Abstract........................................................................................................................................i
Statement of originality ...................................................................................................................v
Acknowledgements ......................................................................................................................vii
Contents ........................................................................................................................................ix
List of Tables ..................................................................................................................................xiii
List of Figures ..................................................................................................................................xv
List of Abbreviations .....................................................................................................................xvii
Definition of Terms ......................................................................................................................xix

1.0 Introduction .............................................................................................................................1
  1.1 Background ...........................................................................................................................2
  1.2 Purpose of study ....................................................................................................................3
  1.3 Research Aims ......................................................................................................................4
  1.4 Significance ..........................................................................................................................4
  1.5 Thesis Structure ...................................................................................................................5

2.0 Literature Review .....................................................................................................................7
  2.1 Search Strategy .....................................................................................................................7
  2.2 Venous Thromboembolism ...................................................................................................8
    2.2.1 Thrombus Pathophysiology .........................................................................................8
    2.2.2 Secondary Complications ..........................................................................................11
  2.3 Incidence of DVT ................................................................................................................13
  2.4 Diagnosis .............................................................................................................................14
    2.4.1 Objective Measures .......................................................................................................14
    2.4.2 Signs and Symptoms .....................................................................................................18
  2.5 Reliability of Measurements ................................................................................................31
    2.5.1 Types of Reliability .......................................................................................................32
  2.6 Rationale for Study ..............................................................................................................35

3.0 Methods ...................................................................................................................................39
  3.1 Introduction ..........................................................................................................................39
5.3 Inter rater reliability.................................................................99
5.4 Repeatability by Location ......................................................101
5.6 Strengths and Limitations ......................................................103
5.7 Implications for Practice ......................................................106
5.8 Implications for Education ....................................................107
5.9 Implications for Further Research ........................................109
6.0 Conclusion .............................................................................113
References .................................................................................119
APPENDICES .............................................................................129
  Appendix A: Informational Flyer ................................................129
  Appendix B: Data collection Form Phase I ....................................131
  Appendix C: Data Collection Form Phase II ..................................133
  Appendix D: Data Collection Form Phase III ...............................135
  Appendix E: Tape Measures Used in Study ...................................137
  Appendix F: Location where Circumference Measurements were performed .... 139
  Appendix G: Office of the Human Research Ethics Committee PAH Approval Letter .................................................................141
  Appendix H: Human Research Ethics Committee PAH Approval Letter for amendment to Examiner Informed Consent Form and Healthy Volunteer Informed Consent Form .................................................................143
  Appendix I: Griffith University HREC Approval ................................145
  Appendix J: Examiner Informed Consent Document ......................147
  Appendix K: Health Volunteer Participant Informed Consent Document........153
List of Tables

Table 1. Summary of Literature Search ................................................................. 7
Table 2. Wells' clinical prediction rule for deep vein thrombosis (DVT) .............. 19
Table 3. Summary table of studies examining calf circumference measurements ....... 30
Table 4. Outline of study design ........................................................................... 42
Table 5. Overview of study .................................................................................. 57
Table 6. Calf measurements using the Birch and Gulick spring tapes (n=804) ........ 61
Table 7. Thigh measurements using the Birch and Gulick spring tapes (n=804) ...... 63
Table 8. ANOVA of Birch tape measurements at calf location (n=402) ............... 72
Table 9. ANOVA of Gulick spring tape measurements at calf location (n=402) ....... 72
Table 10. ANOVA of Birch tape measurements at the calf location (n=402) .......... 74
Table 11. ANOVA of Gulick spring tape measurements at the thigh location (n=402) ...... 74
Table 12. Descriptive statistics at each measurement location (n=348) ................. 77
Table 13. Consistency across landmarks using relative standard deviation (rsd) (n=16) ....... 79
Table 14. Two factor repeated measures ANOVA for each location (n=384) ........... 81
List of Figures

Figure 1. Components of Observed Measure Score ................................................................. 31
Figure 2. Framework for the Reliability of Circumference Measurements .............................. 37
Figure 3. Changes in Tape Measure Length vs. Days in use ................................................. 59
Figure 4. Box plot of distribution of calf measurements (n=402) from each tape ............... 61
Figure 5. Distribution of calf measurement differences between tape measures (n=402) ..... 62
Figure 6. Box plot of distribution of thigh measurements (n=402) for each tape showing outliers ....................................................................................................................... 63
Figure 7. Number of measurement differences in thigh circumferences (n=402) taken with Gulick spring tape and Birch Tape. .................................................................................. 64
Figure 8. Difference versus average of values measured by examiners at the calf location ... 65
Figure 9. Difference versus average of values measured by examiners at the thigh location. 65
Figure 10. Distribution of differences between Birch and Gulick tapes at calf level showing +/- 2SD from mean (n=402) ....................................................................................... 66
Figure 11. Distribution of differences between Birch and Gulick tapes at thigh level showing +/- 2SD from the mean (n=402) ................................................................................... 67
Figure 12. Birch tape calf measurements plotted against Gulick spring tape calf measurements (cm) with line of identity .......................................................... 68
Figure 13. Birch tape thigh measurements plotted against Gulick tape thigh measurements (cm) with line of identity .......................................................... 68
Figure 14. Bland Altman plot of differences between Birch tape measure and Gulick spring tape in relation to the mean of the two tape measures at the calf level (n=402) .......... 70
Figure 15. Bland Altman plot of differences between the Birch tape and Gulick spring tape measures in relation to the mean of the two tape measures at the thigh level (n=402)....71

Figure 16. Bland Altman plot showing reproducibility coefficients for Birch and Gulick spring tapes at calf location (cm)..............................................................73

Figure 17. Bland Altman plot showing reproducibility coefficients for Birch and Gulick spring tapes at thigh location (cm) .................................................................74

Figure 18. Distribution of measurements at each location .................................................................78

Figure 19. Distribution of mean measurements by time point at ankle location (n=348).......83

Figure 20. Mean measurements by time point for each examiner at ankle location ...........83

Figure 21. Distribution of mean measurements by time point at calf location (n=348) ........84

Figure 22. Mean Measurements by Time point for each Examiner at Calf Location ..........84

Figure 23. Distribution of Measurements by Time point at Mid-calf location (n=348) ........85

Figure 24. Mean measurements by Time point for each Examiner at the Mid-calf Location 85

Figure 25. Distribution of Measurements by Time point at the Thigh Location (n=348) ......86

Figure 26. Mean Measurement by Time point for each Examiner at the Thigh Location .....86
List of Abbreviations

AIN    Assistant in Nursing
CI     Confidence Interval
cm     Centimetres
d      Mean
DVT    Deep Vein Thrombosis
EEN    Endorsed Enrolled Nurse
GCS    Graduated Compression Stockings
ICC    Intra-class Correlation Coefficient
LOA    Limits of Agreement
mm     Millimetres
PAH    Princess Alexandra Hospital
PE     Pulmonary Embolism
PTS    Post Thrombotic Syndrome
RN     Registered Nurse
rsd    Relative standard deviation
SCI    Spinal Cord Injury
sd     Standard Deviation
SIU    Spinal Injuries Unit
sw     Within Subject standard deviations
VTE    Venous Thromboembolism
Definition of Terms

**Accuracy**: Accuracy describes how closely the measurement is to the true measurement.

**Agreement**: The degree to which two measurements of the same object agree. Usually the measurements are taken using different devices.

**Deep vein thrombosis (DVT)**: A blood clot that forms in one of the deep veins of the venous system more commonly found in the legs but may occur in the arms or the abdomen.

**Inter-rater reliability**: The extent to which two or more examiners agree.

**Precision**: The degree to which a result is reproduced each time the measurement is repeated.

**Pulmonary Embolism (PE)**: A blood clot that develops in the blood supply to the lungs. A PE may develop spontaneously but more often develops from a piece of a DVT that has become dislodged.

**Reliability**: The degree to which the measurement is free from measurement error.

**Repeatability**: The variability in repeated measurements by one observer when all other factors are constant.

**Reproducibility**: The variability in repeated measurements when one or more factors are varied.

**Stability**: Stability refers to the device or procedure’s uniformity over time.

**Venous thrombo-embolism (VTE)**: Over arching name given to the different types of blood clots that inhabit the venous system.
1.0 Introduction

Venous thrombo-embolism (VTE) has been recognised as a leading cause of morbidity and mortality in the acute spinal cord injury (SCI) population (Aito, Pieri, D'andrea, Marcelli, & Cominelli, 2002). Deep vein thrombosis (DVT), the most common form of VTE, has a reported incidence ranging from a low of 10% to a high of 90-100% in the absence of adequate prophylaxis in the SCI population (Aito et al., 2002; Green et al., 1990). Pulmonary embolism (PE), which is often preceded by a DVT, is a leading cause of death in the first year after injury for this group of patients (Green et al., 1990).

Diagnosis of DVT is established by recognition of clinical symptoms followed by confirmation with an objective diagnostic procedure such as Doppler ultrasonography. The reliability of signs and symptoms in the diagnosis of thrombo-embolism has been debated in medical literature since the 1960s (Brown & Kros, 2003; Haeger, 1969); however, diagnosis of DVT remains problematic.

The most commonly used of these diagnostic signs and symptoms in the SCI group is swelling of the calves and/or thighs as indicated by an increase in circumference measurements that are usually recorded on a daily basis. Previous studies have examined aspects of calf and thigh circumference measurements in a wide range of situations including inpatient and outpatient groups (Johanning, Franklin, Thomas, & Elmore, 2002; Stein, Henry, Gopalakrishnan, & Relyea, 1995), comparisons with objective diagnostic procedures (Diamond & Macciocchi, 1997; Stein et al., 1995) and the effectiveness of numerous cut-off measurements (Swarzinski & Dijkers, 1991), with differing opinions on the validity of the measurements and their usefulness (Haeger, 1969; Stein et al., 1995; Swarczinski & Dijkers, 1991). To date, however, the reliability of circumference measurements has been poorly examined in the literature.
1.1 Background

The practice of calf and thigh circumference measurements is endorsed by the Consortium for Spinal Cord Medicine, an organisation responsible for the publication of practice guidelines for the management of conditions associated with SCI (Green et al., 2002). It is difficult to determine from the literature, however, just how widespread this practice is in spinal units worldwide. Just one other study, which was conducted in the United States, could be found that specifically outlined the protocol used for circumference measurements within a spinal injuries unit (Swarzcinski & Dijkers, 1991). The use of regular calf and thigh circumference measurements for the surveillance and detection of DVT is standard practice within the spinal injuries unit (SIU) at the Princess Alexandra Hospital (PAH).

Despite a few differences, the PAH protocol is mostly in agreement with the protocol identified in the previous study. In the US, the measurements are usually performed on admission to the ward and then daily in the morning, using a standard sewing tape, such as the Birch tape measure, at pre-determined reference points using anatomical landmarks (Swarzcinski & Dijkers, 1991).

At the PAH reference points are determined on admission to the SIU when baseline measurements are performed. The first of these reference points is determined by measuring from the base of the patella to the widest part of the calf with the patient lying in a flat supine position. Two marks 1cm apart are then made using an indelible ink pen. The distance from the patella to the widest part of the calf is then recorded on a calf and thigh measurement log sheet to ensure that all future measurements are made in the same location. The thigh reference point is similarly recorded with measurement made from the top of the patella to the widest part of the thigh. The circumference is measured by placing the tape measure
between these two marks and looping it around the leg. The measurement is taken from the zero mark on the tape measure.

Measurements are performed on a daily basis with the patient lying flat in a supine position, prior to rising from bed. The measurements are recorded on a signed log sheet, which is kept in the patient record at the bedside. The PAH unit protocol states an abnormal reading to be a calf or thigh circumference measurement greater than 1.5cm when compared to baseline measurements, the previous day’s measurements or the other limb. Abnormal findings are reported to the medical officer for further investigation, which in the first instance is often a Doppler ultrasound.

Some major concerns have been identified with the PAH’s current procedure. Firstly, the widest part of the calf and the widest part of the thigh are identified visually by the staff member recording baseline measurements. This estimation may differ between assessors. Secondly, the patella has been shown in previous studies to have poor inter-rater reliability as a landmark. It is suggested that mobility in the patella may result in inconsistency in subsequent measurements. Lastly, the use of sewing tape measures, such as the Birch tape, is not considered to be best practice for circumference measurements. This type of tape measure is thought to stretch with repeated use and poor inter-rater reliability has been identified in some studies using this type of tape, possibly due to varying amounts of tension applied by the users. It appears the procedure outlined above is not based on evidence but rather a practice that has been in place for more than 10 years handed on from one nurse to another.

1.2 Purpose of study

Circumference measurements are a multi-step procedure and, whilst it is considered a relatively simple task, there are many different aspects of the procedure that raise questions of reliability. The purpose of this study was to examine the various components that make up the procedure of calf and thigh circumference measurements and in turn to assess the
reliability of each of those components. The equipment, the procedure and the examiners undertaking the procedure were all identified as possible sources of error in circumference measurements. This study sought to assess the equivalence and stability of these factors in order to add to the knowledge base on this subject and in turn inform current practice.

1.3 Research Aims

The aim of this project was to examine the reliability of calf and thigh circumference measurements by testing the stability and the equivalence of the equipment, the procedure and the measurements. Without reliable measures, the everyday practice of calf and thigh measurements is meaningless with the potential that the diagnosis of DVT in SCI patients is delayed or undetected. With this in mind three research aims were developed:

1. Determine the reliability of the Birch tape measure.
2. Compare the reliability of the current landmark (patella) with landmarks identified in the literature.
3. Assess the inter-rater reliability of circumference measurements at different anatomical locations.

1.4 Significance

As mentioned above in section 1.1, the practice of calf and thigh circumference measurements is endorsed by the Consortium for Spinal Cord Medicine (Green et al., 2002). The Consortium, however, does not give an explanation of how the procedure is to be performed or indeed provide any evidence of the reliability of the measurements. The practice of calf and thigh circumference measurements is a relatively simple surveillance tool requiring few resources but without evidence of its reliability questions may be raised over its true cost effectiveness.

VTE is a costly illness for both the health care system and the patient. The financial cost to the health care system includes the use of diagnostic procedures, increased length of
stay and of course staffing costs for prevention, diagnosis and treatment of VTE. The cost of VTE to the patient may be far greater with the risk of developing chronic vascular insufficiency and PTS or worse still PE and death. Without knowing if circumference measurements are reliable it is difficult to know if some DVT cases are undiagnosed or alternatively if a number of unnecessary diagnostic procedures is performed.

Reliability of results is fundamental to any clinical measurement and the paucity of research in this area highlights the need for further examination of the reliability of this fundamental surveillance tool. In this study it is proposed to test the reliability of calf, thigh and ankle circumference measurements by multiple testers at multiple time-points.

1.5 Thesis Structure

The method and results of a study undertaken to examine the various issues related to the use of calf and thigh circumference measurements for the detection of DVT in spinal cord injured patients are presented in this thesis.

In chapter one the background to the research and the research aims and significance of the study has been outlined. Presented in chapter two is an overview of the literature detailing what is known about VTE and the practice of circumference measurements as well as the process of reliability studies. The quantitative methods used in the study and a detailed description of the participant recruitment, procedures, the analysis methods chosen and the ethical considerations in undertaking each phase of the study are outlined in chapter three. The results of the analysis of the study are presented in chapter four. Each phase is outlined with descriptions of the analysis undertaken and the results are presented in graphical detail. A discussion of the findings of the research in the context of the literature and the limitations of the study is presented in chapter five. The implications for clinical practice, education and further research are also presented here. The study conclusion is presented in chapter six with an overview of the study findings and its contribution to practice and research.
2.0 Literature Review

Calf and thigh measurements are a currently used method of surveillance for DVT in the SCI patient. The question of whether these measurements provide a reliable method of detecting a DVT is one that remains unanswered in the literature. Multiple concepts including the people performing the measurements, the equipment used, the location measured and the clinically significant difference in measurements may influence the reliability of calf and thigh circumference measurements and are discussed in the following chapter.

2.1 Search Strategy

A thorough review of the literature was undertaken using a variety of databases including CINAHL, MEDLINE and Health Reference Centre Academic. Keyword searches of these databases included the terms deep vein thrombosis, circumference measurements, spinal cord injury and reliability in a variety of combinations. Combination searches using MeSH terms were conducted and the resulting lists were refined by excluding non-English publications and expanded to cover the years 1972 to 2012 (see Table 1).

Table 1. Summary of Literature Search

<table>
<thead>
<tr>
<th>MeSH Terms</th>
<th>Limited by</th>
<th>Results</th>
<th>Used in Review</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Spinal cord injuries” and “venous thrombosis”</td>
<td>English only 1972-2012</td>
<td>141 papers</td>
<td>11 papers</td>
</tr>
<tr>
<td>“Reproducibility of Results” and Bland Altman</td>
<td>English only 1972-2012</td>
<td>2,624 papers</td>
<td>12 papers</td>
</tr>
<tr>
<td>“Circumference measurements”</td>
<td>English Only Subject “anthropometry”</td>
<td>32 papers</td>
<td>3 papers</td>
</tr>
<tr>
<td>“Anthropometry” methods and instrumentation</td>
<td>English only Heading “leg”</td>
<td>84 papers</td>
<td>4 papers</td>
</tr>
</tbody>
</table>
There was a lack of original research in the literature examining this topic and further examination of the literature was conducted through the use of references cited in articles. There was a wealth of narrative papers outlining the importance of the topic and supporting the need for further enquiry into the reliability of calf and thigh circumference measurements. The major areas of enquiry presented in the literature review are an outline of venous thromboembolism including thrombus pathophysiology and secondary complications of DVT, the incidence of DVT, the diagnosis of DVT with specific reference to the use of objective measures and clinical signs and symptoms, and the reliability of circumference measurements.

2.2 Venous Thromboembolism

Venous thromboembolism (VTE) is the overarching term used to describe a group of conditions in which a thrombus (blood clot) develops. Most commonly this manifests as a deep vein thrombosis (DVT), a thrombus in a vein of one of the extremities, usually the leg. However it is estimated that 30% of patients diagnosed with a DVT will have a portion of the thrombus dislodge and become a pulmonary embolus (PE) (Studies, 2003). A PE is a life threatening condition where a thrombus partially or fully occludes the main artery or one of its branches that supply blood flow to the lungs. Approximately 30% of patients who suffer from VTE may also develop post thrombotic syndrome (PTS); a condition causing chronic venous insufficiency and leg ulcers which result in protracted and an ongoing need for medical treatment (Morrison, 2006).

2.2.1 Thrombus Pathophysiology

A thrombus consists of erythrocytes, thrombocytes and fibrin (Cranley, Canos, & Sull, 1976; Wallis & Au, 2001). The development of most venous thrombi occurs behind the venous valves or at a vein junction where a large vein branches off to smaller ones (Fahey, 1989). The flow of blood is slower in the small pockets behind the valves and as the vein
walls are slightly larger here pooling can occur allowing coagulation of blood (Wallis & Autar, 2001). As early as the 1800s, three factors were identified as the cause of thrombosis—venous stasis, hyper-coagulability and vessel wall damage (Cranley et al., 1976). The unique nature of SCI means that often all three of these factors are present during the acute phase after injury, thus significantly increasing the risk of clot formation in this group of patients and the subsequent susceptibility to PEs and their complications including possibly death (Green, 1999).

2.2.1.1 Venous stasis

Venous stasis occurs in the absence of peristaltic contraction of the muscles that propel blood through the venous network. Three forces which cause stasis have been identified; (1) decreased velocity of the blood, (2) venous dilation and pooling and (3) obstruction (Cranley et al., 1976; Wallis & Autar, 2001). Immobilisation is the most common reason for stasis to occur, thus combining these three forces together (Cranley et al., 1976). When a person is subjected to prolonged bed rest due to illness or injury such as in SCI several things happen. There is an immediate decrease in the velocity of the blood flow through the vessels. This decrease in velocity causes pooling to occur which in turn causes venous dilation and congestion. When blood pools at the valves, coagulation factors are activated, and a thrombus is formed (Wallis & Autar, 2001).

Possibly the most well known example of venous stasis is as a result of long aeroplane flights (Kennedy, Setnik, & Li, 2001; Wallis & Autar, 2001). Venous stasis may also occur following operations where a general anaesthetic lasts longer than 30 minutes (Cranley et al., 1976). This is due to the lack of muscle activity in the calf that supports venous return. This function may remain impaired for as long as seven days post-operatively (Line, 2001).

Obstructions, which also cause stasis may result from tumours, peripheral vascular disease or trauma that impairs the venous blood flow (Line, 2001). Significantly, however,
Recent studies indicate that stasis alone may not be enough to cause a DVT and that the presence of either hyper-coagulability or vessel wall damage is also required for thrombus formation (Qaseem et al., 2007).

2.2.1.2 Hyper-Coagulability

Coagulability refers to the blood’s capacity to coagulate or clot. Coagulation in healthy individuals is necessary to maintain homeostasis in the blood supply in times of injury or illness (Line, 2001). Blood contains various enzymes in inactive forms which under normal circumstances allow the blood to flow freely. Under the right circumstances, these inactive enzymes may be activated in order to allow coagulation of the blood to prevent further bleeding. Coagulation is a three step process with each step triggering the next; hence it is commonly referred to as the coagulation cascade.

The cascade begins with the triggering of the prothrombin activator, which leads to conversion of prothrombin into thrombin and ultimately conversion of fibrinogen into fibrin. This coagulation process is usually balanced out with fibrin degradation, as the thrombi are absorbed back into the circulatory system (Cranley et al., 1976). In the case of hyper-coagulability the ability to thrombose overrides the fibrin degradation and as a consequence, larger thrombi are formed.

There are numerous reasons why the balance of components of the blood may be altered leading to an increase in coagulation. These include: malignancy, trauma, surgery, dehydration, burns, pregnancy, medication and old age (Cranley et al., 1976; Wallis & Autar, 2001). Hyper-coagulability may also result in the case of some hereditary conditions affecting the circulating anticoagulants such as deficiency of Antithrombin III, protein S, protein C resistance or deficiency and Factor V Leiden. (Cranley et al., 1976; Fahey, 1989)
2.1.1.3 Vessel Wall Damage
Under normal conditions the inside of the vein wall is smooth making it difficult for thrombi to adhere. Once damaged however, the vessel wall attracts platelets and fibrin in an attempt to repair itself and the vein itself may in turn become obstructed through this process (Church, 2000; Cranley et al., 1976; Fahey, 1989). Damage to the wall lining may occur in numerous ways. Orthopaedic surgery poses many risks through various mechanisms such as the use of general anaesthesia which decreases vascular tone, the actual twisting of vessels as in hip surgery and the use of compression bandages for long periods (Church, 2000; Wallis & Autar, 2001). The SCI patient is exposed to many of these mechanisms in undergoing complicated orthopaedic surgery in order to stabilise the spine. Localised damage to blood vessels may also result from infection, intravenous infusion of medication, insertion of peripheral vascular devices or venipuncture (Church, 2000; Cranley et al., 1976; Fahey, 1989).

2.2.2 Secondary Complications
The presence of a DVT in itself is often not clinically significant to a patient. Many patients with distal DVTs, that is a DVT occurring in the veins below the knee (popliteal, peroneal, anterior tibial and posterior tibial), display no clinical symptoms and suffer no long term effects from the thrombus (Day, 2003). The concern however once a DVT is diagnosed, is the prevention of secondary complications which can have serious consequences for a patient (Church, 2000; Day, 2003). The two most common secondary conditions of DVT are PEs and PTS.

2.2.2.1 Formation of PE
A PE occurs when a part or all of the thrombus becomes dislodged from the vein wall and travels through the circulatory system to the lung where it lodges in the pulmonary
circulatory system. Signs and symptoms of a PE include shortness of breath, tachypnoea, tachycardia, chest pain and an impending sense of doom. Depending on the size and location of the embolism, patients may display all or none of these symptoms (Wallis & Autar, 2001). Approximately 80% of clinically significant PEs are thought to arise from proximal DVTs. A proximal DVT is one which is present in the veins of the thigh and lower abdomen usually the femoral and popliteal veins (Stein et al., 1995; Wallis & Autar, 2001). Although PE can develop from thrombi found in the distal veins; these are usually smaller and occur less frequently (Line, 2001). Significantly though, in the acute SCI population the risk of death due to PE is 210 times that of healthy individuals (Green et al., 2002).

PE is thought to affect 30% of those diagnosed with DVT and can be fatal within the first 30 minutes of symptom onset (Studies, 2003). In 1995, PE was reported by DeVivo and Stover as the third leading cause of death among those with paraplegia and it was found to account for 14.9% of all deaths in the first year post injury in SCI patients making early detection of DVT in the SCI population extremely important (Green et al., 2002).

**2.2.2.2 Development of Post Thrombotic Syndrome**

Post thrombotic syndrome (PTS) is the name given to chronic venous insufficiency caused by DVT. Studies estimate one in every three patients with a DVT will develop PTS within five years of the thrombus with symptoms appearing as early as one year post thrombus development (Asbeutah, Riha, Cameron, & Mcgrath, 2004; Elton, 2004; Kahn, 2004; Kolbach, Sandbrink, Hamulyak, Prins, & Neumann, 2003).

Symptoms of PTS range from mild complaints of skin discoloration, discomfort or swelling of the affected limbs to severe life altering complications such as chronic pain, intractable oedema and venous leg ulcers (Kahn et al., 2008; Kolbach et al., 2003). In the SCI population this is most commonly characterised by severe oedema of the affected limbs with many patients having to wear compression stockings during the day to minimise the swelling.
2.3 Incidence of DVT
The Australian annual incidence of DVT is approximately one per 1000 but varies
with age and rises to 10 per 1000 in the elderly (Studies, 2003). VTE is reported as the most
common preventable cause of death in hospitalised patients today. The incidence of VTE is
100 times greater for hospitalised patients than those in the community, and PEs account for
10% of all deaths in hospital in Australia (Clinical practice guideline for the prevention of
venous thromboembolism (deep vein thrombosis and pulmonary embolism) in patients
admitted to australian hospitals., 2009).

Of all hospitalised patients the risk for VTE is greatest in the SCI population with
estimates that VTE occurs in up to a third of patients with SCI (Connelly, 2008). Because of
this known high risk, both mechanical and chemical prophylactic measures are used to
minimise the chance of thrombus formation. Mechanical measures include the use of
properly fitted graduated compression stockings (GCS) and sequential compression devices
(SCD), also known as intermittent pneumatic compression devices (IPCD). Chemical
prophylaxis usually consists of one of a variety of anticoagulant medications available as
either a subcutaneous injection or oral preparation.

In the SCI population the documented incidence of DVT, ranges from a low of 10% to
as high as 90-100% in the absence of adequate prophylaxis (Aito et al., 2002; Green et al.,
1990). More recent reviews put the incidence closer to 10-40% (Aito et al., 2002; Anderson
& Spencer, 2003) which is thought to be due to a greater awareness by health professionals of
the risk of VTE and more consistent use of prophylactic measures (Chen, Apple, Huson, &
Bode, 1999).

It is thought that up to 50% of diagnosed DVT will embolise and that between one
half and two thirds will lead to post thrombotic syndrome (PTS) in SCI patients (Green et al.,
2002). Pulmonary embolism (PE) is the third leading cause of death in the first year post SCI
making VTE prevention, diagnosis and treatment a major concern for all health care workers (Green et al., 2002).

2.4 Diagnosis

Diagnosis of DVTs was solely reliant on the classic clinical signs and symptoms of; Homan’s sign (pain on dorsiflexion of the foot), tenderness, fever, warmth, and erythema and swelling over the site of the DVT. Since the 1960s several significant studies comparing these signs and symptoms to objective measures, such as venography, phlebography, ultrasound and impedance plethysmography have cast doubt on the usefulness of the signs and symptoms in diagnosing DVT. The conclusion was that overall sensitivity using signs and symptoms was no better than by chance (Brown & Kros, 2003; Diamond & Macciocchi, 1997; Haeger, 1969).

The signs and symptoms of DVT may be indicative of other disease processes such as muscle haemorrhage, sciatic pain or cellulitis (Brown & Kros, 2003). For this reason, much of the research surrounding thromboembolism has focused on the utilisation of objective diagnostic tools for diagnosis of DVTs including the use of venograms, Doppler ultrasound, plethysmography and d-dimer assay tests.

2.4.1 Objective Measures

2.4.1.1 Venography

Venography is considered the ‘gold standard’ for definitive diagnosis of DVTs (Cranley et al., 1976). A study of patients post-mortem venograms showed a sensitivity of 95% and a specificity of 97% (Line, 2001). That is, the venograms were able to detect 95% of DVTs present in study participants and correctly identify 97% of participants without DVT. The procedure involves injecting a contrast medium into a vein in the foot during an x-ray of the veins. A tourniquet applied to the limb promotes the filling of the venous system. Positive findings are based on the presence of filling defects in the veins of the affected limb.
There are several limitations of this test however, and consequently the test is rarely used in the clinical setting unless non-invasive studies are ambiguous in the presence of clinical signs (Church, 2000; Fahey, 1989). There is a risk of allergy to the contrast medium, pain associated with the procedure, phlebitis at the injection site and ironically, a small number of patients may develop a DVT (Cranley et al., 1976; Fahey, 1989; Line, 2001; Ramzi & Leeper, 2004). The cost of the procedure may also inhibit its use in some settings (Church, 2000; Cranley et al., 1976). A non-invasive procedure, the Doppler ultrasound is more commonly used to detect DVT and is the procedure currently used at the PAH for detection of DVT in SCI patients.

2.4.1.2 Doppler Ultrasonography
Doppler Ultrasonography studies are a non-invasive procedure designed to measure venous flow through the use of sound waves. A probe that transmits and receives sound waves which bounce off solid objects such as blood cells is passed over the skin above the veins. The Doppler is able to measure changes in the velocity of the blood flow, lack of flow and the effect of compression on flow (Cranley et al., 1976). Sensitivity has been shown to be as high as 95% in proximal DVTs and an equally high specificity in those patients who are symptomatic (Teasall et al., 2006). Sensitivity of distal DVTs is significantly lower, at 73%, but as distal DVTs rarely embolise unless they propagate, this is considered an acceptable rate of sensitivity (Teasall et al., 2006). However, due to the decreased Doppler ultrasound sensitivity in distal DVT diagnosis, many authors advocate the use of serial Doppler ultrasounds in symptomatic patients of high risk groups where a first test is negative to reduce the risk of failure to detect a DVT (Line, 2001; Teasall et al., 2006).

Limitations to the use of Doppler ultrasound include: the expertise of the operator in its use, its inability to distinguish between old and new thrombi and its reduced sensitivity in detecting DVT in the pelvis or smaller vessels of the calf. In addition, a DVT is also difficult
to detect in the presence of oedema or obesity (Fahey, 1989; Ramzi & Leeper, 2004).

Another alternative non-invasive objective measure is the impedance plethysmography.

2.4.1.3 Impedance Plethysmography

Impedance plethysmography measures the electrical resistance caused by changes in venous volume. A blood pressure cuff is inflated around the patient’s thigh occluding venous flow (Line, 2001). Electrodes placed on the skin measure venous flow as the cuff is deflated. If a DVT is present, venous volume will be significantly less than that in the non-infected limb (Line, 2001; Wallis & Autar, 2001).

Plethysmography is not a specific test for DVT and has varied results in terms of sensitivity and specificity, ranging from 65% to 95% (Line, 2001). This is due to numerous other factors which may give false positive results including: pain, incorrect positioning of the patient, vasoconstrictors, congestive heart failure, pregnancy, external venous pressure from a tumour or abscess or increased central venous pressure (Line, 2001; Rosen & Tracy, 2001). False negative results may be due to an old embolus, superficial phlebitis or the occlusion of the popliteal, tibial and peroneal veins (Bland & Altman, 1986).

Most importantly, however, it is suggested that the reliability of impedance plethysmography in patients with SCI is questionable (Yao, 1992). Two important requirements of the test include venous capacitance and venous outflow. These two aspects are affected by muscle weakness making the test problematic in the SCI group, producing potentially false positive results (Yao, 1992). Finally the d-dimer assay, a blood test, is the newest objective measure for detecting a DVT.

2.4.1.4 D-Dimer Assay

D-Dimer is the name given to a product of fibrin degradation that is released into circulation when a DVT forms (Haeger, 1969). There are a variety of haematology tests that screen for the presence of d-dimer with varying degrees of sensitivity (60-96%) and
specificity (13-59%) (Ramzi & Leeper, 2004). Though a relatively simple procedure for the patient (one type of d-dimer test requires just a prick of the finger), studies have revealed poor specificity rates in patients with pneumonia, and heart failure and an equally high level of false negatives in other patient groups including those using chemical prophylaxis for DVT prevention. Consequently, there is a lack of confidence in its use by many physicians (Bland & Altman, 1990) and may be of little value in the SCI population where chemical prophylaxis for DVT prevention is now standard practice.

It should also be noted that d-dimer is not specific for DVTs as there may be other reasons for its presence in the blood, therefore a positive assay test is not definitive (Ramzi & Leeper, 2004). Recent studies have focused on the use of d-dimer tests in conjunction with clinical prediction rules. Clinical prediction rules combine a number of signs and symptoms with risk factors in order to predict the likelihood that a patient may have a DVT. This combination of d-dimer and clinical prediction rules has been shown to have a high negative predictive power meaning that the likelihood of the presence of a DVT can be ruled out (Qaseem et al., 2007; Whitney et al., 1995).

Whilst many of these objective diagnostic tools have varying degrees of reliability it is argued that a diagnosis of DVT in the clinical setting could not be made solely through the use of these objective measures. The serial testing of every patient using one or more of these devices is neither practical nor cost efficient (Morrison, 2006). Furthermore, venography which has a high sensitivity and specificity is invasive and not without considerable risk to the patient (Line, 2001; Whitney et al., 1995).

A more common scenario in the clinical setting is that the health care team notes the presence of clinical signs and symptoms displayed by the patient, and a decision is made as to the likelihood of a DVT. The definitive diagnosis of DVT would then be made using one of the objective measures outlined above with consideration of the patient’s clinical situation.
2.4.2 Signs and Symptoms
Whilst the sole reliance on clinical signs and symptoms for DVT diagnosis is unreliable (Haeger, 1969), more recent studies in ambulant outpatient populations have shown that particular combinations of these signs can be reliably used to identify a high likelihood of DVT with a subsequent confirmation using objective measures (Diamond & Macciocchi, 1997; Goodacre, Sutton, & Sampson, 2005). The combination of signs and symptoms studied varies significantly across studies making comparisons difficult, though the most common combination is the presence of oedema, pain, asymmetry of circumference measurements of the lower limbs and fever. Numerous authors have grouped these signs and symptoms together with the patient’s medical history to formulate prediction rules, the most common of which is called the Wells’ rule (Caprini, 2005; Connelly, 2008; Kahn, 1998; Wells et al., 1995).

2.4.2.1 Clinical Prediction Rules
Clinical prediction rules for DVTs are designed to help anticipate the likelihood of the presence of a DVT. They combine a detailed medical history with a clinical assessment allocating a weighted score to the presence of specific items. The total score provides a degree of likelihood of the presence of a DVT. This is then able to be assessed with the use of a more objective test such as Doppler ultrasound. That is, a low probability of the presence of a DVT combined with a negative ultrasound would give a high likelihood that no DVT is present.

Of these scores the most recognised is the Wells’ score, which uses nine items (see Table 1). A score is developed for each patient according to the table, for the presence of each one of the nine criteria; the resultant total score helps to determine the probability of a DVT. A score of $\geq 3$ indicates a high probability (around 75% chance) of a DVT, a score of $1$
or 2 indicates a moderate probability (around 17% chance) whilst a score of ≤0 indicates that the patient had a low probability of having a DVT (around 3% chance) (Haeger, 1969).

Table 2. Wells' clinical prediction rule for deep vein thrombosis (DVT)

<table>
<thead>
<tr>
<th>Clinical Parameter</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active Cancer (Treatment ongoing or within previous 6 months or palliative)</td>
<td>1</td>
</tr>
<tr>
<td>Paralysis, paresis or recent plaster immobilisation of the lower extremities</td>
<td>1</td>
</tr>
<tr>
<td>Recently bedridden for &gt;3 days because of major surgery (within 4 weeks)</td>
<td>1</td>
</tr>
<tr>
<td>Localized tenderness along the distribution of the deep venous system</td>
<td>1</td>
</tr>
<tr>
<td>Entire leg swelling</td>
<td>1</td>
</tr>
<tr>
<td>Calf swelling by &gt;3cm when compared with the asymptomatic leg (measured 10cm below the tibial tuberosity)</td>
<td>1</td>
</tr>
<tr>
<td>Pitting oedema (greater in the symptomatic leg)</td>
<td>1</td>
</tr>
<tr>
<td>Collateral superficial veins (nonvaricose)</td>
<td>1</td>
</tr>
<tr>
<td>Alternative diagnosis as likely or greater than that of deep vein thrombosis</td>
<td>-2</td>
</tr>
</tbody>
</table>

Scoring Method: score ≥3 High probability, score is 1 or 2 Moderate probability, score is ≤0 Low probability. (Haeger, 1969).

The Wells’ rule was validated in an outpatient population where a DVT was clinically suspected with no other obvious cause of the symptoms. Patients were excluded from the study if they had a previously diagnosed VTE or were taking anticoagulants (Wells et al., 1995). Subsequent testing of the Wells’ rule in other patient groups has proven less successful with Kraaijenhagen et al (2002) and Oudega et al (2005) reporting a sensitivity of 82.8% and 78.9% and specificity of 62.8% and 44.3% respectively for patients in the lowest risk category (Kraaijenhagen et al., 2002; Oudega, Hoes, & Moons, 2005). This compares with 89.5% sensitivity and 64.1% specificity reported by Wells et al (1995). The translatability of the Wells’ rule to other patient
groups has been questioned. When used in combination with the results of a D-dimer blood tests the Wells’ rule showed improved prediction ability. In those patients with a normal D-Dimer and a Wells score of ≤0, 1.8% (Kraaijenhagen et al., 2002) and 2.3% (Oudega et al., 2005) of DVTs were undiagnosed though the Wells study remained the best with 0.9% of DVTs undiagnosed.

The most significant reason given for Wells’ rule’s poor performance in other trials is related to the last item on the checklist; “Alternative diagnosis as likely as or greater than that of deep vein thrombosis”. This is the most subjective component of the prediction rule with the most potential for variability amongst clinicians and it attracts a significant score. The Wells’ score is also less sensitive in the spinal cord injured population as it involves scoring for the presence of pain which may be absent in this group.

2.4.2.2 Pain

Early studies of clinical signs and symptoms of DVT found pain and calf tenderness to be highly sensitive, ranging from 62% to 90% (Kahn, 1998). The level of sensitivity is often dependent on the method used by the examiner. Lisker’s sign is pain produced by percussion of the medial tibia. Approximately 65% of patients with DVT will display pain on percussion of this area which although over half, it leaves a possible 35% of DVTs unrecognised (Kennedy et al., 2001). Pain resulting from the compression of the calf against the tibia is known as Bancroft’s sign. It is thought that the Bancroft sign may produce more pain than other methods and thus it is rarely used (Kennedy et al., 2001). Perhaps the most well known and widely tested method is Homan’s sign.

Homan’s sign, the presence of pain on dorsiflexion of the foot, is the most controversial of all of the detection methods. Useful only in detecting calf thrombosis,
the sign occurs in about 10% of patients and may not be present if the thrombosis is small (Church, 2000). Alternatively, as in the case of an iliofemoral thrombus the pain on dorsiflexion may be extreme. One study found false negatives and false positives occurred as often as 50% of the time (Herzog, 1992). The majority of authors continue to refer to it along with other indicators, however, it lacks sensitivity and specificity (Church, 2000; Fahey, 1989; Herzog, 1992; Kahn, 1998). It is suggested that a positive sign may be elicited in patients on prolonged bed rest, in the case of superficial phlebitis, rupture of the plantar muscle or a gastrocnemius tear.

All of the methods to elicit a pain response lack the specificity required to be truly useful in the diagnosis of DVT. Importantly, however, the use of pain as a clinical sign is limited in the SCI population, particularly in those with motor complete lesions as these patients are not able to feel pain below the level of their injury.

2.4.2.3 Fever

Fever or pyrexia is one of the most common indicators of disease processes. Its duration and characteristics are useful in identifying possible causes. Fever may be constant, meaning it does not vary by more than two degrees Celsius and does not return to normal in a 24 hour period (Castle & Watkins, 1979). Alternatively, it may be remittent, varying by more than two degrees during the 24 hour period without returning to normal. A fever may also be graded as low (37.2-38 degrees Celsius), moderate (38.1-40 degrees Celsius) or high (>40 degrees Celsius) (Castle & Watkins, 1979).

Fever is regarded as a significant clinical sign of DVT (Faris, Rosengarten, & Dudley, 1972; Polit & Hungler, 1999; Wallis & Autar, 2001; Weingarden, Weingarden, & Belen, 1988). It is thought, however, that often DVT is overlooked due to the coexistence of other causes of fever such as urinary sepsis or surgical infection (Polit &
Hungler, 1999). In the SCI patient this is often complicated by thermoregulatory insufficiency known as poikilothermia. Poikilothermia, common in those with high level paraplegia and tetraplegia, is due to injury to the autonomic nervous system (Weingarden et al., 1988). Individuals become greatly affected by changes in the ambient temperature, for example; in extreme heat individuals are likely to have a high temperature.

Despite the long held belief that fever of unexplained causes may be linked with the development of a DVT, few studies have examined the association and those that have, showed poor association between fever and DVT (Diamond & Macciocchi, 1997; Faris et al., 1972; Kazmers, Groehn, & Meeker, 2000). An early study in post-operative patients showed a relationship between low grade temperature for at least two consecutive days and DVT though sensitivity was poor, detecting just half of those with a DVT (Faris et al., 1972). A more recent study of patients undergoing duplex scanning for suspected DVT also showed poor sensitivity (59%) and specificity (56%) (Kazmers et al., 2000), whilst another showed low positive predictive power (0.08) (Diamond & Macciocchi, 1997).

Of note however, is that some authors (Beraldo et al., 1993; Green, 1999) do believe a low grade temperature is significant in SCI patients. The Clinical Practice Guidelines for the prevention of thromboembolism published by the Consortium of Spinal Cord Medicine list low grade fever of unknown origin as a clinical sign of DVT (Green et al., 2002). However, just a few studies have examined the incidence of fever in SCI patients (Beraldo et al., 1993; Weingarden et al., 1988). Authors in one study identified four patients in whom fever was the sole presenting sign. The study identified that spikes in temperature occurred almost always at night and were usually a low grade fever (Weingarden et al., 1988). There are several limitations to the study. The
The incidence of DVT in the population studied was small with 10 patients of the 148 investigated being diagnosed with DVT, and of these just six met the inclusion criteria and had adequate documentation (Weingarden et al., 1988).

2.4.2.4 Oedema and Calf Circumference

Oedema or swelling of the limb may occur in the presence of a DVT when the thrombus has sufficiently occluded the vein to affect venous return. It may also occur when the thrombus has damaged the capillaries resulting in the leakage of fluid into the surrounding tissues distal to the thrombus site (Wallis & Autar, 2001). Oedema, as indicated by an increase in calf circumference, is recognised as a clinically significant sign of DVT (Green, 1999; Landefield, Mcguire, & Cohen, 1990; Line, 2001), though it is argued that oedema is also included in the differential diagnosis of many other disease processes (Kennedy et al., 2001). The distinction for DVT diagnosis lies in whether the oedema is unilateral or bilateral (Line, 2001).

Bilaterally symmetrical oedema is more likely to arise in conditions such as congestive heart failure, pregnancy, liver disease and nephritic syndrome (Kennedy et al., 2001; Line, 2001). Whilst bilateral oedema can occur in the presence of DVT, it is usually in cases where there is involvement in the inferior vena cava or in rare cases of bilateral DVTs (Line, 2001).

Calf circumference as an indicator of the presence of DVT has been examined extensively in many different patient populations. This body of literature is limited in value by the variation in the location on the leg of the measurement, the variation in patient population examined and the measurement difference that is considered to be significant. Of the seven studies reviewed where the size of difference was stated, the measurements chosen for the allowable difference between legs varied from one centimetre to three centimetres or greater (Brown & Kros, 2003; Constans et al., 2003;
Diamond & Macciocchi, 1997; Johanning et al., 2002; Stein et al., 1995; Swarczinski & Dijkers, 1991; Wells et al., 1995).

There is some agreement that a greater than 1cm difference between legs is an abnormal finding (Brown & Kros, 2003; Line, 2001; Stein et al., 1995). However specificity for DVT diagnosis at this level is low (Stein et al., 1995). Studies reporting differences of 2cm or greater have had mixed results (Diamond & Macciocchi, 1997; Johanning et al., 2002). Johanning et al (2002) reported 63.6% of patients with DVT had a circumference measurement >2cm compared with 19.4% without DVT (p>0.003) (Johanning et al., 2002). This contradicts findings from an earlier study by Diamond (1997) where an asymmetry between limbs of greater than 2.5 centimetres had a sensitivity of 20% and was lower than actual visual observation of swelling (25%) (Diamond & Macciocchi, 1997). A difference of 3 centimetres or greater in calf circumference appears to be highly sensitive for DVT (Wells et al., 1995). This finding was confirmed by Constans et al (2003) in a two prospective samples finding a calf circumference ≥3 cm in 40% of patients with DVT compared to 20% without and 79% with DVT compared to 56% without (Constans et al., 2003).

There is a strong recommendation by the consortium for spinal cord medicine based on expert opinion, that calf circumference measurements be recorded for the surveillance of DVT in SCI patients (Green et al., 2002). However few prospective studies relating to the incidence of DVT in the SCI population have examined lower limb circumference measurements and the majority of those were published between 1963 and 1985 (Weingarden, 1992). Despite this, there remains a strong suggestion by some authors that an increase in calf circumference is one of the few reliable clinical signs of VTE in this group of patients (Beraldo et al., 1993; Green, 1999; Green et al., 2002).
Those studies where lower limb circumferences are mentioned are lacking in detail about the measurements. Most state only that a calf circumference measurement was conducted (Beraldo et al., 1993), some are more specific about the location of the measurement, most commonly 10cm below the tibial tuberosity (Whitney et al., 1995). But significantly there is little consensus on what is considered a clinically significant measurement difference.

The most significant study within the SCI population regarding circumference measurements is from over two decades ago and was a retrospective chart review (Swarczinski & Dijkers, 1991). This study compared three separate measurement differences; .5 inches, 1cm and 1.2cm for women/1.5cm for men. They found that regardless of the size of the measurement differences used, there was no association between the calf measurement and the results of the objective test, in this case a fibrinogen uptake test. The authors assumed that fluctuations in the day to day measurements were due to measurement error and hypothesise this error was due to stretching of the tape measure, poor inter-rater reliability, lack of consistency in tensioning of the tape, lack of consistency in positioning the tape at the correct site and even differences in looping the tape over itself to read the measurement.

Significantly a more recent study examining the incidence of DVT in SCI patients contradicts these findings (Green, 1999). The calf measurements were not the focus of the study, but rather collected daily as part of the study protocol along with body temperature. An increase of greater than two centimetres was considered significant for DVT. Unfortunately the authors do not provide further detail about the measurements such as; where on the calf they were measured, what type of tape was used, whether the increase in circumference measurement was from baseline, day to day or between legs or who performed the measurements. This lack of detail makes true
comparison of the studies difficult. However the finding that an increase in circumference measurements had a significant correlation (p<0.001) with the presence of DVT diagnosed by Doppler ultrasound cannot be ignored.

There appears to be just as much variation between studies in the location on the leg where the measurements are taken. Many authors do not give any specific reference point, although there is importance placed on ensuring that measurements should be taken at the same level on both legs (Line, 2001). One early study suggested multiple measurement points using the minimum circumference measurement at the ankle and thigh and the maximum measurement at the calf (Brown & Kros, 2003). More recent studies are specific using either the patella or the tibial tuberosity as points of reference (Johanning et al., 2002; Stein et al., 1995; Wells et al., 1995). For example, the measurements should be taken 10cm below the tibial tuberosity. It should be noted that authors do not provide justification as to the point chosen.

In a survey of 80 physicians regarding what measurement reference points they used to perform circumference measurements, the most common response was the patella, with other responses including the tibial tuberosity and medial malleolus (Tunc et al., 2007). Higher intra and inter-observer agreement was found for the tibial tuberosity (88%, 81%) than either (65%, 57%) the patella or medial malleolus (73%, 65%) (Tunc et al., 2007). Patella mobility was suggested as a reason for poor agreement, whilst the shorter distance to mid calf from tibial tuberosity may account for the better agreement over medial malleolus.

Whilst location and cut-off point are important factors, so too is the tool used to perform the measurement, though this information is notably absent from many studies. The most commonly used tape measure is the standard sewing tape measure though there are questions about its reliability (Swarczinski & Dijkers, 1991). The tape may
become stretched over time; there may be issues with consistent tensioning, lack of consistent positioning on the leg and lack of consistent positioning of the tape over itself for reading the measurement.

It is suggested that the use of a spring tape would avoid these problems (Swarzcinski & Dijkers, 1991). Spring tapes have been shown to have a “high reliability, low process related variance and large between subject variance” (Labs, Tschoepf, Gamba, Aschwanden, & Jaeger, 2000). The use of spring tapes appears to be common place in studies of peripheral oedema but none of the thromboembolism studies reviewed have utilised this tool (Labs et al., 2000).

More recently published studies examining the reliability of tape measures suggest that there is little or no difference in reliability between the standard tape measure and the spring tape measure (Fairclough, Mintowt-Czyx, Mackie, & Nokes, 1994; Te Slaa et al., 2011). In a study of measurement of lower limb amputee anthropometrics seven commonly used measuring devices were tested for accuracy and consistency of measurements including, the standard tape measure and the spring tape measure. When measuring a single known length both the spring tape and the standard tape were deemed to be highly accurate with low error rates of -0.1130 cm and -0.0130 cm respectively (Fairclough et al., 1994).

In terms of consistency of measurement, the participants produced the most consistent measurements using the tape measures in contrast to other measuring devices, however the paper lacks specific detail about the exact standard deviations and ranges of measurements (Fairclough et al., 1994). Although the spring tape did produce very slightly more consistent measurements, this improved consistency was however not clinically significant. The author concludes that tension may be less important than previously reported (Fairclough et al., 1994). The use of foam models instead of actual
limbs in this study makes comparison of findings with other studies, and generalisability to the clinical setting, difficult, particularly in relation to the tensioning of the tape which may be significant in individuals with fleshy limbs.

A study that was published after the current work commenced investigated the reliability and reproducibility of leg circumference measurements using a standard tape measure. Measurements were conducted by 4 observers on 11 volunteers assessing short term, medium term and long term reliability and reproducibility. An Intraclass correlation coefficient for repeatability of 0.90 and a reproducibility index of +/- 4.4% was found using the standard tape measure when repeated measurements were taken for a short term (1 week). Reproducibility was increased to 3.3% if measurements were performed by one observer and reliability decreased when the time interval between measurements increased most likely due to biological fluctuations in the volunteers (Te Slaa et al., 2011). These results show that the standard tape measure is capable of producing highly repeatable and reproducible measurements. The authors of this study concede however that the improvement in reproducibility when only one observer is used may be due to inconsistent tension applied by different observers and the use of a spring tape may improve reproducibility with multiple observers.

The use of signs and symptoms in the diagnosis of VTE remains a contentiously debated issue in the literature. There are many reports of the sensitivity and specificity of these measures, with equally highly variable results. For the SCI population there appears to be more consistent evidence that an increase in lower limb circumference and the presence of fever are clinical symptoms suggestive of VTE.

There is however a lack of consensus in the literature about several aspects of the procedure making comparison of studies very difficult (see Table 3). These measurement aspects include the reliability of the tape measure being used, the
difference in measurements considered clinically significant, inter rater reliability of the measurements and the position on the leg where the measurement is taken. Some of these issues have been considered individually in studies of anthropometric measurement and limb oedema but none have considered all of these aspects together and how they affect the reliability of measurements for diagnosis of DVT. Making comparison of the literature more difficult is the variety of analytical methods used to determine reliability and reproducibility of circumference measurements. Differences in the design of the studies have meant that different aspects of reliability have been the focus of studies. Few studies to date have attempted to consider the various components of measurement error to give a more complete picture of reliability of circumference measurements.
### Table 3. Summary table of studies examining calf circumference measurements

<table>
<thead>
<tr>
<th>Author (Yr)</th>
<th>Paper Type</th>
<th>Subjects</th>
<th>Aim</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wallis and Autar (2001)</td>
<td>Narrative</td>
<td>n/a</td>
<td>Oedema may be caused by thrombosis occluding deep veins or by capillary damage causing leakage into surrounding tissues</td>
<td></td>
</tr>
<tr>
<td>Agarwal and Mathar (2009)</td>
<td>Research study</td>
<td>297</td>
<td>Comparison of incidence of DVT in SCI pts with and without prophylaxis</td>
<td>Increase in Calf girth is highly correlated with DVT (p&lt;0.001)</td>
</tr>
<tr>
<td>Landefield et al (1990)</td>
<td>Research Study</td>
<td>355</td>
<td>Identification of clinical findings useful in estimating the probability of acute proximal DVT</td>
<td>Swelling above the knee independently correlated with DVT</td>
</tr>
<tr>
<td>Line (2001)</td>
<td>Narrative</td>
<td>n/a</td>
<td>Unilateral oedema is one of the best indicators of DVT</td>
<td></td>
</tr>
<tr>
<td>Kennedy et al (2001)</td>
<td>Narrative</td>
<td>n/a</td>
<td>The differential diagnosis of patient with swollen leg is broad</td>
<td></td>
</tr>
</tbody>
</table>

#### Size of Difference considered significant

<table>
<thead>
<tr>
<th>Author (Yr)</th>
<th>Paper Type</th>
<th>Subjects</th>
<th>Aim</th>
<th>Size of Difference Used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cranley et al (1976)</td>
<td>Research</td>
<td>124</td>
<td>Examination of the reliability of Clinical Signs and Symptoms</td>
<td>1 cm</td>
</tr>
<tr>
<td>Diamond and Macciocchi (1997)</td>
<td>Research</td>
<td>61</td>
<td>Examination of the predictive power of clinical symptoms of DVT</td>
<td>&gt;2.5 cm</td>
</tr>
<tr>
<td>Johanning et al (2002)</td>
<td>Research</td>
<td>156</td>
<td>Assessment of the effectiveness of calf circumference combined with d-dimer for diagnosis of DVT</td>
<td>2 cm</td>
</tr>
<tr>
<td>Stein et al (1995)</td>
<td>Research</td>
<td>678</td>
<td>Examination of the asymmetry of calves for diagnosis of PE</td>
<td>&gt;1 cm</td>
</tr>
<tr>
<td>Swarczinski and Djikers (1991)</td>
<td>Research</td>
<td>30</td>
<td>Examination of Calf and Thigh measurements used for detecting DVT</td>
<td>1 cm</td>
</tr>
<tr>
<td>Wells et al (1995)</td>
<td>Research</td>
<td>887</td>
<td>Accuracy of physical signs of DVT</td>
<td>&gt;or = 3 cm</td>
</tr>
</tbody>
</table>

#### Location where measurements are performed

<table>
<thead>
<tr>
<th>Author (Yr)</th>
<th>Paper Type</th>
<th>Subjects</th>
<th>Aim</th>
<th>Where to measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Akman et al (2004)</td>
<td>Research</td>
<td>68</td>
<td>Assessment of the utility of D-dimer testing for diagnosis of DVT</td>
<td>15cm proximal and 10cm distal to the tibial tubercle</td>
</tr>
<tr>
<td>Line (2001)</td>
<td>Narrative</td>
<td>n/a</td>
<td>Examination of the reliability of Clinical Signs and Symptoms</td>
<td>At the same level on both legs</td>
</tr>
<tr>
<td>Cranley (1976)</td>
<td>Research</td>
<td>124</td>
<td>Examination of the reliability of Clinical Signs and Symptoms</td>
<td>The minimum circumference at the ankle and maximum at the thigh</td>
</tr>
<tr>
<td>Johanning et al (2002)</td>
<td>Research</td>
<td>156</td>
<td>Assessment of the effectiveness of calf circumference combined with d-dimer for diagnosis of DVT</td>
<td>10cm below the patella</td>
</tr>
<tr>
<td>Stein et al (1995)</td>
<td>Research</td>
<td>678</td>
<td>Examination of the asymmetry of calves for diagnosis of PE</td>
<td>10cm below the tibial tuberosity</td>
</tr>
<tr>
<td>Wells et al (1995)</td>
<td>Research</td>
<td>887</td>
<td>Examination of the accuracy of physical signs of DVT</td>
<td>10cm below the tibial tuberosity</td>
</tr>
<tr>
<td>Tunc et al (2007)</td>
<td>Research</td>
<td>68</td>
<td>Examination of inter and intraobserver agreement of leg measurements</td>
<td>10 and 20cm under tibial tuberosity</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10 and 20cm above medial malleolus</td>
</tr>
</tbody>
</table>

#### Device used for measurements

<table>
<thead>
<tr>
<th>Author (Yr)</th>
<th>Paper Type</th>
<th>Subjects</th>
<th>Aim</th>
<th>Device used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Swarczinski and Djikers (1991)</td>
<td>Research</td>
<td>30</td>
<td>Examination of calf and thigh measurements used for detecting DVT</td>
<td>Standard sewing tape</td>
</tr>
<tr>
<td>Geil (2005)</td>
<td>Research</td>
<td>n/a</td>
<td>Examination of the accuracy of anthropometric measurements</td>
<td>Seven measurement devices including standard sewing tape and spring tape</td>
</tr>
<tr>
<td>Te Slaa et al (2011)</td>
<td>Research</td>
<td>11</td>
<td>Examination of the reliability and reproducibility of leg measurements</td>
<td>Standard sewing tape</td>
</tr>
</tbody>
</table>
2.5 Reliability of Measurements

Reliability refers to the consistency or repeatability of a measurement of information obtained in a study. Data are said to be reliable if the test can be repeated independently on another set of data and similar results are obtained (Trochim, 2006). All measurements are considered to be made up of the true score and a certain amount of error (Chatburn, 1996). Whilst it is impossible to know the true score, it can be estimated by considering the variance in the differences between repeated measurements. The smaller the variation between measurements the more reliable the measurements are considered to be (Polit & Beck, 2012).

The ‘total’ error consists of predictable (systematic) error and unpredictable (random) error (Burns & Grove, 1997) (See Figure 1) Systematic error is also referred to as measurement bias and is usually associated with equipment calibration or some other constant cause of error. With a systematic error there is a tendency for all of the difference in scores between repeated tests to be either positive or negative (Altman & Bland, 1995; Burns & Grove, 1997). Random error is usually the larger component of error and is often associated with variations in the biology of an individual (for example; fluctuations in oxygen saturations) or variations in the research protocol such as the differences between individual examiners measurement techniques (Burns & Grove, 1997; Chatburn, 1996).

![Figure 1. Components of Observed Measure Score](image-url)
2.5.1 Types of Reliability

The method chosen to assess the reliability of a particular instrument is dependent upon the aspect of reliability being assessed and the reliability estimate being used. Three aspects of reliability were identified in the literature; stability, internal consistency and equivalence (Polit & Beck, 2012). The stability of an instrument refers to its ability to remain consistent under repeated measurements. This involves comparing the results obtained from measuring with an instrument in a test-retest format. Test-retest reliability is used to assess the consistency of a score, measurement or piece of equipment from one time to another (Trochim, 2006). The test-retest format is useful for testing both systematic and random error.

Equivalence involves testing the constancy between different devices and operators. It may be multiple observations of one device or multiple observations of similar devices (Polit & Hungler, 1999). Inter-rater reliability or inter-observer reliability, measures the consistency of scores between two or more different raters or observers about the same phenomenon (Trochim, 2006). It is a way of testing the equivalence between observers in other words the amount of error that is considered unpredictable or random. Internal consistency is an aspect of reliability associated with pen and paper instruments. Commonly used to measure consistency in survey research, (Polit & Hungler, 1999) internal consistency is not a focus of this study.

The statistical tests used to analyse inter-rater reliability and test-retest reliability vary greatly in the literature and whilst all may seem relevant to reliability testing, not all tests will be appropriate for all situations (Burns & Grove, 1997; Chatburn, 1996). The choice of statistical tests should be dependent on the analytical goals of the research (Burns & Grove, 1997). The decision must be made whether to test for systematic error, random error, or both. The most commonly used tests for statistical analysis in reliability studies are; Paired t-tests, Pearson’s correlation coefficients, Intra class correlations, Standard Error of Measurement
and the more recent addition of Bland Altman’s Limits of Agreement (Altman & Bland, 1995).

2.5.1.1 Paired t-tests
The paired t-test is used to compare the means of test and re-test samples in order to assess systematic bias. Although it is a useful test, the lack of measurement of random variation between the tests means that the paired t-test cannot be used alone to measure reliability (Burns & Grove, 1997). The t-test is designed to measure the hypothesis that there is no difference between the two means. In other words, that there is complete agreement between the two samples (test and retest) on what they are measuring. However, because of the design of the test, a systematic bias is less likely to be detected if there is a large amount of random error between the two tests (Burns & Grove, 1997). Malagon and Lamb (1996) caution that the test may lead to the wrong conclusion regardless of whether agreement is poor or good (Malagon, 1996). Consequently, the paired t-test is only appropriate when comparing two sets of related data such as in pre and post testing.

2.5.1.2 Pearson’s Correlation Coefficients
Correlation coefficients are designed to report the degree of association between two sets of data (Altman & Bland, 1995). The correlation coefficient has dominated reliability studies for many years, though it is now considered by many to be an inappropriate measure of reliability when used on its own (Burns & Grove, 1997). Unlike the t-test, Pearson’s correlation coefficient is insensitive to systematic error giving only an indication of random error (Chatburn, 1996).

In studies where the focus is the comparison of methods or devices, it is argued that two devices that are designed to measure the same object will show linear association even if one device consistently measures the object twice as large as the other. The correlation coefficient cannot tell us where the two devices do not agree and therefore it is not
recommended as a method of analysis for studies assessing reliability of measurement techniques.

### 2.5.1.3 Intraclass Correlation Coefficient (ICC)

The Intra-class Correlation Coefficient (ICC) uses analysis of variance to break down the total variance into between subject and within subject variance (Altman & Bland, 1995). The advantage of the ICC in inter-rater reliability studies is its ability to differentiate between systematic and random error (Rankin & Stokes, 1998). The ICC is useful in assessing the repeatability of measurements in different units (Altman & Bland, 1995). However, like Pearson’s correlation coefficient calculation, the ICC has some disadvantages.

There are six different ICC equations and calculation of the ICC requires careful selection of the correct equation dependent on the study design and intent. The data analyst must choose whether the study has scores by a single rater or multiple raters and if the intent of the study is to examine absolute reliability or agreement. Selecting the wrong equation may lead to an over or under estimation of reliability. The ICC is reported simply as a range between zero and one with no indication of acceptable level of reliability or actual measurement values, only that the closer the ICC is to one the greater the reliability (Altman & Bland, 1995; Rankin & Stokes, 1998).

### 2.5.1.4 Standard Error of Measurement (SEM)

The standard error of measurement (SEM) is considered a measure of ‘absolute’ reliability, the smaller the SEM the greater the reliability (Altman & Bland, 1995). The SEM estimates how repeated measures of the same person using the same device are distributed around the true score. The use of SEM is considered useful in studies containing interval data though cautioned in others due to several assumptions. The SEM assumes that there is a population of measurements for each individual and that the population is normally
distributed. It is also assumed that heteroscedascity is not present. In other words it assumes that the error is the same for high scores as it is for low scores (Burns & Grove, 1997).

2.5.1.5 Limits of Agreement
The Limits of Agreement method as proposed by Bland and Altman (1986) is now considered the most appropriate method of analysis in method or device comparison studies (Alanen, 2012; Bland & Altman, 1999; Bland & Altman, 2010; Burns & Grove, 1997; Chatburn, 1996; Germing et al.). The purpose of a method comparison study is to show “a direct comparison of results obtained by alternative methods” and whether the results obtained show that the two methods are comparable enough that one may replace the other in clinical practice (Altman & Bland, 1983).

The differences between measurements made by each of the two methods, is plotted against the mean of the measurements by the two methods for each individual. The limits of agreement with their confidence intervals are then calculated as the mean of the measurements obtained by the two devices plus or minus 1.96 standard deviations. It is expected that the difference between measurements obtained by the two methods in 95% of all future measurements will lie within these limits (Liu et al.). Interpretation of the results is largely by visual inspection of the graphical plots, and ultimately a matter of clinical judgement. The decision about what limit is considered clinically significant should be made a priori and based on the clinical application of the device (Germing et al.).

2.6 Rationale for Study
Venous thromboembolism is a significant illness and compromises patient outcomes both in the short term and long term. There are costs associated with diagnostic procedures to confirm the presence and location of the thrombus and then treatment of the thrombus including administration of anti-coagulants and increased nursing resources to monitor the patient’s condition. It also delays rehabilitation therefore increasing length of stay. Should the
patient then develop post-thrombotic syndrome there may be costs associated with the treatment and management of vascular impairment including venous leg ulcers.

The performance of calf and thigh measurements is a simple, non-invasive procedure which may assist in the early detection of thromboembolism. Lower limb circumference measurements are recommended by the consortium for spinal cord injury (Green et al., 2002), considered to be the leading body of experts in management of patients with spinal cord injury, though the literature does not support or refute the value of lower limb measurements (Swarczinski & Dijkers, 1991). The lack of detail provided in the methods section of many studies hampers true evaluation and comparison of results. Most are not specific about the type of tape measure used, or the location of the measurement which limits study replicability. Fewer still provide information about the investigator performing the measurements and the setting for the study.

There is consensus on one fact within the SCI literature; an increase in lower limb circumference measurements is clinically significant (Green et al., 2002). Despite this consensus of opinion just one study has specifically investigated the use of circumference measurements as an aid to diagnosis in this group of patients (Swarczinski & Dijkers, 1991). The study however raised many concerns over the reliability of the practice. While studies of lower limb circumference measurements have been designed to answer some of the questions, many are focused on one aspect of reliability making it difficult to draw any real conclusions about the reliability of the practice.

From analysis of the literature a framework for the reliability of circumference measurements was developed (see Figure 2). The various aspects of the practice of circumference measurements identified in the literature as being relevant to measurement reliability are outlined in this framework and provide the structure for the current study. Three aspects of the practice of circumference measurements were considered to be of
particular importance, the people performing the measurements, the equipment used, and the measurement location.

![Figure 2](attachment:figure2.png)

**Figure 2.** Framework for the Reliability of Circumference Measurements

In considering the people undertaking the measurements two variables were identified that may affect reliability and reproducibility of circumference measurements. The reproducibility of measurements in previous studies was found to differ depending on the number of examiners used to perform the measurements. Reliability in the measurements may also be affected by the use of non-uniform technique when multiple examiners are used.

The equipment used to conduct circumference measurements has been a major focus of many previous studies with conflicting results about reliability and reproducibility. To date though, few studies have compared the standard tape measure and the spring tape measure for the purpose of circumference measurements. The location the measurements are performed is determined by the site used to undertake the measurement (thigh, calf or ankle) and the landmarks used to identify the location (patella, tibial tuberosity and medial malleolus). There
is a lack of consensus in the literature about which of these sites provides more consistency in measurements.

Further exploration of each of these aspects of circumference measurements is needed to improve the reliability and reproducibility of this practice. A measurement that is unable to be reproduced is unreliable and of little use in clinical practice. Given the high risk of thrombus development in SCI patients and the potentially lethal effects of thromboembolism, a reliable and effective measurement technique is required to assist in the early detection of DVT.
3.0 Methods

3.1 Introduction
The literature review presented in the previous chapter has highlighted several concerns over the reliability of the practice of calf and thigh circumference measurements. Specifically the stability and equivalence of the equipment, the procedure and the measurements was assessed in this study. Presented in chapter three are the research questions formulated from the review of the literature, the design of the study and the procedures used to undertake the study. To examine fully the research questions, this prospective experimental study was undertaken in three phases. For each of the three phases of the study the procedures used, a description of the participants of the study and the method of data analysis will be outlined.

3.2 Research Questions
Highlighted in the literature review were concerns about the reliability of the tape measures used for calf and thigh circumference measurements. It was suggested that the standard tape measure may stretch with repeated use and that a tape measure with a spring loaded attachment would provide more consistent tension of the tape between users. There was also inconsistency in the literature about which location on the leg (calf, mid-calf, ankle or thigh) would provide the most reliable measurements. Possibly the most well published aspect of reliability of circumference measurements is inter rater reliability. Despite numerous studies addressing inter rater reliability, these studies are limited in value because of the contradictory findings and the low number of examiners used in many cases. Four research questions were formulated in an attempt to address these concerns in this study:

Q1. How much does the standard tape measure stretch over time?
Q2. How consistent is the standard (Birch) tape measure compared to the gold standard Gulick spring tape when used for circumference measurements?

Q3. Which location on the leg provides the most consistent measurements?

Q4. How consistently do examiners measure leg circumferences over time?

3.3 Study Design
To answer the research questions outlined above, a mixture of quantitative research methods were employed in a three phased approach. Outlined in Table 2 are the specifications of the research design for each phase of the study. Phase One answered the question of how much the standard tape measure stretched with use over time. A non-experimental prospective design was used to measure the amount of stretch in centimetres that a standard tape measure (Birch tape) had stretched.

Non-experimental designs are appropriate in clinical research for a variety of reasons including; inability to manipulate variables due to physiological or ethical reasons, constraints placed on the research due to time, funding or inconvenience to participants or it might be an exploration of a problem as a precursor to an experimental study (Polit & Beck, 2012). For the purposes of this study, the non-experimental approach was considered appropriate due to time constraints. By using tape measures already in circulation a wider range of results was able to be collected in a short period of time.

Phase Two examined whether measurements produced with the standard tape measure (Birch tape) would agree with measurements taken with the Gulick spring tape sufficiently that it could continue to be used in clinical practice. A prospective cross-sectional experimental design was used in this phase to address the research question. Examination of the independent variables; Birch tape measure, Gulick tape measure, Calf and Thigh locations was performed using a cross-over design with the volunteers serving as their own controls. The advantage of this design of study is that there was equivalence among the volunteers.
exposed to the independent variables as each of the groups was composed of the same individuals (Polit & Beck, 2012).

Phase three of the study was conducted using a prospective experimental repeated measures design or crossover design. The use of this design meant the data collected could be used to answer both research questions three and four. The question of which location on the leg provided the most consistent measurements was answered by using just the baseline measurements from the repeated measures data collected. Again volunteers were used as their own controls in order to ensure equivalence between the independent variables of ankle, mid-calf, calf and thigh location. In answering the question of whether examiners were able to measure circumferences consistently over time data were collected in a longitudinal design. Three independent variables were assessed for this question; the examiners performing the measurements, the time point the measurement was conducted and the location on the leg that the measurement was performed.
### Table 4. Outline of study design

<table>
<thead>
<tr>
<th>Research Question</th>
<th>Phase I</th>
<th>Phase II</th>
<th>Phase III</th>
<th>Phase III</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Q1.</strong> How much does the standard tape measure stretch over time?</td>
<td><strong>Q2.</strong> How consistent is the Standard (Birch) tape measure compared to the gold standard Gulick spring tape when used for circumference measurements?</td>
<td><strong>Q3.</strong> Which location on the leg provides the most consistent measurements?</td>
<td><strong>Q4.</strong> How consistently do examiners measure leg circumferences over time?</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Between Subjects</th>
<th>Within subjects</th>
<th>Within Subjects</th>
<th>Within subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Time frame</strong></td>
<td>Cross sectional</td>
<td>Cross sectional</td>
<td>Cross sectional</td>
<td>Longitudinal</td>
</tr>
<tr>
<td><strong>Variables</strong></td>
<td>DV: Stretch in Tape (cm)</td>
<td>DV: Difference in Measurement (cm)</td>
<td>DV: Circumference Measurement (cm)</td>
<td>DV: Circumference Measurement (cm)</td>
</tr>
<tr>
<td></td>
<td>IV: Time in use (days)</td>
<td>IV:</td>
<td>IV: Location on leg</td>
<td>IV:</td>
</tr>
<tr>
<td></td>
<td>• Birch tape</td>
<td>• Ankle</td>
<td>• Time points</td>
<td>• Examiners</td>
</tr>
<tr>
<td></td>
<td>• Gulick spring tape</td>
<td>• Mid-calf</td>
<td>• Location on Leg</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Location on Leg (calf or thigh)</td>
<td>• Calf</td>
<td>• Thigh</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Control over dependent variable</th>
<th>Non-experimental</th>
<th>Experimental</th>
<th>Experimental</th>
<th>Experimental</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Measurement of variables</strong></td>
<td>Prospective</td>
<td>Prospective</td>
<td>Prospective</td>
<td>Prospective</td>
</tr>
</tbody>
</table>

**IV:** Independent Variable, **DV:** Dependent variable
3.5 Methods

3.5.1 Setting
This research was undertaken at the Spinal Injuries Unit (SIU) of the Princess Alexandra Hospital (PAH). The PAH is a metropolitan tertiary referral teaching hospital and has over 86,000 admissions yearly (Qhealth, 2011). The SIU is a 40 bed unit where patients with SCI are cared for from acute admission through rehabilitation to discharge. On average 86 patients are admitted to the unit each year with traumatic injury accounting for approximately 70% of admissions (SIU Annual Report 2006/07). There are approximately 90 nursing staff employed in the SIU comprising of registered nurses (RN), enrolled nurses (EN) and assistant nurses (AIN).

3.5.2 Sample
A recruitment campaign was undertaken prior to the commencement of the study. This involved distributing informational flyers around the spinal injuries unit in locations that were frequented by staff, i.e. staff room, procedure rooms, nurses’ station. Outlined in the flyers were the aim of the study and the inclusion and exclusion criteria for participants (see Appendix A). Information sessions were held by the investigator to promote the study and give background information outlining the need for the study. The staff were encouraged to approach the investigator confidentially if they would like to participate in the study.

3.5.3 Inclusion Criteria
Inclusion criteria were relevant for phases II and III of the study. Separate inclusion criteria were developed for the examiners performing the circumference measurements and the volunteers whose legs were to be measured. For the examiners the inclusion criteria required that the participant be a registered nurse with a minimum of three months experience working in the SIU. Registered nurses were chosen as these staff are familiar with performing calf and thigh measurements as part of their routine daily practice. For the volunteers who
were to have their legs measured the inclusion criteria required that the participant work within the SIU.

3.5.4 Exclusion Criteria
Exclusion criteria were relevant to both Phases II and III of the study. There were no specific exclusion criteria for examiners in either phase. Participants were excluded from becoming a volunteer if they had a prior history of lower limb injury or vascular disease.

3.6 Data Collection
All of the data collection for the three phases of this prospective study was performed by the investigator. Data collection forms for each of the three phases were designed prior to the commencement of the study (see Appendices B,C and D). Data collected on the forms was then entered into an electronic database by the investigator.

3.7 Procedure
As stated above the nature of the research questions posed for this study dictated that the study be undertaken in three phases each with its own specific study design. The aspects of the research that were common to all three phases have been detailed above. Several aspects of the study design however logically do not overlap between the three phases. The procedures used for each of the three phases of the study are outlined below individually.

3.7.1 Phase I
Phase I of the project involved testing the stability of the equipment currently in use within the SIU. A Birch branded standard dress making style tape measure is currently being used for the purposes of lower limb circumference measurements within the SIU. Previous studies have suggested that tape measures of this style may have a tendency to stretch with repeated use therefore possibly deeming measurements inaccurate. A search of the literature could find no studies where this theory had been tested. The aim of this phase of the study was to test if the tape measure does stretch with repeated use.
3.7.1.1 Devices

The Birch tape measure was made from fibreglass and is 150cm in length and 16mm wide (see Appendix E). The tape was double sided with measurement markings in both centimetres and inches (Liehr, Dedo, Torres, & Meininger, 1995). In line with infection control policies the tape measures are single patient use only. Each patient is supplied with a tape measure on admission to the SIU and the tape measure is disposed of when the patient no longer requires circumference measurements to be performed. The tape measures are not labelled with the patient’s name but are kept at the patient’s bedside for their sole use.

3.7.1.2 Data Collection

All of the data collection in this phase was performed by the investigator. A purpose specific data collection log was designed for recording the measurements. (see Appendix B). Information collected included; a unique study ID number, the patient’s Unit Record (UR) number, the date of admission to the SIU, the date the measurement was performed and the length of stretch in millimetres.

The date of admission to SIU was retrieved from the admission entry in the patient’s progress notes. It was assumed that the tape measure had been in use since this date as it is usual practice that all patients have baseline calf and thigh circumference measurements performed on the day of admission to the unit and then every day throughout admission.

3.7.1.3 Procedure

Twenty tape measures currently in use within the SIU at one hospital were randomly selected by the investigator, from the patients currently having calf and thigh circumference measurements performed. The tape measures were collected and labelled with the patients UR number, the unique hospital identification number. A brand new, Birch tape measure consistent with those currently in use was used as the control for this phase of the study. The control tape was pulled taut and affixed to a flat surface. One at a time each of the tape
measures belonging to the patients was then positioned alongside the control tape and affixed to the flat surface at the zero end of the tape. The patient’s tape was then pulled taut and affixed to the flat surface at the other end of the tape. The difference in length of each tape at the 150cm end of the tape was then measured using a third brand new birch tape measure. The difference in length of each tape was recorded on the log sheet in millimetres. Upon completion of this phase of the project the tape measures were returned to the patients’ bedside for continued use.

3.7.1.4 Data Analysis

Data from the data collection sheets were entered into a Microsoft Excel (2007) spreadsheet for analysis. A visual inspection of the raw data was undertaken to check for the presence of data entry errors prior to analysis. The length of time the tape had been in use was measured in days by subtracting the date of the baseline measurements from the current date plus one (e.g. 01/06/2011 – 01/03/2011 + 1 = days in use). The median mode and range of the variables of days in use and tape measurement were calculated. The days in use were plotted against the tape measurements on a histogram for visual interpretation of the results.

3.7.2 Phase II

Phase II of the study was undertaken to examine the equivalence of the equipment used for calf and thigh circumference measurements. The Birch tape is a standard dress makers tape. The use of such tapes for circumference measurements has been called into question in past studies, but is commonly used throughout many areas of healthcare. One of the main concerns raised in prior studies was the variable tension that may be applied to the tape measure by examiners. The Gulick spring tape has a tensioning device attached to the zero end of the tape measure allowing users to apply a consistent amount of tension to the tape with every measurement (see Appendix E). In studies where the Gulick tape has been used high levels of inter-rater reliability have been found. The aim of this phase of the study
was to test the level of agreement between the Birch tape measure in common use and the Gulick spring tape identified in the literature as being the gold standard measuring device.

3.7.2.1 Devices
Two different tape measures were used in this phase. The first was a standard Birch measuring tape, similar to a dressmaker’s tape, measuring 152cm in length and 1cm in width made from fibreglass (see Appendix E). The measuring tape was identical to those that have been in use in the SIU for more than 10 years. Examiners were asked to loop the tape around the volunteers’ leg and read the measurement using the end of the tape as a zero point. The same tape was used by all examiners on all volunteers.

The second tape was a Gulick spring tape measure. This tape measure is a retractable tape with a spring loaded handle attachment (see Appendix E). The tape itself is approximately 0.7cm in width. The spring loaded handle attachment is metallic and approximately 5cm in length.

The tape measure is wrapped around the leg in a similar fashion to the birch tape measure. To apply tension the spring handle is pulled to a pre-marked notch on the handle. The examiner then reads the measurement from the zero point on the tape. The Gulick spring loaded tape measure has been shown in the literature to allow the examiner to “provide a consistent amount of tension with every measurement” (Bunce, 2009)

All examiners and volunteers were provided with a demonstration of how the tape measure worked at the time of enrolment as this new tape measure was unfamiliar to the staff of the spinal injuries unit. Examiners were also instructed to practice using the device on the volunteers’ alternate leg three times prior to undertaking their first measurement.

3.7.2.2 Anatomical Locations
The anatomical locations the measurements were to be performed for this phase of the study were based on current usual practice. The protocol used in this study included:
1. Have the patient lay in a flat supine position.

2. For calf measurement the mid-point between the base of the patella and the ankle was located and a piece of white surgical tape was placed in this position.

3. The process was repeated for the thigh measurement using the midpoint between the top of the patella and the top of the thigh and a piece of white surgical tape was placed in this position.

The white surgical tape used in this tape was approximately 0.5cm in width. This process was undertaken by the investigator to ensure consistency of placement of the tape. Examiners were instructed to place the tape flat underneath the volunteer’s leg and wrap the tape around the leg at the base of the white tape. The measurement should be read by placing the 0cm mark against the other side of the tape and pulling taut. Examiners were asked to read aloud the measurement to the nearest 0.1cm.

3.7.2.3 Randomization

Randomization for Phase II of the study was conducted on two levels. Volunteers were randomized to either right or left leg to be measured. The leg to be measured was randomly selected by the investigator at the time of volunteer enrolment to the study. The randomization scheme was generated using a random number generator function in Microsoft Excel (2007). In addition examiners were randomised to which measuring device they would use first. At the time of their enrolment to the study examiners were told which measuring device they would use first, either the birch tape measure or the Gulick spring tape measure. This was randomly selected by the investigator using the random number generator function in Microsoft Excel (2007). Examiners subsequently performed all measurements using their assigned device first, then the alternate device.
### 3.7.2.4 Procedures

All measurements were undertaken in a procedure room within the spinal injuries unit. Volunteers were asked to wear shorts or loose fitting pants that would enable a measurement to be taken at the thigh location. Volunteers were asked to lie in a comfortable supine position on a standard hospital bed with legs extended in front of them flat on the bed. The investigator then measured each of the locations using the landmarks noted above placing a piece of adhesive tape at each measuring point. Prior to the commencement of each session volunteers were instructed to try to refrain from making comments about the technique examiners used to perform the measurements or the measurement that the examiner read out aloud.

Examiners were then instructed to measure the circumference of the calf and thigh at the base of the white tape in each location. The examiners were asked to use their randomly assigned measuring device first to measure each location before measuring with the alternate device.

*Example: Examiner 2 has been randomly assigned to use the Birch tape measure first. The examiner would measure the calf circumference first with the Birch tape measure and then measure the thigh with the Birch tape measure. The examiner would then measure the calf circumference first with the spring tape measure and then the thigh with the spring tape measure.*

Examiners were able to start with either the calf or the thigh measurement first but were asked to complete both measurements prior to using the alternate device. Each measurement was read out aloud by the examiner for the investigator to record on a pre-formatted data collection sheet. None of the examiners had access to the data collection sheets. Each of the examiners measured just one volunteer at a time and only one examiner measured the volunteer at a time.
3.7.2.5 Data Analysis
Statistical analysis for Phase II of the study was undertaken with Microsoft Excel (2007), STATA v9.0, and MedCalc v12.5.0.0. All de-identified data was entered into a Microsoft Excel spreadsheet for preliminary analysis. Data were inspected visually for errors and all outliers and missing data were individually assessed for data entry errors.

Data were then entered into Medcalc v12.5.0.0 for further analysis. Univariate analysis was undertaken on each of the variables to calculate measures of central tendency and variability in the data distribution. The difference between each pair of measurements was calculated by subtracting the Birch tape measurement from the Gulick spring tape measurement. The distribution of differences was then assessed for normality by visual inspection of a histogram and a scatter plot showing +/- 2 standard deviations.

In order to assess the agreement between the two tape measures the limits of agreement method proposed by Bland and Altman was used (Bland & Altman, 2010). Bias was assessed visually by plotting the measurements from the two tapes along the line of equality. Two methods in equal agreement will lie along the line perfectly. Bland Altman plots for repeated measures were constructed. The limits of agreement were calculated using the method outlined by Carstensen et al (2008). This analysis was undertaken using STATA v 9.0.

3.7.3 Phase III
The third phase of the study involved examining the reliability of various potential anatomical landmarks to be used by multiple examiners. In testing the equivalence of the procedure two factors need to be examined; 1) the landmarks being used to conduct the measurements and 2) the people performing the measurements. The reliability of the current landmarks used for circumference measurements were compared with the reliability of alternate landmarks identified in the literature and this reliability was examined over multiple
time-points. For the purpose of this study the current method of testing calf circumference was tested against using the midpoint between the tibial tuberosity and medial malleolus and an ankle measurement taken 2 cm proximal to the medial malleolus (see Appendix F).

Inter-rater reliability was assessed by comparing the results of each of the examiners at each of the measurement locations for each time-point. In order to assess inter-rater reliability multiple examiners needed to perform the same measurement at multiple time points. Inter-rater reliability is of particular importance in the clinical setting as circumference measurements will rarely be performed by the same nurse over consecutive days. In a retrospective study of calf circumference measurements undertaken in an SIU, Swarczinski (1991) found an average of 15 nurses performing measurements for one patient (Swarczinski & Dijkers, 1991).

3.7.3.1 Devices
A standard Birch measuring tape, similar to dressmakers measuring tape, measuring 152cm in length and 1cm wide was used for this phase of the study. The same tape was used by all examiners on all of the volunteers.

3.7.3.2 Locations
Four separate locations were measured on each volunteer, referred to as thigh, calf, mid-calf and ankle. Each of these locations was identified by the investigator using anatomical landmarks to ensure consistency in measurement location (see Appendix F).

- Thigh: The thigh point was determined by taking the midpoint between the hip and the top of the patella.
- Calf: The calf point was the midpoint between the base of the patella and the ankle.
- Mid-Calf: The mid-calf point was the midpoint between the tibial tuberosity and the medial malleolus.
- Ankle: The ankle measurement was taken at 2 cm above the medial malleolus.
The distance from landmark to measurement location was recorded on the data collection form for each of the locations. This enabled the investigator to ensure that the tape was placed in the same location at each subsequent measurement or if the white adhesive tape became dislodged during measurements (see Appendix D).

Each of the locations to be measured on each volunteer was marked with a single piece of white adhesive tape approximately 0.5 cm in width. The piece of white tape was placed in the appropriate locations by the researcher on all occasions. The base side of the white tape was placed on the location so that the examiner could measure against the base of the white tape. The white tape was left in place until all of the measurements had been completed on any given day.

3.7.3.3 Randomisation
Randomisation was twofold for this phase of the study. Volunteers were randomised as to which leg would be measured and the order in which examiners measured each location was randomised at each time point. Randomization was undertaken using a random number generator function in Microsoft Excel (2007). The leg to be measured was randomly selected by the investigator at the time of enrolment to the study by the volunteer. To minimize the chance that the examiners would remember measurements between the time points, the order in which measurements were taken was altered by the investigator at each time point.

3.7.3.4 Procedure Phase III
All measurements were undertaken in a procedure room within the spinal injuries unit. Volunteers were asked to wear shorts or loose fitting pants that would enable a measurement to be taken at the thigh location. Volunteers were asked to lie in a comfortable supine position on a standard hospital bed with legs extended in front of them flat on the bed. The investigator then measured each of the locations using the landmarks noted above placing a piece of adhesive tape at each measuring point.
Examiners were asked to measure at each of the marked locations by placing the measuring tape around the limb against the base of the white tape. Examiners were instructed to take the measurement from the end of the tape measure which was equivalent to the zero mark and to state aloud the measurement to the nearest 0.1 cm. This procedure was repeated by each examiner on each volunteer at the three time points – baseline, 15 mins after baseline and 30 mins after baseline. The order in which locations were measured was also randomised by the investigator at each time point. All measurements were recorded on a prepared data collection sheet by the investigator. None of the examiners had access to any of the data sheets. The examiners were asked to engage in other activities, between measurement periods to minimise the chance that they may recall the previous measurement. In many cases the examiners returned to their usual clinical activities between measurements.

### 3.7.3.5 Data Analysis
Statistical analysis for Phase II of the study was undertaken with Microsoft Excel (2007), and MedCalc v12.5.0.0. All de-identified data were entered into a Microsoft Excel spreadsheet for preliminary analysis. Data were inspected visually for errors and all outliers and missing data were individually assessed for data entry errors.

Data were then entered into MedCalc v12.5.0.0 for further analysis. Univariate analysis was undertaken on each of the variables to calculate measures of central tendency and variability in the data distribution. The relative standard deviation for each examiner was calculated based on baseline observations only and ranked for each location.

For analysis of inter rater reliability the mean imputation method was used to impute missing data. The mean of the completed measurements for the volunteer was used where a measurement was missing. The data were then checked for homogeneity, sphericity and normality in order to meet the assumptions of analysis of variance (ANOVA). A two way
ANOVA was carried out with the examiners and time points as factors. A significance level of $p = <0.05$ was used for analysis of variance calculations.

3.8 Ethics

Approval for this study was given by the Management Committee of the Spinal Injuries Unit of the Princess Alexandra Hospital. Ethical approval was granted from the Human Research Ethics Committees of the Princess Alexandra Hospital and Griffith University before commencement of the project (see Appendices G, H and I).

The National Health and Medical Research Council (NHMRC) outline the key principles which are central to conducting ethical research, in the National Statement on ethical conduct in human research (NHMRC, 2007). These are: research merit and integrity, justice, beneficence and respect. The key principles are not merely a set of rules by which to conduct the research project but rather are interwoven in the design of the study and displayed in the conduct of the investigator.

The merit and integrity of a study can be displayed by conducting research that is of value, is conducted by researchers with appropriate skills, is undertaken in a manner that is considerate of resources and that the results will be distributed. This research merit and integrity is displayed in the design of the current study and proof through the literature review of the value of conducting the study. Although the study was undertaken by a novice researcher, there was an appropriate level of supervision in the form of research academics. The plan for distribution of the results of this study includes the fulfilment of this thesis and plans to disseminate the results both at a local level and through publication.

The principle of justice in research relates to the fair recruitment of individuals into the study and that there is no exploitation of participants. The recruitment campaign for this study was an open process inviting all staff members that met the inclusion criteria to apply. All participants in Phases I and II were provided information about the project and their
proposed involvement in the project prior to consent. Participants were informed that participation is voluntary and they were free to withdraw at any time throughout the project. The informed consent process was undertaken by the Investigator and conformed to the guidelines of Informed consent outlined by the Human Research Ethics Committee of the Princess Alexandra Hospital (see Appendices J and K).

The principle of beneficence relates to the perceived benefits of the research and goes hand in hand with the principle of non-maleficence or the idea of doing no harm. As the research did not involve invasive procedures there was minimal risk to participants of this study. A potential risk to both examiners and volunteers in Phase II of the study was that of anxiety related to the performance of circumference measurements. There was the potential for staff to feel as though the standard of their work was in question. This was addressed through comprehensive education about the objectives of the study provided to all of the participants prior to the commencement of the project and at the time of the recruitment.

An increase in staff awareness and knowledge of venous thrombo-embolism and the potential consequences for the SCI patient were one of the benefits for participants in this study. This also translates into a potential benefit for the patients of the SIU and the organisation in terms of having nursing staff perhaps better informed and vigilant to the risks of VTE.

Respect for participants of research is shown not only through the application of the principles of merit and integrity, beneficence and justice, but also through considerations of the participants’ right to confidentiality and privacy. This idea of respect took on many forms throughout the current research and was displayed in the consideration of participants’ privacy, maintaining confidentiality and through appreciation of their participation.

All information collected from participants during the course of the study was de-identified prior to analysis. In each phase of the study participants were given a unique study
identification code which was used on data collection sheets pertaining to the participant. A log sheet containing the code and the corresponding participant’s name was stored in a separate location to the data collection sheets and accessible only to the Investigator.

Hard copies of data were stored in a locked filing cabinet whilst electronic data was stored on a password protected computer. All information collected pertaining to the project will be archived for a period of no less than five years from the completion of the study and then destroyed.

**3.9 Summary**

Outlined in this chapter were the research questions which guided the current study. The methods used in order to conduct each of the three phases of the study were outlined in the context of the research questions. A detailed description of the participant recruitment and selection process was provided as well as the procedures undertaken by participants in the course of the study. Ethical considerations in designing and conducting the study were also discussed. A description of the variables under study was given along with details about analysis of the data. The results of the analysis of the data collected in the study will be presented in Chapter Four.
4.0 Results

The aim of this research was to test the reliability of calf and thigh circumference measurements. Reliability testing involves testing the degree of dependability of an instrument and or process, measuring the object it was designed to measure. The reliability in this study was assessed by examining the equivalence and stability of both the equipment, in Phases I and II, and the procedure in Phase III. Two aspects of the procedure were examined including the landmarks used in the current procedure and the reliability of the person taking the measurement. The results from the three phases of the study will be presented here in the context of the research questions which support the study (see Table 3).

Table 5. Overview of study

<table>
<thead>
<tr>
<th>Study Phase</th>
<th>Research Question</th>
<th>Aspect of Reliability Examined</th>
<th>Study Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase I</td>
<td>Q1. How much does the standard tape measure stretch over time?</td>
<td>Stability of Equipment</td>
<td>Nil</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>20 Birch Tape Measures</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>examined</td>
</tr>
<tr>
<td>Phase II</td>
<td>Q2. How consistent is the standard (Birch) tape measure compared to the gold</td>
<td>Equivalence of Equipment</td>
<td>Examiners: 16</td>
</tr>
<tr>
<td></td>
<td>standard Gulick spring tape when used for circumference measurements?</td>
<td></td>
<td>Volunteers: 32</td>
</tr>
<tr>
<td>Phase III</td>
<td>Q3. Which location on the leg provides the most consistent measurements?</td>
<td>Equivalence of Procedure</td>
<td>Examiners: 8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Volunteers: 16</td>
</tr>
<tr>
<td></td>
<td>Q4. How consistently do examiners measure leg circumferences over time?</td>
<td>Equivalence between observers</td>
<td></td>
</tr>
</tbody>
</table>
Research Q1. How much does the standard tape measure stretch over time?

4.1 Test for stability of equipment (Birch Tape Measure)

Stability of the standard sewing tape measure used in practice was assessed over time by determining if repeated use resulted in stretching of the tape. In order to comply with infectious diseases policies, disposable tape measures are issued to each patient on admission to the SIU. The same tape measure is used each day to perform leg measurements until the practice is ceased usually 8 to 12 weeks later. The tape measure used was a Birch analogical tape measure, 150cm long and 16 millimetres wide made from fibreglass (Liehr et al., 1995).

To assess any change in measurement of the tape over time, 20 tape measures currently in use were compared to a new tape measure identical in brand. Each tape measure was collected from a patient bedside and labelled to ensure it was returned to the correct patient. The 20 tape measures were tested in two batches of ten tapes over two days.

The length of time the tape had been in use was measured in days by subtracting the date of the baseline measurements from the current date plus one (e.g. 01/06/2011 – 01/03/2011 + 1 = days in use). The number of days in use was plotted against the amount of stretch in millimetres per tape measure (see Figure 3).
The twenty tapes had been in use from 5 to 86 days. The change in tape length when compared with a new previously unused tape was 2.0mm (median) with a range of 0mm to 7mm. The amount of stretch was not dependent on time in use with the tapes that had been in use for the shortest and longest days each stretching by the same amount of 1mm.

**Research Q2. How consistent is the Standard (Birch) tape measure compared to the gold standard Gulick spring tape when used for circumference measurements?**

**4.2 Test for equivalence of equipment (tape measure)**

The second aspect of reliability to be examined was equivalence. The standard tape measure was compared against a Gulick spring tape measure which has been cited in the literature as the gold standard for measurements (Tunc et al., 2007). The Gulick spring tape measure allowed the tester to apply the same tension to each measurement performed thus potentially reducing measurement error associated with tensioning (Swarzinski & Dijkers, 1991).
Thirty-two healthy volunteers from the nursing staff of the SIU were recruited for this phase of the study, 24 female and 8 male. Each of the volunteers had the calf and thigh circumference of one of their legs measured by each of the examiners (n=16), registered nurses employed in the SIU. All examiners had been employed in the SIU for a minimum of one year at the time of study participation.

The research plan was for each of the 16 examiners to measure the calf and thigh circumferences of the same one leg of each volunteer using both tape measures. However, due to availability and time constraints not all examiners measured all volunteers. The minimum number of volunteers measured by an examiner was 9 of 32, with the median being 27 of 32. One pair of measurements consisted of a calf or thigh measurement performed with both the Birch tape measure and the Gulick spring tape measure. A total of 804 pairs of measurements were performed.

4.2.1 Calf Measurements
A total of 402 sets of calf circumference measurements were performed using both the Birch and Gulick spring tape measures. Calf measurements ranged from a low of 29.5 cm using the Gulick spring tape to a high of 46.0 cm with the Birch tape, with a normal distribution (see Table 4). The range of measurements between the two tapes was similar as seen in Figure 4 though the Birch tape measurements were consistently larger.
Table 6. Calf measurements using the Birch and Gulick spring tapes (n=804)

<table>
<thead>
<tr>
<th>Descriptive</th>
<th>Birch Tape (n=402) (cm)</th>
<th>Gulick Tape (n=402) (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (SD)</td>
<td>36.6 (3.2)</td>
<td>35.6 (3.2)</td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>36.4 (4.7)</td>
<td>35.5 (4.5)</td>
</tr>
<tr>
<td>Range</td>
<td>29.5 – 46.0</td>
<td>28.0 – 44.7</td>
</tr>
</tbody>
</table>

*Figure 4. Box plot of distribution of calf measurements (n=402) from each tape*

The differences between each pair of measurements taken were calculated by subtracting the measurement taken with the Birch tape from the Gulick spring tape measurement which is considered the gold standard. Measurement differences ranged from -7.0cm to 3.0cm with 96 or 23.9% of measurements falling outside of the +/- 1.5cm range considered to be significant in clinical practice. The difference measurements plotted in
Figure 5 visually approximate a normal distribution with a mean measurement difference of -0.98cm (SD =1.02) with a confidence interval of -1.08 to -0.88cm (95%).

**Figure 5.** Distribution of calf measurement differences between tape measures (n=402)

### 4.2.2 Thigh Measurements
A total of 402 sets of thigh measurements were collected using both the Birch and Gulick spring tape measures. The descriptive statistics for measurements collected from each tape are displayed in Table 5. The smallest thigh circumference measurement was 38.0cm taken with the Gulick tape, and the largest measurement was 69.9cm taken with the Birch tape. Presented in Figure 6 is a box plot of the distribution of measurements taken with each tape indicating the presence of outliers.
Table 7. Thigh measurements using the Birch and Gulick spring tapes (n=804)

<table>
<thead>
<tr>
<th>Descriptive</th>
<th>Birch Tape (n=402) (cm)</th>
<th>Gulick Tape (n=402) (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (SD)</td>
<td>51.96 (6.5)</td>
<td>50.26 (6.4)</td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>51.6 (8.9)</td>
<td>50.0 (7.2)</td>
</tr>
<tr>
<td>Range</td>
<td>38.2 – 69.9</td>
<td>38.0 – 68.0</td>
</tr>
</tbody>
</table>

The difference between each pair of thigh measurements was calculated by subtracting the Birch tape measurement from the Gulick spring tape measurement. The measurement differences plotted in Figure 7 show a normal distribution from -8.0cm to 8.1cm, with 60.7% of measurements ≥-1.5cm or ≤1.5cm. The mean difference for thigh measurements was -1.69cm (SD 1.47) with a 95% confidence interval of -1.7 to -1.5 cm.
Figure 7. Number of measurement differences in thigh circumferences (n=402) taken with Gulick spring tape and Birch Tape.

4.2.3 Agreement between two tapes
Agreement between the two tape measures was assessed by calculating the limits of agreement at both the calf and thigh locations. The critical value for agreement level of +/- 1.5cm was set a priori, as a change in measurements of 1.5cm or greater indicated a need for further medical review within current practice within the SIU.

Calculation of the limits of agreement is dependent on the assumption that the differences follow an approximately normal distribution and that there is no relationship between the difference and the magnitude of the measurement (Bland & Altman, 2010). No relationship was evident between the differences and the magnitude of measurements as shown in Figures 8 and 9 therefore calculation of the limits of agreement was performed.
Figure 8. Difference versus average of values measured by examiners at the calf location

Figure 9. Difference versus average of values measured by examiners at the thigh location
The differences are considered to be normally distributed if 95% of the differences lie within two standard deviations of the mean (Connelly, 2008). It is worth noting that Bland and Altman (1999) point out that it is ‘not necessary for the measurements themselves to follow a normal distribution’ (p.138). For this study, there were just 16 measurements (4%) of differences outside the two standard deviations from the mean at the calf level (see Figure 10). At the thigh level, 17 (5%) measurements were outside two standard deviations of the mean indicating that both the calf and thigh differences can be considered a normal distribution (see Figure 11).

Figure 10. Distribution of differences between Birch and Gulick tapes at calf level showing +/- 2SD from mean (n=402)
Figure 11. Distribution of differences between Birch and Gulick tapes at thigh level showing +/- 2SD from the mean (n=402)

A scatter plot of the individual measurements was prepared to show the differences between the Birch tape and Gulick Spring tape (see Figures 12 and 13). The plot also shows the line of equality, the line all plots would lie on if the two devices always gave the same reading. Visually this plot shows us how well the two devices agree but will not demonstrate the variation in the differences between the two methods (Connelly, 2008). In the current study, the birch tape consistently measured higher than the Gulick spring tape (Figures 12 and 13).
Figure 12. Birch tape calf measurements plotted against Gulick spring tape calf measurements (cm) with line of identity.

Figure 13. Birch tape thigh measurements plotted against Gulick tape thigh measurements (cm) with line of identity.

A Bland Altman plot, using multiple measurements per subject where the true value is constant, was constructed to show the variation in differences between the two tape measures.
The paired data for the 402 sets of calf measurements taken with the Gulick and Birch tapes in shown in Figure 14. The mean of the two measurements for each subject are plotted on the x axis and the difference between the two measurements for each subject is plotted on the y axis.

Analysis of the plots is largely visual utilising the display of the 95% limits of agreement, where 95% of the differences between the two devices should fall (Bland & Altman, 1999). The limits of agreement are defined by the mean difference plus or minus two times the standard deviation of differences \( (\bar{d} \pm 2s_d) \). The closer together the limits are; the greater the agreement between the two devices or methods.

The limits of agreement were calculated using the model proposed by Carstensen and colleagues (2008) for repeated measures with exchangeable variables. For the calf measurements the lower limit of agreement was -3.66cm with a 95% confidence interval of -3.89 to -3.43cm. The upper limit of agreement was 1.66 cm with a 95% confidence interval of 1.43 to 1.89cm.

As well as displaying the level of agreement the plots also indicate the bias and the presence of outliers. Bias tells us how different the new device is from the gold standard. Bias is represented by the mean difference (the middle line) on the graph. The closer this line is to zero the less bias (Germing et al.). For this study the mean difference or bias at the calf level was calculated to be -0.99cm, indicating that the measurements taken with the birch tape were consistently higher than those taken with the Gulick spring tape.

Outliers can indicate a source of measurement error or simply be a mistake in data collection or entry. Connelly suggests that more than three unexplained outliers per 100 measurements is suggestive of a possible problem with the measurement system (Germing et al.). At the calf measurement level in this study no outliers were present outside the limits of agreement for the study.
For thigh measurements, the mean difference or bias was -1.70cm. The lower limit of agreement was plotted at -6.19cm with a confidence interval of -6.01 to -6.37cm. The upper limit of agreement was 2.79cm with a confidence interval of 2.61 to 2.97cm. Again at the thigh level there were no outliers detected (see Figure 15).

The plots show the majority of the measurements clustered around the mean for both the calf and the thigh indicating a high level of agreement between the two tape measures. The limits of agreement for both the calf and the thigh locations are quite wide, well outside the 1.5cm set for agreement.

Figure 14. Bland Altman plot of differences between Birch tape measure and Gulick spring tape in relation to the mean of the two tape measures at the calf level (n=402)
Figure 15. Bland Altman plot of differences between the Birch tape and Gulick spring tape measures in relation to the mean of the two tape measures at the thigh level (n=402)

4.2.4 Reproducibility

Reproducibility is one aspect of precision. It is an important factor to consider in an agreement study because if one instrument has poor precision then the two instruments are unlikely to agree. Reproducibility is the variability in repeated measurements when one or more factors have been altered, for example the instrument, the observer, or time. In this phase of the study the variability in the repeated measures was the examiner. The reproducibility of measurements taken with each of the tape measures was calculated separately at the calf and thigh level.

A one way analysis of variance (ANOVA) was performed for each tape measure at the calf level after checking the assumptions that underlie the ANOVA test (see Table 6 and 7). From this the within subject standard deviations (sw) were obtained by taking the square root of the mean square of the residual (Birch sw = 0.954cm, Gulick sw= 1.024cm). The
reproducibility coefficients for each tape measure were calculated \((2.77 \times s_w)\) to be 2.84cm for the Gulick tape and 2.64cm for the Birch tape. The reproducibility coefficient can be expressed as a limit showing the likely limits in which 95% of the measurements taken with the same tape measure should fall (Bland & Altman, 2010; Mcalinden, Khadka, & Pseudovs, 2011). For the Gulick tape 95% of the differences between measurements taken with this tape would fall +/- 2.84cm. The limits for the Birch tape are +/- 2.64cm.

Bland and Altman recommend comparing these to the limits of agreement to determine if the lack of agreement between methods is explained by a lack of reproducibility (Connelly, 2008) (see Figure 16). As the limits of agreement for the calf level are much wider (1.66 to -3.66cm) than the reproducibility limits for either tape it is likely that there is some other factor lowering the agreement between the two tape measures.

Table 8. ANOVA of Birch tape measurements at calf location (n=402)

<table>
<thead>
<tr>
<th>Model</th>
<th>Sums of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>3825.6893</td>
<td>31</td>
<td>123.4093</td>
<td>135.533</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Within Groups</td>
<td>336.9035</td>
<td>370</td>
<td>0.9105</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>4162.5927</td>
<td>401</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 9. ANOVA of Gulick spring tape measurements at calf location (n=402)

<table>
<thead>
<tr>
<th>Model</th>
<th>Sums of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>3641.2895</td>
<td>31</td>
<td>117.4610</td>
<td>118.831</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Within Groups</td>
<td>388.6268</td>
<td>370</td>
<td>1.0503</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>4029.9162</td>
<td>401</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 16. Bland Altman plot showing reproducibility coefficients for Birch and Gulick spring tapes at calf location (cm)

A one way ANOVA with the volunteers as a factor was calculated for the Birch and Gulick tape measures at the thigh location (see Table 8 and 9). The within subject standard deviations were calculated as Birch= 1.51cm and Gulick = 1.68 cm. The reproducibility coefficients were then calculated as 4.18cm for the Birch tape and 4.65 cm for the Gulick tape with the reproducibility limits being +/-4.18cm and +/- 4.65 cm respectively. In comparing these reproducibility limits to the limits of agreement (-6.19 to 2.79 cm) it can be seen in Figure 17 that the limits of agreement at the thigh location are much wider than the limits of reproducibility.
Table 10. ANOVA of Birch tape measurements at the calf location (n=402)

<table>
<thead>
<tr>
<th>Model</th>
<th>Sums of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>16102.6799</td>
<td>31</td>
<td>519.4413</td>
<td>225.739</td>
<td>0.001</td>
</tr>
<tr>
<td>Within Groups</td>
<td>851.3947</td>
<td>370</td>
<td>2.3011</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>16954.0746</td>
<td>401</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 11. ANOVA of Gulick spring tape measurements at the thigh location (n=402)

<table>
<thead>
<tr>
<th>Model</th>
<th>Sums of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>15121.5631</td>
<td>31</td>
<td>487.7924</td>
<td>171.245</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Within Groups</td>
<td>1053.9446</td>
<td>370</td>
<td>2.8485</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>16175.5077</td>
<td>401</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 17. Bland Altman plot showing reproducibility coefficients for Birch and Gulick spring tapes at thigh location (cm)
Research Q3. Which location on the leg provides the most consistent measurements?

Research Q4. How consistently do examiners leg circumferences over time?

4.3 Test for equivalence of current procedure
In testing the equivalence of the current procedure used in usual practice, two factors needed to be examined; 1) the landmarks being used to conduct the measurements and 2) the people performing the measurements.

The reliability of the current landmarks used for circumference measurements was compared with the reliability of alternate landmarks over multiple time points identified in the literature. For the purpose of this study the current method of testing calf circumference using the patella as a landmark was tested against using the midpoint between the tibial tuberosity and medial malleolus to assess the equivalence of each location.

Inter-rater reliability was assessed by comparing the results of examiners at each of the measurement points for each time-point. In order to assess inter-rater reliability multiple examiners need to perform the same measurement at multiple time points. Inter-rater reliability is of particular importance in the clinical setting as circumference measurements will rarely be performed by the same nurse over consecutive days.

Sixteen healthy volunteers, two male and 14 female, with no prior history of vascular disease or lower limb injury were recruited from the nursing staff of the SIU at the PAH. Eight registered nurses were also recruited from the nursing staff of the SIU at the PAH, as examiners. Due to availability, one of these examiners ultimately withdrew from the study and was replaced. All of the examiners recruited had worked in the SIU for a minimum of one year and performed calf and thigh circumference measurements as part of their daily activities.
In total 116 sets of data of a possible 128 (91%) were collected in phase III resulting in a total of 348 measurements. One complete set of data contains measurements at each location over the three time points. Twelve complete sets of data were unable to be obtained due to time constraints and availability of examiners and or volunteers. Incomplete sets of data were not used in the analysis.

The results displayed in Table 10 are based on 348 individual measurements for each location. The range for each location (ankle: 6.5, mid-calf: 16.7, calf: 17.2, thigh: 30.8) shows the heterogeneity of the sample used in this study. The mean, median and mode coincided at the ankle (22cm) and calf (37cm) locations indicating a symmetric unimodal distribution. The mode and mean differed slightly at the mid-calf (mean: 33.6 cm, mode: 31cm) and thigh (mean: 52.9cm, mode: 49 cm) locations indicating a slightly skewed distribution.

Symmetrical distributions have a skew equal to zero, though distributions for biological data are commonly positively skewed (Kestin, 2006). The distributions of measurements at all four locations for this study were positively skewed (see Figure 18). The calf location also had a positive kurtosis indicating a slightly peaked distribution as indicated in the distribution histograms in Figure 18(c). The lowest standard deviation was 1.4cm at the ankle and the highest 7.1cm at the thigh indicating more variability in measurements at the thigh location.
Table 12. Descriptive statistics at each measurement location (n=348)

<table>
<thead>
<tr>
<th></th>
<th>Ankle(cm)</th>
<th>Mid-Calf(cm)</th>
<th>Calf(cm)</th>
<th>Thigh(cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Range</strong></td>
<td>20.0-26.5</td>
<td>26-42.7</td>
<td>29.3-46.5</td>
<td>40.2 – 71.0</td>
</tr>
<tr>
<td><strong>Mean (SD)</strong></td>
<td>22.7 (1.4)</td>
<td>33.6 (3.7)</td>
<td>37.2 (3.0)</td>
<td>52.9 (7.1)</td>
</tr>
<tr>
<td><strong>Median (IQR)</strong></td>
<td>22.5 (1.9)</td>
<td>33.0 (4.0)</td>
<td>37.0 (5.35)</td>
<td>51.5 (12.5)</td>
</tr>
<tr>
<td><strong>Mode</strong></td>
<td>22</td>
<td>31</td>
<td>37</td>
<td>49</td>
</tr>
<tr>
<td><strong>Skewness</strong></td>
<td>0.45</td>
<td>0.47</td>
<td>0.61</td>
<td>0.67</td>
</tr>
<tr>
<td><strong>Kurtosis</strong></td>
<td>-0.59</td>
<td>-0.48</td>
<td>0.38</td>
<td>-0.25</td>
</tr>
</tbody>
</table>

*Notes*: SD: Standard Deviation, IQR: Inter-quartile Range
Figure 18. Distribution of measurements at each location

a) Ankle Measurement in cms

b) Calf Measurements (cms)

c) Mid-calf Measurement (cms)

d) Thigh Measurement (cms)
4.3.1 Reliability of alternate landmarks

In order to examine the consistency of measurements across landmarks the first obtained measurements from each landmark was analysed as these were considered to be baseline measurements. To assess the variability in scores for each landmark the relative standard deviation for each volunteer was calculated. The relative standard deviation was used as it allows for comparison of variance across different items in a meaningful way.

At each location the relative standard deviation (rsd) was calculated for each volunteer (n=16) and then ranked from lowest to highest. The number of volunteers with the lowest rsd, in other words the lowest variance, at each location was calculated and results are presented in Table 11 below. Based on these results the calf location had the greatest number of volunteers with the lowest rsd thereby indicating it is the most reliable. The ankle location was the next most reliable with the thigh and mid-calf locations the least reliable.

Table 13. Consistency across landmarks using relative standard deviation (rsd) (n=16)

<table>
<thead>
<tr>
<th>Location</th>
<th>Lowest rsd (%)</th>
<th>2nd Lowest rsd (%)</th>
<th>3rd Lowest rsd (%)</th>
<th>4th Lowest rsd (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ankle</td>
<td>6 (37.5)</td>
<td>5 (31.3)</td>
<td>3 (18.8)</td>
<td>2 (12.5)</td>
</tr>
<tr>
<td>Mid-calf</td>
<td>0 (0)</td>
<td>4 (25)</td>
<td>4 (25)</td>
<td>8 (50)</td>
</tr>
<tr>
<td>Calf</td>
<td>9 (56.3)</td>
<td>3 (18.8)</td>
<td>2 (12.5)</td>
<td>2 (12.5)</td>
</tr>
<tr>
<td>Thigh</td>
<td>1 (6.3)</td>
<td>4 (25)</td>
<td>7 (43.8)</td>
<td>4 (25)</td>
</tr>
</tbody>
</table>

4.3.2. Inter-rater Reliability

Different examiners may be able to perform circumference measurements on the same volunteer reliably, but this may be affected when repeated measures are taken over time. To examine this in the current study a two way repeated measures analysis of variance was used to test the reliability of circumference measurements performed on 16 volunteers by 8
examiners over 3 time periods each 15 minutes apart. This analysis was repeated for each of
the four measurement locations; thigh, calf, mid-calf and ankle.

4.3.2.1 Missing Data
Twelve sets of data were unable to be collected in this phase of the study due to
availability of the volunteers. In order to conduct a balanced ANOVA the missing data were
imputed for this phase using the mean imputation method (Bland & Altman, 2003). The
mean of completed measurements for the volunteers at each location and time phase were
imputed where data were missing.

4.3.2.2 Assumptions of ANOVA
The assumptions of ANOVA were checked prior to analysis and some minor
violations were found. Sphericity was violated at the mid-calf location and Huynh-Feldt
correction was used for analysis at this level. The Huynh-Feldt correction method is
recommended where Epsilon is greater than 0.75 (MedCalc software, Ostend, Belgium). For
the mid-calf location in this study Epsilon was 0.943.

D’Agostino – Pearson tests for normal distribution rejected normality at all four
locations due to the positive skewness of all four distributions. ANOVA however is robust
against moderate deviations from the normal distribution, that is both the skew and kurtosis
should be less than +/- 1.0. In this study the skew and kurtosis at all four locations was less
than +/- 1.0 indicating minor deviations from the normal distribution and the analysis was
conducted (see Table 12). Results were considered significant where p=<0.05.
The results of the ANOVA for each location are displayed graphically in Figures 19 - 26. No significant variation in measurements was found between examiners at any of the locations. This is displayed in the figures below which show that at the average ankle measurement for all examiners ranged between 22.5cm and 23.1cm (Figure 19). For calf measurements the average ranged between 36.4 cm and 37.8 cm (Figure 21), the average mid-calf measurement was between 32.9 cm and 34.8 cm (Figure 23), and for the thigh location the average measurement was between 51.1 cm and 54.2 cm (Figure 25).

The effect of time on measurements was found to be significant at the Mid-calf level (p=0.002), and the calf level (p<=0.001) thus indicating poor repeatability of these measurements. Consistency across the three time points by each examiner would be indicated by a horizontal line. At the ankle location differences can be seen in the individual

<table>
<thead>
<tr>
<th>Location</th>
<th>Group</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Examiner</td>
<td>8.407</td>
<td>7</td>
<td>1.201</td>
<td>0.19</td>
<td>0.986</td>
</tr>
<tr>
<td></td>
<td>Time</td>
<td>0.575</td>
<td>2</td>
<td>0.288</td>
<td>2.90</td>
<td>0.057</td>
</tr>
<tr>
<td></td>
<td>Interaction</td>
<td>2.036</td>
<td>14</td>
<td>0.145</td>
<td>1.47</td>
<td>0.125</td>
</tr>
<tr>
<td></td>
<td>Residual</td>
<td>23.83</td>
<td>240</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Examiner</td>
<td>43.936</td>
<td>7</td>
<td>6.277</td>
<td>0.19</td>
<td>0.986</td>
</tr>
<tr>
<td></td>
<td>Time</td>
<td>6.836</td>
<td>2</td>
<td>3.418</td>
<td>12.32</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td></td>
<td>Interaction</td>
<td>5.376</td>
<td>14</td>
<td>0.384</td>
<td>1.38</td>
<td>0.162</td>
</tr>
<tr>
<td></td>
<td>Residual</td>
<td>66.606</td>
<td>240</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Examiner</td>
<td>124.56</td>
<td>7</td>
<td>17.79</td>
<td>0.41</td>
<td>0.894</td>
</tr>
<tr>
<td></td>
<td>Time</td>
<td>3.921</td>
<td>1.885</td>
<td>2.080</td>
<td>6.39</td>
<td>0.002*</td>
</tr>
<tr>
<td></td>
<td>Interaction</td>
<td>7.563</td>
<td>13.197</td>
<td>0.573</td>
<td>1.76</td>
<td>0.050*</td>
</tr>
<tr>
<td></td>
<td>Residual</td>
<td>73.644</td>
<td>226.229</td>
<td>0.326</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Examiner</td>
<td>201.875</td>
<td>7</td>
<td>28.839</td>
<td>0.19</td>
<td>0.988</td>
</tr>
<tr>
<td></td>
<td>Time</td>
<td>1.962</td>
<td>2</td>
<td>0.981</td>
<td>1.14</td>
<td>0.321</td>
</tr>
<tr>
<td></td>
<td>Interaction</td>
<td>8.362</td>
<td>14</td>
<td>0.597</td>
<td>0.70</td>
<td>0.778</td>
</tr>
<tr>
<td></td>
<td>Residual</td>
<td>206.052</td>
<td>240</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. SS: Sum of Squares, df: Degrees of Freedom, MS: Mean Square, F: F-ratio
*p<0.05
measurements by examiners however the average score did not change across the three time periods at time 1, time 2 or time three (see Figure 19). The average score for the calf location significantly decreased from time one to time two but this was followed by a minimal change from time two to time three (see Figure 21).

At the mid-calf location the average score changed substantially from time one to time three but the changes from time one to time two and from time two to time three were minimal (see Figure 23). Lastly at the thigh location the average score decreased from time one to time two but this was offset by increase in the average score from time two to time three (see Figure 25).

The most variability in the measurements between examiners across the time points appears to be at the mid-calf level which can be seen in Figure 24. However, time by group membership interaction was not found to be significant at any of the measurement locations (see Table 12). From the data displayed in Figures 20, 22, 24, and 26 a distinct pattern emerged where some examiners were found to consistently measure higher or lower than others. Examiner one measured the highest across all four locations, followed by high measurements also noted from examiners eight, four and two. Conversely examiners three, five and six were found to produce lower measurements.
Figure 19. Distribution of mean measurements by time point at ankle location (n=348)

Figure 20. Mean measurements by time point for each examiner at ankle location
Figure 21. Distribution of mean measurements by time point at calf location (n=348)

Figure 22. Mean Measurements by Time point for each Examiner at Calf Location
Figure 23. Distribution of Measurements by Time point at Mid-calf location (n=348)

Figure 24. Mean measurements by Time point for each Examiner at the Mid-calf Location
Figure 25. Distribution of Measurements by Time point at the Thigh Location (n=348)

Figure 26. Mean Measurement by Time point for each Examiner at the Thigh Location
4.4 Post hoc analysis of Confidence Intervals

The use of traditional methods of determining sample sizes for studies that involve power calculations have been deemed to be irrelevant in the case of agreement studies as seen in the current study. Proponents of agreement studies argue that it is not the size of the effect that is important but rather the precision with which the effect is determined (Bland, 2010). In the case of a repeatability study the precision with which the within subject standard deviation is estimated is dependent upon the sample size and the number of repeated measures performed (Bland, 2010).

When sample size is examined, it is usually determined a priori using results of similar studies. As no similar studies were identified, a post hoc analysis using data from the current study was conducted to determine the precision of the confidence intervals for phases II and III. The precision of the estimate of the within subject standard deviation was calculated for Phases II and III using the formula outlined by Bland (2010). For phase II there were 32 subjects with two measurements performed on each subject resulting in a confidence interval 25% either side of the within subject standard deviation. In Phase, III for 16 subjects with three repeated measures on each subject a confidence interval of 25% either side of the within subject standard deviation was also found.

Ideally confidence intervals should be narrow, +/-10% is usually acceptable, so as to avoid the potential of finding agreement where a larger sample would not. Disappointingly these results show a lack of precision in the estimates for the current study, however it does provide useful information for future studies of a similar nature in regards to study design and sample size planning. For example, the precision in phase II could have been increased simply by performing two measurements with each tape on each subject. This would have resulted in a 14% CI +/- the within subject standard deviation.
4.5 Summary

Presented in this chapter were the results of statistical analysis of the data collected for the three phases of this study. The stability of 20 Birch tape measures was tested in phase I in order to assess if the Birch tape measure would stretch over time. Birch tape measures that had been in use were found to stretch between 1 and 7mm. However, the amount of stretch was not found to be dependent on the length of time the tape measure had been in use.

For phase two the limits of agreement method was used to assess the level of agreement between the Birch tape measure and the Gulick spring tape measure. Poor agreement between the two tapes was found at both the calf and thigh level with the limits of agreement at the calf: -3.66 cm to 1.66 cm and thigh: -6.19 to 2.79 cm well outside the +/-1.5cm set a priori. Poor reproducibility of measurements was found to account for some of the lack of agreement between the two tape measures. Reproducibility coefficients were calculated for both tape measures (Birch +/- 2.64cm and Gulick +/- 2.84 cm) based on calf measurements. At the thigh level reproducibility was worse with the coefficients for Birch tape +/- 4.15 cm and for the Gulick tape +/- 4.65 cm. At both levels reproducibility was better with the Birch tape than the Gulick spring tape.

Phase three assessed the consistency of measurements performed at four separate locations, the ankle, mid-calf, calf and thigh across three time points. Relative standard deviations calculated from the first time point (baseline) indicated that calf measurements were the most consistently performed followed by the ankle measurements. Mid-calf and thigh measurements were found to be the least consistent.

Two way repeated measures analysis of variance were performed for all four locations to assess inter-rater reliability. No significant variation in measurement was found between examiners at any of the locations. The effect of time by group membership was found to be approaching significance at the mid-calf level only (p=0.05). A significant effect of time on
measurements was found at the mid-calf level (p=0.002) and the calf level (p=<0.001) indicating poor repeatability of measurements at these locations.

A discussion of these findings in the context of the literature will be presented in chapter five. The strengths and limitations of the study as well as the implications for practice, education and future research will also be explored in chapter five.
5.0 Discussion

Presented in this chapter is a discussion of the results presented in the previous chapter in relation to what is already known about the practice of calf and thigh circumference measurements. It outlines how the findings of each of the three phases fits into current clinical practice whilst exploring possible reasons for differences between this study and previous ones. The strengths and limitations of the study will be discussed as well as possible implications for practice, education and future research into this topic.

This study involved an examination of the reliability and reproducibility of calf and thigh circumference measurements used to detect oedema as an indication of deep vein thrombosis (DVT). Calf and thigh circumference measurements are performed daily in many spinal injury units as a method of detecting DVT and is a practice endorsed by the Consortium for Spinal Cord Medicine (Green et al., 2002), a leading body in the field of spinal cord injury medicine. Despite the wealth of literature where an increase in calf circumferences as a sign of DVT is reported (Diamond & Macciocchi, 1997; Goodacre et al., 2005; Haeger, 1969), there have been few studies that have examined the reliability of the measurement process in detail (Bunce, 2009; Swarczinski & Dijkers, 1991).

Daily lower limb circumference measurements are a simple, non-invasive and inexpensive method of DVT surveillance in this high risk group of patients, however the practice is meaningless if reliability and reproducibility of the measurements are poor. In this study three aspects of reliability were explored; the stability of the equipment, the equivalence of the equipment and the equivalence of the measuring procedure performed by nursing staff, currently being used in one spinal injury unit.

5.1 Stretch of Tape Measure
The stability of the equipment was assessed by examining if the Birch tape measure currently used in clinical practice would stretch over time, and if so, by how much. In a
comparison of 20 Birch tape measures it was found that the tapes stretched between 0-7 mm with a median of 2 mm over a period of 5-86 days. During this time the tapes are likely to have been used between 1 and 86 times, however the amount of stretch did not correlate with the period of time over which the tape measure was used. That is, those Birch tape measures that had been in use for the longest duration were not the ones that stretched the most. The stretch of the tape measure was found to be a random occurrence with the tapes used in this study.

Previous studies examining the measurement of calf circumferences have raised the possibility that standard dress making tape measures like the Birch tape may be unreliable due to the possibility that they stretch over time (Swarzinski & Dijkers, 1991; Wacogne, 2005). A search of the literature, however, failed to find any studies that had examined if tape measures do stretch over time. This is important as in clinical practice, changes in measurement of 1.5cm or greater are reportable and may prompt further investigation such as Doppler ultrasound which attract a considerable cost to the health care system. Therefore a reliable measuring device is essential.

The amount of stretch in the tape measure recorded was not found to correlate with the length of time in use. The same amount of stretch was recorded from the two tapes that had been in use the longest and the shortest number of days (1mm). It is possible that some other factors potentially play a role in the stretching of the tape, for example the amount of tension applied to the tape by the user, or the material from which the tape is made. The Birch tape in this study was made from fibreglass, and yet other studies do not always report what the tape measures are made from. The stretch in length of the Birch tape measure in this study was not dependent on the time the tape measure was in use, but the fact that some tapes stretch will affect the readings of measurements, thus affecting the precision of the equipment. This has the potential to prompt further investigations for the patient, but the risk of this may be
considered minimal if the same tape is used for each patient and stretching is minimal from one measurement to the next.

The median amount of stretch of 2 mm found in this study is unlikely to affect usual practice and investigation of the patient in the clinical setting, though the maximum reading of 7 mm could potentially lead to a false positive indication in clinical practice. In other words the patient could be thought to have a DVT when they do not and they may be subjected to unnecessary investigations to rule out a DVT. Given the small numbers in this phase of the study it is suggested further study into the precision of the Birch tape is warranted.

One of the limitations of this phase of the study was that it was not possible to identify if the tape that was examined from each of the patients was in fact the same tape that had been consistently in use since admission. Given that tape measures are considered a disposable item there is often complacency about where the item is stored, making it easy to misplace. Typically in the SIU the tape measure will be wrapped around the foot of the bed and in fact this is where all the tape measures in this study were located. Thus it was assumed that the tape measure had been in use for the full duration of the patients’ stay.

Ideally to examine stretch over time in use, new tape measures would have been labelled and placed in circulation on the same date and then followed up at pre-determined intervals. This would have improved the chance that the same tape measure placed in use was being assessed at follow up. It must also be noted that whilst stretch over time has been raised as an issue in regard to the standard tape measures such as the Birch tape, this has not been considered to be an issue with the Gulick spring tape measure (Fahey, 1989; Swarczinski & Dijkers, 1991; Wacogne, 2005). The Gulick spring tape is made from a flexible vinyl that the makers advertise will not stretch, though no studies could be identified to verify these claims. A comparison of the two tapes in a setting where the Gulick spring tape is routinely used in practice would be beneficial to this debate.
5.2 Agreement between tapes

The equivalence of the equipment currently used in the PAH SIU was tested by comparing the measurements taken with the Birch tape and the Gulick spring tape, in order to determine if there was a difference between the measurements produced by the two tapes. A high level of reliability in circumference measurements has been found in previous studies using spring tape measures and it has been suggested that the use of the standard tape measure may be responsible for inconsistent measurements in other studies (Bunce, 2009; Swarczinski & Dijkers, 1991). However, a study published after the commencement of the current research has contradicted these findings. In the study of short term and long term leg circumference measurements using a standard tape measure on healthy volunteers, measurements taken using the standard tape measure were found to be both reliable and reproducible (Te Slaa et al., 2011).

In the current study reproducibility of measurements between examiners was found to be poor with both tapes at both locations. Examiners were able to reproduce measurements only slightly better with the Birch tape at the calf location (Birch = 2.64cm, Gulick = 2.84cm) and the thigh location (Birch = 4.18cm, Gulick = 4.65cm). The work by Te Slaa et al. (2011) is the only other study to report reproducibility of measurements using a standard tape measure. They reported a reproducibility index of 4.4% for calf measurements taken in the short term and 6.5% for long term measurements however the short term reproducibility index falls to 3.3% if only one observer is performing the measurements.

Based on the mean calf measurement of the current study of 36.6cm, the calf reproducibility measurement would be +/-1.61cm for short term compared with the current finding of +/-2.64cm for the Birch tape. If only one observer were used, reproducibility would improve to +/- 1.21cm. In Te Slaa et al’s (2011) study only four observers were used compared with the 16 examiners used in the current study, suggesting that reproducibility
declines in comparison to the number of examiners used. Comparison of reproducibility of the spring tape measurements is difficult as previously published studies using this tape measure do not report reproducibility results.

The use of the limits of agreement analyses are now considered the gold standard in assessing the level of agreement between two devices (Alanen, 2012; Germing et al.; Ludbrook, 2010). The limits of agreement analysis are useful in comparing the amount of error in the measurements produced by two devices in order to determine the potential causes of the error. That is, whether the error is random or systematic. No previous study that had used this analytical method to compare circumference measurements using different tape measures could be found in the literature. Studies have instead focused on the use of correlational analysis for this purpose (Labs et al., 2000).

Bland and Altman (1999) have strongly cautioned that correlational analysis is not the most appropriate method of determining the level of agreement between two devices or methods of measurement. The purpose of the correlation coefficient is to measure the linear association of two items (Rankin & Stokes, 1998). Bland and Altman (1999) argue that two devices designed to measure the same object should in fact correlate, but this does not necessarily mean that they agree (Connelly, 2008). As described by Connelly (2008) one tape measure may consistently measure an object half the size of the other tape measure. This would show high correlation though the two tape measures would not agree (Germing et al.). It is more important to determine how well they agree in the measurement found (Bunce, 2009; Connelly, 2008).

The limits of agreement method uses bias as an indication of how well two measurements taken with the devices agree. The closer the bias is to zero the better the agreement. In this study there was a clear negative bias at both the calf (-0.99cm) and thigh level (-1.70cm) with the measurements taken by the Gulick spring tape. This indicated that if
the measurements taken with the Gulick tape measure were considered the true measurement then the Birch tape measure overestimated calf measurements by 0.99cm and thigh measurements by 1.7cm. A possible reason for this bias is related to the varying amounts of tension applied by examiners as described by the volunteers whose legs were being measured. In a previous study the Gulick spring tape was thought to be able to provide better results by eliminating error associated with tension (Bunce, 2009) but this was not found to be the case in the current study.

The purpose of the Bland Altman plot is to determine if the two devices agree sufficiently that the new measuring device can replace the other device. In the case of the current study, it was assessed as to whether the Birch tape measure can be used in place of the Gulick spring tape which is considered the “gold standard” device. Bland and Altman (1986) have specified that in order for two devices to show good agreement, 95% of the measurement differences should fall between the mean +/- 2 standard deviations. This is known as the limits of agreement. All clinically significant measurement differences should also be outside the limits of agreement (Bland & Altman, 1986). The clinically significant limit should be set á priori, that is, prior to the commencement of the study. This occurred in the current study with the clinically significant level set at +/- 1.5cm based on usual clinical practice.

In the current study all of the calf and thigh measurement differences between the two tapes fell within the limits of agreement. However, the limits of agreement for both the calf and thigh level were both well outside the 1.5cm limit set á priori, indicating poor agreement between the tape measures at both locations.

At the calf location the limits of agreement indicated that the Birch tape measures between -1.66cm (CI = -1.43cm to -1.89cm) and +3.66cm (CI = 3.89cm to 3.43cm) greater than the Gulick tape measure 95% of the time. The narrow confidence intervals are reflective
of the large sample size used, but they illustrate the extent of the discrepancies between the two tapes and that the level of agreement would not be acceptable for clinical practice.

Agreement at the thigh location was less accurate with the limits of agreement showing that the Birch tape measure would measure between 2.79cm (CI =2.61cm to 2.97cm) less and 6.19cm (CI =-6.01 to -6.37cm) greater than the Gulick tape measure. This poor agreement between the tapes may indicate that the Birch tape measure be abandoned in favour of the Gulick tape measure in clinical practice as the Gulick tape is considered the gold standard. However Bland and Altman (1986) caution that the poor agreement may be attributable to poor repeatability and reproducibility of one or both of the methods (Bland & Altman, 2010). Though repeatability was unable to be assessed in the current study, reproducibility of the individual tape measures was examined.

At the calf level the measurements were more reproducible with the Birch tape measure than the Gulick spring tape (Birch sw= 0.95cm, Gulick sw= 1.02cm). That is, measurements taken by one examiner on one individual were able to be reproduced by another examiner on the same individual using the same tape measure. Though the within subject standard deviation (sw) was high for both tapes, the Birch tape reproducibility coefficient (+/-2.64cm) was slightly better than the Gulick tape (+/-2.84cm) at the calf level. This indicates that the examiners were slightly better at reproducing calf measurements using the Birch tape than with the Gulick tape.

Significantly though, the reproducibility coefficients for both tape measures fall, at least partly, within the limits of agreement for the calf location (1.66cm to -3.66cm). This indicates that reproducibility may account for at least some of the poor agreement between the two tape measures, though there is likely to also be some other factor lowering the agreement between the two tapes at this location. One such factor could be the repeatability of measurements by examiners. Repeatability is measured by having the same examiner
perform multiple measurements on the same individual. Repeatability was not examined in this study.

At the thigh level the within subject standard deviations for both tape measures were greater than the 1.5cm difference set \textit{á priori} (Birch sw =1.52cm, Gulick sw =1.69), with the reproducibility coefficients calculated as +/- 4.18cm for Birch and +/- 4.65cm for the Gulick spring tape indicating poor reproducibility by examiners with both tape measures. In other words, the measurements taken by one examiner were not able to be reproduced by another examiner. This would mean measurements with either tape measure taken at this site would warrant further investigation in clinical practice. Again the reproducibility coefficients were partly within the limits of agreement (-6.19cm to 2.79cm) indicating that poor reproducibility was at least a contributing factor to the poor agreement between tape measures at the thigh location.

Measurements taken using the Birch tape were more reproducible at both the calf and thigh locations than those measurements taken with the Gulick spring tape as evidenced by the within subject standard deviations (Calf: Birch sw =0.95cm, Gulick sw =1.02cm; Thigh: Birch sw = 1.52cm, Gulick sw =1.69cm). Despite examiners being able to reproduce measurements better with the Birch tape, reproducibility of measurements at both locations with both tapes was poor overall. The tensioning of the tape measure and the familiarity with the tape measure are possible reasons for poor reproducibility with the tapes in this study.

Although not specifically asked, many volunteers whose legs were measured in the study commented that they could feel a considerable amount of difference in the tension applied to the Birch tape between examiners. Examiners were instructed to apply the amount of tension they would normally use when conducting measurements in usual practice. The Gulick tape with its spring attachment is thought to reduce the error in measurements associated with tensioning thus improving reproducibility (Bunce, 2009). In the current study
however, neither tape measure was able to display good reproducibility. In usual clinical practice there is the possibility that a different nurse will perform the measurements each day. If measurements are not reliably reproducible then a patient may undergo needless diagnostic procedures due to an inaccurate clinical picture or worse, a DVT may go undetected.

Prior to the study none of the examiners had seen or used the Gulick spring tape. Although examiners were instructed on its use and given the opportunity to practice, many commented on its difficulty in use. The main comment was that the tension mark on the spring was not very clear and difficult to see at a glance. Many felt the device would be impractical for everyday use in the busy ward environment as it was time consuming to ensure an accurate reading.

5.3 Inter rater reliability

The examination of the equivalence of the procedure for DVT detection involved looking at the reliability of the locations being measured and the inter rater reliability of the examiners performing the measurements using the Birch tape measure. This study focused on measurements at four specific locations; the thigh, the calf, mid-calf and ankle. The aim was to determine if one particular location performed better in terms of repeatability and inter-rater reliability.

Based on relative standard deviations of measurements taken at the four locations, the calf measurement was found to be the most reliable followed by the ankle and thigh locations. The mid-calf location was the least reliable location measured.

Inter-rater reliability using the Birch tape was shown to be good with no significant variation in measurements between examiners at any of the locations measured. However, inter-rater reliability over time was found to be poor at both the calf and mid-calf levels. High inter rater reliability of circumference measurements has previously been found in some studies (Bunce, 2009; Labs et al., 2000; Tunc et al., 2007; Whitney et al., 1995). This
contradicts many of the findings of the current study. These disparities can be explained by the fact that previous studies varied in design, ranging from the use of one examiner with 30 subjects to 35 examiners and one subject.

The experience level of the examiners also varied from the use of novices in a study reporting poor inter-rater reliability (Maylia, Fairclough, Nokes, & Jones, 1999) to experienced medical doctors and physiotherapists reporting high inter-rater reliability (Tunc et al., 2007; Whitney et al., 1995). A common recommendation from previous studies is that to ensure accuracy of the measurements there should be consistency in the clinician taking the measurements, meaning the one clinician should perform all measurements on the one patient (Bunce, 2009). It should be highlighted that in most cases the studies in question have examined the use of tape measures for circumference measures to assess atrophy and hypertrophy of muscles, rather than for calf swelling associated with DVT (Bunce, 2009; Maylia et al., 1999; Whitney et al., 1995). Assessment of atrophy and hypertrophy of muscles would undoubtedly occur on a less frequent basis in clinical practice than monitoring for DVT making the use of the same examiner more feasible (Bunce, 2009).

In a retrospective study examining the records of 30 acute care spinal cord injured patients with an average stay of 31 days where nursing staff performed the measurements, it was found that on average 15 different clinicians had taken measurements for each patient (Swarzinski & Dijkers, 1991). In a clinical ward environment where nursing staff work rotating shifts it is not possible, nor is it feasible from a practical point of view, to have the same person undertake calf and thigh measurements on a daily basis for a particular patient. The use of strict protocols and education and training though may minimise this concern on a day to day basis.
5.4 Repeatability by Location

The decision of what part of the leg to measure for detection of DVT is one that has been widely examined in the literature though with little consensus (Bunce, 2009; Tunc et al., 2007). Many authors (Constans et al., 2003; Diamond & Macciocchi, 1997; Johanning et al., 2002; Stein et al., 1995; Swarczinski & Dijkers, 1991) refer to calf swelling as a symptom of DVT though few explain how they assess this in regards to the region of the calf to be assessed and what landmarks are used in identifying the location to be measured. The lack of agreement by authors (Johanning et al., 2002; Stein et al., 1995; Swarczinski & Dijkers, 1991) regarding the exact location to measure may account for some of the reported inconsistency in circumference measurements reported in other studies. In the current study precise landmarks were identified as the hip, the patella, the tibial tuberosity, the ankle and the medial malleolus.

The thigh measurements were taken at the midpoint between the proximal edge of the patella and the hip. The calf measurements were taken at the midpoint between the distal edge of the patella and the ankle. The mid-calf measurement was performed at the midpoint between the medial malleolus and the tibial tuberosity, whilst the ankle measurement was taken 2cm proximal to the medial malleolus. The aim of this phase of the study was to determine if one particular location performed better in terms of repeatability and inter-rater reliability.

Thigh circumference measurements are considered by many to be notoriously difficult to assess and are often deemed inaccurate with poor repeatability, though few studies have tested this assumption (Maylia et al., 1999). In the current study there was no statistically significant effect of time on measurements at the thigh level indicating good repeatability over time by examiners at this landmark site. This is in contrast to an earlier study that showed poor repeatability at the thigh level, though in that study the investigators had no
prior experience using the tape measure (Maylia et al., 1999). Positioning and tension applied to the tape have been suggested to be the main reasons for the lack of repeatability (Maylia et al., 1999).

This study used pre-positioned tape markings to control the measurement site between examiners and thus the lack of consistency is presumed to be related to the tension applied by the examiner and not the placement of the tape measure. The finding in this study of poor inter-rater reliability in measurements taken at the thigh location contradicts the findings of Whitney et al (1995), but is consistent with most other studies (Lohman, Roache, & Martorell, 1992; Maylia et al., 1999; Wacogne, 2005; Whitney et al., 1995).

It is suggested in the literature that when a distal DVT occurs the swelling is first noticeable near the ankle before moving up into the calf (Cranley et al., 1976; Fahey, 1989). For this reason a location 2cm proximal to the medial malleolus was included as a measurement site in this study. Results of this study show that ankle measurements are both reliable and repeatable. This is evidenced by low relative standard deviations for individual examiners and no significant difference in measurements noted between examiners (p=0.986). Just one previous study, published after commencement of the current study, adequately described ankle circumference measurements (Smith, Ma, & Stafford). In a study of oedema assessment methods high inter-examiner agreement (ICC 0.96 right and 0.97 left) and high intra-examiner agreement (ICC 0.98 to 1.0) was found for ankle circumference measurements. Further investigation would need to be conducted examining the sensitivity and specificity of ankle oedema in the diagnosis of DVT prior to adopting this practice in the clinical environment.

The results of this study show that measurements taken at the calf level are consistent when based on just the baseline measurement. For over half of the volunteers measured, the calf location had the lowest relative standard deviation of all four locations. When assessing
repeatability across all three time points; baseline, baseline +15minutes and baseline +30minutes, there was a statistically significant effect of time on measurement at the calf level (p=<0.001) indicating poor repeatability over time. Previously poor repeatability of calf measurements using the patella as a landmark was thought to be due to movement in the patella (Tunc et al., 2007). This meant there was a chance that measurements would not be consistently taken in the same place each time. In this study the movement of the patella was controlled for by the white tape marker, meaning the examiner was not marking out the location each time. Therefore some other factor must be responsible for the poor repeatability over time.

The mid-calf location which measures the circumference of the midpoint between the tibial tuberosity and the medial malleolus was the worst performer of all four locations. This location was the least consistent with regards to rsd and showed poor repeatability over time (p=0.002). These results are in contrast to Tunc et al (2007) who found better intra-observer agreement of calf measurements based on the tibial tuberosity as a reference point (88%) when compared to those based on the patella (65%) as a reference point (Tunc et al., 2007). One reason for the poor repeatability in this study is anatomical form. The mid-calf location on many people is where the bulk of the calf muscle begins to taper off. This may result in difficulty in having the tape lay flat against the leg whilst performing the measurement. This is supported by Geil (2005) who recommends that where measurements are conducted in an area where the limb tapers the proximal edge of the tape should be used for the measurement as this maintains closer contact with the skin.

5.6 Strengths and Limitations
The use of lower limb circumference measurements for the diagnosis of DVT has been reported in publications since the 1960s, however to date, few studies have assessed the reliability of the technique. Despite this lack of evidence, calf circumference measurements
continue to be endorsed as a method of surveillance for DVT in the highly at risk SCI patient population. Based on this recommendation performing calf and thigh circumference measurements has been a routine daily activity undertaken by the nursing staff within the spinal injuries unit at the site for more than ten years. Until now this task was undertaken without consideration of the evidence base behind its practice or the reliability of the measurements obtained. This study has provided a thorough analysis of the current procedure in place within one spinal unit. It has also been a comprehensive exploration of the issues surrounding the reliability of circumference measurements adding to the evidence base on this topic.

The long held belief that tape measures stretch with prolonged use was explored in this study for the first time. Though the number of tape measures examined for stretch over time in this study was small, no other study to date has attempted to explore this issue which is frequently suggested in the literature as a source of error in measurements (Swarzcinski & Dijkers, 1991; Wacogne, 2005). The findings of this study, which showed that the stretch in tape measures was not dependent on the time in use, provided contrary evidence to the theories previously reported in the literature and certainly warrant further investigation.

The reliability of ankle circumference measurements was also investigated in this study. To date, swelling of the ankle as an indicator of DVT in the spinal cord injured population has not been investigated and there is little mention in other patient groups. The results of this study indicate that ankle circumference measurements are repeatable and reproducible. As studies (Church, 2000; Fahey, 2004) suggest a relationship between ankle swelling and the presence of DVT and the ankle circumference measurements are reliable, there is the need for future studies to investigate the sensitivity and specificity of ankle swelling in the diagnosis of DVT.
The design of this study was inherently larger in size and scope compared with previous studies examining this topic, but this was necessary in order to fully examine the various components of the procedure. This was the first study to use the limits of agreement method for analysis in comparing tape measures. Only through the use of the limits of agreement could it be shown that the two tape measures were unable to be used interchangeably in clinical practice. The use of alternative methods of analysis in other studies, however, make meaningful comparisons of these results difficult. The number of participants in this study was larger than other studies examining lower limb circumference measurements. This increased number of examiners was considered appropriate because in usual practice measurements are undertaken by multiple clinicians on the one patient. This study allowed for a more meaningful analysis of inter-rater reliability than previously offered by studies which often only used two examiners.

There were several limitations identified in the conduct of this study. The current study was carried out in a single clinical area where calf and thigh circumference measurements are performed daily for a specific purpose of DVT surveillance and as such the results of this study are not necessarily generalisable to other clinical areas.

Bland and Altman state that if one or both of the devices has problems with repeatability then the agreement between the two devices will be poor (Connelly, 2008). It must therefore be a consideration that some of the disagreement between devices be attributable to poor repeatability and reproducibility of measurements by one or both of the tape measures. The within subject standard deviations suggest that measurements with the Gulick spring tape are less reproducible than with the Birch tape, possibly due to the fact that the examiners were less familiar with the Gulick spring tape. The failure to examine repeatability of measurements using the Gulick spring tape in this study must be accepted as a limitation. Whilst repeatability of the Birch tape was shown to be good, it is impossible to
say if repeatability with the Gulick spring tape may have influenced agreement between the two tape measures.

The issue of missing data in this study should be considered a limitation in that to complete the ANOVA analysis data were imputed. The mean imputation method was used in this case and whilst considered appropriate for this study has some disadvantages. These disadvantages include: the new mean that is generated may be an underestimate of the true value, the distribution of values may be distorted and correlations may be depressed due to the repetition of a single repeated value (Bland & Altman, 2003).

Another study addressing these limitations may be necessary to reach firm conclusions about the usefulness and reliability of calf and thigh circumference measurements in the spinal cord injured population. Certainly a repeated study using these same methods but in a different setting would help to bring about recommendations for a set of guidelines to ensure consistency in the procedure and reduce some of the error associated with measuring leg circumferences.

5.7 Implications for Practice

The aim of the current study was to examine the reliability of a standard tape measure in use, the inter-rater reliability of the measurements, and the reliability of the locations being measured. The results of this study have several implications for the current practice of calf and thigh circumference measurements with spinal injured patients.

In examining the reliability of the tape measure, the focus of this study was on the stability of the equipment over time and the reproducibility of measurements. Stability of the Birch tape measure currently in use was assessed by examining whether the tape would stretch over time in use. Given that there was no relationship between the stretch of the Birch tape measure and the length of time it was in use, there is no indication that regular changing of the tape throughout the period of use would result in more accurate results.
The Birch tape was also compared to the Gulick spring tape which is considered to be the gold standard tape measure. The results of this study indicated that reproducibility of measurements with both tapes was poor though it was slightly better with the Birch tape. Despite being referred to as the gold standard in tape measures, this study has shown that the more expensive Gulick spring tape does not result in better reproducibility of measurements. Therefore the recommendation would be to continue the use of the Birch tape measure for circumference measurements.

The literature suggests that calf and thigh circumference measurement as a practice of detection of DVT has been abandoned by many due to poor inter-rater reliability. This study examined inter-rater reliability at four separate locations, thigh, calf, mid-calf and ankle. Inter-rater reliability was found to be generally good across all four locations. However, it should be noted that upon closer examination of each location the calf location currently used in practice had the least variation in measurements between examiners. This was followed by the ankle, mid-calf and then lastly the thigh location. The recommendation for clinical practice is to continue to undertake calf circumferences, using the midpoint between the base of the patella and the ankle, as the measurement site. To ensure consistency between measurements the site should be marked. The continued use of a thigh circumference measurement should be used with caution given the high variation in measurements between examiners. As the ankle measurement also showed low variation and high inter-rater reliability it suggested that further investigation into the use of this location on a routine basis be conducted.

**5.8 Implications for Education**

The practice of routine lower limb circumference measurements is a relatively simple, cheap and non-invasive method of DVT surveillance in the spinal cord injured patient. Taking the measurements accurately, however, is a multi-step process that involves many
variables. To ensure measurement accuracy; the measurement location must not be altered between measurements, the tension should be consistent between clinicians, the same tape measure should be used for the duration of the measurement period and there must be accuracy in the measurement recorded as well as follow-up if a reportable difference is noted.

The results of this study are based on having many of these variables controlled for the purpose of the research project. In clinical practice the impact of these variables may be reduced by increasing the level of education and skill of the clinician performing the measurements. There appears to be little consensus in the spinal injuries medicine field about the best way to conduct lower limb circumference measurements. The Consortium for Spinal Cord Medicine strongly recommend that circumference measurements be performed but do not provide detail about where on the leg the measurements should be done, what tape measure should be used or even what difference in measurements might be clinically significant (Green et al., 2002). The original guidelines were published in 1999 and no updates have been provided since 2002. Anecdotally it is known that regular calf and thigh circumference measurements are not performed in all spinal units within Australia. With a lack of clinical research to support or refute the practice the decision to perform these measurements appears to be based on the medical officers’ preference.

Currently there is no formal policy for undertaking calf and thigh circumference measurements within the SIU or the greater site hospital. In fact this procedure is only routinely performed within the SIU. Spinal cord injured patients who are admitted to other wards such as trauma, orthopaedic and intensive care units do not routinely have calf and thigh measurements performed for DVT surveillance. A formal policy outlining the purpose and practice of lower limb circumference measurements is urgently needed if this practice is useful and is to continue and be adopted for spinal cord injured patients throughout the care continuum. Such a policy should include when to implement the measurements, position of
the patient, time of day to perform the procedure, the use of landmarks to find the location to
measure, how to record baseline measurements and the necessity to perform measurements
daily as part of the surveillance.

Most importantly, staff should be provided with education on what a clinically
significant change in measurement is and how to look for other signs and symptoms which
may support the change in measurement being related to a DVT. A policy would include the
mandatory notification of the change to a medical officer for follow-up assessment of the
patient.

All new staff should be provided training on undertaking the procedure during
orientation to the ward and ongoing refresher training could be provided with a mandatory
yearly VTE competency. The implementation of a formal policy and staff education is an
inexpensive and easy intervention that may increase inter and intra rater reliability in
measurements. It may also result in early detection of DVT thus reducing the chance for
secondary complications for the patient.

5.9 Implications for Further Research

The aims of this study have been fulfilled as shown by the results. However, the
results have also raised more questions which need to be addressed through further research
of this topic. Implications for further research include the need for a larger study examining
the issue of tape stretch, inter-rater reliability studies examining the repeated use of the
Gulick spring tape and a prospective study examining the specificity of ankle swelling as an
indication of a DVT.

The results of this study indicate there is no relationship between the length of time in
use and stretch of the tape measure, however this study only examined 20 tape measures
currently in use within one ward. It would appear from the lack of literature surrounding this
topic that this is the first study to examine stretch of tape measures over time. Though the
median 2mm stretch found in this sample would not necessarily impact upon clinical decision making, having a tape measure stretch 7mm could significantly affect the measurement results of a patient. A larger study is needed to confirm these results as well as test whether stretch occurs with other types of tape measures, such as the Gulick spring tape.

The current study examined the use of the Gulick spring tape though only in terms of its agreement with the Birch tape and reproducibility of measurements. Agreement between the two tapes was found to be poor in terms of clinical significance. The fact that reproducibility of measurements with both tapes was also poor no doubt impacts upon the poor level of agreement. The poor reproducibility with the Gulick spring tape indicates the less expensive tape measure is as good as the considerably more expensive Gulick tape. What was not examined was inter-rater reliability using the Gulick spring tape. A previous study indicated that inter-rater reliability using a spring tape was high; however that study involved different landmarks and only two testers, both of whom were experienced with using this tape measure.

Further research could be conducted using a similar design to the current study with multiple examiners across multiple time periods to examine both intra and inter-rater reliability of the Gulick spring tape measure. This would allow for a full comparison of results with the Birch tape measure at all of the locations.

Another area for further study is the validity of ankle circumference measurements in the surveillance of a DVT. This study has shown that ankle measurements have a high inter-rater reliability and low variance. What the current study was unable to address is whether an increase in ankle circumference measurements is a specific indicator of the presence of DVT. There is little literature examining the reliability of ankle measurements and although there is a suggestion by some authors that ankle swelling does occur early in DVT formation
(Church, 2000; Fahey, 2004), there is currently no literature examining the specificity of ankle oedema as an indicator of lower limb thrombosis.

The current study has been one of the largest studies examining the reliability of calf and thigh circumference measurements undertaken to date. Importantly, the use of limits of agreement to examine tape measure reliability is unique to this study and is now considered the most appropriate method of examining measurement method studies. The results indicate that examiners using an inexpensive standard dress making tape measure are able to reproduce measurements as effectively as the gold standard spring tape measure.

Importantly, this study has suggested that contrary to popular belief the dress makers tape (Birch tape) does not stretch over time. There was some indication that stretching of the tape measure occurs in clinical practice, however this did not correspond to the length of time the tape measure had been in use. Further studies examining other factors associated with stretching may account for this anomaly.

An examination of the reliability of locations used for circumference measurements showed good inter-rater reliability at all locations. However, the calf location showed the lowest relative standard deviations of the four locations indicating less variation in measurements between examiners making it the most reliable of all four locations.

The most significant finding of this study is that multiple examiners are able to produce reliable measurements in healthy volunteers using the current calf and thigh locations. The findings suggest that the current practice of lower limb circumference measurements in the SIU is a reliable one. The further studies suggested here can only add to the knowledge base and thereby enhance the processes already in place.
6.0 Conclusion

The early use of chemical and mechanical prophylaxis for the prevention of DVT is now standard care for patients with acute spinal cord injury. This practice has seen the incidence of DVT reported in the SCI literature fall from as high as 90% without prophylaxis to 10-30% with prophylaxis (Aito et al., 2002). Though significant reductions have been seen in the incidence rates in SCI patients, the secondary effects of DVT still remain prevalent, with VTE the third leading cause of death in the first year post injury for these patients.

The use of calf and thigh measurements performed daily for surveillance of DVT is recommended by one of the leading bodies of SCI research (Green et al., 2002). Early recognition of the symptoms of DVT followed by diagnosis and treatment remain the key to reducing the risk of a fatal PE or the life long consequences of PTS. Given the nature of spinal cord injury, many of the symptoms displayed by the normal population may be absent or misleading. Pain may be absent in those with motor complete injuries, and fever may be misleading in a patient with urinary tract infection. An increase in circumference measurement from baseline of the calf and/or thigh has been found to be a more reliable symptom in the SCI population (Green, 1999).

The daily circumference measurement of the lower limbs of patients with spinal cord injury is a practice that has continued for many years. Performed for the surveillance of DVT, the current procedure lacks a strong evidence base, thus leading to the current study. More recent anthropometric studies (Bunce, 2009; Maylia et al., 1999; Smith et al.; Te Slaa et al., 2011) however, have shown good inter and intra-rater reliability in the performance of calf and thigh circumference measurements, though many of these studies were limited either in the number of participants or the number of examiners. The aim of the current study was to examine the reliability of the procedure used for lower limb circumference measurements for
the detection of DVTs. With clinicians relying on the measurements to make informed clinical decisions about the need for further investigation, it is essential that the practice be undertaken according to the best evidence available.

Three aspects of reliability of the current procedure were examined in this three phase study; the stability of the equipment, the equivalence of the equipment and the equivalence of the procedure itself. In Phase I the stability of the equipment was examined by testing if the Birch tape measure currently used in this spinal unit would stretch with repeated use. The equivalence of the equipment was examined using the Bland Altman limits of agreement method to test the agreement between the Birch tape and the Gulick spring tape in phase II. The Gulick spring tape is considered by many to be a gold standard in tape measures. Phase III tested the equivalence of the current procedure by testing the inter-rater reliability and the reliability of the location on the limb being measured and comparing this to alternate locations identified in the literature.

In examining the stability of the Birch tape measure, 20 tape measures currently in use were tested against one brand new, unused, Birch tape measure. The length of time the tape measure had been in use was compared to the amount, in millimetres, the tape measure had stretched compared to the unused one. Of the 20 tape measures tested 17 were found to have stretched, (range 1-7mm). Though the median stretch of all the tape measures recorded was 2 mm, this was not found to be related to the length of time the tape had been in use. This finding suggests that repeated use of the Birch tape does not cause it to stretch as suggested in previous literature. Of interest however, was the finding that the tape measure does stretch as much as 7mm which may impact the reliability of measurements taken.

The equivalence of the Birch tape measure was tested by assessing the level of agreement between measurements taken with the Birch tape measure and those taken with the Gulick Spring tape. Sixteen examiners measured the calf and thigh circumferences of 32
volunteers’ legs resulting in 402 data sets. Good agreement between the Birch tape and the Gulick spring tape measure was found. However the limits of agreement were wider than the 1.5cm limit set a priori indicating that the two tape measures could not be used interchangeably. A strong bias of 0.99cm at the calf and 1.7cm at the thigh indicated that the Birch tape measure would consistently measure circumferences greater than the Gulick tape measure. However, the Birch tape showed better reproducibility of measurements than the Gulick tape in the current study, possibly because the examiners performing the measurements were more familiar with the Birch tape.

Phase III of the study examined the equivalence of the procedure by examining the locations being measured as well as the reliability of the examiners. A total of 116 sets of measurements were performed by eight examiners on 16 volunteers over three separate time points in this phase. In examining the equivalence of the current procedure the calf and thigh locations currently used, as well as an ankle measurement and alternate calf location (mid-calf) were tested. The variance of measurements by examiners between volunteers at each location was calculated in order to determine which location showed the least amount of variation. The most reliable location was found to be the calf location, followed by the ankle, thigh and then the mid-calf location was the least reliable.

Measurements by eight examiners on 16 volunteers across three time points each 15 minutes apart were examined to determine the equivalency of the examiners, otherwise known as inter-rater reliability. There was no significant difference in measurements between examiners at any of the four locations indicating good inter-rater reliability at all four locations. However the effect of time on measurements was found to be significant at the calf (p=<0.001) and mid-calf (p=<0.002) locations indicating significant variations in repeated measurements by examiners over time at these locations.
Two important findings from this study may impact upon the current practice of calf and thigh measurements for the detection of DVT. The first significant finding was that the Birch tape showed good stability over time and better reproducibility of measurements than the Gulick spring tape. This indicates that there is no need to change to a more expensive style of tape measure, and that the Birch tape does not need replacement throughout the period of use.

Secondly, the calf location currently used for measurements showed the least variation in measurements of the four locations measured when based on a single measurement. Importantly, the thigh location showed a significant variation in measurements by examiners indicating that this location should be used with caution. The finding of good inter-rater reliability and low variation in measurements of the ankle location indicates the need for further investigation into the specificity of ankle measurements in DVT diagnosis.

Highlighted in this study are several areas where further enquiry may be beneficial. No significant relationship was found between the stretching of the Birch tape measure and the time in use suggesting repeated use of the tape measure is acceptable in clinical practice. Confirmation of this result with a larger study, perhaps including other brands or types of tapes such as the Gulick spring tape would help to support this finding.

The use of Bland Altman’s limits of agreement in this study has highlighted the distinct differences between the Birch tape and the Gulick spring tape. The large bias found in this study, suggesting that measurements taken with the Birch tape will be larger than those taken with the Gulick spring tape, and the poor reproducibility by both tapes, suggest that repeatability could be poor with either one or both tape measures. The unfamiliarity with the Gulick spring tape by the examiners may have contributed to the results of this study. The finding of good inter-rater reliability with the Birch tape at four separate measurement locations suggests that despite these findings calf and thigh measurements with the Birch tape
are repeatable. The lack of the repeatability component in Phase II of this study must be accepted as a criticism. Future studies with both the Birch and Gulick spring tape at all four measurement locations should identify repeatability and reproducibility issues with both tape measures.

A thorough examination of many of the aspects of the practice of lower limb circumference measurements in place for DVT detection has been presented in the current study. This is the first study that has examined the stretch of tape measures calling into question the reliability of equipment currently in use. The use of the Bland Altman method to examine the agreement between two tape measures is unique to this study. It has offered a more comprehensive assessment of the measurement process and allowed for greater transparency of where the differences between the two tape measures may actually lie. The prospective nature of this study using nurses experienced in performing circumference measurements also offered a realistic insight into the current practice. The number of examiners and volunteers used in this study makes it one of the largest studies examining reliability of lower limb circumference measurements undertaken.

Without a reliable measure, the everyday practice of calf and thigh measurements is meaningless with the potential that the diagnosis of DVT in SCI patients is delayed or undetected. The use of lower limb circumference measurements for surveillance of symptoms of DVT remains one of the simplest, non-invasive and least expensive methods for ensuring early intervention and a reduction in the incidence of secondary effects of DVT. The results of this study will add to the existing knowledge base surrounding the reliability of the procedure. It will improve the practice within the spinal injuries unit by assisting in the refinement of the current procedure.
References


Bland, J. M. (2010, 17 May ). How can i decide the sample size for a repeatability study? *Frequently asked questions on the design and analysis of measurement studies.*


Wacogne, I. (2005). Using a tape measure for intra-observer variability. Evidence Based Medicine, 10(4), 1. doi: 10.1136/ebm.10.4.104


**Appendix A: Informational Flyer**

**How do we Measure Up!**

I am looking for QSCIS staff interested in participating in a research project titled:

*Examining the reliability of calf and thigh measurements for the diagnosis of deep vein thrombosis in spinal cord injured patients.*

I require people to act as both Healthy Volunteers and Examiners.

**Criteria Include:**

Healthy Volunteers **must not** have a history of lower limb surgery or vascular disease.

Examiners **must** be Registered Nurses with a minimum of 3 months experience in the Spinal Injuries Unit.

If you would like to participate in the research or require more information about the project please contact:

**Ellen Engarde**  
Clinical Nurse  
Spinal Injuries Unit  
Ph: (07) 3240 2737  
Email: Ellen_Engarde@health.qld.gov.au
Appendix B: Data collection Form Phase I

Data Collection Form Phase I

<table>
<thead>
<tr>
<th>Study ID Number</th>
<th>UR Number</th>
<th>Date of SIU Admission</th>
<th>Date of Examination</th>
<th>Stretch in Tape measure (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Appendix C: Data Collection Form Phase II

**Data Collection Form Phase II**

<table>
<thead>
<tr>
<th>Volunteer ID No.</th>
<th>Tape Position</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Leg Measured:</strong> Right/Left</td>
<td><strong>Calf:</strong> ____cm below patella</td>
</tr>
<tr>
<td></td>
<td><strong>Thigh:</strong> ____cm above patella</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Examiner ID No.</th>
<th>Test Date</th>
<th>Device Used</th>
<th>Calf</th>
<th>Thigh</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Spring/Birch</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Spring/Birch</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Spring/Birch</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Spring/Birch</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Spring/Birch</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Spring/Birch</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Spring/Birch</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Spring/Birch</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Spring/Birch</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Spring/Birch</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Spring/Birch</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Spring/Birch</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Spring/Birch</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Spring/Birch</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Spring/Birch</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Spring/Birch</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Spring/Birch</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Spring/Birch</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

133
Appendix D: Data Collection Form Phase III

<table>
<thead>
<tr>
<th>Data collection Form Phase III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test Date:</td>
</tr>
<tr>
<td>Test Time:</td>
</tr>
<tr>
<td>Examiner ID No.:______________</td>
</tr>
<tr>
<td>Volunteer ID No.:______________</td>
</tr>
</tbody>
</table>

Tape Position

- **Calf:** _________ cm below patella
- **Thigh:** _________ cm above patella
- **Ankle:** _________ cm above medial malleolus
- **Mid-calf:** _________ cm below tibial tuberosity

<table>
<thead>
<tr>
<th>Location</th>
<th>Baseline</th>
<th>&quot;+15 mins&quot;</th>
<th>&quot;+30 mins&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Right</td>
<td>Left</td>
<td>Right</td>
</tr>
<tr>
<td><strong>Calf</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Thigh</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Ankle</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mid-Calf</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix E: Tape Measures Used in Study

Birch tape

Gulick Spring Tape
Appendix F: Location where Circumference Measurements were performed

Thigh Measurement Location
Midpoint between the hip and proximal edge of the patella

Calf Measurement Location
Midpoint between the distal edge of the patella and the proximal edge of the ankle

Mid-Calf Measurement Location
Midpoint between the tibial tuberosity and the medial malleolus

Ankle Measurement Location
2cm above the medial malleolus
Appendix G: Office of the Human Research Ethics Committee PAH Approval Letter

Office of the Human Research Ethics Committee

Enquiries to: Ethics Manager
Phone: (07) 3240 7672
Fax: (07) 3240 7667
E-mail PAH_Ethics_Research@health.qld.gov.au
Date 26 November 2009

Ms Ellen Eugarde
Spinal Injuries Unit
Princess Alexandra Hospital
Woolloongabba
Queensland 4102

Dear Ms Eugarde

HREC Reference number: HREC/09/QPAH/281
Project title: Examining the reliability of calf and thigh measurements for the diagnosis of deep vein thrombosis in spinal cord injured patients

<table>
<thead>
<tr>
<th>Document</th>
<th>Version</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advertisement: Participant recruitment flyer</td>
<td>1</td>
<td>25 September 2009</td>
</tr>
<tr>
<td>Application for low-risk review</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Letter of invitation to participant: Participant recruitment letter</td>
<td>1</td>
<td>07 October 2009</td>
</tr>
<tr>
<td>Patient Information Sheet/Consent Form: Nurse Unit Manager ICF</td>
<td>1</td>
<td>07 October 2009</td>
</tr>
<tr>
<td>Patient Information Sheet/Consent Form: Examiner ICF</td>
<td>1</td>
<td>07 October 2009</td>
</tr>
<tr>
<td>Patient Information Sheet/Consent Form: Healthy Volunteer ICF</td>
<td>1</td>
<td>07 October 2009</td>
</tr>
</tbody>
</table>

At a meeting of the Metro South Health Service District Human Research Ethics Committee (MSHSD HREC) held on 2 November 2009, the Committee reviewed the above research Protocol. The Metro South Health Service District Human Research Ethics Committee is duly constituted, operates in accordance and complies with the current National Health and Medical Research Council’s National Statement on Ethical Conduct in Human Research 2007.

On the recommendation of the Human Research Ethics Committee approval is granted for your project to proceed. This approval is subject to researcher(s) compliance throughout the duration of the research with certain requirements as outlined in the National Statement on Ethical Conduct in Human Research 2007 and Australian Code for the Responsible Conduct of Research.

The following links have been provided for your convenience:
Some requirements are briefly outlined below. Please ensure that you communicate with the HREC on the following:

- **Protocol Changes**: Substantial changes made to the protocol require HREC approval.
- **Problems and SAEs**: The HREC must be informed of any problems that arise during the course of the study which may have ethical implications. Serious adverse events must be notified to the HREC as soon as possible.
- **Lapsed Approval**: If the study has not commenced within twelve months approval will lapse requiring resubmission of the study to the HREC.
- **Annual Reviews**: All studies are required by the NHMRC to be reviewed annually. To assist with reporting obligations an Annual Report template is available on the MSHSD HREC website. This form is required to be completed and returned to the HREC within the 12 month reviewing period.

As this research involves the recruitment of patients from the Metro South Health Service District (MSHSD), it is my responsibility to remind you of your ongoing duty of care for all people recruited into projects or clinical trials whilst public patients. All conditions and requirements regarding confidentiality of public information and patient privacy apply. You are required to comply at all times with any application requirements of Australian and Queensland Laws including the Health Services Act, the Privacy Act, Public Health Act (2005) and other relevant legislation, ethics obligations and guidelines which may be applicable to the MSHSD from time to time including, without limitation, any requirement in respect of the maintenance, preservation or destruction of patient records.

When the study involves patient contact, it is your responsibility as the principal investigator to notify the relevant consultant and request their approval.

Should you have any problems, please liaise directly with the Chair of the HREC early in the program.

A copy of this letter should be presented when required as official confirmation of the approval of the Metro South Health Service District Human Research Ethics Committee.

We wish you every success in undertaking this research.

Yours sincerely

Dr David Theile Snr

DISTRICT CHIEF EXECUTIVE OFFICER

METRO SOUTH

______/___ ___/______
Appendix H: Human Research Ethics Committee PAH Approval Letter for amendment to Examiner Informed Consent Form and Healthy Volunteer Informed Consent Form

Ms Ellen Ugarde
Clinical Nurse
Princess Alexandra Hospital
Ipswich Road
WOOLLOONGABBA 4102

Dear Ms Ugarde

Re: 2009/281


On the 23rd March 2010 the Chair of the Metro South Health Service District Human Research Ethics Committee, reviewed the following amendment for the above study and approval was granted:

<table>
<thead>
<tr>
<th>Document</th>
<th>Version</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Examiner Information &amp; Consent Form</td>
<td>1.1</td>
<td>05 March 2010</td>
</tr>
<tr>
<td>Healthy Volunteer Information &amp; Consent Form</td>
<td>1.1</td>
<td>05 March 2010</td>
</tr>
</tbody>
</table>

It should be noted that all requirements of the original approval still apply.

If you have any queries, please do not hesitate to contact the Metro South Health Service District Human Research Ethics Committee, Executive Support Officer on (07) 3176 7672.

Best wishes for the progress of the study.

Yours sincerely

Dr Jennifer Fleming
Chair
Human Research Ethics Committee
Metro South Health Service District
Centres for Health Research
Princess Alexandra Hospital
Appendix I: Griffith University HREC Approval

Dear Ms Eugarde

I write further to your application for ethical clearance for your project Prior Review: Examining the reliability of calf and thigh measurements for the diagnosis of deep vein thrombosis in spinal cord injured patients." (GU Ref No: NRS/52/09/HREC). This project has been considered by Human expedited review 1.

The Chair resolved to grant this project conditional ethical clearance, subject to you resolving the following matters:

As per the expectation articulated by the National Statement on Ethical Conduct in Human Research (2007) in relation to multi-site research, and Booklet 8 of the Griffith University Research Ethics Manual, this proposal has been reviewed via the special administrative arrangements for multi-site research. This involves an acceptance of the ethical and scientific review already conducted, and only considers Griffith University specific matters.

Please provide an assurance that the Manager, Research Ethics will be promptly notified if any complaints about the ethical conduct of the research are received.

This decision was made on 14-Jan-10. Your response to these matters will be considered by Office for Research.

The ethical clearance for this protocol runs from 14-Jan-10 to 31-Dec-10.

Please forward your response to Dr Gary Allen, Manager, Research Ethics, Office for Research, as per the details below.

Please refer to the attached sheet for the standard conditions of ethical clearance at Griffith University, as well as responses to questions commonly posed by researchers.

It would be appreciated if you could give your urgent attention to the issues raised by the Committee so that we can finalise the ethical clearance for your protocol promptly.

Regards

Dr Gary Allen
Manager, Research Ethics
Office for Research
Bray Centre, Nathan Campus
Griffith University
ph: 3735 5585
Examiner Information and Consent Form
Version 1.1, dated 19th February 2010

Examining the reliability of calf and thigh measurements in the diagnosis of deep vein thrombosis in spinal cord injured patients

Principal Researcher: Ellen Eugarde, Clinical Nurse
Princess Alexandra Hospital and Griffith University

Principal Supervisor: Prof Leanne Aitken, Professor Critical Care Nursing
Princess Alexandra Hospital and Griffith University

Associate Supervisor: Dr Marion Mitchell, Senior Research Fellow Critical Care Nursing, Princess Alexandra Hospital and Griffith University

This Participant Information and Consent Form is five pages long. Please make sure you have all of the pages.

1. Your Consent

We wish to invite you to take part in this research project. This Participant Information and Consent Form contain detailed information about the research project. Its purpose is to explain to you as openly and clearly as possible all the procedures involved in this project before you decide whether or not to take part in it.

Please read this Participant Information and Consent Form carefully. Feel free to ask questions about any information in the document. You may also wish to discuss the project with a relative or friend - feel free to do this. You may take a copy of the Participant Information and Consent Form to keep as a record.
2. Purpose and Background

Patients with acute spinal cord injury (SCI) are at a very high risk of developing a deep vein thrombosis (DVT). Early diagnosis and intervention are important to minimise the risk of developing a life threatening Pulmonary Embolism (PE) or life–long complications associated with post-thrombotic syndrome (PTS).

As part of the routine care of patients with acute SCI in the Spinal Injuries Unit (SIU) at the Princess Alexandra Hospital (PAH), calf and thigh measurements are performed and recorded by the nurse assigned to them each day. The measurements are performed using a standard sewing tape measure at pre-determined reference points using anatomical landmarks.

To date few studies have examined the reliability of the equipment and technique used for the procedure. The published studies are inconclusive and use varied landmarks, assessor numbers and collection time-points. It is also unknown whether the method currently employed in the SIU is consistent with that of other spinal injuries units in Australia and New Zealand as currently there are no standard practice guidelines for this procedure.

The aim of this project is to test the reliability of the method of calf and thigh circumference measurements currently used in the SIU at the PAH. Without a reliable measure, the everyday practice of calf and thigh measurements is meaningless with the potential that the diagnosis of DVT in SCI patients is delayed or undetected.

It is expected that 42 people will be taking part in the various stages of this project. Seven Nurse Unit Managers (NUMs) will be surveyed in Phase I; one from each of the spinal units from Australia and New Zealand. This survey will be used to compare the current practice within the SIU with that of other spinal injuries units.

In Phase II, 32 healthy volunteers and 25 Registered Nurses from the Queensland Spinal Cord Injuries Service will be recruited to examine the reliability of the equipment and the method currently used within the SIU.

The results of this research will be used by the researcher Ellen Eugarde to obtain a Master of Philosophy degree through Griffith University. This research has been funded by the Covidien Venous Thrombo-embolism Management and Prevention Scholarship Program.

3. Procedures

If you choose to participate in the project you will be performing calf and thigh circumference measurements on healthy volunteers recruited from the staff of the Queensland Spinal Cord Injuries Service. You will be randomly allocated by the toss of a coin to one of two parts of the study, either Part A or Part B. The part to which you are allocated to will determine the number of measurements performed, the location the measurements are performed and the length of time your participation is required for.

Part A

In Part A you will be asked to perform calf and thigh measurements as per the current policy of the SIU at the PAH using two different measuring devices; a standard sewing tape
measure and a spring tape (a self-tensioning tape measure). Training on how to use the spring tape measure will be provided to you prior to commencement of the measurements.

With the healthy volunteer lying in a relaxed supine position you will be asked to perform calf and thigh circumference measurements at pre-marked locations on a single limb.

The circumference at each of the positions will be measured using both of the measuring devices. The circumference measurement will be taken by wrapping the tape measure around the limb at the base of the white adhesive tape. You will call the measurement to a third person who will record the measurement on a data collection form. The process will be repeated for each of the 32 healthy volunteers participating in this phase of the project. The order in which the devices are used will be randomly decided by the toss of a coin for each of the volunteers.

In total you will perform 4 measurements on each volunteer, 2 measurements using the standard sewing tape measure and 2 measurements using the spring tape measure. The measurements for each of the individual healthy volunteer’s are expected to take approximately 5mins. Your total participation time therefore including training is expected to be approximately 1hr.

**Part B**

In Part B you will be asked to perform 4 circumference measurements on a single limb of 16 healthy volunteers lying in a relaxed supine position using a standard sewing tape measure. The measurements will be performed at the widest part of the thigh, the widest part of the calf, the ankle and the mid-point between the ankle and knee. Prior to commencement of this part of the study you will be given a demonstration on the correct position to measure the ankle and the mid-point between the ankle and knee as these measurements are not a part of the current SIU protocol.

You will first measure and mark each position using a small piece of white adhesive tape placed directly above the point the circumference is to be measured. The circumference is then measured by wrapping the tape measure around the limb at the base of the white adhesive tape. The measurement will then be called to a third person who will record the indicated measurement on a data collection form. This process will be repeated at three separate time points. Time 1 = Baseline, Time 2 = 15 minutes later, Time 3 = 30 minutes from the first measurement time.

For this part of the study the procedure outlined will be repeated for each of the 16 healthy volunteers. The order in which volunteers are measured will be randomly allocated at each time point by pulling a number between 1 and 5 out of a box. It is anticipated that in total part B will take approximately 2 hours.
4. **Possible Benefits and Risks**

Though you may not benefit personally from your participation in this project, it is hoped that outcomes from this study may benefit future patients with spinal cord injury. The potential benefits include an increased staff awareness of the importance of early detection and intervention of DVT and the formulation of evidence based guidelines for the practice of calf and thigh circumference measurements.

As the procedures undertaken in this project are non-invasive there are few risks to you as a participant. A potential risk to you as a participant is that of anxiety related to being observed. There is the potential for you to feel as though the standard of your work is in question. If you become upset or distressed as a result of your participation in the research, the researcher is able to arrange for counselling or other appropriate support. Any counselling or support will be provided by staff who are not members of the research team. In addition, you may prefer to suspend or end your participation in the research if distress occurs. Should you wish to make a complaint about your treatment as a participant in this project, please speak to the Principle Investigator.

5. **Privacy, Confidentiality and Disclosure of Information**

In any publication and/or presentation, information will be provided in such a way that you cannot be identified, except with your permission. All information collected from you during the course of the study will be de-identified prior to analysis. In each phase of the study you will be given a unique study identification code which will be used on data collection sheets pertaining to you. A separate log sheet will contain the code and your name and demographic details. This log sheet will be stored in a separate location to the data collection sheets and accessible only to the Principal Researcher.

Throughout the course of the study all data collected will be stored in a locked filing cabinet or on a password protected computer accessible only to the principal researcher. At the completion of the study all of the information collected will be archived and stored for a period of five years after the end of the study before being destroyed.

6. **Further Information and Ethical Approval**

If you require further information or have any problems concerning this project you can contact Ellen Eugarde (telephone 07 3240 2737).

The Human Research Ethics Committees of the Princess Alexandra Hospital and Griffith University have approved this study. Should you have any questions or concerns about the ethical conduct of this study please contact the Secretariat of the Human Research Ethics Committee (07 3240 5856) at the Princess Alexandra Hospital.
7. Participation is Voluntary

Participation in any research project is voluntary. If you choose not to take part you are not obliged to. If you decide to take part and later change your mind, you are free to withdraw from the project at any stage. Your decision whether to take part or not to take part, or to take part and then withdraw, will not affect your relationship with the Princess Alexandra Hospital or the SIU.

8. Withdrawal from the Research Project.

If you decide to withdraw, please notify a member of the research team before you withdraw. If you decide to leave the project, the researchers would like to keep the demographic data about you that have been collected. This is to help them make sure that the results of the research can be measured properly.
Examiner Information and Consent Form  
Version 1, dated 7th October 2009

Examining the reliability of calf and thigh measurements in the diagnosis of deep vein thrombosis in spinal cord injured patients

Principal Researcher: Ellen Eugarde, Clinical Nurse  
     Princess Alexandra Hospital and Griffith University

I have read, or have had read to me in a language that I understand, this document and I understand the purposes, procedures and risks of this project as described within it.

I have had the opportunity to ask questions and I am satisfied with the answers I have received.

I freely agree to participate in the research project as described.

I understand that I will be given a signed copy of this document to keep.

The researcher has agreed not to reveal my identity and personal details if information about this project is published or presented in any public form.

Participants Name: (printed) __________________________________________

Signature: ___________________________ Date: ______________

Name of Witness to Participants Signature (printed): ______________________

Signature: ___________________________ Date: ______________

Declaration by Researcher: I have given a verbal explanation of the research project, its procedures and risks and I believe the participant has understood that explanation.

Researchers Name (printed): ______________________

Signature: ___________________________ Date: ______________
Princess Alexandra Hospital
Ipswich Rd, Woolloongabba, QLD, 4102
Ph (07) 3240 2111

Healthy Volunteer Information and Consent Form
Version 1.1, dated 19th February 2010

Examining the reliability of calf and thigh measurements in the diagnosis of deep vein thrombosis in spinal cord injured patients

Principal Researcher: Ellen Eugarde, Clinical Nurse
Princess Alexandra Hospital and Griffith University

Principal Supervisor: Prof Leanne Aitken, Professor Critical Care Nursing
Princess Alexandra Hospital and Griffith University

Associate Supervisor: Dr Marion Mitchell, Senior Research Fellow Critical Care Nursing, Princess Alexandra Hospital and Griffith University

This Participant Information and Consent Form is five pages long. Please make sure you have all the pages.

1. Your Consent

We wish to invite you to take part in this research project. This Participant Information and Consent Form contains detailed information about the research project. Its purpose is to explain to you as openly and clearly as possible all the procedures involved in this project before you decide whether or not to take part in it.

Please read this Participant Information and Consent Form carefully. Feel free to ask questions about any information in the document. You may also wish to discuss the project with a relative or friend - feel free to do this. You may take a copy of the Participant Information and Consent Form to keep as a record.

2. Purpose and Background

Patients with acute spinal cord injury (SCI) are at a very high risk of developing a deep vein thrombosis (DVT). Early diagnosis and intervention are important to minimise the risk of
developing a life threatening Pulmonary Embolism (PE) or life –long complications associated with Post-thrombotic syndrome (PTS).

As part of the routine care of patients with acute SCI in the Spinal Injuries Unit (SIU) at the Princess Alexandra Hospital (PAH), calf and thigh measurements are performed and recorded by the nurse assigned to them each day. The measurements are performed using a standard sewing tape measure at pre-determined reference points using anatomical landmarks.

To date few studies have examined the reliability of the equipment and technique used for the procedure. The published studies are inconclusive and use varied landmarks, assessor numbers and collection time-points. It is also unknown whether the method currently employed in the SIU is consistent with that of other spinal injuries units in Australia and New Zealand as currently there are no standard practice guidelines for this procedure.

The aim of this project is to test the reliability of the method of calf and thigh circumference measurements currently used in the SIU at the PAH. Without a reliable measure, the everyday practice of calf and thigh measurements is meaningless with the potential that the diagnosis of DVT in SCI patients is delayed or undetected.

It is expected that 42 people will be taking part in the various stages of this project. Seven Nurse Unit Managers (NUMs) will be surveyed in Phase I; one from each of the spinal units from Australia and New Zealand. This survey will be used to compare the current practice within the SIU with that of other spinal injuries units.

In Phase II, 32 healthy volunteers and 25 Registered Nurses from the Queensland Spinal Cord Injuries Service will be recruited to examine the reliability of the equipment and the method currently used within the SIU.

The results of this research will be used by the researcher Ellen Eugarde to obtain a Master of Philosophy degree through Griffith University. This research has been funded by the Covidien Venous Thrombo-embolism Management and Prevention Scholarship Program.

3. Procedures

If you choose to participate in the project you will have calf and thigh circumference measurements performed by Registered Nurses from the SIU of the PAH. You will be randomly allocated by the toss of a coin to one of two parts of the study, either Part A or Part B. The part to which you are allocated to will determine the number of measurements you will have performed and the location the measurements. You will be asked to wear a comfortable loose fitting pair of shorts for the study.

Part A
In Part A you will be asked to lie on your back in a comfortable relaxed position on a standard hospital bed in a treatment room in the SIU. On a randomly selected leg the Investigator will measure the distance from the top of the knee cap to the widest part of the thigh and then place a small piece of white adhesive tape, 2cm long and .5cm wide as a marker directly above the position to be measured. The distance from the base of the knee cap to the widest part of the calf will then be measured and a small piece of white adhesive tape placed just above the position to be measured.
The RN will then measure the circumference at each of the two positions using two different types of tape measures, a standard sewing tape measure and a device called a spring tape (a self-tensioning tape measure). The order in which the tape measures are used will be randomised. The circumference measurement will be taken by wrapping the tape measure around the limb at the base of the white adhesive tape. The RN will call the measurement to a third person who will record the measurement on a data collection form. The process will be repeated by each of the 20 RNs participating in this phase of the project. This is anticipated to take between one and two hours.

**Part B**

In Part B you will be asked to lie on your back in a comfortable relaxed position on a standard hospital bed in a treatment room in the SIU. A RN will perform 4 circumference measurements on a single limb using a standard sewing tape measure. The measurements will be performed at the widest part of the thigh, the widest part of the calf, the ankle and the mid-point between the ankle and knee. Each position will be first measured and marked by the RN using a small piece of white adhesive tape placed directly above the point to be measured. The RN will then wrap the tape measure around the limb at the base of the white adhesive tape and call the measurement to a third person recording the indicated measurement on a data collection form. The process will be repeated by the RN at three separate time points. Time 1 = Baseline, Time 2 = 15 minutes later, Time 3 = 30 minutes from the first measurement time. For this part of the study the procedure outlined will be performed by 16 RNs. It is anticipated that in total part B will take approximately 2 hours.

4. **Possible Benefits and Risks**

Though you may not benefit personally from your participation in this project, it is hoped that outcomes from this study may benefit future patients with spinal cord injury. The potential benefits include an increase staff awareness of the importance of early detection and intervention of DVT and the formulation of evidence based guidelines for the practice of calf and thigh circumference measurements.

As the procedures undertaken in this project are non-invasive there are few risks to you as a participant. There is a small risk that you may experience some skin irritation associated with the positioning and removal of the white adhesive tape used. To minimise this risk you will be encouraged to remove the tape yourself where possible and be provided with option of using wipes specifically designed for the purpose of removing adhesive tapes. Should the use of the adhesive tape cause a problem your participation in the project will be discontinued immediately. Should you wish to make a complaint about your treatment as a participant in this project, please speak to the Principle Investigator.

5. **Privacy, Confidentiality and Disclosure of Information**

The data collection forms are anonymous with no identifying data of healthy volunteers or RNs recorded. Any information obtained in connection with this project will remain confidential. It will only be disclosed with your permission, except as required by law.
In any publication, information will be provided in such a way that you cannot be identified. Only summarised data will be made publicly available to maintain confidentiality. Information you provide for this study will be retained for a minimum of 5 years. After this time, all data will be destroyed.

6. Further Information and Ethical Approval

If you require further information or have any problems concerning this project you can contact Ellen Eugarde (telephone 07 3240 2737). The Human Research Ethics Committees of the Princess Alexandra Hospital and Griffith University have approved this study. Should you have any questions or concerns about the ethical conduct of this study please contact the Secretariat of the Human Research Ethics Committee (07 3240 5856) at the Princess Alexandra Hospital.

7. Participation is Voluntary

Participation in any research project is voluntary. If you choose not to take part you are not obliged to. If you decide to take part and later change your mind, you are free to withdraw from the project at any stage. Your decision whether to take part or not to take part, or to take part and then withdraw, will not affect your relationship with the Princess Alexandra Hospital or the SIU.
Healthy Volunteer Information and Consent Form
Version 1.1, dated 19th February 2010

Examining the reliability of calf and thigh measurements in the diagnosis of deep vein thrombosis in spinal cord injured patients

Principal Researcher: Ellen Eugarde, Clinical Nurse
Princess Alexandra Hospital and Griffith University

I have read, or have had read to me in a language that I understand, this document and I understand the purposes, procedures and risks of this project as described within it.

I have had the opportunity to ask questions and I am satisfied with the answers I have received.

I freely agree to participate in the research project as described

I understand that I will be given a signed copy of this document to keep.

The researcher has agreed not to reveal my identity and personal details if information about this project is published or presented in any public form.

Participants Name: (printed) ____________________________________________
Signature: ____________________________ Date: ________________

Name of Witness to Participants Signature (printed): ________________________
Signature: ____________________________ Date: ________________

Declaration by Researcher: I have given a verbal explanation of the research project, its procedures and risks and I believe the participant has understood that explanation.

Researchers Name (printed): ____________________________
Signature: ____________________________ Date: __________________