University of Calgary

PRISM Repository **https://prism.ucalgary.ca**

The Vault Open Theses and Dissertations

2013-01-09

Qualifying the Variability of Surface Topography Indices for Detection of Clinical Progression of Scoliosis

Dubetz, Tyler Paul Arthur

Dubetz, T. P. (2013). Qualifying the Variability of Surface Topography Indices for Detection of Clinical Progression of Scoliosis (Master's thesis, University of Calgary, Calgary, Canada). Retrieved from https://prism.ucalgary.ca. doi:10.11575/PRISM/27490

http://hdl.handle.net/11023/403

Downloaded from PRISM Repository, University of Calgary

UNIVERSITY OF CALGARY

Qualifying the Variability of Surface Topography Indices for Detection of Clinical Progression of Scoliosis

by

Tyler Paul Arthur Dubetz

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF MECHANICAL AND MANUFACTURING ENGINEERING

CALGARY, ALBERTA

DECEMBER, 2012

© Tyler Paul Arthur Dubetz 2012

Abstract

Surface topography (ST) measurements of scoliotic deformities offer the potential to reduce radiation and to provide a more complete description of the deformity when compared to standard X-rays. These ST measurements must be evaluated within a clinical context to determine their feasibility for clinical and research applications. This study quantifies the variability in a set of developed ST indices and estimates a magnitude of clinically important difference based on current clinical standards. The variability is compared to the clinically important difference to determine if the system is adequate for clinical implementation (individual evaluation) or research implementation (group-to-group comparisons). Nine of ten ST indices were found acceptable for typical research implementation. One of ten ST indices (aspect ratio) was found acceptable for typical clinical implementation. Aspect ratio also shows a trend towards differentiating between progressed and non-progressed scoliosis. However, the estimation of what constitutes clinically important difference should be examined further.

Preface

Parts of this thesis have previously been published in:

Dubetz, T.P.A., Smith, K.N., Küpper, J.C., Howard, J.J., Harder, J.A., Joughin, V.E., Ronsky, J.L. (2011a). Detecting clinically significant progression in adolescent idiopathic scoliosis using surface topography indices. Proceedings of the Eleventh Annual Scientific Conference of the Canadian Spine Society, Fairmont Chateau Frontenac, Quebec City, Canada, March 9-12, 2011. (Podium Presentation).

Dubetz, T.P.A., Küpper, J.C., Smith, K.N., Howard, J.J., Harder, J.A., Ronsky, J.L. (2011b). Non-invasive surface imaging for the diagnosis, monitoring, and treatment of scoliosis. Proceedings of the Alberta Graduate Conference – Today's Ideas, Tomorrow's Innovators. MacEwan Hall, University of Calgary, Calgary, Canada, May 5-7, 2011. (Podium Presentation).

The author of this thesis was the primary author and presenter of both of the listed works. As such, the majority of data analysis, presentation of results, and discussion of results was performed by the current author. Other authors listed on these publications assisted in data processing, clinical interpretations and guidelines, and overview of the project as a whole.

Acknowledgements

I would like to thank the following people:

My supervisor, Dr. Janet Ronsky, for supporting me for all these years. She has provided an endless source of enthusiasm and optimism. Dr. Jason Howard, for always being available to bounce ideas off, sharing his clinical knowledge, and always offering a helping hand regardless of the situation or (lack of) time. Jessica Küpper for her extremely valuable feedback on a wide variety of topics. Dr. Tak Fung for his generous help with statistics. Both Ting and Payam for reminding me that I can do this. And my mom for the endless support.

My funding bodies that have allowed me to pursue this graduate degree: the Alberta Ingenuity Fund, NSERC, and the APEGA Education Foundation.

And finally all my friends, family, and lab mates that have helped me to get through this in one piece.

Dedication

To my brother, who always set the bar so high, I've gotten to where I am just by trying to keep up.

And to my grandpa, who taught me it's okay to strive for perfection.

Table of Contents

Abstract ii
Preface iii
Acknowledgements iv
Dedication v
Table of Contents vi
List of Tables viii
List of Figures and Illustrations ix
List of Symbols, Abbreviations and Nomenclature xi

CHAPTER ONE: INTRODUCTION 1
1.1 Scoliosis 1
1.2 Clinical Standard of Care 4
1.2.1 Critique of Cobb Angle 8
1.3 Alternative Systems of Surveillance 10
1.3.1 Radiographic Based Imaging 10
1.3.1.1 Stereo-radiographic techniques 10
1.3.1.2 Computed Tomography scans 11
1.3.1.3 EOS 12
1.3.2 Magnetic Resonance Imaging 13
1.3.3 Surface Topography Evaluation of Scoliosis 14
1.4 Project Overview 16
1.4.1 Surface Topography Imaging System 17
1.4.2 Surface Topography Indices 19
1.4.3 Sources of Variability 24
1.4.4 Application of System 28
1.5 Objectives/Hypothesis/Specific Aims 30
1.6 Thesis Organization 31

CHAPTER TWO: DETERMINATION OF VARIABILITY IN SURFACE TOPOGRAPHY INDICES 32
2.1 Introduction 32
2.1.1 Terminology 33
2.2 Background and Literature 37
2.2.1 Variability of Cobb Angle 37
2.2.2 Variability of Surface Topography System 39
2.3 Measures of Variability 43
2.3.1 Intraclass Correlation Coefficients 45
2.3.2 Minimum Detectable Change 48
2.4 Methods 52
2.4.1 Data Acquisition and Processing Protocol 52
2.4.2 Test-Retest Design 55

2.4.3 Determination of Statistics 56
2.4.3.1 Calculation of ICC, SEM, and MDC 58
2.4.4 Subject Demographics 59
2.5 Results 61
2.6 Discussion 63

CHAPTER THREE: CLINICALLY RELEVANT LEVELS OF CHANGE 66
3.1 Methods 68
3.1.1 MCID of Cobb Angle 69
3.1.2 MCID of ST Indices 70
3.1.3 Calculations and Statistical Analysis 71
3.1.4 Study Design 74
3.1.5 Subject Demographics 75
3.1.6 Assumptions and Discrepancies 78
3.2 Results 80
3.3 Discussion 85

CHAPTER FOUR: DISCUSSION 90
4.1 Assessment of ST Imaging System 91
4.1.1 Quantification of Variability in ST Indices 91
4.1.2 Estimation of Clinically Important Differences in ST Indices 93
4.1.3 Qualification of Variability with respect to Clinically Important Differences in ST Indices 93
4.2 Improvement of ST Imaging System 99
4.2.1 Decreasing Variability of ST Indices 100
4.2.2 Different Estimation of MCID 101
4.3 Limitations 102
4.3.1 Quantification of MDC 102
4.3.2 Comparison of MDC to MCID 103
4.3.3 Estimation of MCID 103
4.4 Summary of Findings 107
4.5 Future Work and Conclusions 108

REFERENCES 113

APPENDIX 1: ST INDEX DATA FROM ANOVA TABLES 121

List of Tables

Table 2.1: Summary of mean and maximal intra-scan standard deviation (SD) for each index curve. The standard deviations are also given as a percentage of the maximum index value of an example subject to illustrate the magnitude. (Swanson, 2008, p. 83) 41

Table 2.2: Intra-observer variability from InSpeck based ST system in Montreal..... 43

Table 2.3: MDC subject demographics 60

Table 2.4: ICC(2,1) of ST indices (modified from Dubetz et al., 2011a)..... 62

Table 2.5: MDC (95% confidence interval) of ST indices (modified from Dubetz et al., 2011a)..... 62

Table 2.6: Comparison of ST index variability across three studies in two different research groups 65

Table 3.1: Clinically progressed group demographics 76

Table 3.2: Non-progressed group demographics 77

Table 3.3: Comparison of progressed vs. non-progressed groups 77

Table 3.4: Changes in ST indices within a clinically progressed AIS population (modified from Dubetz et al., 2011a; Dubetz et al., 2011b) 81

Table 3.5: Changes in ST indices within a non-progressed AIS population 82

Table 3.6: Tests of normality 84

Table 3.7: Independent samples tests..... 85

Table 4.1: ICC values of various measures of scoliosis 92

Table 4.2: MDC and MCID values for the ST indices. The lightest colour of shading represents a clinically acceptable MDC when compared to the mean MCID, the medium colour of shading represents an acceptable MDC when compared to the 37.5th percentile and mean MCID, and the darkest shading represents an acceptable MDC when compared to the 25th percentile, 37.5th percentile, and mean MCID. Un-shaded cells in the table represent situations where the MDC is clinically unacceptable as it is higher than all of the MCID values..... 96

List of Figures and Illustrations

Figure 1.1: A) Shoulder asymmetry (note how the left shoulder is higher in the right) B) Waistline asymmetry and the body shifted to the right C&D) Rib prominence tends to correlate with the size of the curve E) Patients typically look normal when viewed from the side (SRS, 2012, Physical findings) 2

Figure 1.2: Cobb angle example 5

Figure 1.3: Typical flow chart for treatment of AIS in skeletally immature patients..... 7

Figure 1.4: Photographs of four patients each with the same degree of curvature (70 degrees) but each with a different curve pattern. From left to right the curves are lumbar, thoraco-lumbar, combined thoracic and lumbar, and thoracic. (James, 1954, p. 48) 10

Figure 1.5: *Back surface rotation (BSR)* and *principal axis rotation (PAX)*. θ_1 = PAX rotation with respect to the patient PSIS reference axis. θ_2 = BSR, measured using the line joining left and right dual-tangent points T_L and T_R. θ_3 = difference between BSR and PAX rotation. *Rib prominence* = $d_L - d_R$, the difference in distances from left and right dual-tangent points to the reference axis. (modified from Jaremko, 2001, p. 59)..... 20

Figure 1.6: Quarter-areas for index calculation. Rear quarter-areas (A_{QL}, A_{QR}) and their centroids (Q_{CL}, Q_{CR}), defined by quadrants cut through the centroid (O) parallel and perpendicular to the reference axis (PSIS, PAX or BSR). (Jaremko, 2001, p. 60) 21

Figure 1.7: Indices defining left-right asymmetry of half-areas. Half-areas were cut relative to an appropriate reference axis (BSR, PAX or PSIS; PAX shown). Asymmetry of half-centroid locations C_L and C_R was measured in the anteroposterior (dXC) and lateral (Z_{CL} vs. Z_{CR}) directions. θ = angle of rotation of the line joining the half-centroids. dAP/dLat = aspect ratio for each half-area (left and right). Fz = hypothetical unit force applied inward at each half-centroid. In a symmetric torso, the forces would cancel; in a scoliotic patient, a twisting moment proportional to dXC would be generated. (Jaremko, 2001, p. 60) 22

Figure 1.8: Data acquisition and processing protocol..... 23

Figure 1.9: Selection of spinous process line from transverse torso cross sections 28

Figure 2.1: Two limiting distributions, one for a high-precision measurement, the other for a low-precision measurement (modified from Taylor, 1997, p. 129) 34

Figure 2.2: Measurement variability 50

Figure 2.3: Data acquisition and processing protocol 53

Figure 2.4: Surface topography scan 54

Figure 2.5: 3D torso model – deletion of unwanted structures 54

Figure 3.1: Comparison of mean within-subject change scores for all ST indices between progressed and non-progressed groups. Error bars represent ± 1 standard deviation of the between-subject variability within each group. 83

List of Symbols, Abbreviations and Nomenclature

Symbol	Definition
2D	Two Dimensional
3D	Three Dimensional
AIS	Adolescent Idiopathic Scoliosis
ANOVA	Analysis of Variance
AP	Anteroposterior
BSR	Back Surface Rotation
CMM	Coordinate Measuring Machine
CNC	Computer Numerical Control
CR	Computed Radiography
CT	Computed Tomography
DA	Discriminant Analysis
DM	Double Major
DT	Double Thoracic
H	Hypothesis
ICC	Intraclass Correlation Coefficient
IIS	Infantile Idiopathic Scoliosis
JIS	Juvenile Idiopathic Scoliosis
L	Lumbar
MCID	Minimum Clinically Important Difference
MDC	Minimum Detectable Change
ML	Medial-Lateral
MR	Magnetic Resonance
MS_e	Mean Square Error
MT	Main Thoracic
PA	Posteroanterior
PAX	Principal Axis Orientation
PSIS	Posterior Superior Iliac Spine
RMS	Root Mean Square
SA	Specific Aims
SAQ	Spinal Appearance Questionnaire
SEM	Standard Error of Measurement
SP	Spinous Process
SRS	Scoliosis Research Society
SRS-22	Scoliosis Research Society Patient Questionnaire
ST	Surface Topography
TL	Thoracolumbar
U of C	University of Calgary
V	Vertical

CHAPTER ONE: INTRODUCTION

Spinal deformities such as scoliosis rely on measurements of human geometry to diagnose and describe them. Currently, measurements taken as part of routine clinical treatment of these conditions do not fully utilize the recent advancements in technology in capturing or describing free-form shapes in three dimensions. The application of such technology could provide solutions to clinicians by better describing the deformity and enabling more informed treatment decisions. These advancements could also improve patient outcomes by reducing negative consequences associated with current practices (i.e. X-ray radiation exposure).

1.1 Scoliosis

Scoliosis is a skeletal deformity of the spine and ribcage. It is classically defined as a side-to-side (lateral) curvature of the spine by the Scoliosis Research Society (SRS) (SRS, 2012, What is scoliosis) although the true nature of the deformity is complex and three-dimensional (3D). While there is always an abnormal curvature to the spine in the coronal plane, it is accompanied by rotations of the vertebrae in the transverse plane, abnormal curvatures in the sagittal plane, wedging of the vertebrae and intervertebral discs, and rib deformities (Pope et al., 1984; Machida, 1999; Dickson, 1999). Some of the common externally visual results of the internal deformity can include shoulder blade asymmetry, shoulder height asymmetry, waist height asymmetry, rib prominence or hump, and a twisted or curved appearance to the torso as a whole (Figure 1.1). The skeletal deformity associated with scoliosis has been shown to have detrimental effects

on the psychological and psychosocial well being of the patient, such as self-esteem, body image, and perception of limitation (Clayson et al., 1987; Weinstein et al., 2003) as well as reduced likelihood of marriage (Pope et al., 1984). In more extreme cases it can have detrimental effects on the physical well being such as increased incidence of back pain, reduced respiratory function, higher mortality rates, and higher working disability (Weinstein, 1999).

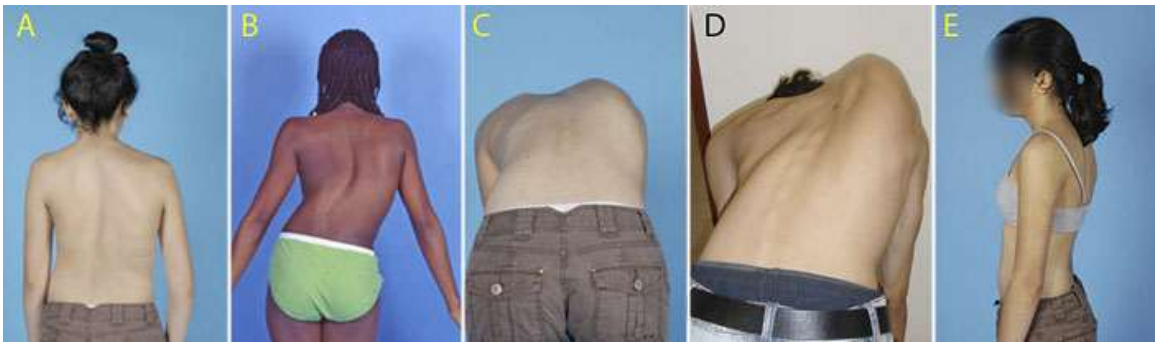

Figure 1.1:
A) Shoulder asymmetry (note how the left shoulder is higher in the right)
B) Waistline asymmetry and the body shifted to the right
C&D) Rib prominence tends to correlate with the size of the curve
E) Patients typically look normal when viewed from the side
(SRS, 2012, Physical findings)

There are various types of scoliosis, classified both by the age of the patient when the condition first presents itself (infantile, juvenile, adolescent, adult) and the cause of the condition (congenital, neuromuscular, idiopathic). Out of all forms of scoliosis, over 70% of cases are idiopathic (Kane & Moe, 1970; Pope et al., 1984), and of these approximately 80% are adolescent (Asher & Burton, 2006). This thesis focuses on the most common type: adolescent idiopathic scoliosis (AIS). This form of scoliosis affects

adolescents, typically between the ages of 10 and 18, and is of unknown cause. The prevalence of AIS in the general population is 2% to 3%, with an overall female preponderance of 3.6:1. This female preponderance greatly increases with Cobb angle (defined in section 1.2), ranging from almost equal numbers in females and males with small curves, and reaching 10:1 for curves greater than 30° (Weinstein, 1999). Of those clinically diagnosed, less than 10% will progress to a point requiring conservative (bracing) or surgical treatment (Weinstein, 1999).

The symptoms particular to AIS are typically milder than other forms of scoliosis, mainly affecting the patient's psychological and psychosocial well-being due to the physical deformity. Rates of back pain and mortality are comparable between AIS and normal populations, and pulmonary function is only affected in extreme thoracic curves reaching 100° or greater (Weinstein, 1999).

AIS is a very heterogeneous disease differing primarily with regard to the probability of curve progression, typically related to the age of onset, gender, curve location, curve magnitude, and curve pattern (James, 1954; Weinstein, 1999; Asher & Burton, 2006). Several classification models have been developed and adopted clinically to assist physicians in treatment decisions. The King system (King et al., 1983) has five different curve types, as well as a miscellaneous category. The Lenke system (Lenke et al., 2001) has six curve types, three lumbar spine modifiers, and three sagittal thoracic modifiers, resulting in up to 54 different combinations.

The multifactorial etiopathogenesis aspect to AIS is acknowledged, and more recently, much work has been done to investigate this component. Wang et al. (2011) provided an extensive literature review, categorizing the various theories of etiopathogenesis of AIS into six subgroups: genetics factors, nervous system abnormalities, abnormal skeletal growth, hormones and metabolic dysfunction, biomechanical factors, and environmental and life style factors. The heterogeneity present in AIS as well as the multifactorial aspect of the disease were strongly emphasized. This notion is summarized best by their final question "Is AIS one disease or heterogeneous grouping?"

Due to this wide variety in presentation of symptoms, patient concerns, and rates of curve progression, the clinical treatment of AIS may vary substantially between patients. This lack of predictability necessitates frequent clinical monitoring of the condition.

1.2 Clinical Standard of Care

The current standard clinical measure of scoliosis severity is the Cobb angle taken from coronal plane X-rays (Cobb, 1948). The Cobb angle is calculated as the angle between the endplates of the vertebrae at either end of the curve under consideration (Figure 1.2), using the superior endplate of the superior vertebra and the inferior endplate of the inferior vertebra. The vertebrae chosen as the ends of the curve are defined as the vertebrae with the greatest angle to the vertical. A perfectly symmetrical spine would have a Cobb angle of 0°. A spine curvature with Cobb angle measurement under 10° is considered normal postural asymmetry and not scoliosis (Kane, 1977; SRS, 2012, Imaging Studies).

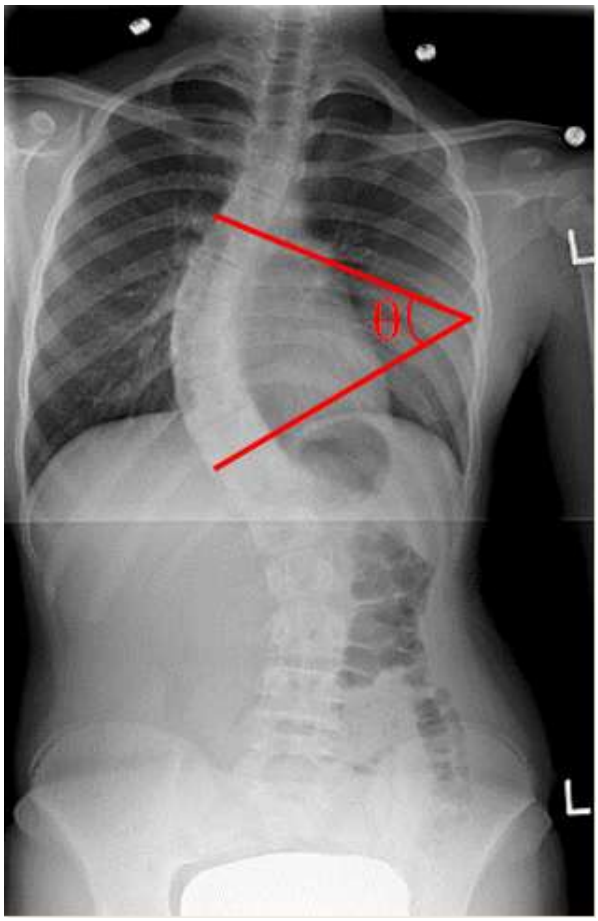

Figure 1.2: Cobb angle example

The current clinical standard of care involves the use of full torso X-rays in both the diagnosis and monitoring of the disease. The typical interval between X-rays is approximately six months. This monitoring is required from initial diagnosis until the patient has reached skeletal maturity or the spine stabilizes, which could be several years. Guidelines established by the SRS (SRS, 2012, Treatment) recommend treatment based on Cobb angles as follows:

a) Monitoring of progression for Cobb angles less than 25° in growing patients and less than 50° in skeletally mature patients.
b) Conservative treatment through bracing for Cobb angles between 25° and 40° during growth.
c) Surgical treatment for Cobb angles exceeding 45° while growing or exceeding 50° when skeletally mature.

However, there are other factors such as curve progression rate or psychological distress that could warrant surgical intervention even if the Cobb angle has yet to reach 45° (Weiss, 2008; Grivas, 2002). The history and development of treatment guidelines and definitions of progression associated to the Cobb angle are relatively arbitrary (Kane, 1977; Lonstein & Carlson, 1984). It is important to remember this fact when attempting to define strict definitions or design studies that rely on Cobb angle measurements.

It is important to note that scoliosis falls along a continuum ranging from healthy non-deformity, up to severe deformities requiring extensive surgery. It is not simply a positive or negative diagnosis. It is this continuum and the progression along it that necessitates the frequent monitoring of the state of the deformity, as demonstrated in the typical flow-chart for the treatment of AIS in a growing patient provided in Figure 1.3. It should be noted that the primary objective of conservative treatment with bracing is to stop curve progression (Veldhuizen et al., 2002; Rigo et al., 2003). It is therefore important to implement such treatment in a timely manner before the opportunity is lost and only invasive surgical treatment is left as a viable option. This is why frequent

monitoring of the deformity is a necessity, especially during the dramatic growth stages associated with adolescence where rapid curve progression can occur (Terver et al., 1980). Currently, with the use of X-rays, there is a trade-off between the benefits of more frequent monitoring and the drawbacks of increased radiation exposure.

The typical radiographic monitoring protocol implemented at the Alberta Children's Hospital, which is the centre where all subjects for the current study were recruited, uses a standing coronal plane X-ray from which the Cobb angle is measured. However, additional views such as side-bending, supine distraction, or lateral radiographs may also be taken at the physician's discretion. These are typically used for surgical planning, and are required for complete Lenke or King classifications.

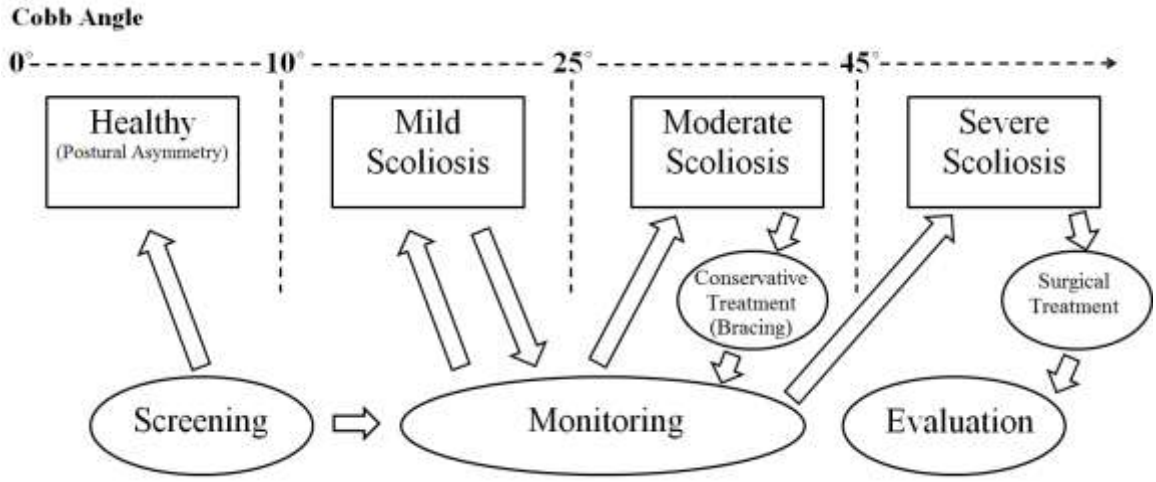

Figure 1.3: Typical flow chart for treatment of AIS in skeletally immature patients

1.2.1 Critique of Cobb Angle

The use of the Cobb angle for the diagnosis and monitoring of AIS patients poses two significant problems, namely an increased risk of cancer and an inadequate description of the scoliotic deformity.

It is well established that exposure to X-ray radiation carries risks of increased cancer rates, and should only be done if medically necessary. The typical AIS patient will receive numerous torso X-rays over the course of treatment, with the reported mean number of radiographic examinations ranging from 10.1 to 24.7 (Doody et al., 2000; Levy et al., 1996). Considering that torso X-rays cover a relatively large area of the body, that the majority of AIS patients are female, and that the X-rays are taken during adolescence when rapid growth and development associated with puberty is taking place, it is prudent to examine the detrimental effects of such radiation exposure. Doody et al. (2000) showed that in a large population of women (5573 patients enrolled) under the age of 20 when first diagnosed with spinal deformity (92.7% of which were scoliosis), there was a statistically significant 1.7-fold increase in the likelihood of dying due to breast cancer compared to the general population. However, this population were all diagnosed between 1912 and 1965, with the majority (97.4%) of X-ray exams taking place before 1976. Levy et al. (1996) modeled the lifetime risk of cancers of the thyroid gland, female breast, lungs, digestive organs, and active bone marrow in an AIS specific population using radiation doses obtained from modern radiographic equipment and techniques, which are lower than the doses used in the study by Doody et al. They reported the highest risk in women, aged 9-13 at time of referral who required surgical correction, to

be 238 excess incident cancers per 100,000 individuals. The number of excess incident cancers decreased with increasing age at time of referral, men compared to women, and tended to decrease with decreasing Cobb angle to a low of 14 excess incident cancers per 100,000. They predicted that by replacing the typical anterorposterior (AP) X-rays with posteroranterior (PA) X-rays, these numbers could be reduced to a high of 96 and a low of 10 excess incident cancers per 100,000. The authors concluded that the cancer risks from full-spinal radiographs for scoliosis are not negligible. In spite of this, the Cobb angle measured from radiographs continues to be the standard of care.

As described earlier, scoliosis is a complex 3D deformity involving the spine and ribcage. The Cobb angle's singular, two-dimensional quantification of scoliosis is a simplified description of the deformity. The need for a more complete 3D description has been stressed for surgical planning (Lenke et al., 2001), often justifying the collection of radiographs in multiple planes. This need has also been stressed by many researchers (Pope et al., 1984; Stokes et al., 1987; Dickson, 1999; Labelle et al., 1995), culminating in the formation of a Scoliosis Research Society working group on 3D terminology of spinal deformity (Stokes et al., 1994).

In AIS where a primary concern is the cosmetic aspect of the deformity (Weinstein, 1999), the Cobb angle, which only quantifies 2D spinal curvature, may not be a good measure of 'severity' of the deformity from the patient perspective. This is dramatically highlighted in Figure 1.4, which shows four patients with the same Cobb angle but varied cosmetic outcomes (James, 1954).

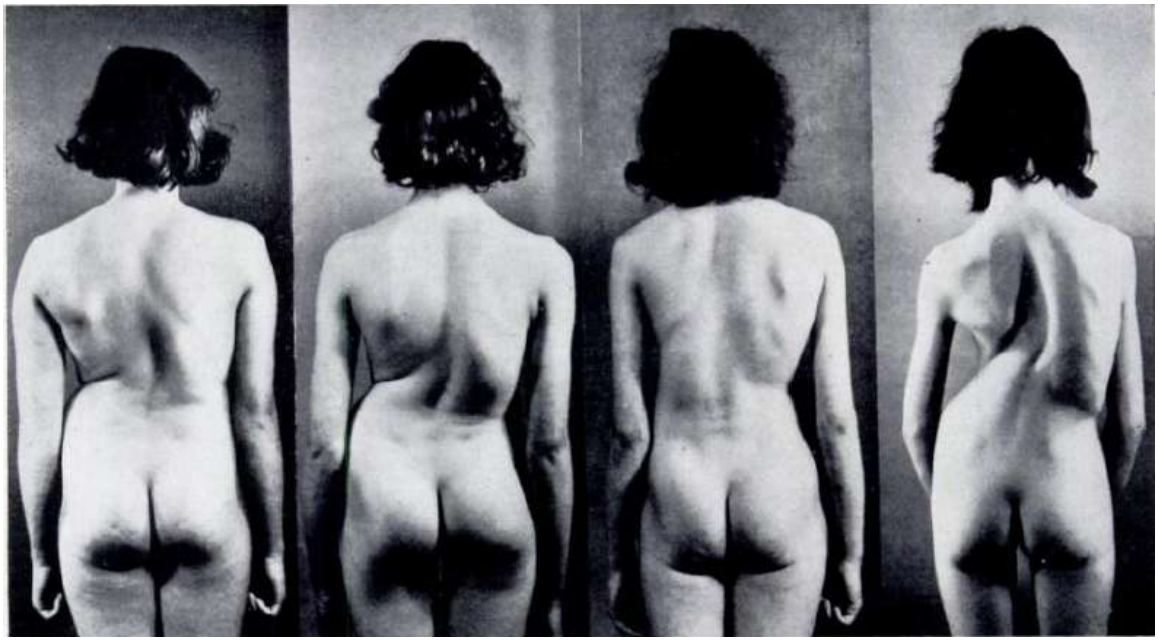

Figure 1.4: Photographs of four patients each with the same degree of curvature (70 degrees) but each with a different curve pattern. From left to right the curves are lumbar, thoraco-lumbar, combined thoracic and lumbar, and thoracic. (James, 1954, p. 48)

1.3 Alternative Systems of Surveillance

Many alternatives to traditional approaches for the surveillance of scoliosis have been developed or suggested. While these typically attempt to address some of the shortfalls of the radiographic Cobb angle method, they all possess various advantages and limitations. Brief descriptions of some of the more prominent alternatives to the current standard are detailed below.

1.3.1 Radiographic Based Imaging

1.3.1.1 Stereo-radiographic techniques

Techniques have been developed which utilize two or more X-rays taken at different angles, with fiducial marks of known dimensions in the field of view. This allows

reconstruction of the 3D geometry of the spinal vertebrae and rib cage through the use of a Direct Linear Transform (Dansereau & Stokes, 1988; Labelle et al., 1995).

Obtaining the 3D spinal geometry enables the quantification of additional 3D descriptions of the scoliotic deformity, such as the plane of maximum curvature (Stokes et al., 1994) and vertebral axial rotation (Labelle et al., 1995). Serial 3D spinal geometry has also been used in the prediction of future scoliosis progression to within 4.1° through the use of artificial progression surfaces (Wu et al., 2010). Used in conjunction with surface topography (ST) methods, these stereo-radiographic techniques have been used to register the 3D external torso geometry with the 3D internal bony geometry (Poncet et al., 2000), allowing for the investigation of correlations between the two.

While these techniques can provide invaluable 3D information, they at least double the X-ray exposure to the patient due to the requirement of at least two X-ray images. Combined with the additional expense of processing twice as many radiographic films, and the resources required to perform the 3D reconstructions, these techniques have mainly been used as the foundation for research purposes and have not yet seen routine clinical implementation.

1.3.1.2 Computed Tomography scans

Computed tomography (CT) scans have an ability to show more detail with higher resolutions than a standard radiograph. Consequently, CT scans have been used to more accurately determine vertebral rotations and vertebral morphology (Yazici et al., 2001;

Krismer et al., 1996; Oestreich et al., 1998; Liljenqvist et al., 2000). CT scans have also been used to evaluate pedicle screw placement in the surgical correction of scoliosis (Liljenqvist et al., 1997).

However, CT scans expose the patient to a much greater radiation dose as compared to a standard X-ray. Don (2004) showed that an abdomen/pelvic CT scan has an effective dose 56 times higher than a scoliosis AP radiograph. Additionally, scoliotic deformity is minimized in a non-weight bearing position, for example in a supine position, as would be required for a standard CT scan (Little et al., 2012; Yazici et al., 2001; Dickson, 1999). This makes it difficult to measure the true severity of the spinal deformity. For these reasons, CT scans are not used for routine monitoring of AIS.

1.3.1.3 EOS

The EOS imaging system (EOS Imaging SA., Paris) is a relatively new imaging modality that has recently become commercially available. It is based on slot scanning radiograph imagers, which allow the reduction of scattered radiation and improved signal-to-noise ratios. The EOS system can therefore theoretically obtain better images with lower doses as compared to standard computed radiography (CR) methods. These claims were examined by Deschenes et al. (2010), who showed that the EOS system produced significantly better images with regard to all but one examined structure, and had reduced doses ranging from 2.9 to 9.2 times less when compared to CR methods, in scenarios where both a PA and lateral image were obtained. The reduction in dosage for the thoracoabdominal region ranged from 6 to 9 times less when compared to CR methods.

The EOS system is designed to obtain stereo images in both the coronal and sagittal planes, allowing for 3D reconstructions similar to the techniques described in section 1.3.1.1. Unlike CT scans, the images are taken in a standing, weight-bearing position. The associated sterEOS software (EOS Imaging SA., Paris) that performs the 3D reconstruction may have potential limitations as it fits the detected bony geometry with models from a proprietary database. If a specific geometry arises in which none of the programmed models accurately describe the imaged geometry, it is unclear how the sterEOS software would handle it. In addition to radiation exposure and fidelity of geometric model fit, other drawbacks include cost, a scan time up to 20 seconds (EOS Imaging, 2010, Workflow), and the fact that only the bony geometry rather than the externally visible torso is imaged. On this last point, it could be argued that the externally visible torso geometry is more important to the patient. In spite of these potential limitations, the EOS system shows promise for application to the routine monitoring of scoliosis.

1.3.2 Magnetic Resonance Imaging

Magnetic Resonance (MR) imaging possesses many positive attributes for the surveillance of scoliosis, such as being non-invasive, preventing radiation exposure, and allowing for the imaging of the internal skeletal structures directly. Unfortunately, it is impractical for routine monitoring of AIS for several reasons. As described in section 1.3.1.2, the scoliotic deformity is minimized if a patient is in a supine position, which is typically required in MR imaging as it is in CT scanning. While standing MR scanners do exist, they are considerably more rare, often lower strength (poorer image quality),

and therefore much less accessible than traditional MR scanners. This poses a problem due to typical budgetary constraints within healthcare systems. Metallic instrumentation used for the surgical correction of scoliosis may also prohibit the use of MR imaging (Oestreich et al., 1998) or result in substantial image artifacts. However, MR imaging does have advantages for specific tasks in the surgical treatment of AIS such as assessment of the contents of the spinal canal (Kotwicki, 2008; Oestreich et al., 1998), and may be used in the planning of surgical procedures.

1.3.3 Surface Topography Evaluation of Scoliosis

Surface topography has had a significant presence in the evaluation of scoliosis for as long as the condition has been described, predating the discovery of X-rays. This is an intuitive fact, as one of the primary symptoms of the disease is the deformity of the torso and rib cage. This aspect of the deformity is typically the paramount concern of the patient.

Part of the clinical examination of a scoliosis patient is a visual assessment by the treating physician, which is inherently reliant on surface topography or shape. A simple hand-held measurement tool called the Scoliometer was introduced by Bunnell (1984), which is based on the forward bending test described by Adams (1882). While it is not directly correlated to Cobb angle (Bunnell, 1984), it is still used in many locations throughout the world for screening tests. The Scoliometer provides a fast and affordable measurement method to make an initial assessment, as it is usually performed by a nurse or trained layperson as opposed to an orthopaedic surgeon.

A wealth of visual scanners that try to quantify either the entire torso or just the posterior surface have been reported (Theologis et al., 1997; Sakka et al., 1995; Poncet et al., 2000; Pazos et al., 2007; Rankine et al., 2012; Gorton et al., 2012). For any such visual measurement system, there are two generic components involved: a method to capture the surface topography; and a set of indices to quantify the shape that was captured. Patias et al. (2010) and Kotwicki (2008) have published thorough review papers on this topic, and their general conclusions regarding all current ST systems are summarized here. Patias et al. (2010) and Kotwicki (2008) both noted that while a direct correlation between novel ST indices and the gold standard of the familiar Cobb angle is desirable as it places the index in terms of a known quantity, as of yet no ST technique has been able to do this reliably. Currently, there are correlations between ST indices and the Cobb angle in the general sense that the more severe the Cobb angle the greater the surface deformity. However, the large standard deviations in measurements within the AIS population preclude a direct relationship between the two. Additionally, factors other than the Cobb angle such as age and curve type influence torso deformity. In younger children a weaker relationship between surface and spinal deformity is reported (Grivas et al. 2007). Similarly, patients with double curves present a significantly reduced trunk deformity when compared to thoracic or thoracolumbar curves of similar magnitude (James, 1954). Because ST measures a much more complete, and different, aspect of the scoliotic deformity than the Cobb angle, it is questionable whether a direct correlation between the Cobb angle and ST indices is desirable, let alone feasible. As stated by Kotwicki (2008), “When debating on the role of the surface topography in the evaluation of the body morphology in children with idiopathic scoliosis, one should begin with

rejecting the dogma of the radiological Cobb angle, as the only gold standard for scoliosis evaluation" (p. 747).

The significant benefits of surface topography in the evaluation of scoliosis, namely that it is non-invasive, non-radiating, and quantifies the aspect of the deformity most concerning to the patient, have recently led to considerable interest in these techniques as a potential solution for understanding and monitoring this complex condition.

1.4 Project Overview

The motivation for this project is to develop a non-invasive clinical tool for the diagnosis and monitoring of AIS patients that minimizes the exposure to radiation. By minimizing the harmful radiation associated with the current clinical practice, such a system would provide a safer method to diagnose and monitor the disease, while still allowing the frequent examination required to make timely treatment decisions. Additionally, it is desired that this clinical tool provide a 3D description of the deformity, to more completely quantify the true nature of the disease.

With these considerations in mind, the Scoliosis Research Group at the University of Calgary (U of C) is pursuing a system based on the surface topography of the torso of the patient. The overall hypothesis associated with this approach is that the skeletal deformations of the spine and ribcage will be reflected by changes in the surface of the torso.

The development of the scoliosis measurement system has been ongoing at the U of C since 1998. Numerous advancements have been made since the project's inception. The pertinent developments will be further described in the following sections.

1.4.1 Surface Topography Imaging System

In the first generation of the imaging system developed by the Scoliosis Research Group at the U of C, the torso geometry was captured by a laser scanner system. This system took approximately 15 seconds to perform a complete torso scan, with a typical accuracy of 1 mm and spatial resolution of 6.7 mm vertical row separation and approximately 1.5 mm horizontal point separation (Poncet et al., 2000; Jaremko, 2001). Due to the time required to capture the 3D data, issues such as postural sway had the potential to negatively affect the accuracy of the reconstructed torso. A faster data acquisition would increase the vertical row separation of the collected 3D surface points. In 2003 the Scoliosis Research Group upgraded from the laser scanner system to an optical imaging system manufactured by InSpeck (InSpeck Inc, Montreal; currently owned by Creaform, Lévis), which remains in use today. With this second generation system, a full torso can be captured in approximately 2.8 seconds, greatly mitigating the effects of postural sway. The system provides a typical accuracy of 1.42 mm and a spatial resolution of under 1 mm in all three orthogonal directions (Robu, 2006).

This system is composed of four optical digitizers, two Mega 3D Capturor LF and two DF Capturor II digitizers, as well as a patient positioning frame. These digitizers are arranged to capture the front, back, left, and right sides of the patient's torso.

Additionally, there is an option of obtaining stereo X-rays of the patient while remaining in the positioning frame. This enables a registration between the external ST data and the internal bony geometry. The details of the operating principles of the InSpeck imaging system, along with an analysis of its performance can be found in Pazos et al. (2005) and Robu (2006). To briefly summarize, the digitizer projects a periodic fringe pattern generated by a halogen lamp and grating slide, and captures four images of the subject, shifting the fringe pattern between images. This procedure allows for the calculation of a phase function for every pixel in the image, based on the intensity of the pixel in each of the four images. There are also a number of distinguishable features projected simultaneously on the subject. These features allow for the use of both phase shifted Moiré techniques and active optical triangulation to determine the 3D surface profile of the acquired image.

Prior to data collection, a research nurse places visible markers on seven anatomical landmarks of the subject: left and right back dimples defining the posterior superior iliac spine (PSIS); the vertebra prominens, assumed to be the midline of the T1 vertebra; bottom of the left and right ribcage; and the bottom and top of the sternum. These markers are used to define a body-embedded coordinate system based on the subject's own anatomy. The medial-lateral (ML) axis is defined by the line connecting the two PSIS markers. The origin is at the midpoint of these two markers. The global vertical defines the reference vertical (V) axis. The anteroposterior (AP) axis is formed through the vector cross-product of the ML and V axes and points anteriorly to complete the right-handed coordinate system.

1.4.2 Surface Topography Indices

Once the 3D geometry of the torso is captured, a method is required to consistently quantify the shape. This is achieved through a set of ST indices. Sixty-five evenly spaced transverse (horizontal) cross sections are taken through the torso from the marker at the vertebra prominens, to the midpoint between the PSIS markers, assumed to be the level of the spinous process of S2 (Moore, 1980). This yields on average, 3.6 cross sections per vertebral level. Eleven basic geometric measures are then made on each cross section, of which the complete details have been previously described (Jaremko et al., 2002; Jaremko, 2001; Robu, 2006; Swanson, 2008). These measures are described briefly here. The first moment of area is used to find the centroid of the cross section (Figure 1.5). The lateral coordinate of the location of the centroid defines the (1) *lateral centroid line*. The AP coordinate of the centroid location defines the sagittal centroid line and the most posterior value along the entire torso is taken as (2) *kyphosis*. The principal axes of inertia for each cross section are found by determining the major and minor axes that would result in a product of inertia equal to zero. The angle formed between the cross section minor axis and the ML reference axis was defined as the (3) *principal axis orientation (PAX)*. The (4) *eccentricity* of the cross section was taken as the eccentricity of an ellipse with the same moments of inertia as the cross section. Dual-tangent points of the back surface are found by determining the most posterior point of the surface on both the right and left sides of the centroid, as sectioned by a line parallel to the AP axis. The angle of the line connecting these two dual-tangent points with respect to the ML axis was defined as (5) *back surface rotation (BSR)*. The difference in AP displacement of the dual-tangent points defined (6) *rib prominence*. The (7) *spinous process line* (SP

line) was manually approximated to follow the spinous processes of the vertebrae. This was achieved by visually selecting the point that best represented the location of the spinous processes, typically by taking the most anterior point between the dual tangent points of the back (common in the thoracic spine), or the posterior bulge caused by the protrusion of the spinous processes (common in the lumbar spine or between the scapulae).

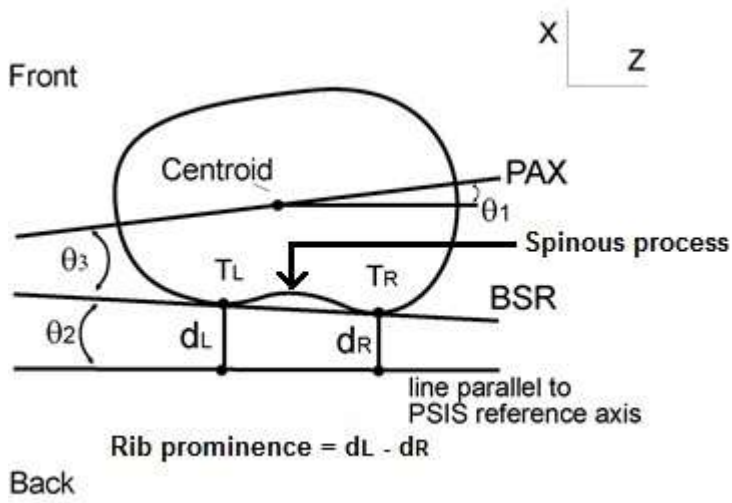

Figure 1.5: ***Back surface rotation (BSR)*** **and** ***principal axis rotation (PAX)*****. θ_1 = PAX rotation with respect to the patient PSIS reference axis. θ_2 = BSR, measured using the line joining left and right dual-tangent points T_L and T_R. θ_3 = difference between BSR and PAX rotation.** ***Rib prominence*** **= $d_L - d_R$, the difference in distances from left and right dual-tangent points to the reference axis. (modified from Jaremko, 2001, p. 59)**

The remaining indices were determined by dividing the cross section into quarter and half areas, respectively (see Figure 1.6 and Figure 1.7). For this study, sectioning was done with lines parallel to the reference axes but running through the centroid. The (8) *quarter area difference* was defined as the difference between the left and right rear quarter

sections, divided by the sum of the rear quarter section areas. The following three indices of left/right asymmetry were normalized such that they were dimensionless with an expected mean of zero in normal subjects. They followed the generic form of:

$$\boldsymbol{asymmetry} = \frac{(\boldsymbol{L} - \boldsymbol{R})}{\left(\frac{\boldsymbol{L} + \boldsymbol{R}}{\mathbf{2}}\right)} \qquad \textbf{Eq. 1.1}$$

The moments of inertia were calculated for the right and left halves separately, and the asymmetry between them in both the lateral and AP direction were defined as (9) *lateral inertia* and (10) *AP inertia* respectively. Finally, (11) *aspect ratio* was the asymmetry between the ratio of the half section AP range ($dAP_{R\ or\ L}$) divided by the half section lateral range ($dLat_{R\ or\ L}$).

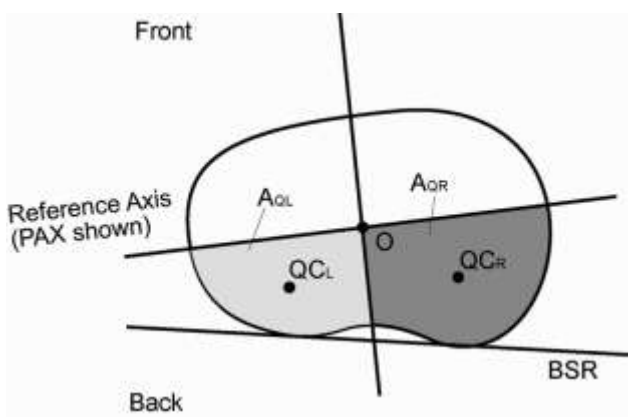

Figure 1.6: Quarter-areas for index calculation.
Rear quarter-areas (A_{QL}, A_{QR}) and their centroids (Q_{CL}, Q_{CR}), defined by quadrants cut through the centroid (O) parallel and perpendicular to the reference axis (PSIS, PAX or BSR). (Jaremko, 2001, p. 60)

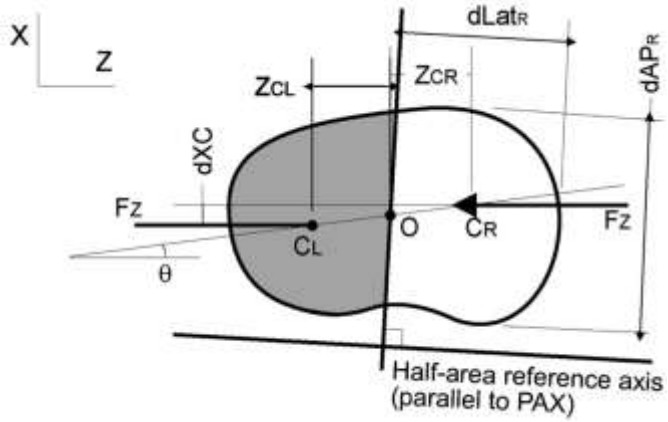

Figure 1.7: Indices defining left-right asymmetry of half-areas.
Half-areas were cut relative to an appropriate reference axis (BSR, PAX or PSIS; PAX shown). Asymmetry of half-centroid locations C_L and C_R was measured in the anteroposterior (dXC) and lateral (Z_{CL} vs. Z_{CR}) directions. θ = angle of rotation of the line joining the half-centroids. dAP/dLat = aspect ratio for each half-area (left and right). Fz = hypothetical unit force applied inward at each half-centroid. In a symmetric torso, the forces would cancel; in a scoliotic patient, a twisting moment proportional to dXC would be generated. (Jaremko, 2001, p. 60)

For this study, 10 of the 11 indices were examined. The ST index of *eccentricity* was not utilized for this study, as Swanson (2008) demonstrated that this index showed little change even between pre/post operative scans. Furthermore, this is the only index of the 11 that doesn't quantify asymmetry or a move towards/away from normal. It is unclear as to what normal *eccentricity* values would be, as the torso is not perfectly circular, and becomes more or less elongated at different thoracic levels.

Figure 1.8 details the steps taken to obtain 3D torso geometry of a subject and then process the data to obtain the final ST indices.

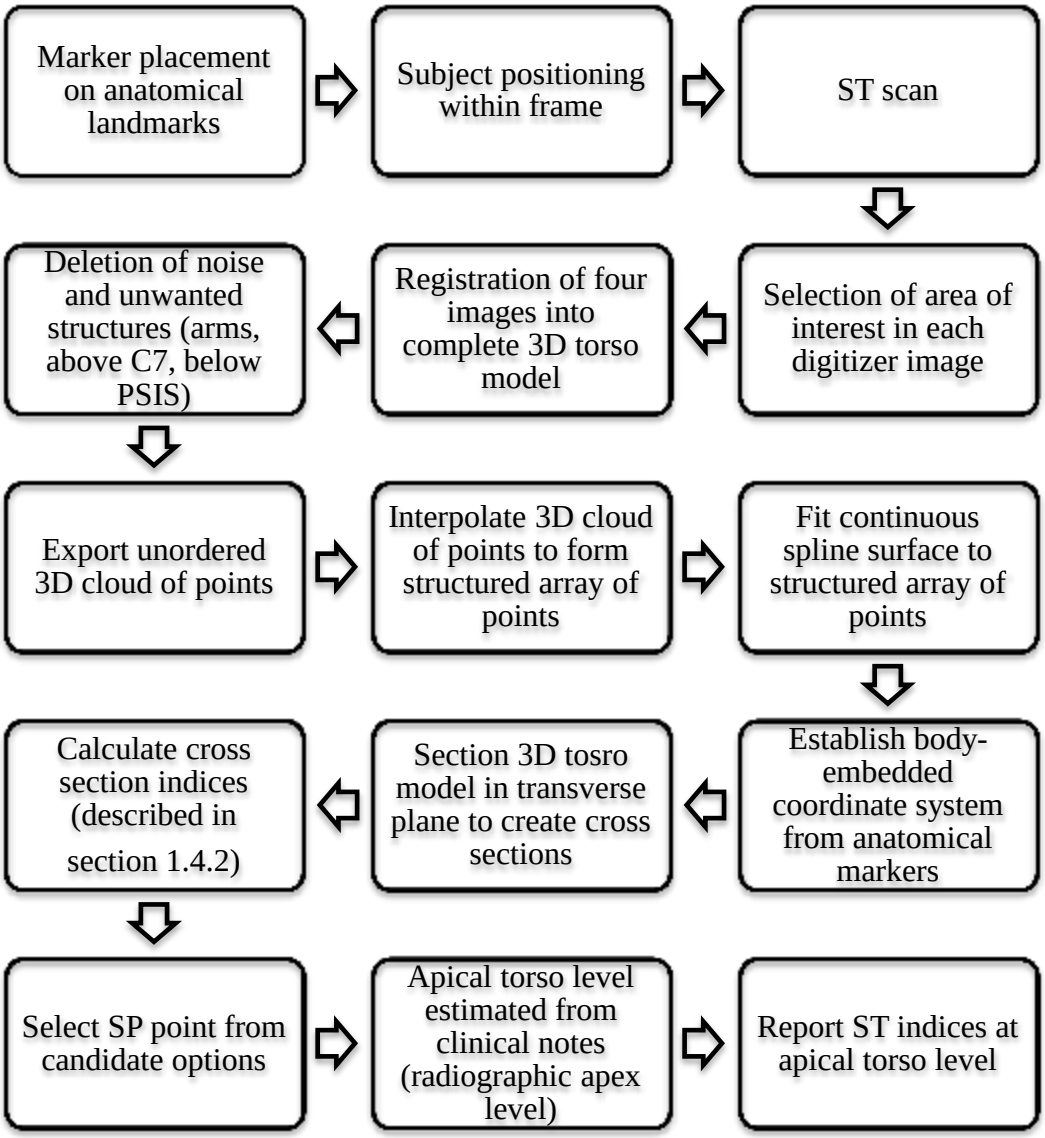

Figure 1.8: Data acquisition and processing protocol

1.4.3 Sources of Variability

With the development of any measuring system, it becomes critical to quantify the accuracy and precision of the measurements such that appropriate interpretations of the recorded measures can be made. As there are inherent errors in any measurement system, it is important to examine where potential sources of error can be introduced in the data collection procedures.

Error can broadly be divided into systematic, random, or gross (blunder) errors. Systematic errors are typically harder to quantify as they are always present and will affect repeated measurements in approximately the same way. Systematic errors can often be described by a deterministic model, however they must first be identified. Systematic errors are typically quantified through the use of a calibration object with known measurements. This procedure has previously been performed on the current ST system with the InSpeck digitizers by Robu (2006) and will not be re-examined in the current thesis. Robu (2006) reported the systematic errors in marker position with the current digitizer setup had a root mean square (RMS) error of 1.42 ± 0.05 mm.

Random errors can be described by a stochastic model and are quantified statistically from multiple measurements. While some attempts to quantify random errors in the InSpeck system have been performed previously in our group (Robu, 2006; Swanson, 2008), these efforts typically focused on the variability in location of markers on a collected surface. What is truly needed to determine the clinical usefulness of the system is determination of the variability in the ST indices themselves. Additionally, an upgrade

in InSpeck software (FAPS 7.5 and EM 6.1, November 2009) provided new advanced methods for the 3D torso reconstruction that may yield superior results over the previous studies. It was therefore deemed prudent to perform a more exhaustive analysis of this important topic.

Gross or blunder errors are measurements that are considered to be mistakes, or measurements that do not belong to the same set of data as the rest of the measurements. Therefore, gross errors should be detected and then rejected prior to processing the data such that these errors do not influence the outcomes of the measurements (e.g. rejection of outlier data points prior to fitting a surface to the torso data). With the current system, gross errors are treated in the InSpeck software (EM 6.1) but due to the proprietary nature of the software, the specific outlier detection algorithm is unknown.

For this system, the sources of variability in the ST indices can broadly be grouped into those occurring before data collection and those occurring after data collection. However it should be noted that it is extremely difficult to identify every source of error in a measurement. Therefore the points listed below are the more prominent sources of variability, but this does not comprise an exhaustive list.

1. Sources of variability before/during data collection
    - Posture within positioning frame
    - Marker placement on anatomical landmarks
    - Motion artifacts between subsequent fringe images (could be caused by postural sway or inhaling/exhaling)
    - Lighting conditions
    - Alignment of digitizers with respect to each other and the subject

2. Sources of variability after data collection
    - Selection of the area of interest in the field of view of each digitizer
    - Deletion of unwanted noise or structures such as arms or head from the 3D torso data
    - Registration of the individual surface models from each of the four cameras into a single surface model
    - Selection of visible markers on the subject's torso
    - Resolution errors due to pixel size
    - Interpolation errors when fitting continuous surfaces or lines to collected 3D torso data for ST index calculation
    - Machine or rounding errors in numerical calculations for ST index calculation
    - Numerical instabilities (e.g. a cross section with eccentricity of zero, or a perfect circle, has an infinite number of principal axes) for ST index calculation

The random variability in the ST indices is mainly influenced at the stages of processing that require operator input. However, it is believed that not all sources of variability will affect the ST indices in the same manner. For example, the final ST indices are calculated on a transverse cross section. Therefore, the sensitivity of the ST indices to variability in placement of the PSIS markers, which define the reference ML axis, will be relatively minor. This is the result of several facts. First, the variability in AP position of the markers is well constrained by the back surface, which is relatively flat around the PSIS markers. Second, the variability in the ML position of the markers will have no effect as this still defines the same reference line between markers. Finally, the variability in the V position of the markers will have minimal effect as the projection of the reference axis onto a transverse cross section would not depend on the elevation of the axis in the coronal (ML-V) plane. However, changes in the V position of the markers would alter the location of the transverse cross sections. All of the ST indices are calculated automatically once the 3D torso geometry is input into our custom software, with the exception of the SP line. The SP line requires user input to select the closest candidate point to the anterior point that follows the depression along the spinous processes of the back (Figure 1.9).

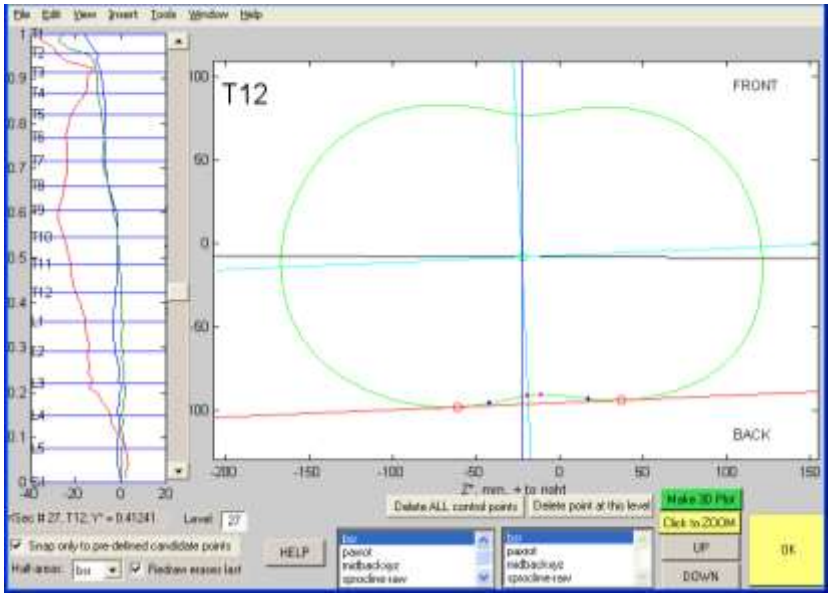

Figure 1.9: Selection of spinous process line from transverse torso cross sections

1.4.4 Application of System

The ST imaging system was developed to improve the clinical monitoring of AIS. As the ST system has developed through the stages of proof-of-concept towards clinical implementation, the application of the system has also evolved. The brief historical development will be outlined. In the first generation of this ST system, the objective was to correlate the ST indices with the radiographic Cobb angle. This was accomplished with limited success, but not to a clinically acceptable degree (Jaremko, 2001). Swanson (2008) investigated the change in ST indices prior to and following surgical correction. The goal of using this patient group was to elucidate the relationship between the Cobb angle and the ST indices that the dramatic change in Cobb angle realized between these two conditions may provide. However, the study results required verification for applicability to the natural history of AIS. Motivated by concerns related to the large standard deviations in ST measurements within the AIS population (section 1.3.3),

further development of this system was directed to not necessarily replace X-rays with ST measures, but to supplement them to reduce the required number of X-rays associated with monitoring of progress. This was done by investigating the relationships between *changes* in ST indices to *changes* in Cobb angle. By doing so, each patient acts as their own baseline, hopefully greatly reducing the large discrepancies found between the magnitudes of Cobb angle versus ST indices. If successful, such a system would require an initial X-ray to determine a baseline Cobb angle, then subsequent monitoring could be done with ST techniques until there was indication that the changes in ST indices had progressed to a clinically significant level (associated with change in Cobb angle equal or greater than 10°). Only then would another X-ray be taken, greatly reducing the lifetime exposure of AIS patients. While a classification rate of 92% success between progressed (change in Cobb angle equal or greater than 10°) and non-progressed groups (change in Cobb angle less than 10°) was achieved with a discriminant analysis (DA) on the ST indices (Swanson, 2008), this result proved to not be sufficiently robust on more detailed analysis. When different random selections of patients were used for training and testing of the DA, the classification rate varied widely and fell far below a clinically acceptable level. This finding may be partly attributed to some unaccounted dependencies in the study data or to the definition of progressed and non-progressed groups. The current thesis aims to further our understanding of the system performance to contribute to the goal of developing a clinically useful surveillance tool for the management of AIS.

1.5 Objectives/Hypothesis/Specific Aims

The objective of this specific project within the system's development is to quantify the random variability in the ST indices introduced after data collection, and to evaluate these values in the context of a clinical data set. This is part of the larger project goal to verify and assess the ST system's capability for clinical application. The thesis objectives are accomplished through completing the following specific aims (SA).

SA1: Quantify the intra, inter, and total-observer variability in ST indices calculated from the same set of input scans.

SA2: Quantify the change of ST indices over time in a clinically progressed AIS group.

SA3: Quantify the change of ST indices over time in a non-progressed AIS group.

SA4: Qualify the intra and total-observer variability in ST indices for clinical acceptability through comparisons to changes of ST indices over time in a clinically progressed AIS group.

SA2 and SA3 are developed to test the following hypothesis (H).

H1: The ST indices will demonstrate significantly more change in a clinically progressed group of AIS patients than a non-progressed group of AIS patients.

1.6 Thesis Organization

This thesis is organized into four chapters. Chapter two focuses on addressing SA1, and details the methods used and results found to quantify the intra, inter, and total-observer variability of the ST indices of interest. Chapter three addresses H1 as well as SA2 – SA3. It details the methods used to examine the changes in ST indices over time, and compares the clinically progressed and non-progressed groups. Chapter four addresses SA4 by providing a detailed discussion relating the results from chapters two and three, the clinical implications of these results, as well as potential future work to further the advancement of the scoliosis imaging system in use by the Scoliosis Research Group at the U of C.

CHAPTER TWO: DETERMINATION OF VARIABILITY IN SURFACE TOPOGRAPHY INDICES

2.1 Introduction

Clear quantification of the variability associated with data processing is an essential step towards validating the clinical use of surface topography as a tool for the detection and monitoring of scoliosis. A thorough understanding of the precision of the measurement system and the variability introduced into the measurements through processing the data is required to establish the level of confidence in the measurements and how this informs the clinical decisions based on those measurements.

Central to the method of quantifying variability is the nature of the data used. The initial goal of the Scoliosis Research Group was to develop a system that utilized surface topography to quantify the scoliotic deformity at a given moment in time, thus replacing X-rays in the diagnosis and monitoring of AIS (Jaremko et al., 2001; Jaremko, 2001; Jaremko et al., 2002). More recently, the potential advantages of applying the technology to detect the change in surface topography in a patient over time have been identified (Swanson, 2008). Specifically, this approach enables the patient to be used as their own baseline for comparison, thus eliminating patient-to-patient variability and providing a better measure of the progression of the scoliotic deformity over the period of time between ST scans. For this purpose, the choice of method to quantify variability should reflect the fact that a *difference between two calculated values* would be used to make

clinical decisions. It is critical to have an a priori understanding of a measurement's variability to be able to have confidence in decisions based off of the measurement.

This chapter provides a description of the statistical approaches used to quantify the variability in the ST indices. Two measures of variability are presented and critiqued as to their limitations and applicability to the intended use of the ST indices. The specific form and calculation of these measures are described. The study design and data used to calculate these measures are detailed. Finally, the results are presented and briefly discussed. The further discussion of these measures of variability with respect to their clinical applicability is elaborated on in Chapter Four.

2.1.1 Terminology

There are a large number of related terms when discussing variability of measurements. These terms are applied inconsistently within the orthopaedic literature. Therefore, clear definitions are presented here and applied consistently throughout the thesis. When references are made to the literature, their results will be placed in the context of these definitions, regardless of the terminology used in the original document, such that accurate comparisons can be made.

Variability is used to describe the variation of measurements made on the same quantity. Variability can be caused by multiple factors. This term is chosen in preference to the term *error*, as error implies a difference from a true value. While there is a true value for torso shape or spinal alignment at a given instant in time, and errors will certainly be

introduced in the process of measuring ST indices or Cobb angle, there is not a true value (either known or unknown) for torso shape or spinal alignment in a general sense. With the vast number of degrees of freedom between the alignment of anatomical structures including vertebrae, ribs, and muscles, everyday occurrences such as inhaling/exhaling deeply or posture will influence the 'true' value of the ST indices or Cobb angle. As the errors introduced by the measurement system will present themselves in the same fashion as the variation of the actual recorded shape at a specific instant, *variability* will be used as it is the more encompassing term. The term *precision* is used in this thesis as well, with the same concept as *variability*, in that it quantifies the variation of measurements made on the same quantity. However, *precision* has the inverse sense to *variability* (i.e. high variability equates to low precision and vice versa). Figure 2.1 illustrates this concept of precision and variability, where x represents the measurement of a given quantity, and f(x) is the limiting distribution representing the hypothetical distribution of infinite repeated measures (Taylor, 1997).

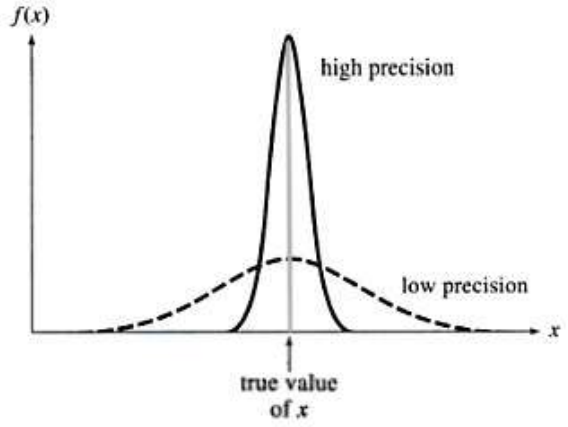

Figure 2.1: Two limiting distributions, one for a high-precision measurement, the other for a low-precision measurement (modified from Taylor, 1997, p. 129)

In this thesis, variability is quantified with respect to the *observer*. This term implies that any variation recorded is introduced by the people and process involved in recording the data and calculating a value from said data. Other terms often found in the literature that are used in the same context are *rater, operator, user*. No tests were performed to quantify variability introduced by the subject, which could include differences in posture, day-to-day weight gain, movement artifacts, or normal growth. This study focuses on the sources of variability introduced after data collection (section 1.4.3), and not those introduced before data collection.

Variability is further qualified by the terms, *intra-observer, inter-observer, and total-observer*. *Intra-observer variability* quantifies the variability in a measurement performed by an individual observer. It is estimated by taking repeated measurements while maintaining all conditions constant, such as operator, equipment, and procedure. Other terms sometimes used in the literature to define this aspect of variability are *repeatability* and *reliability*.

Inter-observer variability is used to quantify systematic differences between observers. While this definition is consistent with the statistical definition of inter-observer variability, it is different than what is commonly described in the orthopaedic literature. For example, it could be possible that a certain repeated measure showed large variability between trials for all of the multiple observers (high intra-observer variability). Yet if the means of the repeated trials were very similar between observers, the inter-observer variability would be very low (and lower than the intra-observer variability), indicating

that there are negligible systematic differences between the observers. This approach differs from the common orthopaedic description of "inter-observer" variability, as used by Morrissy et al. (1990) that typically reports a worst case scenario – where a test is being performed by different observers and with the assumption that a single observer would be more consistent than multiple observers – and quantifies the *total variability*. In this common orthopaedic situation, the reported "inter-observer" variability is always larger or equal to the intra-observer variability, as it is a summation of the intra and inter-observer variability (i.e. the variability introduced individually by those taking the measurements and the systematic differences between them). The differentiation of inter-observer variability as defined for this thesis and as it is sometimes used in the literature is made clear to avoid misinterpretation. Otherwise, a statement such as "variable X has low inter-observer variability" could provide the misunderstanding that readings will always be very consistent regardless of the observer. With the current definition, the proper interpretation is that there are low systematic differences between the observers of this system. This leads to the third qualification of variability used to describe the summation of effects.

Total-observer variability quantifies the cumulative effects of both *intra* and *inter-observer variability*. For the clinical application of our imaging system, this will likely be the most important measure of variability. In a hospital setting it is not always possible to ensure the same observer (e.g. physician, nurse, or trained technician) would be able to work on all scans for a particular patient. Additionally, as clinical implementation of this system expands, it would be advantageous to be able to compare

and combine studies from different medical centres. This quantity is sometimes referred to in the literature as *reproducibility*, or as mentioned above, it is sometimes mislabelled as inter-observer variability.

2.2 Background and Literature

This section provides background information on the variability of the current clinical standard measure of scoliosis severity, namely the Cobb angle. The variability of the Cobb angle influences, to some extent, what is defined as clinical progression of AIS. This definition becomes critical in choosing how to qualify the variability of the ST indices. Additionally, the previously reported variability of the ST system is presented and the areas where further study is required for clinical implementation of the system are identified.

2.2.1 Variability of Cobb Angle

The variability of the radiographic Cobb angle has been extensively reported (Morrissy et al., 1990; Carman et al., 1990; Tanure et al., 2010; Aubin et al., 2011; Srinivasalu et al., 2008; Kuklo et al., 2005). There are several methods to quantify variability, which will be further described in section 2.3. While an effort is made to place the reported Cobb angle variability within the context defined above, not all authors report comparable measures of variability.

The seminal works on Cobb angle variability were performed by Morrissy et al. (1990) and Carman et al. (1990). Morrissy et al. (1990) showed a 95% confidence interval of 4.9° and 7.2° for intra-observer and total-observer variability respectively, on individual measurements of Cobb angle without preselected end-vertebrae. Carman et al. (1990) reported on the expected difference between Cobb angles measured between two time points, as that value is what clinical decisions are based upon. This study also did not define preselected end-vertebrae. Because each measurement has its own variability, the variability of the difference is larger than the variability of an individual reading by a factor of √2 (Carman et al., 1990; Weir, 2005; Stratford 2004). Additionally, the authors utilized a technique called tolerance limits (Carman et al., 1990; Remington & Shorck, 1970) to determine the limits within which 95% of future differences would fall, with a 95% confidence interval given their sample size. With these techniques, they reported values of 9.6° and 10.1° for the intra-observer and total-observer cases, respectively. The proper interpretation of these values, for example in the total-observer case, was stated in their paper as follows: "in the absence of any true change, one can be 95 per cent confident that 95 per cent of the time one observer's reading will be no more than 10.1 degrees more or less than the other observer's reading due to observer error alone" (Carman et al., 1990, p. 331).

More recently, the variability of Cobb angle measurements have been re-examined with the introduction of newer technologies such as digital X-rays, virtual protractors, and other computer aided techniques. Tanure et al. (2010) compared the manual method of computing the Cobb angle with a semi-automated process on digital images. Designed to

represent clinical settings, the radiograph data sets were drawn from patients with a diagnosis of idiopathic scoliosis. No restrictions were placed on age, curve location, or curve magnitude. Additionally, the end-vertebrae of the curves were not preselected. The semi-automated process involved the selection of each extremity on the vertebral plateaus defining the end-vertebrae. No statistically significant differences were found between the manual or digital methods of calculating Cobb angles. For both methods, they reported intra-observer mean absolute differences ranging from 2.06° – 3.46° and standard deviations of 1.69° - 2.73°. The total-observer mean absolute differences ranged from 3.61° - 3.85° with standard deviations of 3.18° - 3.45°. Aubin et al. (2011) found a standard deviation of 4.9° in the total-observer variability in the main thoracic region using a semiautomatic software system on adult scoliosis based on conventional X-ray films, scanned into the software. Srinivasalu et al. (2008) reported an average intra-observer variability of 1.3° and inter-observer variability of 1.26° for a data set of digitally acquired images over a wide range of ages. However, this result does not represent a realistic clinical situation as the end-vertebrae were preselected in their study, removing a large component of variability. Additionally, the authors do not clarify whether the line across the vertebral end-plates is drawn within the software or manually by the observer.

2.2.2 Variability of Surface Topography System

The variability associated with the InSpeck system implemented at the University of Calgary has been quantified and reported previously (Robu, 2006; Swanson, 2008). Robu (2006) quantified the variability in the system by examining the root mean square

(RMS) error in location of physical markers on a stationary scoliotic mannequin. Marker locations measured with the InSpeck system were compared to those obtained with a coordinate measuring machine (CMM) with a known accuracy of ± 0.5 μm. Based on these measurements, using the current system orientation, the intra-observer RMS error was 1.29 mm with a standard deviation of ± 0.45 mm. The maximum mean differences decomposed into X, Y, and Z constituents were under 0.4 mm and 0.5 mm in the intra-observer and inter-observer cases, respectively. Additionally, Robu (2006) examined the intraclass correlation coefficients (ICC) (detailed in section 2.3.1) of the ST indices with three groups: a scoliotic mannequin, five scoliotic patients, and five normal subjects. These tests showed a very high ICC of 0.979 for the mannequin, but dropped substantially to a range of 0.541 – 0.891 with the other two groups. The repeated measurements conducted for this study involved subject repositioning and subsequent reimaging of the torso shape. Therefore, any variability described by these tests constitutes a combination of observer and subject effects (i.e. variability introduced before and after data collection, as listed in section 1.4.3), with no method to delineate between the contributions of each error source. The measures reported by Robu (2006) were helpful as a first step in understanding the InSpeck system. However, they are limited in their use for clinical interpretation (i.e. how the reported variability is realized in the final ST indices and how this influences a clinical decision based on these indices).

Swanson (2008) continued this work and carried the quantification of variability through to the ST indices. Intra-observer variability quantified (Table 2.1) along the entire length of the torso, increased from the inferior to superior direction up the torso. The superior

regions of the 3D torso model are presumably more affected by observer influence as the arms are removed from the model through manual selection and deletion. As a continuous surface is fit to the remainder of the model, the exact location of the cut-lines through the arms affects the torso cross section. Therefore, high variability at levels above the arms may adversely affect these results.

Table 2.1: Summary of mean and maximal intra-scan standard deviation (SD) for each index curve. The standard deviations are also given as a percentage of the maximum index value of an example subject to illustrate the magnitude. (Swanson, 2008, p. 83)

	MEAN SD	MAX SD	MEAN SD AS % OF MAX INDEX VALUE	MAX SD AS % OF MAX INDEX VALUE	UNITS
BSR	0.12	0.28	1.60	3.73	degrees
PAX	0.18	0.61	1.68	5.68	degrees
RibProm	0.30	0.58	1.86	3.60	mm
QuarterArea	0.72	1.18	2.74	4.49	dimensionless
AspectRatio	0.66	2.30	2.23	7.78	dimensionless
LatCenLine	0.37	1.10	1.68	4.99	mm
SpinousProcessLine	1.30	2.83	6.40	13.94	mm
LatInertia	0.52	1.10	3.22	6.82	dimensionless
SagCenLine	0.26	0.38	0.99	1.44	mm
APInertia	0.78	1.20	2.77	4.26	dimensionless

Therefore, the current study contributes by addressing these limitations to extend the variability analysis to include a total-observer component, to focus the quantification of variability to the most clinically relevant areas of the torso, and to present the data in a manner that facilitates easy implementation by clinicians.

Research collaborations with a research group based out of Ecole Polytechnique de Montreal and Sainte Justine Hospital in Montreal have led to their adoption of a very comparable ST system. It is also utilizes InSpeck digitizers, acquires horizontal cross sections through the 3D torso model, and calculates geometric indices from these cross sections. While the entire set of torso indices is not common between both groups, both groups typically report BSR, PAX (termed axial trunk rotation), lateral centroid line, and sagittal centroid line. Pazos et al. (2007) reported ICC and minimum detectable change (MDC – detailed in section 2.3.2) values for two different postures: both free-standing as opposed to inside a positioning frame as is the clinical practice at the U of C. These values are understood to represent intra-observer variability. This study involved patient repositioning between scans, while the anatomical markers remained fixed. Seoud et al. (2012) introduced a novel technique of functional data analysis in which a linear combination of 10 independent basis functions were fit to the index curves plotted against position along the trunk. This approach enables comparison of entire curves as opposed to individual points along the curve. Similarly to the study of Pazos et al., this study also reported intra-observer variability, and involved patient repositioning between scans. Select results from the two studies that are directly comparable to the current study are presented in Table 2.2. Similar to the findings of Swanson (2008), Seoud et al. (2012) reported that variability increases with measurements obtained from the superior portion of the torso. They showed dramatic increases in variability above the lower edge of the shoulders. Consequently, for torso regions below this level, the higher ICC and lower MDC values of the ranges reported in Table 2.2 are more applicable. As the variabilities measured by Swanson (2008) were reported as mean and maximum standard deviations

of the ST indices, and both Seoud et al. (2012) and Pazos et al. (2007) reported ICC and MDC values, a direct comparison cannot be made. The results of the current study will be compared to those presented in Table 2.2 in Chapter four.

Table 2.2: Intra-observer variability from InSpeck based ST system in Montreal

Index	Intra-Observer Variability		Reference
	ICC	MDC	
BSR	0.92 - 0.93	2.39° - 2.48°	Pazos et al. (2007)
PAX	0.97 - 0.97	1.43° - 1.48°	
BSR	0.79 - 0.98	3.3° - 11.1°	Seoud et al. (2012)
PAX	0.80 - 0.98	3.9° - 11.1°	
Lateral centroid line	0.87 - 0.98	8.0 mm - 17.5 mm	
Sagittal centroid line	0.90 - 0.97	7.2 mm - 19.4 mm	

2.3 Measures of Variability

There are numerous statistical methods or quantities that can be used to provide an estimate of variability in a series of measurements. Standard deviation, standard error, 95% inclusion limits, tolerance limits, mean absolute difference, root mean square error, ICC, or MDC can all provide a measure of variability. Reasons to report one quantity over another to express variability should be based on how the measured quantities, in this case ST indices, are being used. While the reported values of variability of Cobb angles and ST indices listed in the previous section may all be technically correct, they are misleading with respect to the intended use. The intent of quantifying the variability

is to allow the clinician to estimate the certainty of his or her measurement within a specified confidence interval, customarily chosen to be 95%. This would be phrased as: "the measured value is within the reported variability of the true value, 95% of the time." When a value of variability is published, stating that the inter-observer reliability of a measurement is X units, it can be reasonably assumed that the interpretation by the clinician is that their recorded measurement is within X units of the true measurement. However, if the quantity X represents a standard deviation, then an interval of ± 1 standard deviation will only cover 68% of data that are normally distributed. If the reported variability is the mean absolute difference, then the clinician will know how much their measurement will differ from the true value on average, but that is of little help to quantify the uncertainty in the single measurement under consideration.

Additionally, if the quantity of interest is the difference of two measurements, then consideration must be given to the fact that there is variability in each of the measurements, and therefore the variability of the difference is compounded. This is the case with the proposed usage of the ST imaging system being developed. It is also often the case when dealing with Cobb angles taken from subsequent X-rays through the longitudinal monitoring of AIS.

The MDC is able to estimate the variability in a difference between two measures. It is largely independent of the population from which it is determined (Weir, 2005; Stratford, 2004) and it can be considered a fixed characteristic of any measure (Nunnally & Bernstein, 1994). For these reasons the MDC is particularly well suited to the intended

implementation of the ST system. As the MDC is a measure of absolute variability, it cannot be used to compare different indices. The ICC is a measure of relative variability, which allows for relative comparisons between different indices. However its relative nature does not allow for an easy interpretation of how precise the measurements are in absolute terms. For these reasons, both the ICC and MDC will be considered in this thesis. Further descriptions along with the advantages and limitations of each statistic are given below.

2.3.1 Intraclass Correlation Coefficients

The ICC is commonly used to quantify the variability of repeated measures (Shrout & Fleiss, 1979; Weir, 2005; Stratford, 2004). The ICC is a measure of relative reliability. It is a unitless ratio of the between-subjects variability and the total variability, which is comprised of the between-subjects variability and error. As such, the lower the variability due to error, the closer this value is to 1.0. The higher the variability due to error, the smaller this value becomes, theoretically approaching 0.0. While there are many versions of the ICC, this basic description holds true for all.

Shrout and Fleiss (1979) defined six different versions of the ICC. While others have added further variations (McGraw & Wong, 1996), an appropriate version for this study was obtained from the definitions of Shrout and Fleiss. The six versions are built on three different models (designated by the first index). Each of the three models can be based on the entries representing individual measures from each observer or an average of repeated measures from a given observer (designated by the second index). Therefore,

the specific version of the ICC is designated ICC(model, # of measures per entry). Within the context of determining intra/inter/total-observer variability, the main differences between the versions are dependent on how the group of observers are treated. In Model 1, each subject is rated by a different group of observers chosen randomly from a larger population. In Model 2, the group of observers is chosen randomly from a larger population, but then proceeds to rate every subject. Finally, in Model 3, the group of observers are the only users of interest (i.e. they represent the entire population of observers and not a random sample of) and rate every subject (Shrout & Fleiss, 1979; Weir, 2005). The decision on the appropriate model dictates whether a 1-way or 2-way analysis of variance (ANOVA) should be performed. For this study, the specific form of ICC used was the ICC(2,1). Further details and associated equations are described in section 2.4.3.

Although the ICC provides a method to compare the variability in dissimilar measurements (e.g. a rotation measured in degrees versus a displacement measured in millimeters), there are several key drawbacks with the ICC that must be addressed. As it is a unitless quantity and a somewhat abstract concept, it requires a solid understanding of the context in which it is being used before an appropriate interpretation can be made. Many variability studies that present an ICC include an unjustified ranking of what constitutes a ‘poor’ or ‘good’ ICC value, or a ranking taken from the literature that was developed for a different context. For example, Pazos et al. (2007) provides no interpretation for an ICC less than 0.90, and ranks values between 0.90 and 0.95 as fair, values between 0.95 and 0.98 as good, and values over 0.98 as very good. On the other

hand, Dang et al. (2005) considered an ICC less than 0.40 poor and equal or greater than 0.75 as excellent. Obviously, such dramatic discrepancies serve to confuse the readers of such literature. Fundamentally, the decision of what constitutes poor, acceptable, and good ICC values is a judgement call based on the context and population for which it was determined.

The relative nature of the ICC dictates that the value is dependent on the between-subjects variability in the data. In practice, what this translates to is that low or "poor" ICCs can be calculated even if the trial-to-trial variability, or error, is low, as long as the between-subjects variability is low. Conversely, very high or "good" ICCs can be achieved with high levels of error as long as the between-subjects variability is high. This can serve to mask poor trial-to-trial consistency (Weir, 2005). As AIS has a very heterogeneous presentation, varying in number, location, severity, and flexibility of the spinal curves, as well as other features of the deformity such as rib prominence, shoulder asymmetry, or waist asymmetry, the selection of the sample population can have a considerable effect on the values of ICC. Therefore, it would be expected that wide inclusion criteria would serve to increase the ICC values regardless of measurement error, while a narrow inclusion criteria such as the study of Morrissy et al. (1990) which only included Cobb angles between 20° and 40° would serve to decrease the ICC even with the same measurement variability. It should be clear that the utility of the ICC is limited to relative comparisons within similar samples (e.g. ST index A shows less variability than ST index B when both values are taken from the same set of subjects) and

not absolute decisions (e.g. ST index A is sufficiently reliable for clinical decision making).

2.3.2 Minimum Detectable Change

The primary drawback with the use of the ICC to quantify variability is its relative nature and dependence on the variability of the subject groups. A more appropriate and useful statistic for clinical interpretation would be measured in absolute terms and be independent of the variability of the subject groups. The statistic that provides these features is the Standard Error of Measurement (SEM). As the intended use of our system relies on taking the difference between two measurements, each of which has an associated error, the SEM is used to directly calculate an MDC based on the desired confidence interval. Since the SEM is simply multiplied by a scalar value to determine the MDC, the two statistics share all of the properties that are highlighted below.

The SEM is a measure of absolute variability. It is measured in the units of the original measurement, in this case the units of the corresponding ST indices. It can be viewed as a typical error associated with that measurement (Weir, 2005; Hopkins, 2000). This makes the SEM easy to understand and very tangible, as opposed to the relative nature of the ICC. Furthermore, the SEM is largely independent of the population from which it is determined (Weir, 2005; Stratford, 2004) and it can be considered a fixed characteristic of any measure (Nunnally & Bernstein, 1994). Thus, the SEM and by extension the MDC, provide measures of variability that are easy to understand and interpret, and will not be influenced by the heterogeneity of the AIS population. It is for these reasons the

MDC was chosen as the primary statistic used to represent the variability of the ST system.

The meaning of the MDC with respect to measurement variability, and how it is used to detect true change in measurements is demonstrated schematically in Figure 2.2 and with the ensuing equations. The black points represent the true value of the measurand at each time point. In general, these true values are unknown. The red points represent the observed measure taken at each time point. When discussing random variability as opposed to systematic variability, as is the case with this thesis, these observed measures are normally distributed about the true measure. Theoretically, if one were to make an infinite number of observations, the mean of these would converge to the true value. The difference between the true values represents the true change score, and the difference between the observed values represents the observed change score.

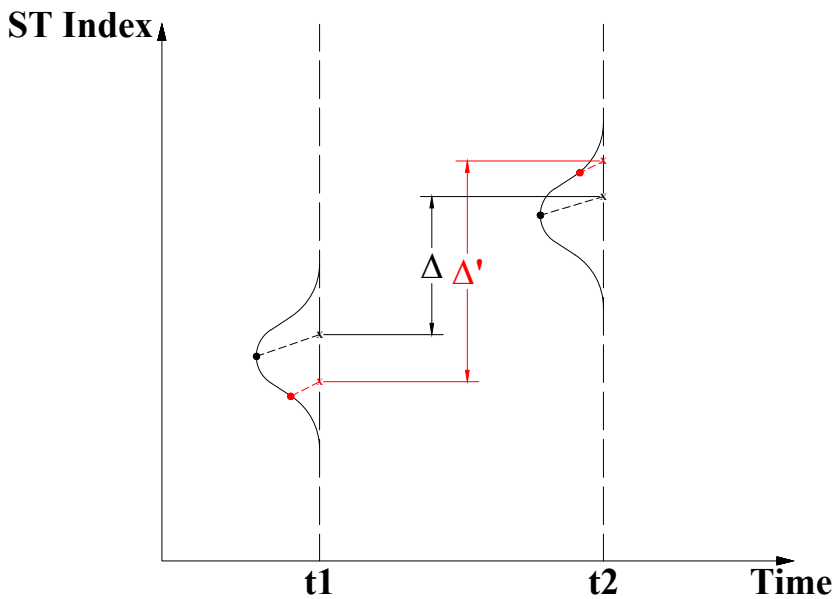

Figure 2.2: Measurement variability

$$\boldsymbol{\Delta}' = \Delta \pm \boldsymbol{\delta}_{random} \qquad \textbf{Eq. 2.1}$$

Where: **Δ'** = observed change score in ST index

Δ = true change score in ST index

$\boldsymbol{\Delta}_{random}$ = random variability in observed change score

In this study, the variability between two measurements is quantified by the MDC, therefore:

$$\boldsymbol{\Delta}' = \Delta \pm \mathbf{MDC} \qquad \textbf{Eq. 2.2}$$

If a true change between measurements has occurred, then:

$$|\Delta| > \mathbf{0} \qquad \textbf{Eq. 2.3}$$

$$|\Delta \pm \boldsymbol{MDC}| > \boldsymbol{MDC} \qquad \textbf{Eq. 2.4}$$

$$|\Delta'| > \boldsymbol{MDC} \qquad \textbf{Eq. 2.5}$$

Thus, to detect a true change in measurements between time points with a given confidence interval, the observed change score must be greater in magnitude than the MDC.

The clinical users at the Alberta Children's Hospital indicated that characteristics of a measurement which allows greatest ease of clinical application include: highly understandable, easy to visualize, and tangible. The MDC satisfies many of these key characteristics, as it is measured in the same units as the ST index it is associated with. However, as the ICC allows for comparisons between ST indices, the two measurements of variability that are evaluated in the following study are the ICC and the MDC.

2.4 Methods

2.4.1 Data Acquisition and Processing Protocol

Figure 2.3 shows the workflow to obtain the final ST indices. Only the steps within the dashed box were repeated for the test-retest design, which correspond to the sources of variability introduced after data collection described in section 1.4.3. This is comparable to the methods used to calculate the variability of the Cobb angle as reported in the referenced studies in section 2.2.1. These studies quantify variability introduced by the observer and measurement system after data collection, whether that be an ST scan or an X-ray. These studies do not quantify variability introduced before data collection, such as the influence of patient posture.

The boxes with bold text in Figure 2.3 indicate the steps where manual intervention is required by the observer. These are the steps most likely to introduce variability into the final ST indices. Figure 2.4 illustrates the data collection process and the projection of the fringe pattern on the subject's torso. Figure 2.5 provides an example of the deletion of unwanted structures from the 3D torso model. The areas in yellow are manually selected for deletion.

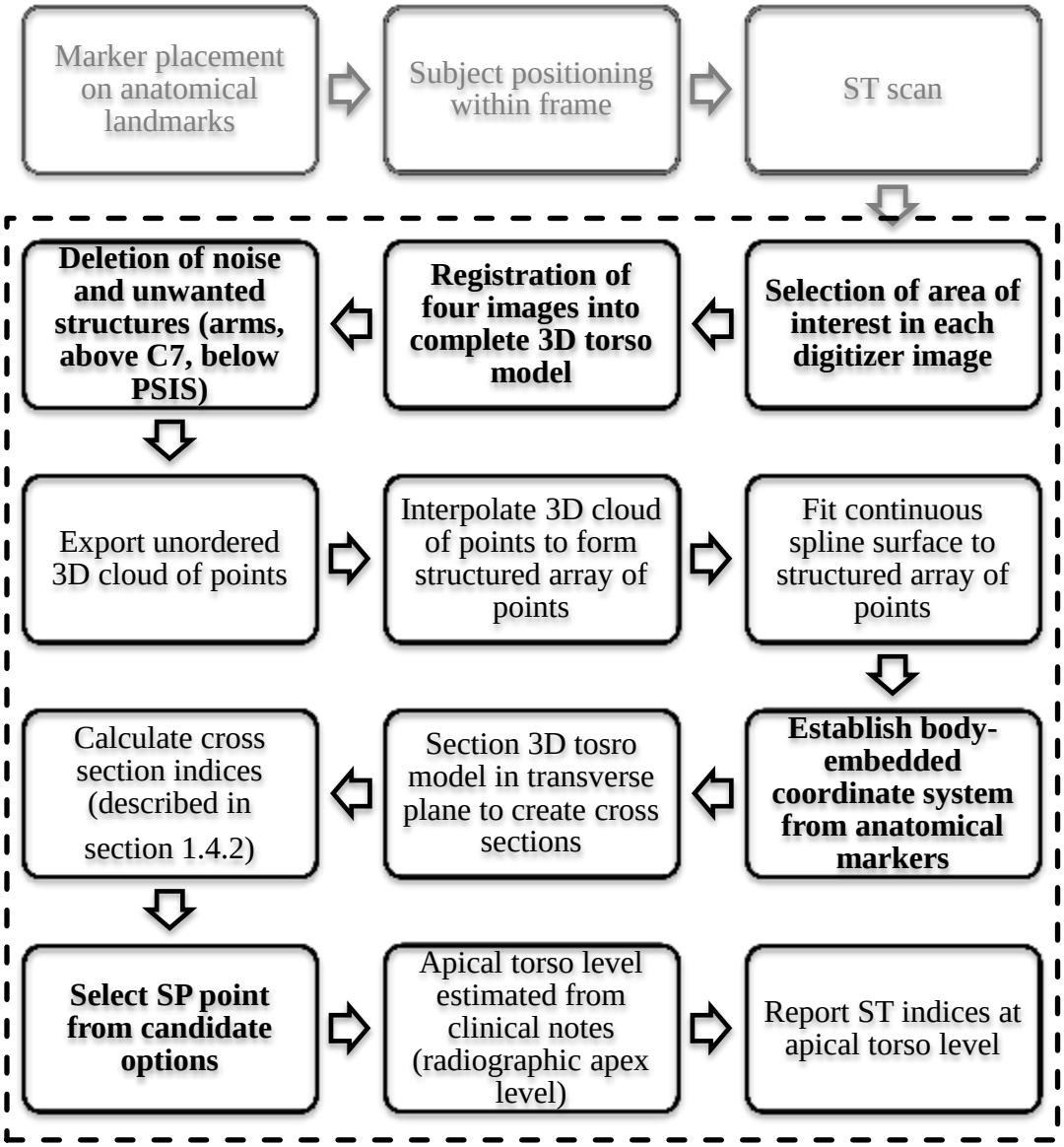

Figure 2.3: Data acquisition and processing protocol

Figure 2.4: Surface topography scan

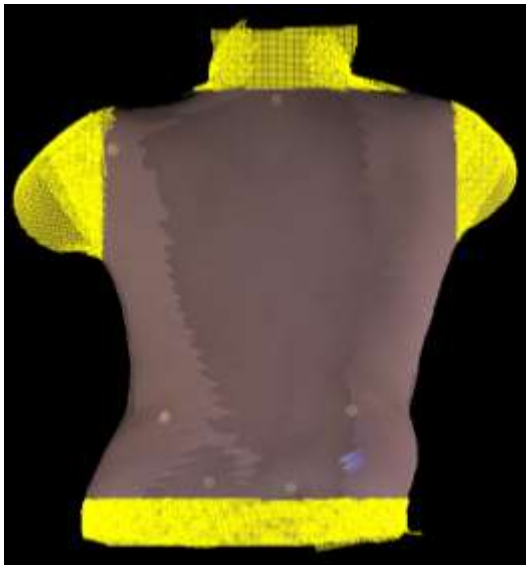

Figure 2.5: 3D torso model – deletion of unwanted structures

2.4.2 Test-Retest Design

The intra, inter, and total-observer variability were quantified with a test-retest study design. All observers participating in this study were well trained in the use of the system. Each observer had knowledge in the background of AIS and the use of surface topography to describe the deformity. Additionally, each observer had specific knowledge and training on the InSpeck system and had processed numerous patient scans prior to this study. For these reasons, no learning effect was anticipated or considered throughout the course of this study. Each trained observer processed the same set of patient scans three times. Originally, there were three trained observers and 15 patient scans in the study design. However, due to a bug with a software upgrade, one of the observer's data sets was processed improperly. Unfortunately, this member had left the Scoliosis Research Group before the data could be reprocessed properly. Additionally, two of the patient scans were discovered to be corrupted and unable to be processed. This left the following parameters for the test-retest study:

- Number of observers, $k = 2$
- Number of repeated trials, $m = 3$
- Number of subjects, $n = 13$

For this study, it was desired to obtain singular values to quantify the variability of each ST index. By doing so, this would facilitate the use of classical statistical tests that are performed in Chapter 3. Additionally, as shown by both Swanson (2008) and Seoud et al. (2012), the variability in the ST indices increases dramatically in the region of the torso superior to the lower edge of the arms. It was therefore desired to quantify the

variability below this level to obtain the variability associated with processing data and not an artifact of where the arms were cut from the torso model. For these reasons, all ST indices other than kyphosis were evaluated only at the level of the apex of the primary curve as determined by X-ray. As described in section 1.4.2, kyphosis is taken at the most posterior portion of the torso. The apex levels were provided from clinical X-rays taken during routine monitoring at the Alberta Children's Hospital. The level of the apex was chosen as it represents the region of greatest interest. The apex also provides a repeatable position along the torso at which to calculate the ST indices.

It should be noted that for the future implementation of the ST system, ST indices will be calculated at all levels of the torso. It is believed the additional information provided by areas other than the apex will prove to be useful in detecting progression of scoliosis.

2.4.3 Determination of Statistics

The version of ICC chosen for this study was what Shrout and Fleiss (1979) designated ICC(2,1) (section 2.3.1). The reason for this decision was based on the experimental design and the intended clinical implementation of the ST imaging system. In the current experimental design, each observer processed every subject, thus eliminating Model 1. Furthermore, this imaging system is ultimately intended to be implemented in multiple clinical centres. Therefore, the group of observers in this experiment must be considered as a sample of a greater population of observers, unless the intent is to process all patient ST scans in one facility, which is not the case. This dictated the use of Model 2. Finally, although taking an average of multiple scans would provide more accurate results as they

converge to the true value, this is not logistically possible in a clinical setting where time and resources are tightly controlled. The system is being designed such that a single scan is adequate for clinical implementation. This situation thus dictates that the number of measurements per entry will be 1.

For all calculations, each ST index was treated separately. As each ST index is calculated and reported independently, no interactions between indices are of interest. Therefore, for each analysis a repeated measures ANOVA table was calculated for each of the ten indices. Three variations of the ANOVA table, and subsequently the ICC and MDC, were calculated to quantify different representations of variability.

The first ANOVA was for the intra-observer variability. In this case, the variability of each individual observer was calculated for each index. This was performed by calculating a random effects ANOVA table, with the three repeated trials for the n = 13 subjects as entries to the ANOVA.

The second ANOVA was for the inter-observer variability. In this case, the output quantifies the systematic differences between observers. This was performed by calculating a random effects ANOVA table, with the average of the three trials done by each observer, for the n = 13 subjects as entries to the ANOVA.

Finally, the third ANOVA was for the total-observer variability. This provides a measure that considers the cumulative effects of both various observers and multiple trials. This

used a random effects ANOVA with all six repeated measures (three from each observer), for each subject.

2.4.3.1 Calculation of ICC, SEM, and MDC

Once the ANOVA tables were constructed for each of the indices and intra, inter, and total situations described above, the ICC(2,1) was calculated in SPSS (Version 19, IBM SPSS, Armonk, USA) in accordance with the description described by Shrout and Fleiss (1979).

The method used to calculate the SEM is shown in Eq. 2.6 (Weir, 2005). This method was shown by Stratford (2004) to converge to the true value as quickly as any other method of calculating the SEM.

$$\boldsymbol{SEM = \sqrt{MS_e}} \qquad \textbf{Eq. 2.6}$$

Where: $\boldsymbol{MS_e}$ represents the mean square error term from the ANOVA table

Once the SEM has been calculated, the MDC is easily calculated based on the desired confidence interval (Weir, 2005; Stratford, 2004; Bland & Altman, 1996). For this study, a 95% confidence interval was chosen, resulting in the equation below, where 1.96 is the z-score associated with a 95% confidence interval:

$$\boldsymbol{MDC = 1.96 \times \sqrt{2} \times SEM} \qquad \textbf{Eq. 2.7}$$

2.4.4 Subject Demographics

This study has received ethics approval from the appropriate ethics board and all subjects recruited to participate in the study and/or their guardians provided informed consent. The subject group consisted of AIS patients ranging in age from 10 to 16 years of age, with nine females and four males. In the strictest sense, the subject group was not randomly chosen, as there was considerable overlap with the subject groups used to determine clinically progressed and non-progressed AIS as described in Chapter 3. This was done for the sake of efficiency and to reduce the considerable number of hours required to process the data. While not a true random sample, the MDC subject group covered a wide range of ages, both genders, and a wide range of curve magnitudes and locations (Table 2.3). For these reasons it is believed that the MDC subject group provides a representative sample to the entire AIS population. As lateral X-rays were not available, it was not possible to determine curve types as per the King (King et al., 1983) or Lenke (Lenke et al., 2001) systems. Therefore curve types were simply defined by the location of the apex of the primary curve, as determined from the coronal plane X-ray. The five curve type options were: (1) double major (DM), (2) double thoracic (DT), (3) lumbar (L), (4) main thoracic (MT), and (5) thoracolumbar (TL).

Table 2.3: MDC subject demographics

Subject	Gender	Age (years)	Cobb (degrees)	Curve Type
1	M	14	31	DM
2	F	11	18	DM
3	F	11	18	DT
4	M	10	6	DM
5	F	14	34	DM
6	F	13	44	MT
7	F	14	34	TL
8	F	13	33	TL
9	M	14	28	DT
10	F	13	23	TL
11	F	16	23	TL
12	M	14	11	congenital
13	F	13	33	MT
Summary	9F / 4M	13.1 ± 1.6	25.8 ± 10.6	

After 15 subjects had been selected for inclusion into the MDC group and the analysis had been performed, some discrepancies in the data were discovered that are disclosed here.

- MDC Subject 4: the reported Cobb angle was 6°, which by definition is not considered AIS. However this subject eventually progressed to Cobb angles greater than 10°, confirming the diagnosis of AIS. This fact, along with the variability in Cobb angle measurements as previously discussed, warranted the inclusion of this subject in the study.

- MDC Subject 12: a hemi-vertebra was discovered in the X-rays when being examined to determine curve type. This indicates a diagnosis of congenital scoliosis as opposed to AIS. The decision was made to maintain this subject in the study based on the clinical opinion that torso shape in a congenital scoliosis patient would not be significantly different to that in an AIS patient. Therefore it was deemed valid to include this subject to determine the variability in the ST indices.

2.5 Results

The ICC(2,1) values for each ST index are reported in Table 2.4. In general, the ICC(2,1) values are extremely high for almost all indices and all variations of observer. The one notable exception is the SP line. While this index showed moderately high ICC values for the intra-observer cases, there was considerable variability between observers, as shown by the inter-observer value of 0.624.

The MDC values for each ST index are reported in Table 2.5. Again, the ST index showing greatest difference between the intra-observer variability and inter-observer variability is the SP line.

Table 2.4: ICC(2,1) of ST indices (modified from Dubetz et al., 2011a)

Index	Intra-Observer 1	Intra-Observer 2	Inter-Observer	Total-Observer
BSR	0.997	0.993	0.975	0.981
PAX	0.991	0.994	0.978	0.981
Rib Prom	0.996	0.985	0.971	0.975
1/4 Area Diff	0.997	0.999	0.998	0.997
Aspect Ratio	0.993	0.994	0.995	0.992
Lat Cent Line	0.994	0.991	0.995	0.991
SP Line	0.990	0.920	0.624	0.744
Lat Inertia	0.978	0.964	0.983	0.967
AP Inertia	0.990	0.993	0.998	0.992
Kyphosis	0.996	0.998	0.998	0.996

Table 2.5: MDC (95% confidence interval) of ST indices (modified from Dubetz et al., 2011a)

ST Index	Intra-Observer 1	Intra-Observer 2	Inter-Observer	Total-Observer
BSR (degrees)	1.129	2.089	3.793	3.301
PAX (degrees)	2.008	1.841	3.349	3.115
Rib Prom (mm)	1.792	3.514	4.861	4.517
1/4 Area Diff (1)	3.804	2.651	3.204	3.842
Aspect Ratio (1)	5.157	4.641	4.610	5.657
Lat Cent Line (mm)	2.775	3.287	2.542	3.358
SP Line (mm)	2.515	7.478	15.520	13.017
Lat Inertia (1)	28.549	35.116	24.204	34.217
AP Inertia (1)	5.437	4.543	2.696	4.944
Kyphosis (mm)	1.626	1.033	1.318	1.590

2.6 Discussion

All of the calculated ICCs are extremely high with the exception of the SP line (Table 2.4), especially for the inter and total-observer variability. It is not surprising that the SP line showed the greatest variability as it requires the greatest amount of user input. This allows for the greatest amount of variability between repeated calculations. Currently, there is work underway within the Scoliosis Research Group to rewrite the custom software used to process the data and this includes a higher level of automation in the determination of the SP line. It is expected that this increased level of automation will reduce the variability in this specific index. Additionally, procedural changes such as the inclusion of visible markers placed on the spinous processes through physical palpation could reduce the variability in this particular ST index.

Conclusions drawn from comparisons between the ICCs of the remaining ST indices are difficult to justify. As the ICC values are extremely high, there is likely a ceiling effect. Due to this, it would be difficult to claim that the total-observer variability of quarter area difference (0.997) is less variable than lateral centroid line (0.991). Appendix 1 lists an expanded table including the 95% confidence interval on the ICC(2,1) values. This shows that the differences between most ST indices are less than the range of the 95% confidence intervals, making it difficult to consider one index superior to the others.

Table 2.6 presents select results of the current study in the context of the two studies carried out by the Montreal based research group. As these studies share some of the ST indices and are calculated from the same ST imaging hardware, comparisons between

them illustrate the effects of the various study protocols. The ICCs from the current study (even the total-observer values) are higher than both of the other studies reporting intra-observer values. This implies that patient repositioning has a greater effect on the relative variability of the ST indices than the choice of observers. However, when examining the MDC values, it can be seen that the absolute variability in BSR and PAX for the total-observer case of the current study is higher than that of the intra-observer case of the study by Pazos et al. (2007). This implies that the choice of observer may have a greater effect that patient repositioning on absolute variability.

The reported MDCs for each of the ST indices (Table 2.5) provide a measure in the units of the corresponding index. As this is an absolute measure of variability, specific to each ST index, comparisons of MDCs of different indices hold no interpretation as to which index is superior. Determining the MDC is an important step in evaluating the clinical utility of the ST imaging system. However, the MDC must be evaluated against what constitutes a clinically relevant or important change for each specific index. This quantity will be examined in the following chapter. Finally the relation between the MDC and the clinically relevant levels of change will be discussed in Chapter 4.

Table 2.6: Comparison of ST index variability across three studies in two different research groups

Index	Intra-Observer Variability		Reference	Observer	Subject Repositioning	Location of ST Indices
	ICC	MDC				
BSR	0.92 - 0.93	2.39° - 2.48°	Pazos et al. (2007)	Intra	Yes	Undisclosed
PAX	0.97 - 0.97	1.43° - 1.48°				
BSR	0.79 - 0.98	3.3° - 11.1°	Seoud et al. (2012)	Intra	Yes	Entire torso
PAX	0.80 - 0.98	3.9° - 11.1°				
Lateral centroid line	0.87 - 0.98	8.0 mm - 17.5 mm				
BSR	0.98 - 1.00	1.13° - 3.79°	Current study	Intra	No	Apex of primary curve
PAX	0.98 - 0.99	1.84° - 3.35°		Inter		
Lateral centroid line	0.99 - 1.00	2.5 mm - 3.4 mm		Total		

CHAPTER THREE: CLINICALLY RELEVANT LEVELS OF CHANGE

The overall goal of examining the variability of the ST measurement system (including associated ST indices) is to understand the applicability for clinical application in detecting scoliosis and monitoring progression. Consequently, it is important to recognize the differentiation between a detectable change and a clinically relevant change. The minimum detectable change, or MDC, evaluated in Chapter Two is strictly an objective quantification of the variability of the measurement system. This value does not incorporate any expert clinical opinion or judgement as to whether the MDC has any clinical significance.

On the other hand, a minimum clinically important difference (MCID) (Beaton et al., 2002) is a subjective quantity that is deemed to have clinical implications. The primary motivation in determining the MCID, in the context of this study, is to provide a measure against which to judge the MDCs. The MDCs quantify the precision of the system, while the MCIDs enable determination of whether the system is sufficiently precise for acceptable clinical use.

A simple example to illustrate this point can be made from measuring body temperature to determine a fever. Consider the case in which normal body temperature is taken as 37.0°C and a temperature exceeding 37.7°C is considered a fever warranting medical intervention. The MCID would be calculated as:

$$\boldsymbol{MCID_{fever} = 37.7℃ - 37.0℃ = 0.7℃} \qquad \textbf{Eq. 3.1}$$

If a mother using the back of her hand placed on a child's forehead can detect changes of 2°C (MDC_{mom} = 2°C), then it is not precise enough. This is because MDC_{mom} > $MCID_{fever.}$ This result means that many true fevers may go untreated as it takes a change in temperature of only 0.7°C to warrant intervention. Conversely, if a high precision thermometer is used that can distinguish differences of 0.001°C ($MDC_{thermometer}$ = 0.001°C), it would be precise enough for clinical implementation as it could reliably detect a 0.7°C change ($MDC_{thermometer}$ < $MCID_{fever}$). However, just because the thermometer can detect a 0.001°C difference does not mean that this difference is clinically significant. The MDC is not a measure of clinical significance because a difference of 0.001°C would not alter the recommendation for clinical intervention. The MCID is dependent on the medical condition but independent of the tool used to measure that condition.

This chapter focuses on the methods used to estimate the MCID of the ST indices and present the results. SA2 and SA3 are addressed and extended to the testing of H1. The MCID of the ST indices are discussed in the context of clinically relevant levels of scoliosis progression as determined by the current standard of care. A detailed comparison and discussion of the relationship between the MDC and MCID is provided in Chapter Four.

3.1 Methods

As the MCID for any condition is a subjective quantity, this amount could vary from physician to physician. The MCID is typically based on experience, either individual or collective. There is considerable lack of consensus with regard to the quantification of the MCID for a given condition (e.g. Beaton et al. 2002). Various considerations must be addressed to properly situate the result within a specific context, including (Beaton et al. 2002): whether the results are determined by looking at differences between subjects in a given group, changes within a subject over time, or a combination of both; whether the results are applicable to an individual's scores or a group's scores; and the perspective from which the results are interpreted (e.g. a change that the patient considers important may differ from that of the clinician, researcher, or policy maker). For this study, the MCID is determined within the context of within-subject changes over time, from the perspective of the clinician, and the interpretation is considered for both the individual and group settings.

Obviously, determination of appropriate MCID values presents a challenge when developing new measurements, such as the ST indices. In this case, the MCID is unknown because there is no prior experience with the clinical application of these indices. Evaluation of whether or not the measurement system is sufficiently precise requires an estimate of the MCID. For the purposes of this study, this estimate was accomplished by relating the newly developed ST indices to established clinical standards of care, namely the Cobb angle and its associated MCID.

3.1.1 MCID of Cobb Angle

Several different criteria have been used to indicate a clinical progression of AIS (Lonstein & Carlson, 1984; Bunnell, 1986; Theologis et al., 1997; Kim et al., 2010; Silva & Lenke, 2009). The most common criteria include: a single change in Cobb angle equal to or greater than 10° between X-rays, a single change in Cobb angle equal to or greater than 5° between X-rays, or two consecutive changes in Cobb angle equal to or greater than 5°. However, studies outlining the clinical progression definition of AIS are at least partially based on what can precisely be detected by current radiographic measurements (i.e. the MDC of the Cobb angle) and not necessarily based on a level of change that influences the course of treatment (Kim et al., 2010). Guidelines set up by the SRS (section 1.2) define various ranges of Cobb angles and typical courses of treatment for those ranges (SRS, 2012, Treatment):

- Postural asymmetry (not scoliosis): $0° \leq$ Cobb angle $\leq 10°$
- Monitoring: $10° <$ Cobb angle $\leq 25°$
- Brace treatment: $25° <$ Cobb angle $\leq 40°$
- Surgical treatment: Cobb angle $\geq 45°$

Treatment decisions are typically made based on the above guidelines as well as consideration of other factors. While it is acknowledged that these guidelines are approximate, they imply that the MCID for Cobb angle is somewhere between 10° and 15°, as these magnitudes of change would alter treatment recommendation. This basis for choosing an MCID does not depend on the precision of the radiographic Cobb angle measurement.

For this study, the radiographic MCID in Cobb angle was defined as 10°. This lower limit was chosen to provide a more stringent and challenging measure to assess the MDCs, and thus provide a conservative criterion on whether or not the ST measurement system has the necessary precision to be clinically useful. It is also consistent with the clinical protocol at the Alberta Children's Hospital, in which a change of 10° or more is considered as "true progression."

3.1.2 MCID of ST Indices

To estimate the MCID of the ST indices, the within-subject change in indices in a group of AIS patients who were shown to have clinically progressed between two time points were measured. An increase in Cobb angle of 10° or larger was used to categorize this group of clinically progressed patients. Therefore, a well-established clinical standard was utilized to aid in the estimation of the MCID of the newly developed ST indices. The crux of many previous studies (section 1.3.3) was to establish a direct correlation between ST indices and Cobb angle. By using the Cobb angle to establish the clinically progressed group, a link between Cobb angle and ST indices is made without requiring a direct correlation.

The selection of a clinically progressed group to define the MCID of the ST indices was based on the hypothesis that the ST indices will demonstrate significantly more change in a clinically progressed group of AIS patients than a non-progressed group of AIS patients. Thus, to test the strength of this hypothesis, an additional group of subjects was required. In addition to the group of clinically progressed AIS patients, a group of

clinically non-progressed AIS patients were also chosen to enable comparisons to the MDC values and the change in ST indices of the clinically progressed group. Non-progressed was defined as those showing a change in Cobb angle less than 5°. As the currently intended use of the system is to detect a progression or change of scoliosis, both a non-progressed AIS population and a non-scoliotic population were considered to be similar. This is due to the consideration that neither group should demonstrate a change in scoliotic deformity. The non-progressed AIS group was chosen over a non-scoliotic group for this study, as the non-progressed AIS group should be more similar to the progressed AIS group than a non-scoliotic population. Additionally, by using a non-progressed AIS population, clinical X-rays were available to confirm the magnitude of change in Cobb angle. Spinal X-rays would not be available for a non-scoliotic population.

3.1.3 Calculations and Statistical Analysis

For both groups, the ST indices were calculated at the level of the apex of the primary curve. As described previously (section 2.4.2), the choice of calculating the ST indices at a single point was to facilitate the use of classical statistical tests and comparisons. The apex provides a repeatable location that avoids the problematic areas above the location of the arms. Additionally, the apex of the primary curve is the region of greatest interest as it is where the scoliotic deformity is most deviated from a healthy spine. The only exception to this was for the kyphosis index, which was determined at the maximum posterior value along the torso (as detailed in section 1.4.2). For the analysis undertaken in this thesis, all ST indices were evaluated independently.

The demographics of the clinically progressed and non-progressed groups were compared. Continuous variables included: age at the first scan, age at the second scan, time between the two scans, Cobb angle at the first scan, Cobb angle at the second scan, and change in Cobb angle between the two scans. These continuous variables were tested with an independent samples t-test for the null hypothesis: that the difference in means between the two groups equals zero. Fisher's exact test was used to check that the categorical variables (gender and curve type) of the two groups were equal between groups. A level of significance of 0.05 was used for all statistical tests.

The within-subject change scores of the ST indices were calculated as per Eq. 3.2:

$$\Delta \boldsymbol{X} = |\overline{\boldsymbol{X_{T2}}} - \overline{\boldsymbol{X_{T1}}}| \qquad \textbf{Eq. 3.2}$$

Where $\boldsymbol{X}$ represents the ST index under consideration, and $\boldsymbol{T1}$ and $\boldsymbol{T2}$ represent the value taken at time point one and time point two, respectively.

Some measurable variability in the ST indices is introduced when processing data (Chapter Two). To mitigate these effects and to obtain better estimates of the true values of the ST indices and consequently better estimates of the within-subject change scores, each scan was processed multiple separate times. The mean value for each index was determined. This procedure was completed at both time points T1 and T2. The absolute value of the difference between the two means was determined (Eq. 3.2). The absolute value was utilized when taking the difference between time points as many of the geometric indices were assigned an arbitrary positive direction when developed

(Jaremko, 2001). For example, a rotation to the right was assigned positive for back surface rotation. Consequently, a positive difference in rotation between an initial time point and a subsequent time point has an ambiguous interpretation. It could mean either an increase in rotation to the right or a decrease in rotation to the left. Further, the AIS population is highly heterogeneous and rotations or curves are seen both to the right and to the left. Therefore, the absolute value is used to simply indicate the magnitude of the change. It does not imply whether that change diverged from a 'normal' or symmetrical torso surface or converged on a 'normal' surface. Examination of the raw data of various patients within the study database, not necessarily included in either of the clinically progressed or non-progressed groups, indicates that in some instances the ST indices show an improvement, or reduction in asymmetry over time. While AIS is generally not known to spontaneously correct itself, this does indicate that a closer examination of the direction of change in ST indices may yield more useful information than simply the magnitude of change. This aspect is currently being explored in an ongoing study, and is beyond the scope of this thesis. For the purpose of this thesis, only the magnitudes of the change in ST indices are utilized.

The within-subject change scores for all indices were tested within each group (clinical progression and non-progressed) for normality. A Shapiro-Wilk test was used as it has been developed for small sample sizes and has been shown to have higher power than other common tests of normality. However, when sample sizes are as small as n = 10 or lower, almost all of the goodness of fit tests have relatively low power, therefore their interpretation must be used with caution (Razali & Wah, 2011).

Differences in the mean within-subject change scores between the clinically progressed and non-progressed groups were tested with both a parametric independent samples t-test and a non-parametric Mann-Whitney U test. If there was a discrepancy in the conclusions drawn from these tests, the results from the Mann-Whitney U test were taken as correct, as they do not rely on a parametric assumption. However, if the conclusions were in agreement, thus confirming the parametric assumption, the results from the t-test were taken as they provide more information. Post-hoc power calculations were performed for specified ST indices.

3.1.4 Study Design

As both the variability and the expected differences in the ST indices between time points for a clinically progressed population were unknown, no sample size calculations were made a priori. Due to the extensive amount of time required to process all the necessary ST scans multiple times, a sample size of n = 10 for both the clinically progressed and non-progressed groups were chosen for convenience. The selection process for the clinically progressed group was based on sorting the study database by ID number and then to indicate patients with Cobb angle changes 10° or greater. The first 10 of such patients were chosen for inclusion, as there were not a large number of patients (13 patients) that met the inclusion criteria for this group. For the non-progressed group, patients were selected at random from the study database and if they showed Cobb angle changes under 5° between X-rays, they were included in the non-progressed group. The first 10 such randomly chosen patients were included in the clinically non-progressed group. However, through the course of the analysis, it was determined that one subject

from the progressed group and two subjects from the non-progressed group needed to be excluded from the study (details given in section 3.1.6).

As a summary, the study design was as follows:

Clinically Progressed

- Number of observers, k = 1*
- Number of time points, T = 2
- Number of repeated trials, m = 5
- Number of subjects, n = 9

Non-Progressed

- Number of observers, k = 1*
- Number of time points, T = 2
- Number of repeated trials, m = 3
- Number of subjects, n = 8

* While only one observer analyzed each group of subjects, it was a different observer between the two groups.

3.1.5 Subject Demographics

The demographics of the clinically progressed group (Table 3.1) indicate that this group was comprised of predominantly female participants (7F and 2M), aged 13.0 ± 1.3 years at T1, and 14.1 ± 1.5 years at T2. Cobb angles at T1 ranged from 11° to 51°, with a mean of 23.7° ± 13.4°. At the second time point, the mean Cobb angle had increased to 36.3° ± 13.2°, with range from 23° to 61°. The curve type was relatively evenly distributed across types: DM (n=2), DT (n=2), MT (n=2) and TL (n=3). The clinically non-progressed group (Table 3.2) was comprised of all females who were slightly younger in mean age at T2 (13.6 ± 1.8 years). With the exception of three individuals (Subjects 5, 6, and 7), all subjects fall clearly within the SRS definition of treatment. Both the mean and

standard deviation of the Cobb angle measurements for this group are very similar between time points, with the values at T1 of 31.2° ± 16.6° and T2 of 31.2° ± 16.7°. This group also demonstrated a relatively even distribution of curve types: DM (n= 2), MT (n= 1), TL (n=4), L (n=1). Statistical analysis of the two groups (Table 3.3) revealed that the only statistically significant difference between groups is in the change in Cobb angle between scans.

Table 3.1: Clinically progressed group demographics

Subject	Gender	Age Time 1 (years)	Age Time 2 (years)	ΔTime (months)	Cobb Time 1 (degrees)	Cobb Time 2 (degrees)	ΔCobb (degrees)	Curve Type
1	M	14	14	10	31	43	12	DM
2	F	11	12	7	18	30	12	DM
3	F	13	14	7	11	26	15	DT
4	F	11	12	7	18	30	12	DT
6	F	13	14	6	37	53	16	MT
7	F	14	15	13	51	61	10	MT
8	F	13	14	17	15	26	11	TL
9	M	13	16	31	21	35	14	TL
10	F	15	16	12	11	23	12	TL
Summary	7F / 2M	13.0 ± 1.3	14.1± 1.5	12.2 ± 7.9	23.7± 13.4	36.3 ± 13.2	12.7 ± 1.9	

Table 3.2: Non-progressed group demographics

Subject	Gender	Age Time 1 (years)	Age Time 2 (years)	ΔTime (months)	Cobb Time 1 (degrees)	Cobb Time 2 (degrees)	ΔCobb (degrees)	Curve Type
1	F	14	15	6	34	33	-1	DM
2	F	13	14	5	36	34	-2	TL
3	F	13	13	6	33	37	4	TL
5	F	12	13	5	16	15	-1	DM
6	F	9	10	17	13	16	3	TL
7	F	16	16	8	20	17	-3	TL
8	F	13	14	11	40	40	0	MT
9	F	14	14	6	65	65	0	L
Summary	8F	13.0 ± 2.0	13.6 ± 1.8	8.0 ± 4.1	32.1 ± 16.6	32.1 ± 16.7	0.0 ± 2.4	

Table 3.3: Comparison of progressed vs. non-progressed groups

Group		Progressed	Non-Progressed	p (2-sided)
Continuous Variables		Mean ± SD	Mean ± SD	
Age Time 1	(years)	13.0 ± 1.3	13.0 ± 2.0	1.000
Age Time 2	(years)	14.1± 1.5	13.6 ± 1.8	0.543
ΔTime	(months)	12.2 ± 7.9	8.0 ± 4.1	0.197
Cobb Time 1	(degrees)	23.7± 13.4	32.1 ± 16.6	0.264
Cobb Time 2	(degrees)	36.3 ± 13.2	32.1 ± 16.7	0.571
ΔCobb	(degrees)	12.7 ± 1.9	0.0 ± 2.4	4.006E-09 *
Categorical Variables		Count	Count	
Gender				0.471
	Female	7	8	
	Male	2	0	
Curve Type				0.675
	DM	2	2	
	DT	2	0	
	L	0	1	
	MT	2	1	
	TL	3	4	

* $p < 0.05$

3.1.6 Assumptions and Discrepancies

There are several key assumptions made in this study. The overall hypothesis of the work performed by the Scoliosis Research Group is that scoliotic changes to the skeletal system will be reflected on the torso (Chapter One). More specifically, Hypothesis 1 (H1) states: The ST indices will demonstrate significantly more change in a clinically progressed group of AIS patients than a non-progressed group of AIS patients.

Therefore, the assumption that is made when selecting this group of clinically progressed AIS patients is that *they define the MCID of ST indices.*

Conversely, since the ST indices were developed to measure asymmetries introduced by scoliosis, they should not show changes in a group of non-progressed AIS patients. It is assumed that theoretically these indices do not describe symmetrical growth. Therefore, comparison of these groups can be used to check these assumptions.

After 10 subjects had been selected for inclusion into the two groups and the analysis had been performed, some discrepancies in the data were discovered that are disclosed here.

- Progressed Subject 5 was excluded from the results: The Cobb angle was erroneously reported from supine X-rays and not standing X-rays. There were no standing X-rays corresponding to the surface topography acquisition.

- Progressed Subject 6: The clinical notes varied over time as to which curve was classified as the primary curve. The curve used in this section of the study was the only curve that progressed by 10° or greater, regardless of its classification.
- Progressed Subject 8: This subject was the only case where the apex of the curve changed significantly between scans (time points T1 to T2), from L1 to the L3-4 disc. As the end vertebrae remained constant and it was determined that indeed it was the same curve. The ST indices were calculated at L1 for both scans (time points T1 and T2), to provide a consistent location for comparison of ST indices.
- Non-progressed Subjects 4 and 10 were excluded from the results: Both these patients had ST scans that could not be processed through the InSpeck software. This problem can be attributed to large occluded areas of the torso, minimal overlap in the field of view between cameras, or other various software bugs.
- Non-progressed Subject 6: This subject was nine years old at the time of the first scan (time point T1). While AIS is typically diagnosed for patients between 10 to 18 years of age, this is a guideline and not a strict requirement as skeletal or developmental age does not correspond exactly with chronological age (Terver et al. 1980; SRS, 2012, Treatment). Corresponding to her diagnosis of AIS by the treating physician, she was included in the study.

3.2 Results

The change in Cobb angle along with the within-subject change scores for each subject as calculated using Eq. 3.2 for the 10 indices, are presented for the clinically progressed (Table 3.4) and non-progressed groups (Table 3.5). The mean within-subject change scores and between-subject standard deviations for each ST index and group are presented (Figure 3.1). In Figure 3.1, the vertical axis represents the numerical values of the ST indices, however it is understood that each ST index has its own specific measurement units. The normality of the data was tested using a Shapiro-Wilk test (Table 3.6). Based on this test, five indices (PAX, rib prominence, quarter area difference, AP inertia, and kyphosis) were found to deviate from normality. Therefore, for these indices only the results from the non-parametric Mann-Whitney U test should be considered. However, the power of the Shapiro-Wilk test for normality is low with the sample sizes of this study (section 3.1.3). Consequently, there is an increased chance of failing to reject the assumption of normality when it is truly false. For this reason, the parametric assumption should be used with caution. The parametric t-tests and non-parametric Mann-Whitney U tests for independent samples for all indices (Table 3.7) confirm that the change in Cobb angle is the only variable that is statistically different between these two groups.

Table 3.4: Changes in ST indices within a clinically progressed AIS population (modified from Dubetz et al., 2011a; Dubetz et al., 2011b)

Subject	ΔCobb (degrees)	ΔBSR (degrees)	ΔPAX (degrees)	ΔRib Prom (mm)	Δ1/4 Area Diff (1)	ΔAspect Ratio (1)	ΔLat Cent Line (mm)	ΔSP Line (mm)	ΔLat Inertia (1)	ΔAP Inertia (1)	ΔKyphosis (mm)
1	12	3.92	4.66	10.92	3.35	0.93	3.03	9.97	4.25	5.97	0.74
2	12	1.74	0.35	4.11	11.76	17.48	6.09	8.88	50.52	2.51	15.98
3	15	0.87	2.59	3.12	29.79	26.78	16.89	16.44	98.87	18.39	2.59
4	12	2.73	1.42	6.88	2.91	6.91	0.25	2.95	1.00	5.66	3.15
6	16	0.69	3.58	0.55	21.96	9.97	14.10	15.71	115.17	15.53	2.75
7	10	0.48	1.19	2.15	1.29	4.95	1.20	1.31	1.90	7.17	9.85
8	11	2.32	0.29	2.67	23.55	16.17	11.04	6.97	63.52	17.29	7.19
9	14	1.06	2.48	0.70	2.54	15.52	1.42	0.21	24.24	1.58	0.62
10	12	3.28	3.61	5.01	29.85	32.28	17.00	12.44	104.71	20.14	0.18
Mean	12.7	1.90	2.24	4.01	14.11	14.56	7.89	8.32	51.57	10.47	4.78
Std Dev	1.9	1.23	1.54	3.28	12.19	10.21	6.93	5.98	46.45	7.29	5.28

Table 3.5: Changes in ST indices within a non-progressed AIS population

Subject	ΔCobb (degrees)	ΔBSR (degrees)	ΔPAX (degrees)	ΔRib Prom (mm)	Δ1/4 Area Diff (1)	ΔAspect Ratio (1)	ΔLat Cent Line (mm)	ΔSP Line (mm)	ΔLat Inertia (1)	ΔAP Inertia (1)	ΔKyphosis (mm)
1	-1	4.77	7.41	14.19	8.04	2.95	9.73	23.47	57.83	6.15	1.90
2	-2	0.70	1.93	3.47	18.50	24.87	14.86	14.35	39.03	8.44	2.48
3	4	2.29	7.03	4.19	11.05	6.78	7.92	7.95	33.04	6.15	13.60
5	-1	3.30	3.03	4.41	9.25	6.26	1.98	9.16	9.93	7.05	18.14
6	3	0.22	0.65	1.57	10.77	9.76	6.25	4.15	45.91	7.44	0.60
7	-3	1.10	0.58	2.64	1.54	3.10	2.93	9.68	5.34	0.24	14.09
8	0	1.36	0.48	3.71	0.82	1.11	0.93	1.35	4.69	0.61	5.48
9	0	0.01	2.42	0.12	10.36	14.25	5.07	10.01	13.17	8.32	5.65
Mean	0.0	1.72	2.94	4.29	8.79	8.63	6.21	10.01	26.12	5.55	7.74
Std Dev	2.4	1.64	2.80	4.25	5.64	7.79	4.60	6.70	20.47	3.28	6.60

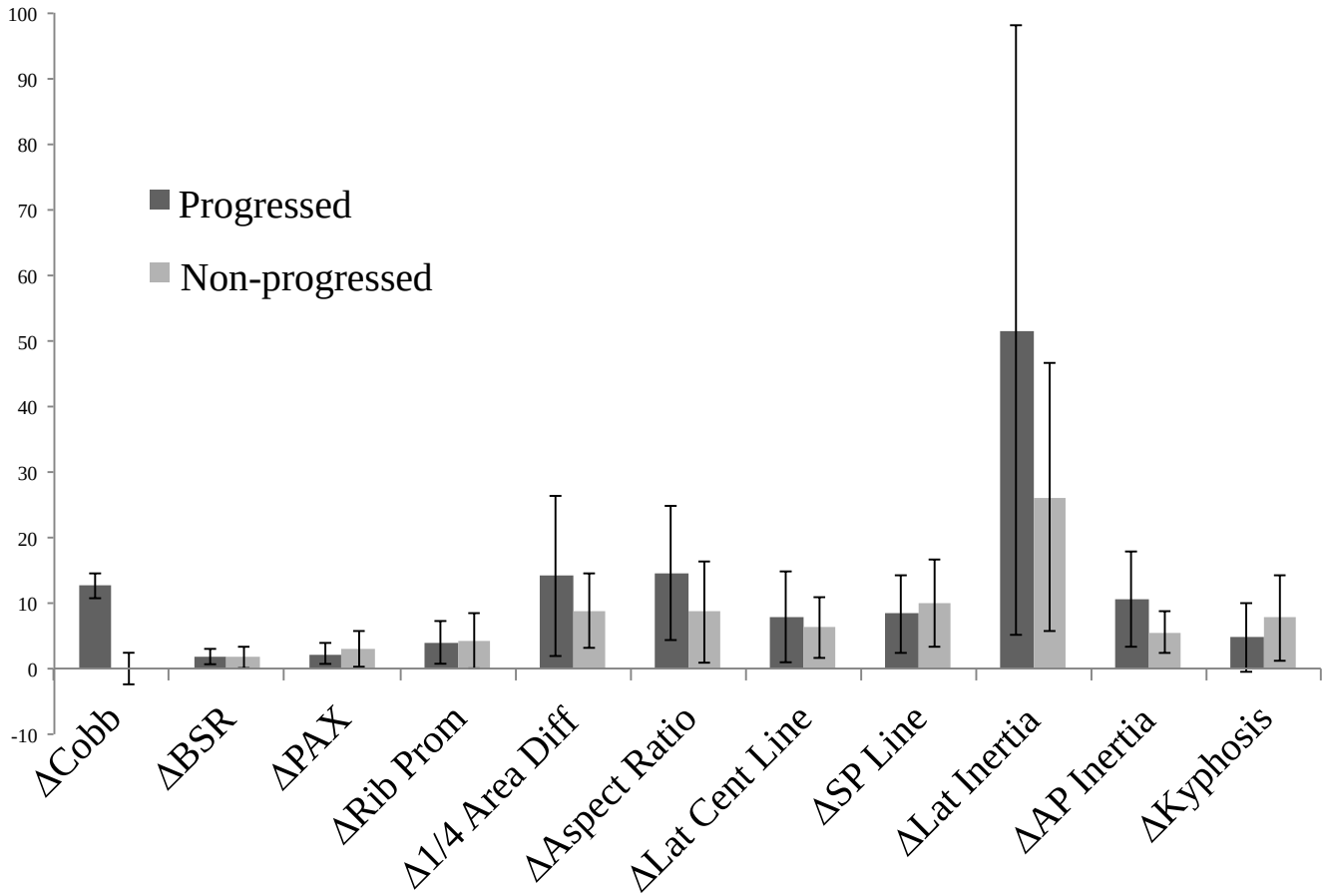

Figure 3.1: Comparison of mean within-subject change scores for all ST indices between progressed and non-progressed groups. Error bars represent ± 1 standard deviation of the between-subject variability within each group.

Table 3.6: Tests of normality

Index	Group	Shapiro-Wilk		
		Statistic	df	p (2-sided)
ΔCobb	Progressed	.910	9	.317
	Non-Progressed	.915	8	.388
ΔBSR	Progressed	.928	9	.464
	Non-Progressed	.913	8	.376
ΔPAX	Progressed	.942	9	.602
	Non-Progressed	.813	8	.039
ΔRib Prom	Progressed	.903	9	.273
	Non-Progressed	.743	8	.007
Δ1/4 Area Diff	Progressed	.835	9	.050
	Non-Progressed	.912	8	.365
ΔAspect Ratio	Progressed	.956	9	.753
	Non-Progressed	.862	8	.126
ΔLat Cent Line	Progressed	.864	9	.106
	Non-Progressed	.944	8	.653
ΔSP Line	Progressed	.939	9	.569
	Non-Progressed	.919	8	.423
ΔLat Inertia	Progressed	.877	9	.146
	Non-Progressed	.894	8	.255
ΔAP Inertia	Progressed	.875	9	.140
	Non-Progressed	.784	8	.019
ΔKyphosis	Progressed	.831	9	.046
	Non-Progressed	.884	8	.207

Indicates a deviation from normality ($p < 0.05$)

Table 3.7: Independent samples tests

Index	t-test	Mann-Whitney U
	p (2-sided)	p (2-sided)
ΔCobb	4.006E-09 *	.000 *
ΔBSR	.801	.700
ΔPAX	.525	.847
ΔRib Prom	.883	1.000
Δ1/4 Area Diff §	.264	.336
ΔAspect Ratio	.203	.149
ΔLat Cent Line	.570	.773
ΔSP Line	.590	.630
ΔLat Inertia §	.164	.501
ΔAP Inertia §	.094	.386
ΔKyphosis	.321	.441

Indicates a deviation from normality

§ Equal variances not assumed

* $p < 0.05$

3.3 Discussion

No statistically significant differences in the magnitude of ST change scores between the clinically progressed or non-progressed groups at the chosen significance level of $p < 0.05$ (Table 3.7) were found. It appears that the change in aspect ratio tends towards a difference between the groups. However, it is not statistically significant with the limited sample sizes and high between-subjects variability. The change in lateral inertia also

shows a relatively small p-value ($p = 0.164$) for the t-test. However, the large disagreement between the parametric and non-parametric results ($p = 0.501$) indicates that the parametric assumption may not hold regardless of the test of normality.

There are several conclusions to be drawn from these results:

1. The assumptions regarding the expected changes in ST indices for the two groups may not hold (section 3.1.6). If this were the case, it would imply that the changes in ST indices witnessed over time are not entirely due to changes in the scoliotic deformity as defined by the change in Cobb angle. This aspect can be further broken down into two components:
    a. Changes in ST indices are affected by a number of factors. These may include changes in: scoliotic deformity, posture, growth, weight gain, marker placement, or a number of other unknown factors.
    b. Changes in the scoliotic deformity can occur that are not reflected by a change in the Cobb angle. The Cobb angle is a simplified representation of the scoliotic deformity. Therefore, it is reasonable to expect aspects of the scoliotic deformity to be neglected by the Cobb angle representation.

In the first case (a), it is reasonable to hypothesize that factors such as posture may largely influence the calculations of ST indices. Possible solutions to this problem could be to identify, control or account for the other factors that influence ST indices. Perhaps a more appropriate assumption would be that all subjects will show change in ST indices

over time due to multiple factors. Then a more appropriate MCID would be the excess change in ST indices shown by the clinically progressed group over the non-progressed group.

In the second case (b), there may be subjects in the non-progressed group who did progress in aspects of the disease that are sensitive to ST indices but not the Cobb angle. Unfortunately there is no way to indicate this situation, as this study has attempted to place the newly developed ST indices in the framework of reference of the existing clinical standard. Addressing this concern would require the rejection of the Cobb angle as the only current gold standard for scoliosis evaluation, as stated by Kotwicki (2008). A potential method to accommodate this situation would be to attempt to determine the MCID of AIS from the patient's perspective. A similar analysis could be conducted using the criterion for inclusion into the progressed/non-progressed groups based on the patient self-report measures as measured by a standardized questionnaire such as the SRS-22 or the Spinal Appearance Questionnaire (SAQ). Such an approach was taken by Gorton et al. (2012) when examining differences in ST indices between preoperative, postoperative, and healthy control subjects.

The low sample sizes and high between-subjects variability are potentially masking the true differences between study groups. As the current sample sizes are under 10, an increase in sample size would very likely show improvements in statistical significance and power. Additionally, attempts could be made to reduce the between-subject variability through some form of sub-group analysis. As progression rates have been

related to curve type (Bunnell, 1986; Lonstein & Carlson, 1984; Weinstein, 1999; Asher & Burton, 2006), curve type becomes a logical sub-group classification. However, this type of analysis requires substantially more subjects such that each sub-type has a sufficient sample size.

This study was completed to provide a clinically relevant estimate of the MCID of the ST indices. In the absence of clinical experience to define the MCID for the new ST indices, this estimate was based on the clinically accepted definition of progression of 10° change in Cobb angle, as this value was used to define the clinically progressed group. The within-subject change scores in ST indices displayed by the clinically progressed group define the MCID for the ST indices. In theory, the threshold to determine the MCID for each index would be the smallest reported within-subject change score for each index in the clinically progressed group. Calculating the MCID in this way would mean that the entire clinically progressed sample showed important levels of change in each ST index. However, examination of results (Table 3.4) reveals a large between-subject variability in the change scores within the clinically progressed group. This large between-subject variability is also demonstrated in the non-progressed group (Table 3.5). Several potential options exist for establishing the MCID. Attempting to include 100% of the clinically progressed group may provide an overly conservative estimation of the MCID for each ST index. Simply using the mean change score as the MCID may be adequate when interpreting results in the context of a group; however it would likely overestimate the MCID in the context of the individual. This situation arises because approximately half of the clinically progressed group would be excluded from indicating clinically

relevant levels of change in their ST indices. An appropriate threshold must be chosen to determine the MCID of each ST index at both the individual and group level. Goldsmith et al. (1993) recommends using the 25th percentile of the clinically progressed group when evaluating for an individual. In future components of this study, based on these options, the two thresholds: mean for the group context and 25th percentile for the individual context, will be taken as the MCID for each index. The MDCs will be interpreted against these MCIDs.

CHAPTER FOUR: DISCUSSION

The overall goal of the Scoliosis Research Group at the U of C is to develop a system for the monitoring of AIS patients that can be clinically implemented. There have been several generations of ST systems as equipment and the application of the system has been refined. Throughout these refinements, the ST system has moved forward from a strictly research tool towards a clinically relevant tool. The specific study undertaken in this thesis was to verify and assess the ST system's capability to be used clinically, and to provide insight into the directions that would assist in the forward development of the system towards this goal.

This current work has quantified the variability of the ST imaging system through the use of the MDC. The clinically relevant levels of change in ST indices have been estimated through the use of the MCID. While these two analyses were done separately, to determine the viability of implementing the current system clinically the MDC must be considered with respect to the MCID. Integrating the results of the two analyses will enable the evolution from reporting the system precision to a discussion on whether or not that level of precision is adequate for clinical use. This discussion addresses SA4 by comparing the variability in the ST indices relative to clinically important differences in ST indices. Limitations of the current system are outlined. Improvements are suggested that will continue to move the system towards a clinically viable solution for non-invasive monitoring of scoliosis progression.

4.1 Assessment of ST Imaging System

4.1.1 Quantification of Variability in ST Indices

The first specific aim (SA1) of the project was to quantify the intra, inter, and total-observer variability in ST indices calculated from the same input scan (Chapter Two). The ICC value was adopted as a measure for evaluating relative variability. While there are several forms of ICC and it is context and population specific, its relative and dimensionless nature provides a means to compare dissimilar systems at least in a generic sense. The limitations of the ICC were discussed in section 2.3.1. While the interpretations of these comparisons should be used cautiously, the ICC provides a means to compare the ST system developed by the U of C Scoliosis Research Group with different systems used for the measurement of the scoliotic deformity. In general, the ICC for all ST indices other than the total-observer case of the SP line compare favourably with those reported for similar systems (Table 4.1). However, it should be noted that most measures throughout the literature show high ICC values. This result suggests that the variability in the various measures, radiographic or ST based, used to quantify scoliosis is rather low.

Table 4.1: ICC values of various measures of scoliosis

Study	System	Indices	Intra-Observer	Inter-Observer	Total-Observer	Condition	Age (years)	Cobb range	Reposition
Current study	InSpeck	ST	0.92 - 1.00	0.62 - 1.00	0.74 - 1.00	AIS	10 - 16	6° - 44°	No
Pazos et al., 2007	InSpeck	ST	0.85 - 0.99	NR	NR	AIS	11.0 - 19.7	NR	Yes
Seoud et al., 2012	InSpeck	ST	0.79 - 0.98	NR	NR	AIS	11 - 18	20° - 75°	Yes
Gorton et al., 2012	Vitus Smart 3D	ST	0.59 - 0.98	NR	NR	JIS or AIS	10.8 - 17.7	49° - 108°	NR
Rankine et al., 2012	Milwaukee Topographic Scanner	ST	0.61 - 0.99	0.82 - 0.99	NR	AIS mannequin	NR	NR	Yes
Dang et al., 2005	X-ray	Radiographic	5/13 indices > 0.8	7/13 indices > 0.8	NR	AIS	11 - 15	20° - 45°	No
Srinivasalu et al., 2008	X-ray	Radiographic	0.77 - 0.99	0.88 - 0.99	NR	IIS, JIS, or AIS	0 , >10	<20°, >40°	No
Zhang et al., 2010	X-ray	Radiographic	0.92 - 0.99	0.91 - 0.98	NR	AIS	14.6 ± 2.7	<90°	No

AIS = Adolescent Idiopathic Scoliosis
JIS = Juvenile Idiopathic Scoliosis
IIS = Infantile Idiopathic Scoliosis

More clinically relevant than the ICC values, the MDC was calculated for each of the ST indices for the intra, inter, and total-observer cases. By quantifying the variability in an absolute sense and in the units of the original measurement, direct comparison with the MCID is enabled.

4.1.2 Estimation of Clinically Important Differences in ST Indices

The second specific aim (SA2) of the project was to quantify the change of ST indices over time in a clinically progressed AIS group. This group was used to estimate the MCID of the ST indices based on Hypothesis 1 (Chapter 3). The mean MCID and the 25th percentile MCID are justified for a research and clinical setting, respectively (Chapter 3). For exploratory purposes, an MCID taken at the 37.5^{th} percentile of the clinically progressed group was calculated to examine the sensitivity of the acceptability of ST indices to the estimation of MCID.

4.1.3 Qualification of Variability with respect to Clinically Important Differences in ST Indices

The successful completion of Specific Aim 4 (SA4) provides the greatest potential contribution to the field in relation to quantifying scoliosis with ST indices. To date, many ST systems have been developed and numerous authors have reported on the variability of these systems. A comparison of the ICC values based on select references was provided in Table 4.1. Additionally, several studies have examined differences in ST indices between pre/post operative subjects (Gorton et al. 2012; Swanson, 2008; Seoud et al. 2012; Pazos et al., 2007). The use of patient data acquired before and after surgical

procedures enables the examination of the ability to discriminate between groups of scoliosis patients using ST indices. However, the data acquired from these sample groups do not adequately represent the natural progression of the disease. To the best of the knowledge of the author, only the study of Seoud et al. (2012) has examined the ability to detect true change (i.e. change greater than the MDC) in the case of natural progression of AIS. However, this study has limited applicability as it was conducted on a single case subject. The study was not expressly designed to estimate an MCID in ST indices for groups representing the natural progression of AIS. For a ST system to be used for the monitoring of AIS, it is detecting the natural progression that becomes critical, as opposed to the more dramatic changes realized when using it to evaluate surgical outcomes.

For a measure to be acceptable, the precision to which true change can be detected – quantified by the MDC, must be smaller than the change that is trying to be detected – quantified by the MCID (MDC < MCID). Table 4.2 presents various iterations of both the MDC and MCID for each ST index for easy comparison. Through this comparison, judgement can be made on the acceptability of the ST indices. All observers that analyzed data in this study were well trained in the use of the system and considered experts. Therefore, any discrepancies between the interpretation of results between intra-observer 1 and intra-observer 2 will be taken as the more conservative of the two. The MDC values (intra-observer 1, intra-observer 2, and total-observer) in comparison to the MCID values (25th percentile, 37.5^{th} percentile, and mean) (Table 4.2) reveal considerable variation across ST indices and between observers. For example, for the

aspect ratio index all three values of MDC (intra-observer 1, intra-observer 2, and total-observer) are lower than the 25th percentile MCID (the smallest and most stringent quantification of MCID). Therefore, aspect ratio is considered acceptable for all interpretations. The examination of the rib prominence index shows that none of the three MDC values are lower than the 25th percentile MCID, indicating this index would not be acceptable for interpretation with respect to an individual. However, as both intra-observer MDCs are less than the mean MCID, rib prominence could be used in a research setting where comparisons to a group are made (justifying the mean MCID) and the same observer processed both scans (as the intra-observer cases were acceptable but the total-observer case was not).

Table 4.2: MDC and MCID values for the ST indices.
The lightest colour of shading represents a clinically acceptable MDC when compared to the mean MCID, the medium colour of shading represents an acceptable MDC when compared to the 37.5th percentile and mean MCID, and the darkest shading represents an acceptable MDC when compared to the 25th percentile, 37.5th percentile, and mean MCID. Un-shaded cells in the table represent situations where the MDC is clinically unacceptable as it is higher than all of the MCID values.

	ΔBSR (degrees)	ΔPAX (degrees)	ΔRib Prom (mm)	Δ1/4 Area Diff (1)	ΔAspect Ratio (1)	ΔLat Cent Line (mm)	ΔSP Line (mm)	ΔLat Inertia (1)	ΔAP Inertia (1)	ΔKyphosis (mm)
MDC										
Intra-Observer 1	1.13	2.01	1.79	3.80	5.16	2.77	2.51	28.55	5.44	1.63
Intra-Observer 2	2.09	1.84	3.51	2.65	4.64	3.29	7.48	35.12	4.54	1.03
Total-Observer	3.30	3.12	4.52	3.84	5.66	3.36	13.02	34.22	4.94	1.59
MCID										
Mean	1.90	2.24	4.01	14.11	14.56	7.89	8.32	51.57	10.47	4.78
37.5th percentile	1.01	1.37	2.54	3.24	9.20	2.63	5.96	19.24	5.89	2.13
25th percentile	0.78	0.77	1.43	2.72	5.93	1.31	2.13	3.07	4.08	0.68

The MCID must be placed in a specific context (Beaton et al., 2002). With the application of the ST imaging system, if the context is to compare groups of patients, then the mean MCID becomes appropriate as it measures the central tendency of the entire group. This situation is more suited to a research scenario than a clinical practice, as it depends on comparing groups of subjects as opposed to an individual patient. Under this context, all ST indices except BSR (9 of the 10) have clinically acceptable intra-observer MDCs, as they are less than the mean MCID. For total-observer MDCs, only six of the ten ST indices, namely quarter area difference, aspect ratio, lateral centroid line, lateral inertia, AP inertia, and kyphosis remain clinically acceptable. This is an encouraging finding as almost all ST indices show some utility in a typical research setting, where the observer processing the ST scans can be kept constant. However, the overall objective of the project is to develop a system for clinical implementation. This criterion requires the ability to properly detect changes on an individual AIS patient.

The conclusions regarding the ST imaging system are dramatically different if the suggestion of Goldsmith et al. (1993) is adhered to. They advocate using the 25th percentile of the changes in ST indices as the MCID to be adequately inclusive of the clinically progressed group. With this MCID, only one index, the aspect ratio, shows adequate MDCs for the context of the individual. This adequacy was maintained regardless of whether the scans were processed by the same observer processed (intra-observer MDC) or a different observer (total-observer MDC). This is an important feature, as in a typical clinical setting it would be difficult to guarantee that the same

observer processed all ST scans of a given patient throughout the course of their treatment.

It is interesting to note that the aspect ratio was also the index that showed the greatest trend towards differentiating between the progressed and non-progressed groups (Table 3.7, section 3.2). If it turns out through further investigation that the ST index that shows the strongest adequacy from a precision perspective can also differentiate between progressed and non-progressed groups, this index may have strong potential as a candidate to simplify the number of indices required. Consequently, this would fulfill the primary objectives of the project, leaving only the logistics of clinical implementation to be resolved.

Selection of the 25th percentile of the clinically progressed group to estimate the MCIDs was based on the recommendation of Goldsmith et al. (1993). MCIDs were determined at the 37.5th percentile to recognize the nascent nature of the ST indices and the lack of associated clinical expertise on which to base the selection of MCID threshold. MDCs for all ST indices were compared to these 37.5th percentile MCIDs to determine whether any further ST indices showed promise. Through increasing the MCID threshold to the 37.5th percentile the AP inertia and kyphosis indices also showed acceptable levels of MDC for both intra and total-observer cases. These two additional ST indices show higher levels of precision (total-observer MDC: AP inertia = 4.94, kyphosis = 1.59 mm) when compared to a clinically relevant change in indices (37.5th percentile MCID: AP inertia = 5.89, kyphosis = 2.13). However, unlike the aspect ratio, both AP inertia and

kyphosis showed little trend towards distinguishing between clinically progressed and non-progressed groups, with p-values of 0.386 and 0.441, respectively.

The preceding discussion centers around the assessment of the ST imaging system with respect to clinical implementation. Ultimately, if through further investigation, the aspect ratio turns out to be able to reliably differentiate between clinically progressed and non-progressed groups and individuals, the remaining ST indices would be rendered irrelevant. A more likely scenario is that a combination of ST indices is optimized (i.e. weighted equation) to delineate between a clinically progressed and a non-progressed group or individual. However, as this ability to differentiate between groups or individuals has yet to be shown at a statistically significant level, steps to improve the ST imaging system as a whole should be taken to further the development of this system into a clinical tool.

4.2 Improvement of ST Imaging System

The assessment of the ST imaging system performed in this thesis evaluated the variability with respect to clinically important changes in the ST indices. This evaluation was performed by examining the MDC to determine whether it was less than the MCID for each index. Therefore in this context, to improve the system would require either a decrease of the MDCs or an increase of the MCIDs.

4.2.1 Decreasing Variability of ST Indices

The reduction of variability in the ST indices and therefore the associated MDCs could be attained through several methods. These are briefly described as follows:

- The ST indices could be calculated through the averaging of a number of repeated trials. This approach would serve to bring the observed values closer to the true values and therefore reduce variability. However, as it currently stands, it takes a trained user approximately one hour of processing to obtain final value for a single trial. This length of processing time has been identified as a potential limiting factor for clinical adoption by the clinical staff at the Alberta Children's Hospital. Therefore multiplying this amount of time by the number of repeated trials would present more limitations to clinical adoption. For the averaging of repeated trials to become a feasible option, a large reduction in processing time would need to be realized. This could likely be achieved through a higher level of automation. This aspect is being worked on in an ongoing study within the Scoliosis Research Group.
- An increased level of automation in the software would likely reduce variability introduced in the stages requiring significant manual input.
- For the SP line index in particular, the total-observer MDC could be lowered through procedural changes in the data processing to ensure a more consistent selection of the SP points between different observers. However, this modification is unlikely to affect the intra-observer MDC. Procedural changes to the data collection could include the addition of visible markers to the spinous

processes through manual palpation by an informed clinical expert (e.g. clinical research nurse). This modification may reduce both the intra and total-observer variability of the SP line.

- A new imaging system to capture the 3D geometry of the torso is being developed in an ongoing project at the U of C, and has passed the proof-of-concept stage. This imaging system has the potential to provide greater accuracy, precision, and automation than the current system. Once a working prototype is developed, an analysis similar to the current project should be performed to quantify the precision of the new system to compare to the existing system.

4.2.2 Different Estimation of MCID

The MCID is dependent on the medical condition but independent of the tool used to measure that condition (Chapter 3). Therefore, an increase in the MCID is not achievable per se. However, it should be recognized that the MCID is estimated by the change in ST indices of a particular group. Therefore, the sample obtained from the larger population of AIS patients to represent this group influences the value of the MCID. A reduction of the large between-subject variability in the clinically progressed group would serve two beneficial purposes. Firstly, reduced variability would result in an increase in the statistical significance of the difference in means between the clinically progressed and non-progressed groups. Secondly, the system precision requirements would be eased to more attainable levels. This would be accomplished because a reduction of variability in the group would serve to bring the 25th percentile closer to the mean value.

A reduction of the between-subject variability may be achieved through an increase of sample size, as n = 10 is quite low. Another approach to achieve this effect may be to perform a subgroup analysis. The curve type is a logical criteria to classify the subgroups as it has been shown that progression is correlated to curve type (Bunnell, 1986; Lonstein & Carlson, 1984; Weinstein, 1999; Asher & Burton, 2006). However, to perform such a subgroup analysis would require a much larger sample size to ensure each subgroup is adequately populated. As there are so few subjects in each curve type category in the current study, it is difficult to provide an accurate estimate of the numbers required for a subgroup analysis. To do so would require accurate estimates of the population means and standard deviations for the ST indices, for each curve type.

4.3 Limitations

In addition to the possibilities presented above to improve the ST imaging system, there are a number of limitations of the current study that should be addressed.

4.3.1 Quantification of MDC

Many sources of variability cumulate in the calculation of the ST indices (section 1.4.3). For this study, a conscious decision was made to only include the sources of variability introduced after data collection (Figure 2.3). There are beneficial reasons to this approach. It assists in identifying the sources of variability and therefore ways to improve upon it can be formulated. Conversely, accounting for sources of variability subsequent to data collection results in variability quantification that is known to be

incomplete. Therefore, in a true clinical setting, the variability would be somewhat greater than that estimated in this thesis by the MDC.

4.3.2 Comparison of MDC to MCID

An average of repeated measures were used to calculate the within-subject change scores to reduce the random variability introduced by processing. However, the change scores require ST scans from different time points. Therefore, these change scores inherently incorporate variability introduced before data collection such as changes in posture or marker placement on anatomical landmarks. These sources of variability were not explicitly taken into account in the MDC and cannot be separated from the changes due to the progression of AIS. In addition to these extra sources of variability, there are influences of time between the scans that do not apply when discussing a single time point, such as significant growth, weight gain, or development such as puberty that affects torso shape.

4.3.3 Estimation of MCID

Hypothesis 1 (H1) states the ST indices will demonstrate significantly more change in a clinically progressed group of AIS patients than a non-progressed group of AIS patients. The strength of H1was tested by comparing the change in ST indices in a clinically progressed group (calculated in SA2) to the change in ST indices in a non-progressed group (calculated in SA3). The choice of definition for the MCID was based on H1 being true (i.e. the change in ST indices demonstrated by the clinically progressed group defined the MCID). Therefore, by extension, testing of H1 also tests the strength of the

assumption that the change in ST indices shown by the clinically progressed group defines the MCID of the ST indices.

Other than the change in Cobb angle between time points, the non-progressed group was statistically similar to the progressed group (Table 3.7). The expectation based on H1 was that the change in ST indices in the non-progressed group would be significantly less than the progressed group. However, the data did not support this hypothesis. No statistically significant ($p < 0.05$) differences between groups were found for any of the ST indices. Any weaknesses in describing the scoliotic deformity associated with the Cobb angle will be propagated to the current study. This is a result of defining the progressed and non-progressed groups by the Cobb angle. As the Cobb angle is fundamentally inadequate to describe the full 3D nature of the scoliotic deformity, the ST indices may detect aspects of the deformity that cannot be detected by the Cobb angle. This could partially explain the unexpected lack of support for H1.

The support for H1 through the results of SA2 and SA3 was not as strong as expected. The ST indices are designed to measure asymmetry. Therefore, the within-subject change scores measure changes in asymmetry. It was assumed that the only asymmetrical changes to the torso shape would be due to scoliosis as opposed to normal growth, weight gain, or pubertal development. Consequently, it was expected that there would be statistically significant differences between the progressed and non-progressed groups. While the results were contrary to the expectation, there are a number of plausible explanations for the study results, which are described below.

1. Other factors affecting torso shape may also be asymmetrical. For instance, posture has many degrees of freedom and is not constrained to being symmetrical. For example, an AIS patient who is self-conscious of her waistline asymmetry may develop the habit of standing with more weight on one foot than the other in an effort to minimize the appearance of her asymmetry. Then, if her scoliotic deformity worsens she may compensate even more with her posture. Or likewise, a patient with a shoulder asymmetry may tend to shrug the lower shoulder to bring it to an equal height. These types of postural adjustments could serve to mask the asymmetry of the torso caused by the scoliotic deformity.
2. The groups may not truly represent mutually exclusive categories of progressed/non-progressed. The choice of using a change in Cobb angle of 10° or greater to determine the clinically progressed group was taken over a change of 15° to be more conservative in the assessment on the ST system's precision. However, some reports state the MDC of the Cobb angle is approximately 10° at a 95% confidence interval (Carman et al. 1990). It is possible that using 10° is too small to be sufficiently confident that only truly progressed subjects were included. Additionally, subjects in the non-progressed group may have true scoliosis progression, but this may not be reflected with respect to the Cobb angle. By using the Cobb angle as the differentiator between groups, the simplifications and assumptions associated with the Cobb angle measurement are propagated through to the study groups. This fact must be stressed as the use of the Cobb angle in differentiating groups – scoliotic vs. non-scoliotic (Kane, 1977) and

progressed vs. non-progressed (Lonstein & Carlson, 1984), is often cautioned to be a guideline and not an absolute.

3. The use of the positioning frame may help to eliminate sway and increase positional repeatability (Robu, 2006). However, there is the potential that the positioning frame forces the AIS patient out of their deformed neutral position and closer to a symmetrical neutral position. This situation is easy to envision if one considers that the subject's pelvis is aligned within the positioning frame, and then if the arm rests were spaced symmetrically about the pelvis, it would require the subject to conform to that posture to some degree. By doing so, this may reduce some of the rotational indices such as BSR or PAX. This situation would be analogous to forcing a healthy subject to stand with their pelvis facing forward and then turning their shoulders 45°, and claiming that the large rotations in BSR or PAX indicate they have a worsening scoliosis.
4. As individual measurements will vary randomly about the true value (as illustrated in Figure 2.1), and the mean value of an infinite number of repeated measures will converge upon the true value, there is the possibility that the calculated MCID values have not sufficiently converged on true value with the limited number of repeated measures.

A more appropriate assumption would be that all subjects would show change in ST indices over time due to multiple factors (section 3.3). If this were adopted, a more appropriate MCID would be the excess change in ST indices demonstrated by the clinically progressed group over the non-progressed group. This quantity would provide

a more stringent measure with which to qualify the MDC. At present, there is no statistically significant difference between the groups. Consequently, it is a challenge to quantify this "excess change." However, a subgroup analysis may also be useful in this circumstance to provide a solution by reducing the between-subject variability within the groups.

4.4 Summary of Findings

In summary, the key findings of this study are:

1. With the current estimation of MCID, all indices except BSR are acceptable for typical research implementation.
2. With the current estimation of MCID, only aspect ratio is acceptable for typical clinical implementation.
3. H1 is not supported. Increasing study sample size or performing a subgroup analysis has potential to strengthen support of H1.
4. If strengthened support of H1 is not obtained, then a different estimation of MCID based on the "excess change" in the clinically progressed group over the non-progressed group should be investigated.

4.5 Future Work and Conclusions

The findings presented in this thesis show exciting potential that should be explored with an expanded study. As this initial study was performed with low sample sizes, it is difficult to show statistical significance. However, it appears that the aspect ratio index shows potential to differentiate between clinically progressed and non-progressed groups of AIS patients. This index also demonstrated the best performance in system precision when compared to the magnitude of change expected in a clinically progressed group. Sample size calculations (Brant, 2012, Inference for Means) on the aspect ratio index show that for the given sample means and standard deviations, a sample size between 28 and 47 would be required for a significance level of 0.05 and a power of 0.80. This is a realistic number of subjects to recruit given the previous recruitment rates at the Alberta Children's Hospital. In line with this suggestion to increase the sample sizes of the study, it is believed that a subgroup analysis based on curve type may prove fruitful. Any ability to reduce the between-subject variability within the progressed and non-progressed groups would serve to both increase the statistical significance of the difference in means between groups, as well as ease the requirements on system precision to more attainable levels.

Future studies regarding the influence of posture should be considered, as it is believed to have a significant ability to affect the asymmetry of the torso. The design of such an investigation must incorporate a method to define a scoliotic patient's neutral or natural position. As soon as one attempts to align or constrain the torso with the bias of what is

normal, it may very well influence the shape of the torso to be more normal or symmetrical due to the multitude of degrees of freedom within the torso.

In line with these concepts regarding posture, perhaps new imaging protocols or ST indices could be developed to help identify changes to the scoliotic deformity more reliably. As illustrated by the forward bend test (Adams, 1882), certain postures emphasize various aspects of the scoliotic deformity. Perhaps imaging an AIS patient in a forward bend position or a voluntary maximal rotation to the right and left may emphasize certain aspects of the deformity that could be detected and quantified with the ST system. Up to this point, all of the imaging protocols used by the Scoliosis Research Group at the U of C have placed the subject in a neutral standing position. With regard to the development of new ST indices, a preliminary study has been undertaken to examine the ability of the first and second derivatives of the current ST indices to indicate change of the scoliotic deformity. The theory behind this concept is that for a rotational index such as PAX, an individual can voluntarily change the magnitude or first derivative (rate of change) simply by rotating their torso. However, it would be more difficult to voluntarily change direction of rotation in different parts of the torso. In scoliosis it has been shown that the vertebrae become more rotated in the transverse plane as they approach the apex of the curve (Dickson, 1999), thus there is a change in the direction of rotation at the apex. Therefore, looking for minimum/maximum and inflection points in the ST index curves may provide better indicators of the scoliotic deformity, or a larger deviation from a non-AIS torso.

The current study could be expanded to include investigation of all sources of variability, introduced both before data collection and after data collection to quantify the components of variability introduced by the subject and the research nurse. To address this limitation, a future study could be designed that initially utilizes a scoliotic mannequin and captures multiple ST scans. This would allow the quantification of day-to-day variability in the ST indices that is more inclusive than the current study. This could account for factors such as changes in positioning of the subject and cameras relative to each other, various lighting and environmental conditions, and the collection of different point clouds of data, all while maintaining the same geometric shape. Subsequent steps could introduce repeated ST scans of human subjects that would allow for the quantification of variability in ST indices introduced by human factors such as changes in posture, motion artifacts, marker placement, and day-to-day growth/weight gain.

Another aspect that could be incorporated into future studies would be to include other measures of progression of AIS, such as patient-centric questionnaires as performed by Gorton et al. (2012). This could lessen the dependence of the study on the Cobb angle measurement and its associated weaknesses.

In addition to the intended use of the ST system as described in the current study, there are other potential applications of the ST system that could be further explored. These include, but are not limited to:

- Screening tool: In this application, the key variables would be the ST indices at a single point in time (as opposed to a change in ST indices over time). The populations of interest would be an AIS population (progressed or non-progressed) and a non-scoliotic or healthy control population. Differences between populations could possibly allow for the detection of the presence of a scoliotic deformity and initiate treatment.
- Treatment evaluation: In this application, the correction of the scoliotic deformity could be quantified by the ST indices and indicate the degree of success of various conservative or surgical interventions. Ideally, this could be extended to be able to predict the expected treatment outcome. Such a tool could allow an AIS patient to visualize the expected outcome before treatment such that they can make an informed decision on whether or not the risks associated with treatment are worth the expected outcome.
- Custom brace design and manufacturing: In this application, the ST system could be used to capture torso geometry in a similar fashion to the manual casting methods that are currently used in brace manufacturing. Then, by allowing for digital modification of the torso shape, a positive form can be manufactured by CNC milling machines. Such a system could reduce the extensive manual labour required to manufacture a custom-made scoliosis brace.

This study has further shown the clinical utility of the ST imaging system being developed by the Scoliosis Research Group at the U of C. It has quantified the variability introduced through processing the data and compared it to an expected change. With the given estimation of the MCID, the ST index of aspect ratio was demonstrated to be sufficiently precise to allow for use in a clinical implementation. This study has shown promise for the clinical utility of the ST system and has yielded encouraging results warranting further investigation. Once the ST system has been refined further, it would allow for the monitoring of progression of AIS without harmful ionizing radiation and would provide a more complete description of the scoliotic deformity. This achievement would overcome the two main drawbacks to the radiographic monitoring of AIS with the Cobb angle, which is the current standard of care. Such a realization would improve the course of treatment for AIS patients and would likely be applicable to other forms of scoliosis.

References

Adams, W. (1882). Lectures on the pathology and treatment of lateral and other forms of curvature of the spine. Churchill.

Asher, M. A., & Burton, D. C. (2006). Adolescent idiopathic scoliosis: natural history and long term treatment effects. Scoliosis, 1(2), 1–10. doi:10.1186/1748-7161-1-2

Aubin, C.-E., Bellefleur, C., Joncas, J., de Lanauze, D., Kadoury, S., Blanke, K., Parent, S., et al. (2011). Reliability and accuracy analysis of a new semiautomatic radiographic measurement software in adult scoliosis. Spine, 36(12), E780–E790. doi:10.1097/BRS.0b013e3181f0825a

Beaton, D. E., Boers, M., & Wells, G. A. (2002). Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research. Current opinion in rheumatology, 14(2), 109–114.

Bland, J. M., & Altman, D. G. (1996). Statistics notes: measurement error. Bmj, 313(7059), 744.

Brant, R. (2012). Inference for means. In Comparing two independent samples. Retrieved October 30, 2012, from http://www.stat.ubc.ca/~rollin/stats/ssize/n2.html

Bunnell, W. P. (1984). An objective criterion for scoliosis screening. The Journal of bone and joint surgery. American volume, 66(9), 1381–1387.

Bunnell, W. P. (1986). The natural history of idiopathic scoliosis before skeletal maturity. Spine, 11(8), 773–776.

Carman, D. L., Browne, R. H., & Birch, J. G. (1990). Measurement of scoliosis kyphosis radiographs. The Journal of bone and joint surgery. American volume, 72(3), 328–333.

Clayson, D., Luz-Alterman, S., Cataletto, M. M., & Levine, D. B. (1987). Long-term psychological sequelae of surgically versus nonsurgically treated scoliosis. Spine, 12(10), 983–986.

Cobb J (1948). Outline for the study of scoliosis. American Academy of Orthopaedic Surgeons: Instructional Course Lectures. Ann Arbor, MI, J. W. Edwards. 5: 251.

Dang, N. R., Moreau, M. J., Hill, D. L., Mahood, J. K., & Raso, J. (2005). Intra-observer reproducibility and interobserver reliability of the radiographic parameters in the Spinal Deformity Study Group's AIS Radiographic Measurement Manual. Spine, 30(9), 1064–1069.

Dansereau, J., & Stokes, I. A. F. (1988). Measurements of the three-dimensional shape of the rib cage. Journal of biomechanics, 21(11), 893–901.

Deschênes, S., Charron, G., Beaudoin, G., Labelle, H., Dubois, J., Miron, M.-C., & Parent, S. (2010). Diagnostic imaging of spinal deformities: reducing patients radiation dose with a new slot-scanning X-ray imager. Spine, 35(9), 989–994. doi:10.1097/BRS.0b013e3181bdcaa4

Dickson, R. A. (1999). Spinal deformity-adolescent idiopathic scoliosis: Nonoperative treatment. Spine, 24(24), 2601–2606.

Don, S. (2004). Radiosensitivity of children: potential for overexposure in CR and DR and magnitude of doses in ordinary radiographic examinations. Pediatric radiology, 34 Suppl 3, S167–172; doi:10.1007/s00247-004-1266-9

Doody, M. M., Lonstein, J. E., Stovall, M., Hacker, D. G., Luckyanov, N., & Land, C. E. (2000). Breast cancer mortality after diagnostic radiography: findings from the U.S. Scoliosis Cohort Study. Spine, 25(16), 2052–2063.

Dubetz, T.P.A., Smith, K.N., Küpper, J.C., Howard, J.J., Harder, J.A., Joughin, V.E., Ronsky, J.L. (2011a). Detecting clinically significant progression in adolescent idiopathic scoliosis using surface topography indices. Proceedings of the Eleventh Annual Scientific Conference of the Canadian Spine Society, Fairmont Chateau Frontenac, Quebec City, Canada, March 9-12, 2011. (Podium Presentation).

Dubetz, T.P.A., Küpper, J.C., Smith, K.N., Howard, J.J., Harder, J.A., Ronsky, J.L. (2011b). Non-invasive surface imaging for the diagnosis, monitoring, and treatment of scoliosis. Proceedings of the Alberta Graduate Conference – Today's Ideas, Tomorrow's Innovators. MacEwan Hall, University of Calgary, Calgary, Canada, May 5-7, 2011. (Podium Presentation).

EOS Imaging. (2010). Workflow. In EOS: A radically new vision for orthopedic imaging. Retrieved October 9, 2012 from http://www.eos-imaging.com/index.php?page=eos&hl=en_EN

Goldsmith, C. H., Boers, M., Bombardier, C., & Tugwell, P. (1993). Criteria for clinically important changes in outcomes: development, scoring and evaluation of rheumatoid arthritis patient and trial profiles. OMERACT Committee. The Journal of Rheumatology, 20(3), 561–565.

Gorton, G. E., Young, M. L., & Masso, P. D. (2012). Accuracy, reliability, and validity of a 3-dimensional scanner for assessing torso shape in idiopathic scoliosis. Spine, 37(11), 957–965. doi:10.1097/BRS.0b013e31823a012e

Grivas, T. B. (2002). Surgery is performed for cosmetic reasons. Studies in health technology and informatics, 116-122.

Grivas, T. B., Vasiliadis, E. S., Mihas, C., & Savvidou, O. (2007). The effect of growth on the correlation between the spinal and rib cage deformity: implications on idiopathic scoliosis pathogenesis. Scoliosis, 2:11, 1–6. doi:10.1186/1748-7161-2-11

Hopkins, W. G. (2000). Measures of reliability in sports medicine and science. Sports Medicine (Auckaland, New Zealand), 30(1), 1–15.

James, J. I. P. (1954). Idiopathic scoliosis; the prognosis, diagnosis, and operative indications related to curve patterns and the age at onset. The Journal of bone and joint surgery. British volume., 36-B(1), 36–49.

Jaremko, J. L. (2001). Estimation of scoliosis severity from the torso surface by neural networks (Doctoral dissertation, University of Calgary, Calgary, Canada).

Jaremko, J L, Poncet, P., Ronsky, J., Harder, J., Dansereau, J., Labelle, H., & Zernicke, R. F. (2001). Estimation of spinal deformity in scoliosis from torso surface cross sections. Spine, 26(14), 1583–1591.

Jaremko, Jacob L, Poncet, P., Ronsky, J., Harder, J., Dansereau, J., Labelle, H., & Zernicke, R. F. (2002). Indices of torso asymmetry related to spinal deformity in scoliosis. Clinical biomechanics, 17(8), 559–568.

Kane, W. J. (1977). Scoliosis prevalence: a call for a statement of terms. Clinical orthopaedics and related research, (126), 43-46.

Kane, W. J., & Moe, J. H. (1970). A scoliosis-prevalence survey in Minnesota. Clinical orthopaedics and related research, (69), 216–218.

Kim, H., Kim, H. S., Moon, E. S., Choon-Sik, Y., Chung, T.-S., Song, H.-T., Suh, J.-S., et al. (2010). Scoliosis imaging: what radiologists should know. RadioGraphics, 30(7), 1823–1842. doi:10.1148/rg.307105061

King, H. A., Moe, J. H., Bradford, D. S., & Winter, R. B. (1983). The selection of fusion levels in thoracic idiopathic scoliosis. The Journal of bone and joint surgery. American volume, 65(9), 1302–1313.

Kotwicki, T. (2008). Evaluation of scoliosis today: examination, X-rays and beyond. Disability and rehabilitation, 30(10), 742–751. doi:10.1080/09638280801889519

Krismer M. Sterzinger W. Haid C. Frischhut B. Bauer R. (1996). Axial rotation measurement of scoliotic vertebrae by means of computed tomography scans. Spine, 21(5): 576-581.

Labelle, H., Dansereau, J., Bellefleur, C., & Jéquier, J.-C. (1995). Variability of geometric measurements from three-dimensional reconstructions of scoliotic spines and rib cages. European spine journal : official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society, 4(2), 88–94.

Lenke, L. G., Betz, R. R., Harms, J., Bridwell, K. H., Clements, D. H., Lowe, G. T., & Blanke, K. (2001). Adolescent idiopathic scoliosis: a new classification to determine extent of spinal arthrodesis. Spine, 83-A(8), 1169–1181. doi:10.1097/00007632-200101010-00009

Levy AR. Goldberg MS. Mayo NE. Hanley JA. Poitras B. (1996). Reducing the lifetime risk of cancer from spinal radiographs among people with adolescent idiopathic scoliosis. Spine, 21(13): 1540-1548.

Liljenqvist, U. R., Halm, H. F., & Link, T. M. (1997). Pedicle screw instrumentation of the thoracic spine in idiopathic scoliosis. Spine, 22(19), 2239-2245.

Liljenqvist, U R, Link, T. M., & Halm, H. F. (2000). Morphometric analysis of thoracic and lumbar vertebrae in idiopathic scoliosis. Spine, 25(10), 1247–1253.

Little, J. P., Izatt, M. T., Labrom, R. D., Askin, G. N., & Adam, C. J. (2012). Investigating the change in three dimensional deformity for idiopathic scoliosis using axially loaded MRI. Clinical biomechanics (Bristol, Avon), 27(5), 415–421. doi:10.1016/j.clinbiomech.2011.12.004

Lonstein, J. E., & Carlson, J. M. (1984). The prediction of curve progression in untreated idiopathic scoliosis during growth. The Journal of bone and joint surgery. American volume, 66(7), 1061–1071.

Machida, M. (1999). Cause of idiopathic scoliosis. Spine, 24(24), 2576–2583.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. Psychological Methods, 1(1), 30–46. doi:10.1037//1082-989X.1.1.30

Moore KL (1980). The Back. Clinically Oriented Anatomy. Baltimore, Williams & Wilkins: 605-673.

Morrissy, R. T., Goldsmith, G. S., Hall, E. C., Kehl, D., & Cowie, G. H. (1990). Measurement of the Cobb angle on radiographs of patients who have scoliosis. Evaluation of intrinsic error. The Journal of bone and joint surgery. American volume, 72(3), 320–327.

Nunnally, J. C., & Bernstein, I. H. (1994). Psychometric theory (3rd ed., pp. 261–262). McGraw-Hill Inc.

Oestreich, A. E., Young, L. W., & Young Poussaint, T. (1998). Scoliosis circa 2000: radiologic imaging perspective. I. Diagnosis and pretreatment evaluation. Skeletal radiology, 27(11), 591–605.

Patias, P., Grivas, T. B., Kaspiris, A., Aggouris, C., & Drakoutos, E. (2010). A review of the trunk surface metrics used as Scoliosis and other deformities evaluation indices. Scoliosis, 5(12), 1–20.

Pazos, V, Cheriet, F., Song, L., Labelle, H., & Dansereau, J. (2005). Accuracy assessment of human trunk surface 3D reconstructions from an optical digitising system. Medical & biological engineering & computing, 43(1), 11–15.

Pazos, V., Cheriet, F., Danserau, J., Ronsky, J., Zernicke, R. F., & Labelle, H. (2007). Reliability of trunk shape measurements based on 3-D surface reconstructions. European spine journal : official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society, 16(11), 1882–1891. doi:10.1007/s00586-007-0457-0

Poncet, P., Delorme, S., Ronsky, J. L., Dansereau, J., Clynch, G., Harder, J., & Zernicke, R. F. (2000). Reconstruction of laser-scanned 3D torso topography and stereoradiographical spine and rib-cage geometry in scoliosis. Computer Methods in Biomechanics and Biomedical Engineering, 4(1), 59–75.

Pope MH. Stokes AF. Moreland M. (1984). The biomechanics of scoliosis. CRC Crit Rev Biomed Eng, 11(3): 157-188.

Rankine, L., Liu, X. C., Tassone, C., Lyon, R., Tarima, S., & Thometz, J. (2012). Reproducibility of newly developed spinal topography measurements for scoliosis. The open orthopaedics journal, 6, 226–230. doi:10.2174/1874325001206010226

Razali, N. M., & Wah, Y. B. (2011). Power comparisons of Shapiro-Wilk , Kolmogorov-Smirnov , Lilliefors and Anderson-Darling tests. Journal of Statistical Modeling and Analytics, 2(1), 21–33.

Rigo, M., Reiter, C., & Weiss, H.-R. (2003). Effect of conservative management on the prevalence of surgery in patients with adolescent idiopathic scoliosis. Pediatric rehabilitation, 6(3-4), 209–214. doi:10.1080/13638490310001642054

Robu, D. (2006). Assessment of 3D reconstruction of scoliotic human torso using optical imaging techniques (Master’s thesis, University of Calgary, Calgary, Canada).

Sakka, S. A., Wojcik, A., Macindoe, S., & Mehta, M. H. (1995). Reproducibility and reliability of the Quantec Surface Imaging System in the assessment of spinal deformity. In Proceedings of the 2nd International Symposium on Three-Dimensional Scoliotic Deformities combined with the 8th International Symposium on Surface Topography and Spinal Deformity. In: D'Amico et al, eds. Three Dimensional Analysis of Spinal Deformity. The Netherlands: IOS Press (pp. 441-5).

Scoliosis Research Society. (2012). Imaging Studies. In Adolescent Idiopathic Scoliosis. Retrieved December 18, 2012, from http://www.srs.org/professionals/conditions_and_treatment/adolescent_idiopathic_scoliosis/imaging_studies.htm

Scoliosis Research Society. (2012). Physical findings. In Idiopathic Scoliosis: Adolescents. Retrieved August 4, 2012, from http://www.srs.org/patient_and_family/scoliosis/idiopathic/adolescents/physical_findings.htm

Scoliosis Research Society. (2012). Treatment. In Adolescent Idiopathic Scoliosis. Retrieved August 3, 2012, from http://www.srs.org/professionals/conditions_and_treatment/adolescent_idiopathic_scoliosis/treatment.htm

Scoliosis Research Society. (2012). What is scoliosis? In Common scoliosis questions. Retrieved August 3, 2012, from http://www.srs.org/patient_and_family/scoliosis/

Séoud, L., Dansereau, J., Labelle, H., & Cheriet, F. (2012). Multilevel analysis of trunk surface measurements for noninvasive assessment of scoliosis deformities. Spine, 37(17), E1045–E1053. doi:10.1097/BRS.0b013e3182575938

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. Psychological bulletin, 86(2), 420–428.

Silva, F.E., & Lenke, L.G. (2009). Adolescent Idiopathic Scoliosis. In T.J. Errico, B.S. Lonner & A.W. Moulton (Eds.), Surgical management of spinal deformities (pp.97-118). Philadelphia, PA: Saunders Elsevier.

Srinivasalu, S., Modi, H. N., Smehta, S., Suh, S., Chen, T., & Murun, T. (2008). Cobb angle measurement of scoliosis using computer measurement of digitally acquired radiographs-intraobserver and interobserver variability. Asian Spine Journal, 2(2), 90–93.

Stokes IA. Bigalow LC. Moreland MS. (1987). Three-dimensional spinal curvature in idiopathic scoliosis. J Orthop Res, 5(1): 102-113.

Stokes IA. (1994). Three-dimensional terminology of spinal deformity. A report presented to the Scoliosis Research Society by the Scoliosis Research Society Working Group on 3-D terminology of spinal deformity. Spine, 19(2): 236-248.

Stratford, P. W. (2004). Getting more from the literature: estimating the standard error of measurement from reliability studies. Physiotherapy Canada, 56(01), 27–30. doi:10.2310/6640.2004.15377

Swanson, S.L. (2008). Estimating scoliosis progression from three dimensional torso shape (Master's thesis, University of Calgary, Calgary, Canada).

Tanure, M. C., Pinheiro, A. P., & Oliveira, A. S. (2010). Reliability assessment of Cobb angle measurements using manual and digital methods. The spine journal : official journal of the North American Spine Society, 10(9), 769–774. doi:10.1016/j.spinee.2010.02.020

Taylor, J. R. (1997). An introduction to error analysis: the study of uncertainties in physical measurements. Univ Science Books.

Terver, S., Kleinman, R., & Bleck, E. E. (1980). Growth landmarks and the evolution of scoliosis: a review of pertinent studies on their usefulness. Developmental medicine and child neurology, 22(5), 675–684.

Theologis TN. Fairbank JC. Turner-Smith AR. Pantazopoulos T. (1997). Early detection of progression in adolescent idiopathic scoliosis by measurement of changes in back shape with the Integrated Shape Imaging System scanner. Spine, 22(11): 1223-1227.

Veldhuizen, A. G., Cheung, J., Bulthuis, G. J., & Nijenbanning, G. (2002). A new orthotic device in the non-operative treatment of idiopathic scoliosis. Medical engineering & physics, 24(3), 209–218.

Wang, W. J., Yeung, H. Y., Chu, W. C.-W., Tang, N. L.-S., Lee, K. M., Qiu, Y., Burwell, R. G., et al. (2011). Top theories for the etiopathogenesis of adolescent idiopathic scoliosis. Journal of pediatric orthopedics, 31(1 Suppl), S14–S27. doi:10.1097/BPO.0b013e3181f73c12

Weinstein, S. L. (1999). Natural history. Spine, 24(24), 2592–2600. doi:10.1097/00007632-199912150-00006

Weinstein, S. L., Dolan, L. A., Spratt, K. F., Peterson, K. K., Spoonamore, M. J., & Ponseti, I. V. (2003). Health and function of patients with untreated idiopathic scoliosis: a 50-year natural history study. JAMA: The Journal of the Amercian Medical Association, 289(5), 559–567.

Weir, J. P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. Journal of Strength & Conditioning Research, 19(1), 231–240.

Weiss, H.-R. (2008). Adolescent idiopathic scoliosis (AIS) - an indication for surgery? A systematic review of the literature. Disability and rehabilitation, 30(10), 799–807. doi:10.1080/09638280801889717

Wu, H., Ronsky, J. L., Cheriet, F., Küpper, J., Harder, J., Xue, D., & Zernicke, R. F. (2010). Prediction of scoliosis progression with serial three-dimensional spinal curves and the artificial progression surface technique. Medical & biological engineering & computing, 48(11), 1065–1075. doi:10.1007/s11517-010-0654-6

Yazici, M., Acaroglu, E. R., Alanay, A., Deviren, V., Cila, A., & Surat, A. (2001). Measurement of vertebral rotation in standing versus supine position in adolescent idiopathic scoliosis. Journal of pediatric orthopedics, 21(2), 252–256.

Zhang, J., Lou, E., Shi, X., Wang, Y., Hill, D. L., Raso, J. V., Le, L. H., et al. (2010). A computer-aided Cobb angle measurement method and its reliability. Journal of Spinal Disorders & Techniques, 23(6), 383–387.

Appendix 1: ST Index Data from ANOVA Tables

	Observer 1 (K)				
	MSe	ICC(2,1)	ICC Low 95% CI	ICC High 95% CI	MDC
BSR	0.166	0.997	0.994	0.999	1.129
PAX	0.525	0.991	0.978	0.997	2.008
Rib Prom	0.418	0.996	0.990	0.999	1.792
Quarter Area Diff	1.883	0.997	0.992	0.999	3.804
Aspect Ratio	3.461	0.993	0.983	0.998	5.157
Lateral Centroid Line	1.002	0.994	0.985	0.998	2.775
Spinous Process Line	0.823	0.990	0.974	0.997	2.515
Lateral Inertia	106.085	0.978	0.945	0.993	28.549
AP Inertia	3.847	0.990	0.974	0.997	5.437
Kyphosis	0.344	0.996	0.991	0.999	1.626

	Observer 2 (T)				
	MSe	ICC(2,1)	ICC Low 95% CI	ICC High 95% CI	MDC
BSR	0.568	0.993	0.983	0.998	2.089
PAX	0.441	0.994	0.985	0.998	1.841
Rib Prom	1.607	0.985	0.964	0.995	3.514
Quarter Area Diff	0.915	0.999	0.996	1.000	2.651
Aspect Ratio	2.803	0.994	0.986	0.998	4.641
Lateral Centroid Line	1.406	0.991	0.976	0.997	3.287
Spinous Process Line	7.278	0.920	0.813	0.973	7.478
Lateral Inertia	160.493	0.964	0.912	0.988	35.116
AP Inertia	2.686	0.993	0.983	0.998	4.543
Kyphosis	0.139	0.998	0.996	0.999	1.033

	Total-Observer				
	MSe	ICC(2,1)	ICC Low 95% CI	ICC High 95% CI	MDC
BSR	1.418	0.981	0.959	0.993	3.301
PAX	1.263	0.981	0.960	0.993	3.115
Rib Prom	2.655	0.975	0.947	0.991	4.517
Quarter Area Diff	1.921	0.997	0.993	0.999	3.842
Aspect Ratio	4.165	0.992	0.982	0.997	5.657
Lateral Centroid Line	1.468	0.991	0.980	0.997	3.358
Spinous Process Line	22.052	0.744	0.555	0.896	13.017
Lateral Inertia	152.382	0.967	0.931	0.988	34.217
AP Inertia	3.181	0.992	0.982	0.997	4.944
Kyphosis	0.329	0.996	0.992	0.999	1.590

	Inter-Observer				
	MSe	ICC(2,1)	ICC Low 95% CI	ICC High 95% CI	MDC
BSR	1.873	0.975	0.920	0.992	3.793
PAX	1.460	0.978	0.930	0.993	3.349
Rib Prom	3.076	0.971	0.908	0.991	4.861
Quarter Area Diff	1.336	0.998	0.993	0.999	3.204
Aspect Ratio	2.766	0.995	0.982	0.998	4.610
Lateral Centroid Line	0.841	0.995	0.983	0.998	2.542
Spinous Process Line	31.352	0.624	0.137	0.868	15.520
Lateral Inertia	76.251	0.983	0.946	0.995	24.204
AP Inertia	0.946	0.998	0.992	0.999	2.696
Kyphosis	0.226	0.998	0.992	0.999	1.318