Schroth Exercises for Adolescent Idiopathic Scoliosis – Reliability, A Randomized Controlled Trial and Clinical Significance

by

Sanja Schreiber

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Rehabilitation Science

Faculty of Rehabilitation Medicine
University of Alberta

Abstract

Background and objectives

In America, care recommendations for adolescents with idiopathic scoliosis (AIS) include observation, bracing, and surgery. In Europe exercises are often prescribed. Schroth scoliosis-specific exercises have demonstrated promising results, but only in studies of suboptimal quality.

Schroth exercise prescription is guided by curve classification. An algorithm for determining Schroth curve types was developed.

Reporting statistical significance should be supplemented with clinical significance estimates, which is rare for research on conservative treatment for scoliosis.

This thesis aimed: 1) to determine the reliability of Schroth therapists in classifying patients with AIS using the proposed algorithm; 2) to investigate the effect of Schroth exercises combined with standard of care on curve severity, quality-of-life (QOL), perceived appearance and back muscle endurance compared to standard of care; and 3) to determine the clinical significance of the effect of Schroth exercises for all outcomes.

Methods

For the reliability study, we recruited 44 participants with AIS and 10 consecutive volunteer Schroth-certified therapists. Therapists rated video assessments presented randomly twice at least seven days apart blinded to identities. The reliability was determined using Gwet's AC1 coefficients.

For the RCT, we randomized 50 patients with AIS, aged 10-18 years, with curves 10°-45° to standard of care (observation or bracing) or supervised Schroth exercises plus standard of care. After introducing Schroth exercises, a daily home program was adjusted during weekly supervised sessions for six months. Assessors and the statistician were blinded. The primary outcome was the change in the largest Cobb (LC) angle and the Sum of Cobb (SOC) angles. Secondary outcomes included Biering-Sorensen back muscle endurance test, Scoliosis Research Society (SRS-22r) and Spinal Appearance Questionnaires' (SAQ) scores. Per protocol and intention-to-treat linear mixed models analyses were reported.

Clinical significance was determined using anchor- and distribution-based methods. Numbers needed to treat (NNT), and proportion of improved, stable and deteriorated patients were reported.

Results

The overall intra-rater AC1 was 0.64 (95%CI 0.53-0.73), 0.70 (0.60-0.78) among well-trained raters, and 0.81 (0.77-0.85) in experienced raters. The weighted intra-rater AC1 averaged 0.75 (95%CI 0.63-0.84) overall, 0.82 (0.73-0.88) in well-trained raters, and 0.89 (0.80-0.94) in experienced raters. Inter-rater AC1 was 0.43 (95%CI 0.28-0.58) overall, 0.50 (0.38-0.61) for well-trained raters, and 0.67 (0.50-0.85) for experienced raters. The weighted inter-rater AC1 was 0.48 (95%CI 0.29-0.67) overall, 0.61 (0.49-0.72) among well-trained, and 0.79 (0.64-0.94) among experienced raters.

After six months, Schroth group had by 3.5° (p<0.01) smaller LC in the per protocol analysis. The difference in the square root of the SOC also favored Schroth group (p<0.05) such that a patient with an average 51.2° SOC at baseline had a 49.3° at six months in the Schroth group,

and 55° in the control group, while the difference increased with severity. Schroth exercises improved patients' back muscle endurance, by 30 seconds (p=0.02). Intention-to-treat results were similar in direction and magnitude, but did not reach statistical significance. No statistically significant differences were observed between groups on QOL and perceived appearance outcomes. Some covariates (age, weight, height, self-efficacy, brace wear, and Schroth classification) had important main effects on outcomes in different models.

Clinical significance was reached for the LC, LOC and SRS-22r function score, with their respective cutoffs at 3.4°, 8.7°, and 0.11. Their NNT were 3 and 4, and 4 respectively. Proportions of improved or stable patients were significantly larger in the Schroth group for all outcomes. Biering-Sorensen test cutoff was 36.15 sec. High rates of ceiling effects were observed on questionnaires' scores, which created problems in determining statistical and clinical significance. Treatment effects on the questionnaires' scores were smaller than the measurement error and not statistically significant suggesting that the questionnaires were not responsive to change over time. Distribution-based methods produced cutoffs, which exceeded the magnitude of changes commonly seen in this population.

Conclusions

1) Reliability of experienced and well-trained therapists was adequate when using the proposed algorithm.
2) Schroth exercises added to the standard of care led to statistically and clinically significant improvements in curve severity in patients completing the program.

3) Clinically significant improvement was observed for the Largest Cobb, Sum of Cobb angles, and SRS-22r function scores. More responsive questionnaires may be needed to quantify treatment effects on QOL and perceived appearance.

Preface

Some of the research conducted for this thesis was completed in collaborations with co-authors identified in the respective chapters. I was primary author of all the chapters in the thesis but received input on my manuscripts from co-authors for chapters 3, 4, 5, and 6. No part of this thesis has yet been previously published.

The algorithm referred to in Chapter 3 was designed by myself, with the assistance of Dr. Eric Parent. The chapters 3 and 6, and the discussion/conclusion in chapter 7 are my original work. Similarly, the introduction and literature review in chapters 1 and 2 are my original work.

The data analyses in chapters 4 and 5 were conducted by Elham Khodayari Moez. I was responsible for the data collection and providing the treatments to the participants. The data interpretation was done by myself, with the assistance of Elham Khodayari Moez.

The research projects included in this thesis, received research ethics approval from the University of Alberta Research Ethics Board, “Schroth Exercise Trial for Scoliosis (SETS study)”, Pro00011552, Sep 16, 2010 and “Schroth Reliability Study”, Pro00019547, Oct 4, 2011.

For Stefan, Sean and Sawyer

…with all my love

Acknowledgements

I would like to thank several people who helped in the process of finalizing my PhD thesis.

First of all, there are no words that can describe my sincere gratitude to my supervisor, Eric Parent. Thank you, Eric for all our research related and personal discussions and for providing constructive critiques, especially during this last year of my PhD. Although, sometimes I did not like seeing a long line of comments in my manuscripts, I am thankful for them now, because I know they made me a better scientific writer. Your insights were always appreciated and I learned a ton as your student. I am forever grateful to your timely responses for everything that needed to be taken care of: numerous scholarship applications, assignments, abstracts, presentations, manuscripts, and finally, this thesis. You were always there. Thank you, Eric for all your support during this journey and for always pushing me further to get me to where I am now!

I am also thankful to my supervisory committee Doug Hedden and Greg Kawchuk, for providing insightful comments and thoughts on my research projects. I would also like to thank Marc Moreau and Sharon Warren for being part of the committee during my candidacy exam. Thank you all for pushing me beyond my comfort zone during the candidacy. I am grateful for all the work I accomplished in the years that followed.

Thank you so much to Joanne Volden and Tammy Hopper for their encouragement and support in times when it was truly needed. Your kindness will be forever remembered!

Here, I would also like to thank Kathleen Shearer, our research coordinator, without whom I would, probably, still be recruiting patients! Thank to Doug Hill for always being so nice and supportive of my research. Also, thank you to Jocelyn Reekie for liaising effectively between her boss and me.

I would like to acknowledge Elham Khodayari Moez for conducting the analyses for one of my studies, and Yachen Wang for helping me understand the mixed model analysis interpretations.

Here, I would also like to thank my colleague, Susan Armijo Olivo for valuable statistic-related and personal discussions.

There were not very many students in our lab at the beginning of my studies, but recently that changed, and it made a big difference. Alan, Elise, Vahid thank you for all the laughs we had together.

Special thanks goes to the members of the Scoliosis Research Group: Edmond Lou, Sarah Southon, Jim Raso, François Roy, Kajsa Duke, and their students, for great discussions and feedback they provided during our weekly meetings.

I would also like to thank the funding sources: Faculties of Medicine and Dentistry and Rehabilitation Medicine joint scholarship, Scoliosis Research Society Small exploratory grant, that helped equip the lab for my Schroth study, Glenrose Hospital Foundation for providing funds for my reliability study, CIHR Institute of Human Development, Child and Youth Health for supporting my research travel.

There is a special person in my life that deserves special thanks! Draga mala mama moja, hvala ti što si uvek verovala u mene, i što si uvek bila moja najveća podrška u svemu što sam ikada postigla. [My dear little mama, thank you for always believing in me, and for being my biggest support in everything I have ever achieved.]

Finally, I would like to thank the best husband and sons in the world, Stefan, Sean & Sawyer! Seany & Scholo, I hope you will forgive mama for taking the time we should have spent together to finish this big work, and thank you for always reminding me that life is beautiful! Schnuto, thank you for always being there for me, and for helping me complete this hardest endurance race of my life! Your love and support kept me pushing forward, even through the most stressful times. I love you all so much! 💙

Table of contents

List of tables

List of figures

CHAPTER 1

Introduction

1.1 Background

Adolescent idiopathic scoliosis (AIS) is defined as "a structural, lateral, rotated curvature of the spine that arises in otherwise healthy children around puberty."[1] The Scoliosis Research Society (SRS) defines the diagnosis of scoliosis as an asymmetry on the forward bending test and a Cobb angle of ≥10°.[2] SRS defines the Cobb angle as an "angle between lines drawn on endplates of the end vertebrae".[3] The prevalence of AIS is between 1% and 12%.[4] The annual incidence is about 2%[5] based on a sample of 26,947 students followed up for 2 years. The highest prevalence of AIS with an equal ratio between genders is seen among smaller curves[2], while a female predominance occurs among patients with larger curves and increases with the curve magnitude up to 8:1 in curves over 50°.[1,2,6]

The management of AIS is limited by our lack of understanding of its etiology and pathogenesis.[1,7-9] Nevertheless, the Scoliosis Research Society (SRS) proposes general evidence-based guidelines for treating the scoliosis symptoms. Observation is suggested for growing patients with curves of <25° and skeletally mature patients with curves of <50°. Treatment by observation consists of regular monitoring of patients for curve progression, but with no specific treatment applied. Hence, observation can be regarded as the natural history. Bracing is recommended for growing patients with curves between 25° and 40° that have documented progression. Bracing may also be suggested for patients on the first visit if the curve is >25° and if they have significant growth remaining (e.g. premenarchal girls), or if the curve is >30° in children with at least one year of growth remaining. Surgery may be recommended for growing patients with curves >45° and for mature patients with curves >50°.[10]

In some patients, scoliosis deformity can progress and generate chronic pain, impact respiratory capacity, quality of life, activity participation, and self-image.[1,11-15] The adverse consequences of progressive scoliosis generally manifest once curves exceed 40°-50° and are lifelong.[1,14,16,17] Therefore, prevention of progression is an important research priority.

In a systematic review on the effect of exercises on scoliosis, Fusco and colleagues[18] reported that asymmetric exercises improved the cosmetic appearance, pain and slowed the progression (deterioration) of scoliosis. However, the exercise studies reviewed were not of adequate scientific quality, because most were retrospective and only one was an RCT, they did not report on compliance or recruitment strategies, and they did not include intention-to-treat analysis or blinded evaluators. The most recent review by Mordecai and Dabke was an independent review[19] that included nine prospective cohort studies, of which only three were controlled and only one used observer blinding. The authors indicated that selection criteria, recommendations, and contraindications to exercise were not clearly determined in any of these papers. Most recently, after this independent review became available, Monticone published results of an RCT that investigated the effect of scoliosis-specific active self-correction and task-oriented exercises compared to traditional spinal exercises on Cobb angles and QOL in patients with AIS.[20] The authors found significant improvement in the outcomes in the exercises group compared to control. This was a well-designed study, and the evaluators were blinded to the treatment allocation. However, the authors did not report on the compliance.

Several scoliosis-specific exercise approaches have been described in the literature including: Schroth, Integrated Scoliosis Rehabilitation (ISR), Dobomed, Side-shift, Lyon, Functional Individual Therapy of Scoliosis and the Scientific Exercise Approach to Scoliosis (SEAS).[18] Of those, the Schroth method has the longest history of existence (since 1921) and has been the most frequently published on. The Schroth method consists of sensorimotor, postural and breathing exercises individualized according to the patient's specific scoliosis curve pattern in order to correct the asymmetric posture in daily activities.[21,22] Several cohort studies demonstrated positive outcomes of Schroth exercises on back muscle strength[23], breathing function[23], slowing curve progression,[24] improving Cobb angles[23,24] and decreasing the incidence of surgery.[25] However, there was only one prospective uncontrolled study on the effect of Schroth exercises in patients with AIS. Therefore, for this thesis, we conducted the first RCT to evaluate the effect of the promising Schroth approach on curve characteristics, back muscle endurance and quality of life in children with AIS.

The Schroth approach uses a specific classification system to categorize patients with scoliosis into four curve patterns (3c, 3cp, 4c and 4cp)[21,26]. Because an appropriate classification guides

adequate exercise selection, the reliability of therapists in classifying people with scoliosis according to Schroth is important. For the purposes of this thesis, we developed and tested a Schroth classification algorithm to unambiguously guide therapists in determining the curve types.

Many authors agree that reporting statistical test results should be supplemented with methods for determining what constitutes clinically significant change.[27-29] Various definitions to determine clinically meaningful effects or changes have been proposed[28,30], but essentially all are used to decide if a study result ‘‘matters in the real world of clinical medicine.’’[31] Clinically important change may be important to patients, clinicians, or both. Discussion about to whom a change is important depends on the study objective. To determine the clinically meaningful effect of the therapy anchor- and distribution-based methods[30-36] are most frequently used. While the anchor-based methods require that the change in the scores be compared to an external measure (anchor), such as the patient's perception of change[33,36-38], distribution-based methods rely on the statistical distribution of the results and the psychometric properties of the outcome measure to determine whether a clinically significant change has occurred.[32,33,36] The clinical significance of the therapy effect observed is typically reported as the number and percentage of patients who improved, stayed unchanged or deteriorated based on whether the patient has experienced change exceeding, either an anchor- or distribution-based threshold for important change. Ideally, a clinically important difference should be meaningful both to patients and clinicians.

In scoliosis research some attempts were made to assess the clinically significant effect of the scoliosis surgery by determining the MCID from the SRS-22r questionnaire, which is routinely used to assess the quality of life of patients with scoliosis.[39,40] However, no research has been done in defining what constitutes a clinically significant effect in the conservative management of scoliosis. The magnitude of clinically important effects can vary between different conservative and surgical treatments for AIS, because their goals, and risks involved differ. Evaluating the clinical significance of outcomes, specifically in the context of an exercise trial is therefore needed.

1.2 Thesis outline and objectives

The present thesis consists of a general literature review, four individual research chapters, as well as, a general discussion and conclusion. The main objective was to determine the effects of the Schroth exercises on curve characteristics, back muscle endurance and quality of life in adolescents with idiopathic scoliosis in a RCT. The purpose was to provide much needed evidence to inform global clinical practice about the effectiveness of the Schroth exercises used as an additional treatment to the standard of care for adolescent idiopathic scoliosis, which generally consists only of monitoring, bracing and surgery.

In the Chapter 2, an overview of general literature pertaining to the questions investigated in this thesis has been provided.

The first research chapter (Thesis chapter 3), "*Reliability of Schroth Curve Type Classification in Adolescents with Idiopathic Scoliosis",* reports the intra- and inter-rater reliability of ten randomly selected certified Schroth therapists in classifying adolescents with idiopathic scoliosis using the Schroth classification algorithm designed for this thesis.

The second research chapter (Thesis chapter 4) is titled *"The effect of Schroth exercises added to the standard of care on the Cobb angle in adolescent with idiopathic scoliosis – an assessor and statistician blinded randomized controlled trial"*. Trial registration: Schroth Exercise Trial for Scoliosis NCT01610908. The objectives were to determine the differences in group changes from baseline, to three months and six-months by comparing the effects of Schroth exercises added to the current standard of care in Canada to standard of care alone on curve severity measured by the Cobb angle.

The third research chapter (Thesis chapter 5) is titled *"The effect of Schroth exercises added to standard of care on quality of life and muscle endurance in adolescent with idiopathic scoliosis – assessor and statistician blinded randomized controlled trial"*. Trial registration: Schroth Exercise Trial for Scoliosis NCT01610908. The objectives were to determine the differences in group changes from baseline to 3 months and 6-months by comparing the effects of Schroth exercises added to the current standard of care in Canada to standard of care alone on back

muscle endurance and quality of life, measured using Scoliosis Research Society 22r and Spinal Appearance Questionnaires.

The fourth research chapter (Thesis chapter 6)**,** *"Clinical Significance of the Curve Severity and Quality of Life in Patients with Adolescent Idiopathic Scoliosis Following the Schroth Treatment",* focuses on assessing the magnitude of clinically important effects for the outcomes of the RCT, using anchor-based and distribution-based methods. The goal of this study is to promote knowledge transfer from rehabilitation science into practice, by reporting and interpreting the RCT results in a manner that is meaningful to end users – patients, clinicians and policy makers.

After the research chapters, a final chapter follows which includes a general discussion and conclusions. In this chapter the entire research work has been synthesized and its strengths and weakness have been discussed. In addition, implications for practice and how the future research may follow-up have been outlined.

CHAPTER 2

Literature review

2.1 Standard of care for scoliosis

Four interventions are commonly used for adolescent idiopathic scoliosis (AIS) including: observation, exercise, bracing, and surgery. In Central and Southern Europe, all four treatments are being prescribed for scoliosis,[41] while in North America, exercises are not part of the standard of care.[42]

Treatment by observation consists of monitoring patients for progression, but without applying a specific treatment. Hence, observation can be regarded as the natural history. Physiotherapy, generally, is not prescribed in North America, but in central Europe (France, Poland, Italy, Spain, Switzerland, Austria and Germany) it is. Scoliosis-specific exercise approaches include Schroth,[43] Integrated Scoliosis Rehabilitation (ISR),[18] DoboMed,[18] Side-shift,[18] Lyon,[18] Functional Individual Therapy of Scoliosis (FITS)[44] and the Scientific Exercise Approach to Scoliosis (SEAS).[45] Depending on the curve characteristics, different braces are prescribed to treat scoliosis among which most are rigid. In North American practice, the following brace types have historically been used: Milwaukee, Wilmington, Charleston, Providence, and thoraco-lumbar-sacral-orthosis (TLSO) including Boston and its derivatives. The SpineCor brace is the only dynamic brace[46] and is not routinely used to treat scoliosis. In North America, the most widely used brace types are TLSO including Boston, the Charleston and Providence braces. Surgical procedures are, nowadays, mostly performed using a posterior spinal fusion and instrumentation approach[47]. However, some surgeons still prefer the anterior approach because of better correction, decreased risk of crankshaft phenomenon (where anterior segments of the vertebrae continue to grow when the posterior segments are fused) and because the fusion levels could be saved in a skeletally immature adolescents. [48]

The goal of observation during adolescence is to monitor the progression of the curve. Physiotherapy aims to prevent or improve curve progression, improve the appearance and the quality of life. The goal of bracing is to prevent curve progression, while surgery aims to correct the curve and maintain the results.

2.2 Natural history

In order to fully appreciate the conclusions about the effectiveness of the different interventions for scoliosis, it is necessary to understand the effect of natural history of scoliosis.

Several early long-term studies on natural history of scoliosis presented a poor prognosis. [1,49] Conclusions were made that all types of scoliosis lead to increased mortality, higher percentage of unmarried women, and higher rates of disability, back pain, poor general health, and cardiopulmonary compromise.[1,49] However, none of the studies were prospective, samples were relatively small and included patients with all causes of scoliosis, the outcomes instruments used were not validated and were not adjusted for the location of the curvature.[50]

Later studies on the natural history of AIS disproved earlier findings, and showed the same prevalence of mortality between controls and patients with AIS, slightly more prevalent back pain, but not more disabling, similar function and perception of general health compared to controls, and that clinically important cardiopulmonary compromise was rare.[50] In addition, even though the curves continue to increase into adulthood, patients with untreated AIS can function well as young adults, become employed, get married, have children, and grow to become active older adults.[1]

On the other hand, untreated patients with AIS can develop significant deformity, and the esthetic appearance can be disturbing to some patients.[51,52] Continued curve progression in adults depends on many factors. Not all untreated patients will have aggressively progressive scoliosis. Many investigators agree that the curves with a thoracic apex have the highest prevalence of progression, ranging from 58% to 100%[1], followed by lumbar, thoracolumbar, and double major, in descending order[50]. In addition, more skeletally and sexually immature patients have greater probability of curve progression, and the likelihood of progression before and after maturity is higher if the curve at presentation is larger.[53,54]

The natural history of AIS is not fully understood. Even though, most authors agree that the incidence of progression depends on the skeletal maturity, curve magnitude at presentation and the location of the curve[1,13], they do not seem to agree on the incidence of spontaneous improvements of the curves during maturation. In most recent studies on natural history of scoliosis, the prevalence of spontaneous improvements is not reported. In Brooks et al's

prospective study on 3,492 adolescents, 474 were diagnosed with AIS. In this sample, there were 22% spontaneous curve improvements at a 1 year follow-up[55]. In Wong and Tan's study, the incidence of patients with AIS who improved naturally was 9.5%. Moreover, 35% of the patients who had left thoraco-lumbar or left lumbar curves experienced a decrease of more than 10°.[56] Rogala, and colleagues, conducted a prospective study of screening for scoliosis, that included almost 27,000 students aged 12-14 years old.[5] The patients who were screened positively for scoliosis (N=1122) were observed and followed-up for two years. The authors reported a 3% incidence of spontaneous improvement of scoliosis curves. They also observed that 20% of patients with curve of ≥20° did not progress. Therefore, some of the positive treatment effects for scoliosis may be attributed simply to the natural history.

As suggested by Weinstein et al.[1], the most problematic unanswered question about scoliosis is the absence of information about its etiopathogenesis. It is uncertain whether one, or multiple factors cause AIS, and we do not fully understand why some scolioses progress and others do not. Therefore, it is difficult to elucidate why some patients improve after being treated, while others do not. All the studies on long-term effects of different therapies for scoliosis may be based on very heterogeneous samples in terms of pathophysiology, because AIS is diagnosed by exclusion criteria, and many unknown variables surrounding the cause of scoliosis are still present. As a result, physiotherapy, bracing and surgical treatments for scoliosis address signs and symptoms of scoliosis, rather than the causes of scoliosis.

Should the cause of scoliosis be revealed, it might improve the prediction of progression for each patient, which would better inform the treatment practice. Until then, patients and parents deserve unbiased well-informed explanation about benefits and risks of a recommended treatment, based on current knowledge.

2.3 Effect of different interventions for scoliosis

Several studies reported on long-term effects of observation, bracing and surgery on the outcomes of scoliosis patients, while only a few on the effects of physiotherapy. The outcomes studied in patients under observation, bracing and surgery, observed at more than 20 years after the treatment, included: radiographic findings[57], magnetic resonance imaging findings[57], pulmonary function[58], general health-related quality of life[52,59-61], effects of pregnancy on

scoliosis progression[62], back pain[58,61,63-65], function[16,61,66,67], spinal mobility and muscle endurance[52,57-59,62-64,68]. Surgical rates after bracing, observation and physiotherapy were also investigated[25,69], as well as the long-term effects on personality development in scoliosis patients treated by braces and surgery[70]. The outcomes investigated in the studies on the effect of exercises for scoliosis included curve magnitudes,[20,71-74] muscle strength,[73,75,76] pulmonary function,[73] prevalence of surgery,[25] prevalence of bracing,[71] postural deficiencies,[73] and quality of life[20,74].

2.3.1 Exercises

Weinstein states that there is no definite evidence to support the effectiveness of physical therapy to reduce the risk of curve progression, improve the curve magnitude, or decrease the need for surgery[1]. Numerous short-term studies have been conducted on determining the effect of exercises on different outcomes for AIS [18,23,45,71,72,77,78], but very little is known on their long-term effects.

Several systematic reviews on the effects of exercises for scoliosis [18,79-82], report promising results, but highlight the need for stronger study designs. Researchers, whose studies were included in the reviews, published three of five of these reviews, which increased the risk of reviewer bias. Those reviews suggest that scoliosis-specific exercises slowed the progression (deterioration) of scoliosis and/or reduced curve severity measured by the Cobb angle. [18,80,83] Some studies also showed improved neuromotor control,[76] respiratory function,[73] back muscle strength,[73] and cosmetic appearance.[73,84] Among all the promising scoliosis-specific exercise approaches reviewed, Schroth exercises were the most studied. Lenssinck et al's earlier review concluded that the exercises may have positive effect on the scoliosis outcome, but more evidence was needed. The most recent review by Mordecai and Dabke was an independent review of 110 publications [19], and included nine prospective cohort studies, of which only three were controlled and only one used observer blinding. The authors indicated that selection criteria, recommendations, and contraindications to exercise were not clearly determined in any of these papers. Moreover, most exercise studies did not report on compliance, intention-to-treat analyses, or on recruitment strategies. The magnitude of changes in the Cobb angles was usually statistically significant, but often within the range of measurement error.

Overview of exercise studies other than Schroth

In the first published randomized controlled trial (RCT) on exercises for scoliosis, Wan et al[72], determined the effect of scoliosis-specific asymmetric exercises. The authors reported improvements with scoliosis-specific exercises (different than Schroth) added to surface electrical stimulation on Cobb angle in patients with scoliosis. All patients received electrical stimulation on the lateral surface of the body, traction, and postural training, while the experimental group also underwent specific scoliosis-specific asymmetric strengthening exercises once a day. Eighty Chinese patients (40/group), aged 15±4 years were treated over a 6 months period. The thoracic curve and lumbar curve were similar in both groups at baseline (23°±10° to 26°±13°). Both groups improved, but a larger effect was observed in the exercise group.

In a recent RCT, Monticone et al found that scoliosis-specific active self-correction and task-oriented exercises significantly improved the Cobb angles (by 5.3° at skeletal maturity) and the QOL measured by SRS-22r questionnaire (by 0.75 to 0.89/5), while the traditional spinal exercises were associated with stable outcomes in 55 skeletally immature patients with AIS and curves <25° per group at baseline. [20]

McIntire et al, examined the effect of quantified trunk rotational strength training on balancing strength asymmetry, increase strength overall, and stabilize curves in adolescent idiopathic scoliosis[75]. Fifteen patients (12 females and three males), with an average age of 13.9 years and an average main Cobb of 33° were enrolled. After four months of supervised strength training, involving an average of 32 training sessions, each lasting about 25 minutes, their strength had significantly increased by 28% to 50%, $P<0.005$ to $P<0.001$. Short-term results (eight months) of the study were positive, as they showed the significant increase in muscle strength and stabilization of the curves, particularly in the group of patients with 20° to 40°. However, the curves did not stay stable at the two years follow-up, and 64% progressed by ≥6°, similar to the progression expected in non-treated patients. On the other hand, most of the patients were at high risk of progression having Risser 0 and 1 (10 patients) and all were less than eight months postmenarchal.

Mooney et al.[76] used the same approach as McIntire. The authors studied 12 adolescent patients

with scoliosis who were 11 to 16 years old and had curvatures ranging from 20° to 60°. After the four months treatment using the quantified trunk rotational strength training the myoelectric asymmetries between trunk sides were corrected. Patients also experienced significant strength gains ranging from 12% to 40%. Only one patient, with the largest curve at baseline (60°) progressed and underwent surgery, while none of the remaining patients progressed, and 4 of 12 had decreases in their curvatures from 20° to 28°. Interestingly, none of the patients used braces during this study.

In a long-term follow up at least six years after the treatment or maturation, of 328 Japanese children with AIS treated with hitch and side-shift exercise in combination with part-time bracing, Maruyama et al reported a 6.1% incidence of surgery[85]. Those who underwent a surgical treatment had significantly larger curves at baseline, than the ones who did not progress. This is consistent with the prognosis of scoliosis in patients with larger curves at the onset of scoliosis.

Bialek published preliminary results of the effects of the scoliosis-specific physiotherapeutic approach called Functional Individual Therapy of Scoliosis (FITS) used alone and combined with braces on the radiological and clinical outcomes in AIS[44]. After a mean 2.08 years follow-up, in the group where only exercises were applied: (1) in single structural scoliosis, 50.0% of patients improved, 46.2% were stable and 3.8% progressed, while (2) in double scoliosis, 50.0% of patients improved, 30.8% were stable and 19.2% progressed. In the group receiving exercise and Cheneau light brace, results were: (1) in single scoliosis, 20.0% of patients improved, 80.0% were stable, no patient progressed, while (2) in double scoliosis, 28.1% of patients improved, 46.9% were stable and 25.0% progressed. The author observed the best results in curves 10°-25° at baseline, and suggested that this was a good indication to start therapy before more structural changes within the spine were established[44].

Overview of Schroth exercises studies

In the only prospective study on Schroth exercises, the back muscle strength, change in the Cobb angle and the pulmonary function were assessed.[73] Otman et al found that the muscle strength increased significantly after a yearlong treatment compared to the pretreatment values. They also found a significant decrease in the Cobb angle after the treatment (form an average angle at the baseline 26.1° to 17.8° at follow-up).

Weiss et al[86] compared the incidence of surgery in patients with scoliosis (N=343) presenting various diagnoses (AIS N=179, early-onset scoliosis, N=106, kypho-scolioses, N=38, congenital scoliosis, N=9 and other etiologies, N=11) treated at the Schroth clinic by a combination of Schroth exercises and Cheneau brace with the incidence of surgery reported by Goldberg et al.[87] He reported overall an 11.95% incidence of surgery in the Schroth clinic vs. 28.1% in the Goldberg study, which was statistically significant[24]. When only AIS patients were observed, the incidence of surgery was only 7.3% overall, being as low as 6.9% in patients 11 to 14 years old and 10 % in patients between 9 and 10 years old.

Weiss et al.[24] compared the incidence of curve progression in a group of girls with idiopathic scoliosis (N=115, age 9 to 15) treated using the scoliosis-specific Schroth intensive in-patient program to that observed in patients with scoliosis not receiving treatments (N=107, age 4 to 15) from another prospective study. The authors matched their sample with the controls based on age, and formed three groups, as follows: I: age <12; II: age 12–14; and IIa: age 12–14 with Cobb angle >30. After a mean follow-up of 33 months, the incidence of progression in the treated group was statistically significantly lower than in the control group. In girls younger than 12 years old, the incidence of progression >5° was 46.7% vs. 71.2% in the control group. In the group of girls aged 12 to 14, this difference was even bigger with 19.2% in the Schroth group vs. 55.8% in the control group.

Similarly, two retrospective studies using Schroth exercises found a lower incidence of surgery in conservatively treated patients (Schroth exercises, and a combination of Schroth exercises and braces) comparing to just observation.[25,86] In 157 patients, Rigo, et al reported a 12.1% incidence of surgery following Schroth outpatient therapy (n=43), and 14.1% following a combination of Cheneau brace and Schroth exercises (n=106)[25]. Eight cases were under observation (5.1%). Mean age was 12.6 years (SD=1.1, range 10–14), and 79 cases were pre-menarchal (50.3%), which identifies this sample as at higher likelihood of progression. Mean initial Cobb angle was 26.7° (SD=12.3, range 11°–65°). The incidence of surgery in Rigo was found lower than in untreated patients from Goldberg et al's study[88], where the incidence was reported to be as high as 28.1%[87].

A retrospective controlled study demonstrated significantly better effects for scoliosis-specific "3D corrective spinal technique" (experimental) than for conventional exercise (control) on

Cobb angle, vertebral rotation, and SRS-22 questionnaire scores.[32] The experimental treatment included a blend of Schroth and spinal stabilization exercises, but the treatment was not fully described. The control treatment included symmetrical stretching and spinal stabilization exercises. Both treatments were delivered with 60 minutes session/day, 2-3 times a week, for a total 30 sessions over the four-months period. Improvement in the Cobb angle was observed in both groups (8.1°±4.5° in the experimental and 4.3°±2.1 in the control group) and the between group differences was significant.

The literature on the effect of exercises for scoliosis has been widely criticized because of numerous methodological weaknesses. The best evidence on the effect of exercises comes from the Monticone et al's RCT. Most of other exercise studies were retrospective, did not report on compliance, intention-to-treat analysis, and did not blind the assessors. Of those few that were prospective, the majority was uncontrolled or not randomized. Some studies, such as Wan et al's RCT [72], did not specify which scoliosis diagnosis was studied, thus the results cannot be generalized to the patients with AIS. Many follow-up reports on physiotherapeutic approaches for scoliosis are combined with brace treatment, which makes it hard to conclude on the effect of exercises independent of the bracing effect. Conducting long-term follow-up studies of exercises for scoliosis can be problematic, because such studies usually require patients to continue the treatment (exercising) on their own, so that the achieved correction can be maintained. It becomes difficult to understand the effect of treatment after cessation of the supervised exercise program.

2.3.2 Bracing

Long-term follow-ups on bracing suggest that patients with scoliosis may have a higher prevalence of back pain compared to untreated patients with scoliosis,[65] and/or more severe back pain compared to controls,[17] and of respiratory compromise if the major thoracic curve becomes greater than 45° or with more rotation[1,58,61]. Smaller curves, however, show no significant differences in back pain outcome between braced and non-treated patients observed 22 years after the cessation of the treatment.[68]

A prospective study on long-term effects of bracing and observation 16 years after the treatment[68] suggested that the incidence of a curve progression of ≥ 6° in the observation group

(40.4%) was significantly higher than in the braced group (0%). This difference, although statistically significant, was within the range of measurement error. Moreover, after reaching maturity, this difference was non-significant, and the average increase of the curves was 5.7° (SD = 5.7, -3° to 21°) and 5.0° (SD = 4.4, -5° to 15°) in the braced and observed groups, respectively. Six patients (10%) in the observed group underwent spinal surgery during the maturation, versus none in the braced group. After maturity was reached, in both groups, none of the patients underwent surgery. These results, which are consistent with several other studies on braces[89-93], imply that the brace treatment might alter the natural history during the growth phase, but the long-term effects on outcomes after the maturity has been reached, curve progression and surgical rate, are not significantly different in patients treated using braces and observation.

The most important conclusion from this prospective study is that 70% (N=40) of the observed patients did not require any other treatment, while 20% were braced (N=11) and 10% underwent surgery (N=6). Those 70% of patients not requiring further treatment had average curves of 30.6° (SD=5.0, 21°–42°) at the end of maturation and 35.0° (SD=6.5, 21°–48°) after a mean of 16 years of follow-up. Considering that the long-term effects of both treatments are virtually the same, and that only 10% of non-treated patients will require surgery, some authors question if it is worthwhile subjecting such large number of patients to the burden of wearing a brace.

MacLean et al investigated the psychological effects of brace wear by comparing 31 patients treated with low profile braces for idiopathic scoliosis with a control group of healthy subjects. [94] They concluded, after a relatively short follow-up (median 10 months) that although bracing has some initial impact on self-esteem[94], "no overt residual differences occurred". On the other hand, the authors found that a significant period of stress and self-esteem change at the initiation of brace-wear occurred in the majority of patients (88%). They noted the problems patients face with brace wear such as soreness, discomfort with activity, torn clothing, limitations of participation in sport, physical activity and social events. Most patients in that study wore a part-time low-profile brace, which might have contributed to the fact that bracing did not have a significant psychological impact.

Many studies have been published on brace treatment for AIS, some support the effect of brace treatment in preventing curve progression, while others suggest that bracing is ineffective[95]. Most studies did not follow the SRS (Scoliosis Research Society) recommendation for bracing

and non-surgical treatment inclusion criteria: (1) ≥ 10 years of age at initiation of bracing, (2) initial curve of 25° to 40°, (3) Risser sign 0 to 2, (4) female premenarcheal or less than 1 year past menarche, (5) no previous treatment, and (6) at least 2 years of follow-up[96]. Not following the SRS criteria can bias the results. For example, if patients aged ≤ 10 were included, that may suggest the juvenile idiopathic scoliosis (JIS) was also included, and JIS has worse prognosis than AIS[17]. Moreover, if the sample consists of patients with lesser likelihood of progression (Risser >2, and postmenarchal more than 1 year[13,97]) the effect of the treatment may be overestimated. If the patients were previously treated using other treatment modalities, a carry-over effect could mask the effect of the current therapy.

Janicki, et al[98] conducted a study that followed the SRS inclusion recommendation. They compared the rate of surgery in 83 patients treated with TLSO, worn 22 hours/day (N=48) and Providence braces, worn 8-10 hours/night (N=35), at least two years after the treatment. Their results revealed that in the TLSO group, only seven patients (15%) did not progress, whereas 41 patients (85%) progressed by ≥6°, including the 30 patients whose curves ultimately exceeded 45°. Of those, 38 patients (79%) required surgery. In the Providence group, 11 patients (31%) did not progress, 24 patients (69%) progressed by ≥6°, including 15 patients whose curves ultimately exceeded 45°, while 21 patients (60%) required surgery. On the other hand, the patients whose curves were smaller at the beginning of the treatment had better prognosis. Five (15%) of 34 patients in the TLSO group and 10 (42%) of 24 patients in the Providence group did not progress, while 26 patients (76%) and 11 patients (46%), respectively, required surgery.

The most recent multicenter study conducted by Weinstein and colleagues[99] investigated the effect of bracing on the scoliosis curve progression to the threshold for surgery. Curve progression to ≥50° was defined as treatment failure, while skeletal maturity without curve progression to this threshold was considered treatment success. The study commenced as a RCT, but after 32 months of recruitment, the researchers introduced preference groups, because the recruitment was slower than anticipated due to a stated patients' preference for one of the treatments leading to their refusing randomization. Inclusion criteria followed the SRS recommendations. There were 116 participants in the randomized and 126 in the preference cohort. Since changing treatment groups was allowed, the as-treated group consisted of 96 observed and 146 braced participants.

The intention-to-treat analysis[100] revealed superior rate of treatment success for brace therapy (75%) compared to observation (42%) in the randomized group. Moreover, longer brace-wear was correlated with increased treatment success, which corroborates the results from other brace studies[101]. Wearing a brace for at least 12.9 hours per day was associated with a success rate of 90%-93%, which implies that current prescribed brace-wear dosage may be overestimated. Combining the randomized and preference cohorts, the success rate in the observation group was 48%, and 41% in patients who wore brace only 0-6 hours/day, suggesting that although brace is more effective than observation there are many unnecessary brace prescriptions, as has been recognized in previous studies[68].

The ≥50° threshold used to define failure in the BrAISt was not fully in agreement with the SRS recommendations to use a surgery threshold of ≥45°. Moreover, the SRS criterion for skeletal maturity in girls is a Risser 4 (75-100% ossification of the iliac apophysis) with more than two years after onset of menarche. In the BrAISt, the use of a cutoff of 50° and defining maturity as Risser 4 without the requirement of the 2 years post menarche in girls, may have increased the risk of false successful outcomes. Similarly, the follow-up in the observation group was 21 rather than the recommended 24 months. Further, the percentage of patients who were recommended or who underwent surgery was not provided. Nevertheless, Weinstein et al's BrAIST study is the newest and a major supportive finding on the effectiveness of bracing, which likely will have a broad positive impact on the rate of brace prescription.

2.3.3 Surgery

In a recent systematic review of the literature, Westrick et al[50], analyzed the long-term effect of surgery on correction and loss of Cobb angle, prevalence of pseudoarthrosis, rate of hardware failure and health related quality of life five to 20 years after the treatment. The results of the review suggested that anterior systems gave an overall average curve correction of 62.4%, Harrington rods gave a 34.1% average correction, and segmental hooks gave a 51.2% average Cobb angle correction. Loss of Cobb angle correction during follow-ups averaged 11% for anterior constructs, 17.5% for Harrington rods, 6.5% for segmental hooks, 3.4% for pedicle screw fixation, 3% for Isola hybrid, and 7% for Wisconsin fixation. The rate of hardware failure was 12.3% for anterior constructs, 15.8% for Harrington rods, 3.9% for segmental hooks, 0% for Isola hybrid, 8.6% in Wisconsin patients, and 7.1% in pedicle screw fixation. The reoperation

rate was 10.2% for anterior constructs, 11.9% for Harrington rods, 7.6% for segmental hooks, 7.1% for pedicle screws, 9.5% for Isola hybrid, and 5.7% for Wisconsin patients. It appears that the rate of hardware failure in newer surgical procedures (segmental hooks, Isola hybrid and pedicle screw) is minimal, and the loss of correction is acceptable in the long run. Six to 10% of patients underwent more than one surgery. These authors did not find any conclusive evidences supporting that the surgery improved quality of life, and suggest that "the indications for surgery remain largely for cosmetic and psychological reasons".[50]

Increase in back pain prevalence and decreased function at a long-term follow-up in surgically treated patients was confirmed in the controlled study conducted by Danielsson and Nachemson[58]. The authors examined the effect of surgery (N=144) at least 20 years after the treatment on back pain and function and compared the results to a matched healthy control group (N=100). They found that a loss of correction occurred in all patients and was on average 3.5°. Also, 4% of patients experienced a loss of correction of more than 11°. An additional curve-related surgical procedure was undergone by 5.1% of the patients. Patients had significantly more degenerative disc changes, and more frequent lumbar pain than the controls (65 vs. 47%, P=0.0079). In terms of back function, the authors reported statistically significant (P<0.0001) reduction in physical functioning, more bodily pain, and less general health in patients compared with the controls. Moreover, the sick leave due to the back pain was significantly more prevalent in patients than in controls (45% vs. 19%, *P*= 0.0040). Although, patients' curves remained stable for most part, in the long term, disability in terms of back pain and function was not negligible.

In a long-term follow-up study on effects of surgery and bracing in patients with AIS after 20 to 28 years of follow-up, Danielsson et al[62] compared consecutive series of females with AIS treated between 1968 and 1977, either using Harrington rods (N=145) or with a brace (N=122), and 90 randomly chosen healthy women without scoliosis or history of back surgery. The authors reported that the scoliosis did not progress during the pregnancy in surgically and braced patients, and it was not correlated with the number of pregnancy. In addition, the patients with AIS treated surgically and those treated with a Milwaukee or Boston brace did not differ from a normal population in terms of marriage or number of children born. However, braced patients were significantly older at first delivery than surgically treated patients and controls (28 vs. 26.6

vs. 25.9). The rate of complications during pregnancy, including back pain and number of cesarean sections, was not higher than in the general population. Sexual function was affected by back problems more in the AIS patients than in the controls, and this difference was statistically significant, but it is not clear whether this was because of the condition itself or because of the treatment administered. The results suggest that patients did not have more problems during pregnancy and delivery than a control group, but the prevalence of vacuum extraction was significantly more frequent in the surgically treated group than in the controls.

In a systematic review of the literature, Hawes[102] identified articles which dealt with the outcome of surgical treatment of adolescent idiopathic scoliosis, included >10 patients in the study, with a follow-up time of >2 years and including average Cobb angle correction and other outcome parameters. From 1979 to 2007, there were 82 articles reporting surgical outcome for 5780 patients. During those last 10 years, when the most advances have been made in surgery for AIS, most surgeons report a mean 40% to 50% correction and some report 60% or higher average Cobb angle corrections at follow-up at least two years after the treatment. Rates of reoperations varied from 0% to 45%, and some patients required three or more operations to achieve a stable result. Pulmonary function in patients decreased during the post-operative period, but in many patients, values returned to levels similar to those found preoperatively within two years. In other words, there was no evidence that surgical treatment for scoliosis improves pulmonary function in the long run. In a recent study, Gitelman[103] argues that the pulmonary function of surgically treated patients depends on the surgical approach used. Several studies with short-to-medium range follow-up (2–5 years) have shown that chest wall-violating procedures, such as anterior approaches and rib-resection thoracoplasty, have worse pulmonary function results than posterior approach procedures that spare the chest wall[103].

Spinal surgery is an irreversible procedure, and the normal active range of movement in the spinal column, including the non-fused segments is lost[58,62,102,104]. It has been shown that pain increases as flexibility decreases in non-surgical cases[104]. When compared with control subjects, the ability of surgical patients to side flex was reduced by 20% to 60%[104]. Surgery has a long-term effect on pain in patients with AIS and it has been shown that surgically treated patients experience more severe pain than controls without AIS.[58] In addition, operated patients have larger pain total body area score, than the controls.[58] Nowadays, surgery reliably halts the

scoliosis progression, permanently achieves significant curve correction and improves appearance.[50,58,102] Several long-term outcome surveys suggest that, for many patients, a reduced magnitude of lateral spinal curvature can be maintained for >20 years[62], although several spine surgeries may be required to achieve a stable improvement.[102] Surgery, while an effective method of improving a cosmetic deformity, can bring significant risks, and, "replaces one abnormality (a flexible, curved spine) with another (a rigid, straighter spine)".[102] Complications are common, and multiple revision surgeries may be needed in some cases to achieve a stable result.[50,102] Like pain and other symptoms, patient satisfaction with surgery is not correlated with curvature magnitude.[105] In a recent study, Westrick and colleagues[50] suggest that there is no medical necessity for surgery based on the current body of literature in terms of health-related quality of life (defined by enhanced function, self-image, or general health).

2.4 Classification of scoliosis

Scoliosis can be classified according to the age of onset into: infantile (0 – 2), juvenile (3 – 9), adolescent (10 – 18) and adult (>18).[42] This classification is most widely used by clinicians. With respect to the age at onset, some further differentiate between the early and late onset scoliosis, with early onset higher end range and late onset lower end range being at the age of 10.[106]

Anatomical classification is based on the location of the curves, and is generally used to describe the patient sample and plan a therapy. According to Ponseti, there are four major patterns: main cervico-thoracic, main thoracic, main thoraco-lumbar, main lumbar and combined thoracic and lumbar.[107] Moe and Kettleson proposed a scoliosis classification system also based on the location of the curves, which includes three single curve types: thoracic, thoracolumbar and lumbar; and four combined curve types: main thoracic/minor lumbar, double major thoracic/lumbar, double major thoracic/ thoracolumbar and thoracic double major.[108] Until King and colleagues[108,109] developed another classification in 1983, this system was routinely used by surgeons.[108] To our knowledge, there were no reports on its reliability.

While Moe and Kettleson's classification was purely descriptive, King's classification informed the selection of fusion level in thoracic idiopathic scoliosis.[110] King's classification system consists of five curve types and was mainly made in consideration for planning lumbar sparing

thoracic surgery.[108,111] Although widely used both for surgical and bracing decisions, King's classification was shown to be poorly reliable with a mean kappa 0.40 for interobserver reliability and 0.62 for the intraobserver reliability.[112,113]

A new classification system was proposed by Lenke et al. in 2001.[114] This classification distinguishes six types of scoliosis, each of which can be subdivided according to the extent of lumbar deviation, including 42 different subtypes.[108,111,114] This classification is more reliable with an overall classification inter-rater reliability reaching Kappa 0.62, with the highest agreement for the sagittal modifier (Kappa=0.91), and intra-rater reliability of 0.73[113] higher than that of the King's classification.[113] Since its presentation, it has been endorsed by the SRS and widely used by surgeons to guide surgical procedures.

Although the Lenke system is very comprehensive, it is based on two planes: sagittal and frontal. Qiu et al proposed the Peking Union Medical College (PUMC) classification system to address three-dimensional characteristics of scoliosis.[115] This system was designed to facilitate the selection of surgical approach and fusion level, and consists of three major categories, termed Types I, II, and III, having single, double, and triple curves, respectively. These categories include a total of 13 subtypes that are based primarily on apex location, curve flexibility, and curve magnitude.[115,116] The PUMC classification was shown to be more reliable than King's classification[115] and, as Lenke's classification, produced good to excellent reliability coefficients[117].

Most recently, Rigo et al. developed a scoliosis classification system to specifically guide brace treatment.[108] Rigo's classification is based on clinical and radiological assessments. Initial clinical diagnosis distinguishes between four basic types called: three curves, four curves, non three-non four and single lumbar or thoracolumbar. Additional radiological assessment further subdivides curves into particular sub-types A, B, C, D and E. He has shown an acceptable intra- and inter-observer reliability of the classification with intra-observer Kappa value of 0.87, and the inter-observer Kappa between 0.61 - 0.8[27]. Rigo's classification somewhat relates to the Schroth classification introduced hereafter, with respect to the clinical assessment, including the position of the pelvis relative to the plumb line, the presence/absence of the rib hump and the lumbar prominence. His classification also relies on the radiological findings, not routinely used for the Schroth classification.

Some of these classification systems were designed either to guide the brace prescription, such as Rigo's classification, or to guide the surgical spinal fusion procedures, such as King's, Lenke's and PUMC's. However, to our knowledge the Schroth classification system remains the only classification developed to guide the exercises treatment for scoliosis.

The Schroth classification consists of four specific curve type categories.[6,21,22] Patients with scoliosis are classified according to their clinical presentation to guide exercise prescription decisions. A clinician classifies a person with scoliosis based on the alignment of four body blocks: (a) the pelvis and lower extremities; (b) the lumbar spine; (c) the thoracic spine and rib cage, and (d) the shoulder girdle, cervical spine, and head. In a healthy person, body blocks are superposed, because the body does not deviate in the frontal plane from the plumb line[26,43] The Schroth method recognizes two thoracic categories(3c and 3cp), in which the thoracic deformity is major, and two lumbar/thoraco-lumbar scoliosis patterns (4c and 4cp), in which the lumbar or thoraco-lumbar deformity is dominant.

Schroth exercise treatment is guided by the specific classification, but the reliability of the therapists' classifying has not been reported previously. Therefore, we proposed the usage of our Schroth classification algorithm to help therapists reach clinical decisions about the curve types. Chapter 3 of this thesis reports on the reliability of the Schroth therapists in classifying patients with AIS into four different Schroth classifications using our standardized algorithm.

2.5 Clinical significance

Many authors agree that reporting statistical test results should be supplemented with methods for determining clinically significant change.[27-29,118] But, how can a clinically significant effect be measured?

Lipsey, W.M. and Wilson, B.D. critically assessed 302 meta–analyses and found "strong skewness towards positive effects" in published studies. Of 302 meta-analyses, only six produced negative mean effect sizes, and relatively few mean effect sizes were in the immediate vicinity of zero. More than 90% of the mean effect sizes were 0.10 or larger, and 85% were 0.20 or larger. The authors concluded that almost every treatment examined had positive effect.[119]

Anchor-based and distribution-based methods are used to measure meaningful change over time and evaluate whether the treatment has had a clinically significant effect on an outcome (patient's overall health, specific domain of health, quality of life, physical achievement, or some other outcome). Anchor-based methods use an external indicator (anchor) of clinical change to interpret the observed change in a target outcome. On the other hand, distribution-based methods are based on the statistical properties of the scale used to measure some outcome of interest and may include effect size measures, [36,120-122] standard error of measurement (SEM)[123], and the reliable change index.[27,124] These methods rely on statistical properties to quantify how much change in a measure is deemed clinically important.

To determine if a clinically significant change has occurred in the patient's response, researchers most commonly ascertain the minimal clinically important difference (MCID), and use it as a threshold for meaningful change. The MCID is "the smallest difference in score in the domain of interest that patients perceive as important, either beneficial or harmful, and which would lead the clinician to consider a change in the patient's management".[34]

Bago et al[39] and Carreon et al[40] made first efforts in determining the MCID in scoliosis research. Bago et al used both anchor- and distribution-based methods to determine the MCID on the SRS-22r questionnaire.[39] The authors recommended using the distribution-based MCID when the aim is to analyze each of the scales of the instrument, because some subscales determined by the anchor-based methods showed an MCID below the measurement error. On the other hand, if the interest is the patients' perception, the authors suggest using only the anchor-based MCID of the sum score, which corresponds to 13.1, while the average-sum=0.6. The MCID based on distribution methods using 95% confidence interval of the minimally detectable change in the score for each of the domains were as follows: Pain=0.6, Function=0.8, Image=0.5, Mental Health=0.4, Raw-Sum=6.8, Average-Sum=0.5.

Carreon and colleagues[40] used only anchor-based method to determine MCID for the Pain, Appearance and Function domains in the SRS-30 questionnaire. They used the last eight questions on the SRS-30 questionnaire to construct the anchors for Pain and Function by summing up the relevant questions pertaining to each domain. For the Appearance domain, together with the relevant questions on the SRS-30, they also used two questions from the Spinal Appearance Questionnaire (SAQ). The anchors were used to construct Receiver Operator

Characteristic (ROC) curves for each domain separately, and determined the following MCIDs: Pain=0.20 (area under the curve [AUC] 0.723, CI=0.70-0.77), Activity=0.08 (AUC 0.648, CI=0.60-0.69), and Appearance=0.98 (AUC 0.629, CI=0.60-0.68).

Most recently, Rushton and Grevitt conducted two reviews of the literature, of which one investigated the clinical significance of the difference in SRS questionnaires scores between untreated adolescents with idiopathic scoliosis compared to adolescents without this condition[125]; and the other to calculate the MCID from published research on surgically treated adolescents with idiopathic scoliosis, and compare it to Bago et al's[39] published MCID on Pain, Image, Function and Mental health domains to determine whether an observed change was clinically significant. The authors found that the Pain and Self-image domains' scores were statistically lower among cohorts with AIS compared to those unaffected, but the observed difference was clinically significant only for the Self-image domain. Similarly, they found that 81% and 94% of the surgically treated cohorts included experienced statistically significant improvements in Pain and Self- image domains, but that only Self- image improved by clinically significant amounts using reported MCIDs.

The work of these authors attempted to determine the clinically important change in the domains measured on the SRS questionnaires in patients who have undergone the scoliosis surgery. Bago et al were the only authors to report MCIDs for all SRS-22 domains, including raw sum score and average sum score, while Carreon et al reported the MCIDs for only Pain, Function and Self-image. To our knowledge no research has been done to assess the clinically important effect of the conservative treatment for scoliosis.

2.6 Concluding remarks

This literature review presented some shortcomings of the current standard of care for AIS, a common pediatric disease associated with chronic consequences on curve progression, and quality of life, including pain, self-image, and psychological issues. Moreover, the review highlighted the limited treatment options in North American routine care, which do not include exercises, although promising results of effect of exercises obtained in studies of suboptimal quality exist. To address that exercise treatment is not more widely adopted for scoliosis due to the lack of strong evidence, a rigorous randomized controlled trial was conducted to determine

the effect of the Schroth exercises combined with standard of care on curve severity, quality of life, self-image and back muscle endurance compared to the standard of care alone. The results of this RCT are presented in Chapters 4 and 5.

This literature review also pointed out that several classifications for scoliosis have been used to guide treatment decisions. Of several scoliosis-specific exercise approaches, only Schroth uses a classification to guide the exercise prescription. However, its reliability is unknown. Therefore, as part of the planning of this RCT, a standardized classification algorithm and the instructions for its use have been developed. The reliability of the Schroth therapists in classifying patients into different curve types using this algorithm has been tested, and presented in Chapter 3.

In modern clinical research reporting clinical significance of the outcomes in support of the statistical analysis is becoming standard. To better inform clinical practice, clinical significance of the outcomes produced in the RCT (Chapters 4 and 5) has been investigated and results are presented in Chapter 6 of this thesis.

CHAPTER 3

Reliability of the Schroth curve type classification in adolescents with idiopathic scoliosis[1]

3.1 Summary

Background: A system of four scoliosis classifications is used to guide Schroth therapists in the specific exercise prescription for patients with adolescent idiopathic scoliosis (AIS). We developed a rule-based algorithm to assist therapists in reliably classifying patients.

Objective: To determine the intra- and inter-therapist reliability in classifying patients with AIS using our proposed classification algorithm.

Design: An international intra- and inter-rater reliability study.

Methods: We recruited 44 participants with AIS, aged 10 to 18 years, with curves between 10°-50° and 10 consecutive English-speaking volunteer Schroth therapists from the registry of international certified therapists. Patients' standing posture and their appearance on the Adam's forward bending test were videotaped. Therapists, blinded to participants' identity, rated video assessments presented randomly on two occasions at least seven days apart. Gwet's AC1 and weighted AC1 coefficients[126] were calculated to determine the intra- and inter-therapist reliability.

Results: The overall intra-rater AC1 was 0.64 (95%CI 0.53-0.73), 0.70 (95%CI 0.60-0.78) among well-trained raters, and 0.81 (95%CI 0.77-0.85) in experienced raters. The weighted intra-rater AC1 averaged 0.75 (95%CI 0.63-0.84) overall, 0.82 (95%CI 0.73-0.88) in well-trained raters, and 0.89 (95%CI 0.80-0.94) in experienced raters. Inter-rater AC1 was 0.43 (95%CI 0.28-0.58) overall, 0.50 (95%CI 0.38-0.61) for well-trained raters, and 0.67 (95%CI 0.50-0.85) for experienced raters. The weighted inter-rater AC1 was 0.48 (95%CI 0.29-0.67) overall, 0.61 (95%CI 0.49-0.72) among well-trained, and 0.79 (95%CI 0.64-0.94) among experienced raters.

[1] Sanja Schreiber, Eric C. Parent, Greg N. Kawchuk, Douglas M. Hedden

Limitations: Reliability might have been limited by insufficient algorithm comprehension of some therapists and inability of therapists to assess the patients in person.

Conclusions: A high level of understanding of the algorithm improved the intra- and inter-rater reliability. Weighted analysis demonstrated raters' adequate reliability, justifying usage of the proposed algorithm.

3.2 Introduction

Adolescent idiopathic scoliosis (AIS) is a 3D deformity of the spine affecting 2-3% of teenagers, mostly females.[1] Scoliosis can lead to chronic consequences including external torso deformity, pain, limited function and poor self-image.[127]

The goal of scoliosis treatments in adolescents is to prevent curve progression during the period preceding skeletal maturity where risk of progression is high. Stopping curves from progressing before maturity is important because 1) the health consequences of scoliosis manifest most typically once curves exceed 50° and 2) it helps prevent/limit progression during adulthood. In North America, the standard of care for scoliosis includes observation, bracing and surgery,[10] while in Europe, scoliosis-specific exercises are also prescribed alone, as an add-on to braces, or as a post-surgical treatment .[128]

Schroth exercises are scoliosis-specific exercises, developed in the 1920's by Katharina Schroth in Germany, which aim to correct patient's posture and curvatures by improving the endurance and control of the muscles affected by scoliosis. Although, in low quality studies, Schroth exercises have demonstrated promise in slowing progression or reversing curve magnitude.[18] They remain the most studied exercises for scoliosis.

Several classifications for scoliosis exist to guide surgical and brace treatment. However, to our knowledge, only the Schroth classification has been proposed to guide scoliosis treatment using exercises.

The Schroth classification seems to be the only one developed exclusively to guide the exercise treatment for scoliosis. The Schroth classification consists of four curve types[21]. During a clinical assessment, the influence of scoliosis on the alignment and posture in each of four body blocks is appraised to determine the curve pattern. The four blocks are (a) the pelvis and lower

extremities; (b) the lumbar spine; (c) the thoracic spine and rib cage, and (d) the shoulder girdle, cervical spine, and head (Figure 3.1). In a healthy person, body blocks can be represented by symmetrical superimposed rectangles, since the body is not deviating laterally from the plumb line and blocks are not rotated relative to each other[26,43]. According to Schroth there are two thoracic (3c and 3cp) and two thoracolumbar/ lumbar (4c and 4cp) curve patterns described in *Appendix 1*[26,43] (Figure 3.1.). The reliability of the Schroth classification system is unknown.

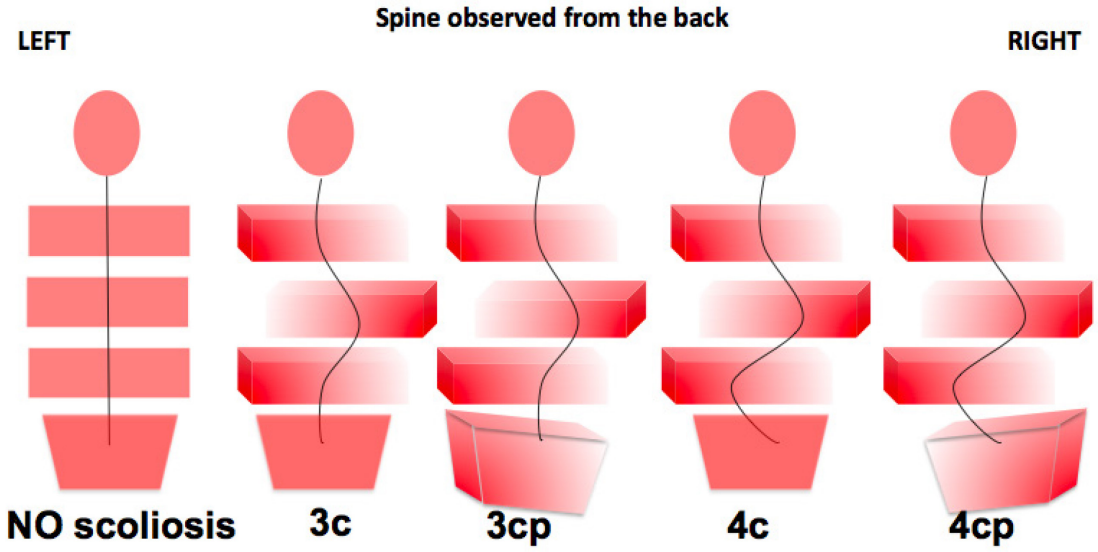


Figure 3.1. From left to right, the first illustration shows a body without scoliosis; the3c type is represented with a major thoracic curve, a small or non-existent lumbar curve and a balanced pelvis; the 3cp type key features are a thoracic curve with the lumbar and hip blocks coupled and deviated to the opposite side of the thoracic curve; 4c is the most balanced curve type with thoracic and lumbar curves similar in magnitude and a balanced pelvis; the 4cp type presents with a major thoraco-lumbar or lumbar curve with or without presence of a thoracic curve, and with the pelvic block shifted to the opposite side of the major curve.

Recently, Rigo developed a classification to specifically guide brace treatment using the Cheneau rigid brace and its derivatives.[108] Rigo's classification is based on the Schroth classification, but also uses radiological criteria. The initial clinical assessment distinguishes between five curve types based on the location of external deformity and pelvis position. Radiological assessment allows to further sub-classify the curves based on the lumbar and the thoracic configurations to arrive at eight curve patterns. To acknowledge a presence of an upper thoracic structural curve, a modifier label is added to the curve pattern. The Rigo classification

has acceptable intra-rater reliability with a Kappa value of 0.87, and an inter-rater Kappa between 0.61 - 0.81[27].

Similarly, Weiss recently developed a classification to guide the Cheneau Light brace construction, which relies heavily on the original Schroth classification[129]. Weiss' differentiates further between curve patterns when the pelvis is imbalanced to create two additional categories. To our knowledge, there are no publications on this classification's reliability or where this classification was used clinically.

Several other classifications to guide treatment for scoliosis exist including King's,[109] Lenke's[114] and Peking Union Medical College (PUMC).[115] King's classification was developed to guide surgery, but has also been used to guide brace design. Lenke's and PUMC classifications are used only for planning surgeries. The PUMC classification was found to be more reliable than King's (Kappa=0.64)[115] and Lenke's classifications (Kappa=0.73)[113], with excellent intra- and inter-rater reliability (Kappa=0.90).[117] However, the Lenke's classification has been endorsed by the Scoliosis Research Society (SRS) and is used most widely.

Schroth exercise prescription is based on the classification. Therefore, for an ongoing randomized controlled trial, our team proposed a rule-based algorithm to assist therapists in classifying patients. Algorithms have been successfully used to maximize reliability of Cobb angle measurements and classifying patients with scoliosis.[130] Optimizing reliability is important to improve the treatment standardization.

The purpose of this study was to determine the intra- and inter-rater reliability of 10 certified Schroth therapists in classifying adolescents with idiopathic scoliosis using a proposed rule-based classification algorithm. We hypothesized that the algorithm would help therapists classify patients reliably.

3.3 Methods

Research Design

This was an intra- and inter-rater reliability study with 10 international Schroth therapists who rated video assessments of 44 volunteers with AIS on two occasions at least a week apart. All participants provided an informed consent.

Patients

Volunteers meeting the selection criteria were recruited consecutively from a specialized multidisciplinary clinic and provided informed consent. The inclusion criteria were: (1) diagnosis of AIS, (2) all curve types, (3) 10-18 years of age, and (4) curves between 10° and 50°. Exclusion criteria were: (1) patients with scoliosis other than AIS, and (2) history of scoliosis surgery.

Raters

In addition to three Schroth therapists participating in the Schroth Exercise Trial for Scoliosis (ClinicalTirals.gov NCT01610908) and the Schroth instructor from the Asklepios-Katharina Schroth-Klinik in Germany, six additional consenting therapists were recruited. All therapists in the registry of those previously trained in English at the Schroth clinic in Germany were invited to participate. The therapists' inclusion criteria were: (1) Schroth certified at the clinic in Germany, and (2) fluent in English. All consenting therapists were enrolled, but only the first six consecutive therapist volunteers who completed the ratings were included.

Sample size

Sample size was estimated based on the Gwet's recommendations.[131] Tolerating no more than 15% error margin on the coefficient of variation of the percent agreement (standard error/ percent agreement) and 20% relative error on the percent agreement (width of 95% CI around the percent agreement), 44 patients and 10 raters were needed.[131]

Physical Examination

Patients were videotaped while standing 2.5 meters from the camera in their habitual posture presenting each side of the body for 10 sec, as well as while performing Adam's forward bend test (about 20 sec). This test consists of bending forward with the feet together and knees straight while dangling the arms with the palms pressed together. The maximal thoracic and lumbar rotations observed during the test were measured using a scoliometer by SS. Patients stood in a 71 cm by 122 cm frame, embedded with white non-reflective linen sheets to ensure consistent lighting conditions. A posture assessment grid was positioned between the camera and the patient. The participants' identity was hidden using iMovie for Mac (v8.0.6 Apple, Cupertino,

CA) and the scoliometer measurements were included in the video. Edited videos lasted about 60 seconds.

Algorithm

The Schroth classification algorithm and operational definitions were designed based on information from the 2011 Schroth training manual by consensus of two research therapists.[21] The international Schroth instructor reviewed the material before it was piloted. The algorithm guides therapists in detecting the key postural features of the body blocks relevant to determine the Schroth curve type (Figure 3.2). Each step leads to a "yes/no" decision, while the final step leads unambiguously to one of four Schroth scoliosis classifications. To assist therapists make decisions and to control for their subjective interpretations, operational definitions and instructions were provided (*Appendix 1)*.

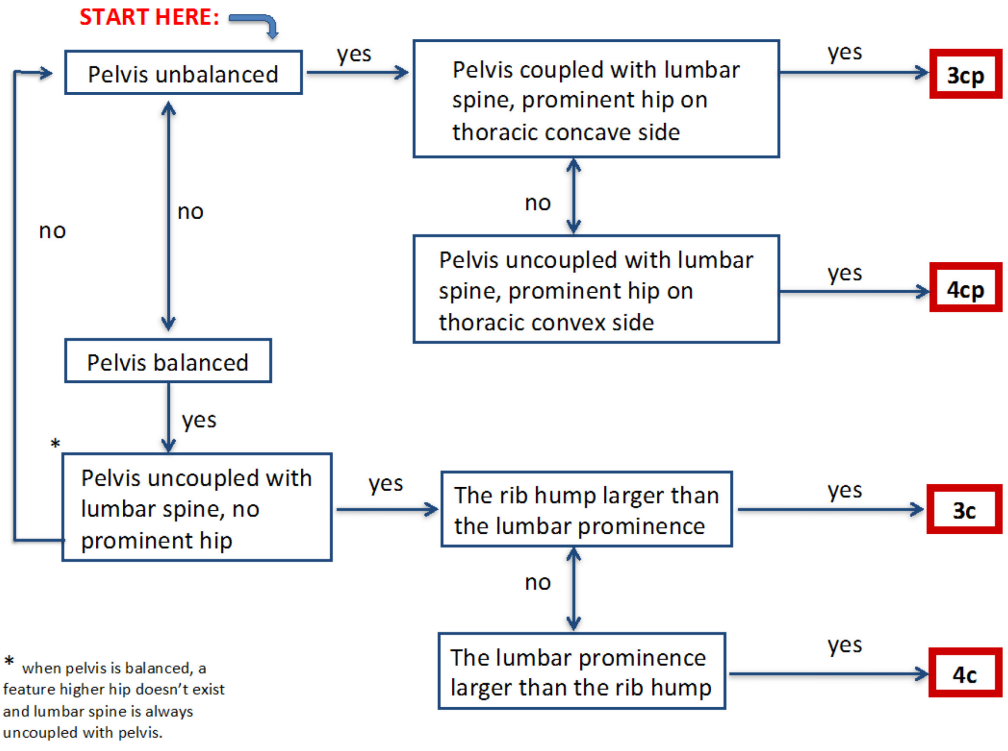


Figure 3.2. Schroth classification algorithm

Therapists begin using the algorithm (Figure 3.2) by determining if the pelvis is balanced or not (a displaced pelvis is considered unbalanced). If the pelvis is unbalanced, the relationship between the pelvis and the lumbar spine is then assessed. If the lumbar spine and the pelvis block are judged to be coupled (deviated in the same direction), the classification is 3cp. If the lumbar and pelvis blocks are uncoupled (deviated in opposite directions) the classification is 4cp. In the first step, if the pelvis is deemed balanced, the therapist determines if the pelvis is uncoupled from the lumbar spine, and if there is a prominent hip observed. The next step is to assess the relative importance of the rib and lumbar prominences. If the thoracic prominence is judged more important, the classification is 3c. If the lumbar prominence is found dominant the classification is 4c.

Rater training

Each therapist was provided with secured access to the study website, where the algorithm, the user's guide, and the operational definitions *(Appendix 1)* were available for download. Therapists were presented with four videos streamed from the website to practice using the rating procedures and the algorithm. After viewing each video, therapists provided their decisions for each algorithm branch, their classification decision, and comments. The primary author reviewed comments and prepared explanatory videos for all four training assessments to highlight the algorithm decisions in reference to the relevant section of the instructions and operational definitions. To address comments on video quality we consulted a professional photographer to improve video capture and lighting.

Rating procedures

After training, the 44 patient videos were made available for the first rating. The therapists could view the videos as many times as needed before submitting their answers. After the first rating, the access was blocked for a week. Then the same 44 videos were streamed in a different random order. Therapists were blinded to their first ratings, to the subject identification, and to other therapists' ratings.

Analyses

Cohen's Kappa[132] is widely used for nominal measurements, but has two limitations, known as Kappa paradoxes[126,133,134]. Kappa is influenced by trait prevalence and by rater's classification probabilities. Kappa produces lower values when overall agreement among raters is high. Also, given a certain overall agreement, when raters do not agree similarly on the number of patients they classify into each of the groups (asymmetrical marginal totals) Kappa is larger than when they agree more consistently (symmetrical marginal totals).

To guard against these paradoxes, Gwet's AC1 statistic has been proposed[126]. In estimating chance agreement, AC1 considers the number of observations, raters, and categories, whereas Kappa only accounts for the number of observations[126]. The AC1 and weighted AC1 reliability coefficients were calculated using commercial software (AgreeStat 2013.1 for Excel Windows Advanced Analytics, Maryland, USA)

Intra- and inter-rater percentage of agreement, AC1 and weighted AC1 reliability coefficients with their 95% confidence intervals[126] were calculated. Reliability was also estimated separately for two therapists who conceptualized the algorithm and who used it with 50 patients enrolled in the SETS trial (labeled "experienced therapists") (ClinicalTirals.gov NCT01610908). We also estimated reliability for the six therapists who self-reported full understanding of the algorithm (labeled "well-trained"). Based on the Landis and Koch's benchmarking system,[135] most widely used in clinical studies, 0.61-0.80 reliability estimates are considered substantial. Therefore, acceptable reliability was set a priori to $\geq$0.61, judged sufficient to recommend clinical use. Mean reliability coefficients and their standard deviations were calculated using Fisher's transformation.

The weighted analysis was justified because differences in Schroth therapy plans have a greater significance between some classification categories. For the "three-curve" patterns (3c, 3cp) the recommended treatment emphasis is on the thoracic curve, and for the "four-curve" patterns (4c, 4cp), the emphasis is on the thoracolumbar/lumbar curve. When the pelvis is displaced in the opposite direction relative to the major curve, the classification includes the "p" (3cp and 4cp). Treatment for 3cp curves emphasizes shifting the pelvis towards the thoracic convexity. In contrast, treatment for 4cp curves emphasizes shifting the pelvis towards the lumbar convexity (opposite of 3cp curves). Curve patterns 3c and 4c are more similar because the pelvis remains relatively unaffected and pelvis corrections are not emphasized. However, the 3c classification is

characterized mainly by a thoracic curve while, 4c by a double curve[21]. Because 3c and 4c curve patterns share thoracic curves and balanced pelvis, the exercise prescription will not differ as drastically as between 3cp vs. 4cp (opposite pelvis corrections), 4cp vs. 3c (emphasis mainly on pelvis and lumbar vs. thoracic curves) or 3cp vs. 4c curve patterns (emphasis on pelvis and thoracic vs. thoracic and lumbar curves only).

Since some classifications exhibit some similarities in their recommended exercise prescription, different weights can be assigned to disagreements between specific pairs of curve types to reflect partial agreement. Full agreement weighting (1.00) was assigned to the same group ratings. The weights for each pairing are summarized in Table 3.1.

Table 3.1. Custom weights

	3c	3cp	4c	4cp
3c	1.00	0.75	0.50	0.00
3cp	0.75	1.00	0.00	0.00
4c	0.50	0.00	1.00	0.75
4cp	0.00	0.00	0.75	1.00

A benchmarking model that does not account for random agreement arising from the number of subjects, rating categories and raters could be overly optimistic.[126] Therefore, Gwet proposed a benchmarking method, which considers only the part of the estimated agreement coefficient that confidently reflects genuine agreement among raters. That part, termed the "significant part" is obtained by subtracting Gwet's calculated "critical value" based on the number of subjects, raters and rating categories from the estimated AC1[126]. For our study of 10 raters classifying 44 participants into four categories, the critical value is 0.08[126]. In other words, there is 5% chance that a pure random rating would yield an AC1 of 0.08. The "significant part" of AC1 coefficients were then qualified as fair, moderate, substantial and almost perfect reliability.[135]

While referring to the instruction in the operational definitions *(Appendix 1),* the therapists begins the algorithm (Fig. 3.2) by determining if the pelvis is balanced on not. If the pelvis is deemed unbalanced, the relationship between the pelvis and the lumbar spine is being assessed. If the lumbar spine and the pelvis block are coupled, the classification is 3cp. If the lumbar and pelvis blocks are uncoupled (i.e. deviate in opposite directions) the classification is 4cp. If in the first step the pelvis was deemed balanced, the therapist determines if the pelvis is uncoupled

from the lumbar spine, and if there is a prominent hip observed. The next step is to assess the relative importance of the rib and lumbar prominences. If the thoracic prominence is judged more important the classification is 3c. If the lumbar prominence is found dominant the classification is 4c.

Rater training

Each therapist was provided with a user name and password to access the secured study website. First, therapists downloaded the instruction documents containing the study methodology, the explanations of the Schroth classification algorithm, as well as the operational definitions *(Appendix 1)*. Therapists were then presented with four videos streamed from the secure web site to provide training in using the study rating procedures and the algorithm. After viewing each video the therapists provided their classification decision and specified their decisions for each algorithm branch as well as comments on whether clarifications were needed. The primary author reviewed comments and prepared a review video explaining each of the algorithm decisions by referring to the relevant section of the instructions and operational manual for all four training cases. To address comments on video quality we consulted a professional photographer to improve video capture set-up and lighting.

Rating procedures

After training, the 44 patient videos were made available for the first round of ratings by providing a password to access the streaming videos from a secure server. The therapists could view the videos as many times as needed before submitting their answers. After submitting the first ratings, access was blocked for a week. Then, the same 44 videos were made available in a different random order for a second round of ratings. Therapists did not have access to their first ratings or to the subject identification during the second ratings to ensure blinding. Therapists were also blinded to other therapists' ratings.

The Landis and Koch's benchmarking system is most widely used in clinical research. AC1 coefficients will be compared against the Landis and Koch's benchmarks, where 0.0-0.20 slight, 0.21-0.40 fair, 0.41-0.60 moderate, 0.61-0.80 substantial and 0.81-1.0 almost perfect reliability. Our acceptable reliability coefficient value was set a priori to >0.61, which is considered sufficient to recommend clinical use. However, a benchmarking model that does not account for

the number of subjects, rating categories and raters in the study could assign overly optimistic benchmarks to an agreement coefficient, which could still carry a large error margin.[126] Because, the calculated reliability coefficient consists of parts considered to be obtained by random and non-random agreements, Gwet proposed a benchmarking method that takes into account only the part of the estimated agreement coefficient that can be trusted as measuring a genuine agreement among raters. That part, termed the "significant part" is obtained by subtracting Gwet's calculated "critical value" based on the number of raters and rating categories from the estimated AC1[126]. For our study of 10 raters classifying 44 participants into four categories, the critical value is 0.08[126].

3.4 Results

Patients

Participant's age was 14.2 (SD=2.0) years. On the radiographs, 17 patients had major thoracic with thoracolumbar/lumbar curves, two had a single major thoracic curve, 17 had major thoracolumbar/lumbar with a thoracic curve, seven had single thoracolumbar/lumbar curves and one patient had a single upper thoracic curve. Of those, eight were left proximal thoracic curves with a mean 25.0° (SD=7.1°), 36 were right thoracic with a mean 26.7° (SD=10.3°), and 37 were left thoracolumbar/lumbar curves with a mean 25.8° (SD=10.0°) (Table 3.2.).

Table 3.2. Description of the participants with AIS

	N	Mean	SD	Min	Max
Age (years)	44	14.2	2.0	10.0	20.0
Upper thoracic (°)	8	25.0	7.1	15.0	38.0
Major thoracic (°)	36	26.7	10.3	10.0	49.0
Thoracolumbar/lumbar (°)	37	25.8	10.0	10.0	48.0

Therapists

Sixty-five therapists were invited to participate, 16 consented and the first six to complete the ratings in addition to the three research therapists and the Schroth instructor were included. Therapists' mean age was 46.7 (SD=12.3) years ranging from 31 to 72. There were six females and four males, with a mean of 6.6 (SD=6.4) years working as Schroth therapists (range 3 to 23

years). Five therapists had treated >75 patients in their career, two between 26 and 75, two between 11 and 25, while one had treated 4 to 10 patients. The therapists' self-reported understanding of the classification algorithm ranged from 50% to full understanding, with a median of 100%. All therapists found the algorithm useful and 7/10 stated they would continue using it clinically.

Intra-rater AC1 and weighted AC1 reliability estimates

Considering all therapists, the mean intra-rater AC1 was 0.64 (95%CI 0.53-0.73) (range 0.34-0.83) and the mean weighted intra-rater AC1 was 0.75 (95%CI 0.63-0.84) (range 0.41-0.92)(Table 3.). Experienced raters reached higher reliability estimates with a mean AC1 of 0.81 (95%CI 0.77-0.85) (range 0.79-0.83), and weighted AC1 of 0.89 (95%CI 0.80-0.94) (range 0.85-0.92) (Table 3). The six well-trained therapists reached a mean intra-rater AC1 of 0.70 (95%CI 0.60-0.78) (range 0.47-0.83), and a weighted intra-rater AC1 of 0.82 (0.73-0.88) (range 0.60-0.92) (Table 3.3).

Table 3.3. Intra-rater AC1 and weighted AC1 reliability estimates and percent agreements with associated 95% confidence intervals

	AC1 (95% CI)	Percent Agreement (95% CI)	Weighted AC1 (95% CI)	Weighted Percent Agreement (95% CI)
Exp1	0.79 (0.64-0.94)	84.1 (73.0-95.2)	0.92 (0.87-0.98)	96.0 (93.2-98.9)
Exp2	0.83 (0.69-0.96)	86.3 (75.9-96.8)	0.85 (0.71-0.99)	91.5 (83.7-99.2)
R1	0.47 (0.27-0.67)	59.1 (44.1-74.0)	0.60 (0.41-0.79)	78.4 (68.5-88.3)
R2	0.52 (0.32-0.71)	63.6 (49.0-78.3)	0.58 (0.36-0.79)	78.4 (67.5-89.4)
R3	0.73 (0.57-0.89)	79.5 (67.3-91.8)	0.89 (0.82-0.96)	94.3 (90.7-97.9)
R4	0.51 (0.30-0.71)	61.5 (45.8-77.2)	0.57 (0.35-0.80)	76.3 (64.2-88.4)
R5	0.61 (0.43-0.80)	70.5 (56.6-84.3)	0.68 (0.50-0.87)	83.0 (73.1-92.8)
R6	0.34 (0.14-0.54)	48.8 (33.5-64.2)	0.41 (0.18-0.64)	67.4 (54.8-80.1)
R7	0.72 (0.55-0.88)	77.2 (64.5-90.0)	0.80 (0.66-0.95)	88.0 (79.9-96.2)
R8	0.64 (0.46-0.82)	72.7 (59.2-86.3)	0.83 (0.72-0.94)	90.9 (85.3-96.5)
Mean (overall)	*0.64 (0.53-0.73*	*72.0 (63.9-78.6)*	*0.75 (0.63-0.84)*	*86.7 (79.9-91.3)*
Mean (experienced)	*0.81 (0.77-0.85)*	*85.0 (82.9-86.9)*	*0.89 (0.80-0.94*	*94.0 (86.8-97.3)*
Mean (well-trained)	*0.70 (0.60-0.78)*	*76.6 (69.4-82.3)*	*0.82 (0.73-0.88)*	*90.1 (84.8-93.6)*

Inter-rater AC1 and weighted AC1 reliability estimates

The inter-rater AC1 for the entire sample was 0.43 (95% CI 0.28-0.58). The corresponding value for the experienced raters was 0.67 (95% CI 0.50-0.85)(Table 4). The well-trained therapists' inter-rater AC1 was 0.50 (95% CI 0.38-0.61). Overall, the weighted AC1 was 0.48 (95% CI 0.29-0.67). For experienced raters, the weighted AC1 was 0.79 (CI 0.64-0.94), and for the well-trained therapists it equaled 0.61 (95% CI 0.48-0.71)(Table 3.4).

Table 3.4. Inter-rater AC1 and weighted AC1 reliability estimates and percent agreements with associated 95% confidence intervals

	AC1 (95% CI)	Percent Agreement (95% CI)	Weighted AC1 (95% CI)	Weighted Percent Agreement (95% CI)
All raters (N=10)	0.43 (0.28-0.58)	56 (45-67)	0.48 (0.29-0.67)	73 (63-82)
Experienced raters (N=2)	0.67 (0.50-0.85)	75 (62-88)	0.79 (0.64-0.94)	89 (81-96)
Well-trained raters (N=6)	0.50 (0.38-0.61)	61 (53-70)	0.61 (0.49-0.72)	79 (73-85)

Acknowledgments A clinical research grant from the Glenrose Rehabilitation Hospital Foundation supported the study. The Interfaculty Graduate Studentship awarded jointly between the Faculty of Rehabilitation Medicine and the Faculty of Medicine and Dentistry, University of Alberta supported Sanja Schreiber. We would like to thank: Kathleen Shearer for research coordination; the Rehabilitation Medicine Technology Group for the website development and IT support; Axel Hennes, international Schroth instructor for reviewing the algorithm and operational definition prior to training, and all the patients and therapists for participation in our study.

3.5 Discussion

We proposed a standardized algorithm to achieve adequate rater's reliability in using the Schroth curve type classification. For the first time, we presented the intra- and inter-rater reliability of Schroth therapists in determining the Schroth curve types in adolescents with idiopathic scoliosis who were conservatively treated.

Applying Gwet's benchmarking method described in the analysis section, the overall intra-rater AC1 did not meet our threshold of 0.61 set a priori to recommend clinical use, but the weighted AC1 did. The intra-rater AC1 and weighted AC1 for well-trained raters (the raters who reported full algorithm understanding) were substantial.

Overall inter-rater AC1 ranged from fair, to moderate for the weighted AC1. The inter-rater AC1 and weighted AC1 for well-trained raters were moderate. The two raters with the most experience using the algorithm were most reliable in classifying patients, with substantial intra-rater AC1 and moderate inter-rater AC1. Their weighted intra-rater AC1 was almost perfect and inter-rater AC1 was substantial. The intra-rater AC1 and weighted AC1 for the experienced and well-trained therapists met our a priori threshold to be considered sufficient for recommending the use of our algorithm, and so was the inter-rater AC1 for the experienced therapists.

Because the Schroth classification is the only system solely used to guide the exercise prescription, comparisons with other reliability publications is difficult. Rigo's[108] and Weiss'[129] classifications were derived from the Schroth classification to guide brace design. Although Weiss suggested his classification could guide exercise treatment, he did not explain the exercise prescription for each category and reliability is unknown. The King's[109], Lenke[114] and PUMC[117] classifications were proposed to guide surgical procedures, thus meant for patients with more significant deformities. Algorithms proposed for King's and Lenke's classifications improved reliability and accuracy. With the usage of an algorithm the intra-rater reliability of the King's classification improved from previously published Kappa of 0.64[112] to 0.85[130] and the inter-rater from 0.44[112] to 0.82.[130]. When compared to the gold standard (software that automatically classifies curve types of AIS cases according to the Lenke classification system, based on radiographic measurements of Cobb angles), the accuracy of the Lenke's classification improved from 77.1% to 92.9%[136] by using an algorithm.[136].

Because the reliability of the Schroth classification has not been reported without the use of an algorithm it is not possible to determine if reliability was optimized by the use of our algorithm. The lower inter-rater reliability observed for some of our raters may be related to difficulties in detecting small external deformities in our sample. Our patients were treated conservatively and presented with smaller deformity than what would have been encountered when planning surgery. Nevertheless, we demonstrated that better understanding and experience with the usage

of our algorithm improved the reliability to levels approaching the best surgical classifications.

Because the treatments are similar for some Schroth curve classifications in terms of the direction of the corrective exercises used, some classification disagreements could be viewed as partial agreements. Further, during the course of treatment, a therapist would be able to confirm or change classification and adjust the treatment if needed. Our results may therefore reflect a lower limit of the reliability because therapists were shown only one assessment on which they had no control. In contrast, in case of full disagreement, the prescribed exercises not intended for the classification would cue postural corrections in opposite and unfavorable directions, so the patients could theoretically deteriorate. The weighted analysis is therefore more relevant for clinical practice when treating patients using Schroth exercises.

In the present study, there were two broad sources of variability: the raters and the patients. Variability associated with raters can be due to (1) their interpretation of the classification process, or (2) their perception of the patients. To minimize differences in therapists' interpretation of the classification process, we proposed a standardized Schroth classification algorithm with operational definitions and instructions to appraise the key characteristics of patients. However, therapists' differing comprehension of the algorithm remains a possible source of error. Indeed, we found that the two most experienced therapists in using the algorithm had the highest reliability. Overall, the self-reported comprehension of the algorithm was between 50%-100% and correlated significantly with the intra-rater reliability coefficient in this sample of only 10 therapists (Spearman's Rho=0.68, $p<0.05$). Lower levels of understanding observed in some therapists, perhaps calls for additional training. Moreover, the algorithm itself may need to be modified to improve reliability. Although suggestions for improving the algorithm were solicited, the therapists did not provide any.

Most therapists thought that a radiograph would be helpful in assisting the classification process. Although, we agree that radiographs could help, they are generally not available to physical therapists in most jurisdictions, and the German Schroth School continues to teach the classification based only on the clinical presentation. In contrast, the Barcelona Scoliosis Physical Therapy School uses radiographs in addition to the clinical exam to classify patients. Not providing the radiographs with the video assessment helped ensure the generalization of our findings to a large group of practicing Schroth therapists.

All raters found the algorithm user-friendly, and seven of 10 stated they would use it clinically. One therapist would not use the algorithm because of having now adopted a different treatment approach. Another therapist developed her own system, which includes radiograph assessment. Finally, a therapist believed that because curve type may change during treatment, the need for classifying at baseline was small. Although, curve type may change over time, it is unlikely to be drastic (e.g. from 3cp to 4cp). Schroth instructors continue suggesting that treatment be planned at baseline based on the curve classification.

Patient-specific characteristics might have also influenced reliability. To avoid sampling bias and maximize generalizability, patients were consecutively selected from our specialized scoliosis clinic. However, Lenke 1, 2 and 5 curve types that partially correspond to 3c, 3cp and 4cp curve types according to Schroth, reportedly account for 83% of all scoliosis patterns[137]. Thus, there might have been some underrepresentation of the 4c Schroth curve type. If some curve types are more difficult to classify, this imbalance might have affected reliability. The lower intra-rater reliability observed for some of our raters may be related to difficulties in detecting small external deformities in our conservatively treated sample.

Video quality could have also been a source of variability. After training, a professional photographer helped improve the quality of the videos by recommending background colors to improve contrast. We also adjusted the lighting to avoid shadows affecting the 3D perception of the deformity. Despite our efforts, videos were not always perfectly clear, but most therapists reported the quality was adequate.

Lastly, although video-assessment is used in clinical reliability studies[138-140], the inability to rate in person, might have influenced the reliability. Difficult-to-rate patients may be assessed more appropriately, if the therapists could observe and palpate patients in different positions or while performing movements in person. Palpation and other types of observation are not taught as being necessary to classify patients, thus we believe the current standardized video assessment adequately reflects the assessment of most practicing Schroth therapists.

Results from this study are generalizable, because the patients and the therapists were enrolled consecutively among all eligible volunteers. Patients were typical of those for which exercises

are indicated. Therapists' recruitment targeted all international therapists and ultimately included representation from seven countries.

Although, the results of this study are promising, more research is warranted. First, a reliability study without using the algorithm could be conducted to assess the true value of using the algorithm. Second, the sources of disagreement between therapists should be determined by calculating the reliability on each step within the algorithm, and finding which one influenced the reliability most significantly. Another step should aim at discovering whether there were some difficult patients to rate, and what were their characteristics. The algorithm should also be tested with therapists assessing patients in person. Finally, it would be important to assess whether the reliability changes with the severity of the scoliosis.

In summary, we proposed a simple standardized Schroth curve type classification algorithm to maximize the reliability of curve classification and to therefore help prevent treatment errors that could stem from inappropriate classification choices. The algorithm was well accepted with a majority of raters planning to use it clinically. Overall intra- but not inter-rater weighted reliability estimates were adequate to recommend clinical use of the algorithm and classification with the training provided. Raters with more experience using the algorithm achieved adequate reliability suggesting that with more training we can recommend usage of our standardized algorithm clinically.

CHAPTER 4

The effect of Schroth exercises added to the standard of care on the Cobb angle in adolescents with idiopathic scoliosis – an assessor and statistician blinded randomized controlled trial[2]

4.1 Summary

Background The North American non-surgical standard of care for adolescent idiopathic scoliosis (AIS) includes observation or bracing, but not exercises. This randomized controlled trial (RCT) aimed to determine the effect of Schroth exercises added to standard of care (Experimental group) on the Cobb angle compared to standard of care alone (Control group) in patients with AIS.

Methods Fifty patients with AIS aged 10-18 years, with curves of 10°-45°, recruited from a scoliosis clinic were randomized (1:1) to the Experimental or Control group. A Schroth home program was taught over five individual sessions, and adjusted weekly according to an algorithm. The assessors and the statistician were blinded to allocation. The primary outcome was the change in the Cobb angle of the largest curve (LC) from baseline to six months. Per protocol and intention-to-treat analyses are reported. Trial registration: Schroth Exercise Trial for Scoliosis NCT01610908.

Findings After six months, in the per protocol analysis, the Schroth group had significantly smaller LC than controls (-3·5°, 95%CI -1·08 to -5·92, $p<0.01$). Likewise, the difference in the square root of the sum of curves (SOC) in each group was -0·40, (95%CI -0·02 to -0·78, $p<0.05$), suggesting that an average patient with a 51·2° SOC at baseline, will have a 49·3° SOC at six months in the Schroth group, and 53·1° in the control group with the difference between groups increasing with severity. Intention-to-treat analyses produced similar differences, but they were not statistically significant. Completers attended 85% of visits and completed 82·5% of the prescribed home program.

[2] Sanja Schreiber, Eric C. Parent, Elham Khodayari Moez, Douglas M. Hedden, Doug Hill, Marc Moreau, Edmond Lou, Elise M. Watkins, Sarah Southon

Interpretation Schroth exercises added to the standard of care may help improving the curve severity in patients adhering to the program.

Funding Scoliosis Research Society, Glenrose Hospital Foundation, Sick Kids Foundation-Canadian Institute of Health Research and University of Alberta.

4.2 Introduction

Adolescent idiopathic scoliosis (AIS), a three-dimensional spinal deformity, is the most common (84%-89%) form of scoliosis[141] with a prevalence between 0·47 and 5·2%.[142] There is a high preponderance of AIS among girls. The risk of progression is linked to the remaining growth potential and initial curve magnitude.[13] Scoliosis may lead to decreased self-esteem, mental health concerns, pain, respiratory complications, and limited function. The negative consequences generally manifest once curve exceeds 40°-50°.[1] Therefore, early treatment is recommended throughout pubertal growth.

In North America, the standard of care for growing patients with AIS includes observation (curves between 10° to 25°), bracing (curves between 25° to 45°), and elective surgery (curves >45°).[10] The efficacy of exercise treatments is controversial. Although evidence suggests that scoliosis-specific exercises could improve some outcomes,[18] exercise therapy has not yet been widely accepted. European guidelines recommend exercises used alone and as an add-on to bracing for patients with curves <45° to prevent further curve progression, and the need for surgery.[128,143] Differences between the guidelines may be due to cost, culture, social standards or differing appraisals of the quality of research involving exercises.

Bracing can induce stress[94], fear of injury[94], discomfort[94], limitation in activities,[94] impair lung function,[144] and negatively affect self-esteem.[94] While surgery reduces deformity, patients fear surgery due to its invasiveness, risk of complications, post-surgical pain, and long recovery. Surgery permanently limits the active range of movement,[102] and scoliosis may still progress in the non-fused spinal segments. Conversely, exercises are well received,[145] and frequently requested by patients and their parents.[146]

Several systematic reviews on exercises for scoliosis[18,79-81] report promising results on curve severity, improving neuromotor control, respiratory function, back muscle strength, and cosmetic

appearance. However, most reviews have risk of reviewer bias as they were published by authors of studies included in these reviews. In a recent independent review,[147] nine prospective cohort studies were included, of which only three were controlled and only one used observer blinding. Other limitations of exercise studies included unclear reporting of patient selection criteria, recommendations for, and contraindications to exercise, not reporting on compliance, intention-to-treat analyses, or recruitment strategies. Change in Cobb angles was usually statistically significant, but often within the measurement error.

Among the promising scoliosis-specific exercise approaches reviewed, Schroth exercises were the most studied. The Schroth method consists of sensorimotor, postural and breathing exercises aimed at recalibration of normal postural alignment, static/dynamic postural control, and spinal stability.[21] Several studies of limited quality demonstrated positive outcomes of Schroth exercises on back muscle strength,[23] breathing function,[23] slowing curve progression,[24] improving Cobb angles,[23,24] and decreasing the prevalence of surgery.[25]

Therefore, the objective of this first RCT on Schroth exercises was to determine the effect of a six-month Schroth exercise intervention added to standard of care (observation or bracing) on the Cobb angle, compared to the standard of care alone in patients with AIS. We hypothesized that Schroth exercises would improve scoliosis curves.

4.3 Methods

Methods

Study design

This was a parallel, phase II, assessor and statistician blinded, randomized controlled clinical trial. The full protocol for this study along with the CONSORT checklist is available in Appendix 1.

Participants and therapists

Between April 2011 and November 2013, 50 patients with AIS were enrolled from the Stollery Children's Hospital scoliosis clinic in Edmonton, Alberta, Canada. Inclusion criteria were: 10-18 years old, both genders, curves between 10°-45°, Risser 0 to 5 and the ability to attend weekly

visits. Exclusion criteria were: patients with diagnosis other than AIS, having completed brace treatment, scheduled surgery, a follow-up later than 6±2 months, and previous spine surgery. We obtained assent from patients and informed parental consent. This study was approved by the local ethics review board (Pro00011552).

Prior to the study, the main therapist had three years of Schroth therapy experience and provided about 95% of the therapy sessions. A second certified therapist filled-in as needed.

Randomization and masking

A study coordinator invited consecutive eligible patients attending the scoliosis clinic to participate. Within two weeks from this visit, a researcher obtained consent and booked an evaluation to confirm eligibility and collect baseline data. After this exam, participants were randomized using a computer-generated sequence in pre-sealed envelopes into the Schroth exercises or the control group. We used random size (4-8) blocked randomization stratified for the four Schroth curve types to ensure a balanced allocation of curve types in both groups (25/group).

Therapists and participants could not be blinded to the treatment. Participants were asked not to reveal their group allocation to ensure evaluator blinding. The statistician was also blinded to coding of group allocation. Radiographs were obtained during routine clinic visits by a trained technician blinded to study participation. An experienced evaluator masked to groupings and timing of scanning measured the radiographs.

Intervention

Experimental group

The six-month supervised Schroth exercise intervention included five one-hour long private sessions delivered during the first two weeks, followed by weekly one-hour long group classes combined with a 30-45 min daily home exercise program. Exercises are described in Appendix 3, with the corrective movements required, the targeted curve type, the level of passive support involved, whether static or dynamic, and the dosages recommended. A Schroth curve classification algorithm[148] and algorithms to guide the exercise prescription and progression for each Schroth curve type[149] were developed to standardize treatment and ensure reproducibility.

Compliance was monitored using logbooks, and verified daily by a parent and weekly by the therapist. Therapists assessed adequate exercise performance weekly using a checklist. Attendance was calculated as a percentage of prescribed visits, and compliance as a percentage of the prescribed exercise dose completed over six months.

Control group

Control subjects received the standard of care including observation or bracing if the Scoliosis Research Society (SRS) bracing criteria were met, and attended only study assessments.

Measurements

The Cobb angle of the largest curve (LC) was the primary outcome. The secondary outcome was the sum of all curves measuring ≥10° (SOC) to ensure capturing changes affecting all curves. Standing posterior-anterior radiographs were obtained using a positioning frame at baseline and six months. Cobb angles were measured for each curve using semi-automated software with measurement error ≤2.5°.[150]

The self-efficacy questionnaire (SEQ) score was collected at baseline as a covariate for the analyses. This validated questionnaire measures self-efficacy for overcoming barriers to physical activity (defined as corrective exercises) using eight items rated from one (Disagree a lot) to five (Agree a lot).[151]

Baseline and six months evaluations included a physical exam where height, weight, trunk rotation using scoliometer, Schroth curve classification, back muscle endurance, demographics, and the SEQ scores were recorded. The back muscle endurance, Spinal Appearance Questionnaire and SRS-22r questionnaires were used as secondary outcomes and will be reported separately.

Statistical analysis

To assess differences between groups in changes from baseline to six months while adjusting for important covariates, we used per protocol (PP) and intention-to-treat (ITT) linear mixed models analysis. ITT analysis was performed using linear interpolation.[152] Separate analyses were conducted for each outcome. Covariates considered included age, weight, height, self-efficacy,

brace-wear (yes/no), and scoliosis classification. For covariates selection, we used a stepwise method with Akaike information criterion (AIC).[153] Outcomes were transformed as needed to meet normality assumptions.

To detect a 0.50 effect size when comparing the change in the primary outcome between two groups with 80% power using a two-tailed 0.05 hypothesis test, and considering a 0.6 correlation between repeated measures in two time points, 50 patients per group was needed.[154] However, the study ended after recruiting 50 participants when funding was received to continue the study as a multicenter RCT with slightly different participants' criteria (Trial registration NCT01610908). Statistical analyses were performed using the statistical program R.[155]

Role of the funding source

Funding included a small exploratory grant from the Scoliosis Research Society, the Glenrose Rehabilitation Hospital Foundation, and a SickKids Foundation – CIHR Institute of Human Development, Child and Youth Health new investigator award. SS received a scholarship from the Faculties of Rehabilitation Medicine and Medicine and Dentistry at the University of Alberta. EW was funded by CIHR. Sponsors had no role in study design, data collection, data analysis, interpretation, or writing the report. SS and EP had access to all data and final responsibility for the decision to submit for publication.

Table 4.1. Baseline characteristics of the study population

	Schroth exercises + Standard of care (95% Confidence interval), N=25	Standard of care (95% Confidence interval), N=25
Age (years)	13·5 (12·7-14·2)	13·3 (12·7-13·9)
Girls	23	24
Braced patients	17	17
Height (m)	1·60 (1·6-1·6)	1·60 (1·6-1·6)
Weight (kg)	45·9 (42·6-49·1)	50·5 (47·1-54·0)
Largest curve (°)	29·1 (25·4-32·8)	27·9 (24·3-31·5)
Sum of curves (°)	48·1 (39·1-57·2)	54·3 (44·9-63·6)
Risser sign (0 to 5)	1·76 (1·10 to 2·45)	1·44 (0·77 to 2·11)
Risk of progression (%)	65	65

4.4 Results

Groups did not differ at baseline for height and age. However, controls were 4·4 kg heavier than Schroth participants. Forty-seven girls and three boys were evenly distributed between groups. The mean height, weight and age were 1·60m (SD=0·1), 48·2kg (SD=8·3), and 13·4 years (SD=1·6), respectively. The mean LC was 28·5° (SD=8·8°) and the mean SOC was 51·2° (SD=22·3°)*(Table 1)*.

Schroth curve types were as follows: 3c (n=7) affecting the thoracic spine without pelvis imbalance, 3cp (n=15) thoracic dominant deformity with imbalanced pelvis observed on the thoracic concave side, 4c (n=5) with a thoracolumbar/lumbar dominant deformity without pelvis imbalance and 4cp (n=23) with a thoracolumbar/lumbar dominant deformity with pelvis displaced to the lumbar concave side. Curve types were balanced between groups with no more than one subject difference for each type.

Dropouts

Attrition was 12% (6/50), with four dropouts in the Schroth and two in the control group. Of those, there were four girls (one control and three in the Schroth group) and two boys (one per group). The LC (23°, SD=5·3) and SOC (38°, SD=17·5) of patients who dropped out were smaller than for the remaining patients.

Compliance

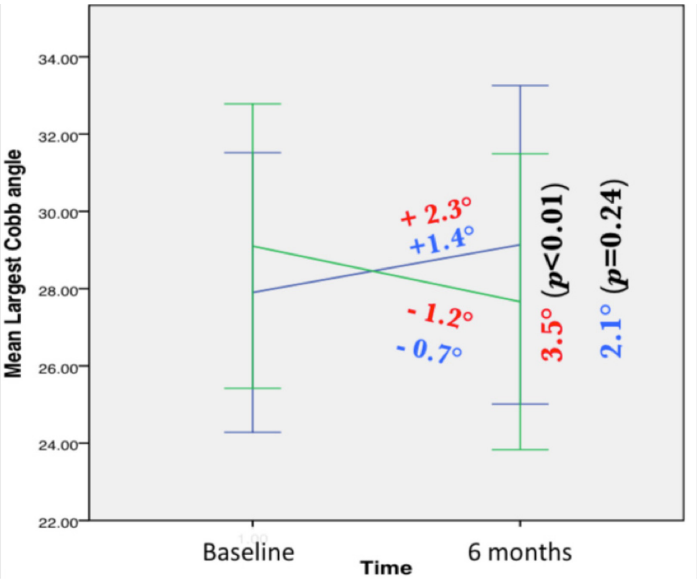
Patients with complete follow-up attended 85% of prescribed visits and completed 82·5% of the home program. In the ITT analysis, 76% of visits were attended and 73% of the prescribed home exercises were completed.

Largest curve

In the PP analysis, the difference in LC between groups at six months was -3·5° (CI -5·01° to -1·09°, p=0.007). On average, after adjusting for confounders the LC decreased by 1·2° in the Schroth and increased by 2·3° in the control group over six months. The covariates weight, the curve classifications 3cp and 4cp had significant main effects on the LC (p=0·02, p=0·0008, p=0·01, respectively)(*Table 4.2*).

The ITT analysis produced similar values for the LC outcome, but the difference between groups was not significant (-2·08°, CI -5·55° to 1·39°, p=0.24). (*Table 4.2*) The weight and classification had significant main effects on the LC (p<0·01). The interaction between group and time, as well as the individual growth curves are shown in the *Figure 4.1*. The individual growth curves are presented in the Appendix 4.

Figure 4.1. The interaction between groups and time shows an increase of the LC in the control and decrease in the Schroth group over the six months follow-up. Groups are colored in green (Schroth) and in blue (Control). Numbers in red represent the results of the PP and in blue of the ITT analysis.



Sum of curves

In the PP analysis, after adjusting for confounders, the difference between groups over time was statistically significant favoring the Schroth group (-0.40, 95% CI -0·77° to -0·03°, p=0.048). (*Table 4.2*) To meet the normality assumption, the SOC was transformed to its square root. The square root of SOC decreased by -0·14 in the Schroth, and increased by 0·26 in the control group over six months. This difference in square roots of the SOC between the groups indicate that a patient with characteristics corresponding to the baseline mean SOC of 51·2° and the selected covariate set will have a SOC of 49·3° after six months in the Schroth, and a SOC of 55° in the

control group. Moreover, the difference between groups increased with severity. Weight and classification 3cp had significant main effects on the SOC (p=0.009, p=0.02, respectively).

Although the model suggests decreasing curve severity in the Schroth group over six months and increasing in the control group, in the ITT analysis, the difference between groups was not statistically significant (-0·39°, CI -1·02° to 0·24°, p=0.23). The main effect of weight and classification was statistically significant (p<0·01). (*Table 4.2.*) The interaction between group and time are shown in the *Figure 4.2*. The individual growth curves are presented in the Appendix 4.

Figure 4.2. The interaction between groups and visits in the per protocol analysis, shows an increase of the SOC in the control group and a decrease in the SOC in the Schroth group. Groups are colored in green (Schroth) and in blue (Control). Numbers in red represent the results of the PP and in blue of the ITT analysis.

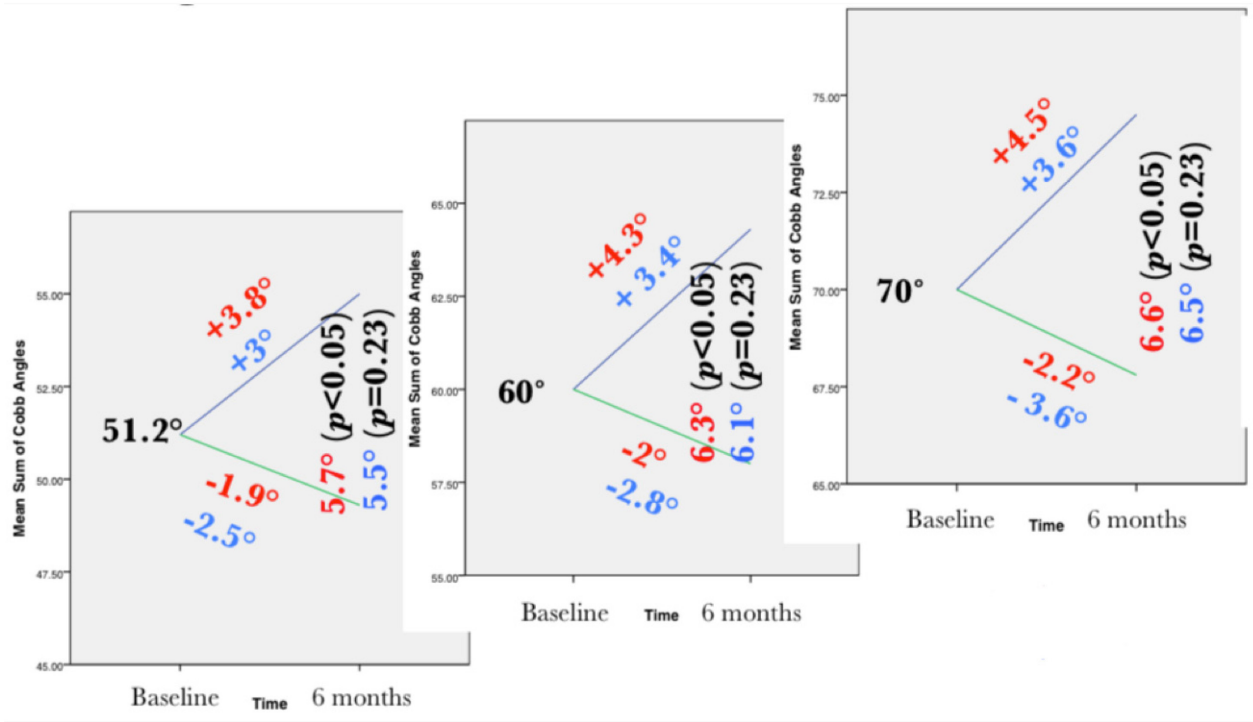

Mean Sum of Cobb Angles
55.00
52.50
50.00
47.50
45.00
51.2°
+3.8°
+3°
-1.9°
-2.5°
5.7° (p<0.05)
5.5° (p=0.23)
Baseline
Time
6 months
Mean Sum of Cobb Angles
65.00
62.50
60.00
57.50
55.00
60°
+4.3°
+3.4°
-2°
-2.8°
6.3° (p<0.05)
6.1° (p=0.23)
Baseline
Time
6 months
Mean Sum of Cobb Angles
75.00
72.50
70.00
67.50
65.00
70°
+4.5°
+3.6°
-2.2°
-3.6°
6.6° (p<0.05)
6.5° (p=0.23)
Baseline
Time
6 months

Table 4.2 Linear mixed model coefficients for the significance of each coefficient and its associated p-values in the per protocol and the intention-to-treat analyses

	Per protocol					Intention-to-treat				
	Value	95% Confidence interval	t-value	DF	p-value	Value	95% Confidence interval	t-value	DF	p-value
Largest Cobb (°)										
Interaction group * time	- 3·50	-5·91 to -1·09	- 2·84	41	0·007	-2·08	-5·55 to 1·39	-1·17	48	0·244
Group	6·63	1·14 to 12·12	2·36	42	0·023	4·96	-1·39 to 11·31	1·53	48	0·133
Time	2·29	0·53 to 4·05	2·54	41	0·015	0·71	-1·74 to 3·16	0·56	43	0·574
***Predictors*:**										
Height	- 32·21	-69·25 to 4·83	- 1·70	42	0·096	- 22·36	-58·33 to 13·61	-1·22	43	0·226
Weight	0·47	0·08 to 0·86	2·31	42	0·026	0·35	-0·02 to 0·72	1·76	43	0·082
Classification 3cp	12·19	5·57 to 18·81	3·61	42	0·001	11·06	4·53 to 17·59	3·32	43	0·001
Classification 4c	1·87	-6·70 to 10·44	0·43	42	0·670	1·68	-6·77 to 10·13	0·39	43	0·700
Classification 4cp	8·76	2·41 to 15·11	2·70	42	0·010	8·26	2·05 to 14·47	2·61	43	0·011
Sum of all curves (°)										
Interaction group * time	- 0·40	-0·77 to -0·03	-2·04	41	0·048	- 0·39	-1·02 to 0·24	-1·21	48	0·233
Group	0·45	-0·53 to 1·43	0·90	42	0·372	0·39	-0·77 to 1·55	0·67	48	0·506
Time	0·26	-0·01 to 0·53	1·84	41	0·073	0·21	-0·24 to 0·66	0·93	43	0·355
***Predictors*:**										
Height	- 5·64	-12·48 to 1·20	-1·61	42	0·114	- 3·82	-10·42 to 2·78	-1·14	43	0·262
Weight	0·10	0·03 to 0·17	2·71	42	0·010	0·08	0·01 to 0·15	2·22	43	0·032
Classification 3cp	1·49	0·27 to 2·71	2·40	42	0·021	1·37	0·17 to 2·57	2·25	43	0·030
Classification 4c	-1·15	-2·74 to 0·44	-1·43	42	0·160	-1·19	-2·74 to 0·36	-1·51	43	0·139
Classification 4cp	0·31	-0·87 to 1·49	0·52	42	0·607	0·25	0·89 to 1·39	0·42	43	0·673

4.5 Discussion

This RCT demonstrated positive effect of Schroth exercises added to standard of care (observation and bracing) on the LC and the SOC in patients with AIS. The difference in SOC between groups increased with larger baseline SOC. After six months, the LC decreased in the Schroth group by 1·2°, but increased in the control group by 2·3°. That 3·5° difference was statistically significant in the PP, but not in the ITT analysis. The same PP vs. ITT trend was noted for the SOC.

Many consider a 5° change in Cobb angle clinically important. This threshold is based on reported standard errors of measurement (SEM) for manual Cobb angle measurements. The SEM for our semi-automated method is <2·5°.[150] According to natural history, scoliosis curves progress on average by 0·9°/month, with a range of 0·3°-1·6°/month.[156] This corresponds to an average expected progression of 5·4° over six months (range 1·8°-9·6°). Bracing was recently reported effective at preventing progression to the surgical range but did not produce curve improvements on average.[99] In our trial 17 participants per group wore a brace. Therefore, observing curve improvements in the Schroth and deteriorations in the control group beyond the SEM and bracing effect after only six months seems clinically important.

Assuming that all patients with missing values experienced curve progression, three (12%) deteriorated by >5° in the Schroth group, four improved (16%), and 18 remained stable (72%). In the control group, 10 deteriorated (40%), one improved (4%) and 14 (56%) remained stable. More patients were successfully treated (improved + stable) in the Schroth than in the control group (88% vs. 60%, Chi-square 5.1, p=0.024). Our results demonstrate the benefit of Schroth exercises.

Several controlled studies on scoliosis-specific exercises for AIS have reported significant effects on curve severity. Wan et al's RCT report larger Cobb angle improvements (from 26°±12° to 10°±7°) in the group treated daily with scoliosis-specific exercises (not Schroth) added to standard care (surface electrical stimulation, traction and postural training) than with standard care alone (from 25°±13 to 18°±9°) over six months.[157] Comparison with our results is difficult because the type of scoliosis investigated in this RCT is unclear.

Monticone et al's RCT found that scoliosis-specific active self-correction and task-oriented exercises improved Cobb angles by 5·3° at skeletal maturity and that traditional exercises were associated with stable curves.[158] Their larger Cobb angle improvement could be explained by the lower risk of progression of their sample, and a longer intervention (until skeletal maturity vs. six months in our study). Their sample had smaller curves than ours at baseline (19·3°±3·9° vs. 28·5°±8·8°), but lower age (12·5±1·1 vs. 13·4±1·6), and Risser grades (0·55 vs. 1·60). According to Lonstein and Carlsson's formula ((Cobb – 3 x Risser)/Age)[13], Monticone's sample had a 35% risk of progression vs. 65% in our study.

Negrini et al's[159] prospective study found that one year of the Scientific Exercises Approach to Scoliosis (SEAS) improved the LC by 0·33°, and the SOC by 0·67° while in the "usual" rehabilitation program the LC worsened by 1·12° and the SOC by 1·38°. Our larger improvements compared to Negrini et al's possibly arose because of our more intense therapy (daily home sessions and weekly visits over six months vs. twice/week of home exercises and 4-6 visits over one year).

Noh et al's retrospective study found better effects of scoliosis-specific "3D corrective technique" including Schroth and stabilization exercises compared to symmetrical stretching and stabilization exercises on the Cobb angle.[160] Treatment dose (60-minute sessions, 2-3 times a week, for 30 sessions over four months) was lower than in the present study. Authors reported improvement in Cobb angle of 8·1°±4·5° in the experimental and 4·3°±2·1° in the control group, which were larger than in our study. However, their sample had lower (10%) estimated risk of progression[13] compared to ours (65%).

The only prospective cohort study that specifically focused on Schroth exercises[23] showed improved Cobb angles in 49/50 adolescents and one stable curve after one year. Treatment was intensive consisting of four-hour sessions, five days/week for six weeks, followed by the same program at home with biweekly follow-ups until six months, and then bimonthly until one year. Mean LC decreased from 26·1° to 19·2° over six months. The higher intensity might explain the lower compliance (74%), high dropout rate (25%), and larger change observed in the Cobb angle compared to our study.

Several features of this RCT helped reduce the risk of bias. Randomization balanced number of

patients wearing a brace in both groups, and curve types distribution. The evaluators and statistician were blinded. We standardized curve classification[148] and exercise prescriptions[149] using algorithms. Patients reported not using co-interventions at follow-ups. Exercise dosage led to high adherence monitored via patients/parent/therapist logbook to minimize the overestimation. Compliance and attendance rates were reported using ITT. We reported ITT analysis and acknowledged reasons for missing data. The main reason for non-compliance and dropout was "time constraint due to homework". The PP analysis using linear mixed models with maximum likelihood estimation is justified when missingness is unrelated to outcomes as in our study.[154]

Ours was the first study to stratify randomization using curve types. Patients with major thoracic curves and deviated pelvis to the thoracic concavity (3cp) had the largest curve magnitudes, possibly because of their worst prognosis for progression.[1] In contrast, patients with double major curves (corresponding to 4c) had the smallest curve magnitude all including left lumbar and right thoracic, which have a better prognosis.[1] Differences between patterns emphasize the importance of accounting for curve type in randomization.

A limitation of this study includes limited statistical power due to early termination, and the 12% attrition rate. Regardless, we detected a large effect (Cohen's d of 0·92) amongst compliant patients. Subject heterogeneity could be another limitation. We included patients of all maturity levels. More mature patients (Rissers 3-5) have lower risk of progression, and potentially better treatment success. Nevertheless, our sample's estimated risk of progression was higher than in most exercise trials. Lastly, this study could not determine the effect of only Schroth exercises, because exercises were combined with standard of care. Ethically, we could not withhold the bracing from the patients meeting the SRS criteria.

In conclusion, in this first RCT, six-months of Schroth exercises added to standard of care improved curve severity in adolescents with idiopathic scoliosis compared to standard of care. Statistical significance was reached in the PP, but not in the ITT, which emphasizes the importance of compliance with the exercise program. Low dropout and adequate compliance rates indicate the feasibility of adding Schroth intervention to the standard of care in North America. This trial increases the level of evidence on the benefits of Schroth exercises for AIS

by its methodological rigor, and justifies continued investigation in the ongoing Multicenter Schroth Exercise Trial.

CHAPTER 5

The effect of Schroth exercises added to the standard of care on the quality of life and muscle endurance in adolescents with idiopathic scoliosis – an assessor and statistician blinded randomized controlled trial[3]

5.1 Summary

Background In America, care recommendations for adolescents with small idiopathic scoliosis (AIS) curves include observation or bracing. Schroth scoliosis-specific exercises have demonstrated promising results on various outcomes in uncontrolled studies. This randomized controlled trial (RCT) aimed to determine the effect of Schroth exercises combined with the standard of care on quality-of-life outcomes and back muscle endurance compared to standard of care alone in patients with AIS. Trial registration: Schroth Exercise Trial for Scoliosis NCT01610908.

Methods Fifty patients with AIS, aged 10-18 years, with curves 10°-45°, recruited from a scoliosis clinic were randomized to receive standard of care or supervised Schroth exercises plus standard of care for six months. Schroth exercises were taught over five sessions in the first two weeks. A daily home program was adjusted during weekly supervised sessions. Assessors and the statistician were blinded. Outcomes included the Biering-Sorensen back muscle endurance, Scoliosis Research Society (SRS-22r) and Spinal Appearance Questionnaires (SAQ) scores. Per protocol and intention-to-treat linear mixed models analyses are reported.

Findings After six months, the groups did not differ significantly for any questionnaire score. However, Schroth exercises improved the back muscle endurance compared to standard of care alone in adolescents with AIS by 30 seconds on average (p=0.02). Age, weight, height, self-efficacy, brace wear, and scoliosis classification were retained as covariates in different models.

[3] Sanja Schreiber, Eric C. Parent, Elham Khodayari Moez, Douglas M. Hedden, Doug Hill, Marc Moreau, Edmond Lou, Elise M. Watkins, Sarah Southon

Interpretation Supervised Schroth exercises provided added benefit to the standard of care by improving muscle endurance, but not questionnaires' scores. Given the high prevalence of ceiling effects on both questionnaires' domains, we hypothesize that in the AIS population receiving conservative treatments, different QOL questionnaires with adequate responsiveness are needed.

Funding Scoliosis Research Society, Glenrose Hospital Foundation, Sick Kids Foundation, Canadian Institute of Health Research and University of Alberta.

5.2 Introduction

Adolescent idiopathic scoliosis (AIS) is a progressive three-dimensional deformity of the spine of unknown cause that occurs during the puberty in otherwise healthy children.[1] The prevalence in the general adolescent population has been reported to be up to 5.2%[142] and the annual incidence about 2%.[5] The severity of scoliosis is monitored using the Cobb angle, which reflects the spinal curvature in the frontal plane, to inform treatment decisions. Scoliosis progresses most rapidly during the pubertal growth spurt. The negative consequences of progressive scoliosis generally manifest more importantly once scoliosis curves exceed 40°-50°.[1] Scoliosis leads to decreased self-esteem,[161] mental health concerns,[162] pain,[58,63,163,164] respiratory complications[165] and limited function.[58,63] These observations justify efforts to start the treatment early before the pubertal growth.

In North America the standard of care for scoliosis includes: observation for patients with curves between 10° to 25°, and who are still growing; bracing for patients with curves between 25° to 45° during the growth phase, and spinal fusion for patients with curves >45° while the patients are still growing and >50° if the growth has ceased.[42] In contrast, European scoliosis treatment guidelines include exercises to prevent further progression of the deformity, and to prevent the need for bracing or surgery.[128,143,166] Cost, culture,[11] social standards[146] or, possibly, differing appraisals of the quality of research involving exercise treatments[42] could explain these differences in treatment recommendations.

Standard care consisting of observation, bracing and surgery has its shortcomings. Patients with curves ≤25° are routinely radiographed to assess the progression every 4-6 months, but no

treatment is offered. Over six months, scoliosis curves are expected to progress 5.4° on average, with some fast progressive curves predicted to increase by as much as 9.6°.[156] With quick curve progression, body shape asymmetries can also develop, quickly affecting the trunk, pelvis, ribs, shoulder, lumbar and waist areas. Symptoms such as pain, psychological issues, and a decreased quality of life (QOL) can appear or the likelihood of experiencing such symptoms during adulthood can increase during observation.[50,105,110] Brace-wear may induce stress, negatively affect self-esteem,[94] produce soreness, discomfort with activity, torn clothing, as well as, limitations in sport, physical activity and social events.[51,94] Although, brace treatment alters the natural history and limits progression during the growth phase[99], the long-term effects on curve progression and surgical rates after maturity are not significantly different in patients treated with braces and observation.[68] While surgery reduces deformity and prevents further curve progression,[50] in the long term, surgically treated patients have more degenerative disc changes than controls, more frequent lumbar or bodily pain, reduced physical functioning and general health as well as more sick-leaves due to back pain.[58]

Generally, exercises are well received by patients.[145] Exercise is frequently sought by patients and their parents.[146] It has been reported that persons with scoliosis who exercise regularly, show higher self-esteem and have better psychological outcomes.[51] This strengthens the importance of physical therapy as a treatment alternative for AIS.

Several systematic reviews on the effects of exercises for scoliosis have reported promising results [18,79-82] Fusco et al's systematic review[18] concluded that asymmetric exercises slowed the progression (deterioration) of scoliosis and/or reduced curve severity measured by the Cobb angle, improved neuromotor control,[76] respiratory function,[73] back muscle endurance, [73] and cosmetic appearance.[73,84] However, QOL outcomes were not routinely assessed.

Since Fusco's review, three studies on exercises were published. A prospective controlled study of 32 patients demonstrated significantly better effects for scoliosis-specific "3D corrective spinal technique" than for conventional exercise on the radiological and SRS-22 questionnaire scores.[160] However, the conventional exercise group also experienced significant effect on QOL reinforcing the evidence suggesting that various types of exercise can benefit QOL of patients with AIS.[51] In a recent RCT, Monticone et al found that scoliosis-specific active self-correction and task-oriented exercises significantly improved the Cobb angles and the QOL measured at

skeletal maturity by the SRS-22r questionnaire (by 0.75 to 0.89/5) while the traditional spinal exercises were associated with stable outcomes.[158] Participants were 110 skeletally immature patients with AIS and curves <25° at baseline. In a third recent RCT, preoperative aerobic training improved the QOL in 40 surgical candidates with AIS.[167]

Among all scoliosis-specific exercise approaches, the Schroth method is among the most studied and widely used. The Schroth method consists of scoliosis-specific sensorimotor, postural and breathing exercises.[26] Auto-correction, the patient's ability to reduce the spinal deformity through active postural realignment of the spine in three dimensions,[18] is a fundamental component of the Schroth method. Auto-correction is achieved through self-elongation and postural corrections specific for each curve pattern, and is eventually integrated in daily activities. In several cohort studies, the Schroth method demonstrated positive outcomes on back muscle strength,[73] breathing function,[73] slowing curve progression,[24] improving Cobb angles[24,73] and decreasing the prevalence of surgery.[25]

The results from a large case-control study[64] suggest that the back muscle endurance of scoliosis patients is significantly lower than in people without scoliosis. Paraspinal muscles are needed to maintain spinal alignment throughout the day. To our knowledge this outcome has not been investigated in any Schroth exercise studies even though exercises may affect back muscle endurance.

The promising effect of exercises on QOL should be confirmed in a RCT conducted by independent researchers to limit investigator bias, and to address the methodological limitations of prior studies. To our knowledge no RCT or prospective controlled studies have been conducted on the effect of Schroth exercises, and none blinded the assessors. Moreover, most exercise studies did not report on compliance, intention-to-treat analyses, or on recruitment strategies.

Therefore, the objective of this RCT on Schroth exercises was to determine the effect of a six-month program of Schroth exercises in conjunction with standard of care (observation and bracing) on QOL, perceived appearance and back muscle endurance, compared to the standard of care alone in patients with AIS. We hypothesized that Schroth exercises would improve these outcomes.

5.3 Methods

Study design

This was a phase II assessor- and statistician-blinded, randomized controlled clinical trial comparing a six-month Schroth exercise intervention added to standard of care (experimental group) to the standard of care alone (control group). The full protocol for this study is available in Appendix 2.

Participants and therapists

We consecutively enrolled 50 patients with AIS from the scoliosis clinic at our institution. Inclusion criteria were: 10-18 years old, both genders, curves 10°- 45°, Risser 0-5 (all skeletal maturities) and ability to travel to weekly visits. Exclusion criteria were: diagnosis other than AIS, planning surgery, having had surgery, previously weaned from brace, being scheduled for clinical follow-up later than in 6±2 months or being discharged from the clinic when approached to participate. We obtained assent from the patients and informed parental consent prior to the enrolment. This study was approved by the local ethics review board (Pro00011552).

The main Schroth-certified therapist had three years of Schroth therapy experience and provided 95% of the therapy sessions. Another certified therapist filled in as needed.

Randomization and masking

A research coordinator invited eligible patients attending regular scoliosis clinic visits to participate in the study. Within two weeks from the visit, a researcher contacted interested patients to obtain consent and book a baseline evaluation. After an initial exam confirming eligibility and collecting baseline data, participants were randomized using a computer-generated sequence in pre-sealed envelopes into the Schroth exercises or the control group (standard of care). We used random size (4-8) blocked randomization stratified for the four Schroth curve types[168] to ensure allocation of a balanced number of participants in both arms of the study (25 per group) for each curve type.

Therapists and patients could not be blinded when offering or receiving the Schroth treatment. However, participants were asked not to reveal their group allocation to ensure blinding of the evaluator. The statistician was not aware of the data coding.

Intervention

Schroth exercises added to standard care (experimental) group

The six-months supervised Schroth exercise intervention included five initial one-hour long private training sessions delivered during the first two weeks after baseline, followed by weekly one-hour long group classes combined with a 30-45 min daily home exercise program. Exercises are presented in Appendix 3 with a description of the corrective movements required, the curve type for which they are recommended, the level of passive support involved, whether they offer a static or dynamic challenge and the dosages recommended. A Schroth curve type classification algorithm[168] and algorithms to guide the exercise prescription and progression for each of the four Schroth curve types[149] were developed for this trial to standardize exercise delivery.

Compliance was monitored using exercise logbooks, verified daily by a parent and weekly by a therapist. During each class, adequate performance of prescribed home exercises was assessed using a checklist. To maximize compliance, we provided home equipment, access to facilities, and promoted parental involvement. When compliance dropped below 70%, we tried to resolve the issues cooperatively with patients and parents[169,170]. Attendance was calculated as a percentage of visits to our lab, and compliance as a percentage of the prescribed home exercise dose completed over the course of the six months of treatment.

Standard care (control) group:

Control subjects received the standard of care, consisting of observation or bracing if the SRS bracing criteria[171] were met. Controls attended study assessments, but no therapy sessions.

Measurements

The outcomes were collected at baseline, three and six months, and included: Biering-Sorensen back muscle endurance, Scoliosis Research Society (SRS-22r) and Spinal Appearance Questionnaire' (SAQ) domains' scores.

The Biering-Sorensen test is a validated test assessing the isometric endurance of the trunk extensor muscles. This test measures the duration (in seconds) a subject is able to hold the unsupported upper body horizontal, while having the legs and buttocks fixed to a table, with the upper body hanging over the edge of the table and the arms crossed over the chest. The test is stopped when a subject can no longer control the posture or when 240 seconds have been reached.[172] The test-retest reliability was shown to be adequate ICC=0.85(CI 0.76-0.90) with a SEM of 15.6 seconds.[173] The test was validated for measuring back muscle fatigue.[174]

The SRS-22r questionnaire is a scoliosis-related QOL questionnaire that assesses five domains: function, pain, self-image, mental health (five questions each), and satisfaction with care (two questions).[175] Each question is scored from 1-5, where 1 is the worst, and 5 the best. The SRS-22r has adequate test-retest reliability and validity.[176] We analyzed the total score, function, pain and self-image domains, because those outcomes are deemed the top priority in conservative treatment for scoliosis.[177]

The SAQ measures changes in patients' perception of their deformity using 20 questions including standardized drawings. It assesses the following domain scores: trunk shift, waist, kyphosis, prominence, chest, shoulders, general and curve.[178] Each item is scored from 1-5, where 1 is the best, and 5 the worst. In surgically treated patients, the SAQ was responsive, and it has adequate psychometric properties (test-retest reliability of 0.57 to 0.99 for the different scale items and a Cronbach alpha of 0.7). [178] For the purposes of our study, we considered all but the SAQ kyphosis domain because this study did not focus on kyphosis corrections.

Self-efficacy scores were collected at baseline as a covariate for the analysis. This validated questionnaire measures self-efficacy for overcoming barriers to physical activity (defined as corrective exercises) using eight items rated from one (Disagree a lot) to five (Agree a lot).[151,179] Physical activity levels in adolescent girls are related to self-efficacy beliefs.[180,181] Self-efficacy was found to be a moderator of the relationship between declines in physical activity and perceived social support.[180]

Statistical analysis

Descriptive statistics were calculated for baseline demographics and radiographs, for the entire sample, and for the patients who dropped out.

To detect a 0.50 effect size when comparing the change in the primary outcome between two groups with 80% power using a two-tailed 0.05 hypothesis test, and considering a 0.6 correlation between repeated measures, 50 patients per group were needed.[154] However, the study ended after recruiting 50 participants when funding was received to continue the study as a multicenter RCT with slightly different participants' criteria (Trial registration NCT01610908).

All outcomes, except the SAQ curve domain, were treated as continuous outcomes. To assess differences in group changes from baseline to six months for the continuous outcomes while adjusting for important covariates, we used linear mixed models analysis. Both a per protocol (PP) and an intention-to-treat (ITT) analysis were performed. For the ordinal outcome, SAQ curve, which is based only on one item with five levels, we used generalized linear mixed model analysis. The ITT analyses were conducted using the linear interpolation method,[152] in which values immediately surounding the missing data are joined by a line representing the average progression of the individual trajectory. Separate analyses were conducted for each outcome to assure the best covariate set was selected in the model. Covariates considered included age, weight, height, self-efficacy, whether a person wore a brace or not, and Schroth scoliosis classification. We used the stepwise selection method for covariates using the Akaike information criterion (AIC),[153] a goodness of fit measure that favors smaller residual error in the model, but penalizes for including additional predictors and helps avoiding over fitting. Outcome variables were transformed as needed to ensure meeting the normality assumptions. All final models included group, time and their interaction even if they were not retained by the stepwise selection methods. Time was divided into covariates – Time and Time3, where Time denoted an effect of time in general, and Time3 the effect of follow-ups (2^{nd} and 3^{rd} time points).

Statistical analyses for the continuous variables were performed using the statistical program R.[155] For the ordinal outcome we used the GLIMMIX procedure within the SAS program.

Role of the funding source

The trial was funded by a small exploratory grant from the Scoliosis Research Society, Glenrose Rehabilitation Hospital Foundation, and the Sick Kids Foundation and CIHR Institute of Human Development, Child and Youth Health. The first author was supported by a scholarship from the Faculties of Rehabilitation Medicine and Medicine and Dentistry at the University of Alberta.

The sponsors of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. SS and EP had full access to all the data in the study and had final responsibility for the decision to submit for publication.

5.4 Results

Treatment groups did not differ at baseline for gender, age, height and Risser sign (Table 1). Controls had higher mean weight by 4.4 kg than the experimental group. The mean age was 13.4 years (SD=1.6) (Table 5.1). The mean largest curve was 28.5° (SD=8.8°), and the mean Risser was 1.60 (Table 5.1).

Curve types based on the Schroth classification were as follows: 3c (n=7) affecting the thoracic spine without pelvis imbalance, 3cp (n=15) thoracic dominant deformity with imbalanced pelvis observed on the thoracic concave side, 4c (n=5) with thoracolumbar/lumbar dominant deformity without pelvis imbalance and 4cp (n=23) with thoracolumbar/lumbar dominant deformity with the pelvis displaced to the lumbar concave side. The number of patients with each of the 4 curve types was balanced between groups with no more than one subject difference for a given curve type.

Table 5.1. Baseline characteristics within each group

	Schroth exercises + Standard of care (95% Confidence interval), N=25	Standard of care (95% Confidence interval), N=25
Age (years)	13·5 (12·7-14·2)	13·3 (12·7-13·9)
Girls	23	24
Braced patients	17	17
Height (m)	1·60 (1·6-1·6)	1·60 (1·6-1·6)
Weight (kg)	45·9 (42·6-49·1)	50·5 (47·1-54·0)
Largest curve (°)	29·1 (25·4-32·8)	27·9 (24·3-31·5)
Sum of curves (°)	48·1 (39·1-57·2)	54·3 (44·9-63·6)
Risser sign (0 to 5)	1·76 (1·10 to 2·45)	1·44 (0·77 to 2·11)
Risk of progression[13] (%)	65	65

Baseline, three and six months mean scores of the outcomes muscle endurance, SRS-22r and SAQ scores in each group are presented in Table 5.2.

Table 5.2. Muscle endurance, Quality of life and Spinal Appearance Questionnaire scores by visit (baseline, three and six months) and group

		Baseline			Three months			Six months		
		Mean	**Standard deviation**	**95% confidence interval**	**Mean**	**Standard deviation**	**95% confidence interval**	**Mean**	**Standard deviation**	**95% confidence interval**
Sorensen	**Control**	112.3	65.0	85.5-139.2	121.8	66.7	91.5-152.2	125.7	63.6	98.2-153.2
	Schroth	109.6	53.1	87.7-131.5	141.7	58.8	116.3-167.1	149.2	60.3	122.5-175.9
SRS-22r:										
Function	**Control**	4.55	.39	4.38-4.72	4.56	.33	4.41-4.71	4.56	.41	4.38-4.73
	Schroth	4.60	.33	4.46-4.74	4.58	.41	4.40-4.76	4.69	.36	4.52-4.85
Pain	**Control**	4.19	.65	3.92-4.47	4.44	.53	4.21-4.68	4.36	.57	4.11-4.60
	Schroth	4.46	.50	4.26-4.67	4.60	.40	4.42-4.77	4.72	.48	4.51-4.94
Self image	**Control**	3.82	.63	3.55-4.08	3.85	.54	3.62-4.09	3.73	.58	3.48-3.98
	Schroth	3.91	.62	3.65-4.17	3.78	.56	3.54-4.02	3.91	.62	3.63-4.20
Total	**Control**	4.14	.41	3.96-4.31	4.22	.37	4.05-4.39	4.16	.43	3.97-4.34
	Schroth	4.25	.37	4.09-4.40	4.30	.39	4.14-4.48	4.40	.33	4.25-4.54
SAQ:										
General	**Control**	2.89	.94	2.49-3.28	2.85	1.09	2.38-3.33	2.91	1.05	2.46-3.37
	Schroth	2.92	.90	2.55-3.29	2.93	1.00	2.49-3.36	2.92	1.06	2.44-3.40
Curve	**Control**	2.21	.41	2.03-2.38	2.09	.73	1.77-2.40	2.17	.89	1.79-2.56
	Schroth	2.16	.47	1.96-2.36	2.41	.73	2.08-2.73	2.29	.64	1.99-2.58
Prominence	**Control**	1.71	.53	1.48-1.93	1.69	.52	1.47-1.92	1.91	.98	1.49-2.33
	Schroth	1.64	.55	1.41-1.87	2.00	.77	1.66-2.34	2.00	.63	1.71-2.29

Trunk Shift	**Control**	1.90	.51	1.68-2.11	1.85	.77	1.51-2.18	2.06	.80	1.72-2.41
	Schroth	1.90	.52	1.68-2.11	2.09	.72	1.77-2.41	2.09	.54	1.85-2.34
Waist	**Control**	2.64	1.61	1.94-3.33	2.52	1.37	1.93-3.11	2.74	1.48	2.10-3.38
	Schroth	2.75	1.67	2.06-3.44	2.88	1.53	2.22-3.55	3.19	1.53	2.49-3.89
Shoulders	**Control**	2.56	.95	2.16-2.96	2.52	.95	2.11-2.93	2.67	.91	2.28-3.07
	Schroth	2.48	1.06	2.04-2.92	2.84	1.00	2.39-3.29	2.79	.90	2.37-3.20
Chest	**Control**	2.19	1.40	1.60-2.78	2.11	1.40	1.50-2.72	2.66	1.50	1.99-3.32
	Schroth	2.14	1.58	1.49-2.79	2.43	1.60	1.74-3.13	2.62	1.61	1.88-3.35

Dropouts

Only six of 50 randomized patients dropped out (12%): four in the Schroth and two in the control group. Of those, there were four girls (one in the control and three in the Schroth group) and two boys (one per group). The largest curve of the dropouts was smaller on average (mean Cobb 23°, SD=5.27) than for the remaining sample.

Compliance

Patients who completed the study had high compliance: 85% of visits were attended and 82.5% of the home exercise program was completed. In the ITT analysis, the compliance estimate dropped to 76% of visits attended and 73% of the prescribed home program exercise completed.

Results from both the PP and the ITT analyses are presented in Tables 5.3 for the Biering-Sorensen test, 5.4 for the SRS-22r, and 5.5 for the SAQ.

Table 5.3 Linear mixed model coefficients and associated significance estimates in the per protocol and the intention-to-treat analyses for the Biering-Sorensen test. *Interaction group*time3* represents the effect of the treatment; *Group* main effect of group, *time* main effect of time, *time3* main effect of the follow-ups, *age* main effect of age, *SEQ* main effect of the SEQ score, *brace wear* main effect of the brace-wear.

	Per protocol				Intention-to-treat			
	Value	*t-value*	*DF*	*p-value*	*Value*	*t-value*	*DF*	*p-value*
Biering-Sorensen test								
Interaction group * time3	30.00	2.29	83	0.02	25.17	1.78	96	0.08
Group	-0.64	-0.29	45	0.77	-1.40	-0.09	46	0.93
Time	6.58	0.98	83	0.32	6.95	0.97	96	0.34
Time3	-2.55	-0.21	83	0.84	0.96	0.07	96	0.94
Covariates:								
Age	8.94	1.85	45	0.07	8.89	1.98	46	0.05
SEQ	-1.68	-1.60	83	0.11	-2.12	-1.97	96	0.05
Brace wear	24.35	1.42	45	0.16	21.59	1.36	46	0.18

Table 5.4 Linear mixed model coefficients and associated significance estimates in the per protocol and the intention-to-treat analyses for SRS-22 function, pain, self-image and total domains. *Interaction group*time3* represents the effect of the treatment; *Group* main effect of group, *time* main effect of time, *time3* main effect of the follow-ups, *weight* main effect tof weight, *Classification 3cp* main effect of the 3cp classification, *Classification 4c* main effect of the 4c classification, *Classification 4cp* main effect of the 4cp classification, *age* main effect of age, *brace wear* main effect of the brace-wear.

	Per protocol				Intention-to-treat			
	Value	*t-value*	*DF*	*p-value*	*Value*	*t-value*	*DF*	*p-value*
SRS-22r function[4]								
Interaction group * time3	24.54	0.69	85	0.49	-1.107	-0.03	97	0.98
Group	35.27	0.98	43	0.33	29.61	0.83	44	0.41
Time	27.16	1.42	85	0.16	13.70	0.72	97	0.47
Time3	-49.98	-1.41	85	0.16	-12.95	-0.37	97	0.71
Covariates:								
Weight	3.26	1.76	43	0.08	2.75	1.58	44	0.12
Classification 3cp	-126.60	-2.72	43	0.01	-135.22	-3.05	44	0.00
Classification 4c	-126.08	-2.10	43	0.04	-122.68	-2.14	44	0.03
Classification 4cp	-106.14	-2.38	43	0.02	-99.01	-2.37	44	0.02
SRS-22r pain[4]								
Interaction group * time3	-21.67	-0.60	83	0.55	-15.68	-0.41	97	0.68
Group	134.32	3.06	41	0.00	94.67	2.17	47	0.03
Time	28.12	1.51	83	0.13	12.73	0.65	97	0.52
Time3	21.05	0.63	83	0.53	50.00	1.40	97	0.16
Covariates:								
Age	-26.25	-1.92	41	0.06	-20.52	-1.76	47	0.08
Classification 3cp	- 91.55	-1.37	41	0.18				
Classification 4c	-24.23	-0.30	41	0.76				
Classification 4cp	-133.30	-2.12	41	0.04				
SRS-22r self-image								
Interaction group * time3	-0.15	-1.02	86	0.31	-0.15	-1.07	97	0.29
Group	0.16	1.01	46	0.32	0.13	0.83	47	0.41
Time	-0.02	-0.28	86	0.78	-0.006	-0.08	97	0.93
Time3	0.05	0.33	86	0.74	0.03	0.25	97	0.80

Covariates:								
Brace wear	0.29	1.01	46	0.06	0.31	2.19	47	0.03
SRS-22r total								
Interaction group * time3	0.05	0.70	84	0.48	-0.02	-0.28	97	0.78
Group	0.12	1.21	47	0.23	0.13	1.26	48	0.21
Time	0.03	0.85	84	0.40	-0.004	-0.10	97	0.92
Time3	-0.02	-0.39	84	0.70	-0.09	1.29	97	0.20

Table 5.5 Linear mixed model coefficients and associated significance estimates in the per protocol and the intention-to-treat analyses for SAQ domains. *Interaction group*time3* represents the effect of the treatment; *Group* main effect of group, *time* main effect of time, *time3* main effect of the follow-ups, *height* main effect of height, *brace wear* main effect of the brace-wear, *SEQ* main effect of SEQ score, *SEQ2* main effect of the SEQ scores when ≥35, *Age 10-11* main effect of the age 10 and 11, *Classification 3cp* main effect of the 3cp classification, *Classification 4c* main effect of the 4c classification, *Classification 4cp* main effect of the 4cp classification, *age* main effect of age, *I age* main effect of the quadratic function of age, *Age 13* main effect of age 13.

	Per protocol				Intention-to-treat			
	Value	*t-value*	*DF*	*p-value*	*Value*	*t-value*	*DF*	*p-value*
SAQ general								
Interaction group * time3	0.05	0.23	87	0.82	0.11	0.55	97	0.58
Group	0.04	0.17	45	0.86	0.01	0.03	46	0.98
Time	0.03	0.30	87	0.77	0.03	0.29	97	0.77
Time3	-0.03	-0.16	87	0.87	-0.11	-0.57	97	0.57
Covariates:								
Height	2.23	1.76	45	0.91	2.49	2.02	46	0.05
Brace wear	-0.74	-3.21	45	0.08	-0.74	-3.25	46	0.00
$(SAQ\ waist)^{-0.3}$								
Interaction group * time3	-0.003	-0.10	78	0.92	-0.03	-0.84	95	0.40
Group	-0.02	-0.58	43	0.56	-0.02	-0.59	47	0.56
Time	-0.02	-1.06	78	0.28	-0.02	-1.18	95	0.24
Time3	-0.002	-0.08	78	0.94	0.03	0.78	95	0.43
Covariates:								
SEQ	0.01	2.02	78	0.05	0.007	2.03	95	0.04
SEQ 2	0.11	3.20	78	0.00	0.10	3.18	95	0.00
Brace wear	0.08	2.08	43	0.04	0.08	2.39	47	0.02
SAQ shoulder								
Interaction group * time	0.26	1.08	86	0.28	0.41	1.75	86	0.28

Group	-0.13	-0.46	47	0.65	-0.13	-0.50	47	0.65
Time	0.08	-0.60	86	0.55	0.08	0.70	86	0.55
Time3	-0.10	-0.45	86	0.65	-0.17	-0.77	86	0.65
SAQ trunk shift								
Interaction group * time	0.27	1.80	82	0.07	0.11	0.62	97	0.53
Group	-0.08	-0.57	40	0.57	-0.01	-0.05	43	0.96
Time	0.09	1.12	82	0.27	0.15	1.52	97	0.13
Time3	-0.23	-1.48	82	0.14	-0.16	-0.84	97	0.40
Covariates:								
Age 10-11	-0.76	-3.38	40	0.00	-0.82	-3.36	43	0.00
Height	-1.73	-2.47	40	0.01	-2.01	-2.68	43	0.01
Classification 3cp	0.49	2.87	40	0.01	0.47	2.51	43	0.01
Classification 4c	0.11	0.54	40	0.59	0.10	0.42	43	0.67
Classification 4cp	0.32	2.04	40	0.05	0.35	1.99	43	0.05
log (SAQ chest)								
Interaction group * time	-0.19	-1.26	86	0.21	0.22	1.62	97	0.11
Group	0.04	0.28	41	0.78	-0.13	-0.79	42	0.43
Time	0.05	0.76	86	0.45	0.14	1.97	97	0.05
Time3	0.18	1.36	86	0.18	-0.19	-0.46	97	0.15
Covariates:								
I Age	-0.06	-2.82	41	0.01	-0.06	-2.76	42	0.01
Age 13	-0.50	-2.74	41	0.01	-0.45	-2.60	42	0.01
Brace wear	-0.40	-2.86	41	0.01	-0.45	-3.33	42	0.00
Classification 3cp	0.44	2.06	41	0.04	0.39	1.81	42	0.06
Classification 4c	0.63	2.35	41	0.02	0.58	2.23	42	0.03
Classification 4cp	0.49	2.48	41	0.02	0.41	2.10	42	0.04
$\sqrt{SAQ\ prominence}$								
Interaction group * time	0.10	1.53	86	0.13	0.12	1.87	97	0.06
Group	-0.04	-0.75	45	0.45	-0.04	-0.73	46	0.47
Time	0.03	0.88	86	0.38	0.05	1.41	97	0.16
Time3	-0.02	-0.31	86	0.76	-0.05	-0.86	97	0.41
Covariates:								
Age	0.06	3.07	45	0.00	0.06	3.07	46	0.00
Height	-0.81	-2.09	45	0.04	-0.84	-2.25	46	0.03

Biering-Sorensen test

While controlling for age, self-efficacy and brace wear, in the PP analysis, after 6-months, the Schroth group had longer hold time than the controls by 30 seconds (p=0.02). Assuming covariate values corresponding to the average characteristics in the sample, on average the hold time decreased by about 2.5 seconds in the control group, and increased by about 27.5 seconds in the Schroth group over six months.

In the ITT analysis, the adjusted mean difference between groups at six months was similar (25 seconds), but not significant (p=0.08).

SRS-22r function

To meet normality assumption the function variable was transformed to a power of four. The covariate set included weight and curve classifications. Schroth exercises did not have a significant effect on the SRS-22r function scores in both the PP (p=0.49) and the ITT analyses (p=0.98).

In the PP and ITT analysis, the curve classification had a significant effect. The best function score was observed for the 3c curve type. The differences in the function domain between patients classified as 3c vs. 3cp, 4c or 4cp were all statistically significant (p=0.009, p=0.04, and p=0.02, respectively). The 3cp and 4c curve types did not differ in function. Patients with 3cp and 4c curve types had slightly lower function than the patients with 4cp classification, but not significantly. Average function based on the model estimates across the sample for the 3c pattern was 4.84. The corresponding function for 3cp, 4c and 4cp was 4.51, 4.52 and 4.57, respectively.

SRS-22r pain

The outcome was transformed to its power of four. In both the PP and the ITT models the groups did not differ significantly after 6-months (p=0.55; p=0.68, respectively). Group had a significant main effect on the pain (p=0.004), suggesting that the Schroth group had consistently higher pain score than the controls. Patients with a 3c curve type had the best pain score, while the worst score was observed in patients with 4cp curves (p=0.04). For example, if an average patient classified as 3c had an average pain score of 4.7, then the

patients classified as 3cp, 4c and 4cp, would have pain scores of 4.46, 4.54 and 4.34, respectively.

The main effect of group, but not classification, remained significant in the ITT analysis (p=0.03).

SRS-22r self-image

In the PP and ITT the effect of Schroth was not statistically significant (p=0.31 and p=0.29, respectively). In the ITT, patients who wore brace had on average significantly better self-image by 0.31 than the ones who did not (p=0.03).

SRS-22r total

The PP and ITT analyses did not retain any covariates. The Schroth program did not significantly affect the SRS-22r total scores (p=0.48 for the PP, and p=0.78 for the ITT).

SAQ trunk shift

In both PP and the ITT analysis, the Schroth intervention did not have a significant effect on SAQ trunk shift scores (p=0.08, p=0.54, respectively). The PP model retained age, height and curve type as covariates. Analysis of the model found that that patients aged 10 and 11 behaved differently. To address this difference age was divided into covariates Age and Age 10-11.

Patients who were 10 and 11 years old had better scores on average by 0.75 points than their older counterparts (p=0.002). Taller patients also had better scores (p=0.02). On average, for every 1 cm increase in height, patients had better score on the SAQ trunk shift by 0.02. Patients with 3c curve types had statistically significant better SAQ trunk shift score compared to patients with 3cp and 4cp curve patterns (p=0.006 and p=0.047, respectively).

In the ITT analysis the same significant effects of covariates were observed (height (p=0.009), age of 10 and 11 (p=0.001) and classifications 3cp (p=0.014) and 4cp (p=0.049)).

SAQ waist

In both the PP and the ITT analysis, the Schroth intervention did not have a significant effect on SAQ waist scores (p=0.92 for PP and p=0.40 for ITT*)*. The model retained self-efficacy and brace-wear as covariates. Self-efficacy was divided into covariates SEQ (overall effect of SEQ) and SEQ2 (the effect of SEQ scores when ≥35).

The patients who had higher self-efficacy scores (stronger belief in their own ability to perform the required exercise program) also had better SAQ waist scores (p=0.04). The difference was largest once the self-efficacy score reached 35, such that with the score <35 on the self-efficacy questionnaire much worse SAQ waist results were observed compared to scores ≥35 (p=0.002). Those wearing a brace had worse SAQ waist scores than those without brace (p=0.04).

The ITT analysis produced the same significant effects for covariates.

SAQ prominence

To meet the normality assumption, the SAQ prominence was transformed to its square root. In both the PP and the ITT analysis, the Schroth intervention did not have a significant effect on SAQ prominence scores (p=0.13 for PP and p=0.06 for ITT*)*.

The PP analysis showed that taller persons had better SAQ prominence scores (p=0.04). If we assume that the SAQ prominence of a patient is 2 (lower is better) then the SAQ prominence of persons who are 10 cm and 15 cm shorter than that patient will be 2.23, and 2.36, respectively. We also found that older patients had worse scores on this outcome (p=0.004). If we assume that the SAQ prominence of a patient is 2, then the SAQ prominence of persons who are 2 and 4 years younger than that patient will be 1.67 and 1.38, respectively. In the ITT analysis, results for covariates were similar (age p=0.003 and height p=0.027).

SAQ chest

To meet the normality assumption, the SAQ chest variable was transformed using the logarithmic transformation. In both the PP and the ITT analysis, the Schroth intervention did not have a significant effect on SAQ chest scores (p=0.21 for PP and p=0.11 for ITT).

Patients who were 13 years old had statistically better scores compared to all the other ages (p=0.009) presented by the covariate Age13. Our model also determined, through a quadratic age function presented by a covariate I Age that the lowest (best) SAQ chest scores were among 13 year olds, but as the age was decreasing towards youngest or increasing towards oldest patients, the scores on the outcome were increasing, and this difference was significant (p=0.007). We also found that those who wore a brace had statistically better scores than those who did not (p=0.007). Curve classification had a significant main effect on the SAQ chest score, such that the 3c curve type had statistically better scores than the 3cp (p=0.046), 4c (p=0.02) and 4cp (p=0.02) curve types. The worst scores were observed for the 4c curve pattern.

The ITT analysis produced similar results, with significant effects of age (p=0.007), 13 year olds (p=0.011), brace wear (p=0.002) and the classifications 4c (p=0.028) and 4cp (p=0.039) when compared to the 3c curve type.

SAQ shoulders

In both PP and the ITT analysis, the Schroth intervention did not have a significant effect on SAQ shoulder scores (p=0.28 for PP, and p=0.08 for ITT). No covariates were retained in the model.

SAQ general

In both the PP and the ITT analysis, controlling for height and brace-wear, the Schroth intervention did not have a significant effect on SAQ general score (p=0.82 for PP, and p=0.58 for ITT). In the PP analysis the patients who wore a brace had significantly better SAQ general scores by 0.74 on average (p=0.002).

In the ITT analyses, in addition to the significant main effect of brace on the SAQ general score, the effect of height was also significant (p=0.046), such that taller patients had

worse SAQ general score. For every 1 cm increase in height the SAQ general score worsened by 0.02 points.

SAQ curve

There was no statistically significant effect of the Schroth program on SAQ curve score (p=0.095). The generalized mixed model selected classification and brace wear as the covariates. The 3cp classification had a significant effect on the outcome (*p*=0.02). Patients classified as 3cp were 9.2% more likely to have an SAQ curve score of 3 or worse.

5.5 Discussion

This was the first RCT investigating the effect of Schroth exercises on back muscle endurance, SRS-22r, and SAQ questionnaires' scores. We found that Schroth exercises combined with standard of care, consisting of observation and bracing, improved the back muscle endurance compared to standard of care alone in adolescents with AIS. However, the Schroth exercises did not have significant effect on the SRS-22r and SAQ questionnaires' scores.

In the only prospective study on Schroth exercises that examined the back muscle properties, strength, rather than endurance, was assessed using Lovett's manual muscle testing with scores ranging from 1-5.[73] Otman et al found that the muscle strength increased significantly after a yearlong treatment compared to the pretreatment values. In two other studies, the authors found that supervised resistive rotational exercises significantly increased strength after four months.[75,76] To our knowledge, the present study is the first to assess the change in back muscle endurance after an exercise treatment for AIS. The Schroth exercises combined with the standard of care increased the back extension hold time by about 27.5 seconds, while the hold time in the control group decreased by about 2.5 seconds. This difference of about 30 seconds was statistically significant in the per protocol analysis.

Although Schroth exercises improved endurance, such positive outcome could not be found for the SRS-22 or the SAQ outcomes. Only one other study investigating Schroth exercises (in conjunction with spinal stabilization) tested their effect on the QOL using SRS-22r, but not the SAQ questionnaire.[74] The authors, in this recent retrospective controlled clinical trial, reported better SRS-22 results at four months for both the corrective spinal technique exercise intervention (experimental) and the conventional exercise (control) groups. The experimental group demonstrated statistically significant greater benefits but only for self-image (from 3.3±1.2 to 4.2±1.0) and the total score (from 3.8±1.8 to 4.5±0.4). Monticone et al's recent RCT found positive effects of scoliosis-specific active self-correction and task-oriented exercises on changes in the Cobb angles and QOL measured by the SRS-22r questionnaire at skeletal maturity in 110 patients with AIS and curves <25° compared to standard rehabilitation exercises.[20] To our knowledge other controlled scoliosis-specific exercises studies have not monitored quality of life or perceived appearance outcomes.

The larger effect on QOL observed in these studies may be due to the higher frequency of visits in Noh et al's trial (2-3/week vs. 1/week in the present study)[74] and to the larger duration of Monticone's trial (until maturity vs. six months in our study).[20] As with these trials, we also observed a statistically significant difference in satisfaction with the treatment between the groups. At the end of the treatment the Schroth group scored on average 4.55 (SD=0.61) and the controls 3.89 (SD=0.84) in the SRS-22r satisfaction domain (p=0.005).

Patients examined in both studies[20,74] above had smaller curves than our sample but, interestingly, their scores on the SRS-22r questionnaire's domains at baseline were worse and comparable with SRS-22r scores in surgical candidates.[182] The higher baseline SRS-22r scores and the smaller follow-up duration of our trial might have limited our ability to detect changes on the SRS-22r. Other studies also demonstrated a high prevalence of ceiling effects in patients with smaller curves treated conservatively on the SRS-22r[183] and the SAQ.[184] Patients with AIS with curves <45° normally are in good health, and have high level of function.[16,61] Moreover, the SRS-22r and the SAQ questionnaires were originally designed for the surgically treated patients with AIS, who generally experience more scoliosis-related adverse symptoms that affect their QOL.

In the present study, the baseline values on the SRS-22r and SAQ questionnaires' domains demonstrated high ceiling effects across the study sample: SRS-22r (pain = 18.4%, function = 28.6%), and SAQ (prominence = 26.5%, waist = 29.2%, chest = 46.9%, trunk shift = 12.2% and shoulders = 18.4%). The percentage of patients who scored ≥4 on the SRS-22r for which the best score is 5 and ≤2 on the SAQ domains where the best scores is 1 was also very high: SRS-22 (total = 71.3%, image = 47%, pain = 77.5% and function = 100%) and SAQ (general = 18.3%, curve = 79.6%, prominence = 89.7%, waist = 48%, chest = 65.3%, trunk shift = 77.5% and shoulders = 38.8%). High scores possibly limited the ability of these questionnaires to measure improvements. This finding is consistent with the results of a recent study that investigated the responsiveness of the SRS-22r questionnaire in patients with AIS treated with braces and exercises.[185]

To our knowledge, no alternative validated and more responsive questionnaires capable of capturing improvements in patients with AIS curves <45° and treated conservatively exist. Parent et al,[186] studied the Scoliosis Quality of Life Index which had been developed for use in

younger children and found even more problems with ceiling effects than with the SRS-22r. Others have suggested using the Trunk Appearance Perception Scale (TAPS) which is completed by patients[187] or the Trunk Aesthetic Clinical Evaluation (TRACE)[188] judged by a clinician as possibly more responsive in conservatively treated patients with AIS. However they have not been routinely used. The development of a new tool specific for this group of patients with AIS may be required to monitor quality of life and perceived appearance.

Although the effect of treatment on questionnaires was not significant, some covariates had significant main effects. For the SRS-22r function domain, we found patients classified as 3c according to Schroth, which corresponds to King III or Lenke 1 or 2 classifications, had the best scores across the entire sample. The worst function was observed among patients with 3cp pattern (King IV, V, Lenke 1, 2), followed by 4c (King II, Lenke 3, 4C, 6, 1C), and 4cp (King I, Lenke 5, 6). The curve type 3c characterizes a thoracic major curve, with an imbalanced trunk, but aligned head and hips. It may be that the function in these patients remained unaffected by scoliosis because of this relatively balanced posture. Moreover, in our sample, patients classified into the 3c group had the smallest curve magnitude which also may explain their highest function score.

Similarly, for the SRS-22r pain domain, patients with 3c curve type again had the best pain score, while patients with 4cp curves had significantly worst score. The features of the 4cp curve patterns include a thoracolumbar or lumbar major curve, with or without a smaller thoracic curve, but with the pelvis displaced opposite of the lumbar convexity. The natural history of scoliosis suggests that, regardless of the curve magnitude, the thoracolumbar curves are the most likely to be associated with increased pain[17], which is consistent with our finding.

Surprisingly, we found that patients who wore a brace had better SRS-22r self-image score than the patients who did not. This is not in agreement with previous studies, which suggests that patients having been braced have significantly more distorted perceived body appearance compared to the patients who do not wear a brace and have similar curve magnitudes.[186,189] This difference may be because we report short-term (six months of treatment) rather than long-term results (16 years after maturity). However, at our institution, a cross-sectional study reported that Scoliosis Quality of Life Index self-esteem scores, a questionnaire closely related to the SRS22 self-image, were lowest for braced patients compared to other management strategies. This

would suggest that the self-image of the braced participants in the present trial might be less affected by scoliosis than others.

While the effect of the Schroth program on SAQ scores was not significant, as with SRS-22r function and pain, classification was an important covariate for the SAQ trunk shift, chest scores, and curve. This is consistent with the curve types definitions, which focus on pelvis position (opposite to trunk shift) and whether the thoracic spine (chest area) is most affected by the scoliosis or not. For all three scores, the 3c curves had the best perceived appearance compared to patients with other classifications. The SAQ chest domain is based on items focused on wanting more even breasts and more even chest in the front. In our sample, the patients with 3c curve types had smallest curves, thus also possibly explaining better scores on the SAQ chest domain. For the SAQ curve score, the 3cp curve types were more likely to have worse SAQ curve score than other groups. The 3cp curve pattern consists of a larger thoracic major curve accompanied by a pelvis deviated opposite of the thoracic convexity. In our sample, patients with the 3cp classification had largest curves, which might explain the propensity toward worse scores on the SAQ curve domain.

Three SAQ scores were affected by age: trunk shift, chest and prominence. The younger patients (≤13 years old) generally had better scores on these outcomes. However, the effect of age was not linear for those SAQ domains. The 10-11 years olds had better SAQ trunk shift scores than others, while 13 year olds had better SAQ chest scores. Younger patients might not be yet sensitized to the perception of their posture at such an early age. Or, as a recent study suggests, adolescents may have difficulties in understanding the questions and drawings used in the SAQ questionnaire.[190]

Three SAQ scores were affected by height: general, prominence, and trunk shift. Taller patients had better SAQ trunk shift, and prominence scores. In our sample, taller patients had smaller curves on average compared to others, thus likely better SAQ trunk shift and prominence scores. The SAQ prominence domain is focused on representing the perception of the rib and lumbar prominences. Patients with smaller curves have smaller rib and lumbar prominences due to smaller vertebral rotations. In contrast, SAQ general scores were worse in taller patients, but only significantly in the ITT analysis. In contrast to the other SAQ domains above, the SAQ

general score are based on items with no reference to a specific anatomical body part, but rather represent a patient's general appearance expectations.

Contrary to the findings related to the SRS-22r self-image domain, we found that patients who wore a brace had statistically worse SAQ waist outcome than the persons who did not wear a brace. This discrepancy may be due to the fact that the SAQ waist assesses the appearance only of the waist, while the SRS-22r self-image gives an overall score of the appearance of the entire body. We also found that the patients who wore brace had higher likelihood of having a worse score on the SAQ curve domain. Bracing is prescribed once the curve reaches a certain severity criteria; it was therefore expected to find worse SAQ curve scores in braced patients. In contrast, those who wore a brace had better SAQ chest scores. Similarly, patients who wore a brace had statistically better scores on the SAQ general domain. Our finding that wearing a brace influenced our models for some outcomes does not mean that bracing has an effect on the outcomes over time. Merely, this alerts us to the fact that patients meeting criteria for wearing a brace and wearing one during the study present SAQ scores that differ from patients not wearing a brace. This observation applies to patients braced in both the control and the Schroth groups. The present trial did not randomize patients meeting criteria to be prescribed a brace to a no-brace group, which would be required to conclude about the effectiveness of bracing.

The best SAQ waist scores were observed among patients who had higher self-efficacy scores. More specifically, the ones who scored ≥35 (out of 40) on the self-efficacy questionnaire had better SAQ waist results. This indicates that the children with stronger belief in their own ability to adequately perform the required exercise program would score better. Patients, whose waist is misaligned, due to a pelvic displacement, generally observe the obvious immediate effect of the correcting their pelvis misalignment. In contrast, patients whose waist and pelvis are not affected as much by scoliosis may not feel they have the ability to correct their deformity because it is most obvious from the back. Given the high prevalence of patients with 3cp (15) and 4cp (23) curve patterns (pelvis is displaced opposite of the largest curve) in both groups, the better score on the SAQ waist could be explained by the higher self-efficacy score, since these patients might also have had a stronger belief in their ability to affect their appearance using exercises.

This was a RCT, presenting strengths in studying the effect of Schroth exercises on the selected outcomes in patients with AIS. The assessors and the statistician were blinded to the treatment

allocation. As in most clinical trials involving exercises, the therapists and the patients could not be blinded. The compliance was monitored using patient/parent/therapist logbooks, which is novel in exercise studies for scoliosis. The completers' and the overall compliance rates were high. The attrition rate was low (12%), suggesting the current study protocol is feasible. We made efforts to standardize the treatment by developing the classification and exercise prescription algorithms. None of our participants reported using co-interventions.

There are also limitations. The SRS-22r is the most frequently used questionnaire assessing the QOL in patients with AIS after a treatment. The SAQ is increasingly used in the same population. Our results suggested that, due to a high ceiling effects, and high rate of scores close to the best values, in both these questionnaires, perhaps different QOL tools should be used in patients with AIS treated conservatively. The study design, does not allow determining if exercises could replace bracing. To answer such a question, our study would have to randomize patients meeting the brace prescription criteria into an exercise only or a brace only group. Not offering a brace treatment to the patients meeting criteria is an ethical concern.[99] The primary goal was to determine the effect of the Schroth exercises as an add-on to the standard of care, and not as a stand-alone therapy.

In summary, the Schroth exercises in conjunction with the standard of care improved back muscle endurance in patients with AIS over a six-months treatment period, compared to standard of care alone. It is possible that the Schroth intervention did not have a significant added benefit to the standard of care on scoliosis-related QOL and perceived spinal appearance. However, given the high prevalence of ceiling effects on SRS-22r and SAQ outcomes, and clustering of scores near the best status on both questionnaires, we argue that these two questionnaires may be inadequate to reliably capture improvements in conservatively treated patients with AIS and with curves $\leq 45°$. Perhaps different scoliosis-specific QOL instruments would be more adequate in this population. QOL has not been routinely documented in patients with AIS treated conservatively with brace or exercises. We recommend that QOL be included as standard outcome for patients with AIS, not only because the curve magnitude does not correlate well with the QOL,[191] especially in smaller curves,[192] but also because QOL is an important outcome for patients[177] and clinicians.

CHAPTER 6

Clinical significance of the effect of a six-month Schroth exercise intervention in adolescents with idiopathic scoliosis[4]

6.1 Summary

Background Appraising clinical significance of outcomes is important, but it has not been standard in research on conservative treatments for adolescents with idiopathic scoliosis (AIS).

Objective To determine the clinical significance of the effects of Schroth exercises in children with AIS.

Design Data obtained from a randomized controlled trial (RCT) was used to estimate anchor- and distribution-based thresholds to determine the clinical significance of the targeted outcomes.

Patients Fifty children with AIS, aged 10-18, with curves ranging from 10°-45°, with or without brace, and all maturity levels were included.

Intervention A supervised Schroth exercise intervention added to standard care (bracing or observation), was compared to standard care for six months.

Measurements Anchor-based minimally clinically important differences (MCID) and distribution-based minimal detectable difference (MDD_{95}), standard error of measurement (SEM) and effect sizes (ES) have been determined for the following outcomes: Largest Cobb angle, combined Cobb, Biering-Sorensen test, SRS-22r and Spinal Appearance Questionnaire (SAQ) domains. Numbers needed to treat (NNT) and proportions of patients improved, stable and deteriorated are also reported.

Results All, but one of the questionnaires' anchor-based cut-offs were below the respective SEM, while most of their MDD_{95} produced values higher than commonly seen in this population. MCIDs for the radiographic outcome and Biering-Sorensen test were larger than their prospective SEMs, but smaller than the MDD_{95}. ES for the radiographic measures were large,

[4] Sanja Schreiber, Eric C. Parent, Douglas M. Hedden

while others ranged from small to medium, mainly favoring the Schroth group. NNT for radiographic outcomes and the SRS-22r function were small (<4). Proportions of patients improved and stable were significantly larger in the Schroth group.

Limitations Most patient-reported outcomes did not correlate with the anchor, suggesting that the change in these outcome measures did not reflect the change in the patients' Global rating of change (GRC).

Conclusions Schroth exercises had clinically significant effects on radiographic outcomes and the SRS-22r function domain, but not on other outcomes.

6.2 Introduction

Many authors agree that reporting statistical test results should be supplemented with methods for determining clinically significant change[27-29,118] and in some health journals this has become standard.

Two categories of approaches are commonly used to determine what constitutes a clinically significant change in an outcome of interest: *anchor-* and *distribution-based methods.* Anchor-based methods use an external indicator (anchor) to assign subjects into several groups reflecting the perceived importance/magnitude of their changes in a clinical outcome.[193] A health outcome is selected as a target tool and the change in this target is then linked to the perceived change using the anchor[34,36]. Anchor-based approaches must satisfy two basic requirements: 1) the anchor must be interpretable by the end users, and 2) the outcome measure (target) and the anchor must be appreciably correlated, because the change in the target must reflect the change in the anchor to suggest that it can be used as an interpretable tool.[34] Ideally, the anchor should correlate ($r \geq 0.3$) with the change score observed in the target outcome[32,34,36,194]. Researchers also often construct receiver operating characteristics (ROC) curves to identify thresholds on the target outcome that presents the best balance in identifying patients truly improved (sensitivity) and truly not improved or deteriorated (specificity) based the anchor [195].

Distribution-based methods, on the other hand, are determined by the statistical properties of the tool used to measure some outcome of interest and may include effect sizes (ES) [36,120-122], standard error of measurement (SEM)[123], minimal detectable difference (MDD)[195,196] or the

reliable change index[27,124]. These methods rely on statistical properties to quantify how much change in a measure is deemed clinically important. The rationale for using distribution-based methods is based on requirement that a change must exceed the magnitude of measurement errors documented on a target tool before considering that clinically important change has occurred.

Bago et al[39] and Carreon et al[40] made first efforts in determining the clinical significance of changes in scoliosis-related quality-of-life (QOL). Bago et al. used both anchor- and distribution-based methods to determine the clinical cutoffs on the SRS-22r questionnaire in surgically treated adolescents with idiopathic scoliosis with a mean follow-up of 45.6 months. The authors recommended using the distribution-based cutoff when the aim is to analyze each of the scales of the SRS-22, because anchor-based methods estimated the MCID below the measurement error on some subscales. On the other hand, if the focus is the global patients' perception of their quality-of-life, the authors suggested using only the anchor-based MCID estimates of the SRS-22 sum score (MCID=0.6) because it exceeded the SEM. The anchor in their study corresponded to the highest level on the 4-point Likert scale of GRC, which represented a "Much better" improvement. In this surgically treated group, the cutoff based on the SEM distribution-based method for each of the domains were as follows: Pain=0.6, Function=0.8, Image=0.5, Mental Health=0.4, Average-Sum=0.5. Carreon and colleagues[40], also in surgically treated patients, using ROC curves, determined the MCID for the pain, appearance and function domains in the SRS-22 questionnaire. The patients had filled out SRS-22r before and SRS-30 and SAQ questionnaires one year after surgery. To construct anchors for pain and function they summed the relevant questions from the last eight questions on the SRS-30 questionnaire. The anchor for the appearance domain was constructed by summing the relevant questions from the SRS-30 with two questions from the Spinal Appearance Questionnaire (SAQ). The anchors were used to construct ROC curves for each domain separately, and obtain the following MCIDs: Pain=0.20 (area under the curve [AUC] 0.72, CI=0.70-0.77), Activity=0.08 (AUC 0.65, CI=0.60-0.69), and Appearance=0.98 (AUC 0.629, CI=0.60-0.68). The authors also calculated the SEM and MDD_{90}, which were as follows: Image (SEM=0.21, MDD_{90}=0.47), Activity (SEM=0.17, MDD_{90}=0.41) and Pain (SEM=0.15, MDD_{90}=0.33)

Most recently, Rushton and Grevitt conducted two literature reviews. The first one[125]

investigated the clinical significance of the differences between scores reported using SRS-22r questionnaires in untreated groups of patients with AIS and published normative data on adolescents without scoliosis. Differences in the mean SRS-22r domain scores for each AIS cohort compared with published normative data for unaffected Western adolescents were assessed using 95% confidence intervals (CI).[197] To assess whether any difference was clinically significant, the 95% CI was compared with an estimate of clinically significant change for that domain defined by a distribution-based method and calculated as $1.96 \times SEM \times \sqrt{2}$. If the lower bound of the CI for the difference in means was greater than the threshold estimated from the normative data it was considered clinically significant. The authors found that the pain (81% of cohorts) and self-image (91% of cohorts) domain scores were statistically significantly lower among cohorts with AIS than those unaffected. However, when comparing to Bago et al's[39] MCID estimates only the observed differences in self-image scores were consistently clinically significant (73% of cohorts). Cohorts with AIS typically scored well in function and mental health domains and differences compared to unaffected subjects were rarely clinically significant. While this review suggests that some patients with AIS with indications for conservative treatments have clinically significant self-image and rarely pain deficits compared to controls, it is still unknown whether the effect of conservative treatments is clinically important.

The second review used published research on cohorts of surgically treated adolescents with idiopathic scoliosis to calculate 95% confidence intervals for the changes reported from pre-op to 2 years post-op and compared these values to Bago et al's[39] published MCID estimates for pain, image, function and mental health domains to determine whether an observed change was clinically significant. They found that 81% of the surgically treated cohorts experienced statistically significant improvements in pain and 94% in self-image domains. In contrast, only the changes in self-image were found consistently clinically significant using Bago's clinical significance thresholds determined by the distribution-based methods (4 out of 5, 80% of cohorts). Changes in the pain domain were found to be clinically significant in only 1 out of 12 examined study cohorts (8%). No clinically significant changes were observed in the mental health or the activity domains.

These authors reported the clinical importance of change in SRS-22r domains in patients who

have undergone the scoliosis surgery. To our knowledge no research has been done to assess the clinical importance of effects of conservative treatments for scoliosis. Bago et al reported MCID and MDD_{95} for all SRS-22 domains, while Carreon et al reported the MCIDs for SRS-22r questionnaire's pain, function and image domains. The MCID or the MDD_{95} of the SAQ domains, radiographic outcomes and in the Biering-Sorensen test of back endurance have not been established. The objectives of this study are: 1) to determine the anchor-based and distribution-based estimates of clinically significant change thresholds for radiographic outcomes, the Biering-Sorensen back muscle endurance test, as well as the SRS-22r and SAQ questionnaire domains; 2) to use these estimates to determine the clinical significance of the effect of a 6-month Schroth exercise program in patients with adolescent idiopathic scoliosis.

6.3 Methods

Design Overview

This is a measurement study using data from the Schroth Exercises RCT for Scoliosis (ClinicalTrials.gov NCT01610908) and applying anchor- and distribution-based methods to determine the clinical significance of the treatment effects over six months documented using radiographic, back endurance, SRS-22r and SAQ questionnaires measurements.

Setting and Participants

Consecutive participants were recruited from a local scoliosis clinic. The first 50 participants with AIS from the Schroth exercise trial for scoliosis formed the sample for this study. We included adolescents with AIS (males and females), age 10-18, all curve types, with curve magnitudes between 10° - 45° according to the Cobb method, treated with or without brace, and all maturity levels (Risser=0-5). Surgical candidates, adolescents who have had surgery, had completed a brace treatment and patients with forms of scoliosis other than AIS were excluded. The local ethics review board approved the study, participants provided assent and parents consent.

Randomization and Interventions

Participants were randomized using a computer-generated sequence in pre-sealed envelopes into the Schroth exercises or the control (standard of care) group, so that each group included 25 subjects.

The Schroth intervention consisted of weekly 1-hour long supervised Schroth exercises sessions combined with a 45-minute long daily home exercise program delivered over a period of six months in addition to the standard of care.[198] Controls received only standard of care (observation or bracing) during the trial time.

Outcomes and Follow-up

Radiographic measurements (Largest Cobb and Combined Cobb), Biering-Sorensen back muscle endurance time, and SRS-22r and SAQ questionnaires' domain scores were used as targets and the GRC was used as anchor.

The ***Cobb angle*** is the angle between the most tilted upper and lower end vertebrae of the scoliosis curvature observed on a posterior-anterior radiograph[10]. Our semi-automated digital measurement demonstrated excellent reliability[150] with error (SEM) better than most published values at $\leq 2.5°$ for the Cobb angle[150] and <0.3 levels for end-vertebra identification[150]. In addition, there was no difference between more and less experienced raters. This computer-aided method uses the Hough Transform to detect the orientations of the vertebral endplates from which the Cobb angle is calculated. Users simply identify the upper and lower end-vertebrae for the curve and software automatically extracts the angle[150]. The intra-class correlation coefficient for estimating the intra-rater reliability was ICC=0.99 (CI 0.987 - 0.992) and for the inter-rater reliability ICC=0.981(CI 0.977 – 0.983). The evaluator extracted all Cobb angles larger than 10^{o} while blinded to groupings, image time point, prior measurements and subject identity. The largest Cobb angle measured was used as the Largest Cobb and the sum of all Cobb angles exceeding 10^{o} was used as the Combined Cobb variable in the analyses.

The ***Scoliosis Research Society (SRS-22r) questionnaire*** is the most commonly used scoliosis-related quality of life assessment with 22 questions assessing 5 domains: function, pain, image, mental health (5 questions each), and satisfaction (2 questions).[175] Each domain score ranges from 1 to 5, with higher values indicating better outcomes[42]. The SRS-22r can detect meaningful changes in response to treatments in surgically treated patients,[39,199] has adequate test-retest

reliability, measured by interclass correlation coefficient (ICC=0.85 to 0.96)[40,200] and construct validity with good correlations with SF-36, SF-12, and Oswestry scores.[201] However, to our knowledge, the responsiveness of the SRS-22r questionnaire in conservatively treated patients with AIS has not been reported.

The Spinal Appearance Questionnaire (SAQ) measures the patients' perception of their spinal deformity's appearance using standardized drawings and questions[178]. The SAQ consists of 20 questions with responses ranging from 1-5, and it addresses nine domains: general, curve, prominence, trunk shift, waist, shoulders, kyphosis, chest and surgical scar. Domains are calculated as means of the corresponding items, with lesser values indicating a better result. In this study, kyphosis and surgical scar domains were not used, because our sample is made of patients with scoliosis without surgical treatment. We calculated the test-retest reliability of the SAQ domains in 34 subjects who completed the questionnaire as part of a routine visit to the scoliosis clinic and again within two weeks as part of an evaluation in the trial. The test-retest reliability estimates ($ICC_{3,1}$) were as follows: General 0.74 (95%CI 0.47 to 0.88), Prominence 0.77 (95%CI 0.53-0.90), Curve 0.34 (95%CI 0.00 to 0.65), Trunk Shift 0.68 (95%CI 0.38 to 0.85), Waist 0.84 (95%CI 0.66 to 0.93), Shoulders 0.80 (95%CI 0.56 to 0.92), and Chest 0.72 (95%CI 0.42 to 0.88). The SAQ has not been tested for its responsiveness in patients conservatively treated for AIS.

The ***Biering-Sorensen test*** is a validated test assessing the isometric endurance of the trunk extensor muscles, most commonly used in people with low back pain[173,202,203]. This test measures how many seconds a subject is able to hold the unsupported upper body horizontal, while having the legs and buttocks fixed to a table, with the upper body hanging over the edge of the table and the arms crossed over the chest. The test is stopped when a subject can no longer control the posture or when 240 seconds have been reached[172]. The test-retest reliability was shown to be adequate ICC=0.85(CI 0.76-0.90) with a SEM=15.6 seconds[173]. The Biering-Sorensen test has also been routinely used in cohorts without low back pain[174,204], while in people with scoliosis a modified Biering-Sorensen test has been used[63,64]. We used the original Biering-Sorensen test described above.

The ***Global Rating of Change (GRC)***[34,205,206] was used as the anchor. The following GRC was administered at the 3 and 6 months follow-ups: "Please rate the overall condition of your back

from the time you began the treatment until now". The scale ranges from –7 (a very great deal worse) to +7 (a very great deal better). The test-retest reliability is high, ICC=0.90[207], as well as its face validity based on patient-rated importance of change, Pearson's r = 0.72 – 0.90[207]. For construct validity evidence, it reached significant correlation with change on Roland Morris, Oswestry, Pain rating, Euroqol, asthma quality of life and hop test[207].

Statistical Analysis

Anchor-based methods

1. For determining the MCID using the anchor-based approach, we first assessed the correlation between each of the targets (Largest Cobb, Combined Cobb, Biering-Sorensen test, and SAQ and SRS-22r domains) and the anchor (GRC). Then, mean changes on each target were estimated for the subgroups of patients who rated their change on the GRC as follows: deteriorated (-2 to -7), no change (-1, 0, +1) and improved (+2 to +7). On the GRC scale -1 corresponds to "A tiny bit worse (almost the same)", 0 to "About the same" and +1 to "A tiny bit better (almost the same)". Consistent with Cella et al work[194], we considered these GRC ratings to represent an insignificant change.

2. ROC curves were constructed for each target separately. ROC curves plot the sensitivity (true positive rate) against one minus the specificity (false positive rate) for each possible change in outcome to identify the ability of different amount of change on the target outcome to discriminate between patients who felt improved or not, as defined by the GRC. The amount of change in the target outcome with the best balance between sensitivity and 1 minus specificity is identified (typically the closest to the top left corner of the ROC graph) and represents the MCID. The area under the ROC curve represents the ability of the target outcome to discriminate between improved and unchanged/deteriorated patients and allows comparison between outcomes.

Distribution-based methods

Using distribution-based methods, we estimated the MDD_{95}, which is based on the SEM, in addition to the ES for each outcome.

1. The *SEM* is an indicator of the spread of the distribution of measurement errors in repeated measurements.[195] The formula for SEM is: $SEM=SD\times\sqrt{1-r}$, where *SD* is the standard deviation of the score before the treatment and *r* is a reliability estimate for the measurement obtained from previous studies. For the Cobb angle outcomes we used intra-rater test-retest ICC=0.98[150]. For SRS-22r pain, self-image, function and total scores we used test-retest ICC=0.96[200], 0.90[200], 0.90[200], 0.89[208], respectively. For the SAQ domains, we used our calculated test-retest ICC estimates reported above, and for the Biering-Sorensen a test-retest ICC of 0.85[173].

2. MDD_{95} represents the minimal detectable difference in the score using a 95% confidence interval, and was calculated as: $MDD_{95} = z \times SEM \times \sqrt{2}$ [209], where *z* is the z-score corresponding to the desired confidence interval (CI) in a normal distribution, set at 1.96 for a 95%CI, using α=0.05.

3. *ES* were calculated using Cohen's *d* formula: $Cohen's\ d = \frac{M1-M2}{SD_{pooled}}$, where $SD_{pooled} = \sqrt{\frac{SD1^2+SD2^2}{2}}$. M1 and M2 represent the mean change observed in the interval of interest in the experimental and control group, respectively. SD_1 and SD_2 are the standard deviations of the changes between baseline and follow-up observed in the experimental and control group, respectively. ES=0.2-0.5 were considered small, ES=0.5-0.8 medium, and ES>0.8 large[210].

In addition, the percentage of patients who improved, stayed unchanged and deteriorated were also estimated for each of the subgroups based on whether the individual patient's change in the target outcome exceeded the adopted threshold defining clinically important change for that outcome. The distribution of patients among each improvement subgroup was compared between the Schroth and the control group using a Chi-square test.

To determine clinical significance of the selected outcomes, some authors recommend using multiple anchor- and distribution-based methods and then examining them to identify a single value or range of values for what is deemed to be a clinically significant threshold, a process known as triangulation[32,33,193]. In the present study, we used two anchor-based (mean difference and ROC curve) and three distribution-based methods (SEM, MDD_{95} and Cohen's d) to infer the

clinical significance of the outcomes. Because it is expected that the cutoffs determined by different methods will result in a range of values, to determine the recommended clinical significance threshold, we triangulated the results to converge the estimates to a single value. Although, the anchor-based methods provide important information about the value of change in a patient-reported outcome, when a cutoff determined by a distribution-based methods yields a larger value than a cutoff determined by an anchor-based method, it cannot be guaranteed that the observed change is not simply due to a measurement error. In such cases, the recommended clinical significance threshold was a distribution-based estimate (MDD_{95} or SEM). In cases where the MDD_{95} were much larger than the SEM estimates, such that for most measurements if adopting MDD_{95} it would be unrealistic to ever achieve clinically significant changes in comparison with previous studies showing significant effects using the target outcomes, the SEM was adopted.

Numbers needed to treat (NTT) were calculated as: $NNT = \frac{1}{ARR}$, where ARR is an absolute risk reduction, calculated as the difference between the control event rate (CER) and the experimental event rate (EER). The CER and EER corresponded to the proportion of patients in the control and the experimental group who deteriorated by an amount exceeding the clinical significance threshold. When reporting NNT it is important to include the confidence intervals to adequately interpret the results[211,212]. Wald's method for calculation of the CI is the most frequently used. Despite its general usage, the Wald method has several documented shortcomings, including dependency on sample size and producing unreliable theoretically impossible results when probabilities of events are close to 0 and 1.[211,213] When a treatment effect is not significant, the CI for the NNT calculated by Wald method will include a negative number, suggesting that the treatment has a harmful effect, which is known as number needed to harm (NNH). For example, NNT=6, CI -14.3 to 2.6 includes a negative number and does not seem to provide a good estimate of 6. That is because the continuity of the CI has been violated and should be represented using two disjoint regions (NNT=6, CI -14.4 to -∞, 2.6 to +∞; or NNT 3 to ∞ and NNH 15 to ∞). Although, in this example the NNT is relatively low, the CIs are disjointed suggesting that the NNT should be interpreted cautiously. Therefore, we estimated the 95% confidence intervals using the Wilson score method, which provides improved CI and interpretation compared to the Wald method. Tandberg's calculator for confidence intervals for

the NNT was used[214].

Role of the Funding Source

This study was supported by a Scoliosis Research Society small exploratory grant and by a Glenrose Rehabilitation Hospital Foundation clinical research grant. Sanja Schreiber was supported by the Interfaculty graduate studentship awarded jointly by the Faculty of Rehabilitation Medicine and Faculty of Medicine and Dentistry at University of Alberta. The outcomes or the reporting of this investigation were not influenced by the funding sources.

6.4 Results

Sample description

Six of the 50 randomized patients dropped out (12%), 4 in the Schroth and 2 controls. The mean age was 13.4 years (SD=1.6) (Table 6.1). The mean largest curve was 28.5° (SD=8.8^{o}) and the mean of the combined curves was 51.2° (SD=22.3^{o}) (Table 6.1). The average height and weight are reported in *Table 1*. Curves types were determined based on the Schroth classification as follows: 3c (n=7) affecting thoracic spine without pelvis imbalance, 3cp (n=15) thoracic dominant with imbalanced pelvis observed on the thoracic concave side, 4c (n=5) with thoracolumbar/lumbar dominant curves without pelvis imbalance and 4cp (n=23) with thoracolumbar/lumbar dominant curves with pelvis displaced to the lumbar concave side. Using the anatomical curve classification, four patients had proximal thoracic, thoracic and thoracolumbar curves, 10 had proximal thoracic, thoracic and thoracolumbar/lumbar curves, five had double thoracic curves, eight had thoracic and thoracolumbar/lumbar curves, 12 had thoracic and thoracolumbar curves, nine had a single thoracolumbar curve, one had a single thoracolumbar/lumbar curve and one had a proximal thoracic and a thoracolumbar curves.

Table 6.1. Summary of the descriptive statistics

Variables	Group	N	Mean	SD	95% CI	Min	Max
Combined Cobb (°)	**Control**	25	54.3	22.6	44.9-63.6	11.7	95.1
	Schroth	25	48.1	21.9	39.1-57.2	11.3	86.0
Major Cobb (°)	**Control**	25	27.9	8.8	24.3-31.5	11.7	42.0
	Schroth	25	29.1	8.9	25.4-32.8	11.3	44.3
Age (years)	**Control**	25	13.3	1.52	12.7-13.9	10	16
	Schroth	25	13.5	1.76	12.7-14.2	10	17
Height (m)	**Control**	25	1.60	0.07	1.57-1.63	1.46	1.75
	Schroth	25	1.60	0.10	1.56-1.64	1.38	1.80
Weight (kg)	**Control**	25	50.5	8.27	47.1-54.0	34.60	71.10
	Schroth	25	45.9	7.91	42.6-49.1	29.30	64.90

Anchor-based estimates of clinical significance

The ***mean differences*** calculated as post-treatment minus baseline outcome for the patients who perceived their change on the GRC scale as "a little bit better" to "a great deal better" (+2 to +7) are reported in *Table 6.2* for each target outcome.

The AUC of the ***ROC curves***, to identify patients deemed improved using the GRC (≥2) as anchor are reported in *Table 6.2*. Only the MCID for the Largest Cobb and Combined Cobb were estimated because they were the only outcomes for which the correlation with the GRC met the recommended threshold r≥0.3. The highest correlations between the targets and the anchor were found for the Largest Cobb ($r= –0.34$, $p<0.05$), Combined Cobb ($r= –0.35$, $p<0.05$). All the ROC-based estimates were smaller than their corresponding mean difference estimates.

Distribution-based estimates of clinical significance

The ***MDD_{95}*** and the ***Cohen's d*** effect sizes are reported in *Table 6.2*. Estimated MDD_{95} were from 1.7 to 30 times larger than the corresponding MCIDs corresponding to the means of the patients who perceived themselves as improved. The effect sizes for the Cobb measurements were large favoring the Schroth group. The effect size determined for the Sorensen test was medium also favoring the Schroth group, while effect sizes for the remaining outcomes were small, all favoring the Schroth group except the SAQ Curve, Prominence and Waist.

Both anchor- and distribution-based cutoffs are reported as percent of baseline values in *Appendix 5* to help the reader judge their magnitude.

Clinical significance cutoffs for the outcomes

Largest Cobb and Combined Cobb. The MDD_{95} was adopted as clinical significance threshold for these outcomes. Anchor-based methods produced values that were much smaller than the distribution-based methods for Cobb measurements. The distribution-based methods cutoffs were consistently more stringent, with MDD_{95} for the Largest Cobb (3.42°) and Combined Cobb (8.68°) being about 3 times larger than the estimated anchor-based cutoffs.

Table 6.2. Mean change in subjects who felt improved based on reporting a GRC (≥2), MCID estimates from ROC curves with the areas under the curve (AUC), standard error of measurement (SEM), minimum detectable change at a 95% confidence interval (MDD_{95}), and Cohen's d effect sizes (ES) for each of the target outcomes. *Indicates effect sizes favoring the Schroth exercises group. N/A refers to a ROC curve cutoff that was not determined due to correlations <0.3 between the target and the GRC.

	Mean change (GRC =+2 to +7)	MCID from ROC (AUC, 95% CI)	SEM	MDD_{95}	Cohen's d
Largest Cobb (°)	-1.33	-0.9 ° (0.69, 0.52-0.86)	1.24 °	3.42 °	-0.92*
Combined Cobb (°)	-2.74	-1.35 ° (0.73, 0.57-0.88)	3.15 °	8.68 °	-0.77*
SRS-22r:					
Pain /5	0.16	N/A	0.12	0.32	0.10*
Image /5	-0.04	N/A	0.20	0.54	0.20*
Function/5	-0.04	N/A	0.11	0.31	0.25*
Total /5	0.07	N/A	0.13	0.36	0.43*
SAQ:					
General /5	-0.05	N/A	0.41	1.12	-0.10*
Curve /5	0.05	N/A	0.32	0.90	0.20
Prominence /5	0.25	N/A	0.29	0.79	0.10
Trunk shift /5	-0.05	N/A	0.34	0.94	-0.15*
Waist /5	0.12	N/A	0.60	1.65	0.05
Shoulders /5	0.05	N/A	0.45	1.23	-0.01*
Chest /5	0.15	N/A	0.79	2.19	-0.04*
Biering-Sorensen (Sec)	36.15	N/A	22.77	62.79	0.54*

SRS-22r and SAQ domains. All, except one anchor-based cutoff (SRS-22r Pain) were below the SEM of the instrument, while the MDD_{95} for the questionnaires were much larger than the SEM estimates. The SEM was adopted as an estimate of a clinically important threshold for the SRS22-r and SAQ domains.

Biering-Sorensen test. The adopted clinical significance threshold estimate for the Sorensen test was the anchor-based mean change threshold (36.15 sec) because it was larger than the SEM and the MDD_{95} estimates would correspond to changes exceeding the magnitude reported in several studies on the effect of back endurance specific therapy[215-217].

Clinical significance of the effect of the Schroth Exercise program

Using the clinical significance threshold adopted for each outcome, the percentage of improved, deteriorated and stable patients was calculated in each group (Table 6.3) Differences in these three proportions between treatment groups were not statistically significant for all scores except the Largest Cobb angle and the SAQ curve domain where fewer deteriorations and/or more improvements were observed in the Schroth group. For the subjective outcomes (SRS-22r and SAQ questionnaires), the difference in percent of improved patients between both groups was not large. Although differences (3% to 32%) did not consistently reach statistical significance, the percent of deteriorated patients in the control group was higher on all outcomes compared to the Schroth group except for the SAQ prominence domain.

For the Largest Cobb and the Combined Cobb, the NNT were low (3 and 4, respectively) with large corresponding ARR of 38% and 26% (Table 6.4). In addition, the NNT for the SRS-22r function was 4 with an ARR of 25%. The Chi-square test comparing the proportion of deteriorated patients vs. not (combining improved and stable) in each group revealed that significantly more patients in the Schroth group avoided deterioration on these three outcomes over six months than in the control group.

Table 6.3. The percentage of improved, deteriorated and stable patients in each group based on the corresponding clinical significance cutoff: MDD_{95} for Largest Cobb and Combined Cobb, SEM for the SRS-22r and SAQ domains, and MCID cutoff for the Biering-Sorensen test.

Outcomes	Clinical Sig. Cutoff	Schroth			Control			
		Impr. (%)	Deter. (%)	Stab. (%)	Impr. (%)	Deter. (%)	Stab. (%)	Chi^2 (p)
Largest Cobb (°)	**3.42°**	6 (26.1)	3 (13.0)	14(60.9)	1 (5.0)	9 (45.0)	10 (50)	7.06 (0.03)
Combined Cobb (°)	**8.68**	2 (8.7)	2 (8.7)	19 (82.6)	1 (5)	7 (35)	12 (60)	4.05 (0.10)
SRS-22r/5								
Pain	**0.12**	13 (61.9)	3 (14.3)	5 (23.8)	11 (47.8)	4 (17.4)	8 (34.5)	0.91 (0.63)
Image	**0.20**	6 (28.6)	8 (38.1)	7 (33.3)	6 (26.1)	10 (43.5)	7 (30.4)	0.13 (0.94)
Function	**0.11**	8 (38.1)	2 (9.5)	11 (52.4)	8 (34.8)	8 (34.8)	7 (30.4)	4.41 (0.11)
Total	**0.13**	12 (57.2)	2 (9.5)	7 (33.3)	10 (43.5)	6 (26.1)	7 (30.4)	2.09 (0.35)
SAQ/5								
General	**0.42**	4 (19)	3 (14.3)	14 (66.7)	5 (21.7)	7 (30.4)	11 (47.8)	1.98 (0.37)
Curve	**0.31**	0 (0)	2 (9.5)	19 (90.5)	6 (26.1)	4 (17.4)	13 (56.5)	7.72 (0.02)
Prominence	**0.19**	0 (0)	9 (42.9)	12 (57.1)	5 (21.7)	6 (26.1)	12 (52.2)	5.52 (0.06)
Trunk shift	**0.28**	6 (28.6)	7 (33.3)	8 (38.1)	7 (30.4)	9 (39.2)	7 (30.4)	0.30 (0.86)
Waist	**0.46**	3 (14.3)	8 (38.1)	10 (47.6)	4 (18.2)	8 (36.4)	10 (45.5)	0.07 (0.97)
Shoulders	**0.33**	6 (28.6)	8 (38.1)	7 (33.3)	7 (30.4)	12 (52.2)	4 (17.4)	1.03 (0.60)
Chest	**0.59**	2 (9.5)	5 (23.8)	14 (66.7)	4 (18.2)	8 (36.4)	10 (45.5)	2.00 (0.37)
Biering-Sorensen (sec)	**36.15**	12 (54.5)	1 (4.5)	9 (40.9)	6 (26.1)	3 (13)	14 (60.9)	4.07 (0.13)

Table 6.4. Number needed to treat (NNT) with 95% CI, absolute risk reduction (ARR) with 95% CI and the Chi-square and its p-value for comparing the distribution of deterioration between groups. * denotes that the continuity of the NNT confidence interval has been violated because the ARR CI includes 0, and should be interpreted as "NNT(Harm) to -∞ and NNT (Benefit) to ∞".

Outcomes	Clinical Sig. Cutoff	NNT (95% CI)	ARR % (95% CI)	Chi-square (*p*)
Largest Cobb (°)	**3.42**	3 (1.8-20.4)	32 (0.05 to 0.54)	5.4 (<0.05)**
Combined Cobb (°)	**8.68**	4 (2.0-64.0)	26 (0.02 to 0.49)	4.47 (<0.05)**
SRS-22r				
Pain/5	**0.12**	33 (-5 to -∞, 4 to ∞)*	3 (-0.20 to 0.25)	0.08 (0.78)
Image/5	**0.20**	19 (-5.54 to -∞, 3.12 to ∞)*	5 (-0.22 to 0.32)	0.13 (0.72)
Function/5	**0.11**	4 (2.1-726.1)	25 (0 to 0.47)	3.99 (<0.05)**
Total/5	**0.13**	6 (-14.29 to -∞, 2.63 to ∞)*	17 (-0.07 to 0.38)	2.02 (0.15)
SAQ				
General/5	**0.42**	7 (-11.11 to -∞, 2.59 to ∞)*	16 (-.09 to 0.39)	1.63 (0.20)
Curve/5	**0.31**	13 (-7.14 to -∞, 3.45 to ∞)*	8 (-0.14 to 0.29)	0.58 (0.45)
Prominence/5	**0.19**	-6* (-2.44 to -∞, 9.09 to ∞)*	-17 (-0.41 to 0.11)	1.37 (0.24)
Trunk shift/5	**0.28**	18 (-4.71 to -∞, 3.23 to ∞)*	6 (-0.21 to 0.32)	0.16 (0.69)
Waist/5	**0.46**	58 (-3.50 to -∞, 3.96 to ∞)*	2 (-0.28 to 0.25)	0.07 (0.79)
Shoulders/5	**0.33**	8 (-6.94 to -∞, 2.53 to ∞)*	14 (-0.14 to 0.39)	0.76 (0.38)
Chest/5	**0.59**	8 (-6.92 to -∞, 2.70 to ∞)*	12 (-0.14 to 0.37)	0.80 (0.37)
Biering-Sorensen (sec)	**36.15**	12 (-9.09 to - ∞, 3.57 to ∞)*	8 (-0.11 to 0.28)	1.00 (0.32)

6.5 Discussion

This study provides data on the clinical significance thresholds for the radiographic measurements of scoliosis curves, SRS-22r and SAQ questionnaires' domains, as well as the Biering-Sorensen back muscle endurance test in patients with AIS who underwent conservative treatment using Schroth exercises in combination with standard of care (observation or bracing). To our knowledge, this is the first study to determine the clinical significance of the outcomes using standardized instruments for scoliosis in conservatively treated patients with AIS.

The number needed to treat is a useful way of interpreting the results of clinical trials[212]. In the present study, NNT for the Largest Cobb, the Combined Cobb, and the SRS-22r function domain were low (3-4) and the ARR were important (25% to 38%). Further, for these outcomes, the difference in the proportion of patients deemed deteriorated or not between treatment groups was also statistically significant ($p<0.05$), suggesting that the Schroth exercises combined with the standard of care led to clinically significant protection against deterioration, which occurs with the progression of scoliosis over time.

There is no consensus as to which method is the most appropriate to use to determine the clinically significant change thresholds.[30,38] Each method has its strengths and weaknesses. In a study where the goal is determining clinical significance from the perspective of a patient, it is recommended to use anchor-based methods[193]. The limitation of anchor-based methods is their dependency on the GRC and its correlation with the target. Most of our outcomes (targets) did not correlate substantially ($r<0.3$) with the anchor creating difficulty to determine valid anchor-based cutoffs. On the other hand, the distribution-based SEM relies on the reliability index of the measurement tool and may not reflect the perception of participants. Because all outcomes in this study (Cobb angle, Biering-Sorensen test, SRS-22r and SAQ) have good reliabilities, the SEM led to small clinical significance threshold estimates. An advantage of the SEM is that it is not dependent on the sample size. In contrast, the ES is largely dependent on the distribution of a sample, so if the sample is heterogeneous, the baseline SD will be large and will produce small effect size. In other words, the same amount of individual change will result in different effect sizes depending on the heterogeneity of the sample at baseline[38]. Further, even though Cohen suggested a convention on cutoff scores for effect sizes[210], they are arbitrary and the true connection to the clinical significance is not known. The general limitation of distribution-based

methods is their lack of linkage to the actual clinical measures, and it is difficult to interpret their clinical importance. While none of the methods is perfect, combining the information from multiple methods is helpful for data interpretation.

In this study, as expected, and as previously reported (see Table 6.2.) anchor- and distribution-based methods produced different cutoffs. The SEM was used as a clinical significance cutoff for the questionnaires' domains based on the fact that all except one (SRS-22r pain) of the questionnaire's anchor-based cutoffs were below the SEM. MDD_{95} were deemed too stringent because of the presence of high ceiling effects affecting questionnaires' scores. For example, the SAQ waist and chest domains had calculated MDD_{95} of 2.46 and 2.24 corresponding to nearly 50% of the full scale with ceiling effects in about 30% and 47% of our participants. Using such a large estimate would make it nearly impossible to reach the amount of change based on the MDD_{95} suggesting that this questionnaire may not be ideal to monitor response to conservative interventions. In addition, the MDD_{95} for the SRS-22r image domain was 0.54, which was slightly higher than the MDD_{95} reported in the study by Bago et al (MDD_{95}=0.50) in surgically treated patients with AIS. Surgical candidates generally have lower score on the image domain than patients with smaller curves. Thus, this threshold can be exceeded more easily after a surgical procedure that aims to correct the curve.

For the objective outcomes, including Largest Cobb and Combined Cobb, correlations with the GRC were adequate relative to the ($r>0.3$) criteria. However, the MCIDs for the Largest Cobb and Combined Cobb were very low, (1.33° and 2.74° respectively). The Schroth approach aims to improve the postural awareness. Because their postural awareness might have improved after the intervention and, consequently the ability for correcting the posture, some patients might have perceived improvement, even if the Cobb angle had not decreased. Therefore, based on our a priori criteria we adopted the stricter MDD_{95} (3.42° and 8.68°, respectively). Although the MDD_{95} for the radiographic measures were larger than the MCID, they were still smaller than the most commonly used threshold corresponding to a 5° change observed on the radiograph. Most authors have considered 5° change between two consecutive radiographs clinically important, although this is an arbitrary value based on previously reported SEMs obtained from the manual Cobb angle measurements[218]. The SEM produced by our semi-automated measurement is lower than most published values (≤2.5°)[150]. Moreover, the natural history of scoliosis suggests that the

curves increase on average at a rate of 0.9°/month [156], with a range of 0.3°-1.6°/month, which is on average 5.4° in six months (range 1.8°-9.6°). In addition, the purpose of bracing is cessation of scoliosis progression. In our study 34/50 patients wore a brace, of which half were in the Schroth and half in the control group. Given that scoliosis tends to progress over time and that bracing aims only at halting the curves from progression, we believe that a curve correction of 3.42° and 8.68° over a six months intervention period is clinically meaningful. Assuming that all patients with missing values experienced curve progression, three (12%) deteriorated by >5° in the Schroth group, four improved (16%), and 18 remained stable (72%). In the control group, 10 deteriorated (40%), one improved (4%) and 14 (56%) remained stable. More patients were successfully treated (improved + stable) in the Schroth than in the control group (88% vs. 60%, Chi-square 5.1, p=0.024).

For the Biering-Sorensen test, we adopted the largest MCID estimate, which was based on the mean change for a GRC≥2, because although the correlation with the anchor was smaller than recommended (r=0.23), the MCID (36.15 sec) was larger than the SEM (22.77 sec). The MDD_{95} (62.79 sec) corresponded to 43% of the baseline values (110.96 sec). A change such as MDD_{95} would likely be unrealistic given that publications on the effect of focused endurance training suggests only an 18% increase in hold time on Biering-Sorensen test is possible in 3 months in low back pain patients[215], a 22% in six weeks in female students[216] and a 32%[217] in ten weeks in female gymnasts. This suggests that improving the reliability of the test by averaging multiple repetitions may be necessary in the future to minimize the distribution-based estimate and align them more closely with the anchor-based estimates.

Table 6.5. Summary of the anchor- and distribution-based cutoffs for the SRS-22 scores in previous studies compared to the present study. GRC stands for Global rating of change; N/A not available.

	Anchor used	Pain	Function	Image	Mental health	Raw sum	Average sum
Bago et al (MDD_{95})[39]	N/A	0.60	0.80	0.50	0.40	6.80	0.50
Bago et al (MCID)[39]	GRC (4 on a 4-point Likert scale)	0.60	0.30	1.30	0.30	13.10	0.60
Bago et al (ROC curves)[39]	GRC (4 on a 4-point Likert scale)	0.20	0.00	1.60	0.40	10	0.40
Carreon et al (MDD_{90})[40]	N/A	0.33	0.41	0.47	N/A	N/A	N/A
Carreon et al (ROC curves)[40]	SRS-30 for the Pain and Function; SRS-30 and SAQ for the Image domain	0.20	0.08	0.98	N/A	N/A	N/A
Schreiber et al MDD_{95}	N/A	0.35	0.31	0.54	N/A	N/A	0.36
Schreiber et al (MCID)	GRC (2 to 7 on a 15-point Likert scale)	0.16	0.04	0.04	N/A	N/A	0.07
Schreiber et al (ROC curves)	GRC (2 to 7 on a 15-point Likert scale)	Could not be determined because of the imbalance between true and false positive rates					

Bago and colleagues were the first to establish the clinical significance cutoffs for all SRS-22r questionnaire domains in surgically treated patients with AIS. Carreon et al determined the cutoffs also in surgically treated patients but only for the SRS-22r Pain, Function and Image domains. In contrast, we estimated the clinically significant cutoffs for the Pain, Function and Image domains, as well as the Total score on the SRS-22r questionnaire in conservatively treated patients. The clinical significance thresholds from these three studies are presented in *Table 6.5*. In the present study, one of the anchor-based MCID was defined as the difference between means at baseline and 6-months in the group of patients who classified themselves in the range of minimally to much better GRC scores (+2 to +7) after the Schroth treatment. Bago et al, used a different 4-point Likert scale as anchor and their MCID corresponded to the change in patients who stated as being "Much better" after the surgical treatment. In contrast, Carreon et al[40] used a different anchor for each of the outcomes to estimate the most optimal cutoff using the ROC curves in surgically treated patients with AIS. For the image score they defined five groups and

selected a ROC curve cutoff corresponding to the patients who felt "better" and "much better" with regards to their constructed anchor, which is similar to the GRC >2 adopted in the present study. For pain and function they identified three groups: improved, no change and deteriorated. The ROC curve cutoffs for pain and function corresponded to the improved group of patients. The main goal of conservative treatment is to stop the curve progression[143], whereas the goal of surgery is not only to arrest the progression, but also to maximize curve correction.[1] Therefore, since the goals and the invasiveness of the two treatments differ drastically, the expected patients' change also differs. Selecting a cutoff score for clinical significance that corresponds to the patients who perceive themselves as minimally to much better is a clinically sound choice, and can be considered fairly stringent in this group where natural evolution of scoliosis would normally suggest continued deterioration or a stable condition[17].

Cohen[210] and other authors[193] recommend that the correlation between the anchor and the target should be $r \geq 0.3$ to justify the anchor is adequate for determining clinical significance. This was achieved only for the Largest Cobb and the Combined Cobb. Correlation for Biering-Sorensen test was slightly below this recommended correlation. As a result, the ROC curves constructed for all the other outcomes were non-informative as they produced similar true positive (sensitivity) and false positive (1- specificity) rates. Measurements with low correlations to the anchor could not discriminate between patients who felt improved, vs. the ones who did not. The lack of correlation between the anchor and the targets suggests that the change in the questionnaires' domains did not reflect the change in the GRC instrument. It can be that the anchor chosen, the GRC, with its 15-point scale ranging from -7 to +7 ("a very great deal worse" to "a very great deal better"), had too broad a meaning for patients. On the other hand, the GRC presents a single measurement of how patients perceive a change that has occurred[38], so whatever aspect of change is important to patients, be it a cosmetic improvement, muscle endurance increase or pain decrease, they should be able to rate it on the GRC, without being restricted to any specific change characteristic. Measurements with low correlations to the anchor may present scaling issues or may not be adequate to monitor the condition, as it matters to patients conservatively treated for scoliosis. Alternatively, the GRC can be biased by the patient's recall. Patients might not remember well their health status before the start of the treatment if the time elapsed between the baseline and the final measurement is long. Nevertheless, a recent study suggests that the reliability of retrospective assessments, such as the

GRC used in this study, is "moderate to substantial" and higher in younger respondents[38]. Therefore, since the subjects in our study belong to a group of younger respondents (10 – 18 y/o), and the GRC has been collected longitudinally three months apart, we believe that the GRC remained an appropriate instrument.

Another solution for the lack of association between the anchor and the outcomes measured might be to examine the clinical significance of the treatment from the perspective of an unbiased assessor, or a parent who likely will have different expectations than the patients[207,219]. Primary caregivers or policy makers are constantly interested in improving care or decreasing health care related costs. Perhaps determining the clinical significance of the treatment from their perspective might be of interest.

Another explanation for the lack of correlation between the anchor and the targets could be in the questionnaires' poor responsiveness to change in the tested population. A high documented rate of ceiling effects for the questionnaires supports the observation that the SRS-22r and SAQ questionnaires do not perform well in the patients with AIS treated conservatively. We observed the following ceiling effects: SRS-22r pain = 18.4%, SRS-22r function = 28.6%, SAQ prominence = 26.5%, SAQ waist = 29.2%, SAQ chest = 46.9%, SAQ trunk shift = 12.2% and SAQ shoulders = 18.4%. SAQ waist and chest domains had important flooring effects at 16.7%, and 14.3%, respectively. We also calculated the percentage of patients who scored ≥4 on the SRS-22r and ≤2 on the SAQ questionnaire domains (i.e. with only 1 point available to possibly show improvements) and determined that majority of patients had high scores: total = 71.3%, image = 47%, pain = 77.5% and function = 100%, general = 18.3%, curve = 79.6%, prominence = 89.7%, waist = 48%, chest = 65.3%, trunk shift = 77.5% and shoulders = 38.8%. Previous studies suggest that patients with AIS who have curves <45°, on average do not experience pain[16,61], have high level of function[16,61] and do not have negative self-image regarding back appearance[220], which was corroborated in this study with high scores on corresponding domains. The SRS-22r[200] and the SAQ[178] questionnaires were originally designed for the surgically treated patients with AIS. We suggest that due to a high ceiling effects and lack of correlation with the anchor, these instruments are not sensitive enough to capture change in conservatively treated patients with AIS using exercises, whose curves are between 10°-45°. Moreover, a recent research suggests that adolescents have difficulties in understanding the questions and drawings

used in SAQ questionnaire[190]. Given this, it may be worthwhile exploring different scoliosis-specific quality of life instruments in the population of conservatively treated patients, such as Trunk Appearance Perception Scale (TAPS) that measures patients' appearance perception[187] or the Trunk Aesthetic Clinical Evaluation (TRACE) that evaluates the appearance as judged by a clinician[221].

High ceiling effects give little room to capture patients' improvements, which could explain the relatively low rate of improvement in both the exercise and the control group in this study. This is however consistent with AIS being a progressive chronic condition affecting otherwise healthy teenagers. There was a consistent higher rate of deteriorated patients in the control compared to the Schroth group, which indicates that the Schroth exercises combined with standard of care was better than standard of care alone. Effect sizes for Cobb Combined and Largest Cobb were large (-0.77 and -0.92 respectively) favoring the Schroth group. The Biering-Sorensen test also favored the Schroth group with medium effect sizes. Despite the criticism of the ability of the questionnaires to capture changes, small effect sizes were observed for the SAQ general and trunk shift domains and for the SRS-22r function, total, pain and image scores, again in favor of the Schroth group. For the other SAQ scores, small effect sizes for curve and prominence domains favored the control group, while waist, shoulders and chest domains captured no effect (0.05, -0.01 and -0.04 respectively).

In conclusion, for the first time the clinical significance was determined for the SAQ questionnaire domains in patients with AIS. We showed that the SRS-22r and the SAQ questionnaires have limited ability to demonstrate clinically significant improvements in the population of adolescents with AIS, who have smaller curves (10°-45°), and who are generally in good health. Until more appropriate instruments for assessing change in this group of patients have been constructed, we recommend using distribution-based methods to determine the clinically meaningful effect of a therapy, with understanding that the results might not convey the perspective of patients.

Our study determined for the first time anchor-based and distribution-based thresholds for clinical significance in outcomes used to measure the effect of a six-months long conservative treatment using Schroth exercises for patients with AIS combined with standard of care consisting of observation and bracing. Based on our findings, we conclude that the Schroth

intervention had clinically significant effects on the radiographic outcomes (Largest Cobb and Combined Cobb), as well as, on the SRS-22r function domain.

CHAPTER 7

General discussion and conclusions

7.1 Summary of the results

The main focus of this thesis was to determine the effect of Schroth exercises added to the standard of care on curve severity, quality of life, perceived body image, and back muscle endurance compared to standard of care alone in a randomized controlled trial (RCT). Further, in order to standardize the treatment and to provide an appropriate therapy plan for the patients, a Schroth classification algorithm was developed and tested in a sample of 10 Schroth therapists. The reliability was determined for the entire sample of Schroth therapists (N=10), experienced (N=2), and well-trained raters (N=6). Finally, to better inform clinical practice about the meaning of the results produced in the RCT, clinical significance of the Schroth intervention has been determined by identifying thresholds for clinically important changes on each outcome using anchor- and distribution-based methods. The numbers needed to treat (NTT), as well as the percent of improved, stable and deteriorated patients were also reported for each outcome.

7.1.1 On the reliability of Schroth curve type classification decisions

Intra-rater AC1 and weighted AC1

- The intra-rater reliability of the ***entire sample*** of the therapists using the proposed Schroth classification algorithm was substantial with AC1 0.64 (95% CI 0.46-0.82). The weighted intra-rater AC1 was substantial 0.75 (95%CI 0.63-0.84).
- The intra-rater AC1 of the ***experienced therapists*** was almost perfect at 0.81 (95%CI 0.77-0.85), and the weighted AC1 was substantial 0.89 (95%CI 0.80-0.94).
- The intra-rater AC1 among the ***well-trained*** therapists was substantial at 0.70 (95%CI 0.60-0.78), and weighted AC1 was almost perfect at 0.82 (95%CI 0.73-0.88).

Inter-rater AC1 and weighted AC1

- ***The Entire sample*** had a moderate inter-rater AC1 of 0.43 (95%CI 0.28-0.58), and moderate weighted AC1 of 0.48 (95%CI 0.29-0.67).

- *The **experienced therapists*** had a substantial inter-rater AC1 of 0.67 (95%CI 0.50-0.85), and substantial weighted AC1 of 0.79 (95%CI 0.64-0.94).
- ***Well-trained therapists*** had a moderate inter-rater AC1 of 0.50 (95%CI 0.38-0.61), and substantial weighted AC1 of 0.61 (95%CI 0.49-0.72).

7.1.2 On the statistical significance of the effect of Schroth exercises in the RCT.

- The completers in the Schroth group had significantly smaller largest curve (LC) by 3.5° (p<0.01), compared to the completers in the control group. Intention-to-treat results were similar in direction and magnitude, but did not reach statistical significance;
- The difference in the square root of the sum of all curves (SOC) was significantly smaller among completers in the Schroth group by 0.34 (p<0.05), such that a patient with an average SOC of 51.2° at baseline had a 49.3° SOC at the end of treatment in the Schroth group, and of 55° in the control group. Intention-to-treat results were similar in direction and magnitude, but did not reach statistical significance.
- The Schroth intervention improved the back muscle endurance of the completers in the experimental group by 30 seconds on average (p=0.02), compared to the control group. Intention-to-treat analysis produced similar results, but did not reach significance.
- The effect of Schroth exercises did not reach statistical significance on other outcomes which included the following SRS-22r scores (pain, function, self-image and total score), and SAQ scores (curve, prominence, trunk shift, shoulders, waist, chest and general score).

7.1.3 On the influence of covariates on outcomes in the Schroth trial

- Heavier persons had larger LC, so that with every 1 kg increase in weight, patients had on average a 0.5° larger LC.
- Heavier patients had larger SOC. For a hypothetical patient with a SOC of 60°, the predicted SOC for persons who are 5 kg, 10 kg and 15 kg heavier, but with similar other characteristics, would be 65°, 75° and 84°, respectively.
- Patients with 3cp curve type had on average larger LC by 12° compared to the 3c type, and on average 8.8° larger LC compared to 4cp curve types

- Patients with a 3cp Schroth classification had a larger SOC compared to the 3c classification. For a person with 3cp curve type and with a SOC of 50°, a similar person on other model characteristics, but classified into the 3c category will have a SOC of 31°.
- Curve type classification influenced SRS-22 pain scores. For both groups, patients with a 3c curve type (main thoracic curve with aligned pelvis) had the best SRS-22r pain score, while the worst score was observed in patients with 4cp curves (main thoracolumbar/lumbar curve with pelvis displaced opposite of the major curve), p=0.04.
- Patients with 3c curve types had a statistically significantly better SAQ trunk shift score compared to patients with 3cp (main thoracic curve with pelvis displaced opposite to the major curve) (p=0.006) and 4cp (p=0.047) curve patterns.
- Patients with 3c curve type had better SAQ chest scores than the ones with 3cp (p=0.046), 4c (main thoracolumbar/lumbar curve with aligned pelvis) (p=0.02), and 4cp (p=0.02) curves;
- Patients classified as 3cp were 9.17% more likely to have an SAQ curve score of 3 or worse (p=0.02).
- Patients who wore a brace had on average significantly better SRS-22r self-image score by 0.31 than the ones who did not (p=0.03).
- Patients wearing a brace had worse SAQ waist scores than those without brace (p=0.04).
- Patients who wore brace had better SAQ chest scores than the ones who did not (p=0.007).
- Patients who wore a brace had better SAQ general scores by 0.74 on average (p=0.002).
- Patients who were 10 and 11 years old had better SAQ trunk shift scores on average by 0.75 points compared to older participants (p=0.002).
- Older patients had worse SAQ prominence scores (p=0.004).
- Thirteen-years-old patients had better SAQ chest scores compared to all other ages (p=0.009).
- Taller patients had better SAQ trunk shift scores (p=0.02).
- Taller patients had better SAQ prominence scores (p=0.04).
- Taller patients had worse SAQ general score (p=0.046).
- In general, patients who had higher self-efficacy scores (stronger belief in their own ability to perform the required exercise program), had better SAQ waist scores (p=0.04).

7.1.4 On the clinical significance of findings in the Schroth trial

- The effect of Schroth exercises reached clinical significance thresholds for the LC and SOC, based on distribution-based cutoffs, as well as SRS-22r function score, based on the anchor-based cutoff. The corresponding cutoffs for these outcomes were: LC=3.4°, SOC=8.7°, and SRS-22r function=0.11, respectively. The NNT were 3, 4 and 4, respectively. The combined percentage of improved and stable patients in the Schroth group compared to the control group was 87% vs. 55% for the LC, 91% vs. to 65% for the LOC, and 90.5% vs. 65% for the SRS-22r function score;
- Based on the cutoffs produced in the analysis of clinical significance, Schroth intervention did not have clinically significant effect on all other outcomes.

7.2 Reliability of the Schroth curve type classification in adolescents with idiopathic scoliosis

Although the Schroth therapy plan for patients with AIS depends on the preliminary curve type classification, the reliability of the Schroth therapists in classifying patients within the four Schroth curve types was unknown. Classification algorithms have been shown to improve raters' reliability estimates. Examples include the algorithms proposed for King's[130] and Lenke's[136] classifications.

We developed a novel algorithm based on the Schroth manual to assist therapists in classifying patients participating in the Schroth RCT, and to minimize the treatment errors that may stem from inappropriate classification. There are four Schroth classifications: two with major thoracic curve, having an aligned (3c) or misaligned pelvis (3cp), and two with major thoracolumbar/lumbar curve, having an aligned (4c) or misaligned pelvis (4cp). The disagreement within the two main classifications (3c and 3cp, or 4c and 4cp) should be penalized less, because the treatment would be similar. Thus, we also tested the weighted reliability to address these similarities in treating patients within the major thoracic or major thoracolumbar/lumbar curves categories.

The weighted intra-rater AC1 overall met our a priori threshold to be considered sufficient for recommending the use of our algorithm. In addition, the intra-rater AC1 and weighted AC1 for the experienced and well-trained therapists were also sufficient, as well as the weighted inter-rater AC1 for the experienced therapists. Having a better understanding of the algorithm and being more familiar with its usage was associated with higher reliability estimates. The high reliability observed for the therapists involved in the Schroth trial show that the choice of exercises prescribed as a function of patient's curve type was adequately standardized. We hypothesized that the reliability of the entire sample would be higher with the usage of our standardized algorithm. Lower estimates than expected may suggests that additional training of the raters in using the algorithm is necessary, given the substantial to almost perfect reliability of the most experienced and well-trained raters; or the algorithm itself should be improved. However, because the reliability of the Schroth classification has not been reported without the use of an algorithm it is not possible to determine if reliability was optimized by the use of our algorithm.

7.3 The effect of Schroth exercises added to the standard of care on curve severity, quality of life, perceived body image and back muscle endurance in adolescents with idiopathic scoliosis – an assessor and statistician blinded randomized controlled trial

In America, patients with AIS with curves <45° have very few treatment alternatives that only include observation (no specific treatment applied) and bracing, prescribed based on the estimated risk of progression. [42] In Europe, on the other hand, exercises are frequently prescribed for this group of patients.[128,166,222] Exercise therapy has been widely ignored as a treatment option for scoliosis in America although several studies, among which there are now two RCTs[20,72], demonstrated a positive effect for scoliosis exercises on various outcomes, including decreasing curve magnitudes,[20,71-74] improving muscle strength,[73,75,76] pulmonary function,[73] reducing prevalence of surgery,[25] reducing the prevalence of bracing,[71], improving postural deficiencies,[73] and quality of life,[20,74]. Interestingly, bracing has been part of the standard of care for decades, although up until very recently,[99] the effect of bracing had also been

controversial.[69,98] There are several scoliosis-specific exercise approaches described in the literature,[18] including Schroth, Integrated Scoliosis Rehabilitation, Dobomed, Side-shift, Lyon, Functional Individual Therapy of Scoliosis and the Scientific Exercise Approach to Scoliosis. Among those, most studies researched the Schroth method, and although they all demonstrated promising results on scoliosis outcomes, they were of suboptimal quality.[83,223] To provide stronger evidence of the effect of Schroth exercises on selected scoliosis outcomes we conducted a RCT.

The major finding in the RCT was that after the six-months long Schroth intervention combined with standard of care (observation or bracing), the exercise group achieved a significant improvement in curve severity (LC and SOC), and back muscle endurance compared to the controls. This improvement in the Cobb angles is not surprising, given that the previous research on Schroth exercises, of which one prospective (not controlled),[73] and one retrospective controlled study[74] provided similar results. In another prospective study that tested the effect of inpatient Schroth intervention using matched pairs from another prospective trial of non-treated patients with AIS,[24] the authors found that incidence of curve progression was significantly smaller in the treatment group. Similarly, two long-term retrospective studies from Europe suggest that there is less prevalence of surgery in conservatively treated patients (exercises, and a combination of exercises and braces) compared to patients simply observed.[25,86] In one study the prevalence of surgically treated patients with AIS was 7.3%[86] after an intensive inpatient Schroth intervention plus bracing. The other study showed a 12.1% surgery rate after outpatient Schroth therapy, and 14.1% after a combined outpatient Schroth treatment with bracing.[25] The results from these studies directly or indirectly indicate that treatment using Schroth exercises as an add-on to braces or a stand-alone intervention may lead to curve improvement or stabilization.

Our study contributes to the evidence about efficacy of the Schroth treatment by providing the results of the first RCT, which corroborate previously published findings on Cobb angle change. In this RCT, we also investigated the effect of the treatment on back muscle endurance, SRS-22r and SAQ questionnaires' scores. The Schroth intervention improved back muscle endurance, but not the questionnaires' scores.

Several exercise studies, including the Schroth exercises[73] and supervised resistive rotational exercises[75,76] investigated their effect on the muscle strength. In their one-year-long prospective

uncontrolled study using Schroth exercises, Otman et al showed that the muscle strength significantly increased from baseline to discharge.[73] Likewise, in two other studies that examined the effect of four months of supervised resistive rotational exercises on torso rotation strength, the authors also found a significant increase in back muscle strength.[75,76] To our knowledge, the present study is the first to assess and show improvements in back muscle endurance after an exercise treatment for AIS.

Only two recent exercise studies, published after our trial began, investigated the quality of life measured using SRS-22r questionnaire scores, in patients with AIS[20,74] and both found significant improvements in some of the SRS-22r domains. The retrospective controlled study by Noh et al. that investigated the effect of the Schroth exercises combined with spinal stabilization,[74] found better SRS-22r results at four months on self-image and the total scores, compared to the conventional exercises. The results of the recent Monticone et al. RCT found positive effects of scoliosis-specific active self-correction and task-oriented exercises on changes in SRS-22r scores at skeletal maturity compared to standard rehabilitation exercises.[20] The discrepancy between the results reported in these two studies and our RCT may be because of the more intensive frequency of visit in the Noh et al. study (2-3 times/week for four months vs. once a week for six months), or the longer intervention in the Monticone et al. study (until maturity vs. six months). On the other hand, in our sample, we also found a high ceiling effect in both SRS-22r and SAQ questionnaires' scores, suggesting that those clinical tools might have inadequate responsiveness to the changes in these outcomes in our population. In contrast, participants in both the Noh et al.[74] and the Monticone et al.[20] study had worse baseline scores on the SRS-22r domains than we observed in our RCT, and therefore more room for capturing improvements on these scales.

To our knowledge there were no studies that used the SAQ scores to assess the perceived image over time in patients treated conservatively. Danielsson et al.[189] used only the pictorial part of the SAQ questionnaire, but not to determine the pre/post treatment change in the scores, rather to assess the correlation with the curve size, trunk rotation measured with the scoliometer, and SRS-22 domain scores in 77 patients 16 years after they had been braced or observed for AIS. The authors found that despite similar curve magnitudes in both groups, previously braced patients had worse perceived spinal appearance compared to the patients who were only

observed. This finding is supported by the results of our study, but only for the SAQ waist score. For the SRS-22r self-image, SAQ chest and general score, better scores were observed among braced patients. The comparison between the results of these two studies should be carried out cautiously, as Danielsson et al[189] used the sum of only the pictorial portion of the questionnaire and the study was conducted 16 years after the adolescence.

7.4 Clinical significance of the effect of a six-month Schroth exercises intervention in adolescents with idiopathic scoliosis

Many authors agree that reporting statistical test results should be supplemented with methods for determining the clinical significance of changes observed[27-29,118]. This is because frequently, by means of statistical inference, one cannot judge upon what is important for a patient, a therapist, a stakeholder or any other end user. As argued by Osoba et al, [224] sometimes very small differences in the clinical outcome can reach statistical significance in large trials, but not all statistically significant difference is indeed clinically meaningful. Likewise, a statistically significant difference in the outcome might not be found, while even a small difference might be meaningful for a patient. We tested clinical significance of the results using anchor- and distribution-based methods since there is no consensus in the scientific community on which methods should be used.[34,38] To determine the anchor-based cutoffs, a global rating of change (GRC) was used to link the mean difference in the outcomes with the responses of patients' who perceived their change from "a little bit better" to "a great deal better" (+2 to +7 on the GRC). We also used the receiver operating curves (ROC) to identify cut-offs on outcomes with best sensitivity and specificity to detecting patients perceiving improvements. The distribution-based methods used were standard error of measurement (SEM), minimal detectable difference with 95% confidence interval (MDD_{95}) and the effect size estimated using Cohen's d.

The clinical significance cutoff for the curve magnitude outcomes was determined using the distribution-based method MDD_{95} (3.42° for LC and 8.68° for LOC), because anchor-based methods produced values deemed too small (-1.33° for LC and -2.74° for LOC), and given our high reliability coefficient in measuring Cobb angles, the SEM cutoff (1.24° for LC and 3.15° for SOC) was also considered too small. All questionnaires' scores cutoffs were determined using the SEM, because all anchor-based cutoffs were below the SEM of the instrument, while the

MDD_{95}, produced values, which, given a high rate of ceiling effects observed for the questionnaires' scores, would be impossible to reach (MDD_{95} cutoffs for the questionnaires' score, ranged from 6.8% to 108% with an average of 41.7% of the mean baseline values). The Biering-Sorensen back muscle endurance test cutoff was determined using the anchor-based mean change difference observed in patients who perceived their change as "a tiny bit better" to "a great deal better" on the GRC (36.15 sec). This was because the MDD_{95} estimate (62.79 sec) corresponded to about 43% of increase in back endurance from the baseline (110.96 sec), which exceeds the magnitude of changes reported in several studies that investigated the effect of an active endurance therapy on back muscle endurance using the Biering-Sorensen test, [215-217] while the SEM (22.77 sec) was deemed too small. In this study, using the cut-off estimated as above, we found that the differences corresponding to a positive effect of the Schroth exercise program in the largest Cobb, sum of all Cobb angles and the SRS-22r function score reached clinical significance. In addition, significantly more patients in the Schroth group avoided deterioration on these three outcomes compared to the control group, determined by the Chi-square test comparing the proportion of deteriorated patients vs. not (combining improved and stable) in each group.

Since this was the first study to determine the clinical significance of the outcomes in patients with AIS treated conservatively, it is difficult to discuss the potential differences and similarities with other exercise studies. However, there were two other studies that investigated the clinical significance in surgically treated patients with AIS. Bago and colleagues[39] were the first to establish the clinical significance cutoffs for all SRS-22r questionnaire domains, while Carreon et al[40] determined the cutoffs only for the SRS-22r pain, function and self-image domains. Interestingly, although the study population in these two studies were surgically treated patients, our SRS-22r image domain MDD_{95} was slightly higher than the MDD_{95} reported in the study by Bago et al. With regards to Bago et al's and Carreon et al's studies, the cutoffs on the same domains were consistently smaller in the Carreon's study, perhaps because this group used MDD with 90% confidence interval, compared to 95% in the Bago's study. Carreon et al also used different anchors (sum of the relevant questions from the last eight questions on the SRS-30 questionnaire for the pain and function, and the sum of the relevant questions from the SRS-30 with two questions from the SAQ for the self-image), compared to the use of a GRC in the Bago at al's study. This suggests that even in the same population of patients, determining clinical

significance of the outcome is challenging. Our results direct researchers towards using different QOL tools for the patients treated conservatively, because all, except one domain anchor-based cutoff were below the SEM. Moreover, we observed a high rate of ceiling effects on both questionnaires, ranging from 12.2% to 47%. This population also exhibited a high preponderance towards the best scores (i.e. with only one point available to possibly show improvements), ranging from 39% to 100%.

7.5 Contribution to the knowledge and future work

The study that determined the reliability of the Schroth therapists in classifying patients with AIS using our standardized algorithm, the RCT that evaluated the effect of the Schroth intervention combined with the standard of care on the curve magnitude, quality of life, perceived body-image and back muscle endurance change compared to standard of care alone, and finally, the study in which the clinical significance of the results produced in the RCT was determined, present novelties in research on Schroth exercises for AIS.

7.5.1 Reliability study

There exist several scoliosis classifications, that were developed to guide surgical treatment, including King's,[112] Lenke's[114] and PUMC.[115] Classifications developed for surgical planning have also been used to guide brace design. Rigo's[108] and Weiss'[225] classifications were developed recently specifically to guide the design of the Cheneau light braces, while SpineCor's classification[226] is used for their soft brace design. In contrast, the Schroth classification is the only one to exclusively guide exercise prescriptions. Algorithms to help guide the scoliosis curve classification decisions are not novel.[130,136] However, until now, algorithms to help clinical decisions in conservative treatment for scoliosis did not exist. The standardized algorithm to help Schroth therapists classify scoliosis patients into four different categories is the first of this kind. Together with the algorithm, we provided the operational definitions to help therapists follow the algorithm step by step. Establishing the reliability of Schroth therapists in classifying patients provided a novel and direct contribution to the evidence base on the Schroth exercise approach.

By determining the non-weighted and weighted reliability estimates, we provided insight into different levels of disagreement. The weighted agreement was improved compared to non-weighted for both the intra- and inter-rater reliabilities. Different weights were assigned to

different levels of disagreement, such that the disagreements were not penalized as harshly when a therapist disagreed on whether a pelvic component was present, but agreed on the dominant area of deformity. This is important for future studies on Schroth exercises, especially when more than one therapist will be providing the treatment, as it reassures that patients with the same dominant deformity will be taught similar corrections. Nevertheless, it is recommended that the reliability among the therapists involved in other studies be determined before new studies begin. By using our algorithm, the classification task should be simplified, especially since rater responses suggested that it is well accepted among therapists. Moreover, with an adequate training in using the algorithm, high reliability estimates are likely. Our training may not have been sufficiently long and practice cases sufficiently complex to ensure that all therapist reach 100% understanding of the algorithm before using it, which appears to have introduced a relationship between the classification reliability estimates and reported understanding of the algorithm. We hypothesize that appropriate classification decision should assure appropriate treatment for patients, maximizing the likelihood of obtaining an optimal treatment effects. Newly certified Schroth therapists frequently report difficulties in distinguishing between the curve types. Our algorithm may help transition between the Schroth training and clinical practice. Thus, we propose its use during the Schroth certification, with adequate opportunity to practice its application because it is simple and easy to follow, but requires adequate training.

7.5.1.1 Future research related to the Reliability study

Future studies should focus on determining the sources of disagreement among therapists. We learned from our study that experience in using the algorithm improved the reliability estimates. We also learned that the lowest intra-rater reliability was associated with a higher number of patients previously treated by a therapist. We speculate that less experienced therapists in treating patients using Schroth, are more likely to strickly follow the instructions taught during the Schroth training. On the other hand, different underlying factors could be associated with low reliability. It would be important to test each question on the algorithm separately to learn if a certain step contributed most to low reliability. In determining the curve classification, there are two to five questions a therapist has to answer in order to reach a unique classification decision. There are seven questions on the algorithm in total, each asking about different body features, starting with whether the pelvis is unbalanced or balanced. It would be interesting to learn which

one of the questions was most problematic. Once it is known which question led to the lowest reliability estimate, it would be possible to adjust the operational definitions, or the wording that was used in the algorithm. Testing each of the seven questions will help improving the algorithm. Another focus could be testing how the algorithm's performance changes when used in patients with more obvious deformities.

7.5.2 Randomized controlled trial

The literature on the effect of exercises for scoliosis has been criticized for the lack of strong evidence stemming from the poor designs of these studies. Indeed, most previous exercise studies for scoliosis were not randomized or controlled, the assessors were not blinded, the attrition rate or the reasons for dropouts were not reported, intention-to-treat analyses were not used, and the study protocols were not readily available.[18,79,83,223] However, strong evidence of positive effect of exercises on scoliosis outcomes is emerging. The recent RCT by Monticone et al. found that the scoliosis-specific active self-correction and task-oriented exercises significantly improved the Cobb angles (by 5.3° at skeletal maturity) and SRS-22r questionnaire scores (by 0.75 to 0.89/5), while the traditional spinal exercises were associated with stable outcomes.[20] This was a well designed, RCT study with a long follow up and with blinded assessors. However, the attrition rate or the reasons for dropout have not been reported, and whether there were differences between the dropouts and the completers is unknown.

Our study is the first RCT to investigate the effect of the Schroth exercises on the curve severity, quality of life, perceived body-image and back muscle endurance in patients with AIS. This study also contributes to the body of knowledge about the effect of exercises for scoliosis and could potentially influence the treatment guidelines, especially once the long-term effects are tested.

This study provides rigorous short-term evidence to support the usage of Schroth exercises in treating patients with AIS who are being observed or are treated with a brace. The risk of bias in our study is lower than in previous studies. The patients were randomized using a computer-generated sequence in pre-sealed envelopes into the Schroth exercises combined with standard of care or the control groups (standard of care only). The study groups were similar in all descriptive characteristics at baseline (curve magnitude, Risser, age, weight, and height), an

equal number of patients was allocated to each arm (25:25), similar curve type prevalence was observed in both groups, and equal number of patients were braced in each group. Blinding of the therapists and the patients was not possible as in many exercise studies. However, the assessors of objective outcomes (Cobb angles and Biering-Sorensen back muscle endurance test) were blinded to group allocation. Self reported questionnaires were completed by the patients, and checked for completeness by the masked assessor. The attrition rate was relatively low (6/50 patients, or 12%) and the overall compliance was high (76% of visits attended and 73% of completed exercise home program). There were two dropouts in the control group and four in the Schroth group. The reasons for dropping out were not related to the design of the trial or the outcomes, and included: relocation and travelling for more than three months during the trial in the control group, and time constraint in the exercise group. We reported all results announced in the proposal regardless of the significance reached. We performed the intention-to-treat analysis, using the linear interpolation method,[152] deemed the most appropriate for our data analysis. All patients who completed the study, remained in the group they were originally assigned to. Per protocol analysis was also carried out and the results were reported. Finally, our study protocol was previously published.[198]

Due to using a rigorous randomized controlled study design and the attrition being due to reasons unrelated to the study design or outcomes, the results from the per protocol analysis could be considered valid. [154] Although the statistical estimates were similar to the analysis of completers, the intention-to-treat analysis results did not reach significance emphasizing the importance of patient compliance to ensure the effects of Schroth exercises can be detected.

7.5.2.1. Future research related to the RCT

Future research should include a larger population of patients with AIS. Therefore, our ongoing multicenter trial that uses the same study protocol is justified. However, the high rates of ceiling effects on both SRS-22r and SAQ questionnaires, also reported in previous studies,[183,184,186] that were found in this PhD study warrant the use of different quality of life questionnaires in future studies. Recently, it has been reported that adolescents lack in understanding of the SAQ questionnaire's both pictorial and verbal items.[190] These authors found that the percent of subjects who encountered at least one problem ranged from 16% to 96%.[190] Using an interview administration of the questionnaire could possibly circumvent the misunderstanding of the

questions.

Positive short-term (six months) effects were demonstrated on the curve severity and endurance in the present study. Future research should focus on determining the long-term effects of the Schroth intervention. In particular, the effect until the maturity is reached. In addition, it is necessary to establish guidelines for the assessment of the exercise effectiveness, as has been proposed for braces.[227] Referring to the Scoliosis Research Society guidelines for assessing bracing effectiveness, the guidelines for the assessment of exercise effectiveness should include similar criteria. One example of such criteria could include: reporting the percentage of patients who have had bracing recommended/undergone for patients with smaller curves, or if exercises were used combined with bracing, the percentage of patients who have had surgery recommended/undergone before reaching skeletal maturity[227]; the percentage of patients who progressed beyond the calculated SEM for the Cobb measurements in the particular study, and the percentage of those who remained within the SEM; if using exercises and braces, the percentage of patients who progress beyond 45°, indicating the possible need for surgery; a minimum two-year follow-up beyond skeletal maturity for each patient who did not deteriorate with exercises or exercises plus braces to determine the percentage of patients who subsequently required or had surgery recommended. Finally, a clinical prediction rule to identify patients who would benefit from the Schroth treatment is also a matter for future work. That way, the overtreatment may be avoidable.

While the results may suggest that Schroth therapy be made available to patients with scoliosis in North America, our trial may not yet have a large impact on practice. For example, since the Schroth treatment is currently not routinely available in the health care system in North America, patients cannot be guaranted treatment by adequately trained therapists with regular follow-ups. To our knowledge in Canada, prior to our organizing a Schroth certification course there were fewer than five practicing therapists offerring the Schroth therapy. In order to establish such care for patients with AIS, a strong knowledge translation plan should be carried out before this treatment becomes accepted and implemented into standard of care. Providing training for practicing physiotherapists will be required. A cost-benefit study may also be required before a widespread change in practice is implemented.

7.5.3 Clinical significance

Determining the clinical significance of the outcomes using standardized instruments for scoliosis in conservatively treated patients with AIS is novel. The data from the Schroth RCT was used to investigate the clinical significance cutoffs of the largest Cobb angle, the sum of all Cobb angles, the Biering-Sorensen back muscle endurance test, and the SRS-22r and SAQ questionnaires scores.

The most important finding in this study was that the Schroth exercises had clinically meaningful effect on the curve severity (LC and SOC), and the SRS-22r function score. Also, the percent of patients who improved and remained stable was higher in the Schroth group on all outcomes, but did not reach clinical significance based on the cutoffs predetermined in this study. The NNT for the largest curve and the sum of all curves, suggest that only three and four patients, respectively, would need to be treated in order to prevent a progression of the curves beyond the estimated clinical importance cutoffs in one patient over six months of supervised Schroth intervention added to the standard care. For the SRS-22r function the NNT was also four. Such small numbers indicate that the effect of Schroth exercises as an add-on to standard of care had a large clinical importance.

On the other hand, the clinical significance cutoffs were determined using the distribution-based methods exceeding the anchor-based estimates, which means that the patients' perceptions about the treatment effect were not taken into account. This was because the calculated anchor-based mean pre/post treatment differences observed in patients who perceived their change as "a tiny bit better" to "a great deal better" for the radiographic outcomes were within the SEM and deemed too small to be considered clinically important. The anchor-based cutoffs for all the other outcomes (except for the Biering-Sorensen test) were also below the SEM. The cutoff for the SRS-22r pain was within the range of measurement error. Another problem with the anchor-based methods was the lack of correlation of all the target outcomes except the Cobb measurements with the anchor (GRC).

Although we were not able to estimate the clinical significance for all the outcomes based on the anchor-based methods, which would be preferable, because we were mainly interested in the patients' perception of the change, we thoroughly examined the data to determine the cause of

such a result. As previous research studies have already noted, there was a high ceiling effect on the SRS-22r[183] and SAQ questionnaires[228] in the younger population of patients with AIS who were treated using different standard of care therapies. Our findings corroborated the results from these previous studies, and gave contribution about the rate of ceiling effects in the conservatively treated patients with AIS.

7.5.3.1 Future research related to the Clinical Significance paper

Future research on quality of life in patients with AIS should concentrate on exploring the usage of different tools for assessing the quality of life outcomes. Because, the most frequently used quality of life questionnaire, the SRS-22r, seems to be inadequately responsive to change over time in the population of patients with AIS treated conservatively and with curves <45°, developing a more sensitive questionnaire could be another important work. Future work should also focus on examining different anchor definitions, because the GRC might be affected by recall bias or because its meaning could be too broad. An alternative strategy to prevent the effect of recall bias could be including more frequent follow-ups.

7.6 Conclusion

In conclusion, this thesis provided novel and important findings about the effect of Schroth exercises added to standard of care on selected outcomes in patients with AIS through a rigorous randomized controlled trial research design.

- The results of this randomized controlled trial provide important evidence that could have a big impact on clinical practice in North America, where exercise treatment is not a part of standard of care.
- A new standardized classification algorithm has been tested and the results indicate its usage could be a helpful tool for Schroth therapists. Appropriate training in its usage improves the reliability of the therapists, thus likely assisting with standardizing the optimal treatment delivery for patients globally.
- This thesis also provides a unique contribution in determining clinical significance of the outcomes in patients with AIS, and informs clinical practice in a clinically relevant way, rather than just reporting the results of the statistical tests.

- Lastly, the work presented here, provides insights into future research based on the findings described in the previous chapters.

Bibliography

1. Weinstein SL, Dolan LA, Cheng JCY, Danielsson AJ, Morcuende JA. Adolescent idiopathic scoliosis. *Lancet*. 2008;371(9623):1527–1537. doi:10.1016/S0140-6736(08)60658-3.

2. Lonstein JE. Scoliosis: surgical versus nonsurgical treatment. *Clin Orthop Relat Res*. 2006;443:248–259. doi:10.1097/01.blo.0000198725.54891.73.

3. Stokes IAFC. 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*. 1994;19(2):236.

4. Grivas TB, Vasiliadis E, Mouzakis V, Mihas C, Koufopoulos G. Association between adolescent idiopathic scoliosis prevalence and age at menarche in different geographic latitudes. *Scoliosis*. 2006;1:9. doi:10.1186/1748-7161-1-9.

5. Rogala EJ, Drummond DS, Gurr J. Scoliosis: incidence and natural history. A prospective epidemiological study. *J Bone Joint Surg Am*. 1978;60(2):173–176.

6. Hennes A. *Schroth-Method*. Bad Sobernheim: Asklepios Katharina Schroth Klinik; 2008:70.

7. Bagnall KM, Grivas TB, Alos N, et al. The International Research Society of Spinal Deformities (IRSSD) and its contribution to science. *Scoliosis*. 2009;4:28. doi:10.1186/1748-7161-4-28.

8. Cheung KMC, Wang T, Qiu GX, Luk KDK. Recent advances in the aetiology of adolescent idiopathic scoliosis. *Int Orthop*. 2008;32(6):729–734. doi:10.1007/s00264-007-0393-y.

9. Burwell RG, Dangerfield PH, Moulton A, Grivas TB. Adolescent idiopathic scoliosis (AIS), environment, exposome and epigenetics: A molecular perspective of postnatal normal spinal growth and the etiopathogenesis of AIS with consideration of a network approach and possible implications for medical therapy. *Scoliosis*. 2011;6(1):26. doi:10.1186/1748-7161-6-26.

10. Scoliosis Research Society. Adolescent Idiopathic Scoliosis - Treatment. *Scoliosis Research Society*. Available at: http://www.srs.org/professionals/education/adolescent/idiopathic/treatment.php. Accessed April 13, 2011.

11. Wong MS, Liu WC. Critical review on non-operative management of adolescent idiopathic scoliosis. *Prosthet Orthot Int*. 2003;27(3):242–253.

12. Peterson LE, Nachemson AL. Prediction of progression of the curve in girls who have adolescent idiopathic scoliosis of moderate severity. Logistic regression analysis based

on data from The Brace Study of the Scoliosis Research Society. *J Bone Joint Surg Am*. 1995;77(6):823–827.

13. Lonstein J, Carlson J. The prediction of curve progression in untreated idiopathic scoliosis during growth. *Journal of Bone and Joint Surgery*. 1984;66(7):1061.

14. Weiss H-R, Goodall D. The treatment of adolescent idiopathic scoliosis (AIS) according to present evidence. A systematic review. *Eur J Phys Rehabil Med*. 2008;44(2):177–193.

15. Cordover A, Betz R, Clements D, Bosacco S. Natural history of adolescent thoracolumbar and lumbar idiopathic scoliosis into adulthood. *J Spinal Disord*. 1997;10(3):193–196.

16. Weinstein SL, Dolan LA, Spratt KF, Peterson KK, Spoonamore MJ, Ponseti IV. Health and function of patients with untreated idiopathic scoliosis: a 50-year natural history study. *JAMA*. 2003;289(5):559–567.

17. Asher MA, Burton DC. Adolescent idiopathic scoliosis: natural history and long term treatment effects. *Scoliosis*. 2006;1(1):2. doi:10.1186/1748-7161-1-2.

18. Fusco C, Zaina F, Atanasio S, Romano M, Negrini A, Negrini S. Physical exercises in the treatment of adolescent idiopathic scoliosis: An updated systematic review. *Physiother Theory Pract*. 2011;27(1):80–114. doi:10.3109/09593985.2010.533342.

19. Efficacy of exercise therapy for the treatment of adolescent idiopathic scoliosis: a review of the literature. 2011;21(3):382–389. doi:10.1007/s00586-011-2063-4.

20. Active self-correction and task-oriented exercises reduce spinal deformity and improve quality of life in subjects with mild adolescent idiopathic scoliosis. Results of a randomised controlled trial. 2014;23(6):1204–1214. doi:10.1007/s00586-014-3241-y.

21. Hennes A. *Schroth-Method*. Bad Sobernheim: Asklepios Katharina Schroth Klinik; 2011:70.

22. Weiss H. *Best Practice in Conservative Scoliosis Care*. 2nd ed. Munich: Pflaum; 2008:86.

23. Otman S, Kose N, Yakut Y. The efficacy of Schroth s 3-dimensional exercise therapy in the treatment of adolescent idiopathic scoliosis in Turkey. *Saudi Med J*. 2005;26(9):1429–1435.

24. Weiss H-R, Weiss G, Petermann F. Incidence of curvature progression in idiopathic scoliosis patients treated with scoliosis in-patient rehabilitation (SIR): an age- and sex-matched controlled study. *Pediatr Rehabil*. 2003;6(1):23–30. doi:10.1080/1363849031000095288.

25. Rigo M, Reiter C, Weiss H-R. Effect of conservative management on the prevalence of surgery in patients with adolescent idiopathic scoliosis. *Pediatr Rehabil*. 2003;6(3-

4):209–214. doi:10.1080/13638490310001642054.

26. Lehnert-Schroth C. *Three-Dimensional Treatment for Scoliosis*. 7 ed. Palo Alto, California, USA: The Martindale Press; 2007:278.

27. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol*. 1991;59(1):12–19.

28. Jacobson NS, Follette WC, Revenstorf D. Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Behavior Therapy*. 1984;15(4):336–352.

29. Kazdin AE. The meanings and measurement of clinical significance. *J Consult Clin Psychol*. 1999;67(3):332–339.

30. Wyrwich KW, Bullinger M, Aaronson N, et al. Estimating Clinically Significant Differences in Quality of Life Outcomes. *Quality of Life Research*. 2005;14(2):285–295.

31. Sloan JA. Assessing the minimally clinically significant difference: scientific considerations, challenges and solutions. *COPD*. 2005;2(1):57–62.

32. Cella D, Eton DT, Lai J-S, Peterman AH, Merkel DE. Combining anchor and distribution-based methods to derive minimal clinically important differences on the Functional Assessment of Cancer Therapy (FACT) anemia and fatigue scales. *J Pain Symptom Manage*. 2002;24(6):547–561.

33. Yost KJ, Eton DT. Combining Distribution- and Anchor-Based Approaches to Determine Minimally Important Differences The FACIT Experience. *Eval Health Prof*. 2005;28(2):172–191. doi:10.1177/0163278705275340.

34. Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR, Clinical Significance Consensus Meeting Group. Methods to explain the clinical significance of health status measures. *Mayo Clin Proc*. 2002;77(4):371–383.

35. Lemieux J, Beaton DE, Hogg-Johnson S, Bordeleau LJ, Goodwin PJ. Three methods for minimally important difference: no relationship was found with the net proportion of patients improving. *J Clin Epidemiol*. 2007;60(5):448–455. doi:10.1016/j.jclinepi.2006.08.006.

36. Musselman KE. Clinical significance testing in rehabilitation research: what, why, and how? *Phys Ther Rev*. 2007;12(4):287–296.

37. Yost KJ, Sorensen MV, Hahn EA, Glendenning GA, Gnanasakthy A, Cella D. *Using Multiple Anchor- and Distribution-Based Estimates to Evaluate Clinically Meaningful Change on the Functional Assessment of Cancer Therapy-Biologic Response Modifiers (FACT-BRM) Instrument.*; 2005:117–127. doi:10.1111/j.1524-4733.2005.08202.x.

38. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol*. 2003;56(5):395–407.

39. Bago J, Pérez-Grueso FJS, Les E, Hernández P, Pellisé F. Minimal important differences of the SRS-22 Patient Questionnaire following surgical treatment of idiopathic scoliosis. *Eur Spine J*. 2009;18(12):1898–1904. doi:10.1007/s00586-009-1066-x.

40. Carreon LY, Sanders JO, Diab M, et al. The minimum clinically important difference in Scoliosis Research Society-22 Appearance, Activity, And Pain domains after surgical correction of adolescent idiopathic scoliosis. *Spine*. 2010;35(23):2079–2083. doi:10.1097/BRS.0b013e3181c61fd7.

41. Negrini S, Aulisa AG, Aulisa L, et al. 2011 SOSORT guidelines: Orthopaedic and Rehabilitation treatment of idiopathic scoliosis during growth. *Scoliosis*. 2012;7(1):3. doi:10.1186/1748-7161-7-3.

42. Scoliosis Research Society, ed. Adolescent Idiopathic Scoliosis Treatment. *http://wwwsrsorg*. 2011:1–12. Available at: http://www.srs.org/professionals/conditions_and_treatment/adolescent_idiopathic_scoliosis/treatment.htm. Accessed November 3, 2011.

43. Lehnert-Schroth C. Introduction to the three-dimensional scoliosis treatment according to Schroth. *Physiotherapy*. 1992;78(11):810–815.

44. Bialek M. Conservative treatment of idiopathic scoliosis according to FITS concept: presentation of the method and preliminary, short term radiological and clinical results based on SOSORT and SRS criteria. *Scoliosis*. 2011;6(1):25. doi:10.1186/1748-7161-6-25.

45. Negrini S, Atanasio S, Fusco C, Zaina F. Effectiveness of complete conservative treatment for adolescent idiopathic scoliosis (bracing and exercises) based on SOSORT management criteria: results according to the SRS criteria for bracing studies - SOSORT Award 2009 Winner. *Scoliosis*. 2009;4:19. doi:10.1186/1748-7161-4-19.

46. Fayssoux RS, Cho RH, Herman MJ. A history of bracing for idiopathic scoliosis in North America. *Clin Orthop Relat Res*. 2010;468(3):654–664. doi:10.1007/s11999-009-0888-5.

47. Kadoury S, Cheriet F, Beauséjour M, Stokes IA, Parent S, Labelle H. A three-dimensional retrospective analysis of the evolution of spinal instrumentation for the correction of adolescent idiopathic scoliosis. *Eur Spine J*. 2009;18(1):23–37. doi:10.1007/s00586-008-0817-4.

48. Hoashi JS, Cahill PJ, Bennett JT, Samdani AF. Adolescent scoliosis classification and treatment. *Neurosurg Clin N Am*. 2013;24(2):173–183. doi:10.1016/j.nec.2012.12.006.

49. Donnelly MJ, Dolan LA, Grande L, Weinstein SL. Patient and parent perspectives on treatment for adolescent idiopathic scoliosis. *Iowa Orthop J*. 2004;24:76–83.

50. Westrick ER, Ward WT. Adolescent idiopathic scoliosis: 5-year to 20-year evidence-based surgical results. *J Pediatr Orthop*. 2011;31(1 Suppl):S61–8. doi:10.1097/BPO.0b013e3181fd87d5.

51. Tones M, Moss N, Polly DW. A review of quality of life and psychosocial issues in scoliosis. *Spine*. 2006;31(26):3027–3038. doi:10.1097/01.brs.0000249555.87601.fc.

52. Danielsson AJ, Hasserius R, Ohlin A, Nachemson AL. Health-related quality of life in untreated versus brace-treated patients with adolescent idiopathic scoliosis: a long-term follow-up. *Spine*. 2010;35(2):199–205. doi:10.1097/BRS.0b013e3181c89f4a.

53. Lonstein JE, Carlson JM. The prediction of curve progression in untreated idiopathic scoliosis during growth. *J Bone Joint Surg Am*. 1984;66(7):1061–1071.

54. Focarile FA, Bonaldi A, Giarolo MA, Ferrari U, Zilioli E, Ottaviani C. Effectiveness of nonsurgical treatment for idiopathic scoliosis. Overview of available evidence. *Spine*. 1991;16(4):395–401.

55. Brooks HL, Azen SP, Gerberg E, Brooks R, Chan L. Scoliosis: A prospective epidemiological study. *J Bone Joint Surg Am*. 1975;57(7):968–972.

56. Wong H-K, Tan K-J. The natural history of adolescent idiopathic scoliosis. *Indian J Orthop*. 2010;44(1):9–13. doi:10.4103/0019-5413.58601.

57. Danielsson AJ, Nachemson AL. Radiologic findings and curve progression 22 years after treatment for adolescent idiopathic scoliosis: comparison of brace and surgical treatment with matching control group of straight individuals. *Spine*. 2001;26(5):516–525.

58. Danielsson AJ, Nachemson AL. Back pain and function 23 years after fusion for adolescent idiopathic scoliosis: a case-control study-part II. *Spine*. 2003;28(18):E373–83. doi:10.1097/01.BRS.0000084267.41183.75.

59. Danielsson AJ. What impact does spinal deformity correction for adolescent idiopathic scoliosis make on quality of life? *Spine*. 2007;32(19 Suppl):S101–8. doi:10.1097/BRS.0b013e318134ed0e.

60. Helenius I, Remes V, Lamberg T, Schlenzka D, Poussa M. Long-term health-related quality of life after surgery for adolescent idiopathic scoliosis and spondylolisthesis. *J Bone Joint Surg Am*. 2008;90(6):1231–1239. doi:10.2106/JBJS.G.00114.

61. Haefeli M, Elfering A, Kilian R, Min K, Boos N. Nonoperative treatment for adolescent idiopathic scoliosis: a 10- to 60-year follow-up with special reference to health-related quality of life. *Spine*. 2006;31(3):355–66; discussion 367. doi:10.1097/01.brs.0000197664.02098.09.

62. Danielsson AJ, Nachemson AL. Childbearing, curve progression, and sexual function in women 22 years after treatment for adolescent idiopathic scoliosis: a case-control study.

Spine. 2001;26(13):1449–1456.

63. Danielsson AJ, Nachemson AL. Back pain and function 22 years after brace treatment for adolescent idiopathic scoliosis: a case-control study-part I. *Spine*. 2003;28(18):2078–85; discussion 2086. doi:10.1097/01.BRS.0000084268.77805.6F.

64. Danielsson AJ, Romberg K, Nachemson AL. Spinal range of motion, muscle endurance, and back pain and function at least 20 years after fusion or brace treatment for adolescent idiopathic scoliosis: a case-control study. *Spine*. 2006;31(3):275–283. doi:10.1097/01.brs.0000197652.52890.71.

65. Weinstein SL, Zavala DC, Ponseti IV. Idiopathic scoliosis: long-term follow-up and prognosis in untreated patients. *J Bone Joint Surg Am*. 1981;63(5):702–712.

66. Cochran T, Irstam L, Nachemson A. Long-term anatomic and functional changes in patients with adolescent idiopathic scoliosis treated by Harrington rod fusion. *Spine*. 1983;8(6):576–584.

67. Maruyama T. Bracing adolescent idiopathic scoliosis: a systematic review of the literature of effective conservative treatment looking for end results 5 years after weaning. *Disabil Rehabil*. 2008;30(10):786–791. doi:10.1080/09638280801889782.

68. Danielsson AJ, Hasserius R, Ohlin A, Nachemson AL. A prospective study of brace treatment versus observation alone in adolescent idiopathic scoliosis: a follow-up mean of 16 years after maturity. *Spine*. 2007;32(20):2198–2207. doi:10.1097/BRS.0b013e31814b851f.

69. Dolan LA, Weinstein SL. Surgical rates after observation and bracing for adolescent idiopathic scoliosis: an evidence-based review. *Spine*. 2007;32(19 Suppl):S91–S100. doi:10.1097/BRS.0b013e318134ead9.

70. Fällström K, Cochran T, Nachemson A. Long-term effects on personality development in patients with adolescent idiopathic scoliosis. Influence of type of treatment. *Spine*. 1986;11(7):756–758.

71. Negrini S, Negrini A, Romano M, Verzini N, Negrini A, Parzini S. A controlled prospective study on the efficacy of SEAS.02 exercises in preventing progression and bracing in mild idiopathic scoliosis. *Stud Health Technol Inform*. 2006;123:523–526.

72. Wan L, Gx W, Bian R. Exercise therapy in treatment of essential S-shaped scoliosis: evaluation of Cobb angle in breast and lumbar segment through a follow-up of half a year. *Chinese Journal of Clinical Rehabilitation*. 2005;9(34):82–84.

73. Otman S, Kose N, Yakut Y. The efficacy of Schroth s 3-dimensional exercise therapy in the treatment of adolescent idiopathic scoliosis in Turkey. *Saudi Med J*. 2005;26(9):1429–1435.

74. Effects of novel corrective spinal technique on adolescent idiopathic scoliosis as

assessed by radiographic imaging. 2013:1–8. doi:10.3233/BMR-130452.

75. McIntire KL, Asher MA, Burton DC, Liu W. Treatment of adolescent idiopathic scoliosis with quantified trunk rotational strength training: a pilot study. *J Spinal Disord Tech*. 2008;21(5):349–358. doi:10.1097/BSD.0b013e318145b7e9.

76. Mooney V, Gulick J, Pozos R. A preliminary report on the effect of measured strength training in adolescent idiopathic scoliosis. *J Spinal Disord*. 2000;13(2):102–107.

77. Weiss H-R. Influence of an in-patient exercise program on scoliotic curve. *Ital J Orthop Traumatol*. 1992;18(3):395–406.

78. Negrini S, Atanasio S, Zaina F, Romano M, Parzini S, Negrini A. End-growth results of bracing and exercises for adolescent idiopathic scoliosis. Prospective worst-case analysis. *Stud Health Technol Inform*. 2008;135:395–408.

79. Lenssinck M-LB, Frijlink AC, Berger MY, Bierman-Zeinstra SMA, Verkerk K, Verhagen AP. Effect of bracing and other conservative interventions in the treatment of idiopathic scoliosis in adolescents: a systematic review of clinical trials. *Phys Ther*. 2005;85(12):1329–1339.

80. Romano M, Minozzi S, Bettany-Saltikov J, et al. Exercises for adolescent idiopathic scoliosis. *Cochrane Database Syst Rev*. 2012;8:CD007837. doi:10.1002/14651858.CD007837.pub2.

81. Negrini S, Antonini G, Carabalona R, Minozzi S. Physical exercises as a treatment for adolescent idiopathic scoliosis. A systematic review. *Pediatr Rehabil*. 2003;6(3-4):227–235. doi:10.1080/13638490310001636781.

82. Negrini S, Atanasio S, Zaina F, Romano M. Rehabilitation of adolescent idiopathic scoliosis: results of exercises and bracing from a series of clinical studies. Europa Medicophysica-SIMFER 2007 Award Winner. *Eur J Phys Rehabil Med*. 2008;44(2):169–176.

83. Negrini S, Fusco C, Minozzi S, Atanasio S, Zaina F, Romano M. Exercises reduce the progression rate of adolescent idiopathic scoliosis: results of a comprehensive systematic review of the literature. *Disabil Rehabil*. 2008;30(10):772–785. doi:10.1080/09638280801889568.

84. Solberg G. Plastic changes in spinal function of pre-pubescent scoliotic children engaged in an exercise therapy programme. *South African Journal of Physiotherapy*. 1996;52(1):19–2224.

85. Maruyama T, Kitagawa T, Takeshita K, Mochizuki K, Nakamura K. Conservative treatment for adolescent idiopathic scoliosis: can it reduce the incidence of surgical treatment? *Pediatr Rehabil*. 2003;6(3-4):215–219. doi:10.1080/13638490310001642748.

86. Weiss H-R, Weiss G, Schaar H-J. Incidence of surgery in conservatively treated patients with scoliosis. *Pediatr Rehabil*. 2003;6(2):111–118. doi:10.1080/13638490310001593446.

87. Goldberg C, Moore D, Fogarty E. Adolescent idiopathic scoliosis: the effect of brace treatment on the incidence of surgery. *Spine*. 2001.

88. Goldberg CJ, Moore DP, Fogarty EE, Dowling FE. Adolescent idiopathic scoliosis: the effect of brace treatment on the incidence of surgery. *Spine*. 2001;26(1):42–47.

89. Bulthuis GJ, Veldhuizen AG, Nijenbanning G. Clinical effect of continuous corrective force delivery in the non-operative treatment of idiopathic scoliosis: a prospective cohort study of the TriaC-brace. *Eur Spine J*. 2008;17(2):231–239. doi:10.1007/s00586-007-0513-9.

90. Miller JA, Nachemson AL, Schultz AB. Effectiveness of braces in mild idiopathic scoliosis. *Spine*. 1984;9(6):632–635.

91. Wiley JW, Thomson JD, Mitchell TM, Smith BG, Banta JV. Effectiveness of the boston brace in treatment of large curves in adolescent idiopathic scoliosis. *Spine*. 2000;25(18):2326–2332.

92. Negrini S, Marchini G. Efficacy of the symmetric, patient-oriented, rigid, three-dimensional, active (SPoRT) concept of bracing for scoliosis: a prospective study of the Sforzesco versus Lyon brace. *Eura Medicophys*. 2007;43(2):171–81; discussion 183–4.

93. Lange JE, Steen H, Brox JI. Long-term results after Boston brace treatment in adolescent idiopathic scoliosis. *Scoliosis*. 2009;4:17. doi:10.1186/1748-7161-4-17.

94. MacLean WE, Green NE, Pierre CB, Ray DC. Stress and coping with scoliosis: psychological effects on adolescents and their families. *J Pediatr Orthoped*. 1989;9(3):257–261.

95. Shaughnessy WJ. Advances in scoliosis brace treatment for adolescent idiopathic scoliosis. *Orthop Clin North Am*. 2007;38(4):469–75, v. doi:10.1016/j.ocl.2007.07.002.

96. Sponseller PD. Bracing for adolescent idiopathic scoliosis in practice today. *J Pediatr Orthop*. 2011;31(1 Suppl):S53–60. doi:10.1097/BPO.0b013e3181f73e87.

97. Sanders JO, Khoury JG, Kishan S, et al. Predicting scoliosis progression from skeletal maturity: a simplified classification during adolescence. *The Journal of Bone and Joint Surgery*. 2008;90(3):540–553. doi:10.2106/JBJS.G.00004.

98. Janicki JA, Poe-Kochert C, Armstrong DG, Thompson GH. A comparison of the thoracolumbosacral orthoses and providence orthosis in the treatment of adolescent idiopathic scoliosis: results using the new SRS inclusion and assessment criteria for bracing studies. *J Pediatr Orthoped*. 2007;27(4):369–374. doi:10.1097/01.bpb.0000271331.71857.9a.

99. Weinstein SL, Dolan LA, Wright JG, Dobbs MB. Effects of bracing in adolescents with idiopathic scoliosis. *N Engl J Med*. 2013;369(16):1512–1521. doi:10.1056/NEJMoa1307337.

100. Weinstein SL, Dolan LA, Wright JG, Dobbs MB. Effects of Bracing in Adolescents with Idiopathic Scoliosis (Supplementary Appendix). *N Engl J Med*. 2013;369(16):1512–1521. doi:10.1056/NEJMoa1307337.

101. Katz DE, Herring JA, Browne RH, Kelly DM, Birch JG. Brace wear control of curve progression in adolescent idiopathic scoliosis. *The Journal of Bone and Joint Surgery*. 2010;92(6):1343–1352. doi:10.2106/JBJS.I.01142.

102. Hawes MC, O'Brien JP. A century of spine surgery: what can patients expect? *Disabil Rehabil*. 2008;30(10):808–817. doi:10.1080/09638280801889972.

103. Gitelman Y, Lenke LG, Bridwell KH, Auerbach JD, Sides BA. Pulmonary Function in Adolescent Idiopathic Scoliosis Relative to the Surgical Procedure A 10-Year Follow-up Analysis. *Spine*. 2011;36(20):1665–1672. doi:10.1097/BRS.0b013e31821bcf4c.

104. Weiss H-R, Bess S, Wong MS, Patel V, Goodall D, Burger E. Adolescent idiopathic scoliosis - to operate or not? A debate article. *Patient safety in surgery*. 2008;2(1):25.

105. Hawes M. Impact of spine surgery on signs and symptoms of spinal deformity. *Pediatr Rehabil*. 2006;9(4):318–339.

106. Dangerfield PH. The classification of spinal deformities. *Pediatr Rehabil*. 2003;6(3-4):133–136. doi:10.1080/13638490310001649543.

107. Ponseti IV, Friedman B. Prognosis in idiopathic scoliosis. *J Bone Joint Surg Am*. 1950;32A(2):381–395.

108. Rigo MD, Villagrasa M, Gallo D. A specific scoliosis classification correlating with brace treatment: description and reliability. *Scoliosis*. 2010;5(1):1. doi:10.1186/1748-7161-5-1.

109. King HA, Moe JH, Bradford DS, Winter RB. The selection of fusion levels in thoracic idiopathic scoliosis. *J Bone Joint Surg Am*. 1983;65(9):1302–1313.

110. Rigo M. Patient evaluation in idiopathic scoliosis: Radiographic assessment, trunk deformity and back asymmetry. *Physiother Theory Pract*. 2011;27(1):7–25. doi:10.3109/09593985.2010.503990.

111. Hefti F, Brunner R, Hasler CC, Jundt G. *Pediatric Orthopedics in Practice*. 1st ed. Springer; 2007.

112. Cummings RJ, Loveless EA, Campbell J, Samelson S, Mazur JM. Interobserver reliability and intraobserver reproducibility of the system of King et al. for the classification of adolescent idiopathic scoliosis. *J Bone Joint Surg Am*.

1998;80(8):1107–1111.

113. Ogon M, Giesinger K, Behensky H, et al. Interobserver and intraobserver reliability of Lenke's new scoliosis classification system. *Spine*. 2002;27(8):858–862.

114. Lenke LG, Betz RR, Harms J, et al. Adolescent idiopathic scoliosis: a new classification to determine extent of spinal arthrodesis. *J Bone Joint Surg Am*. 2001;83-A(8):1169–1181.

115. Qiu G, Zhang J, Wang Y, et al. A New Operative Classification of Idiopathic Scoliosis: A Peking Union Medical College Method. *Spine*. 2005;30(12):1419–1426.

116. Smith JS, Shaffrey CI, Kuntz C, Mummaneni PV. Classification systems for adolescent and adult scoliosis. *Neurosurgery*. 2008;63(3 Suppl):16–24. doi:10.1227/01.NEU.0000320447.61835.EA.

117. Qiu G, Li Q, Wang Y, et al. Comparison of reliability between the PUMC and Lenke classification systems for classifying adolescent idiopathic scoliosis. *Spine*. 2008;33(22):E836–42. doi:10.1097/BRS.0b013e318187bb10.

118. Wampold BE, Jenson WR. ScienceDirect - Behavior Therapy, Volume 17, Issue 3, Pages 197-312 (June 1986). *Behavior Therapy*.:302–311.

119. Lipsey MW, Wilson DB. The efficacy of psychological, educational, and behavioral treatment. Confirmation from meta-analysis. *Am Psychol*. 1993;48(12):1181–1209.

120. Guyatt G, Jaeschke R, Heddle N, Cook D, Shannon H, Walter S. Basic statistics for clinicians: 2. Interpreting study results: confidence intervals. *CMAJ*. 1995;152(2):169–173.

121. Guyatt GH, Kirshner B, Jaeschke R. Measuring health status: what are the necessary measurement properties? *J Clin Epidemiol*. 1992;45(12):1341–1345.

122. Atkins DC, Bedics JD, McGlinchey JB, Beauchaine TP. Assessing clinical significance: does it matter which method we use? *J Consult Clin Psychol*. 2005;73(5):982–989. doi:10.1037/0022-006X.73.5.982.

123. Wyrwich KW, Nienaber NA, Tierney WM, Wolinsky FD. *Linking Clinical Relevance and Statistical Significance in Evaluating Intra-Individual Changes in Health-Related Quality of Life.*; 1999:469–478.

124. Jacobson NS, Roberts LJ, Berns SB, McGlinchey JB. Methods for defining and determining the clinical significance of treatment effects: description, application, and alternatives. *J Consult Clin Psychol*. 1999;67(3):300–307.

125. Rushton PRP, Grevitt MP. Comparison of untreated adolescent idiopathic scoliosis with normal controls: a review and statistical analysis of the literature. *Spine*. 2013;38(9):778–785. doi:10.1097/BRS.0b013e31827db418.

126. Gwet K. *Handbook of Inter-Rater Reliability*. 2nd ed. Gaithersburg, MD, USA: Advanced Analytics, LLC; 2010.

127. Asher MA, Burton DC. Adolescent idiopathic scoliosis: natural history and long term treatment effects. *Scoliosis*. 2006;1(1):2. doi:10.1186/1748-7161-1-2.

128. Negrini S, Aulisa I, Ferraro C, et al. Italian guidelines on rehabilitation treatment of adolescents with scoliosis or other spinal deformities. *Europa Medicophysica*. 2005;41(2):183–201.

129. Weiss H-R, Werkmann M. "Brace Technology" Thematic Series - The ScoliOlogiC® Chêneau light™ brace in the treatment of scoliosis. *Scoliosis*. 2010;5:19. doi:10.1186/1748-7161-5-19.

130. Stokes IAF, Aronsson DD. Computer-assisted algorithms improve reliability of King classification and Cobb angle measurement of scoliosis. *Spine*. 2006;31(6):665–670. doi:10.1097/01.brs.0000203708.49972.ab.

131. Gwet KL. Estimating the Number of Subjects and Number Raters When Designing an Inter-Rater Reliability Study. *http://agreestatcom/blog_irr/sample_size_determinationhtml*. 2013. Accessed November 1, 2014.

132. Cohen J. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*. 1960;20(1):37–46.

133. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. *J Clin Epidemiol*. 1990;43(6):543–549.

134. Gwet KL. Computing inter-rater reliability and its variance in the presence of high agreement. *Br J Math Stat Psychol*. 2008;61(Pt 1):29–48.

135. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159–174.

136. Phan P, Mezghani N, Nault M-L, et al. A decision tree can increase accuracy when assessing curve types according to Lenke classification of adolescent idiopathic scoliosis. *Spine*. 2010;35(10):1054–1059. doi:10.1097/BRS.0b013e3181bf280e.

137. Lenke LG, Betz RR, Clements D, et al. Curve prevalence of a new classification of operative adolescent idiopathic scoliosis: does classification correlate with treatment? *Spine*. 2002;27(6):604–611.

138. Minick KI, Kiesel KB, Burton L, Taylor A, Plisky P, Butler RJ. Interrater reliability of the functional movement screen. *J Strength Cond Res*. 2010;24(2):479–486. doi:10.1519/JSC.0b013e3181c09c04.

139. Smits-Engelsman BC, Fiers MJ, Henderson SE, Henderson L. Interrater reliability of the

Movement Assessment Battery for Children. *Phys Ther*. 2008;88(2):286–294. doi:10.2522/ptj.20070068.

140. Sutherland CA, Albert WJ, Wrigley AT, Callaghan JP. The effect of camera viewing angle on posture assessment repeatability and cumulative spinal loading. *Ergonomics*. 2007;50(6):877–889. doi:10.1080/00140130701237675.

141. Schlösser TPC, van der Heijden GJMG, Versteeg AL, Castelein RM. How "idiopathic" is adolescent idiopathic scoliosis? A systematic review on associated abnormalities. *PLoS ONE*. 2014;9(5):e97461. doi:10.1371/journal.pone.0097461.

142. Konieczny MR, Senyurt H, Krauspe R. Epidemiology of adolescent idiopathic scoliosis. *J Child Orthop*. 2013;7(1):3–9. doi:10.1007/s11832-012-0457-4.

143. Kotwicki T, Durmała J, Czaprowski D, et al. Conservative management of idiopathic scoliosis--guidelines based on SOSORT 2006 Consensus. *Ortop Traumatol Rehabil*. 2009;11(5):379–395.

144. Refsum HE, Næss-Andresen CF, Lange JE. Pulmonary Function and Gas Exchange at Rest and Exercise in Adolescent Girls with Mild Idiopathic Scoliosis during Treatment with Boston Thoracic Brace. *Spine*. 1990;15(5):420.

145. Negrini S, Carabalona R. Social acceptability of treatments for adolescent idiopathic scoliosis: a cross-sectional study. *Scoliosis*. 2006;1:14. doi:10.1186/1748-7161-1-14.

146. Negrini S. Approach to scoliosis changed due to causes other than evidence: patients call for conservative (rehabilitation) experts to join in team orthopedic surgeons. *Disabil Rehabil*. 2008;30(10):731–741. doi:10.1080/09638280801889485.

147. Mordecai SC, Dabke HV. Efficacy of exercise therapy for the treatment of adolescent idiopathic scoliosis: a review of the literature. *Eur Spine J*. 2011;21(3):382–389. doi:10.1007/s00586-011-2063-4.

148. Schreiber S, Parent EC, Watkins EM, Hedden DM. An algorithm for determining scoliosis curve type according to Schroth. *Scoliosis*. 2012;7(Suppl 1):O53. doi:10.1186/1748-7161-7-S1-O53.

149. Watkins EM, Bosnjak S, Parent EC. Algorithms to prescribe Schroth exercises for each of four Schroth curve types. *Scoliosis*. 2012;7(Suppl 1):P22. doi:10.1186/1748-7161-7-S1-P22.

150. Zhang J, Lou E, Shi X, et al. A computer-aided Cobb angle measurement method and its reliability. *J Spinal Disord Tech*. 2010;23(6):383–387. doi:10.1097/BSD.0b013e3181bb9a3c.

151. Dishman RK, Motl RW, Saunders RP, et al. Factorial invariance and latent mean structure of questionnaires measuring social-cognitive determinants of physical activity among black and white adolescent girls. *Prev Med*. 2002;34(1):100–108.

doi:10.1006/pmed.2001.0959.

152. Genolini C, Écochard R, Jacqmin-Gadda H. Copy Mean: A New Method to Impute Intermittent Missing Values in Longitudinal Studies. *OJS*. 2013;03(04):26–40. doi:10.4236/ojs.2013.34A004.

153. Venables WN, Ripley BD. *Modern Applied Statistics with S*. Springer Science & Business Media; 2002.

154. Fitzmaurice GM, Laird NM, Ware JH. *Applied Longitudinal Analysis*. 2nd ed. John Wiley & Sons; 2011.

155. Team TRDC. *R: a Language and Environment for Statistical Computing*. 2010:1–1731.

156. Sanders JO, Browne RH, McConnell SJ, Margraf SA, Cooney TE, Finegold DN. Maturity Assessment and Curve Progression in Girls with Idiopathic Scoliosis. *The Journal of Bone and Joint Surgery*. 2007;89(1):64–73. doi:10.2106/JBJS.F.00067.

157. Wan L, Gx W, Bian R. Exercise therapy in treatment of essential S-shaped scoliosis: evaluation of Cobb angle in breast and lumbar segment through a follow-up of half a year. *Chinese Journal of Clinical Rehabilitation*. 2005;9(34):82–84.

158. Monticone M, Ambrosini E, Cazzaniga D, Rocca B, Ferrante S. Active self-correction and task-oriented exercises reduce spinal deformity and improve quality of life in subjects with mild adolescent idiopathic scoliosis. Results of a randomised controlled trial. *Eur Spine J*. 2014;23(6):1204–1214. doi:10.1007/s00586-014-3241-y.

159. Negrini S, Zaina F, Romano M, Negrini A, Parzini S. Specific exercises reduce brace prescription in adolescent idiopathic scoliosis: a prospective controlled cohort study with worst-case analysis. *J Rehabil Med*. 2008;40(6):451–455. doi:10.2340/16501977-0195.

160. Noh DK, You JS-H, Koh J-H, et al. Effects of novel corrective spinal technique on adolescent idiopathic scoliosis as assessed by radiographic imaging. *J Back Musculoskelet Rehabil*. 2014;27(3):331–338. doi:10.3233/BMR-130452.

161. Kahanovitz N, Snow B, Pinter I. The comparative results of psychologic testing in scoliosis patients treated with electrical stimulation or bracing. *Spine*. 1984;9(5):442–444.

162. Payne WK III, Ogilvie JW, Resnick MD, Kane RL, Transfeldt EE, Blum RW. Does scoliosis have a psychological impact and does gender make a difference? *Spine*. 1997;22(12):1380.

163. Upasani VV, Caltoum C, Petcharaporn M, et al. Adolescent idiopathic scoliosis patients report increased pain at five years compared with two years after surgical treatment. *Spine*. 2008;33(10):1107–1112. doi:10.1097/BRS.0b013e31816f2849.

164. Sato T, Hirano T, Ito T, et al. Back pain in adolescents with idiopathic scoliosis:

epidemiological study for 43,630 pupils in Niigata City, Japan. *Eur Spine J*. 2011;20(2):274–279. doi:10.1007/s00586-010-1657-6.

165. Martínez-Llorens J, Ramírez M, Colomina MJ, et al. Muscle dysfunction and exercise limitation in adolescent idiopathic scoliosis. *Eur Respir J*. 2010;36(2):393–400. doi:10.1183/09031936.00025509.

166. Weiss H-R, Negrini S, Rigo M, et al. Indications for conservative management of scoliosis (guidelines). *Scoliosis*. 2006;1:5. doi:10.1186/1748-7161-1-5.

167. Santos Alves Dos VL, Alves da Silva RJAL, Avanzi O. Effect of a preoperative protocol of aerobic physical therapy on the quality of life of patients with adolescent idiopathic scoliosis: a randomized clinical study. *Am J Orthop*. 2014;43(6):E112–6.

168. Schreiber S, Parent EC, Watkins EM, Hedden DM. An algorithm for determining scoliosis curve type according to Schroth. *Scoliosis*. 2012;7(Suppl 1):O53. doi:10.1186/1748-7161-7-S1-O53.

169. Friedrich M, Gittler G, Halberstadt Y, Cermak T, Heiller I. Combined exercise and motivation program: effect on the compliance and level of disability of patients with chronic low back pain: a randomized controlled trial. *Arch Phys Med Rehabil*. 1998;79(5):475–487.

170. Strecher V. Improving physician-patient interactions: A review. *Patient Counselling and Health Education*. 1983;4(3):129–136. doi:10.1016/S0190-2040(83)80002-0.

171. Emans JB, Hedequist D, Miller R, et al. Reference Manual For The Boston Scoliosis Brace. *Scoliosis Research Society*. 2003. Available at: http://www.srs.org/professionals/education_materials/SRS_bracing_manual/section5.pdf . Accessed May 16, 2011.

172. Biering-Sørensen F. Physical measurements as risk indicators for low-back trouble over a one-year period. *Spine*. 1984;9(2):106–119.

173. Latimer J, Maher CG, Refshauge K, Colaco I. The reliability and validity of the Biering–Sorensen test in asymptomatic subjects and subjects reporting current or previous nonspecific low back pain. *Spine*. 1999;24(20):2085.

174. Coorevits P, Danneels L, Cambier D, Ramon H, Vanderstraeten G. Assessment of the validity of the Biering-Sørensen test for measuring back muscle fatigue based on EMG median frequency characteristics of back and hip muscles. *J Electromyogr Kinesiol*. 2008;18(6):997–1005. doi:10.1016/j.jelekin.2007.10.012.

175. Asher MA, Min Lai S, Glattes RC, Burton DC, Alanay A, Bago J. Refinement of the SRS-22 Health-Related Quality of Life questionnaire Function domain. *Spine*. 2006;31(5):593–597. doi:10.1097/01.brs.0000201331.50597.ea.

176. Glattes RC, Burton DC, Lai SM, Frasier E, Asher MA. The Reliability and Concurrent

Validity of the Scoliosis Research Society-22r Patient Questionnaire Compared With the Child Health Questionnaire-CF87 Patient Questionnaire for Adolescent Spinal Deformity. *Spine*. 2007;32(16):1778–1784. doi:10.1097/BRS.0b013e3180dc9bb2.

177. Negrini S, Grivas TB, Kotwicki T, et al. Why do we treat adolescent idiopathic scoliosis? What we want to obtain and to avoid for our patients. SOSORT 2005 Consensus paper. *Scoliosis*. 2006;1:4. doi:10.1186/1748-7161-1-4.

178. Sanders JO, Harrast JJ, Kuklo TR, et al. The Spinal Appearance Questionnaire: Results of Reliability, Validity, and Responsiveness Testing in Patients With Idiopathic Scoliosis. *Spine*. 2007;32(24):2719–2722. doi:10.1097/BRS.0b013e31815a5959.

179. Motl RW, Dishman RK, Trost SG, et al. Factorial validity and invariance of questionnaires measuring social-cognitive determinants of physical activity among adolescent girls. *Prev Med*. 2000;31(5):584–594. doi:10.1006/pmed.2000.0735.

180. Dishman RK, Saunders RP, Motl RW, Dowda M, Pate RR. Self-efficacy moderates the relation between declines in physical activity and perceived social support in high school girls. *J Pediatr Psychol*. 2009;34(4):441–451. doi:10.1093/jpepsy/jsn100.

181. Motl RW, Dishman RK, Ward DS, et al. Comparison of barriers self-efficacy and perceived behavioral control for explaining physical activity across 1 year among adolescent girls. *Health Psychol*. 2005;24(1):106–111. doi:10.1037/0278-6133.24.1.106.

182. Ersberg A, Gerdhem P. Pre- and postoperative quality of life in patients treated for scoliosis. *Acta Orthop*. 2013;84(6):537–543. doi:10.3109/17453674.2013.854667.

183. Parent EC, Dang R, Hill D, et al. Score distribution of the scoliosis research society-22 questionnaire in subgroups of patients of all ages with idiopathic scoliosis. *Spine*. 2010;35(5):568–577. doi:10.1097/BRS.0b013e3181b9c9c0.

184. Roy-Beaudry M, Beauséjour M, Joncas J, et al. Validation and Clinical Relevance of a French-Canadian Version of the Spinal Appearance Questionnaire in Adolescent Patients. *Spine*. 2011;36(9):746–751. doi:10.1097/BRS.0b013e3181e040e7.

185. Negrini S, Donzelli S, Dulio M, Zaina F. Is the SRS-22 able to detect Quality of Life (QoL) changes during conservative treatments ? *Stud Health Technol Inform*. 2012;176:433–436.

186. Parent EC, Hill D, Moreau M, Mahood J, Raso J, Lou E. Score distribution of the Scoliosis Quality of Life Index questionnaire in different subgroups of patients with adolescent idiopathic scoliosis. *Spine*. 2007;32(16):1767–1777. doi:10.1097/BRS.0b013e3180b9f7a5.

187. Bago J, Sanchez-Raya J, Perez-Grueso FJS, Climent JM. The Trunk Appearance Perception Scale (TAPS): a new tool to evaluate subjective impression of trunk deformity in patients with idiopathic scoliosis. *Scoliosis*. 2010;5(1):6. doi:10.1186/1748-7161-5-6.

188. Zaina F, Negrini S, Atanasio S. TRACE (Trunk Aesthetic Clinical Evaluation), a routine clinical tool to evaluate aesthetics in scoliosis patients: development from the Aesthetic Index (AI) and repeatability. *Scoliosis*. 2009;4:3. doi:10.1186/1748-7161-4-3.

189. Danielsson AJ, Hasserius R, Ohlin A, Nachemson AL. Body appearance and quality of life in adult patients with adolescent idiopathic scoliosis treated with a brace or under observation alone during adolescence. *Spine*. 2012;37(9):755–762.

190. Mulcahey MJ, Chafetz RS, Santangelo AM, et al. Cognitive testing of the spinal appearance questionnaire with typically developing youth and youth with idiopathic scoliosis. *J Pediatr Orthop*. 2011;31(6):661–667. doi:10.1097/BPO.0b013e318221ea8b.

191. D'Andrea LP, Betz RR, Lenke LG, et al. Do radiographic parameters correlate with clinical outcomes in adolescent idiopathic scoliosis? *Spine*. 2000;25(14):1795–1802.

192. Parent EC, Wong D, Hill D, et al. The association between Scoliosis Research Society-22 scores and scoliosis severity changes at a clinically relevant threshold. *Spine*. 2010;35(3):315–322. doi:10.1097/BRS.0b013e3181cabe75.

193. Revicki D, Hays R, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol*. 2008;61(2):102–109. doi:10.1016/j.jclinepi.2007.03.012.

194. Cella D, Hahn EA, Dineen K. Meaningful change in cancer-specific quality of life scores: differences between improvement and worsening. *Qual Life Res*. 2002;11(3):207–221.

195. Portney LG, Watkins MP. *Foundations of Clinical Research*. Prentice Hall; 2009.

196. Stratford PW, Binkley J, Solomon P, Finch E, Gill C, Moreland J. Defining the minimum level of detectable change for the Roland-Morris questionnaire. *Phys Ther*. 1996;76(4):359–65; discussion 366–8.

197. Verma K, Lonner B, Hoashi JS, et al. Demographic factors affect Scoliosis Research Society-22 performance in healthy adolescents: a comparative baseline for adolescents with idiopathic scoliosis. *Spine*. 2010;35(24):2134–2139. doi:10.1097/BRS.0b013e3181cb474f.

198. Schreiber S, Parent EC, Hedden DM, Moreau M, Hill D, Lou E. Effect of Schroth exercises on curve characteristics and clinical outcomes in adolescent idiopathic scoliosis: protocol for a multicentre randomised controlled trial. *Journal of Physiotherapy*. 2014;60(4):234. doi:10.1016/j.jphys.2014.08.005.

199. Bridwell KH, Berven S, Glassman S, et al. Is the SRS-22 instrument responsive to change in adult scoliosis patients having primary spinal deformity surgery? *Spine*. 2007;32(20):2220–2225. doi:10.1097/BRS.0b013e31814cf120.

200. Asher M, Min Lai S, Burton D, Manna B. The reliability and concurrent validity of the

scoliosis research society-22 patient questionnaire for idiopathic scoliosis. *Spine*. 2003;28(1):63–69. doi:10.1097/01.BRS.0000047634.95839.67.

201. Bridwell KH, Cats-Baril W, Harrast J, et al. The validity of the SRS-22 instrument in an adult spinal deformity population compared with the Oswestry and SF-12: a study of response distribution, concurrent validity, internal consistency, and reliability. *Spine*. 2005;30(4):455–461.

202. Keller A, Hellesnes J, Brox JI. Reliability of the isokinetic trunk extensor test, Biering-Sørensen test, and Astrand bicycle test: assessment of intraclass correlation coefficient and critical difference in patients with chronic low back pain and healthy individuals. *Spine*. 2001;26(7):771–777.

203. Pitcher MJ, Behm DG, Mackinnon SN. Neuromuscular fatigue during a modified Biering-Sørensen test in subjects with and without low back pain. *J Sports Sci Med*. 2007;6(4):549–559.

204. Evans K, Refshauge KM, Adams R. Trunk muscle endurance tests: reliability, and gender differences in athletes. *J Sci Med Sport*. 2007;10(6):447–455. doi:10.1016/j.jsams.2006.09.003.

205. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials*. 1989;10(4):407–415.

206. Jaeschke R, Guyatt G, Shannon H, Walter S, Cook D, Heddle N. Basic statistics for clinicians: 3. Assessing the effects of treatment: measures of association. *CMAJ*. 1995;152(3):351–357.

207. Kamper SJ, Maher CG, Mackay G. Global rating of change scales: a review of strengths and weaknesses and considerations for design. *J Man Manip Ther*. 2009;17(3):163–170.

208. Parent EC, Westover L, Moreau MJ, Hedden DM, Hill D. Test-retest Reliability of Scoliosis Research Society 22r Quality of Life Questionnaire Scores in Adolescents with Idiopathic Scoliosis. *COA 2013 Annual Meeting Abstract Supplement*. 2013.

209. Eliasziw M, Young SL, Woodbury MG, Fryday-Field K. Statistical methodology for the concurrent assessment of interrater and intrarater reliability: using goniometric measurements as an example. *Phys Ther*. 1994;74(8):777–788.

210. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates; 1988.

211. Bender R. Calculating confidence intervals for the number needed to treat. *Control Clin Trials*. 2001;22(2):102–110.

212. Altman DG. Confidence intervals for the number needed to treat. *BMJ*. 1998;317(7168):1309–1312.

213. Newcombe RG. Interval estimation for the difference between independent proportions: comparison of eleven methods. *Stat Med*. 1998;17(8):873–890.

214. Tandberg D. Improved Confidence Intervals for the Difference Between Two Proportions and Number Needed to Treat (NNT). Microsoft Excel File, version 1.49. *Center for Evidence Based Medicine - cebmnet*. 2009.

215. Mannion AF, Taimela S, Müntener M, Dvorak J. Active therapy for chronic low back pain part 1. Effects on back muscle activation, fatigability, and strength. *Spine*. 2001;26(8):897–908.

216. Moffroid MT, Haugh LD, Haig AJ, Henry SM, Pope MH. Endurance training of trunk extensor muscles. *Phys Ther*. 1993;73(1):10–17.

217. Durall, C. J., Udermann, B. E., Johansen, D. R., Gibson, B., Reineke, D. M., & Reuteman, P. The effects of preseason trunk muscle training on low-back pain occurrence in women collegiate gymnasts. *Journal of Strength and Conditioning Research*, 2009;23(1):86–92. doi:10.1519/JSC.0b013e31818b93ac.

218. Morrissy RT, Goldsmith GS, Hall EC, Kehl D, Cowie GH. Measurement of the Cobb angle on radiographs of patients who have scoliosis. Evaluation of intrinsic error. *J Bone Joint Surg Am*. 1990;72(3):320–327.

219. Childs JD, Piva SR, Fritz JM. Responsiveness of the numeric pain rating scale in patients with low back pain. *Spine*. 2005;30(11):1331–1334.

220. Watanabe K, Hasegawa K, Hirano T, Uchiyama S, Endo N. Use of the scoliosis research society outcomes instrument to evaluate patient outcome in untreated idiopathic scoliosis patients in Japan: part II: relation between spinal deformity and patient outcomes. *Spine*. 2005;30(10):1202–1205.

221. Zaina F, Negrini S, Atanasio S. TRACE (Trunk Aesthetic Clinical Evaluation), a routine clinical tool to evaluate aesthetics in scoliosis patients: development from the Aesthetic Index (AI) and repeatability. *Scoliosis*. 2009;4:3. doi:10.1186/1748-7161-4-3.

222. Kotwicki, T., Durmała, J., Czaprowski, D., Głowacki, M., Kołban, M., Snela, S., et al. Conservative Management of Idiopathic Scoliosis – Guidelines Based on SOSORT 2006 Consensus. *Ortopedia, Traumatologia, Rehabilitacja, 11(5), 379–395.*

223. Mordecai SC, Dabke HV. Efficacy of exercise therapy for the treatment of adolescent idiopathic scoliosis: a review of the literature. *European Spine Journal*. 2012. doi:10.1007/s00586-011-2063-4.

224. Osoba D, Rodrigues G, Myles J, Zee B, Pater J. Interpreting the significance of changes in health-related quality-of-life scores. *J Clin Oncol*. 1998;16(1):139–144.

225. Weiss H-R. “Brace technology” thematic series - the Gensingen brace™ in the treatment of scoliosis. *Scoliosis*. 2010;5:22. doi:10.1186/1748-7161-5-22.

226. Rivard CH, Coillard C. SpineCor System. *Scoliosis Research Society*. 2002. Available at: http://www.srs.org/professionals/bracing_manuals/section13.pdf. Accessed May 6, 2011.

227. Richards BS, Bernstein RM, D'Amato CR, Thompson GH. Standardization of criteria for adolescent idiopathic scoliosis brace studies: SRS Committee on Bracing and Nonoperative Management. *Spine*. 2005;30(18):2068–75; discussion 2076–7.

228. Roy-Beaudry M, Beauséjour M, Joncas J, et al. Validation and Clinical Relevance of a French-Canadian Version of the Spinal Appearance Questionnaire in Adolescent Patients. *Spine*. 2010. doi:10.1097/BRS.0b013e3181e040e7.

229. Grivas TB, de Mauroy JC, Negrini S, et al. Terminology - glossary including acronyms and quotations in use for the conservative spinal deformities treatment: 8th SOSORT consensus paper. *Scoliosis*. 2010;5:23. doi:10.1186/1748-7161-5-23.

Appendices

Appendix 1

Operational definitions for Schroth Classification Algorithm

NOTE: The algorithm has been developed in a way so that it detects the presence or absence of certain category.

Main (major) curve:

The main curve is a structural scoliosis curvature, which is more pronounced and less flexible compared to other existent scoliotic curves in the spine of a patient with scoliosis. [21]According to the SOSORT terminology, "major scoliosis curvature is the curve with the largest Cobb measurement on upright long cassette coronal x-ray of the spine."[229] According to Schroth method, the main curve influences the body static the most. [21]

Apex:

The apex, or apical vertebra, is the most laterally displaced and rotated vertebra relative to the vertical axis of the body that passes through the patient's sacrum, and compared to the other vertebrae included in the scoliosis curve. [229]

Thoracic curve:

A thoracic scoliosis curvature is any spinal curvature in which the apex of the curve is between the second (T2) and eleventh thoracic (T11) vertebrae. [21,229]

Lumbar curve:

A lumbar scoliosis curve is a spinal curvature whose apex is between the first (L1) and fourth lumbar (L4) vertebrae (also known as lumbar scoliosis). [21,229]

Thoracolumbar curve:

A thoracolumbar curve is any spinal curvature that has its apex at the twelfth thoracic (T12) or first lumbar (L1) vertebra. [229]

Concave:

Curving inward or curved like the inner surface of a sphere. In scoliosis, the concave side is the inner side of the curve.[229]

Convex:

A surface or boundary that curves or bulges outward, as the exterior of a sphere. In scoliosis, the convex side is the outer side of the curve.[229]

Rib hump:

A rib hump is observed on the patient's upper back side of the trunk in the level of the thoracic curve. It is located only on one side of the upper back, and is the most prominent area compared to the opposite side in the same level. A rib hump is located on the thoracic convex side of the body. [6,21,26]

Steps in assessment of the rib hump:

a) A patient assumes her/his habitual posture, while the therapist observes the upper back, searching for asymmetry between left and right side of the patient's body.
b) In the presence of a rib hump, one side of the upper back will be more prominent and the shoulder blade will be protracted on that side.
c) The Adam's bending test in standing should be performed on a patient, and asymmetry should be searched for between the left and right sides. In this test, the rib hump is the most visible as it protrudes the most while trunk is bent and palms are firmly pressed against each other.
d) If a clear decision cannot be made, use the scoliometer to assess the largest vertebral rotation in the upper back. If the number reads 7 or more, the rib hump is present. Positive numbers on the scoliometer, refer to the right asymmetry, and negative to the left asymmetry.

Lumbar prominence:

A lumbar prominence is located in the lower back at the level of the lumbar scoliosis curve. It is the most prominent (protruding) area of the lower back observed unilaterally. It is positioned on the thoracic concave side of the body. [6,21,26]

Steps in the assessment of the lumbar prominence:

a) A patient assumes her/his habitual posture, while the therapist observes the lower back, searching for asymmetry between left and right sides of the patient's body.
b) In the presence of a lumbar prominence, one side of the lower back is more prominent and the waist becomes asymmetric. Also, the lumbar convex side appears as if swollen, while the lumbar concave side is indented.
c) The Adam's bending test in standing should be performed on a patient, and asymmetry should be searched for between the left and right lower back sides.
d) If a clear decision cannot be made, use the scoliometer to assess the largest vertebral rotation in the lower back. If the number reads 7 or more, the lumbar prominence is present. Positive numbers on the scoliometer, refer to the right asymmetry, and negative to the left asymmetry.

Prominent hip:

A prominent hip represents the compensation, for either thoracic or lumbar major curves.[21] When the pelvis laterally deviates in the frontal plane, the prominent hip will occur. Because the pelvis is laterally shifted, the hip becomes adducted (on the side to which the pelvis is shifted) and the iliac crest on the same side becomes raised. [21] A prominent hip, according to the Schroth classification system, only appears in the 3cp and 4cp scoliosis curve patterns. [6,21,26]

Steps in assessment of the prominent hip:

a) A patient assumes her/his habitual posture, while the therapist observes the hips and pelvis, searching for asymmetry between the left and right sides of the patient's body.
b) The therapist should look for triangles bounded by the arms hanging loosely at the patient's sides and the corresponding side of the pelvis. If the pelvis is closer to the

arm on the thoracic concave side, the prominent hip will be present on the same side. If the pelvis deviates more to the arm on the thoracic convex side, the prominent pelvis will be present on that side.

c) In case the asymmetry is not so obvious, a therapist should use a plumb line suspended from the spinous process of C7 to the gluteal cleft to observe lateral hip deviation. The prominent hip will be observed on the side to which the hip deviates, with reference to the plumb line.

d) In the presence of a prominent hip, weight bearing is asymmetrically distributed. A Schroth therapist could use two scales of the same brand and ask a patient to step on each with one leg. If the scales read differently, and more weight is distributed on the thoracic convex side, the prominent hip is on the opposite side. If more weight occurs on the thoracic concave side, the prominent hip is observed on the thoracic convex side.

Pelvis balanced:

Pelvic balance is closely related to the prominent hip. In the absence of a prominent hip, the pelvis is referred to as balanced.

Pelvis unbalanced:

Pelvic balance is closely related to the prominent hip. In the presence of a prominent hip, the pelvis is referred to as unbalanced.

Pelvis coupled with the lumbar spine:

Pelvic displacement can be coupled with the deviation of the lumbar spine, due to scoliosis. The pelvis is considered coupled with the lumbar spine if it deviates to the same side as the lumbar spine.

Steps in the assessment of the coupled mechanism of the pelvis and lumbar spine:

a) If the pelvis is judged as unbalanced, there are two possible outcomes:

    1. When the prominent hip is observed on the thoracic concave side, the pelvis will be coupled with the lumbar spine, as it will deviate in the same direction as the lumbar convexity.

2. When the prominent hip is observed on the thoracic convex side, the pelvis will be uncoupled with the lumbar spine, as it will deviate in the opposite direction of the lumbar convexity.

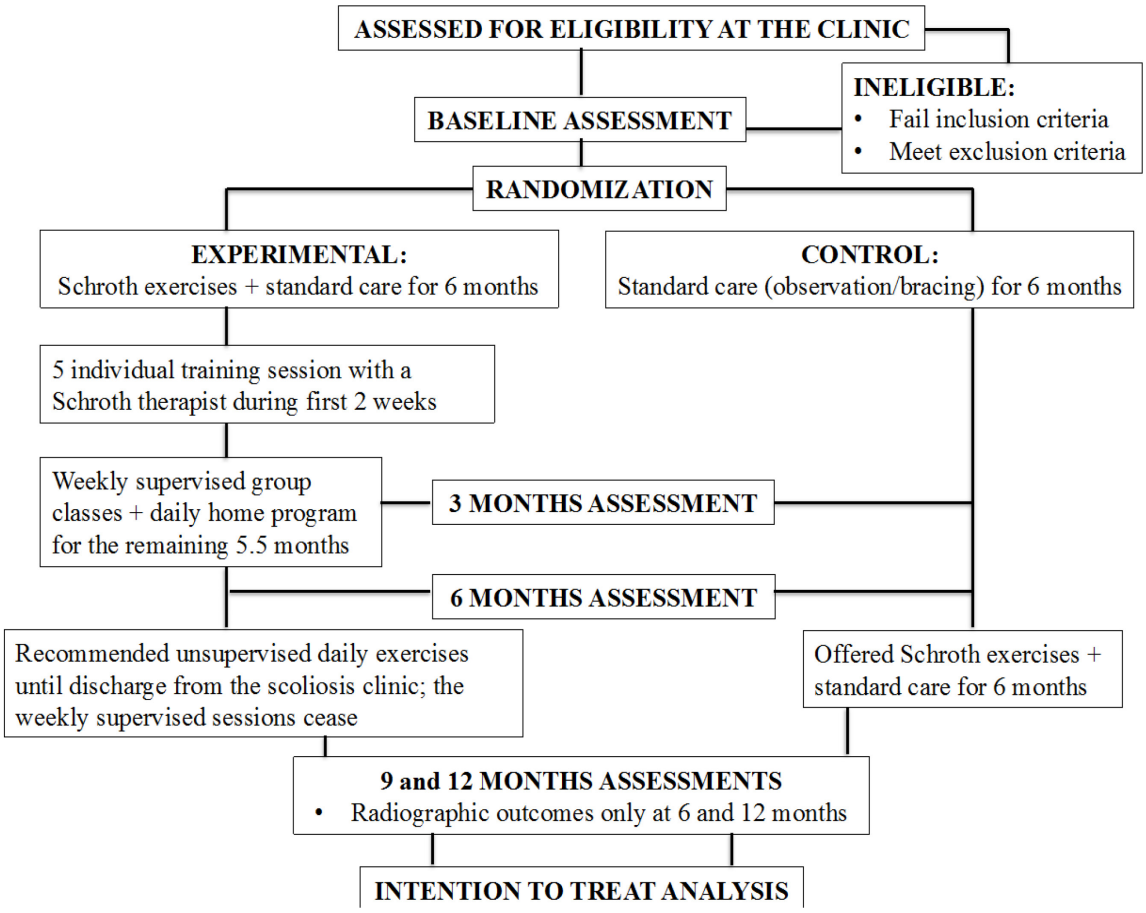
Pelvis uncoupled with the lumbar spine:

Pelvic displacement can be uncoupled with the deviation of the lumbar spine, due to scoliosis. The pelvis is considered uncoupled with the lumbar spine if it is deviated to the opposite side of the lumbar spine. If uncoupled, the pelvis can deviate towards the midline, or further away, in the opposite direction of the lumbar spine deviation.

Steps in the assessment of the uncoupled mechanism of the pelvis and lumbar spine:

a) If the pelvis is deemed as balanced in the physical assessment, then it will always be uncoupled with the lumbar spine, as it leans towards the midline, and away from the lumbar convex scoliosis curve.

Appendix 2


ASSESSED FOR ELIGIBILITY AT THE CLINIC
INELIGIBLE:
• Fail inclusion criteria
• Meet exclusion criteria
BASELINE ASSESSMENT
RANDOMIZATION
EXPERIMENTAL:
Schroth exercises + standard care for 6 months
CONTROL:
Standard care (observation/bracing) for 6 months
5 individual training session with a Schroth therapist during first 2 weeks
Weekly supervised group classes + daily home program for the remaining 5.5 months
3 MONTHS ASSESSMENT
6 MONTHS ASSESSMENT
Recommended unsupervised daily exercises until discharge from the scoliosis clinic; the weekly supervised sessions cease
Offered Schroth exercises + standard care for 6 months
9 and 12 MONTHS ASSESSMENTS
• Radiographic outcomes only at 6 and 12 months
INTENTION TO TREAT ANALYSIS

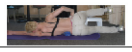
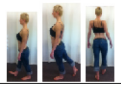
Appendix 3

The Schroth method

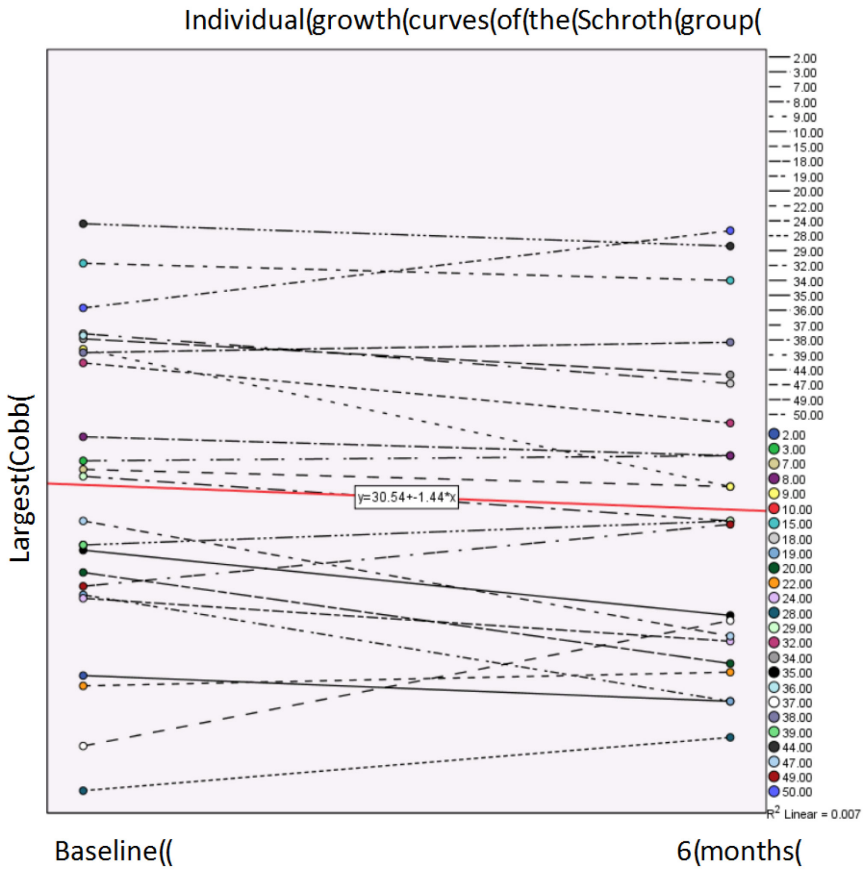
Schroth exercise components common to all exercises

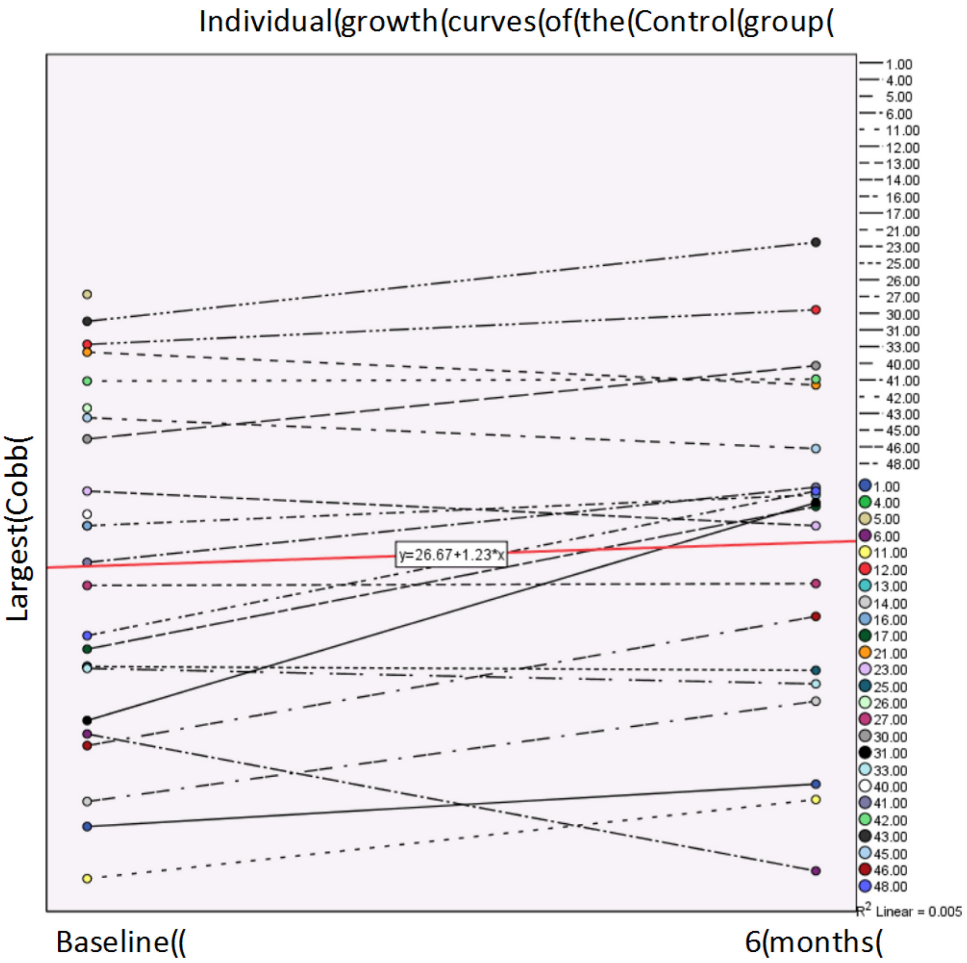
Schroth exercises differ in: the level of passive support provided; whether static or dynamic; dosage; and the curve type for which they are intended. All exercises share five common elements.

1) **Elongation** consists of a conscious self-lengthening.
2) **Side-shift** is performed by translating the rib cage towards the curve concavity.
3) For deep **rotational breathing cycles** (5-10 seconds each), during inhalation the patient actively side-shifts and de-rotates of the torso towards the concavity. During, forced exhalation the patient actively contracts accessory and shoulder muscles, to depress areas of convexity.
4) **Shoulder-counter tractions** aim to open the thoracic concavity by actively moving the shoulder on the concave side of the thoracic spine upwards and sideways thereby straightening the curve. The other shoulder is moved horizontally by contracting the serratus anterior, and rhomboid muscles, which provides de-rotation of the rib hump.
5) **Pelvic corrections** consists of:
    a) aligning in the sagittal profile by balancing over the center of the feet;
    b) adjusting the lumbar spine into a lordosis position;
    c) translating the pelvis in the frontal plane to correct a prominent hip;
    d) rotating the pelvis horizontally to correct excess rotation;
    e) elevating or depressing one side of the pelvis in the frontal plane to correct a prominent higher hip if present.

Exercise	Picture	Curve type	Passive support	Movement type	Start dose*	Target dose*
1. Side-bending in side-lying		3c, 3cp	Substantial	Static	2 min.	10 min.
2. Side-lying + shoulder counter-traction and muscle cylinder		All	Substantial	Static	5 sets x 6 breaths	5 x 8
3. Shoulder counter-traction with psoas activity		All	Substantial	Static	4 x 5	5 x 8
4. Sitting on a ball		All	Medium	Static	4 x 6	5 x 10
5. Hip flexion in side-lying		4c, 4cp	Substantial	Static	5 x 5	5 x 8
6. Hip Resistance		3cp, 4cp	None	Static	5 x 5	5 x 8
7. Chest Twister		3c, 3cp	Medium	Static	4 x 5	5 x 6
8. Hip flexion against a ball		4c, 4cp	Medium	Static	3 x 6	3 x 8
9. Standing between two poles		All	None	Static	5 x 4	5 x 6
10. Foot (or knee) under a bar		4c, 4cp	None	Static	4 x 6	5 x 8
11. Side-Hanging		3c, 3cp	None	Static	2 x 4	2 x 4
12. St. Andrew's Cross		3c, 3cp	None	Static	2 x 4	2 x 4
13. Muscle cylinder		4c, 4cp	None	Static	4 x 5	5 x 10
14. Muscle cylinder with a pole		3c, 3cp	None	Static	4 x 5	5 x 10
15. The Circle		All	None	Dynamic	3	5
16. Conscious walking		All	None	Dynamic	4 x 10m	-
17. Schroth walking		All	None	Dynamic	-	4 x 10m

Appendix 4


Individual(growth(curves(of(the(Schroth(group(
Largest(Cobb(
y=30.54+-1.44*x
2.00
3.00
7.00
8.00
9.00
10.00
15.00
18.00
19.00
20.00
22.00
24.00
28.00
29.00
32.00
34.00
35.00
36.00
37.00
38.00
39.00
44.00
47.00
49.00
50.00
R2 Linear = 0.007
Baseline((
6(months(


Individual(growth(curves(of(the(Control(group(
Largest(Cobb(
y=26.67+1.23*x
1.00
4.00
5.00
6.00
11.00
12.00
13.00
14.00
16.00
17.00
21.00
23.00
25.00
26.00
27.00
30.00
31.00
33.00
40.00
41.00
42.00
43.00
45.00
46.00
48.00
R2 Linear = 0.005
Baseline((
6(months(

Appendix 5

The anchor- and distribution-based cutoffs reported as percent of baseline values

	Baseline mean	MCID percent of baseline values (%)	ROC percent of baseline values (%)	SEM percent of baseline values (%)	MDD_{95} percent of baseline values (%)
Largest Cobb (°)	28.5	4.7	3.2	4.4	12
Combined Cobb (°)	51.2	5.3	2.6	6.1	17
SRS-22r:					
Pain /5	4.33	3.7	N/A	2.8	7.4
Image /5	3.86	1	N/A	5.2	14
Function/5	4.58	0.9	N/A	2.4	6.8
Total /5	4.19	1.7	N/A	3.1	8.6
SAQ:					
General /5	2.90	1.7	N/A	17.2	47.2
Curve /5	2.18	2.3	N/A	11	30.7
Prominence /5	1.67	15	N/A	18	48.5
Trunk shift /5	1.90	2.6	-0.25	14.7	40.5
Waist /5	2.69	4.5	N/A	33.1	91.4
Shoulders /5	2.52	2	N/A	21.4	59.5
Chest /5	2.16	7	N/A	37.5	104
Biering-Sorensen (Sec.)	110.96	32.6	12.5	20.5	56.6