Spine and Thoracic Height Measurements have Excellent Interrater and Intrarater Reliability in Patients with Early Onset Scoliosis

Nicole Michael, BA; Children’s Hospital Colorado, Aurora, CO
Patrick Carry, MS; Children’s Hospital Colorado, Aurora, CO
Mark Erickson, MD, MMM; Children’s Hospital Colorado, Aurora, CO
Nikki Bloch, BA; Children’s Hospital Colorado, Aurora, CO
Steven Gibbons, MD; Tulane University, New Orleans, LA
Courtney O’Donnell, MD; Children’s Hospital Colorado, Aurora, CO
Sumeet Garg, MD, Children’s Hospital Colorado, Aurora, CO

Corresponding Author:
Sumeet Garg, MD
13123 East 16th Avenue, Box 060
Aurora, CO 80045
Phone: (720) 777-3253
Fax: (720) 777-7268
Sumeet.garg@childrenscolorado.org

The manuscript submitted does not contain information about medical device(s)/drug(s). No funds were received in support of this work. Relevant financial activities outside the submitted work: consultancy, royalties, payment for expert testimony.
Structured Abstract

Study Design: Reproducibility of measurements

Objective: This study investigates the reliability and standard error of measurement of spine and thoracic height radiographic measurements in patients with early onset scoliosis (EOS).

Summary of Background Data: Spine and thoracic height radiographic measurements are often used as a surrogate for pulmonary development in patients with EOS. There is limited literature validating the reliability of spine and thoracic height measurements in the EOS population.

Methods: Using pilot data, we determined measuring 49 unique radiographs would provide 80% power to obtain a 95% confidence interval width of 0.05 for the interclass correlation coefficients (ICC). A random sampling strategy, stratified by underlying diagnosis from the Classification of Early Onset Scoliosis (C-EOS), was used to distribute the diagnoses in the study sample. Two attending pediatric spine surgeons, two pediatric orthopaedic fellows, and two research assistants measured coronal spine (T1-S1) and thoracic (T1-T12) height on digital radiographs using imaging software (Surgimap®; Nemaris, Inc; New York) on two separate occasions at least three weeks apart. Order of images was randomized for the second iteration. Linear mixed model regression analyses were used to estimate interrater and intrarater reliability.

Results: The study sample included subjects (N=48) with idiopathic (N=17, 35%), congenital (N=16, 33%, 1 patient excluded), neuromuscular (N=11, 23%), and syndromic (N=4, 8%) scoliosis. Overall interrater reliability estimates for spine height [ICC: 0.894, 95% CI: 0.847-0.932] and thoracic height [ICC: 0.890, 95% CI: 0.844-0.929] were excellent. Intrarater reliability estimates for spine height [ICC: 0.906, 95% CI: 0.830-0.943] and thoracic height [ICC: 0.898, 95% CI: 0.817-0.938] were also excellent.
Conclusions: There is excellent interrater and intrarater reliability for radiographic measurements of spine and thoracic height in the EOS population at our institution.

Key Words
Spine, thoracic, height, reliability, agreement, early onset scoliosis, radiograph, measurement, X-ray, scoliosis, pediatric, length, deformity, growth-friendly treatment

Level of Evidence: 2
Introduction

Scoliosis diagnosed before the age of 10 is considered early onset scoliosis (EOS). EOS is classified into four major etiologies by the Classification of Early Onset Scoliosis (C-EOS): congenital, neuromuscular, syndromic, or idiopathic scoliosis.[1, 2] Coronal and sagittal radiographic spine measurements are important tools in assessing spine deformities and efficacy of treatment.[3] Many of these measurements have been validated in the literature. Specifically regarding coronal Cobb angle, numerous studies report the interrater and intrarater reliability of this digital measurement with intraclass correlation coefficients (ICC) greater than 0.9, assuring strong reliability in this measurement.[4-7]

Serial spine height measurements are a common outcome measure when investigating the success of growth friendly treatment in patients with EOS.[8, 9] The spine height is measured from T1-S1 vertebrae and the height of the thoracic segment of the spine spans T1-T12 vertebrae. Literature concerning children with EOS suggests a decrease in both trunk and spine final height and growth rate compared to normal ranges for spine growth.[10, 11] Optimizing growth of the spine is one of the core principles of treatment in EOS.[12] It is believed this may correlate with improved pulmonary development and function.[11, 13-15]

The reliability of thoracic and spine height measurements needs to be validated to be able to appropriately assess both the deformity of the spine and progression of the deformity, and interpret studies that report this outcome metric for the EOS population.[10, 12] The purpose of this study is to ascertain the reliability of the thoracic and spine height measurements in an EOS population at our institution.

Materials and Methods
A pilot study was used to estimate the intraclass correlation coefficients representing the degree of interrater reliability associated with spine height and thoracic height measurements among patients with EOS. Three different surgeons measured spine and thoracic height measurements on anterior/posterior radiographs obtained from a random sample of 10 patients with EOS. ICC estimates of 0.95 and 0.94 were obtained for the spine height and thoracic height measurements, respectively. Based on the pilot study, we determined that 49 radiographs would provide 80% power to obtain a 95% confidence interval width of 0.05. This estimate assumes an ICC value of 0.94, 6 raters per subject, and that the raters are randomly sampled from a population of raters. The lesser of the two ICC values from the pilot study was conservatively used in the sample size calculation. The methods described by Bonett were used to estimate the sample size based on our desired level of precision, 0.05.[16]

The selected radiographs were retrospectively accessed from a registry of patients with EOS at our institution. Forty-nine radiographs were selected from a total of 1,101 available radiographs using a block randomization scheme based on EOS etiology to ensure a distribution similar to that observed in clinic, as demonstrated by Table 1.

The raters included two board certified pediatric orthopedic spine surgeons (attendings), two pediatric orthopedic fellows (fellows), and two pediatric orthopedic spine research assistants (RAs) from a single pediatric institution. The 49 radiographs were uploaded and measured through Surgimap Spine software (Version 2.2.9.8, New York, http://surgimapp spine.com/). Surgimap is a secure tool used to measure radiographs of the spine.[17-19] Raters were asked to independently measure each of the 49 radiographs by calibrating the software, measuring the thoracic height (T1-T12) from the middle of the superior endplate of T1 to the middle of the inferior endplate of T12, and measuring the spine height from the middle of the superior endplate of T1 to the middle of the superior endplate...
of S1 (T1-S1). The software was calibrated in one of three methods: using the ruler at the bottom of the radiograph (variable measurement based on rater discretion), drawing a line through the thinnest point of a VEPTR rod (6mm), or drawing a diameter line through circular scale marker (25mm). Figure 1 exhibits the measurements of spine and thoracic height using Surgimap software.

The radiographs were measured twice by each rater three weeks apart in order to assess intrarater reliability. All radiographs were de-identified and assigned a unique identification number. Raters measured the radiographs following the numerical order of their identification numbers. Between iterations, the radiographs were shuffled and given a new identification.

Descriptive statistics were used to summarize the C-EOS etiology of the patients’ radiographs measured by the raters. Linear mixed model regression analyses were used to estimate interrater and intrarater reliability of thoracic and spine height measurements (SAS 9.4, SAS Institute Inc., Cary, NC, USA). Random variance components specified in the linear mixed model (between patient error, $\sigma_P^2$, between rater error, $\sigma_R^2$, patient*rater random error, $\sigma_{P(R)}^2$, and residual or within patient error, $\sigma_e^2$) were used to calculate the ICC values representing interrater ($\sigma_P^2/\sigma_P^2 + \sigma_R^2 + \sigma_{P(R)}^2 + \sigma_e^2$) and intrarater reliability ($\sigma_P^2 + \sigma_R^2 + \sigma_{P(R)}^2 + \sigma_e^2$). The 95% confidence intervals were calculated based on previously described methods.[20-22] The standard error of measurement (SEM) was calculated for interrater and intrarater reliability as $\sqrt{\sigma_P^2 + \sigma_R^2 + \sigma_{P(R)}^2 + \sigma_e^2}$ and $\sqrt{\sigma_e^2}$, respectively. Secondary analyses were used to calculate interrater and intrarater reliability for the different rater classifications.

Results
The study population (n=48) included subjects with idiopathic (35%, n= 17), congenital (33%, n=16), neuromuscular (23%, n=11), and syndromic (8%, n=4) scoliosis. One congenital radiograph was excluded due to improper stitching of the radiograph, identified by one of the attending raters.

The ICC value for overall interrater reliability was 0.894 [95 CI: 0.847-0.932] for spine height and 0.890 [95% CI: 0.844-0.929] for thoracic height. Intrarater reliability was 0.906 [95% CI: CI: 0.830-0.943] for spine height and 0.898 [95% CI: 0.817-0.938] for thoracic height. When stratified by rater type, reliability estimates were higher among the attending raters relative to the fellow and RA raters (Table 2 and Figure 2).

The standard error of measurement (SEM) based on intrarater reliability was 1.95 cm for spine height and 1.31 cm for thoracic height. The SEM based on interrater reliability was 2.07 cm for spine height and 1.36 cm for thoracic height.

Discussion

The purpose of this study was to evaluate the reliability of spine and thoracic height radiographic measurements within the EOS population at our institution. The spine and thoracic height measurements demonstrated excellent reliability both overall and among each rater type (attendings, fellows, and RAs). The RAs demonstrated a greater reliability in measuring spine and thoracic height relative to the fellows, perhaps due to having more practice performing radiographic spine measurements on a day to day basis. The varying reliability between rater types highlights the importance of understanding the level of expertise of the person performing measurements in outcome based studies.

One of the challenges to treating the growing spine is to preserve the normal growth rate of the trunk or the spine.[11] For children ages 5-10 without scoliosis, their thoracic spine grows approximately 0.7 cm/year and their spine grows approximately 1 cm/year.[11]
This growth rate presents measurement challenges when evaluating longitudinal changes in spine height. Based on the intrarater standard error of measurement estimates, caution should be used when interpreting changes in spine height measured to be less than 1.95 cm and changes in thoracic height less than 1.31 cm. The measurement error for thoracic and spine height is greater than the normal yearly thoracic and spine respective growth rate of a pediatric patient, and thus must be taken into consideration when evaluating treatments over a short duration of time (1-2 years). When evaluating spine and thoracic height measures at a discrete time point, group comparisons are needed to consider the magnitude of measurement error expected to occur due to random chance alone. Based on the interrater SEM estimates, between-group differences in spine height less than 2.07 cm and thoracic height measurements less that 1.36 cm should scrutinized.

A concern with growth friendly treatment is the “law of diminishing returns,” an observed trend where an initial net gain in spine height, tends to decrease with subsequent growth friendly expansions.[23, 24] Sankar et al. attributes the “law of diminishing returns” to the auto fusion of the spine limiting its growth ability.[24] However, Spurway et al. attributes the “law of diminishing returns” to the consideration of the spine height measurement to be only in the coronal plane.[23] Their study asserts, in order to understand true spine height and growth, the spine needs to be measured 3-dimensionally following the sagittal curved height of the spine.[23] In either scenario, the authors do not mention the SEM inherent to spine height. When interpreting spine height measurements with standard error, considering the error inherent to this measurement while investigating the outcomes of growth friendly treatment is essential.

Results of this study include the following limitations. We only assessed the reliability of spine and thoracic height measurements in the coronal plane of the deformity and did not include the sagittal plane. A similar study by Spurway et al. assessed the
reliability of the coronal and sagittal spine length radiographic measurement and vertical height radiographic measurement of the thoracic spine in patients with EOS.[23] Spurway et al. demonstrated similarly excellent interrater and intrarater reliability of the coronal thoracic height measurements by 5 raters measuring 23 radiographs; however, their study did not include coronal spine height reliability or SEM.[23] We also did not investigate the accuracy of the spine and thoracic height measurement in this study. We do not know the true spine and thoracic heights of these patients and therefore cannot compare to a known value. Lastly, this study was not powered to determine the reliability among individual EOS etiologies.

In conclusion, we assessed the reliability of pediatric spine and thoracic height radiographic measurements in a population that included subjects with idiopathic, congenital, neuromuscular, and syndromic EOS. This study was powered using a distribution of EOS radiographs similar to the distribution observed in clinic, allowing the authors to generalize the results to the larger EOS population. Spine and thoracic height measurements demonstrated excellent overall inter and intrarater reliability at our institution. SEM estimates should be considered when utilizing or interpreting spinal and thoracic height measurements.
References


Figure Legends

Figure 1. Surgimap radiographic measurements of spine and thoracic height. Raters used one of three possible calibration methods depicted below to calibrate software. (Legend: A. Calibration Rule, 50.00mm; B. Calibration VEPTR Rod, 6.00mm; Calibration Circular Scale Marker, 25.00mm; D. Superior Endplate T1; E. Inferior Endplate T12; F. Superior Endplate S1; G. Thoracic Height, T1-T12; H. Spine Height, T1-S1).
**Figure 2.** Inter and Intrareliability Estimates by Rater (Legend: * = spine height measurement; ‡ = thoracic height measurement)
Table 1. Distribution of etiology among randomly selected radiographs from patients in an Early Onset Scoliosis registry.

<table>
<thead>
<tr>
<th>C-EOS Etiology</th>
<th>Registry (n)</th>
<th>Ratio (%)</th>
<th>Radiographs (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Congenital</td>
<td>191</td>
<td>34</td>
<td>17</td>
</tr>
<tr>
<td>Neuromuscular</td>
<td>123</td>
<td>22</td>
<td>11</td>
</tr>
<tr>
<td>Syndromic</td>
<td>50</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>Idiopathic</td>
<td>195</td>
<td>35</td>
<td>17</td>
</tr>
<tr>
<td>Total EOS</td>
<td>559</td>
<td>100</td>
<td>49</td>
</tr>
</tbody>
</table>
Table 2. Estimates by Rater Type.

<table>
<thead>
<tr>
<th></th>
<th>Interrater Reliability</th>
<th>Intrarater Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ICC</td>
<td>95% CI</td>
</tr>
<tr>
<td><strong>Spine Height</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attendings</td>
<td>0.968</td>
<td>0.940-0.982</td>
</tr>
<tr>
<td>Fellows</td>
<td>0.828</td>
<td>0.747-0.891</td>
</tr>
<tr>
<td>RAs</td>
<td>0.873</td>
<td>0.724-0.936</td>
</tr>
<tr>
<td><strong>Thoracic Height</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attendings</td>
<td>0.941</td>
<td>0.907-0.965</td>
</tr>
<tr>
<td>Fellows</td>
<td>0.811</td>
<td>0.716-0.882</td>
</tr>
<tr>
<td>RAs</td>
<td>0.916</td>
<td>0.862-0.950</td>
</tr>
</tbody>
</table>

Copyright © 2017 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.