A call for consistent radiographic definition of lumbar lordosis

Mark B. Frenkel, MD,1 Casey D. Frey, BS,2 Jaclyn J. Renfrow, MD,1 Stacey Q. Wolfe, MD,1 Alexander K. Powers, MD,1 and Charles L. Branch Jr., MD1

1Department of Neurosurgery, Wake Forest Baptist Medical Center, Winston-Salem, North Carolina
2Wake Forest University School of Medicine, Winston-Salem, North Carolina

Over the past several years there has been increased interest in lumbar lordosis (LL) and its relationship to pathology and other spinopelvic parameters, especially pelvic incidence (PI). Lumbar lordosis has been shown to correlate strongly enough with clinical outcome measures in patients with adult spinal deformity (ASD) that some authors suggest its restoration should be a primary objective of deformity surgery.18 Unfortunately, the measurement of LL with other spinopelvic parameters has become such a norm that some studies11 neglect to even mention it in their Methods section. We have identified multiple variations in the definition of LL throughout the literature, and hypothesize that these variations may impact the conclusions and clinical decision making that derives from this measurement. In this paper we attempt to determine how pervasive the variability of the definition of LL is within the current spine literature.

Methods

Search Strategy

A literature search of PubMed was conducted by searching for the term “Lumbar Lordosis” in the Title, Keyword, and/or Abstract fields. The search was restricted to papers published between January 1, 2016, and December 31, 2016. Publications that originally appeared in the query results as “published ahead of print” but that were subsequently updated with later publication dates were not removed. The search was also restricted to the following spine-related journals: Spine; Clinical Spine Surgery; The Spine Journal; Journal of Neurosurgery: Spine; Neurosurgery; Asian Spine Journal; International Journal of Spine Surgery; Spine Deformity; Scoliosis; Scoliosis and Spinal Disorders; and Global Spine Journal. Papers were screened by 2 authors and included only if they measured LL in at least 1 patient for the purposes of the study.

Data Extraction

The entirety of each paper was reviewed twice (once each by 2 authors) using a standardized form. The authors recorded whether each study explicitly specified the upper and lower vertebrae from which LL was measured and, if so, whether the study specified if the inferior or superior endplates of each vertebra was used (with the exception of those measurements terminating at S-1, in which case it was only recorded whether the endplate of the rostral vertebra was specified). Discrepancies between reviewers prompted a reevaluation of the article in question by both reviewers.

Results

Of the 118 papers initially identified, 5 (2 editorials, 2 review articles, and 1 research study) were excluded for not measuring LL in at least 1 patient. A total of 113 papers were included and critically reviewed. Thirty-four studies (30%) failed to describe which vertebrae were used to measure LL. One mischaracterized segmental lordosis as LL. One used a nonconstrained measurement. Three measured surface topography. Three studies1,6,8 had diagrams illustrating LL but did not explicitly state the vertebrae used for measurement in the text of the manuscript or figure legend. These studies were considered to be lacking a definition, because a nonconstrained measurement could have been used. Of the 79 papers that did specify the rostral and caudal vertebrae from which LL was measured, 21 did not specify which endplate was used for the measurement, leaving only 58 studies (51%) with complete definitions of LL.

In those papers that did define which vertebrae were used to measure LL, 12 different definitions of lumbar lordosis were used (Table 1). This lack of consistency...
hold true even when comparing papers published within individual journals; only the 4 papers published in Neurosurgery had a consistent definition within the same journal (although this could not be assessed for International Journal of Spine Surgery or Global Spine Journal because these each had only 1 paper that met inclusion criteria).

The included papers used the measure of LL for several different purposes. Thirty-nine (34.5%) of the included papers used LL to assess the impact of an intervention or compare interventions (Supplemental Table 1). Twenty-one (18.6%) used LL to assess the reliability of spinopelvic measures or the relationship between spinopelvic and other physiological or quality of life measures (Supplemental Table 2). Thirty-two (28.3%) used LL measurements to assess whether some values of LL were risk factors for disease, predictors of a complication or outcome, or correlated with disease severity (Supplemental Table 3). Twenty-one (18.6%) used LL measurements to define a disease state or normal parameters in a population (Supplemental Table 4). There was no consistent definition of LL used within any of these groups. These studies also examined several different cohorts: 10 were focused on healthy individuals, 32 on nonoperatively treated or preoperative patients with disease, 70 on postoperative patients, and 1 study was performed on cadavers. Of the 10 studies focused on healthy individuals, 9 of them described how LL was measured, and 5 of these used the same definition (one used the superior endplate of L1–S1; one did not specify which endplates of L1–S1 were used; another used a nonconstrained measurement; and two used surface topography). All of these 10 were used to either assess the reliability of or relationship between spinopelvic measures, or to define normal parameters in a population.

Twenty-one papers focused primarily on some form of scoliosis, with 7 focusing on either adolescent idiopathic or congenital scoliosis. There was no consistency in the measure of LL used within any of these groups.

**Discussion**

**Impact of Different Definitions**

The lack of a consistent definition of LL was first brought to light by Polly et al. more than 20 years ago when they identified 4 different definitions of LL in the popular literature and examined the interobserver reliability of each.\(^{15}\) We found 12 different definitions of LL used among the studies we analyzed. The most obvious implication of these differences is that uncertainty is introduced when making comparisons between studies. For illustration, the T-12 and L-1 vertebral bodies have each been shown to contribute an average of approximately 4° of kyphosis due to anterior wedging.\(^7\) This has been corroborated by other studies, with at least one population demonstrating statistically significant differences in vertebral body angulation between sexes.\(^2,14\) Similarly, the L5–S1 disc space has been shown to contribute an average of 11.7° of lordosis in at least one population,\(^3\) and the segmental angle between the superior endplate of L-5 and the superior endplate of S-1 (i.e., both the L-5 vertebral body and L5–S1 disc space combined) has been shown to contribute an average of approximately 25° of lordosis,\(^12\) suggesting an approximately 13° contribution from the L-5 vertebral body. As a result, the difference between measuring from the superior endplate of T-12 to S-1 versus the superior endplate of L-1 to the inferior endplate of L-5 (as was done in 9 and 5 of the studies reviewed, respectively) can easily exceed the 11° PI-LL mismatch commonly accepted to be clinically significant for LL\(_{\text{L1-S1}}\).\(^19\) In fact, one group found an 18.3° difference between the means of 90 measurements each of LL\(_{\text{S1-T12}}\) and LL\(_{\text{L1-L5}}\) on 30 radiographs.\(^10\)

These discrepancies in the definition of LL create an inconsistent platform for the use of PI-LL as a clinical tool. This was demonstrated by one study in which logistic regression was used to determine a cutoff for a PI-LL value that had adequate sensitivity and reliability for predicting the risk of developing adjacent-segment disease in patients with prior lumbar fusions. Calculations were performed for both LL\(_{\text{L4-L5}}\) and LL\(_{\text{L5-S1}}\), and the authors found a 55% higher cutoff value for the latter (9.8° and 15.2°, respectively), further illustrating the impact of different LL definitions on PI-LL and the clinical decisions made from this calculation.\(^16\)

Unfortunately, the lack of a consistent definition of LL is pervasive throughout the literature. In some cases, the same primary author or large spine study group have used different definitions in different publications.\(^2,21\) As a result of both how common the use of different LL definitions is and the potential impact of different definitions on a study’s findings, questions are raised by those studies that fail to specify the definition of LL used. Take for example a large multicenter trial in which it was concluded that there was no correlation between LL and patient satisfaction scores in postoperative patients with ASD. The authors did not specify the definition of LL they used for measurement, raising the question of whether uniform measurements were used between study sites.\(^9\)

**A Call for Consensus**

We have found that the literature is rife with studies whose methodology or conclusion has been negatively