Predicting Scoliosis Progression from Skeletal Maturity: A Simplified Classification During Adolescence

By James O. Sanders, MD, Joseph G. Khoury, MD, Shyam Kishan, MD, Richard H. Browne, PhD, James F. Mooney III, MD, Kali D. Arnold, MD, Sharon J. McConnell, MS, Jeanne A. Bauman, MD, and David N. Finegold, MD

Investigation performed at Shriners Hospitals for Children, Erie, Pennsylvania

Background: Both the Tanner-Whitehouse-III RUS score, which is based on the radiographic appearance of the epiphyses of the distal part of the radius, the distal part of the ulna, and small bones of the hand, and the digital skeletal age skeletal maturity scoring system, which is based on just the metacarpals and phalanges, correlate highly with the curve acceleration phase in girls with idiopathic scoliosis. However, these systems require an atlas and access to the scoring system, making their use impractical in a busy clinical setting. We sought to develop a simplified system that would correlate highly with scoliosis behavior but that would also be rapid and reliable for clinical practice.

Methods: A simplified staging system involving the use of the Tanner-Whitehouse-III descriptors was developed. It was tested for intraobserver and interobserver reliability by six individuals on thirty skeletal age radiographs. The system was compared with the timing of the curve acceleration phase in a cohort of twenty-two girls with idiopathic scoliosis.

Results: The average intraobserver unweighted kappa value was 0.88, and the average weighted kappa value was 0.96. The percentage of exact matches between readings for each rater was 89%, and 100% of the differences were within one unit. The average interobserver unweighted kappa value was 0.71, and the average weighted kappa value was 0.89. The percentage of exact matches between two reviewers was 71%, and 97% of the interobserver differences were within one stage or matched. The agreement was highest between the most experienced raters. Interobserver reliability was not improved by the use of a classification-specific atlas. The correlation of the staging system with the curve acceleration phase was 0.91.

Conclusions: The simplified skeletal maturity scoring system is reliable and correlates more strongly with the behavior of idiopathic scoliosis than the Risser sign or Greulich and Pyle skeletal ages do. The system has a modest learning curve but is easily used in a clinical setting and, in conjunction with curve type and magnitude, appears to be strongly prognostic of future scoliosis curve behavior.

Level of Evidence: Prognostic Level I. See Instructions to Authors for a complete description of levels of evidence.

Prognostic studies related to the risk of curve progression in patients with idiopathic scoliosis have been hampered by poor measurements of skeletal maturity and arbitrary definitions of curve progression. In a previous study[1], we compared a number of potential skeletal maturity indicators, including the Risser sign, the Oxford stage, the Greulich and Pyle and Tanner-Whitehouse-III skeletal maturity assessments, and a number of serologic skeletal maturity markers in girls

Disclosure: In support of their research for or preparation of this work, one or more of the authors received, in any one year, outside funding or grants in excess of $10,000 from the Scoliosis Research Society. Neither they nor a member of their immediate families received payments or other benefits or a commitment or agreement to provide such benefits from a commercial entity. No commercial entity paid or directed, or agreed to pay or direct, any benefits to any research fund, foundation, division, center, clinical practice, or other charitable or nonprofit organization with which the authors, or a member of their immediate families, are affiliated or associated.

TABLE I Key Findings of the Simplified Tanner-Whitehouse-III Skeletal Maturity Assessment

Stage	Key Features	Tanner-Whitehouse-III Stage	Greulich and Pyle Reference	Related Maturity Signs
1. Juvenile slow	Digital epiphyses are not covered.	Some digits are at stage E or less.	Female 8 yr + 10 mo, male 12 yr + 6 mo (note fifth middle phalanx)	Tanner stage 1
2. Preadolescent slow	All digital epiphyses are covered.	All digits are at stage F.	Female 10 yr, male 13 yr	Tanner stage 2, starting growth spurt
3. Adolescent rapid—early	The preponderance of digits are capped. The second through fifth metacarpal epiphyses are wider than their metaphyses.	All digits are at stage G.	Female 11 and 12 yr, male 13 yr + 6 mo and 14 yr	Peak height velocity, Risser stage 0, open pelvic triradiate cartilage
4. Adolescent rapid—late	Any of distal phalangeal physes are clearly beginning to close (see detailed description in the text).	Any distal phalanges are at stage H.	Female 13 yr (digits 2, 3, and 4), male 15 yr (digits 4 and 5)	Girls typically in Tanner stage 3, Risser stage 0, open triradiate cartilage
5. Adolescent steady—early	All distal phalangeal physes are closed. Others are open.	All distal phalanges and thumb metacarpal are at stage I. Others remain at stage G.	Female 13 yr + 6 mo, male 15 yr + 6 mo	Risser stage 0, triradiate cartilage closed, menarche only occasionally starts earlier than this
6. Adolescent steady—late	Middle or proximal phalangeal physes are closing.	Middle or proximal phalanges are at stages H and I.	Female 14 yr, male 16 yr (late)	Risser sign positive (stage 1 or more)
7. Early mature	Only distal radial physis is open. Metacarpal physeal scars may be present.	All digits are at stage I. The distal radial physis is at stage G or H.	Female 15 yr, male 17 yr	Risser stage 4
8. Mature	Distal radial physis is completely closed.	All digits are at stage I.	Female 17 yr, male 19 yr	Risser stage 5

with idiopathic scoliosis. In that study, we identified the Tanner-Whitehouse-III RUS skeletal maturity assessment method[2] as the measurement most closely related to curve behavior, or, more specifically, to the timing of curve take-off in early adolsecence, which is termed the curve acceleration phase (CAP). The curve acceleration phase is distinguished by a change in the rapidity of curve progression in early adolescence. Curve progression averages 0.2° per month before the curve acceleration phase, but it may approach 1.0° to 2.0° per month immediately following the start of the curve acceleration phase.

The Tanner-Whitehouse-III RUS (radius, ulna, small bones of the hand) method specifically uses the distal radial and ulnar epiphyses and the metacarpal and phalangeal epiphyses of the first, third, and fifth digits for determination of skeletal age. Each bone is assigned a specific maturity score, and then the individual bone scores are added to obtain a total RUS score. Our previous work identified the radial and ulnar epiphyses as having the least correlation of all of the RUS growth centers with scoliosis behavior. Eliminating the radial and ulnar scores provides a digital skeletal age (DSA) score, which correlates very highly with the timing of the curve acceleration phase. However, the DSA score is clinically cumbersome as it requires access to the Tanner-Whitehouse-III atlas for individual bone scores, which then have to be totaled. We sought to develop a rapid and reliable staging system, based on the radiographic findings that are used to determine the DSA score, that would correlate with scoliosis progression. We then related these stages of skeletal maturity to the probability of scoliosis progression beyond 50°.

Materials and Methods

The present study included the same population as our previous study[1], in which a cohort of twenty-two girls with idiopathic scoliosis was followed through their growth spurt with use of serial spine radiographs, skeletal age radiographs, and a number of clinical and biochemical markers of maturation, all of which were assessed every six months. Patients underwent scoliosis brace treatment according to accepted criteria (initiation of bracing for curves of ≥25° or for curves of 20° to 24° with 5° of documented progression in patients with a Risser stage[3] of ≤2). The key findings of the

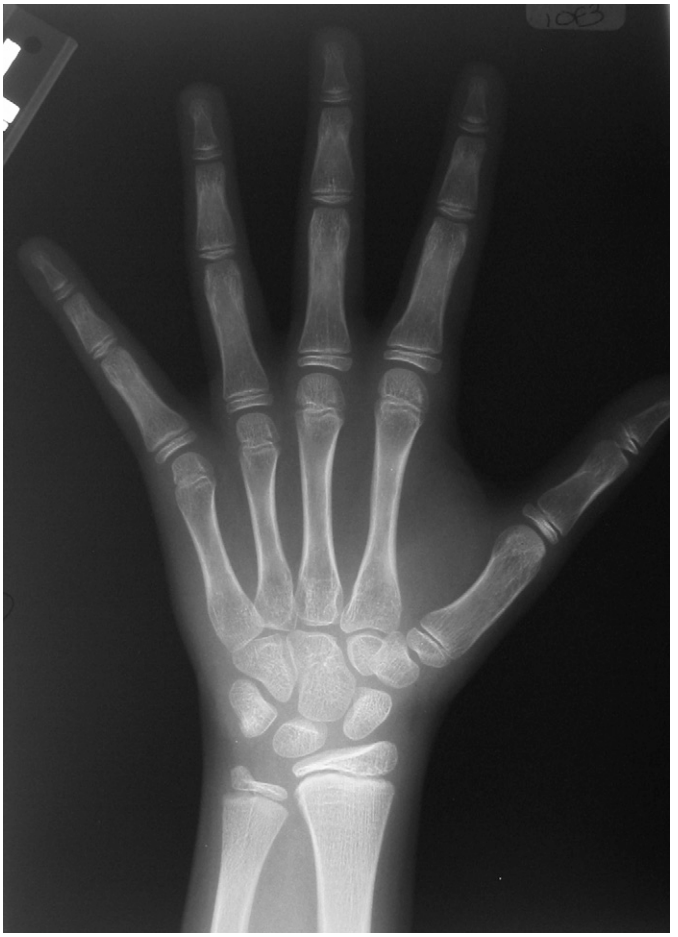

Fig. 1-A

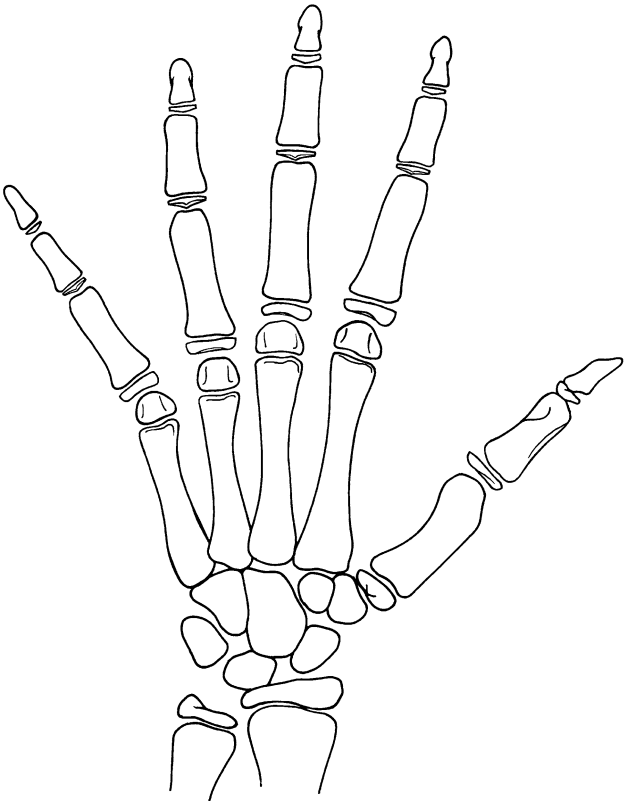

Fig. 1-B

Figs. 1-A and 1-B The key finding in stage 1 is that all of the digital epiphyses are not covered. In this case, the patient is at the end of stage 1. Particularly noticeable at the third middle phalanx, the epiphysis is not as wide as the metaphysis. More often, this finding is most noticeable on the fifth middle phalanx and metacarpal head.

various skeletal maturity stages from the Tanner-Whitehouse-III RUS descriptions were developed into a radiographic skeletal maturity classification system based on when the individual bone changes occurred in relation to the curve acceleration phase in girls with idiopathic scoliosis. It was found that, with increasing age, the various bones underwent changes in an orderly sequence compared with each other. Specifically, the digits on the ulnar side of the hand (the fourth and fifth digits) are the last to have fully covered epiphyses (Tanner-Whitehouse-III stage F), the proximal epiphyses cap their metaphyses slightly before the distal epiphyses do (stage G), the distal phalangeal physes close before the proximal and middle phalangeal physes do (stages H and I), the digital physes close before the metacarpal physes do, and the distal radial physis closes last. As all of the digits are available on the skeletal age anteroposterior radiograph of the hand and wrists, there was no need to restrict the evaluation to the first, third, and fifth digits as is done in the Tanner-Whitehouse-III method. The modified staging system underwent several iterations to clarify the descriptions of the stages, and the specific Tanner-Whitehouse-III descriptions were then related to specific Greulich and Pyle atlas[4] radiographs (Table I). All iterations underwent similar testing with use of different radiographs from the same subject cohort. A detailed description of the method relating the specific Tanner-Whitehouse-III descriptors and references to Greulich and Pyle atlas findings, a self-instructional PowerPoint presentation, and a self-assessment examination with detailed explanations were developed and provided to six reviewers. The reviewers included the originator of the current simplified method (J.O.S.), three additional practicing pediatric orthopaedic surgeons (J.G.K., S.K., J.F.M. III), one radiologist (J.A.B.), and one third-year orthopaedic surgery resident (K.D.A.).

Detailed Description of the Method

We divided the skeletal maturity stages in our classification system in a manner corresponding with typical maturity stages. The infantile rapid stage (stage 0) was not studied in this series. The juvenile slow stage (stage 1) is the stage before the

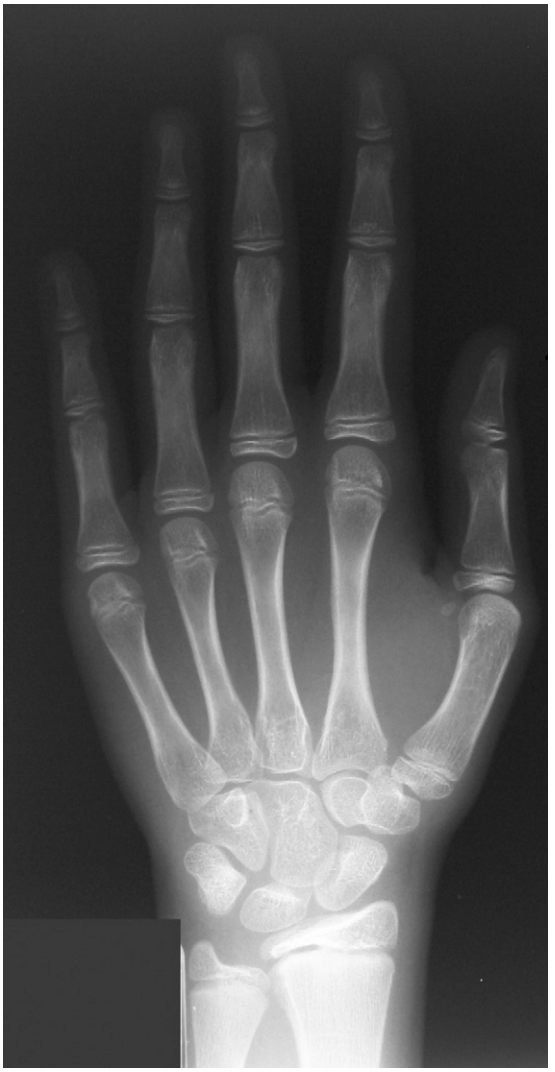

Fig. 2-A

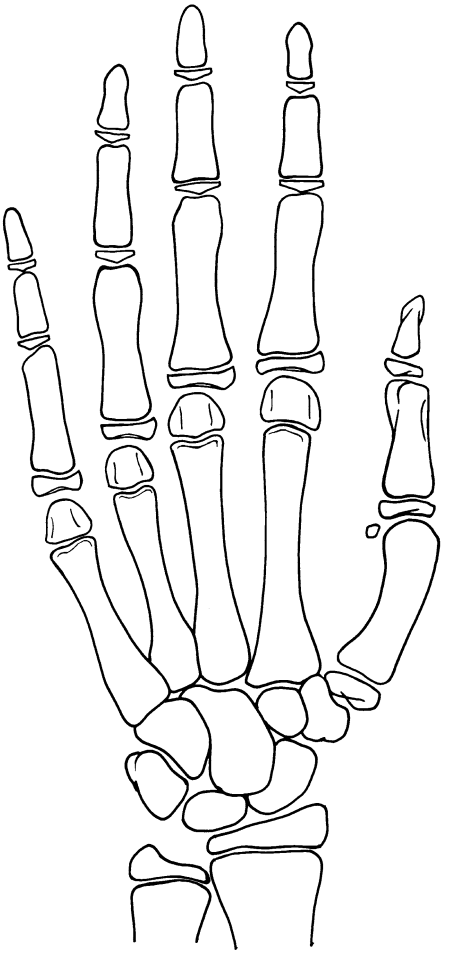

Fig. 2-B

Figs. 2-A and 2-B The key finding in stage 2 is that all of the digital epiphyses are covered. In this case, the patient is near the end of stage 2. The epiphyses are now all as wide as their metaphyses (covered). The dorsal and palmar surfaces of the metacarpal heads are clearly delineated. There is some capping of the second through fifth proximal phalanges but nowhere else.

beginning of the adolescent growth spurt. Secondary sexual characteristics are in Tanner stage 1[5]. The preadolescent slow stage (stage 2) is the stage during which the adolescent growth spurt has started but before the peak height velocity has occurred. Secondary sexual characteristics are in Tanner stage 2. The adolescent rapid stage—early (stage 3) is the stage in which the scoliosis curves start the curve acceleration phase, corresponding with the time of the peak height velocity. Girls may be in Tanner stage 2 or 3. In the adolescent rapid stage—late (stage 4), scoliosis curves continue to increase rapidly. Girls are typically in Tanner stage 3 but are still usually in Risser stage 0, with an open triradiate cartilage of the acetabulum. In the adolescent steady progression stage—early (stage 5), girls are still typically in Risser stage 0, but at this stage the triradiate cartilage is closed. Menarche rarely occurs before this stage. During the adolescent steady progression stage—late (stage 6), the Risser sign is usually positive. Menarche usually has occurred. The early mature stage (stage 7) usually corresponds to Risser stage 4, and scoliosis progression can still occur. The mature stage (stage 8), which corresponds to Risser stage 5 and completion of growth, was not studied.

Radiographic Features of the Stages

Unlike the Tanner-Whitehouse-III method, which uses just the first, third, and fifth digits, this simplified method takes advantage of the natural concordance of all of the digits to provide every possible clue for accurate maturity staging. The Tanner-

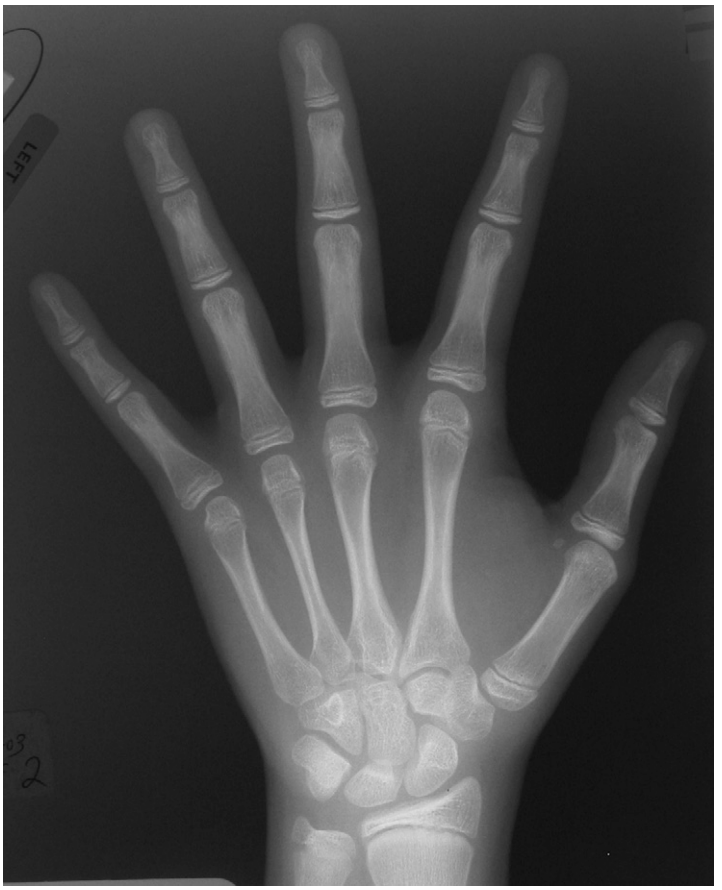

Fig. 3-A

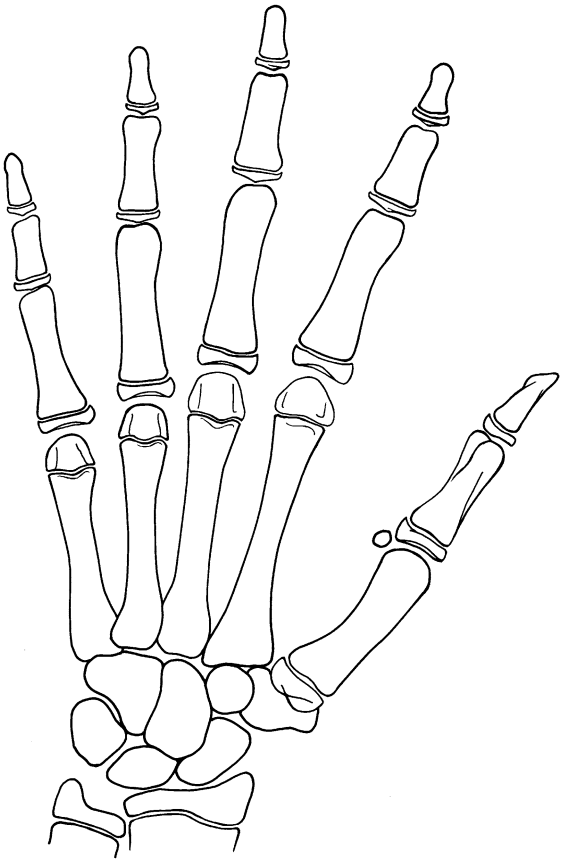

Fig. 3-B

Figs. 3-A and 3-B The key finding in stage 3 is that the preponderance of the epiphyses cap their metaphyses. The capping is a small bend over the metaphyseal edge. In the metacarpals, the second through fifth heads are wider than the metaphyses. The epiphyses cap the thumb metacarpal and all of the digits.

Whitehouse-III descriptors for the third and fifth digits should also be used for the second and fourth digits. Because skeletal maturity occurs as a continuum rather than as discrete stages, every effort has been made to describe the beginning and end of each stage, which is essential for the classification's reliability. If a patient has not fully reached the next radiographic stage, he or she is still considered to be in the less mature stage.

Individual Stages

Stage 1: Juvenile Slow Stage

The key feature of this stage is that all of the digital epiphyses are not covered (Figs. 1-A and 1-B). This stage lasts until all of the epiphyses are covered. The term "covered" means that the epiphysis is as wide as the metaphysis (Tanner-Whitehouse-III stage F) (see Appendix). Once the epiphyses are fully covered, the patient has entered the next phase.

The best place to look is often the middle phalanx of the fifth digit because the epiphysis often seems to be the last to be covered.

This stage can be compared with Greulich and Pyle atlas Female Standard 17 (skeletal age, eight years and ten months), which shows that the epiphyses of the middle and proximal phalanges of the fifth digit and those of all of the metacarpals are not covered. This is similar to Male Standard 22 (skeletal age, twelve years and six months).

Stage 2: Preadolescent Slow Stage

The key feature of this stage is that all digital epiphyses are covered (Figs. 2-A and 2-B). This stage lasts until all of the epiphyses cap their metaphyses (Tanner-Whitehouse-III stage G) (see Appendix).

The metacarpal dorsal and volar surfaces are clearly delineated; this finding corresponds with "covered" epiphyses.

This stage can be compared with Greulich and Pyle atlas Female Standard 18 (skeletal age, ten years), which shows all of the epiphyses covering their metaphyses. This is similar to a male skeletal age of thirteen years, at which there is early epiphyseal capping in some, but not in all, of the digits.

Stage 3: Adolescent Rapid Stage—Early

The key feature of this stage is that the preponderance of the epiphyses cap their metaphyses (Tanner-Whitehouse-III

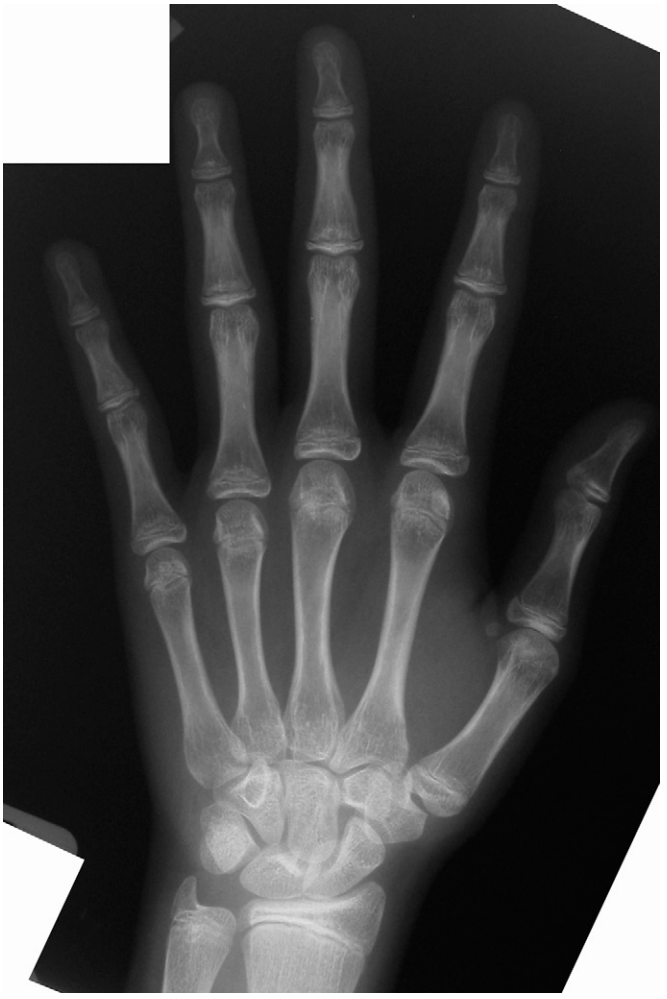

Fig. 4-A

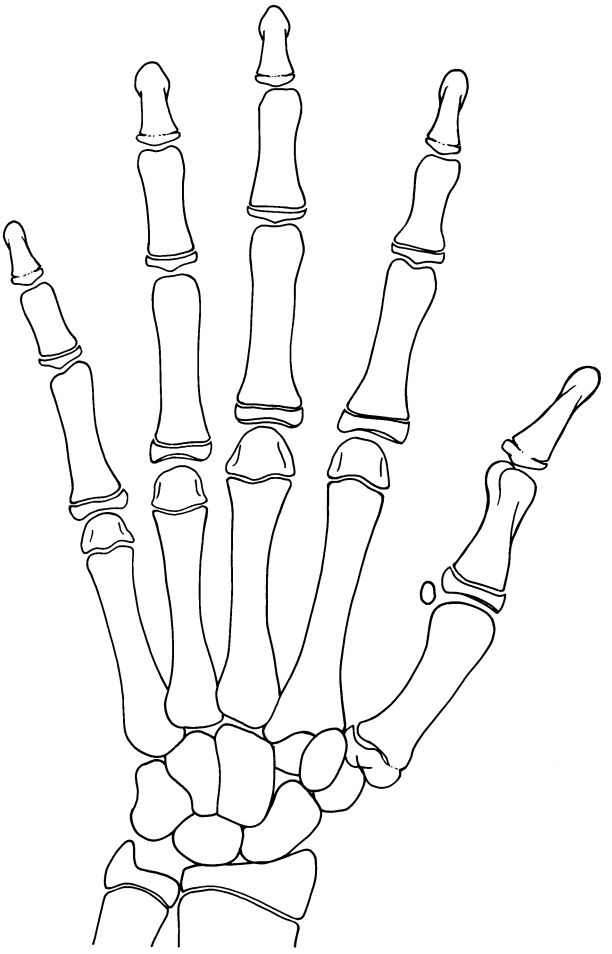

Fig. 4-B

Figs. 4-A and 4-B The key feature of stage 4 is the beginning of distal phalangeal physeal closure. In this case, the distal phalangeal physis of the thumb appears closed on the radiograph, but its rotation makes it more difficult to see than the other digits. The distal phalangeal physes of the second through fifth digits are beginning to close. Physeal closure starts in the center of the physis. The remainder of the digits are fully capped, and the metacarpal heads are wider than their metaphyses as in stage 3.

stage G) (see Appendix) (Figs. 3-A and 3-B). Epiphyses normally correspond with each other. If all but a couple of the epiphyses are capped, then the radiograph falls into this stage.

This is the crucial stage to identify, and it is important to familiarize oneself with all of the Tanner-Whitehouse-III stage-G criteria (see Appendix). Certain features should be seen. Specifically, in this stage, the middle and proximal phalangeal and thumb metacarpal epiphyses are capped. In addition, the heads of the second through fifth metacarpals are wider than their metaphyses and have clear dorsal and volar surfaces.

There are some important nuances. Specifically, it can be difficult to identify capping on the distal phalanges, particularly those of the thumb, which is usually rotated. If the distal phalanges cannot be seen well, they should not be used. However, if the distal phalanges are clearly capped, then the other phalanges probably are capped as well, and the patient is in this stage.

The Greulich and Pyle atlas Female Standards 19 and 20 (skeletal age, eleven and twelve years, respectively) and Male Standards 24 and 25 (skeletal age, thirteen years and six months and fourteen years, respectively) correspond with this stage.

Stage 4: Adolescent Rapid Stage—Late

The key feature of this stage is that one or more of the distal phalanges are beginning to demonstrate physeal closure (Figs. 4-A and 4-B). All of the features of the early adolescent rapid stage are present, except that one or more of the distal phalanges are clearly beginning to demonstrate physeal closure (Tanner-Whitehouse-III stage H). Stage H means that a portion but not the entire physis is clearly closed. A narrowed physis without partial closure is *not* in this stage (see Appendix).

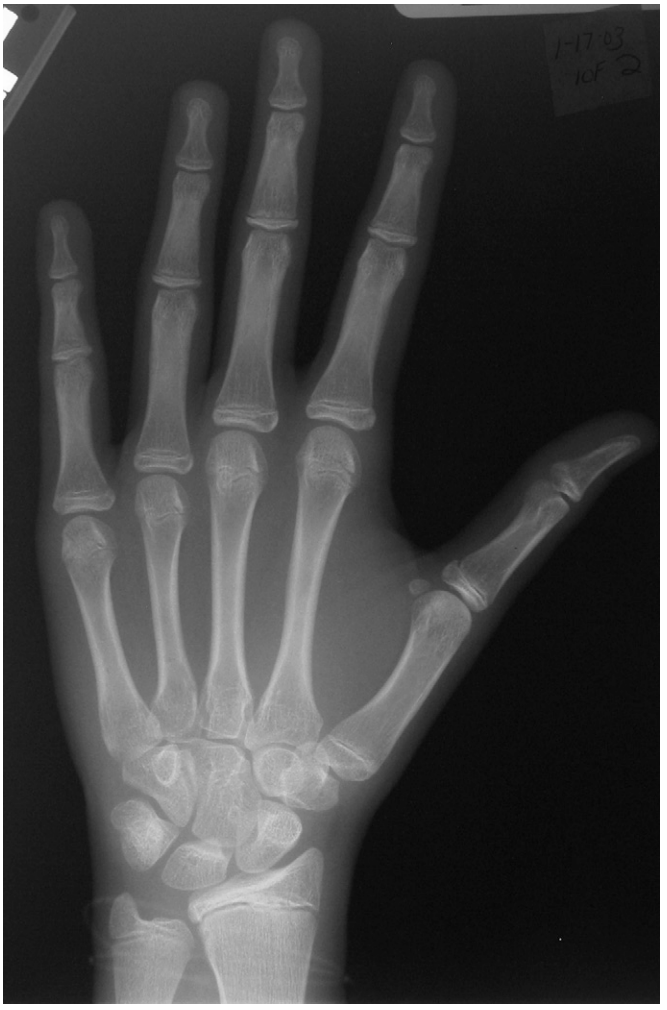

Fig. 5-A

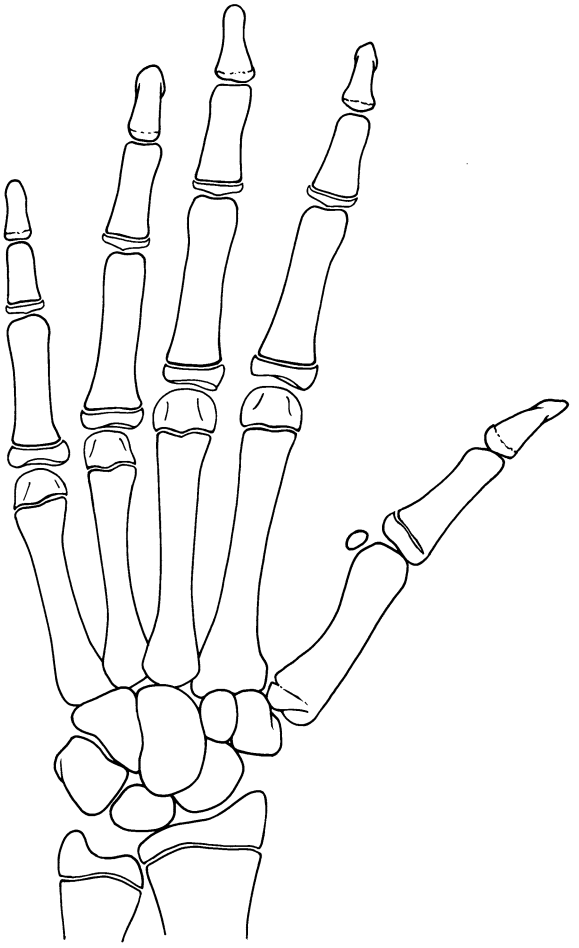

Fig. 5-B

Figs. 5-A and 5-B The key feature of stage 5 is that all of the distal phalangeal physes are closed. If there is any black (growth cartilage) rather than white (physeal scar) in the physis, the radiograph remains as stage 4. The remainder of the epiphyses are capped, and the metacarpal heads are wider than the metaphyses (see stage 3 as demonstrated in Figs. 3-A and 3-B).

If the distal phalanx of the thumb is clearly visible, it may be used. Often, it is too rotated to see clearly and should not be used as the sole criterion for this stage.

These features are best seen on Greulich and Pyle atlas Male Standard 15 (skeletal age, fifteen years) for the fourth and fifth digits. The Female Standard 21 (skeletal age, thirteen years) shows these features for the second, third, and fourth digits but overall is not in this stage because the thumb has a closed distal physis and the fifth digit is cut off from the radiograph.

Stage 5: Adolescent Steady Progression Stage—Early

The key feature of this stage is that all distal phalangeal physes are closed (Tanner-Whitehouse-III stage I) whereas the proximal phalangeal physes are open (Figs. 5-A and 5-B). Tanner-Whitehouse-III stage I indicates there is a white (not black) physeal scar (see Appendix). The first metacarpal physis closes at about the same time or a little later than the distal phalangeal physes do. This stage lasts until either the middle or proximal phalangeal physes are clearly closing (Tanner-Whitehouse-III stage H).

The Greulich and Pyle atlas Female Standard 22 (skeletal age, thirteen years and six months) and Male Standard 27 (skeletal age, fifteen years and six months) are in this stage.

Stage 6: Adolescent Steady Progression Stage—Late

The key feature of this stage is that the proximal and middle phalangeal physes are clearly closing (Tanner-Whitehouse-III stage H) (Figs. 6-A and 6-B). This stage lasts until the phalangeal and metacarpal epiphyses are closed. The finger metacarpals may have a prolonged physeal scar. Please note that a fully white physeal scar is closed (stage I), whereas any remaining black physis is not fully closed (stage II or earlier).

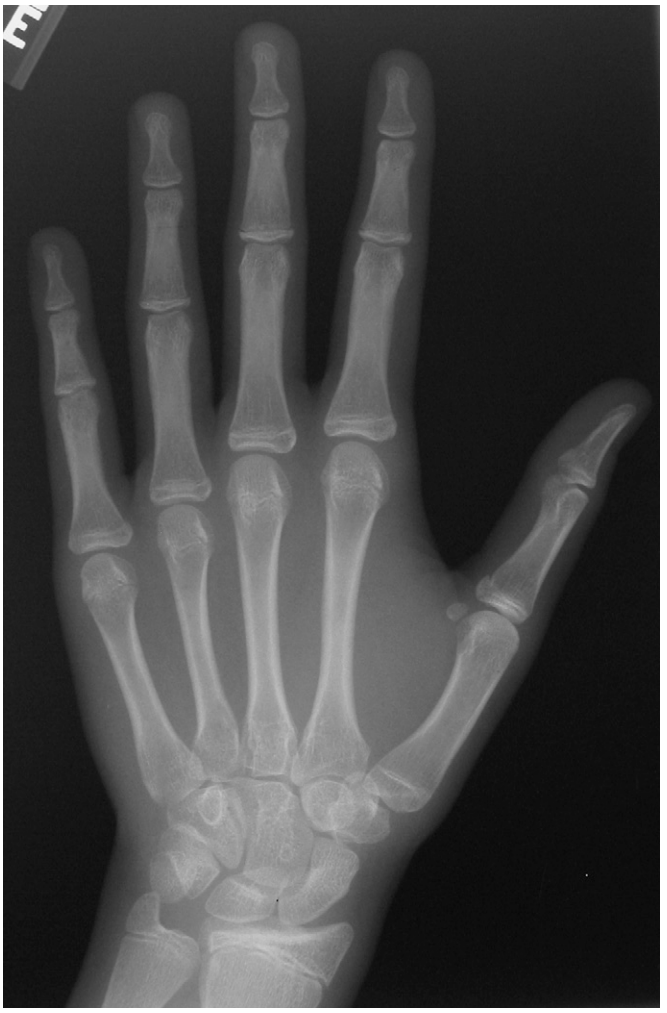

Fig. 6-A

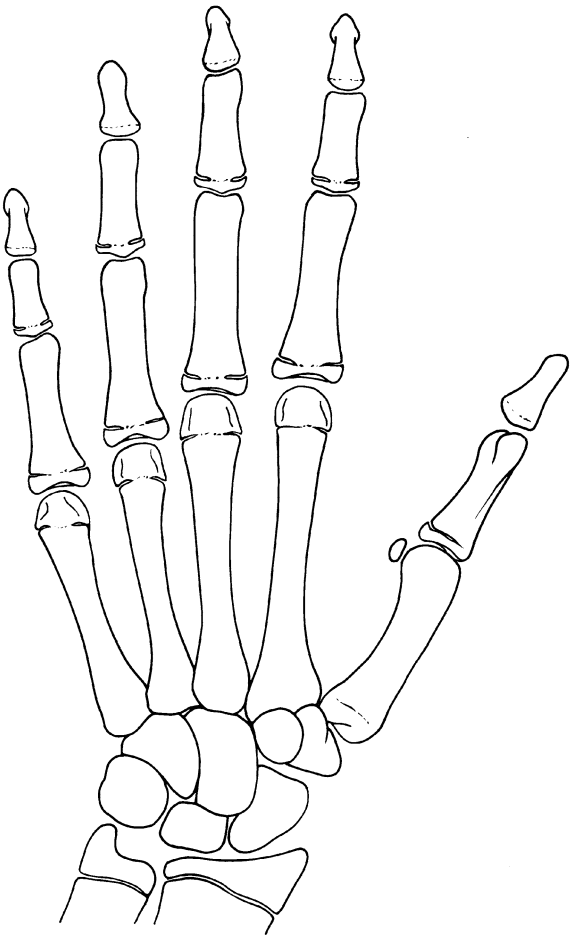

Fig. 6-B

Figs. 6-A and 6-B The key feature of stage 6 is that some of the proximal or middle phalangeal physes are closing. In this stage, all of the distal phalangeal physes are closed (see stage 5 as demonstrated in Figs. 5-A and 5-B) and some of the proximal and middle phalangeal physes are closing. The second through fifth metacarpal physes typically stay open longer than those of the other small bones. When the metacarpal physes go from black (physis) to white (physeal scar), then the physis is considered ''closed'' or ''fused'' for this staging system.

The Greulich and Pyle atlas Female Standard 23 (skeletal age, fourteen years) and Male Standard 28 (skeletal age, sixteen years) correspond with this stage.

Stage 7: Early Mature Stage

The key feature of this stage is that only the distal radial physis is open; the ulna is disregarded (Figs. 7-A and 7-B). All of the digital physes are closed (no black remains, corresponding with Tanner-Whitehouse-III stage I) (see Appendix).

The distal radial physis remains open (Tanner-Whitehouse-III stage G or H, i.e., some black remains).

The metacarpal white physeal scars may persist. If a black physis remains, the patient is still in stage 6.

Greulich and Pyle atlas Female Standard 24 (skeletal age, fifteen years) and Male Standard 29 (skeletal age, seventeen years) correspond with this stage.

Stage 8: Mature Stage

The key feature of this stage is that the distal radial physis is completely closed, i.e., no black remains (Tanner-Whitehouse-III stage I) (Figs. 8-A and 8-B) (see Appendix). The Greulich and Pyle atlas Female Standard 26 (skeletal age, seventeen years) and Male Standard 31 (skeletal age, nineteen years) correspond with this stage.

Reliability Testing

Thirty skeletal age radiographs, representing the maturity spectrum of the subjects followed through their growth spurt, were selected independently of the reviewers, digitized, and blinded. For the entire cohort, there were a total of 161 skeletal age radiographs, with an average of 7.3 radiographs (range, four to eleven radiographs) per subject. As the radiographs were randomly selected across the spectrum, the number of

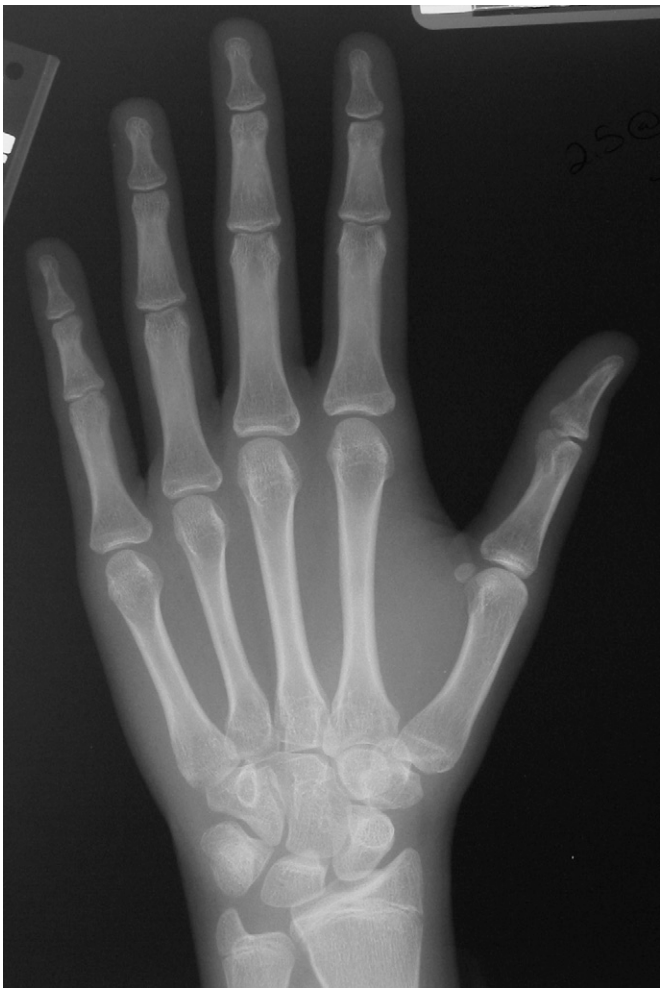

Fig. 7-A

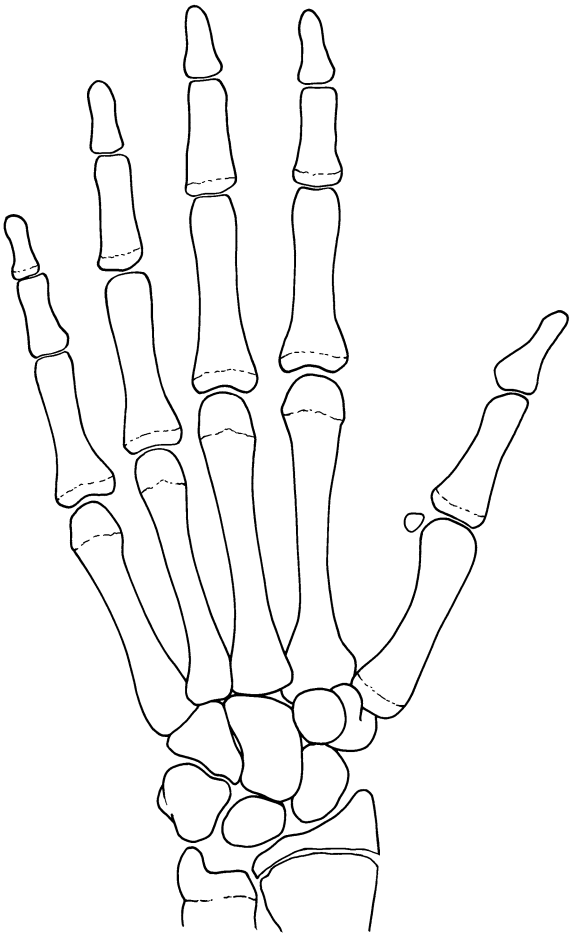

Fig. 7-B

Figs. 7-A and 7-B The key feature of stage 7 is that all of the physes, except for those of the distal parts of the radius and ulna, are closed. There are no black physeal lines remaining for any digits or metacarpals. Stage 7 ends when the distal radial and ulnar physes are white (physeal scar) rather than black (persistent physis).

radiographs from a single subject varied for the different trials. No subject had more than two radiographs in any particular trial. The reviewers were instructed to view the radiographs on a high-resolution computer monitor (at least 1024 × 768 pixels) and to refer to the reference materials. Two batches of the radiographs, with separate numbering systems, were reviewed no sooner than twenty-four hours apart to determine interobserver and intraobserver reliability. An additional interobserver variability test was performed with a separate selection of radiographs and an atlas to identify whether use of a specific atlas improved interobserver reliability.

Validity Testing

All 161 skeletal age radiographs from our previous study[1] were then classified according to the simplified skeletal maturity staging system and were compared with the timing of the curve acceleration phase. The magnitude of the main scoliotic curve at each skeletal maturity stage was then related to the final curve magnitude and the Lenke curve type[1,6]. Correlation of the stages with the curve acceleration phase was compared with the values from our previous study[1].

Statistical Methods

All statistical analyses, including statistical analysis of the exact matches and determination of the weighted and unweighted kappa coefficients, were performed with use of SAS software (version 9.1; SAS Institute, Cary, North Carolina). Pearson correlation coefficients were computed with use of linear regression for the maturity scores compared with time relative to the curve acceleration phase. Curve-fitting was performed with use of TableCurve 2D (version 5.01; SPSS, Chicago, Illinois).

Logistic regression analysis of curve magnitude and maturity stage to predict the probability of the final curve being >50° was done for the most common curve types (Lenke types 1 and 3) and separately for all other curve types. To estimate the probability of the final curve being >50°, we

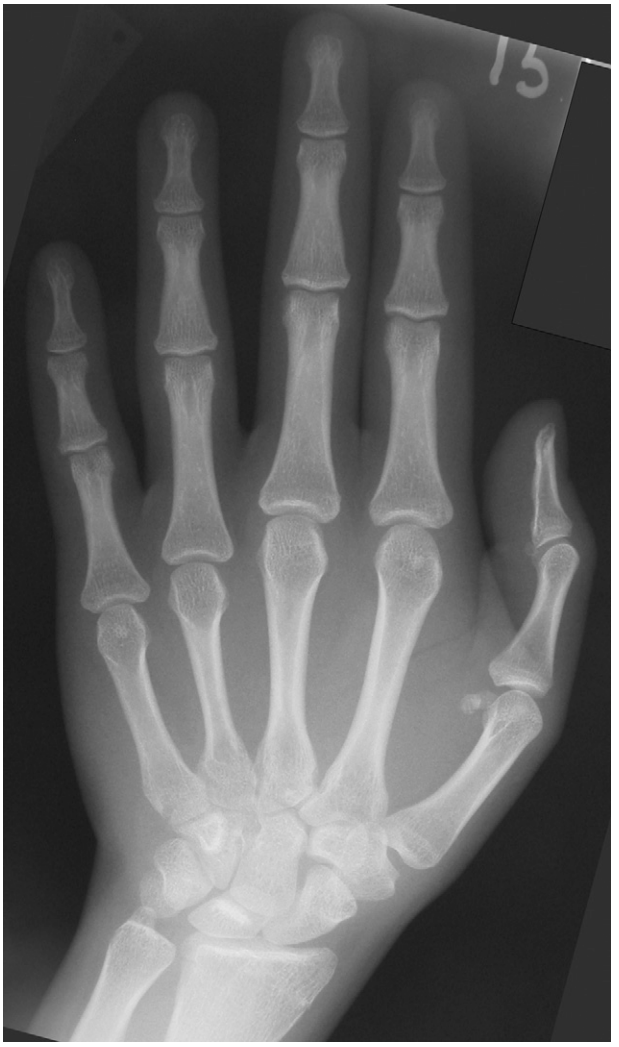

Fig. 8-A

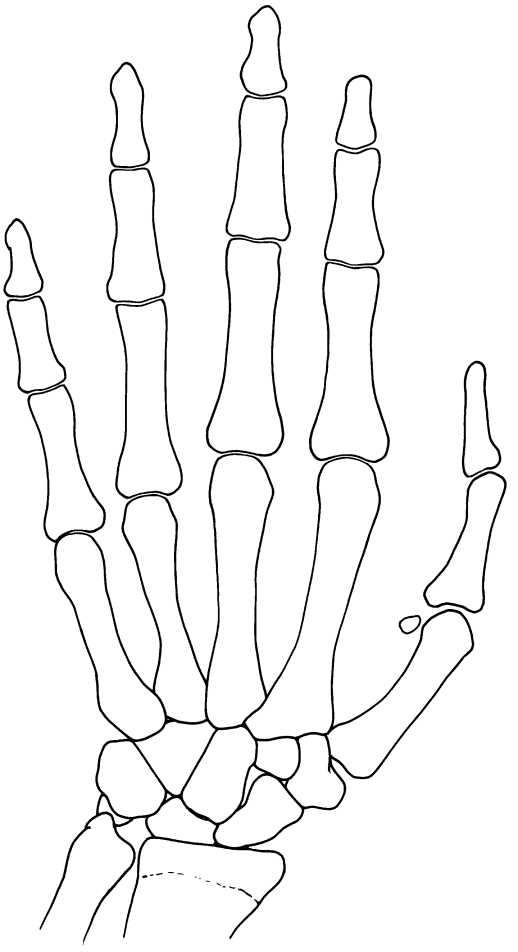

Fig. 8-B

Figs. 8-A and 8-B The key finding of stage 8 is closure of all physes. The residual physeal lines may appear white, but there are no black physeal lines.

first computed the logistic regression equation and then used that equation to calculate the probability for all combinations of curve magnitude, maturity stage, and Lenke curve type.

Results

Reliability

The samples were well distributed across the spectrum except for stage 8 (the mature stage, with all physes closed), which was not included in the study. The intraobserver and interobserver reliability results, with test-retest values given as kappa coefficients, are shown in Table II. The simple or unweighted kappa coefficient becomes larger (approaching 1.0) with increased exact matches. The weighted kappa coefficient does not require exact matches but gives greater importance to closer matches[7].

As shown in Table II, the simple or unweighted intraobserver kappa for the originator was 0.96 and the weighted kappa was 0.99. The simple or unweighted interobserver kappa for the originator with Orthopaedist 1 was 0.84 for trial 1 and 0.72 for trial 2; the weighted kappa values for those same trials were 0.93 and 0.90, respectively. The average intraobserver simple or unweighted kappa for all pairs of observers was 0.88, and the average weighted kappa was 0.96. The percentage of exact matches between readings for each rater was 89%, and 100% of the differences were within one unit. The average inter-rater simple kappa value was 0.71, and the average weighted kappa was 0.89. Seventy-one percent of the time, a pair of reviewers agreed completely, and 97% of the time, the two reviewers either agreed completely or differed by only one stage. The agreement was highest between the most experienced raters. An atlas created specifically for this staging system was made available to the reviewers during a second set of tests to see if it improved inter-rater reliability, but we found no substantive change in the reliability numbers.

TABLE II Interobserver and Intraobserver Reliability*

	Kappa Value					
	Originator	Orthopaedist 1	Radiologist	Orthopaedist 2	Orthopaedist 3	Resident
Originator						
Trial 1	*0.96*	0.84	0.88	0.72	0.80	0.80
Trial 2	*0.99*	0.72	0.88	0.64	0.80	0.63
Orthopaedist 1						
Trial 1	0.93	*0.88*	0.80	0.84	0.64	0.72
Trial 2	0.90	*0.96*	0.76	0.68	0.75	0.56
Radiologist						
Trial 1	0.96	0.92	*0.88*	0.68	0.83	0.76
Trial 2	0.96	0.91	*0.96*	0.60	0.84	0.51
Orthopaedist 2						
Trial 1	0.91	0.95	0.89	*0.84*	0.60	0.68
Trial 2	0.86	0.87	0.85	*0.95*	0.52	0.48
Orthopaedist 3						
Trial 1	0.93	0.87	0.95	0.87	*0.88*	0.75
Trial 2	0.92	0.90	0.94	0.82	*0.96*	0.60
Resident						
Trial 1	0.87	0.80	0.86	0.78	0.83	*0.92*
Trial 2	0.88	0.86	0.87	0.75	0.88	*0.97*

*Unshaded cells without italics = simple interobserver kappa values; unshaded cells with italics = simple intraobserver kappa values; shaded cells without italics = weighted interobserver kappa values; and shaded cells with italics = weighted intraobserver kappa values.

Length of Stages and Correlation with Curve Behavior (Validity)

The stages are not linear. The number of sequential six-month visits for each stage that could be identified as having a lower stage on a previous visit and a higher stage on a subsequent visit was counted. Curve-fitting demonstrated the progression of the stages relative to the curve acceleration phase. The overall progression from stage 2 to stage 8 occurred over four years, starting about twelve months before the curve acceleration phase and lasting three years beyond the start of the curve acceleration phase. Stage 4 occurred about twelve months after the curve acceleration phase began, and stage 8 occurred about thirty-six months after the curve acceleration phase began. Once the patient reached stage 4, each stage progressed rapidly, typically within six months of the previous stage, until stage 7. The Pearson correlation coefficient of the maturity measurements with the curve acceleration phase was 0.91 ($p < 0.001$).

Prediction of Curve Behavior (Logistic Regression)

The risk of progression is determined with use of logistic regression methods. The estimated probability of the final curve being >50° for Lenke type-1 curves (single thoracic curves) and type-3 curves (double major curves with a predominant thoracic curve) is shown in Table III. The average risk of progression is that determined by logistic regression. Where there was no corresponding data point, a logistically created one was added so that the probability could be estimated for all combinations of main curve magnitude and maturity stage. The 95% confidence intervals are shown in parentheses. Wide confidence intervals reflect a small sample size, indicating that the probability may require modification with a larger sample, and therefore those particular cells should be interpreted with caution.

Table III shows the relationship between maturity stage, curve size, and the probability of the curve progressing to >50°. The unshaded cells correspond with combinations for which surgery would be a plausible treatment if >50° at maturity is accepted as a threshold for surgical treatment. For example, if a patient presents with a 25° curve, the prognosis would clearly depend on the maturity stage of the patient. For stages 1 and 2, surgery is quite likely; conversely, for stages 4 to 8, surgery is unlikely. For Stage 3, the probable outcome is not so clear.

As shown in Table III, a 25° curve had substantially different chances of reaching 50°, depending on the skeletal maturity stage. Since stages 1 through 5 all typically occur before

TABLE III Logistic Projection of the Probability of Lenke Type-1 and Type-3 Curves Progressing to Surgery Assuming a >50° Threshold*†

Curve	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5	Stage 6	Stage 7, 8
10°	2% (0% to 40%)	0% (0% to 15%)	0% (0% to 0%)	0% (0% to 0%)	0% (0% to 0%)	0% (0% to 0%)	0% (0% to 1%)
15°	23% (4% to 69%)	11% (1% to 58%)	0% (0% to 2%)	0% (0% to 0%)	0% (0% to 0%)	0% (0% to 0%)	0% (0% to 7%)
20°	84% (40% to 98%)	92% (56% to 99%)	0% (0% to 14%)	0% (0% to 1%)	0% (0% to 1%)	0% (0% to 1%)	0% (0% to 26%)
25°	99% (68% to 100%)	100% (92% to 100%)	29% (3% to 84%)	0% (0% to 5%)	0% (0% to 5%)	0% (0% to 2%)	0% (0% to 64%)
30°	100% (83% to 100%)	100% (98% to 100%)	100% (47% to 100%)	0% (0% to 27%)	0% (0% to 22%)	0% (0% to 11%)	0% (0% to 91%)
35°	100% (91% to 100%)	100% (100% to 100%)	100% (89% to 100%)	0% (0% to 79%)	0% (0% to 65%)	0% (0% to 41%)	0% (0% to 98%)
40°	100% (95% to 100%)	100% (100% to 100%)	100% (98% to 100%)	15% (0% to 99%)	0% (0% to 94%)	0% (0% to 83%)	0% (0% to 100%)
45°	100% (98% to 100%)	100% (100% to 100%)	100% (100% to 100%)	88% (2% to 100%)	1% (0% to 99%)	0% (0% to 98%)	0% (0% to 100%)

*Unshaded cells correspond with combinations of curve size and maturity stage for which surgery would be a plausible treatment if >50° at maturity is accepted as the threshold for surgical treatment. Shaded cells correspond with combinations for which surgery would not be a plausible treatment. †Cells with wide 95% confidence intervals (shown in parentheses) correspond with groups that had too few patients for accurate estimates (or groups that had no patients) and should be interpreted with caution.

iliac apophyseal ossification, a 25° curve in a girl who is in Risser stage 0 and skeletal stage 2 has a 100% (confidence interval, 92% to 100%) chance of reaching >50° despite bracing. On the other hand, the same 25° curve in a girl who is in Risser stage 0 but skeletal stage 4 has an essentially 0% (confidence interval, 0% to 5%) chance of reaching >50°. A 25° curve in a girl who is in skeletal stage 3 has a 29% chance of reaching >50°, but the confidence interval (3% to 84%) is too wide for accurate estimation. To predict the probability of the curve being >50° to within ±10% (with 95% confidence), a power analysis suggests thirty-six or more patients must be followed through to skeletal maturation for each curve type.

Discussion

Reasons for the Method

Previous work identified the DSA scores from the Tanner-Whitehouse-III RUS method as the maturity indicator most closely correlated with curve behavior in a cohort of girls with idiopathic scoliosis. The DSA score, however, is cumbersome to obtain and is more appropriate for a research study than for clinical use unless reliable, rapid computerized measurements can be obtained, as suggested by several investigators[8-11]. Because of these limitations, we sought to develop a reliable but rapid skeletal maturity assessment system based on the findings used to determine the DSA score. Like any maturity indicator, skeletal maturity assessment can be problematic[9]. Skeletal maturity is a continuum and the changes over time can be subtle, yet clinical practice requires precise determination. In their description of the requirements of an orthopaedic classification system, Garbuz et al.[7] indicated that a good classification system should be reliable as evidenced by strong kappa values for both intraobserver and interobserver reliability through testing that is balanced across the spectrum of the classification. Reliability testing demonstrated that the skeletal maturity system developed here is highly reliable as a clinical staging system, and our experience indicates that only a few seconds are required to classify the maturity once the system is learned in detail.

Learning the Method

Users should familiarize themselves with the system by reviewing the detailed description of the method, viewing the presentation, performing the self assessment, and referring to the table as needed. Whether or not to have standardized radiographs available for use during the grading will depend on the reader's experience. Both the figures in this article and the referenced plates in the Greulich and Pyle atlas demonstrate the findings at each stage. Readers of the radiographs were very consistent with themselves when just using the descriptions and the Greulich and Pyle atlas as needed, and a specific atlas for this new method did not improve the reliability of the readings. We believe that this was because the Greulich and Pyle atlas provides excellent figures illustrating the various stages, and a specific atlas does not improve on what can be found in the more readily available Greulich and Pyle atlas. Also, once the reader is comfortable with the method, the Tanner-Whitehouse-III descriptors are very clear and do not require access to any radiographic atlas. Similar descriptors are found in the section on the Greulich and Pyle atlas describing the maturity indicators of individual bones.

When possible, a patient's skeletal maturity radiographs should be viewed in sequence rather than as isolated measurements because, as Tanner et al. noted[3], the changes are more obvious sequentially. While this was not possible for the test-retest, it is often possible in clinical practice and improves the assessments. We did find that there was a learning curve for the method. The orthopaedic resident in particular had substantial scoring improvements in the early iterations as compared with the other reviewers. While the kappa values were stronger for the more experienced reviewers, they remained good for the least experienced as well.

The Greulich and Pyle Atlas

While the Greulich and Pyle atlas provides excellent radiographs illustrating the various stages and is widely available, it must be used with caution. The atlas was developed from representative radiographs at various chronological ages in an era before timing relative to the peak height velocity, or even secondary sexual characteristics, was recognized as an important, quantifiable maturity indicator. Maturity near adolescence is better based on timing relative to the peak height velocity as normal adolescents have as much as a four-year variation in the timing of their growth spurt[12]. We concur with Tanner et al.[3] that the concept of skeletal maturity is more appropriate than skeletal age, and we prefer a system that does not have a specific "age" attached to the radiograph. Bones do not have a different age than patients, but their maturities differ just as children's do. Because of this, the time-intervals of the standards in the Greulich and Pyle atlas are problematic during adolescence, as noted by DiMéglio et al.[13]. In the present series, the mode of stage-3 six-month visits was two (range, one to three) six-month visits, which is substantially shorter than the Greulich and Pyle atlas would suggest. We found that the maturity stages, which are separated by about a year in the Greulich and Pyle atlas, actually occur much more rapidly in the middle stages. In our simplified skeletal maturity system, stage 3 is the most important phase because of its tight correlation with the curve acceleration phase. This stage corresponds with the Greulich and Pyle female skeletal ages of both eleven and twelve years because of the atlas's over-reliance on the carpal bones. The simplified staging system does not use the carpal bones because our previous work demonstrated that the carpus matured before CAP 0. Another problem with the Greulich and Pyle atlas is that the descriptors were designed to highlight many subtleties rather than being designed for reliability. The Tanner-Whitehouse-III classification system avoids the subtleties in favor of clarity. Our modification is therefore based on simple but precise findings to enhance both usability and reliability.

Correlation of the Stages with Scoliosis Curve Behavior

In addition to reliability, a useful maturity staging system for scoliosis must correlate with curve behavior. Similar to the DSA scores from which they are derived, the stages of this system are highly correlated with the timing relative to the curve acceleration phase (Pearson $r = 0.91$). This correlation is stronger than those associated with either the Risser sign or the Greulich and Pyle atlas[1]. The correlation is even stronger using an S-shaped curve fit (a fourth-order polynomial [$r = 0.94$, $r^2 = 0.88$]), reflecting the rapid skeletal maturity at the middle stages (stages 4 through 6) and the slower maturity changes at either end.

Determining Curve Prognosis

Despite the relatively small number of patients in the present study, the curve magnitude from stage 2 onward was highly prognostic of the eventual curve magnitude. This was particularly true for the Lenke type-1 curves (main thoracic curves) and Lenke type-3 curves (double curves with the thoracic curve being larger than the lumbar curve)[6], for which the numbers were sufficiently large for analysis. In order to have estimates of the probability of the final curve being >50° within a 10% confidence interval, approximately thirty-six subjects with each curve pattern need to be followed through the stages to maturity to provide accurate prognosis for each curve pattern, maturity stage, and curve magnitude. From Table III, it is clear that those at greatest risk are patients with curves of ≥20° at stage 2 and of ≥30° at stage 3, for whom bracing is unlikely to succeed. The curve magnitude of >50° was chosen because it often represents a minimum threshold for surgery. The same procedure can be done for any threshold value, whether 45°, 60°, or ultimate curve size at maturity. Because the purpose of the present study was to identify proper maturity markers in patients with idiopathic scoliosis, and because commonly accepted bracing criteria were used, the present study should not be viewed as a natural history study but as a prospective study of maturity and curve progression typical of patients with similar maturity and curve sizes seen and treated in many North American clinical practices. While the design of the current study makes it possible to identify bracing failures, it cannot identify bracing successes. It remains important to further test these findings in both untreated patients and compliant patients who have been managed with bracing for each specific curve pattern and magnitude in a larger prospective study. As we do not know how our results were affected by bracing, specific recommendations regarding bracing must await the completion of a well-designed bracing study that is properly stratified, similar to the current study, according to curve magnitude, pattern, and skeletal maturity.

The simplified Tanner-Whitehouse-III system for the classification of skeletal maturity for patients with idiopathic scoliosis is rapid and reliable, and it correlates highly with curve behavior. It also has the advantages of the general familiarity with hand skeletal maturity, reference to the ubiquitous Greulich and Pyle atlas, and the avoidance of a complex scoring system. All of these qualities appear to make it useful as a clinical maturity indicator in patients with idiopathic scoliosis. In addition, the probability estimates of progression to surgery can assist clinicians in discussing the likely outcome of scoliosis management with the patient and parents.

Appendix

eA The Tanner-Whitehouse-III descriptions along with radiographs selected to demonstrate the method are available with the electronic versions of this article, on our web site at jbjs.org (go to the article citation and click on "Supplementary Material") and on our quarterly CD-ROM (call our subscription department, at 781-449-9780, to order the CD-ROM). ■

James O. Sanders, MD
Department of Orthopaedics and Rehabilitation,
University of Rochester,
601 Elmwood Avenue, Rochester, NY 14624.
E-mail address: james_sanders@urmc.rochester.edu

Joseph G. Khoury, MD
University of Alabama at Birmingham,
316 A.C.C., 1600 7th Avenue South, Birmingham, AL 35233-1711.
E-mail address: joseph.khoury@ortho.uab.edu

Shyam Kishan, MD
Loma Linda University Medical Center, 11406 Loma Linda Drive,
Loma Linda, CA 92354

Richard H. Browne, PhD
Texas Scottish Rite Hospital for Children,
2222 Welborn Street, Dallas, TX 75219

James F. Mooney III, MD
Department of Orthopedic Surgery,
Medical University of South Carolina, 96 Jonathan Lucas Street,
CSB 708, Charleston, SC 29425

Kali D. Arnold, MD
Department of Orthopedic Surgery, Allegheny General Hospital,
1307 Federal Street, Pittsburgh, PA 15212

Sharon J. McConnell, MS
Shriners Hospitals for Children, 1645 West 8th Street, Erie, PA 16505

Jeanne A. Bauman, MD
Hamot Medical Center, 201 State Street, Erie, PA 16550

David N. Finegold, MD
Department of Medical Genetics, University of Pittsburgh
Medical Center, 3705 Fifth Avenue, Pittsburgh, PA 15213

References

1. Sanders JO, Browne RH, McConnell SJ, Margraf SA, Cooney TE, Finegold DN. Maturity assessment and curve progression in girls with idiopathic scoliosis. J Bone Joint Surg Am. 2007;89:64-73.

2. Urbaniak JR, Schaefer WW, Stelling FH 3rd. Iliac apophyses. Prognostic value in idiopathic scoliosis. Clin Orthop Relat Res. 1976;116:80-5.

3. Tanner JM, Healey MJR, Goldstein H, Cameron N. Assessment of skeletal maturity and prediction of adult height (TW3 method). 3rd ed. London: Saunders; 2001.

4. Greulich WW, Pyle SI. Radiographic atlas of skeletal development of the hand and wrist. 2nd ed. Stanford, CA: Stanford University Press; 1959.

5. Tanner JM. Growth and endocrinology of the adolescent. In: Gardner LI, editor. Endocrine and genetic diseases of childhood. 2nd ed. Philadelphia: W.B. Saunders; 1974. p 14.

6. Lenke LG, Betz RR, Harms J, Bridwell KH, Clements DH, Lowe TG, Blanke K. Adolescent idiopathic scoliosis: a new classification to determine extent of spinal arthrodesis. J Bone Joint Surg Am. 2001;83:1169-81.

7. Garbuz DS, Masri BA, Esdaile J, Duncan CP. Classification systems in orthopaedics. J Am Acad Orthop Surg. 2002;10:290-7.

8. Cao F, Huang HK, Pietka E, Gilsanz V. Digital hand atlas and web-based bone age assessment: system design and implementation. Comput Med Imaging Graph. 2000;24:297-307.

9. Cox LA. Tanner-Whitehouse method of assessing skeletal maturity: problems and common errors. Horm Res. 1996;45 Suppl 2:53-5.

10. Tanner JM, Oshman D, Lindgren G, Grunbaum JA, Elsouki R, Labarthe D. Reliability and validity of computer-assisted estimates of Tanner-Whitehouse skeletal maturity (CASAS): comparison with the manual method. Horm Res. 1994;42:288-94.

11. Tanner JM, Gibbons RD. A computerized image analysis system for estimating Tanner-Whitehouse 2 bone age. Horm Res. 1994;42:282-7.

12. Parent AS, Teilmann G, Juul A, Skakkebaek NE, Toppari J, Bourguignon JP. The timing of normal puberty and the age limits of sexual precocity: variations around the world, secular trends, and changes after migration. Endocr Rev. 2003;24:668-93.

13. Diméglio A, Charles YP, Daures JP, de Rosa V, Kaboré B. Accuracy of the Sauvegrain method in determining skeletal age during puberty. J Bone Joint Surg Am. 2005;87:1689-96.