ANTHROPOLOGIE • XLV/1 • pp. 97–102 • 2007

LENKA DUCHAJOVÁ, PETR SEDLAK, MARCELA DVOŘÁKOVÁ

THE BIOLOGICAL AGE DETERMINATION OF CHILDREN WITH GROWTH AND DEVELOPMENT DISORDERS – A VALIDATION OF THE BONE AGE METHODS

ABSTRACT: The authors used the TW2 and TW3 methods of bone age assessment to statistically evaluate a representative group of 94 roentgenograms of the hand and distal part of the forearm of children with growth and development disorders. These methods were compared and validated for diagnostic use in the determination of skeletal maturation in a recent population of children. Data obtained from a group of 36 patients were used to compare the established Czech Kapalin method of bone age determination (Kapalin, Picko 1964) with the TW2 and TW3 foreign standard methods (Tanner et al. *1975, 2001). Initial assessment of methods was made across all groups of roentgenograms. Further analysis and validation was performed within the divisions of gender, age and diagnostic groups. A significant development trend was observed in all roentgenograms and verified by statistical methods. The study showed a significant difference in skeletal age determination between the TW3 (RUS) and TW2 (RUS) methods, with the TW3 (RUS) method having an average estimated value of one year less than the TW2 (RUS) method. Based on the author's evaluation, the TW3 method (RUS compartment) is considered a more accurate method of determining skeletal maturation in clinical practice. Significant differences in all 36 roentgenogram groups (based on the 1% alpha-power) were made with TW3 (RUS) and Kapalin methods. Following the division of the roentgenograms into diagnostic groups, the differences between the TW3 and Kapalin methods were not statistically significant.*

KEY WORDS: Bone age – Skeletal maturation – Growth disorders – TW3 method – TW2 method – Kapalin

INTRODUCTION

Bone age is a way of describing the degree of maturation of children's bones. The skeletal maturation, or bone age of children can be assessed by comparing the ossification and maturation of the epiphysis of the hands and wrists against standards. The most commonly used method is based on a single roentgenogram of the fingers, hand and wrist. A hand is easily x-rayed with minimal radiation and shows many bones in a single view.

The following methods are currently used clinically:

– GP (The Greulich and Pyle method [Greulich, Pyle 1959]), an American method that compares the radiograph of the entire hand and wrist with images in the Atlas.
– The British TW2 (The Tanner and Whitehouse method [Tanner *et al.* 1975]), assigns a score of 20 to the epiphyses in the hand and wrist. The scores are summed and compared to standard Atlas tables.
– TW3 (The Tanner, Healy, Goldstein etc. method [Tanner *et al.* 2001]) is an innovative version of the TW2 process, and is based on the same principles. This method is considered the most accurate in current clinical practice.
– Kapalin (Kapalin, Picko 1964), a method predominantly established on the Czech population, is analogous to

the Greulich and Pyle technique (Riegrová *et al.* 2006). The Kapalin Atlas is of interest because it has not been published, and there is no bone age assessment comparison available between Kapalin values and TW2 and TW3 values.

This study compared bone age data from roentgenograms of children with growth and development disorders. Growth disorders were subdivided into two diagnostic groups; growth retardation, and growth acceleration. The delayed bone age (retardation) group encompasses children with statures that fall below the 3rd percentile for their age, or a growth speed under the 25th percentile. The advanced bone age (acceleration) group is comprised of children with statures above the 95th percentile, or with growth speeds greater than the 75th percentile. Growth disorders in children are manifold, and include chronic disorders (diseases of the kidneys, heart, gastrointestinal tract, lungs and bones), endocrine diseases (diseases involving hormones), primary disorders in growth and development (achondroplasia, hypochondroplasia), aberrations of the chromosomes and genetic syndromes (Turner syndrome etc.).

Some children may just grow more slowly than others based on heredity. 80% of all children can be identified as being of idiopatic short stature or idiopathic tall stature and can be divided into familial short/tall stature or children with constitutional delay of growth and puberty (Lebl *et al.* 2004). Familial short stature is a condition in which shorter parents tend to have shorter children. This term applies to short children who do not have any symptoms of diseases that affect their growth. Children with familial short stature still have growth spurts and enter puberty at normal ages, but they usually will only reach a height similar to that of their parents. Constitutional growth delay is a condition that describes children who are small for their ages but who are growing at a normal rate; they usually have a delayed bone age. These children do not show any signs or symptoms of diseases that affect growth. They tend to reach puberty later than their peers do, with delay in the onset of sexual development and the pubertal growth spurt. But, because they continue to grow until an older age, they tend to catch up with their peers when they reach adult height.

MATERIALS AND METHODS

In this study, 94 roentgenograms of the hand and distal part of the forearm of children with growth and development disorders were evaluated by the TW3 (RUS) and TW2 (RUS) methods (including CARP, TW20 systems of the TW2 method). The roentgenograms were derived from the Auxology Ambulance of the Endocrinology Institute in Prague. 70% of the all roentgenograms studied were taken from children with growth retardation.

36 roentgenograms were assessed by the Kapalin method in the Radiology Department of Children's Surgery at Thomayer Hospital in Prague-Krč. These rentgoenograms were also assessed by the TW3 and TW2 methods. An initial assessment of the methods used was made on the entire sample of roentgenograms prior to their separation into sub-groups. The agreement between the development trend with the results in all groups of the roentgenograms was monitored. These results were validated following a detailed division of the roentgenograms into gender, age and diagnostic groups. The groups of roentgenograms were further divided into the following diagnostic groups of children with: idiopatic small stature (ISS) and growth hormone (GH) deficiencies, constitutional acceleration in growth and development, constitutional retardation in growth and development, familial small stature and pubertas tarda.

The diagnostic classifications were subdivided into gender and age groups; males of age 4.0 to 10.9, and 11.0 to 17.9, and females of age 4.0 to 8.9, and 9.0 to 16.9. The defining age limits (10.9 for males, and 8.9 for females) were based on the TW3 method (norm of skeletal maturation score) with proven effect of secular trend on the current European population (Tanner *et al.* 2001, Krásničanová 2006).

The roentgenograms were scanned into a computer and evaluated by the skeletal atlas for the TW2 and TW3 methods. Conversion of the roentgenograms to computer files enabled a more accurate, objective assessment of bone age. Minimum and maximum intervals between assessments of each roentgenogram ranged from one week to six months. The roentgenograms were evaluated repeatedly to establish intra- and inter- individual errors in measurement. Basic statistical characterizations and pair tests (program NCSS) were used to compare the methods of bone age assessment. Reliability was determined using a selective correlative coefficient with a significance level of 5% or lower. Statistical significance was established using sign convention.

RESULTS

The main objective of the present study was the evaluation of bone age from roentgenograms, and comparison of the values for bone age using the TW2 (RUS) and TW3 (RUS) methods.

The TW3 (RUS) method gave always estimated lower values than the TW2 method in all groups of the roentgenograms (difference +0.8 years ± 0.5 SD based on the 0.1% alpha-power) (*Table 1*). Based on analysis and validation performed within the divisions of gender, age and diagnostic groups, a statistical difference between the TW2 (RUS) and TW3 (RUS) methods was also confirmed. The TW3 (RUS) method has shown to have an average estimated value of one year less than the TW2 (RUS) method, primarily in the older age category. The biggest differences between the TW2 (RUS) and TW3 (RUS) methods were validated for the children with a constitutional acceleration in growth and development with

a difference of +1.0 year (based on the 0.1% alpha-power). Children with pubertas tarda showed a difference of +1.0 year. The smallest differences between TW2 (RUS) and the TW3 (RUS) methods were found in the group of children with familial short stature, where the average difference was +0.5 years (*Table 2, Figure 1*).

In the second part of the study, values of bone age were compared using the Czech Kapalin method of bone age determination with the TW2 and TW3 foreign standard methods. In assessing all groups of 36 roentgenograms, there was no statistically significant difference between the Kapalin method and the TW20 system (+0.1 year) as one of all methods (systems) (*Table 3*). Also, no statistically significant differences between the Kapalin and TW3 (RUS) methods, or the Kapalin and TW20 system of the TW2 method could be found when assessing the roentgenograms divided by gender, age and diagnostic groups. The smallest differences between the Kapalin and the TW3 methods were found in children with constitutional disorders of growth (advanced or delayed, +0.2 years, n=10); and in children with familial short stature (+0.3 year, n=6). The biggest differences were found for children with idiopatic small growth and GH deficiency, which show a difference of +0.6 years (n=13) based on the 1% alpha-power. This result demonstrates the heterogeneity of the diagnostic unit. The biggest differences between Kapalin and the other methods (systems) surveyed were found following an analysis of the roentgenograms groups with the CARP system with

TABLE 1. Results of the comparison of differences in using methods TW2 (RUS), TW3 (RUS) and CARP, TW20 systems in the group of all roentgenograms (n=94). Legend: *=H_0 reject based on the 5% alpha-power, **=H_0 reject based on the 1% alpha-power, ***=H_0 reject based on the 0.1% alpha-power, n=count, d=average difference, SD=standard deviation.

Methods n=94 rtg	**TW2 (RUS)–TW3 (RUS)**	**TW2 (RUS)–CARP**	**TW2 (RUS)–TW20**	**TW3 (RUS)–CARP**	**TW3 (RUS)–TW20**
d	0.7585106	1.041489	0.5234042	0.2829787	–0.2351064
SD	0.4711938	0.9639089	0.5600356	0.7858477	0.5523571
Statistical significance	0.10%	0.10%	0.10%	0.10%	0.10%

TABLE 2. Comparison of differences in using TW2 (RUS) – TW3 (RUS) methods by subdivided diagnostic groups. Legend: ISS=idiopatic small stature and growth hormone (GH) deficiencies, CoA=constitutional acceleration in growth and development, CoR=constitutional retardation in growth and development, FSS=familial small stature, pubertas tarda), n=count, d=average difference, SD=standard deviation, $\bar{x}$=mean.

Diagnostic groups	**TW2 (RUS)**	**TW3 (RUS)**	**TW2 (RUS)–TW3(RUS)**
	$\bar{x}$±sd (year)	**$\bar{x}$±sd (year)**	**d±sd (year)**
ISS+GH (n=32)	10.9±4.5	10.3±4.0	**0.7±0.5*** **
CoA (n=10)	13.6±2.1	12.6±2.1	**1.0±0.1*** **
CoR (n=14)	11.9±2.5	11.0±2.2	**0.9±0.4*** **
FSS (n=8)	8.5±3.9	8.0±3.4	**0.5±0.5***
Pubertas tarda (n=12)	13.2±1.9	12.2±1.6	**1.0±0.4*** **

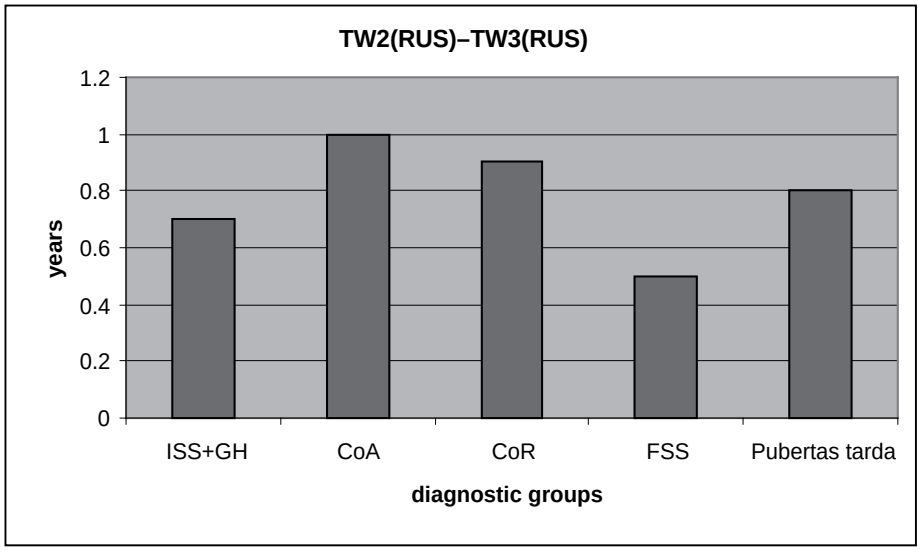

FIGURE 1. Graphic view of the results of differences in using TW2 (RUS) and TW3 (RUS) methods by subdivided diagnostic groups. Legend: ISS= idiopatic small stature and growth hormone (GH) deficiencies, CoA = constitutional acceleration in growth and development, CoR = constitutional retardation in growth and development, FSS = familial small stature, pubertas tarda.

TABLE 3. Statistical significance of differences in using methods by Kapalin, TW3 (RUS), TW2 (RUS), CARP, TW20. Legend: *=H_0 reject based on the 5% alpha-power, **=H_0 reject based on the 1% alpha-power, ***=H_0 reject based on the 0.1% alpha-power, CA=chronological age.

Validity	d	SD
Kapalin – CA	–0.754025**	1.548945
Kapalin – TW3 (RUS)	0.33625**	0.728186
Kapalin – TW2 (RUS)	–0.41625**	0.8881259
Kapalin – CARP	0.80625***	1.153156
Kapalin – TW20	0.11875	0.9118654

TABLE 4. Results of the intra-individual and inter-individual errors in measuring, count of the measured materials.

R=RELIABILITY	TW3 (RUS)	TW2 (RUS)	CARP	TW20
	(n)	(n)	(n)	(n)
Inra-individual error measuring	0.998514	0.995718	0.988133	0.973805
n	91	91	92	30
Inter-individual error measuring	0.998625	x	0.988462	x
n	84	x	82	x

TABLE 5. Null intra-individual differences in using TW3 (RUS), TW2 (RUS) methods and CARP, TW20 systems.

Null difference in %	TW3 (RUS) (n=91)	TW2 (RUS) (n=91)	CARP (n=92)	TW20 (n=30)
0	18	18	34	23
0–0.2	62	36	60	69
0–0.4	79	59	72	87
Krásničanová 2006	65	62	x	x

TABLE 6. Null inter-individual differences in using TW3 (RUS), TW2 (RUS) methods and CARP, TW20 systems.

Null difference in %	TW3 (RUS)	TW2 (RUS)	CARP
0	36	x	33
0–0.2	74	x	60
0–0.4	88	x	79
Krásničanová 2006	44	49	x

TABLE 7. The correlation coefficients of the intra- and inter-individual differences.

Methods		TW3 (RUS)	TW2 (RUS)	CARP	TW20
Differences					
Intra-individual difference	**correlation coefficient**	**0.9926**	**0.9876**	**0.9926**	**0.9875**
	standard error	0.1411	0.202	0.137	0.3372
Inter-individual difference	**correlation coefficient**	**0.9953**	x	**0.9959**	x
	standard error	0.1197	x	0.1137	x

+0.8 years, difference was observed after division of the roentgenograms into diagnostic groups (e.g. +1.3 years in children with idiopatic small growth and GH deficiency, +0.6 years in children with familial short stature).

The reliability of the measuring process for comparing systems of bone age determination was established prior to this evaluation. The results of intra- and inter-individual errors demonstrated high reliability of bone age (0.99) in both cases (*Table 4*). These values also correlated well with results for high percent frequencies of the null inter-individual differences by this method (*Tables 5* and *6*). We found no variation in 74% of the cases using the TW3 (RUS) method for null inter-individual difference, or in 62% cases for null intra-individual difference. Null differences in the range of 0–0.2 years were used in the values of bone age (Tanner *et al.* 2001). Correlation analyses gave high correlation coefficients (0.98–0.99) of the null intra- and inter-individual differences for these methods (*Table 7*).

DISCUSSION

Many authors have compared methods for the assessment of bone age in healthy populations. They have shown that the TW3 (RUS) method estimates values that are one year less than the TW2 method (Tanner *et al.* 2001). Frish *et al.* (1996) assessed these methods using patients with Turner syndrome, growth hormone (GH) deficiencies and familial small stature. Castriota-Scanderbeg *et al.* (1998) researched a different diagnostic group, e.g. patients with

familial small stature, constitutional retardation in growth and puberty, and with Sotos syndrome (Krásničanová 2006). These assessments demonstrate differences between TW2/GP methods, e.g. 0.92 years (Acheson *et al.*1966), 1.25 years (Fry 1968), 0.92 years (Korbáčková 1978), 1.1 years (Frish 1996), 0.82 years (Castriota-Scanderberg 1998).

Krásničanová (2006) acknowledged in her study the differences between the GP/TW2 methods and the congruence between GP/TW3 methods in the Czech Republic. In her study, the average difference between TW2/GP methods was 0.95–1.06 years. Krásničanová compared these methods of bone age assessment using different diagnostic units (constitutional retardation in growth and puberty, and mental anorexia), and weighed them against the use in clinical practice.

In this paper we present the results of our comparison of the values for bone age assessment, using the RUS systems (TW2/TW3 methods), the CARP system (TW2/TW3 methods), and the TW20 system (TW2 method). In total, 94 roentgenograms were evaluated in the study, divided into diagnostic groups, and sub-groups for gender and age resulting in a significant decrease in occurrence. In many cases we can interpret results only casuistically. We could conclude that an average difference of one year exists between TW2/TW3 methods in a representative group of all roentgenograms from children with growth and development disorders. This was also validated when the roentgenograms were divided by sub-groups, then by diagnostic groups, and by gender and age. In all cases the TW2 (RUS) method had higher values than the CARP and the TW20 systems in the different diagnostic units. The TW3 (RUS) method had different results when compared the CARP and the TW20 systems.

Krásničanová's study of children with constitutional retardation in growth and development showed a statistically significant difference of +0.9 years ±0.4 SD (in males, $p \leq 0{,}1$), and +0.9 years ±0.2 SD (in females, $p \leq 0{,}5$), for children in the older age category, between TW2 (RUS)–TW3 (RUS) methods. Those in the lower age category had minimal difference (–0.1 years for males, and +0.2 years for females). This study compared the same diagnostic unit and verified the same differences: TW2 (RUS) – TW3 (RUS) for children in the older age category +0.9 years ±0.4 SD (in males, $p \leq 0.01$) and +0.9 years ±0.1 SD (in females, $p \leq 0.5$). Children in the lower age category were not represented.

Comparison of the intra- and inter-individual differences for the TW3 methods with Krásničanová's results showed the same outcome (Duchajová's result: 62.0%, Krásničanová's result: 65.0%) in intra-individual difference, but different results in inter-individual difference (Duchajová's result: 74.0%, Krásničanová's result: 44.0%) (*Tables 5* and *6*). (This observation occurred twice as often as in Krásničanová.)

The second part of the study deals with the comparison of the TW2/TW3 (RUS) methods with the Kapalin method, which has not been done before in the Czech Republic. In the Radiology Department of Children's Surgery at Thomayer Hospital in Prague-Krč, the Kapalin method has been used for the assessment of bone age. We decided to verify the aplicability of the method in clinical practice.

It was considered that the TW20 system of the TW2 method would most closely aproximate to the Kapalin method for the valuation of the compartments the RUS or the CARP as complex. Both the TW2 and Kapalin methods originated at the same time (70 years ago), and the system TW20 is methodically the most similar to Kapalin [unlike compartment RUS (TW2, TW3 methods), or the CARP system]. This hypothesis was acknowledged with a pair of statistical tests in the group of all roentgenograms (n=36) and in the subgroup divided into diagnostic units. The average difference between TW20 system and Kapalin was +0.1 years for all roentgenograms – this difference was assessed as being statistically insignificant. After the division of the group into diagnostic units, it was verified that there was no statistical difference (+0.2 to +0.4 years) between Kapalin and TW3 (RUS), and between Kapalin and the TW20 system.

Both the GP and Kapalin methods are qualitative, and we can compare Kapalin with the results of GP method. Krásničanová's results (Krásničanová 2006) show no statistical difference (+0.48 years) between the TW3 and GP methods in males younger than 11 years of age. Also no significant statistical difference (+0.02 years) was found for males older than 11 years of age, for females younger than 9 years (+0.15 years), and for females older than 11 years (–0.08 years) (with constitutional retardation in growth and puberty. Our results of the differences between TW3 and Kapalin are similar to Krásničanová's results (TW3/GP), for children with constitutional retardation in growth and puberty (Krásničanová 2006).

In conclusion we can say that values of the TW3 (RUS) and the GP methods are in good agreement, and that the GP, Kapalin and the TW20 systems (of the TW2 method), have similar methodologies. We can say that Kapalin can be used confidently in current clinical practice (from the results of divisions into diagnostic units). However it should be noted that the examined group contained only 36 roentgenograms of the hand and distal part of the forearm in children with growth and development disorders. After division into diagnostic units and sub-groups as outlined, there was a marked decrease in occurrence. We would suggest that, in order to gain greater confidence in these results, a further study using a larger group of roentgenograms be carried out.

REFERENCES

ACHESON R., VICINUS J. H., FOWLER G. B., 1966: Studies in the reliability of assessing skeletal maturity from X-rays. III. Greulich-Pyle Atlas and Tanner-Whitehouse methods contrasted. *Hum. Biol.* 38: 204–218.

CASTRIOTA-SCANDERBEG A., EMBERTI-GIALLORETI L, FRARACCI L., 1998: Skeletal age assessment in children

and young adults: comparison between a newly developed sonographic method and conventional method. *Skeletal Radiology* 27: 271–277.

DUCHAJOVÁ L., 2006: Stanovení biologického věku u dětí s poruchou růstu a vývoje – validita metod kostního věku a růstového věku. Unpublished M.A. Thesis. Charles University, Praha. 122 pp.

FRISCH H., RIEDL S., WALDHOR T., 1996: Computer-aided estimation of skeletal age and comparison with bone age evaluations by the method of Greulich-Pyle and Tanner-Whitehouse. *Pediatric Radiology* 26: 226–231.

FRY E. I., 1968: Assessing skeletal maturity: Comparison of the atlas and individual bone techniques. *Nature* 23: 496–497.

GREULICH W. W., PYLE S. J., 1959: *Radiographic Atlas of Skeletal Development of the Hand and Wrist.* Stanford University Press, California. 256 pp.

KAPALIN V., PICKO V., 1964: Kostní věk u dětí školního věku a jeho závislost. *Československá hygiena* 9, 6: 342–356.

KORBAČKOVÁ A., 1978: Porovnanie metód pre určenie kostneho veku na základe osifikácie skeletu ruky. Unpublished M.A. Thesis. Jan Amos Komenský University, Bratislava. 89 pp.

KRÁSNIČANOVÁ H., KUCHYŇKOVÁ I., 2002: Nová metoda hodnocení kostního věku TW3 a první výsledky jejího použití u nás. *Československá pediatrie* 57: 62–65.

KRÁSNIČANOVÁ H., 2006: *Exaktní hodnocení skeletální zralosti a určování kostního věku pro potřeby klinické pediatrie***.** Iga MZ ČR, Praha. 69 pp.

LEBL J., KRÁSNIČANOVÁ H., 1996: *Růst dětí a jeho poruchy*. Galén, Praha. 160 pp.

LEBL J., ZAPLETALOVÁ J., KOLOUŠKOVÁ S., 2004: *Dětská endokrinologie*. *Trendy soudobé pediatrie.* Sv. 3. Galén, Praha. 479 pp.

RIEGEROVÁ J. , PŘIDALOVÁ M., ULBRICHOVÁ M., 2006: *Aplikace fyzické antropologie v tělesné výchově a sportu (příručka funkční antropologie*), HANEX, Olomouc. 262 pp.

TANNER J. M.,WHITEHOUSE R. H., MARSHALL, W. A., HEALY, M. J. R., GOLDSTEIN, H., 1975: *Assessment of skeletal maturity and prediction of adult height (TW2 method)*. Academic Press, London. 99 pp.

TANNER J. M., HEALY M. J. R., GOLDSTEIN H., CAMERON N., 2001: *Assessment of skeletal maturity and prediction of adult height (TW3 method)*. W. B. Saunders, Harcourt Publishers Limited, London. 110 pp.

Lenka Duchajová
Petr Sedlak
Department of Anthropology
and Human Genetics
Faculty of Science
Charles University
Viniční 7
128 44 Prague 2, Czech Republic
E-mail: l.duchajova@seznam.cz
E-mail: petr.sedlak-uk-prf@seznam.cz

Marcela Dvořáková
Endocrinology Institute in Prague
Národní třída 8
116 94 Prague 1, Czech Republic
E-mail: mdvorakova@endo.cz