Adult Height Prediction Models

Hans Henrik Thodberg[1], Anders Juul[2], Jens Lomholt[3], David D. Martin[4],

Oskar G. Jenni[5], Jon Caflisch[5], Michael B Ranke[4], Luciano Molinari[5], Sven Kreiborg[3]

[1]Visiana, Denmark
[2]Rigshospitalet, Department of Growth and Reproduction, Denmark
[3]Copenhagen University, Denmark
[4]University Children's Hospital Tübingen, Germany
[5]Child Development Centre, University Children's Hospital Zürich, Switzerland

The Handbook of Growth and Growth Monitoring in Health and Disease
Springer (c) Nov 2009

Corresponding author: Dr. H.H. Thodberg, Visiana
Søllerødvej 57 C, DK-2840 Holte, Denmark
Tel.: +45 2144 7087, E-mail: thodberg@visiana.com. Web: www.BoneXpert.com

Abstract

We review seven methods for adult height prediction (AHP) based on bone age, ranging from the Bayley-Pinneau method, published in 1952, to the BoneXpert method, published in 2009. These models are based on four different methods for bone age assessment including Greulich Pyle, Tanner Whitehouse, Fels, and the automated BoneXpert method.

The aim of this chapter is to convey an understanding of the various parameters which contribute to AHP and how to best incorporate them into the AHP methods. The starting point is the Bayley-Pinneau method which predicts the fraction of adult height achieved from the bone age. Children with advanced bone age (early maturers) tend to have a stronger growth spurt, and late maturers have a weaker growth spurt. Accordingly, Bayley and Pinneau provided special AHP tables for early, average and late maturers.

The other five AHP methods reviewed are the three variants of the Tanner Whitehouse method, TW Mark I (1975), TW Mark II (1983) and TW3 (2001), and the RWT methods from 1975 and 1993. They all model the expected adult height of children at each age using a linear model of height and bone age, and for the RWT models, also by using terms with midparental height and body weight. The main short-coming of these models is that the linear bone age dependence is unable to describe children with constitutional delay of growth and puberty or precocious puberty.

The recently developed automated BoneXpert method improves the Bayley-Pinneau method by modelling the growth potential (the fraction of adult height left to grow) as a nonlinear function of two variables, bone age and bone age delay. The BoneXpert AHP method was based on the original images from the First Zürich Longitudinal Study, and was subsequently validated on the more recent Third Zürich Longitudinal study of 198 Swiss children. An additional validation study on 164 Danish children is also presented. The main advantage of the BoneXpert method is that it is based on an automated bone age which removes rater variability.

Abbreviations

AHP:	Adult height prediction	1ZLS:	First Zürich Longitudinal Study
BA:	Bone age	3ZLS:	Third Zürich Longitudinal Study
CA:	Chronological age	GH:	Growth Hormone
GP:	Greulich-Pyle	GHD:	Growth Hormone Deficiency
TW:	Tanner-Whitehouse	RMSE:	Root Mean Square Error
RUS:	Radius, Ulna and Short bones	$\boldsymbol{h}$:	Current height
BP:	Bayley-Pinneau	$\boldsymbol{H}$:	Adult height
RWT:	Roche-Wainer-Thissen	$\boldsymbol{gp}$:	Growth potential = $(\boldsymbol{H} - \boldsymbol{h})/\boldsymbol{H}$
BX:	BoneXpert	SDS:	Standard Deviation Scores
BMI:	Body Mass Index		

1 Introduction

It is not unusual for families to speculate about the expected adult height of their children based on the child's current age and height and the parents' heights. However, as Tanner expressed in 1975, knowledge of the tempo of growth or maturation, i.e. the bone age, is important for making such a prediction (Tanner et al. 1975a):

Children differ greatly in the rate at which they pass through the various phases of growth; some have a rapid tempo of growth and attain adult status at a relatively early age; others have a slow tempo and finishes growing relatively late. A child's height at any age reflects both how tall he will ultimately become and how advanced he is towards that goal.

This review focuses on the seven most important methods for adult height prediction (AHP) based on bone age (BA). A new generation of methods for BA assessment has appeared approximately every twenty years, and each new BA method has led to a new generation of AHP methods. These AHP methods are listed in Table 1, according to the bone age methods on which they are based.

Table 1: The four generations of bone age methods and the seven methods for adult height prediction.

Bone age Generations		Bone age methods	Adult height prediction methods	Bone age used in AHP
1	1946-59	Todd / Greulich-Pyle	Bayley-Pinneau (1946/52)	Todd/GP
			RWT (1975)	GP bone-specific
2	1962-83	Tanner Whitehouse, TW1, TW2, TW3	TW Mark I (1975)	TW2
			TW Mark II (1983)	TW2
			TW3 (2001)	SMS
3	1987-93	Fels	RWT/Khamis (1993)	Fels
4	2008-	BoneXpert	BoneXpert (2009)	BoneXpert

1.1 The first generation – Bayley-Pinneau

The first AHP model was formulated by Bayley (Bayley 1946) and was based on Todd's method for the determination of BA. The Greulich-Pyle (GP) BA method is a refinement of Todd's atlas method, and both are based on data from the Brush foundation study of children from Ohio. The GP atlas (Greulich and Pyle 1959) contains plates corresponding to a series of selected ages. Each plate is from a child having the median maturity at that age.

Together with Pinneau, Bayley revised her AHP method in 1952 using the GP atlas, thus completing the famous Bayley-Pinneau (BP) method (Bayley and Pinneau 1952) which was based on 192 Berkeley children who were followed longitudinally, and validated on another 46 children from Berkeley. The second and final edition of the GP atlas appeared in 1959 and did not require an update of the BP method, since the 1952 article on the BP method appears as an appendix to the 1959 edition of the GP atlas. The GP method is by far the most common BA rating method used today, and the BP method is still widely used all over the world.

1.2 The second generation – Tanner

In Europe, an alternative to the American GP/BP system was developed over four decades by Tanner and co-workers. The first description of the Tanner-Whitehouse (TW) skeletal maturity assessment system appeared in 1962 (Tanner et al. 1962). It was based on twenty bones. The operator assigns maturity stages A, B, .., I to each bone, and the TW system assigns a score to each stage, from which a *summed maturity score* (SMS) is formed ranging from 0 (immature) to 1000 (adult). The TW system then translates the SMS into a bone age based on data from a selected population. The bone age is defined as the age at which the observed SMS is at the 50th percentile. A more complete account of the TW system appeared in book-form in 1975, (Tanner et al. 1975b) and it included a revision of the TW bone age system, called TW2. In addition to the 20-bone system, a 13 bone system called RUS (radius, ulna and short bone) was defined, and the first TW AHP model, called TW Mark I, was based on TW2-RUS bone age.

The second edition of the book (Tanner et al. 1983a) introduced a revised AHP model called TW Mark II, still based on TW2 bone age. This model was based on normal children, short statured children and tall statured girls (Tanner et al. 1983b).

Finally, the third edition (Tanner et al. 2001) presented a new bone age version called TW3 and a new AHP, also called TW3, which predict adult height from SMS rather than from bone age. The model was based on 226 children from the First Zürich Longitudinal Study (1ZLS), who were born between 1954 and 1956 (Prader et al. 1989). These images had been BA-rated many years earlier by Prader's group.

1.3 The third generation – Roche

Roche and collaborators developed the third BA method, the Fels method, in 1988 (Roche et al. 1988). This method was based on the Fels longitudinal study which has recruited, on average, 18 children per year since 1933. This method is similar to the TW method, but involves more bones, more maturity features, and more advanced mathematics.

The Fels BA method came late in Roche's career. Previously, he had based his work on a bone-specific version of the GP bone age, where each bone is assigned a bone age, rather than the usual fast approach using an overall intuitive match with the atlas, as practiced by radiologists. Based on this bone-specific GP BA, Roche developed an AHP method called the RWT model (Roche-Wainer-Thissen), in 1975, based on approximately 200 Fels children (Roche et al. 1975). In developing this model, Roche at first tried to use the bone ages of the hand, foot and knee, but he settled on using simply the median GP bone age of the hand. This is well-approximated by the fast, atlas-based GP rating method (based on an overall intuitive match), and the RWT method has been used in this way by many investigators (see Section 1.5).

In 1993, a modified version of the RWT method was presented (Khamis and Guo 1993). It featured several improvements including the use of the Fels BA method for its basis, as well as

the fact that its mathematical underpinnings were more refined. In addition, the number of Fels study children used for this method was increased to a staggering 433.

The Fels method for BA assessment and the associated AHP came rather late and these methods have not been as popular. This is primarily because the Fels BA method is the most laborious of all BA methods and interest in improved manual ratings appears to have declined, perhaps due to hopes that a computerized BA method was not far away.

1.4 The fourth generation – BoneXpert

Despite the early recognition by many, in particular Tanner (Tanner 1989), that BA assessment was a task well suited for computerization, it took more than twenty years to realize this vision. The complexity in the development of such a method was underestimated, and only by using very advanced mathematical methods and taking advantage of the 10,000-fold increase in computer power over the last 20 years, was it possible to develop such a system, the BoneXpert (BX) method (Thodberg et al. 2009b). The BX method, which constitutes the fourth generation of bone age methods, expresses the bone age based on the GP scale, i.e., it agrees, *on average,* with the manual GP BA, but the standard deviation from the manual rating is 0.5 to 0.7 yrs (depending on the reliability of the manual rater). It is, therefore, considered a completely new kind of bone age rating.

The BX system determines the GP bone age as the average GP bone age of the 13 RUS bones, i.e., it excludes the carpal bones. The computer extracts visual information in a manner different from the human eye, so it was important to conduct extensive validation studies on both healthy children, and children with the common diagnoses of short stature (Turner syndrome, GHD, etc.). Such validation studies have shown that the system is robust and consistent (Martin et al. 2009a;Thodberg 2009;van Rijn et al. 2009).

Finally, the BX AHP model, based on BX bone age, was presented in 2009 (Thodberg et al. 2009a). Remarkably, this model was based on the same study, the 1ZLS, as the TW3 AHP method.

1.5 Previous comparative studies

These competing methods, which were used for the same prediction, led to a number of comparative studies, including the five which we will now discuss.

In 1978, the three existing methods were validated on both normal children from the 1ZLS and on children with Turner syndrome (Zachmann et al. 1978), however, the Turner syndrome children were outside the scope of the models.

It was not until the 1990's that most of the comparative studies were performed, focusing primarily on untreated tall stature (Binder et al. 1997;Brämswig et al. 1990;de Waal et al. 1996;Joss et al. 1992). These studies reported the average (signed) error, i.e., the bias, and the SD of the prediction errors. It is, however, difficult to interpret these results because there were three potential contributing errors including (a) that the bone age rating could be poor or biased; (b) that the population could be different than the one used to estimate the AHP model; and (c) that the model itself could be inadequate for the group of children being studied, e.g., tall stature.

These studies were not able to separate out such effects, and, therefore, did not lead to any progress in our understanding of AHP but merely assessed the errors and biases to be expected under various conditions.

As a result, the clinical community still handles AHP in a manner that is somewhat inefficient and ambiguous. Often, the X-rays are rated according to both the GP and TW method and the predictions of the BP and TW Mark II methods are made and some conclusions are drawn.

1.6 Chapter outline

The seven AHP models all use current height, ***h***, age, BA and gender to predict adult height, ***H***. They differ mainly in the following three aspects:

- The type of BA, i.e., the weight assigned to different bones, and the reliability of the BA rating.
- The population used for the estimation of the model and for its validation, e.g., the range of variation in height and BA delay.
- The mathematical model, i.e., whether it is a linear or non-linear model, and how other information like parental height is incorporated.

These aspects will be addressed, in detail, in this review, which will focus on providing a fundamental understanding of AHP, including the predictors of ***H***, and the interpretation of their contribution. This insight was used to motivate the design of the BX method. We build up an understanding of AHP step by step:

- Section 2 considers the simplest possible AHP models in the spirit of Bayley and Pinneau and describes the improvements made in the BX method.
- Section 3 demonstrates how knowledge of the height distribution of the population is included in a natural way in the BX AHP method, and how one can also include, as an option, the heights of the parents. The performance and validation of the BX method are also reported.
- Section 4 discusses the TW and RWT methods and compares them to the BX method.
- Section 5 explains how information on menarche, BMI and Tanner stages can contribute to AHP.
- Section 6 gives guidance to the practical application of the AHP methods and describes their limitations.

Sections 4 and 5 can be skipped during the first reading.

1.7 Data used in this work

Many aspects of AHP will be illustrated with the data used to develop and validate the BX method. These data, which are summarized in Table 2, are described in this subsection.

The results on the Zürich studies have been reported previously (Thodberg et al. 2009a), while the results on the Björk study are new.

The 1ZLS of growth (Prader et al. 1989) included healthy children which were followed from birth to adulthood with annual height and weight measurements and hand X-rays, which have survived from approximately 5 years and up. In

this study, only left hand X-rays were used. Each were rated manually at the time of the study according to both the GP and TW methods, and by digitizing the films, a BX BA rating was also produced. A bone health index was also derived from the cortical thickness of the metacarpal bones. The adult height of the children was defined as the height when growth was less than 0.5 cm during the last two years. A skinfold measure was formed as the average of the four skinfold measurements (biceps, triceps, sub-scapular, and supra-iliac), and Tanner stages of the secondary sexual development were recorded as two separate scores, one for pubic hair and one for genitals or breast development.

The mean (SD) of the parents' heights was 173.2 (6.8) cm for the fathers and 162.0 (6.2) cm for the mothers. The mean adult heights of the children were 178.2 (7.0) cm for the boys and 165.0 (5.9) cm for the girls (also listed in Table 2), so the secular trend was 5 cm for males and 3 cm for females.

The Third Zürich Longitudinal Study (Thodberg et al. 2009a) of growth and development included children having one parent in the 1ZLS. This study, which is still ongoing, follows the children with annual visits until the age of 18. X-rays are taken at selected ages with most of the children receiving X-rays at ages 7, 10, 12, 14, 16 and 18 yrs. The last height measurement is at age 18, and the adult height is defined as the height at 18 yrs plus a constant which is 0.9 cm for boys and 0.3 cm for girls. This is done to ensure that the adult height is as compatible as possible with the definition used in the 1ZLS. These constants were derived from the 1ZLS data.

The Björk growth study (Björk 1968) enrolled healthy Danish children for orthodontic treatment at the Royal Dental College in Copenhagen in a study designed to provide information on craniofacial growth in relation to somatic growth. From the annual data collection we used the height, the weight and the hand X-ray from which the BX bone age is derived.

The first available X-ray was, on average, at age 7, and the last annual visit was typically at 21, and for many subjects there were also later visits, at 25 yrs or 30 yrs.

In order to be similar to the definition used in the 1ZLS, the adult height was defined at the first visit after age 19 for boys (and age 18 for girls). The median age at adult height was 19.6 yrs for boys and 18.6 yrs for girls.

Visits at ages above 20 yrs and BA less than 6.5 yrs were excluded from the analysis. This resulted in a total of 1286 observations from 83 boys and 81 girls. Weight data were available for all females and for 66% of the males.

Table 2: The studies used to develop and validate the BoneXpert AHP method.

Study	Birth years	N	Adult height (SD) (cm)		Bone age delay (yrs)	
			Boys	Girls	Boys	Girls
1ZLS	1954-56	231	178.2 (7.0)	165.0 (5.9)	0.1	0.2
3ZLS	1973-91	198	178.5 (6.4)	165.6 (5.2)	0.4	0.1
Björk	1939-64	164	178.8 (7.2)	166.1 (6.4)	0.5	0.7

2 The Bayley-Pinneau model

Given the concept of bone age as a measure of how far the bones have progressed from immaturity to maturity, it is natural to attempt a "*naïve AHP model*" with a simple, direct relation between bone age and the *fraction* of adult height achieved. The relation is expected to be non-linear, because we know that the growth velocity varies with maturity.

We define the *growth potential* of a child with current height, $\boldsymbol{h}$, and adult height, $\boldsymbol{H}$, as

$$\boldsymbol{gp} = (\boldsymbol{H} - \boldsymbol{h}) \,/\, \boldsymbol{H}$$

The naïve AHP model then states that we can predict $\boldsymbol{gp}$ from BA alone through some non-linear function $\boldsymbol{gp}_{\mathrm{pred}}$(BA). Since BA is gender-specific, there must be two versions of this function, one for each gender. This is the starting point of the Bayley-Pinneau (BP) method, which models $\boldsymbol{gp}$ primarily as a function of BA. (Bayley-Pinneau, in fact, modelled the percent of mature height which is 100 (1 – $\boldsymbol{gp}$), i.e. the "mirror image" of $\boldsymbol{gp}$). But Bayley-Pinneau soon found that $\boldsymbol{gp}$ also depends on the difference between CA and BA, i.e. *the BA delay*. They, therefore, divided the children into three groups according to the BA delay: normal BA (|CA–BA| < 1 yr), advanced BA, and delayed BA. This division is an oversimplification, because these groups of children form a continuum, and a child can switch from one group to another over time (which causes some counterintuitive effects when applying the BP method).

Figure 1 illustrates the naïve AHP model. It shows BP's $\boldsymbol{gp}$ function for children with normal BA, and it also shows the $\boldsymbol{gp}$ of the BX model for children with CA = BA for comparison. Indeed, the relation is non-linear; in particular, the male curve steepens at the growth spurt.

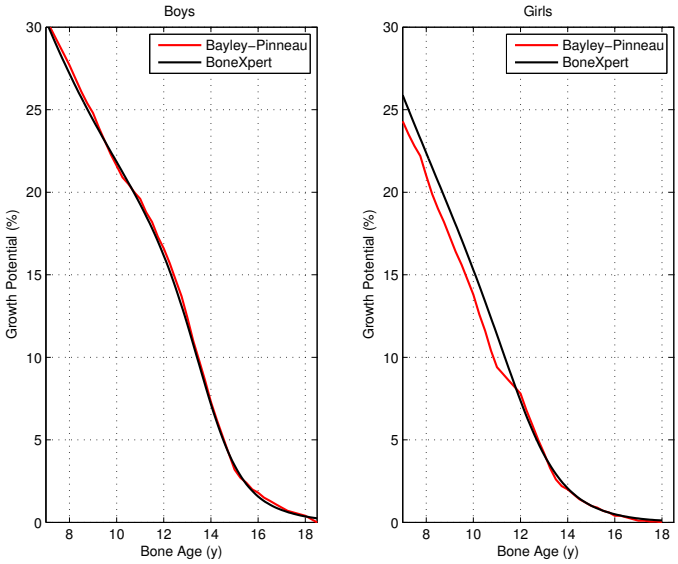

Figure 1: Illustration of the "naïve" AHP model, using a direct relationship between BA and $\boldsymbol{gp}$. The Bayley-Pinneau curves are the BP model for |(BA–CA|<1 yr, and the BoneXpert curves are the BX model for CA = BA. All figures in this chapter, except Figures 4, 5, 11, 15, 16 and 17, are adapted from (Thodberg et al. 2009a).

The BX model can be considered an elaboration of the BP model. Here $\boldsymbol{gp}$ is modelled as a nonlinear function of two variables, the BA and the BA delay, written as $\boldsymbol{gp}_{\mathrm{pred}}$(BA, CA–BA). The function is implemented as a neural network with 21 adjustable parameters for the boys and 9 for the girls, and is visualized in Figure 2.

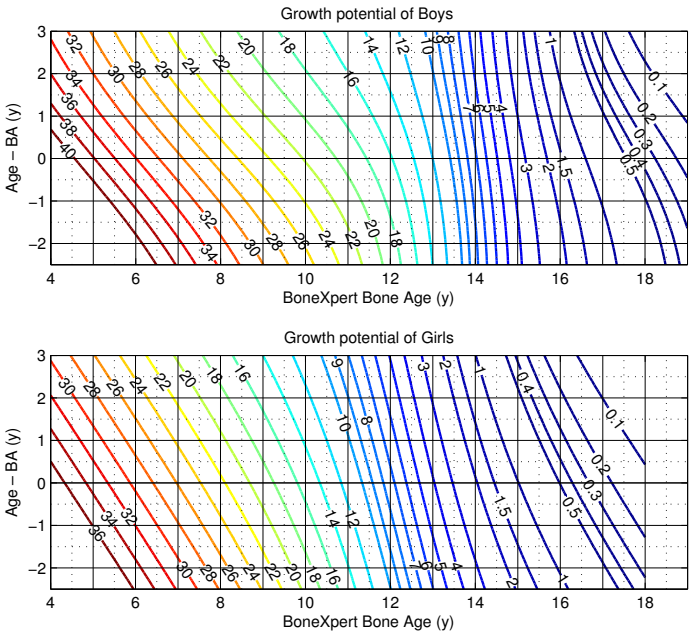

Figure 2: BoneXpert's model of the growth potential for boys (top) and girls (bottom). From here one can read off the ***gp*** for any values of BA and BA delay.

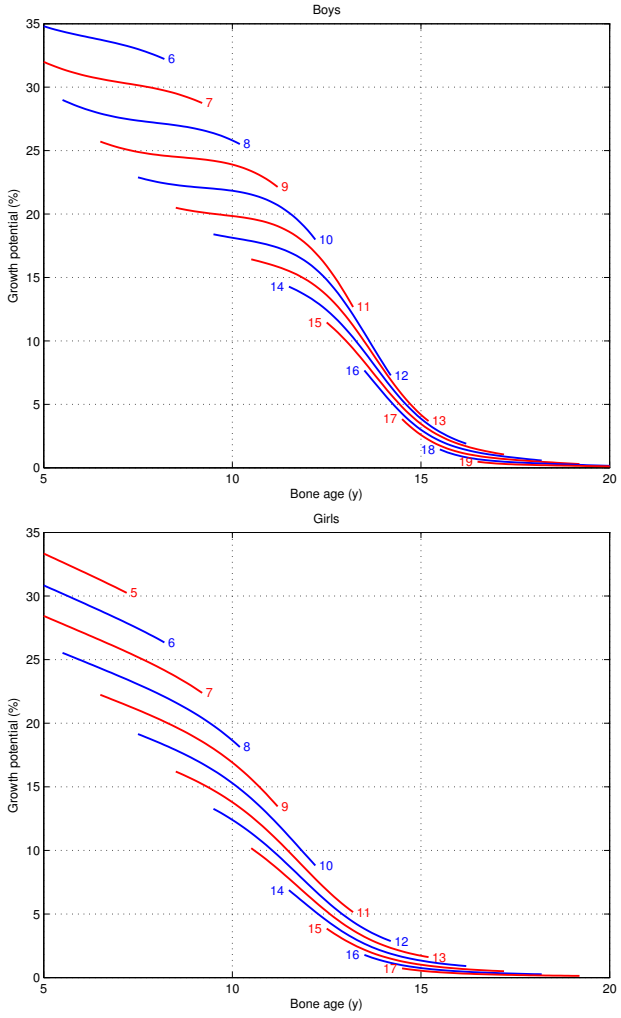

Figure 3: BoneXpert's model of the growth potential as a function of bone age for boys (top) and girls (bottom) at fixed ages, indicated in years next to each curve.

To demonstrate the reasons why the naïve AHP model fails to describe nature, we start by considering Figure 3 which shows the BX ***gp*** model as a series of curves, one for each CA. If CA did not matter, the curves for different ages would coincide, and indeed they tend to do so at the upper end of the BA scale. At lower bone ages, the curves are increasingly spaced, more so for the boys, where for BA ≤ 11 yrs, a one-year change in age induces a larger change in ***gp*** than a one-year change in BA.

More specifically, consider three boys of the same BA = 10, and CA 9, 10 and 11, respectively. Then, the boy at 9 yr has larger ***gp***, and the boy at 11 smaller ***gp***. This is really surprising. When first introduced to the concept of bone age, one cannot help being disappointed because bone age alone is not a good predictor of ***gp***. To understand why, we will formulate the following "unfaithful BA" hypothesis, which may explain the failure of BA to predict ***gp***:

Stature is the length of the axial skeleton, whereas BA is a measure of the morphology of the hand bones. The hypothesis states the hand BA is an unfaithful estimate of the "true" BA of the axial skeleton. Since CA can be viewed as an alternative estimate of "axial BA", a better estimate of ***gp*** *is obtained by "drawing" BA towards CA. This is what we see in the* ***gp*** *model: it appears that when we observe a BA different from CA, we cannot "believe" the BA 100% – we must draw the BA value towards the CA.*

We can test this hypothesis, because if it is true, the pubertal growth spurt in stature must occur at an earlier BA for the BA-delayed children. To perform this test we turn to the 1ZLS data and divide the subjects into tertiles: advanced BA, normal BA, and delayed BA, based on the BA at age 9 yrs for boys and 8 yrs for girls. We then *reparameterize the history of stature for each child to become a function of BA*. Organizing the data in this way is what the naïve AHP model suggests; BA is the clock of the physiological development of the child, so using this timescale, the children should look more similar compared to using CA. We then study *the growth velocity with respect to BA*, i.e. the increment in stature per year of BA, as shown in Figure 4. If BA is the natural clock of the skeleton, we should observe the same growth per year in BA for all three groups.

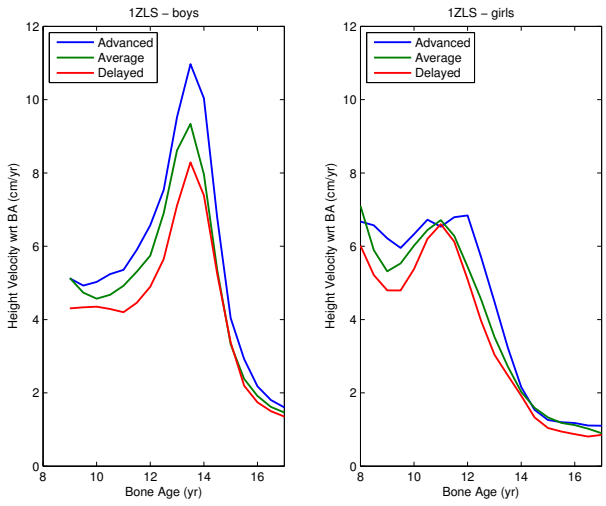

Figure 4: The height increase per BX BA of children in the 1ZLS. The children are divided into tertiles of advanced, normal and delayed bone age according to the BA at age 9 for boys and age 8 for girls.

Figure 4 demonstrates that the growth spurt does occur at the same BA for the three tertiles, namely at 13.5 yrs for boys and

11.0 yrs for girls. Hand BA *does* describe the timing of the growth of the axial skeleton accurately. *This refutes the "unfaithful BA" hypothesis*.

But, Figure 4 also shows that the growth curves are *not* the same for the three tertiles. The strength of the spurt is greater in early maturers who grow more per BA than late maturers. This is the reason for the need to include the BA delay in the prediction of ***gp***. Children do not have the same growth curve, but BA delay reveals the strength of the growth spurt.

For the advanced boys, this effect is very clear in Figure 4; the growth curve shifts upwards. In contrast, the advanced girls do not have a strong peak in the growth spurt, but rather a prolonged spurt, so that the integral under the curve is considerably larger than for the two other tertiles.

Interestingly, Bayley and Pinneau failed to reach this insight in 1952. Instead they speculated more in the direction of the unfaithful BA hypothesis, which they formulated as follows (page 247 in the GP atlas):

"Those children selected as most retarded (or advanced) in one area are likely to be somewhat nearer the average in the other areas. Therefore a fair proportion of the children selected as skeletally deviant can be expected to have a general physical maturity age which is nearer the norm than their skeletal age".

Of course, Tanner knew better, and he has explained the effect very clearly in several contexts.

What we have demonstrated is that there is a considerable variability in the strength of the pubertal spurt, and the challenge of AHP is to predict the strength of the spurt, and the BA delay contributes to that.

To summarize, we have discussed the $\boldsymbol{gp}_{\text{pred}}$ (BA, CA – BA) function, originally estimated crudely by BP, and shown how it is estimated more accurately as a true function of two variables in the BX model. A second advantage of the BX method is that the manual BA is replaced by the automated values, and a third advantage is that ***gp*** is estimated based on the more modern population of the 1ZLS.

To conclude, given ***h***, BA, and CA of a child, the BP-like prediction is

$$\boldsymbol{H}_{\text{raw}} = \boldsymbol{h} \;/\; (1 - \boldsymbol{gp}_{\text{pred}})$$

It is called "raw" because the BX method refines it, as explained in the next section.

3 Including population and parental height

3.1 Combining beliefs about the adult height

The next improvement that the BX method makes to the BP method is to correct a bias that occurs as a result of modelling in terms of the growth potential. The BP approach computes $\boldsymbol{gp}_{\text{pred}}$ from CA and BA, which means that it effectively computes ***h*** from ***H***, rather than ***H*** from ***h***. This would not matter much were it not for the large uncertainties in the relation between ***h*** and ***H***.

To make progress we need to resort to Bayesian inference, a branch of statistics that allows us to model our *belief* about the child's ***H*** as a *probability distribution*. We distinguish between the conditional probabilities $p(\boldsymbol{h}|\boldsymbol{H})$ and $p(\boldsymbol{H}|\boldsymbol{h})$. It is a common occurrence in mathematical modelling that it is easier to model one of these, while it is the other that is actually needed, but, fortunately, the two are related by Bayes' theorem.

We want to compute our belief about ***H*** given ***h*** (and BA and CA), and we denote this $p(\boldsymbol{H}|\boldsymbol{h})$. By Bayes' theorem this is given by

$$p(\boldsymbol{H}|h) \sim p(\boldsymbol{H})\, p(\boldsymbol{h}|\boldsymbol{H})$$

Here $p(\boldsymbol{H})$ is our *a priori* belief about the adult height (the ~ sign means "proportional to" since we have not bothered to normalize the right hand side). We consider these probabilities as functions of the unknown adult height, ***H,*** which we want to predict. The $p(\boldsymbol{h}|H)$ is modelled as a gaussian distribution with the center, $\boldsymbol{H}_{\text{raw}}$, and an SD determined from our estimate of the model from the 1ZLS. The $p(\boldsymbol{H})$ describes what we know about ***H*** prior to observing ***h***, and we model this as a gaussian distribution with its center on the population mean, $\boldsymbol{H}_{\text{pop}}$. In order to arrive at $p(\boldsymbol{H}|\boldsymbol{h})$, we need to multiply the two gaussians, an operation illustrated in Figure 5. The wide gaussian represents the belief derived from the population, here assumed to have SD = 6 cm, and the narrow gaussian is the probability distribution from the BP-like prediction, here assumed to have an SD = 3 cm. These two gaussians represent the two pieces of knowledge that we want to fuse. To do that, we define the (Bayesian) precisions of these beliefs as $1/\text{SD}^2$, so the precisions have the ratio of 1:4. The product of two gaussians is a gaussian whose center, $\boldsymbol{H}_{\text{pred}}$, can be computed as the average of the component centers, weighted with their precisions, i.e.,

$$\boldsymbol{H}_{\text{pred}} = 0.8\, \boldsymbol{H}_{\text{raw}} + 0.2\, \boldsymbol{H}_{\text{pop}}$$

The precision of the result is the sum of the component precisions; the resulting gaussian has SD = 2.7 cm.

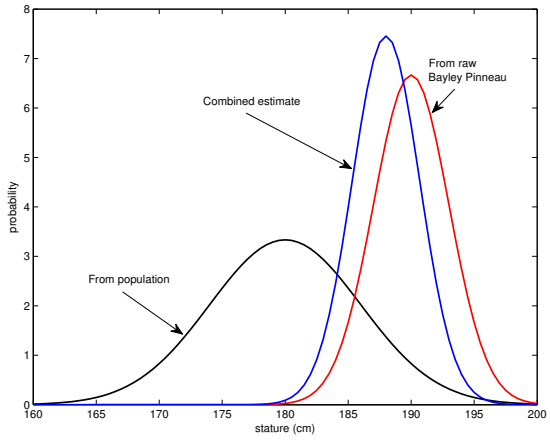

Figure 5: Illustration of the combination of the raw Bayley-Pinneau prediction with information on the general population.

3.2 Drawing towards the population mean

In Figure 5 the raw prediction was drawn towards the population mean with a certain weight which depends on the relative precisions of the two sources of information. Figure 6 shows the error of the raw BP-type prediction as a function of BA; it decreases rapidly after puberty. The uncertainty of the population-based belief is, however, the same for all ages, so the weight of the population mean, shown in the lower plot of Figure 6, drops quickly to zero after puberty.

If we did not use the correction, i.e., if we just use the raw BP expression, there would be a tendency to overpredict tall

children and underpredict short children. Figure 7 illustrates this effect by plotting the error of the raw prediction, $\boldsymbol{H} - \boldsymbol{H}_{\text{raw}}$, as a function of $\boldsymbol{H}_{\text{raw}}$, in the left plot. By fitting a regression line to the data, we can quantify the effect and see that for every cm that $\boldsymbol{H}_{\text{raw}}$ is below the mean ($\boldsymbol{H}_{\text{pop}}$), a correction of + 0.11 cm would remove this bias. This is exactly the effect which we have derived using a principled approach and the plot on the right shows that the $\boldsymbol{H}_{\text{pred}}$ model has no significant bias.

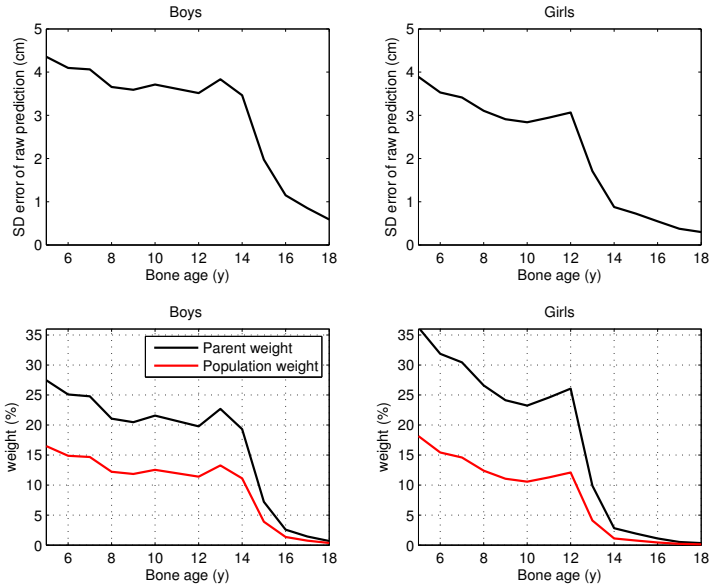

Figure 6: The top plots show the SD of $\boldsymbol{H} - \boldsymbol{H}_{\text{raw}}$, i.e. the prediction error of the raw, BP-like prediction. The bottom plots show the weights to use when forming the final prediction as a weighted average of the raw prediction and the prediction from parental or population height. These weights were computed according to Bayesian inference.
(Note: In order to obtain the overall correct magnitude of these weights (verified in plots like those in Figure 7), two fudge factors were applied: the population SDs were multiplied by a factor 1.4 and the SDs of the $\boldsymbol{H}_{\text{P}}$ prediction by 1.2)

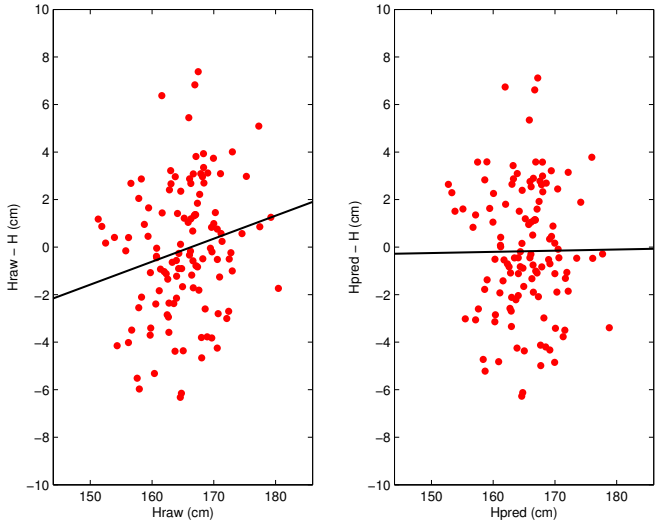

Figure 7: Comparison of the two predictions $\boldsymbol{H}_{\text{raw}}$ (left) and $\boldsymbol{H}_{\text{pred}}$ (right) for girls of bone age 9.5-10.5 yr. The y-axis shows the error of the predictions. The slope in the left plot is 0.11, so by drawing the raw prediction 11% of the way towards the population mean, the slope disappears, as seen in the right plot. $\boldsymbol{H}_{\text{raw}}$ corresponds to the raw Bayley-Pinneau-type prediction, which overestimates the adult height by 1.1 cm for every 10 cm above the mean.

3.3 Incorporation of parental height

The knowledge of the parents' heights should provide us with a more reliable *a priori* belief of the child's adult height, $p(\boldsymbol{H})$, and the elegance of the BX model is that this is implemented in the same way as described above for the population height.

First, we form the midparental height

$$\boldsymbol{H}_{\text{mid}} = \tfrac{1}{2}\,(\boldsymbol{H}_{\text{mother}} + \boldsymbol{H}_{\text{father}})$$

It is customary in clinical practice to invoke the parental height as the *target height*, defined according to Tanner as $\boldsymbol{H}_{\text{mid}} \pm 6.5$ cm for boys and girls respectively. This is sometimes used as a literal prediction of the height (Brämswig et al. 1990), which is an oversimplification.

To set up a proper prediction, $\boldsymbol{H}_{\text{P}}$, of the adult height based on parental height, the BX model uses the form

$$\boldsymbol{H}_{\text{P}} = a\,\boldsymbol{H}_{\text{mid}} + b + \boldsymbol{sec}$$

Here a and b are estimated from the 1ZLS data with separate formulae for boys and girls. The secular trend, $\boldsymbol{sec}$, is separated from the constant term, so that the model can be estimated using the 1ZLS and then generalized to populations with a different secular trend, and the result is

$$\boldsymbol{H}_{\text{P}} = 0.7884\,\boldsymbol{H}_{\text{mid}} + 42.2 \text{ cm} + \boldsymbol{sec}$$

for boys and

$$\boldsymbol{H}_{\text{P}} = 0.7186\,\boldsymbol{H}_{\text{mid}} + 40.3 \text{ cm} + \boldsymbol{sec}$$

for girls, and the SDs of the prediction residuals are 5.9 cm and 4.3 cm, respectively. This prediction model was first formulated by Galton, who found that a is less than one (in 1ZLS the average is 0.75) and named this phenomenon *regression towards mediocrity* (Galton 1886). Children of extraordinary parents are, on average, less extraordinary than their parents, i.e., they regress towards the average of the general population. Since then we have called these models *regression models*.

In order to avoid a bias when the model is applied to populations with a different mean, $\boldsymbol{H}_{\text{pop}}$, the BX method adds a term that removes the expected bias, so the final model is

$$\boldsymbol{H}_{\text{P}} = a\,\boldsymbol{H}_{\text{mid}} + b + (1 - a)(\boldsymbol{H}_{\text{pop}} - \boldsymbol{H}_{\text{pop1ZLS}}) + \boldsymbol{sec}$$

where $\boldsymbol{H}_{\text{pop1ZLS}}$ is 178.2 cm for boys and 165 cm for girls.

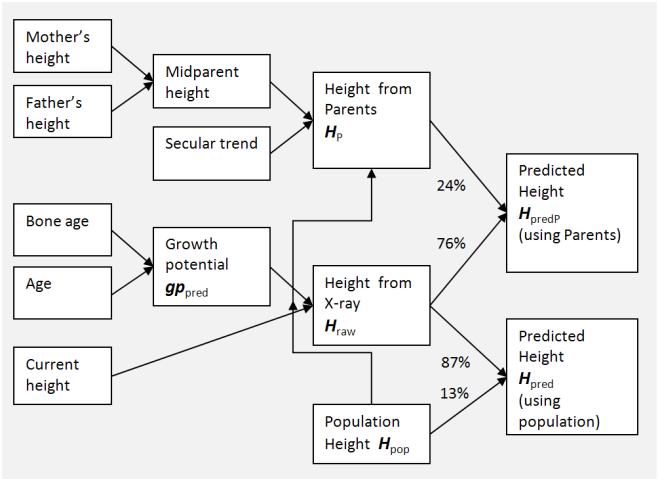

Figure 8: The information flow in the BX AHP method.
The weights at the arrows are approximate values for pre-pubertal children; the exact weights depend on BA, as shown in Figure 6.

The prediction errors of $\boldsymbol{H}_{\mathrm{P}}$ are smaller than the population SDs so from the Bayesian inference argument we can expect the weights of the parents' prediction to be larger, and indeed the weights are about 24% before puberty, as shown in Figure 6.

The final prediction of the BX method, $\boldsymbol{H}_{\mathrm{pred}}$ or $\boldsymbol{H}_{\mathrm{predP}}$, is formed by combining the raw BP-like prediction $\boldsymbol{H}_{\mathrm{raw}}$ with either the population or parental information using the weights in Figure 6:

$$\boldsymbol{H}_{\mathrm{pred}} = (1 - w_{\mathrm{pop}})\, \boldsymbol{H}_{\mathrm{raw}} + w_{\mathrm{pop}}\, \boldsymbol{H}_{\mathrm{pop}}$$

$$\boldsymbol{H}_{\mathrm{predP}} = (1 - w_{\mathrm{P}})\, \boldsymbol{H}_{\mathrm{raw}} + w_{\mathrm{P}}\, \boldsymbol{H}_{\mathrm{P}}$$

The entire flow of information in the model is shown in Figure 8. Notice that when the parents' heights are used for *a priori* knowledge, the population mean is still used, because it enters into the calculation of $\boldsymbol{H}_{\mathrm{P}}$.

In summary, the BX model separates the modelling of information sources into independent modules. The advantages are that each module is relatively simple to design and that the modules can be combined in different ways.

3.4 Performance of the BX model

The BX model was estimated on 231 children of the 1ZLS, and Figure 9 shows the root mean square errors of prediction observed with these data. There is a characteristic plateau with an approximately constant error from BA 7 to puberty, and the error actually exhibits a mild maximum at puberty. This is counterintuitive. One would expect that the prediction error decreases as we approach the target. The growth potential of boys drops from 30% to 11% from BA 7 to 13 yrs (Figure 1), but the prediction error is approximately the same. We speculate that the large uncertainty at puberty is due to the growth spurt; if there is a given uncertainty in our estimate on the whereabouts on the growth curve, this gives a larger error in height prediction when the curve is steeper. The clinical recommendation derived from this phenomenon is, that it is preferable to perform AHP at BA < 12 in boys and BA < 10.5 in girls, and unless the child is treated, there is little rationale for repeating an AHP during puberty.

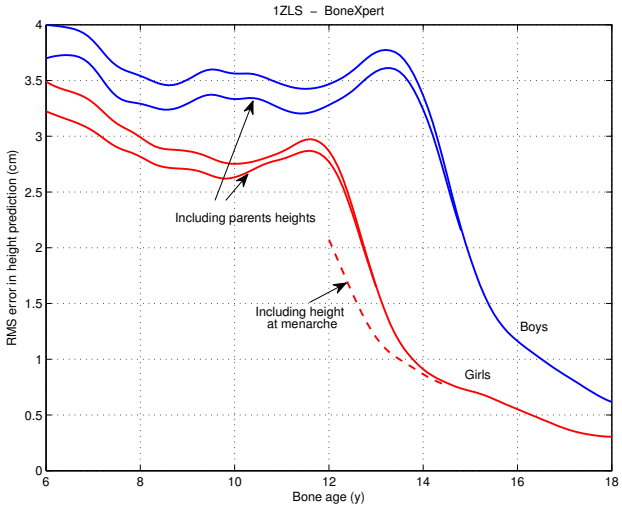

Figure 9: The observed RMS errors of the BX method. There are two solid lines for each sex; the lower ones include parents' heights. The dashed line for girls includes height at menarche (but not parents' heights)

When applying the BX method to new data, the prediction is given in terms of the center value, $\boldsymbol{H}_{\mathrm{pred}}$, or $\boldsymbol{H}_{\mathrm{predP}}$, and the associated expected prediction error, which is based on the performance on 1ZLS (Figure 9). Thus, the prediction uncertainty is an integral part of the prediction.

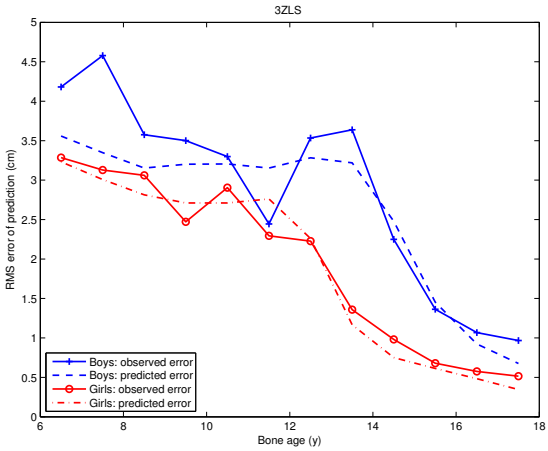

Figure 10: Validation of the BX AHP method on the 3ZLS. The observed errors (solid curves) agree well with the errors predicted from the model (dashed lines).

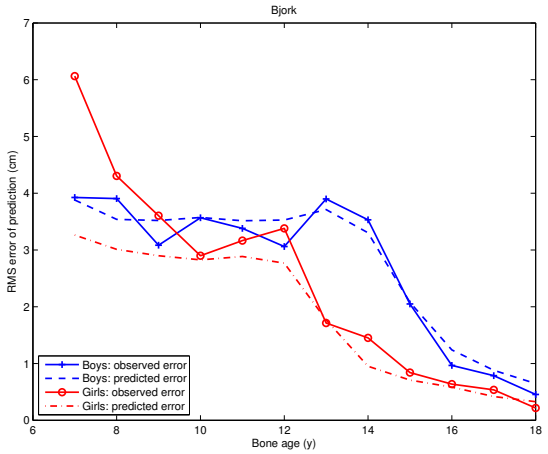

Figure 11: Validation of the BX AHP method on the Björk study. The observed errors (solid curves) agree well with the errors predicted from the model (dashed lines) except for girls below 9 yrs.

The BX method is validated by studying the RMS error of the predictions. This is convenient because it includes both the SD of the prediction and any bias. The performance on the 3ZLS (Thodberg et al. 2009a) is shown in Figure 10. Here we effectively predict the height at age 18 rather than the height when the increase is less than 0.5 cm over two years (the adult height according to the definition in the 1ZLS). Based on analyses on the 1ZLS, one can derive that such a prediction is easier and one should reduce the *expected* errors by a factor 1.11 for boys and 1.05 for girls, which has been done in Figure 10. The observed errors agree well with this, except for a tendency for larger errors than expected for boys below BA 7, although this is not statistically significant. The 3ZLS data represent similar genetic material as the 1ZLS, but with a median shift in time of 28 years.

Figure 11 demonstrates the validation of the BX model using the Björk study, presented here for the first time. Again the errors are as predicted, except for girls with BA < 9 yrs, where there is a statistically significant excess. The Björk data stem from approximately the same time as the 1ZLS (see Table 2),

but they represent a different population with about 0.4 yrs larger average BA delay.

Given these validation studies, it is reasonable to expect that the BX will work well for all healthy Caucasian children in Central and Northern Europe.

The fact that one loses virtually no accuracy in prediction when moving from the data set used to design the BX method to data taken 28 years later, or 1200 km North of Zürich, is a remarkable success for the BX method. It is due to the objective BA rating underlying this method. Had one used manual BA rating, there would inevitably have been some deterioration due to BA rater variability.

4 Discussion of the TW and RWT methods

4.1 Models at fixed CA

Up until now we have extensively covered the BP and BX methods, which are both based on a model of the growth potential. We now turn to a comparison with the TW and RWT methods. They both appeared in 1975, and they use nearly the same mathematical framework. These methods were estimated using longitudinal studies where the X-rays were taken close to the children's anniversaries, so it was expedient to estimate one regression model at each chronological age (CA). The observations for each regression are then statistically independent, and the age does not enter as an explicit variable.

Thus for each value of CA, the adult height is modelled as a linear function of the form:

$$\boldsymbol{H}_{\text{pred}} = a(\text{CA}) \times \boldsymbol{h} + b(\text{CA}) \times \text{BA} + c(\text{CA}) \times \boldsymbol{H}_{\text{mid}} + d(\text{CA})$$

The RWT method has an additional term, linear in the body weight (discussed in Section 5), while the TW Mark I and II models do not include the $\boldsymbol{H}_{\text{mid}}$ term. We have indicated that the coefficients *a*, *b, c* and *d* are associated with the specific integer CA. The challenge in this framework is to make this family of models, estimated at different CAs, consistent. Tanner used graphical methods to ensure that the coefficients *a*, *b,* etc., vary gently with CA, while RWT used mathematical smoothing methods, and an exploration of these techniques completely dominates the latest paper (Khamis & Guo 1993).

4.2 The problem with linear BA dependence

The dependency on BA is linear through the term $b(\text{CA}) \times \text{BA}$. However, from Figure 3, which represents the data at fixed CA values, we see that the dependence of ***gp*** on BA is far from linear. At low BA, the curvature of the curve-snippets is negative, while, at high BA it is positive. It is obvious that ***H*** cannot, in general, be considered a *linear* function of BA for fixed CA and ***h***. In other words**,** the TW & RWT methods oversimplify the BA-dependency. This might not be so severe for the bulk of normal children with BA near the CA. But, in clinical practice, there is a much higher frequency of children with severely delayed BA (constitutional delay of growth and puberty), or severely advanced BA (pubertas praecox), and these groups are poorly described if the model is linear in BA. The non-linear modelling of ***gp*** in the BX method is, therefore, one of its most important advantages.

4.3 Incorporation of parental height

The RWT model includes mid-parental height as one of its linear terms. The coefficient of this term drops by a factor of almost two during puberty, a drop which is not quite as steep as seen in the BX model, as shown in Figure 6. This is because the RWT models are calculated at fixed CA values, whereas Figure 6 is parameterized by BA. The coefficients $c(\text{CA})$ of the mid-parental height are 0.39 and 0.21 before puberty for boys and girls, respectively. These coefficients can be compared to the values used in the BX model of 0.79×0.22 = 0.17 and 0.71×0.25 = 0.18 (the product of the *a*-coefficient of the $\boldsymbol{H}_{\text{P}}$ model with the prepubertal level in Figure 5). Thus, the contribution is about the same for girls, while for boys the RWT model has a remarkably strong contribution from the parents. Indeed the boys of the Fels study seem to behave differently compared to the Zürich data. The bone age contributes significantly to AHP only at age 13 and above.

If the parental heights are unavailable, Roche proposes to insert the population mean instead, but this will overemphasize the population mean – the strategy used in the BX model is more correct.

Tanner was struggling with the use of the parental heights. He did not include a linear term as did RWT. In TW Mark I he had a rule similar to the drawing rule of the BX model, but he was unclear about how the correction due to the parents should drop off after puberty, and in the Mark II method he rejected the drawing rule altogether as "unwise". In the TW3 model, he reintroduced the midparental height as an explanatory variable at each CA as in the RWT model. He found that the prediction error was reduced by 0.23 cm for boys and 0.13 for girls before puberty; exactly the same effect as found in the BX model (Figure 9), also based on the 1ZLS but using BX BA. Tanner decided, however, that the improvement was too small to be worth a presentation of the equations.

4.4 Other features of TW models

The TW models featured other ideas. The TW Mark II model was presented in various versions that included increments in height and/or BA over the last year as explanatory variables. A version with height increment is also presented in the TW3 model, but these models are rarely used in clinical practice due to their complexity.

The TW3 model featured a slightly different form of the regression equations:

$$\boldsymbol{H}_{\text{pred}} = \boldsymbol{h} + b(\text{CA}) \times \text{SMS} + c(\text{CA})$$

This amounts to forcing $a(\text{CA})$ to be 1, which appears like a numerical coincidence. SMS is just a non-linear transformation of BA, but it is unclear why this should be better than using simply BA.

Table 3: The prediction errors of the TW3 AHP method (based on manual bone age) and the BX method (based on BoneXpert GP bone age) in the 1ZLS. The errors are averages over the indicated chronological age ranges.

	Age range (yrs)	TW3 Residual SDs (cm)	BoneXpert Residual RMSs (cm)
Boys	10-15	3.5	3.3
Girls	8-13	3.1	2.7

4.5 Comparison of TW3 and BX performance

The TW3 and BX methods for adult height prediction are both based on the 1ZLS, and this allows a direct comparison of their performance. The TW3 method was based on manual TW ratings (performed at the time of the study by a group of several experienced operators) and its performance is derived from Tables 10 and 12 in (Tanner et al. 2001). The result is shown in Table 3. The BX method has a slightly better performance with boys, and significantly better performance with girls ($p < 0.005$).

5 Other predictors of adult height

This section presents a fairly exhaustive account of other parameters, which could possibly contribute to AHP in addition to age, BA, ***h*** and parental height.

5.1 Menarche

The first menstrual bleeding (menarche) is the result of several years of accumulated exposure of the endometrium to estrogen. Such exposure will concomitantly affect the glandular breast tissue resulting in breast development, as well as influence bone, inducing a pubertal growth spurt, and subsequently controlling the fusion of the growth zones. Thus, it is not surprising that the age at menarche is highly predictable from BA. One can consider menarche to be a timestamp of maturation similar to BA, and one can expect that the growth potential is predictable at menarche.

Figure 12 shows the remaining height growth at menarche in the 1ZLS. The taller girls tend to have less remaining growth compared to the shorter girls, but the BX method summarizes these data in a simple model where the average remaining height is 6.6 cm with an SD of 2.2 cm. This additional piece of information is then combined with the estimate from the radiograph using Bayesian inference, leading to a considerably smaller AHP error for post-menarchal girls, as is shown by the "menarche" curve in Figure 9.

The TW method implements the menarche information by setting up separate formulae for pre- and post-menarchal girls. This adds to the complexity of the method and does not use information regarding the *time* of menarche, once it has occurred.

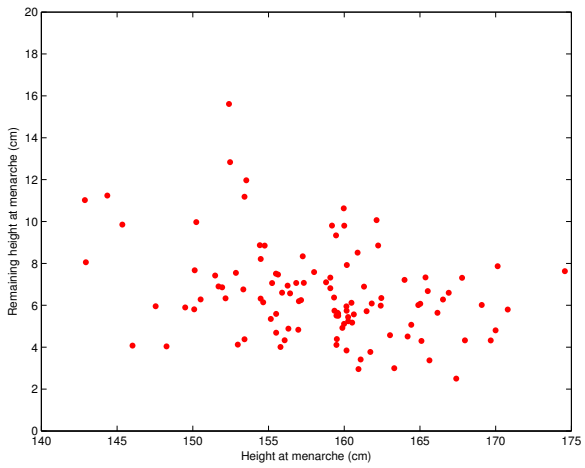

Figure 12: The remaining height growth at menarche is 6.6 cm ± 2.2 cm (SD) in the 1ZLS

5.2 Body weight

This section presents empirical data on the role of body weight as an additional parameter in the BX model. This is implemented by forming the standard deviation score (SDS) of the body mass index (BMI) for BA and computing a correction to the AHP as a linear function of BMI SDS.

Figure 13 shows the height correction per BMI SDS for the 1ZLS. There is a pronounced effect for boys, and the correction is negative with a magnitude of approximately 1.5 cm per BMI SDS up to approximately 13 years. Thus, a boy with positive BMI SDS needs a negative correction of his AHP and vice versa. For girls, the effect is smaller, at less than 0.5 cm per BMI SDS. Figure 13 also demonstrates that the correction looks similar when one uses the skinfold SDS instead of the BMI SDS.

The BX method implements a correction for BMI but only for boys, and the correction is the same for the population-based estimate $\boldsymbol{H}_{\text{est}}$, and the parents-based estimated $\boldsymbol{H}_{\text{estP}}$, (see Thodberg et al. 2009a for details).

Several studies have shown that a higher BMI in childhood leads to an earlier puberty and a lower adult height (Sandhu et al. 2006). This could lead one to assume that BMI affects adult height only via BA. To illustrate this effect, Figure 14 shows the BA advancement per BMI and skinfold SDS in the 1ZLS. The advancement at BA = 10 yrs of about 0.3 yr induces a reduction in the predicted AHP of 0.3 cm for boys and 1.1 cm girls (BA has a rather small effect on the predicted adult height for boys at this age, as can be seen from Figure 3).

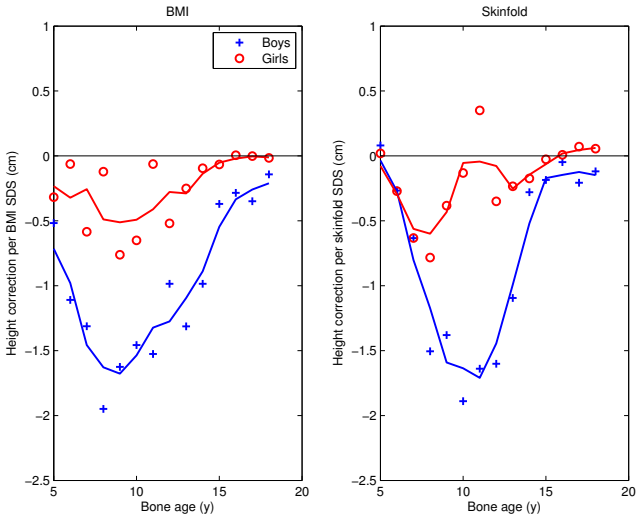

Figure 13: The height correction to be applied to $\boldsymbol{H}_{\text{pred}}$ or $\boldsymbol{H}_{\text{predP}}$ of the 1ZLS in order to take BMI (left plot) or skinfold thickness (right plot) into account. The curves are a smoothed version of the year-by-year data.

The 1ZLS study showed that when predicting adult height from current height, CA and BA are not sufficient for capturing all the effects of a higher BMI, in particular for boys. In other words, BMI has an effect on final height, which is *independent* of its effect on BA shown in Figure 14. One achieves a more accurate prediction of adult height by the explicit adjustment for BMI SDS. Since the *independent* BMI-effect on adult height disappears at puberty, we can assert that BMI in childhood predicts the strength of the pubertal growth spurt for boys. Once more we see that the challenge of AHP is to find predictors of the strength of the growth spurt, which varies from child to child, in particular for boys. In section 2 we saw

that a BA delay indicates a smaller spurt; now we see that high BMI does also.

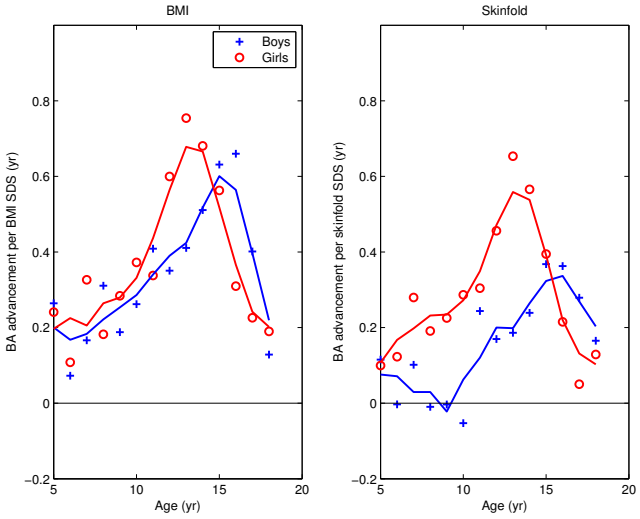

Figure 14: The association of BMI and skinfold with bone age in the 1ZLS. The curves are a smoothed version of the year-by-year data.
All figures in this chapter, except Figures 4, 5, 11, 15, 16 and 17, are adapted from (Thodberg et al. 2009a).

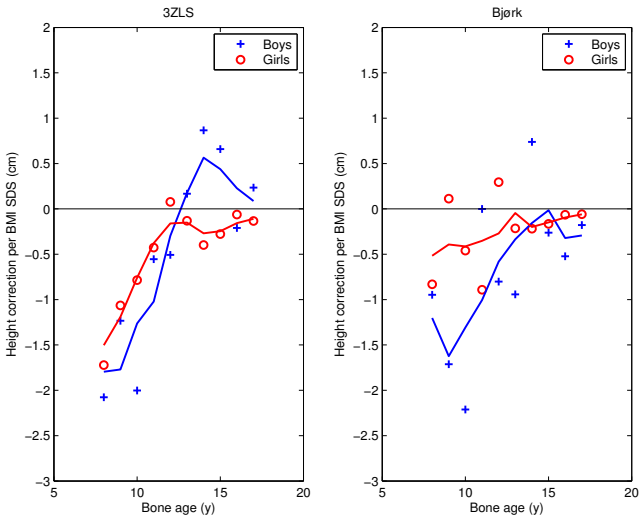

Figure 15: The height correction to be applied to $\boldsymbol{H}_{\text{pred}}$ of the 3ZLS (left) and Björk study (right) in order to take BMI into account. The curves are a smoothed version of the year-by-year data.

The RWT method included weight, but the use of the height and weight, in the same regression equation, makes it difficult to understand the mechanism, because these two variables are highly correlated. The BX method, in contrast, applies the weight adjustment after the height has explained as much as it can, and uses weight in the guise of BMI, which, to a first approximation, is independent of height.

Figure 15 shows the effect of BMI on AHP in the 3ZLS and Björk studies. The effect for boys is again –1 to –1.7 cm per BMI SDS with BA below 12 yrs. But for 3ZLS, there is also a surprisingly large effect for girls.

However, when validating the BX method using the 3ZLS and the Björk study, we found no significant improvement when including the BMI correction. We think, therefore, that it is unwise to include BMI for AHP in clinical practice. Not only is the effect fairly small, but obesity and dieting can interfere with such a correction. There is no doubt that BMI plays a role in AHP, but more research is needed to clarify it.

5.3 Tanner stages

The sexual development is traditionally evaluated in terms of Tanner scores ranging from 1 to 5, and separate scores are assigned to pubic hair development and to genital or breast development. There are no studies on the rater variability of these scores, but they are regarded as fairly reliable measures of the progression of puberty, and it is, therefore, natural to ask whether these scores can further improve the AHP.

This topic has been studied by (Onat 1983) who found that information on sexual development *does* contribute to AHP in addition to BA. Tanner stages are, in general, strongly related to BA, but when they disagree with BA, they can be used to give a correction to the AHP. Thus, one can form the "Tanner stage SDS for BA" and derive a post-processing to the standard BX method, analogous to the treatment of BMI outlined above. Preliminary studies on the 1ZLS show that this turns out to reduce the prediction RMS error by 0.2 to 0.3 cm for boys with BA 11-15 yrs and girls with BA 10-13 yrs, thus confirming Onat's findings. This is the BA range where the standard BX model has relatively large errors, so the Tanner stages supplement BA where it seems most needed.

However, we do not recommend using Tanner stages for AHP in clinical practice because we have observed this effect in the high-quality 1ZLS, and in clinical practice these stages are likely to be less reliable. In addition, the BX method is an *objective* AHP method, and we consider the introduction of Tanner stages with unknown rater variability to represent a step backwards.

5.4 Determining which bone age and which bones should be used

When setting up the TW Mark I model, Tanner discovered that the carpal bones do not contribute useful information to the AHP, so he based the method on TW-RUS BA. But Tanner did not go on to examine the relative merit of the radius and ulna (the wrist) versus the short bones. He always used the standard weighting of TW RUS which assigns a weight of 20% each to the radius, ulna, ray 1, ray 3 and ray 5.

In the GP system, all bones have the same weight, and if one omits the carpals, i.e., uses the 19 short bones plus the radius and ulna, the weight on each bone is close to 5%. This means that the wrist accounts for only about 10%, in contrast to the 40% chosen in the TW-based methods. This is the largest methodological difference between the TW and the GP BA, and this difference is inherited by the AHP methods which are based on these two BA systems.

In order to investigate the relative merit of TW and GP BA and the optimal weight for the wrist, a special study was performed based on the 1ZLS (Thodberg et al. 2009c). The aim was to quantify how well different bone age methods can predict the growth potential, and this was measured by the ***gp*** prediction error (GPPE), averaged over the age range of 10-18 yrs for boys and 8-16 yrs for girls.

Four different BA systems were compared including manual TW-RUS, manual GP, BX, and BX/short, the latter being the BX BA averaged over the 11 short bones, excluding the wrist. The manual TW and GP ratings were performed by the same raters,

as part of the original 1ZLS. The results are shown in Table 4. Manual GP was significantly better than the manual TW rating. The BX rating was slightly, but not significantly, better than manual GP rating, and excluding the wrist, actually improved the prediction marginally. In other words, the wrist does not contribute useful information to AHP in the BX method, when we have many short bones. It is, therefore, not unreasonable to assume that the poor performance of manual TW relative to manual GP is due to the large weight of the wrist, which "contaminates" the TW BA with information irrelevant for AHP.

Table 4: Growth potential prediction error (GPPE) of four different bone age methods.

Bone age system	Both sexes [95% CI]
Manual TW	1.32 [1.28; 1.36]
Manual GP	1.26 [1.22; 1.30]
BX	1.23 [1.19; 1.27]
BX / short	1.22 [1.18; 1.26]

This study confirms that GP BA, with its relatively small weight on the wrist, is close to the optimal BA for AHP.

Finally, a study was performed, again using the 1ZLS, to compare the BX BA for the left versus the right hands (Martin et al. 2009b). The average (signed) difference was found to be consistent with zero, which implies that it does not matter which hand is used for BA, in general, and AHP, in particular, (but it is advised to use the same hand consistently in longitudinal studies to obtain the most reliable BA increments).

5.5 Other parameters

It has been suggested that body shape or somatotype is related to bone age and growth potential, so one could query whether such information is relevant for AHP in a manner not already taken into account by BA and BMI.

Therefore, the following parameters were examined to test whether they explain some of the residual errors in the BX AHP models.

- Bone health index SDS for BA, based on cortical thickness of the metacarpals (Thodberg et al. 2009d).
- Aspect ratio of hand bones using the ratio of the average length to the average width of ten short bones (5 metacarpals, 3 proximal and 2 middle phalanges).
- Relative size of the hand using the ratio of the average length of ten short bones to the stature.

Based on the data in the 1ZLS, 3ZLS and Björk study, no significant contribution to AHP from these parameters was found.

6 Practical methods and techniques

6.1 Key advantages of the BoneXpert method

To summarize, the key advantages of the BX method are the following:

1. The BX method is based on an automated BA, thereby removing the rater variability, which has been the main problem in AHP so far.
2. The BX method is based on more recent data than the methods most often used today, i.e. the BP method from 1952, the RWT method from 1975 and the TW Mark II methods from 1983. The BX method has been validated in a large recent study from Zürich, and also in a study from Denmark.

From the methodological point of view, the BX method appears to have two advantages, which have, however, not yet been demonstrated empirically:

3. The BX method employs a non-linear BA dependence, which enables it to accommodate, within the same overall model, both normal children as well as children with severely advanced or delayed BA.
4. The BX method embodies a better understanding of the role of the population mean and parental heights, which enables it to accommodate normal, short and tall stature within the same model.

6.2 Practical application of BX AHP

The practical application of the BX AHP method is a two-step procedure.

1. The first step is to perform the BX BA determination. This is done using the commercially available BoneXpert program for BA determination (Visiana, Holte, Denmark, www.BoneXpert.com). This is less burdensome if performed by the radiology department, which has easy access to the hand radiograph as a Dicom file, but the program can also be operated by the pediatrician.
2. The second step is to use the BoneXpert GP BA in combination with gender, age, height and additional parameters, if desired. This can be performed using the calculator available on-line at www.BoneXpert.com. An example is shown in Figure 16. BA, age, ***h***, etc. are typed in and the AHP is shown immediately, including the variants with and without parental height, etc. The calculator is freely available and can also be integrated as a component in other software, e.g., an electronic patient journal system.

If BX BA is not available, one can replace the first step, as a second-best solution, by a *manual* GP BA determination. However, the uncertainty in the AHP will then be larger than that indicated by the BX method. How much larger is difficult to say, because it depends on the reliability of the manual rating.

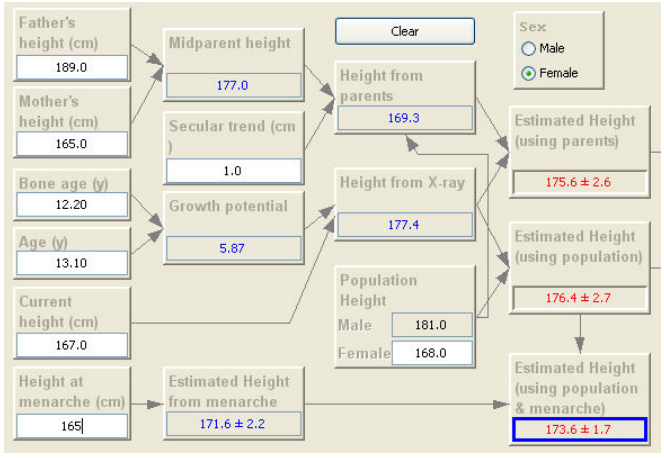

Figure 16: The on-line calculator for AHP, available at www.BoneXpert.com. This example illustrates the AHP for a post-menarchal girl.

6.3 Application area

There are two important limitations of the BX AHP model, which also applies to all the previous AHP methods:

- The BX method was developed and validated on Caucasian children. More studies are needed to verify whether the BX model is valid for other ethnicities, and, if not, to develop special versions of it for these other ethnicities.
- The BX AHP model was developed only for healthy children and extreme variations of healthy children which have no pathologies and are untreated.

By extreme variations of normal we mean children with a large BA delay or advancement and children with extremely low or high adult height. This two-dimensional space of conditions is visualized in Figure 17, indicating the common names for these conditions. These children present in pediatric endocrinology with exceptional tall or short stature for age or with early signs of puberty (or absence of the expected signs of puberty). In these situations, it is customary to perform a bone age determination and an AHP. It is important to stress that the AHP is reliable only if no pathology is observed and if the child is untreated.

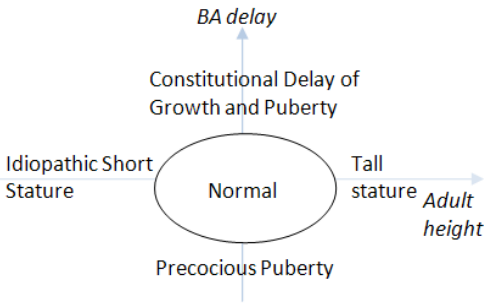

Figure 17: The area of intended use of the BX AHP model.

6.4 Applications to other areas of health and disease

Notwithstanding the previous section, AHP is frequently applied to children with Growth Hormone Deficiency (GHD), Turner Syndrome and other pathologies, or to children treated with growth hormone (GH), i.e., clearly beyond the scope of these models. Tanner has expressed some sympathy for the use of AHP in such conditions during treatment. One can monitor the *predicted* $\boldsymbol{H}$, and if this increases, it can be taken as an indication that the treatment is successful. The rationale behind this assumption is that GH speeds up $\boldsymbol{h}$, but sometimes also speeds up BA. If BA increases excessively, it could jeopardize the positive effect of GH on $\boldsymbol{H}$, because $\boldsymbol{gp}$ decreases with BA.

We emphasize that such use of AHP methods must be considered experimental, and whether this is a sound practice requires dedicated studies on such patients. We suspect that it would be more productive to construct mathematical models using a more principled approach to address these issues.

If the BX AHP model should be extended to encompass GHD children, one could attempt to add further parameters, like the severity of the deficiency, the GH dose, and, possibly, some other parameters that characterize the patient's sensitivity to GH, such as IGF-1 levels, etc.

Finally – to close on a lighter note – AHP has also been applied to ballet girls to predict whether they will reach an adult height in the range required by the corps. This is clearly within the scope of the model, and data on ballerinas were, in fact, included in the design of the TW Mark II model.

6.5 Summary

In a greater perspective, we summarize this review, as follows.

- Adult height prediction is an important tool for pediatric endocrinologists, but it is also a tool for improving our understanding of bone growth, maturation and puberty in general, because there are good quantitative data to challenge the models.
- Children are different and do not follow a universal growth curve. Although the peak height velocity seems to occur at the same BA, the *strength* of the growth spurt varies. The challenge of AHP is, therefore, to unravel predictors of the strength of the pubertal spurt. We have seen that prepubertal BA delay and high BMI are such predictors, affecting the spurt strength negatively, in particular for boys.
- The new BoneXpert method for AHP, which is based on automated BA, is an example of *evidence-based medicine*, in the sense that it replaces a currently accepted subjective procedure with a more accurate, objective method.
- We believe that mathematical/statistical modelling is a key to progress in the understanding of growth and the management of growth-related conditions, and it is our hope that this chapter can serve as a paradigm for such developments.

Acknowledgements

We thank Elisabeth Kaelin for data management and Julia Neuhof for scanning of X-rays in Zürich.

We acknowledge Novo Nordisk for providing the scanner used to digitize the Zürich and Björk studies.

Appendix: A manual AHP with the BX model

This appendix demonstrates how to perform the calculation of Figure 16 manually using information in this chapter.

We consider a 13.1 year old girl with BA 12.2 yr. According to Figure 2, she has a $\boldsymbol{gp}$ = 5.9%. With a current height of 167 cm, this yields a raw adult height prediction

$$\boldsymbol{H}_{\text{raw}} = 167 \text{ cm}/(1 - 5.9/100) = 177.4 \text{ cm}.$$

The population mean $\boldsymbol{H}_{\text{pop}}$ is assumed to be 168 cm, which is incorporated with a weight 9% from Figure 6. The difference between $\boldsymbol{H}_{\text{raw}}$ and $\boldsymbol{H}_{\text{pop}}$ is 11.4 cm and 9% of this is 1.0 cm. So the final result, with the SD read from Figure 9, is

$$\boldsymbol{H}_{\text{pred}} = 176.4 \pm 2.7 \text{ cm}$$

The parent's heights are 165 cm and 189 cm, so the mid-parental height is 177 cm. The secular trend is assumed to be 1 cm, and the height prediction from the parents is made according to the formula in Section 3.3:

$$\boldsymbol{H}_{\text{P}} = 177 \text{ cm} \times 0.7186 + 40.3 \text{ cm} + 1 \text{ cm} + (1 - 0.7186)(168 - 165) \text{ cm}$$
$$= 169.3 \text{ cm}.$$

According to Figure 6, this is combined with $\boldsymbol{H}_{\text{raw}}$ with a weight 22%, i.e., $\boldsymbol{H}_{\text{raw}}$ is corrected by 22% of 8.1 cm = 1.8 cm (the SD read from Figure 9):

$$\boldsymbol{H}_{\text{predP}} = 175.6 \pm 2.6 \text{ cm}.$$

The girl has passed her menarche, and the height at menarche was 165 cm, from which an independent AHP of 165 + 6.6 cm = 171.6 ± 2.2 cm is derived. This is combined with $\boldsymbol{H}_{\text{pred}}$ using a weighted average, where the weights are given by the precisions of each estimate, 0.21 cm^{-2} and 0.14 cm^{-2} respectively, so 60% of the weight is from the menarche information. The SD is read off from Figure 9, and the best prediction becomes

$$\boldsymbol{H}_{\text{predM}} = 173.6 \pm 1.7 \text{ cm}$$

References

Bayley, N. 1946. Tables for predicting adult height from skeletal age and present height. *Journal of Pediatrics*, 28, 49-64

Bayley, N. & Pinneau, S.R. 1952. Tables for predicting adult height from skeletal age: revised for use with the Greulich-Pyle hand standards. *J Pediatr*, 40, (4) 423-441

Binder, G., Grauer, M.L., Wehner, A.V., Wehner, F., & Ranke, M.B. 1997. Outcome in tall stature. Final height and psychological aspects in 220 patients with and without treatment. *Eur.J Pediatr.*, 156, (12) 905-910

Björk, A. 1968. The use of metallic implants in the study of facial growth in children: method and application. *American Journal of Physical Anthropology*, 29, (2) 243-254

Brämswig, J.E., Fasse, M., Holthoff, M.L., Von Lengerke, H.J., Von Petrykowski, W., & Schellong, G. 1990. Adult height in boys and girls with untreated short stature and constitutional delay of growth and puberty: accuracy of five different methods of height prediction. *The Journal of pediatrics*, 117, (6) 886-891

de Waal, W.J., Greyn-Fokker, M.H., Stijnen, T., van Gurp, E.A., Toolens, A.M., de Munick Keizer-Schrama SM, Aarsen, R.S., & Drop, S.L. 1996. Accuracy of final height prediction and effect of growth-reductive therapy in 362 constitutionally tall children. *J Clin.Endocrinol.Metab*, 81, (3) 1206-1216

Galton, F. 1886. Regression towards mediocrity in hereditary stature. *Journal of the Anthropological Institute of Great Britain and Ireland*, 15, 246-263 http://galton.org/essays/1880-1889/galton-1886-jaigi-regression-stature.pdf

Greulich, W.W. & Pyle, S.I. 1959. *Radiographic Atlas of Skeletal Development of the Hand and Wrist, 2.ed* Stanford, Stanford University Press.

Joss, E.E., Temperli, R., & Mullis, P.E. 1992. Adult height in constitutionally tall stature: accuracy of five different height prediction methods. *Arch.Dis.Child*, 67, (11) 1357-1362

Khamis, H.J. & Guo, S. 1993. Improvement in the Roche-Wainer-Thissen Stature Prediction Model: A Comparative Study. *American Journal of Human Biology*, 5, 669-679

Martin, D.D., Deusch, D., Schweizer, R., Binder, G., Thodberg, H.H., & Ranke, M.B. 2009a. Clinical application of automated Greulich-Pyle bone age in children with short stature. *Pediatr.Radiol.*, 39, (6) 598-607

Martin, D.D., Neuhof, J., Jenni O G, Ranke, M.B., & Thodberg, H.H. 2009b. Automatic Determination of Left and Right Hand Bone Age in the First Zurich Longitudinal Study. *Hormone Research* accepted for publication

Onat, T. 1983. Multifactorial prediction of adult height of girls during early adolescence allowing for genetic potential, skeletal and sexual maturity. *Human biology; an international record of research*, 55, (2) 443-461

Prader, A., Largo, R.H., Molinari, L., & Issler, C. 1989. Physical growth of Swiss children from birth to 20 years of age. First Zurich longitudinal study of growth and development. *Helv.Paediatr.Acta Suppl*, 52, 1-125

Roche, A.F., Chumlea, W., & Thissen, D. 1988. *Assessing the Skeletal Maturity of the Hand-wrist: Fels Method* Springfield, Illinois, Thomas.

Roche, A.F., Wainer, H., & Thissen, D. 1975. The RWT Method for the Prediction of Adult Stature. *Pediatrics*, 56, (6) 1026-1033

Sandhu, J., Ben-Shlomo, Y., Cole, T.J., Holly, J., & Davey, S.G. 2006. The impact of childhood body mass index on timing of puberty, adult stature and obesity: a follow-up study based on adolescent anthropometry recorded at Christ's Hospital (1936-1964). *Int J Obes (Lond)*, 30, (1) 14-22

Tanner, J.M. 1989. Review of "Assessing the skeletal maturity of the hand-wrist: FELS method". *American Journal of Human Biology*, 1, (4) 493-494

Tanner, J.M., Healy, M.J.R., Goldstein, H., & Cameron, N. 2001. *Assessment of skeletal maturity and prediction of adult Height (TW3 Method).* London: WB Saunders

Tanner, J.M., Landt, K.W., Cameron, N., Carter, B.S., & Patel, J. 1983a. Prediction of adult height from height and bone age in childhood. A new system of equations (TW Mark II) based on a sample including very tall and very short children. *Arch.Dis.Child*, 58, (10) 767-776

Tanner, J.M., Whitehouse, R.H., Cameron, N., Marshall, W.A., Healy, M.J.R., & Goldstein, H. 1983b. *Assessment of skeletal maturity and prediction of adult height* London, Acad. Press.

Tanner, J.M., Whitehouse, R.H., & Healy, M.J.R. 1962. *A new system for estimating skeletal maturity from the hand and wrist, with standards derived from a study of 2,600 healthy British children.* Paris: Centre International de l'Enfance

Tanner, J.M., Whitehouse, R.H., Marshall, W.A., & Carter, B.S. 1975a. Prediction of adult height from height, bone age, and occurrence of menarche, at ages 4 to 16 with allowance for midparent height. *Archives of Disease in Childhood*, 50, (1) 14-26

Tanner, J.M., Whitehouse, R.H., Marshall, W.A., Healy, M.J.R., & Goldstein, H. 1975b. *Assessment of skeletal maturity and prediction of adult height* London, Acad. Press.

Thodberg, H.H. 2009. Clinical Review: An automated method for determination of bone age. *Journal of Clinical Endocrinology & Metabolism*, 94, (7) 2239-2244

Thodberg, H.H., Jenni O.G., Caflisch J., Ranke, M.B., & Martin, D.D. 2009a. Prediction of adult height based on automated determination of bone age. *Journal of Clinical Endocrinology & Metabolism* epublication ahead of print, November 19, 2009, doi:10.1210/jc.2009-1429

Thodberg, H.H., Kreiborg, S., Juul, A., & Pedersen, K.D. 2009b. The BoneXpert method for automated determination of skeletal maturity. *IEEE Trans.Med.Imaging*, 28, (1) 52-66

Thodberg, H.H., Neuhof, J., Ranke, M.B., Jenni O G, & Martin, D.D. 2009c. Validation of bone age methods through their ability to predict adult height. *Hormone Research* accepted for publication

Thodberg, H.H., Rijn, R.R., Tanaka, T., Martin, D.D., & Kreiborg, S. 2009d. A Pediatric Bone Index derived Automated Radiogrammetry. *Osteoporos.Int.* accepted for publication

van Rijn, R.R., Lequin, M.H., & Thodberg, H.H. 2009. Automatic determination of Greulich and Pyle bone age in healthy Dutch children. *Pediatric Radiology*, 39, (6) 591-597

Zachmann, M., Sobradillo, B., Frank, M., Frisch, H., & Prader, A. 1978. Bayley-Pinneau, Roche-Wainer-Thissen, and Tanner height predictions in normal children and in patients with various pathologic conditions. *The Journal of pediatrics*, 93, (5) 749-755