Aus der Universitäts-Klinik für Kinder- und Jugendmedizin Tübingen

Abteilung Kinderheilkunde I mit Poliklinik

(Schwerpunkt: Allgemeine Pädiatrie, Hämatologie/ Onkologie)

Ärztlicher Direktor: Professor Dr. Rupert Handgretinger

Validation of Automatic Greulich-Pyle Bone Age on Children with short stature of various diagnoses

Inaugural-Dissertation
zur Erlangung des Doktorgrades
der Medizin

der Medizinischen Fakultät
der Eberhard Karls Universität
zu Tübingen

vorgelegt von
Dorothee Nyembo, geb. Deusch
aus
Pforzheim

2011

Dekan: Professor Dr. I. B. Autenrieth

1. Berichterstatter: Professor Dr. M. B. Ranke

2. Berichterstatter: Privatdozent Dr. J. Schäfer

Dem Einen

TABLE OF CONTENTS

Abbreviations

autBA	automatic bone age ratings; BoneXpert value
BA	bone age
BXp	BoneXpert
BXpBA	BoneXpert bone age ratings; equates to autBA
CA	chronological age
CASAS	computer-assisted skeletal age score
CASMAS	computer-aided skeletal maturity assessment system
CBA	CASAS bone age
CI	confidence interval
CR	computed radiography
DICOM	Digital Imaging and Communications in Medicine; data format
DP1	distal phalanx of finger 1
DP3	distal phalanx 3
DP5	distal phalanx 5
dpi	dots per inch
GH	growth hormone
GHD	growth hormone deficiency
GP	Greulich-Pyle
h	height
ISS	idiopathic short stature
manBA	original manual bone age ratings

meta1	metacarpal of finger 1
meta3	metacarpal 3
meta5	metacarpal 5
metaL	average length in mm of metacarpal 2-4
MP3	middle phalanx 3
MP5	middle phalanx 5
New manBA	average of three blind manual re-ratings
PP1	proximal phalanx of finger 1
PP3	proximal phalanx 3
PP5	proximal phalanx 5
RefBA	reference bone age; batch of original manual bone age replaced with New manBA values, if given
RUS	*R*adius, *U*lna and 11 *S*hort bones, used for TW maturity score
SD	standard deviation
secGHD	secondary growth hormone deficiency: caused by tumour, surgery, radiation, or chemotherapy
SGA	small for gestational age
SRS	Silver-Russell Syndrome
Std	Standard
TW	Tanner-Whitehouse
UTS	Turner Syndrome
vs.	versus
y	year, years

1 Introduction

Skeletal maturation (bone age; BA) is one of various parameters for the development of an organism in its entirety. Physiological development depends on a multiplicity of endogenous and exogenous factors. An imbalance in these parameters results in developmental retardation or acceleration and, important for diagnostics, retardation or acceleration of skeletal maturity.

1.1 Bone age determination

Skeletal maturity is routinely determined for the diagnostic evaluation of children affected by short stature in pediatric endocrinology clinics. BA is also used in regular intervals to monitor various forms of hormonal treatments (14;27;29;30;32;53). Several methods of BA determination have been developed, mentioned below (see chapter 1.1.2).

1.1.1 Problems with bone age determination

One would have expected BA to become part of the prediction models for response to growth hormone treatment in short children born small for gestational age (SGA) (27), children with growth hormone deficiency (GHD), or children with turner syndrome (UTS) (32), but this was not the case. The reason for this may be that BA was assessed by different radiologists with different methods all having had a subjective component (3;4;7;12;13;16;17;19;34-36;52;55).

The intra- and inter-individual variations of BA assessment in clinical and research practice impose a major handicap on its usefulness in clinical trials and research. Further, well-trained bone age raters are increasingly rare.

For these reasons, substantial efforts have been made to find ways of reducing the intra- and inter-observer differences in BA assessment by systematization (46) and automation (43;44).

1.1.2 The history of bone age determination

Commonly there are two methods used in clinical practice of bone age determination. The score method of Tanner and Whitehouse (TW) uses continuous stages (42) and the atlas method of Greulich and Pyle (GP) works with morphologically describing standards (15). The more time-consuming TW method has been shown to correlate less well with growth potential than the GP method (23).

A semiautomatic method CASAS (computer-assisted skeletal age score) was presented in 1992 (47). It analyzes the 13 bones of the TW-RUS system (RUS is a short for Radius, Ulna and 11 Short bones - the short bones used are those in ray 1, 3 and 5). CASAS requires a considerable amount of manual work, but yields a smaller inter-observer variation than manual rating, and the results are more precise (the longitudinal development is more continuous) (47).

In 1997 an automated method CASMAS (computer-aided skeletal maturity assessment system) was presented (39), which was reported to analyze 90% of the films. It analyzes four bones: radius and the phalanges of finger 3.

These methods have not found use in clinical practice, and many experts simply continue to use the GP Atlas because of its intuitive simplicity and rapidity. A fully automatic system that considers the routine, while rejecting the abnormal for the expert to assess, as with blood counts, has long been called for (45).

BoneXpert, a radically new method (48), may have the potential of meeting this need. It is a fully-automated program that can read digital hand X-rays and produce a calculation of BA based on the automatic geometrical assessment of the maturity of each of the 13 RUS bones in the hand. BoneXpert automatically rejects radiographs with bad exposure, bad pose of the hand, or abnormal bone shapes. Being based on the individual BA of each bone, it can provide information on the congruity of ossification between the various bones of the hand.

BoneXpert has been developed mainly on a relatively small set of normal children (48), and has been validated on a large study of normal children from Rotterdam (51) and on a set of short stature children from Tübingen (21;22). However, the latter two validation studies had the weakness that the images used had also been used to calibrate BoneXpert. They were thus able to assess the precision (standard deviation, SD) between BoneXpert and the manual ratings but not any bias.

In the present study, BoneXpert is validated without calibration (i.e. in the way it would be applied in clinical routine) in a much larger set of children with short stature of various etiologies. The X-rays were taken while children were not treated as well as while some have been under growth hormone treatment.

1.2 Aim of this study

The aim of this study is to validate BoneXpert in a large set of children with short stature of diverse diagnoses.

We survey the coverage of BoneXpert by analyzing characteristics of images accepted and rejected respectively by BoneXpert, the accuracy beyond BoneXpert's intended age ranges as well as its interoperability in use of digital or printed X-rays.

Subsequently we analyzed the reliability of BoneXpert. On the one hand we investigate the validation of BoneXpert compared with other determination methods:

- comparing with former manual atlas ratings
- comparing a set of UTS girls with additional semi-automatic BA

On the other hand we analyzed the validation of BoneXpert's use in clinical practice:

- using BXpBA examining bone maturation after the first year of growth hormone application.

1.3 Most prevalent diagnoses of short stature within this study

Short stature is defined as -2 SD of middle height within a coeval cohort. It is a collective term, a symptom, caused by diverse diagnoses. The seven most highly frequented diagnoses within this study are shown and described below in order of frequency.

1.3.1 Growth Hormone Deficiency

Human Growth Hormone (hGH, GH) is produced in the pituitary gland situated in the mid-line of cerebrum. Among other effects GH stimulates longitudinal growth of long bones. In case of isolated Growth Hormone Deficiency (GHD), patients normally have a proportionate growth lagging behind in height and weight according to their age cohort.

Stunted growth caused by GHD can be developed prematurely or during childhood. There are several etiologies. On the one hand there are organic causes (Secondary GHD, secGHD) with problems of the hypothalamo-hypophyseal system, congenital malformation, mid-line structure defects, perinatal problems or brain tumors. On the other hand there are genetically determined congenital forms of GHD. And finally, there are so-called idiopathic cases. (8)

1.3.2 Ullrich-Turner Syndrome

The Ullrich-Turner Syndrome (UTS), also called Turner's Syndrome, was first described in 1930 by Ullrich (50) and acknowledged as independent diagnosis in 1938 by Turner (49). It's caused by a numerical chromosome aberration in chromosome X, which can be missing partially or completely. Genome of girls with UTS is monosomy X (45,X) in approximately 50% of UTS girls, furthermore, mosaics (45,X/ 46,XX) and, less frequently, structural defects of

the X chromosome (46,X iso (Xq); 46,XXq- ; 46,XXp-; 46,X, r(X)) are described. The prevalence of Ullrich-Turner Syndrome is 1 in 2,000 newborn girls worldwide (25).

Phenotype main characteristics are, on the one hand, dysplastic ovaries causing absence of puberty and, on the other hand, stunted growth. Further phenotypic characteristics can be congenital heart failures, postnatal oedemas at backs of hands and feet, pterygium colli and other dysmorphism. Final height without GH-therapy is around 20 cm below the female average. Stunted growth in UTS is not a primary problem caused by GH deficiency. But if started early, GH treatment has positive effects in final adult height (26;31;40).

1.3.3 Children born small for gestational age

Children born small for gestational age (SGA) are defined by a birth weight which is two SD below the mean for other children with same gestational age. There is an incidence of SGA births at 2.3-10% (38). SGA can be caused by diseases of mother, influenced by medication during pregnancy, by drug, alcohol or tobacco consumption, multiple pregnancy or other growth development influencing factors, which often are unknown. Approximately 15% of children born SGA do not achieve with their genetic height potential. The majority of children born SGA are not GH deficient. Nevertheless GH treatment is known to improve growth rate and final height in the children born SGA who develop severe short stature (38).

1.3.4 Silver-Russell Syndrome

The Syndrome has been described and named by Silver (41) and Russell (37) in 1953 and 1954 respectively. The Silver-Russell Syndrome (SRS) has an estimated incidence of 1 in 3,000 to 10,000 live births (1), most cases with sporadic occurrence. 10% are caused by uniparental disomia of chromosome number 7 (1;10). The SRS has characteristic clinical symptoms like intrauterine

growth retardation and stunted birth growth, characteristic facial features like micrognathy and a triangular face, body asymmetries as clinodactyly, moreover urogenital dysmorphism and gastrointestinal abnormalities like reflux or esophagitis. Also, further growth is proportionally decelerated, on average -4.3 SD below the mean. Final adult height without growth hormone therapy is around 150 cm in boys and 140 cm in girls (54).

1.3.5 Noonan Syndrome

The Noonan is an autosomal dominant dysmorphic syndrome named after publication by Noonan (24). The incidence is quoted in 1:1,000 to 1:2,500 births (24). Clinical characteristics are: short stature, a downward eye slant, hypertelorism, a short neck, low set nipples, axillary's webbing and cardiac abnormalities. These characteristics are similar to those described by Ullrich and Turner (UTS) based on monosomy X, described above. Delayed bone age and final adult height below the 3rd percentile has been reported in up to 50% of female and nearly 40% of male patients (24).

1.3.6 Idiopathic Short Stature

Idiopathic short stature is defined as a height – 2 SD of national mean of height in absence of specific causative disorders (28).

1.3.7 SHOX Deficiency

SHOX stands for "Short statute HOmeboX-containing gene". It was discovered during research for genes inducing the growth retardation in Turner syndrome. SHOX is located on the distal end of the X and Y chromosomes at Xp22.3 and Xp11.3. This region does not undergo X inactivation, therefore healthy individuals express two copies of the SHOX gene. Several studies provided evidence that the haploinsufficience of the SHOX gene may be responsible for

the growth failure as well as typical skeletal abnormalities such as short fourth metacarpals, cubitus valgus, and Leri-Weil Dyschondrosteosis in Turner syndrome females (9;18;20;33). The incidence of SHOX deficiency is 1 in 1,000 in the total population (20).

The efficacy of GH treatment in subjects with SHOX deficiency was equivalent to that in subjects with UTS (33).

2 Methods

2.1 The BoneXpert Method

BoneXpert has been described in detail in (22;48). Its core technology is the ability to automatically and accurately reconstruct the bone borders using a generative model of each bone, i.e. it is able to generate artificial images of bones of all allowed shapes and appearances. BoneXpert has been trained on a large database of normal children supplemented by children from a pediatric endocrinology practice. Presented with a new, unseen hand, the computer invokes the following three computational steps:

In step 1, a deformable template of each bone adapts to the image and evaluates whether the observed bone morphology is acceptable.

In step 2, BoneXpert computes an intrinsic bone age value for each accepted bone. If a bone age value deviates more than 2.4 years from the average of the hand, it is deemed unacceptable. If less than 8 bones have been accepted, the image is rejected, and no bone age values are reported. Otherwise, bones with no accepted bone age are assigned the average bone age of the accepted bones.

In step 3, the intrinsic bone ages are transformed to agree on average with various bone age systems (GP, TW2, TW3, TW-Japan). The interval scale in the manual method has been replaced by a continuous scale in BoneXpert. The GP bone age is formed as the average over the 13 RUS bones with equal weights, and the result is formed non-linearly to agree with manual ratings performed by 5 different raters and with the GP atlas from 1959. This is the BA used in the present evaluation study, since our radiologists and pediatric endocrinologists use GP. The current version of BoneXpert covers the GP bone age range 2.5-17 for boys and 2-15 years for girls.

The whole process is fully automatic and takes approximately 10 seconds. The only input needed are the X-ray and the gender.

2.2 Data set

This study includes all the available rated X-rays of the left hand taken between 1973 and 2005 of patients with short stature from a single pediatric endocrinology center (University Children Hospital Tübingen). The original complete data consists of a total of 5865 roentgenograms of the left hand and wrist of 1152 patients (613 males, 3094 X-rays and 539 girls, 2771 X-rays), and the respective complete clinical, auxological and laboratory data for each patient. The range of validity of BoneXpert is 2.5-17 for boys and 2-15 years for girls, but in order to test the system at its extremes our dataset covered the manual bone age (manBA) range of 1.0 to 17.2 years in boys and 0.5 to 19 years in girls.

Since 1973 to date, for more than three decades of practice, a single radiologist (K. Drews) has assessed the bone age of 55% of the images in this study. A further 26% were assessed by P. Haber and 19% by other examiners (mainly M. B. Ranke).

78% of the films are original X-ray films, the rest are print-outs of digital images originally recorded on computed radiography (CR, using phosphor storage plates).

2.3 Analysis Method

The films were scanned and digitized in 300 dpi with 12 bits per pixel using a Vidar Diagnostic Pro Advantage scanner (Vidar, Hemdon, VA, USA) with software version TWAIN 5.2. Afterwards the films were reduced to 150 dpi, which is the resolution used by BoneXpert.

2.3.1 Process quality control

The images were scanned as one batch per patient in chronological order and simultaneously processed by BoneXpert. For quality control and error detection (e.g. wrong order of films), BoneXpert immediately produced a chart of average metacarpal length (metaL) by BoneXpert-BA (Figure 2.3.1).

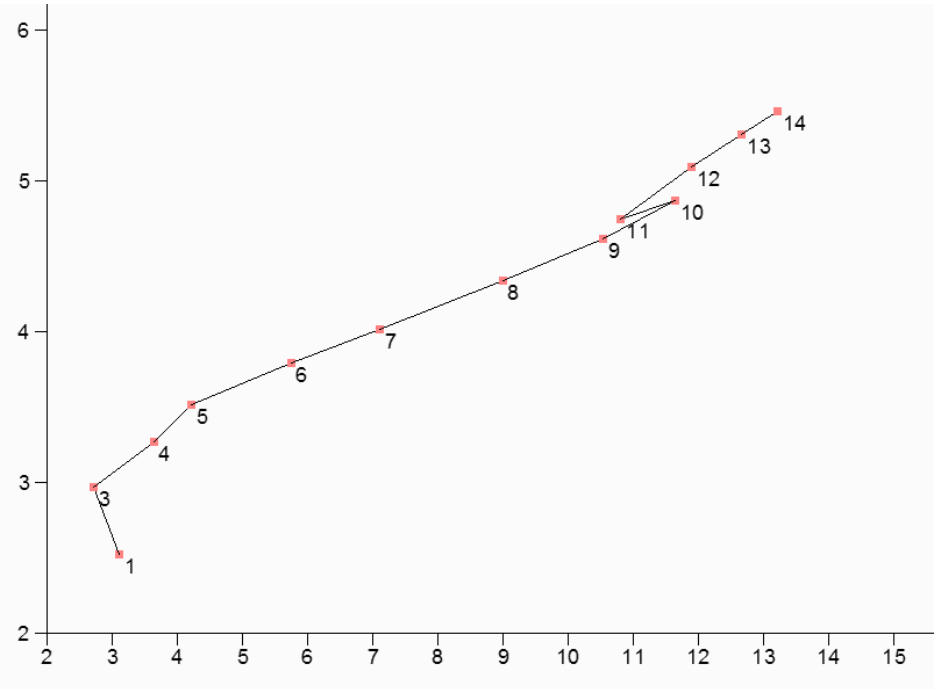


Figure 2.3.1: Exposure of error in image scanning.
Denoted figures indicate the order of scanned images.
The first two X-rays are out of the validity range for BoneXpert (No. 1 was computed incorrect, No. 2 analyzed unsuccessfully). No. 10 and 11 expose error in order of image scanning.

2.3.2 Coverage

Analysis out of BoneXpert's age range

The range of validity of BoneXpert is 2.5-17 for boys and 2-15 years for girls, but in order to test the system at its extremes our dataset covered the manual bone age (manBA) range of 1.0 to 17.2 years in boys and 0.5 to 19 years in girls. We compared the agreement with the original manBA and investigated the accuracy in- and outside BoneXpert's age range respectively.

Comparison: DICOM files versus scanned data of printed X-rays

BoneXpert was designed to analyze DICOM (Digital Imaging and Communications in Medicine; data format) images. In order to analyze a large number of images from our clinic we scanned the films from our archive. We were able to compare the performance of BoneXpert on the scanned files with that of the digital images because some of these films had been both printed out and kept as DICOMs. The printouts were scanned and processed by BoneXpert. So we had the possibility to analyze whether there was a difference in using BoneXpert in printed or digital X-rays.

2.3.3 Reliability

Study of deviation between manual and automatic determination

Both the BoneXpert ratings and the human ratings can potentially be erroneous, and a special analysis method is designed to accommodate this situation.

The manual rating is associated with observer variability, and could also be associated with more severe mistakes, in particular rating an X-ray using the wrong sex, which would give an error of approximately 2 years.

To define and quantify a bone age rating error, the true rating of a hand radiography is defined as the average of many manual ratings. In this work this true rating is approximated by the average of three new ratings (by MBR, DDM, and DD), denoted as new manual bone age (“New manBA”).

In order to avoid re-rating all the images, the following method is applied: Only images where the original manual and the automatic method deviate by more than a chosen threshold T are re-rated. The same method was applied in (52). In the present study the threshold T was defined as 2.05 years, twice the standard deviation for the normal bone age range per age bin.

The new ratings (New manBA) were performed blindly, i.e. knowing neither the age nor any of the previous manual or automatic ratings, so this reference can be considered an independent third rating that treats the two methods on an equal footing - and thus is used to determine which of the two methods is more reliable.

Study of deviation between semi-automatic and automatic determination

A further aim was to compare the values of BoneXpert as automatic system additional with a semi-automatic system. Some X-rays of untreated UTS girls have been analyzed in a former study. Then the computer-assisted skeletal age score system (CASAS, using the TW2-RUS method) was tested. We compared those values with the present BoneXpert ratings and the original manual GP ratings respectively.

2.3.4 Clinical Application in SGA children

Manual and automatic BA progression was analyzed in a subset of 52 x-rays from 26 SGA children for whom we had BA at start and after 12 months of GH treatment.

2.4 Statistical Methods

The core of this study is a comparison between manual and automatic rating. The accuracy of the BoneXpert method with regard to the manual method is defined as the standard deviation (SD) of the mean of the differences between the two methods. This is slightly larger than the root-mean-square (rms) deviation of a line fit, because the SD of the mean also includes the bias.

The agreement between automatic BoneXpert BA and manual New man BA was studied in terms of a Bland-Altman plot (11). It places the findings on an equal basis. For this purpose we plotted the difference between the two methods vs. the average of the two methods.

For all statistical analysis we used SAS jmp® 5.0.1 (SAS Institute Inc., SAS Campus Drive, Cary, NC, USA; http://www.jmp.com/).

3 Results

3.1 Coverage: Formal analysis of BoneXpert findings

3.1.1 Bone recognition – Characteristics of accepted images

The survey of all 5668 images from 1009 patients successfully analyzed by BoneXpert is shown in Table 3.1. For further details see Figure A.25 to Figure A.2.21 (in appendix).

Table 3.1: Patients' characteristics.
Survey of all 5668 images successfully analyzed by BoneXpert. No significant deviation between the diagnoses. For further details see Figures A.22a and A.22b (in the appendix).

Diagnosis	X-rays N (♀)	X-rays %	Patients N (♀)	Age range CA [y] (mean)	SD ♂ (♀)	rms ♂ (♀)
GHD	1654 (497)	29	309 (94)	1.4-20.1 (9.6)	0.76 (0.76)	0.70 (0.73)
UTS	1134	20	-- (173)	1.9-22.0 (12.0)	-- (0.79)	-- (0.72)
secGHD	639 (260)	11	86 (36)	2.1-22.4 (12.0)	0.78 (0.90)	0.73 (0.81)
SGA	436 (130)	8	99 (34)	1.6-18.7 (8.5)	0.81 (0.80)	0.76 (0.76)
SRS	165 (70)	3	32 (16)	2.6-17.4 (9.3)	0.92 (0.72)	0.77 (0.71)
Noonan	120 (41)	2	18 (4)	2.1-19.1 (11.6)	0.84 (0.66)	0.81 (0.62)
ISS	109 (53)	2	32 (14)	2.8-14.3 (8.3)	0.66 (0.67)	0.64 (0.64)
SHOX	17 (12)	>1	6 (5)	4.0-14.5 (9.1)	0.77 (0.40)	0.88 (0.42)
Other	1394 (517)	25	256 (96)	1.5-22.2 (10.0)	0.78 (0.75)	0.72 (0.71)
Collective	**5668 (3314)**	**100**	**1009 (471)**	**1.5-22.4 (10.3)**	**0.72 (0.74)**	**0.72 (0.73)**

The percentages of diagnoses associated with the films were: GHD (29%), UTS (20%), secondary GHD (11%), SGA (8%), SRS (3%), Noonan (2%), ISS (2%) and others (25%), such as Prader-Willi Syndrome, brain tumors, abnormalities of pituitary (like craniopharyngioma or panhypopituitarism), hyperparathyroidism, Crohn's disease, hypochondrodysplasia, 18p-Syndrome, Pierre-Robin-Syndrome.

2921 X-rays were taken while the children were treated with growth hormone. The other images were either taken before or after GH treatment. 298 patients did not undergo any GH treatment during the period of this study.

3.1.2 Bone recognition – Analysis of rejected images

BoneXpert either provides a bone age value or it rejects the image. The rejection is based on the number of successfully analyzed bones. BoneXpert requires at least 8 successful bones. If more than 5 bones are not processed the determination of all other bones is invalid. This yields 13 unprocessed bones and the operator gets a bone age value of “0”. The distribution of unprocessed bones is shown in Figure 3.1.1. BoneXpert assigned “0” to 197 (3%) out of 5865 X-rays.

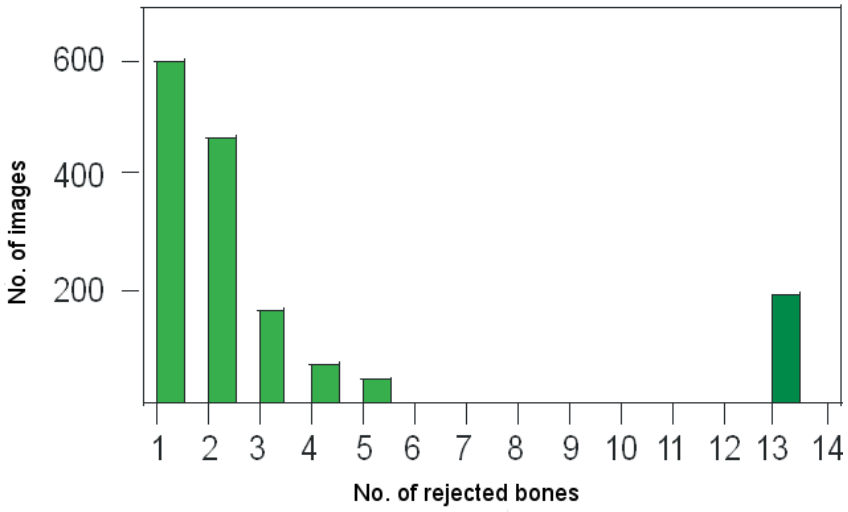


Figure 3.1.1: Number of unprocessed bones out of original data set with 5865 X-rays. In 25% less than 13 RUS bones within one X-ray were processed. While up to 5 unsuccessfully determined bones result a valid bone age, more than 6 rejected bones lead to unsuccessful analysis represented by 13 rejected bones.

The following reasons were found for the 197 exclusions:

- *bone age of subject was too young (out of validity range of BoneXpert)* (n=125), see Figure 3.1.2 below.
- *maturity was too advanced* (n=1)
- *the images were of bad quality* (n=51), examples shown in Figure A.2.1 (see appendix)
- *abnormal bone structure* (n=3), examples shown in Figure A.2.2 (see appendix)
- *Images were magnified or reduced printouts of digital images.* (n=6)
- *Images of very small hands but with a manual bone age within the validity range of BoneXpert.* These were analyzed successfully when magnified by 20%. Examples are shown in Figure A.2.3 (see appendix).

One example for an image that was rejected by BoneXpert because it was out of the validity range is represented here. A closer look at this X-ray reveals that it has the peculiarity of having a radius and carpal bone age of 3.5 y; a metacarpal and proximal phalanx bone age of 2 y and a middle and distal phalanx bone age of 1.25y, hence the regions BoneXpert analyzes were under the age of 2 years and therefore outside its scope. See Figure 3.1.2 below.

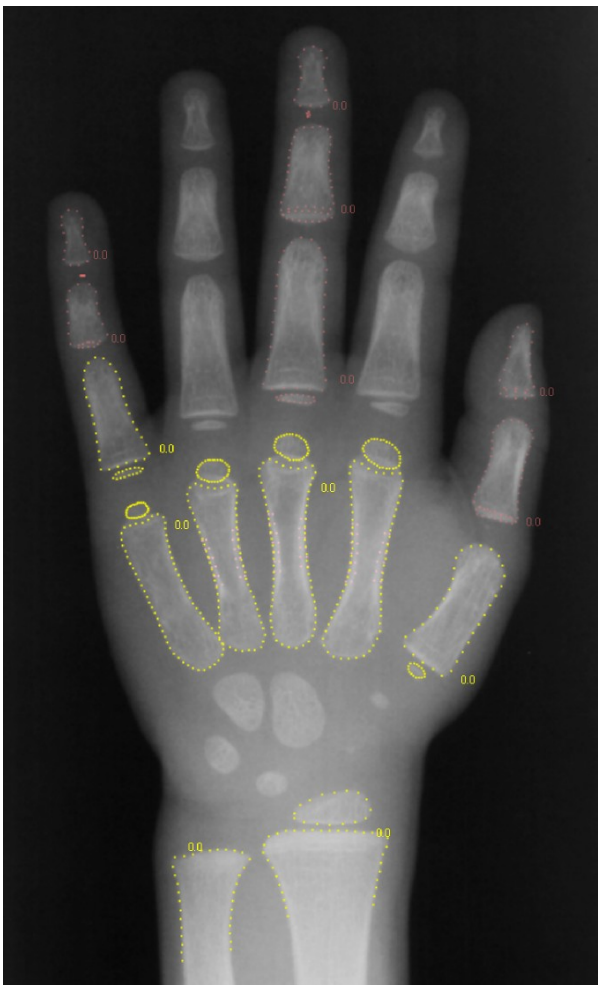


Figure 3.1.2: X-ray rejected by BoneXpert because serious maturity variances in different areas of interest (image number f1353-4). It has the peculiarity of having a radius and carpal BA of 3.5 y; a metacarpal and PP bone age of 2 y and a MP and DP BA of 1.25y. Diagnosis: GHD.

The subsequent analyses pertain to the 5668 X-rays accepted by BoneXpert. All 13 RUS bones were analyzed in 75% of the 5668 X-rays. Bones were rejected with the following frequency: MP5 (7.4%), ulna (6.6%), PP1 (5.8%), DP5 (5.2%), radius (4.9%), DP1 (4.6%), PP5 (2%) meta1 (1.1%), DP3 (1.1%), MP3 (0.7%), meta5 (0.6%), meta3 (0.5%), PP3 (0.4%).

3.1.3 Analysis of outlier

BoneXpert BA deviated from the manual finding by more than 2.05 years in 92 X-rays (1.6%). These cases are of interest, because they can expose errors in the original manual ratings or errors in BoneXpert. They were re-rated by three raters (DD, DDM and MBR) not knowing the chronological age, the manual or BoneXpert rating (Table A.1, see appendix). The average of these three blind ratings is used as reference value ("New manBA").

For these 92 outliers, the original manual rating was differed from New manBA by 0.27 years mean (1.6 y SD). 23 ratings still differed by more than >2.05 years after blind re-rating.

The BoneXpert values for the 92 outliers differed from the new manual bone age (New manBA) by 0.25 years (1.03 y) and all except three were within the 2.05 year limit after this re-rating (compare Figure 3.1.3 and Figure 3.1.4). We directly compared the respective differences of the automatic BA and the original rating to New manBA. This showed the original manual rating to differ more from New manBA in 61% (n=56).

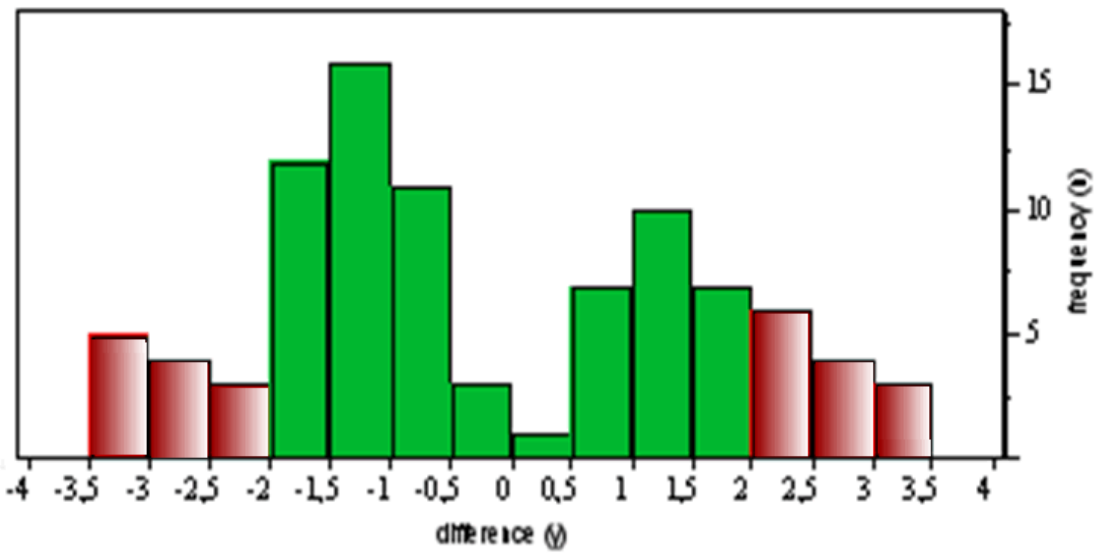


Figure 3.1.3: Distribution of difference between ratings before blind re-rating.
Difference of BXpBA minus origin manual BA (manBA), values out of T are highlighted.

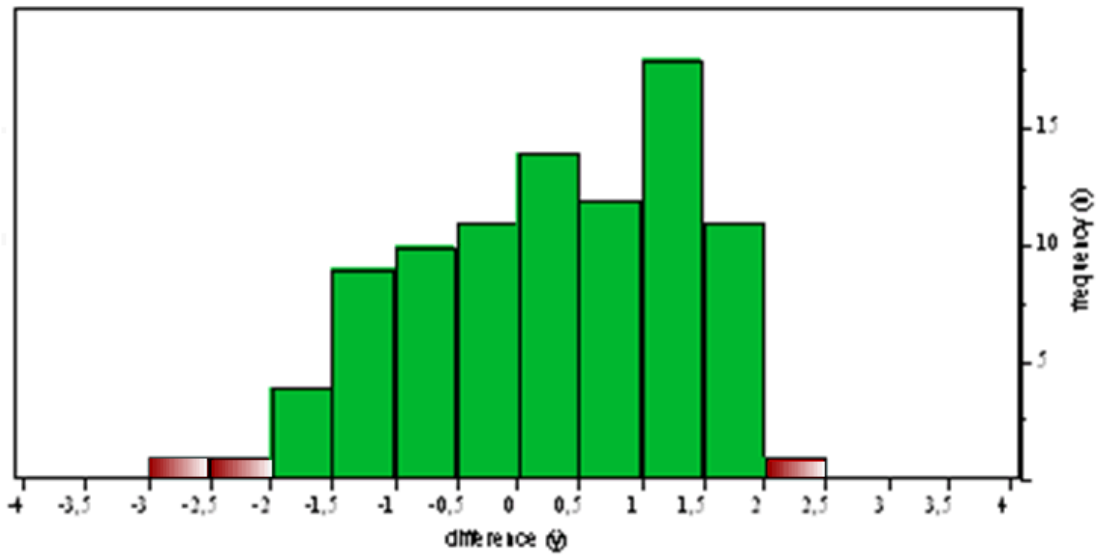


Figure 3.1.4 Distribution of difference between ratings after blind re-rating.
Difference BXpBA minus New manBA, values out of T are highlighted.

When the images were reviewed to find the source of the differences between the manual and the bone expert reading, two main factors were found:

- *Bias by the original manual observer having known the CA*

Re-rating without knowledge of the CA leads to results more similar to those of BoneXpert (Figure 3.1.4 above).

- *Placing the emphasis on the carpals*

The original manual bone age rating corresponded to the carpals whereas the BoneXpert rating corresponded to the rest of the hands. BoneXpert disregards the carpals as recommend by Tanner and Whitehouse (42) (Figure 3.1.5 and Figure 3.1.6 below).

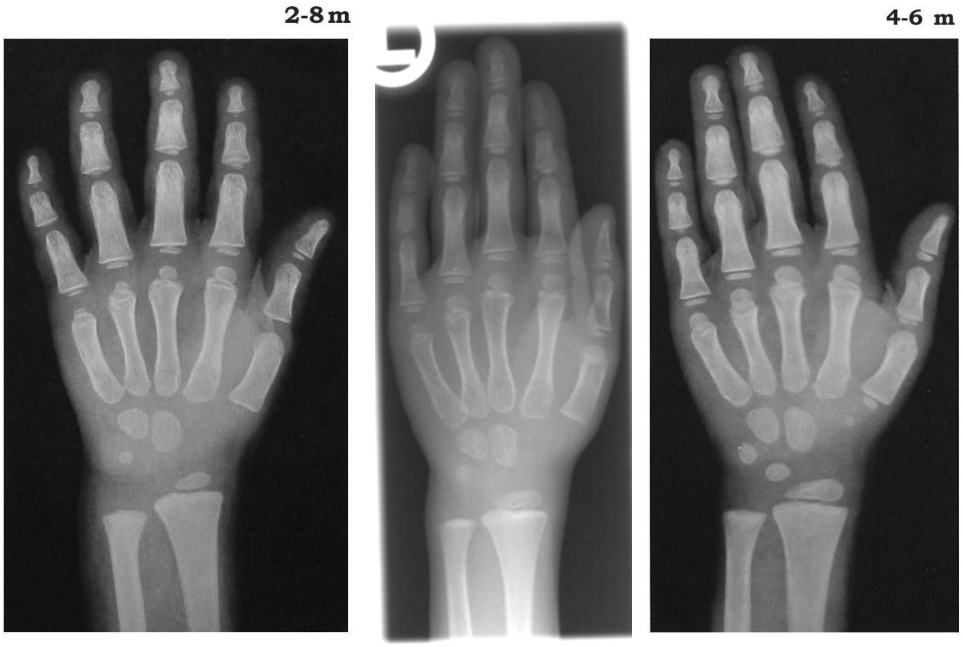


Figure 3.1.5: Carpals may be misleading (image number m1624-3).
Original manual rating was 2.7y, being very influenced by the carpals; BoneXpert, disregarding the carpals rated 4.8y. Manual re-raters disregarding the carpals rated between 3 and 4.5y (average 4.2y). The boy was 6.8y old.
Left: reference picture from the atlas of GP for a 2 years and 8 months old boy.
Middle: patient's x-ray.
Right: reference picture from the atlas of GP for a 4 years and 6 months old boy.

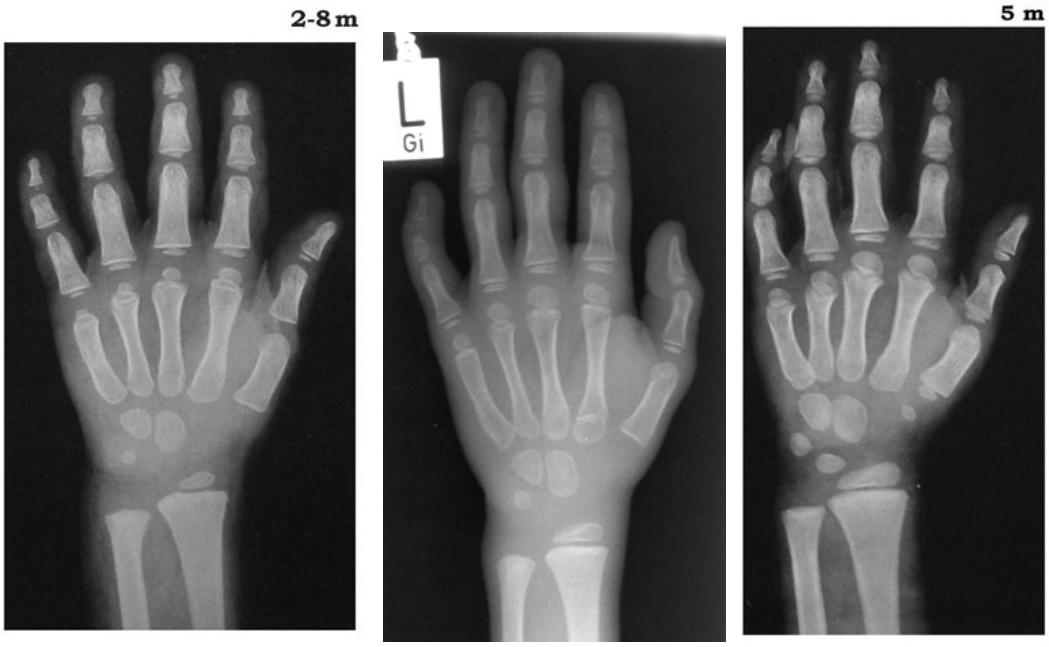


Figure 3.1.6: Carpals may be misleading (image number m1613-2).
Original manual rating was also 2.7y, influenced by the carpals; BoneXpert, disregarding the carpals rated 4.7y. Manual re-rating disregarding the carpals was 5.0y. The boy was 6.8y old,too.
Left: reference picture from the atlas of GP for a 2 years and 8 months old boy.
Middle: patient's x-ray.
Right: reference picture from the atlas of GP for a 5 years old boy.

3.1.4 Accuracy out of BoneXpert's intended bone age range

In order to test the system beyond its age range our dataset covered the manual bone age range of 1.0 to 17.2 years in boys and 0.5 to 19 years in girls. Although 125 of the images with BA younger than BoneXpert's BA range and one of the images with BA older than BoneXpert's BA range had been rejected, BoneXpert accepted 119 images that were beyond its age range.

A total of 77 X-rays were from boys with manual BA less than 2.5 years. Thereof two X-rays had a difference of more than 2.05 years (T) between BoneXpert results and manual BA. After the blind manual re-rating ("New manBA", see above) both were within the limit value of 2.05 years (1010-1 and 1209-1, Table A.1, see appendix). 9 X-rays of boys with manual BA above 17 years were included. One of them showed a difference in its automatic rating which is outside of T (408-7, see Table A.1 in appendix). A closer look at this x-ray showed a completed skeletal maturation (Figure A.2.4, see appendix). In the subsequent BoneXpert version this error was eliminated and the X-ray is now rejected as invalid.

Among the girls, 22 X-rays had a manual BA of less than two years. All of them are within T. Eleven images had a manual BA greater than 15 years. One of them showed a difference between manual and automatic rating greater than T (1235-15, see Table A.1 in appendix). After re-rating it was within the 2.05 year limit.

3.1.5 Comparing the analysis of digital and printed X-rays

804 of the scanned files were print-outs of digital images. So we were able to compare the analyses of BoneXpert on scanned films with that of the digital images (DICOM).

BoneXpert is intended to be used with images in 100% size and a 2% deviation is considered acceptable.

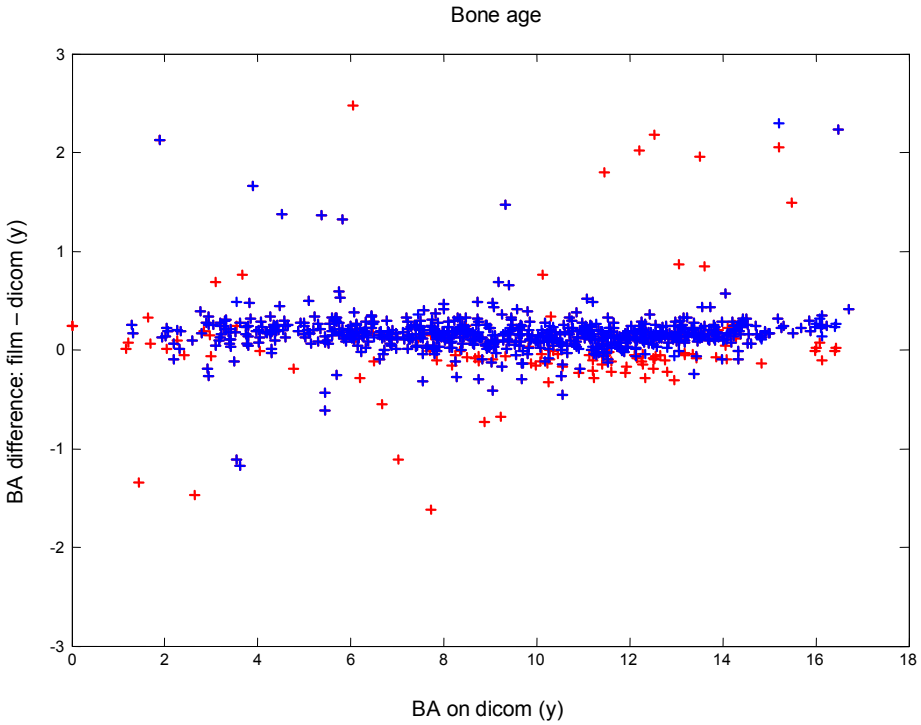


Figure 3.1.7: The bone age difference between digital and printed X-rays (n=804).
The red dots indicate images with abnormal amplification.

Figure 3.1.7 shows the bone age difference between DICOMs and printed X-rays. The agreement for the bulk of the data was good (2.8% outliers), and there was a median difference of 0.1 years, i.e. hardcopies gave 0.1 years smaller bone age. This is below the precision error in a BoneXpert determination, which is 0.2 years (23). The correlation coefficient between the BA values from the DICOMs and the films was 0.999; $P < 0.0001$.

The red markings designate images that were magnified or reduced when printed and showed fairly large bone age differences.

3.2 Reliability of BoneXpert

3.2.1 Accuracy by comparison with manual bone age

Figure 3.2.1 and Figure 3.2.3 show the relation between automatic BoneXpert BA and Reference BA (original manBA values replaced with New manBA, if given) in boys and girls, respectively. The squared correlations were R2 = 0.967 and 0.957 for boys and girls respectively, the standard deviation 0.73 years [0.66;0.84] 95% CI (Figure 3.2.5). The mean difference was 0.09 years. The agreement between automatic BXpBA and RefBA was studied in terms of a Bland-Altman plot (Figure 3.2.2 and Figure 3.2.4), where the difference between two measurements is plotted versus the average of the two measurements. It was 0.16 + 1.0 (y) for boys and 0.32 + 0.97 (y) for girls.

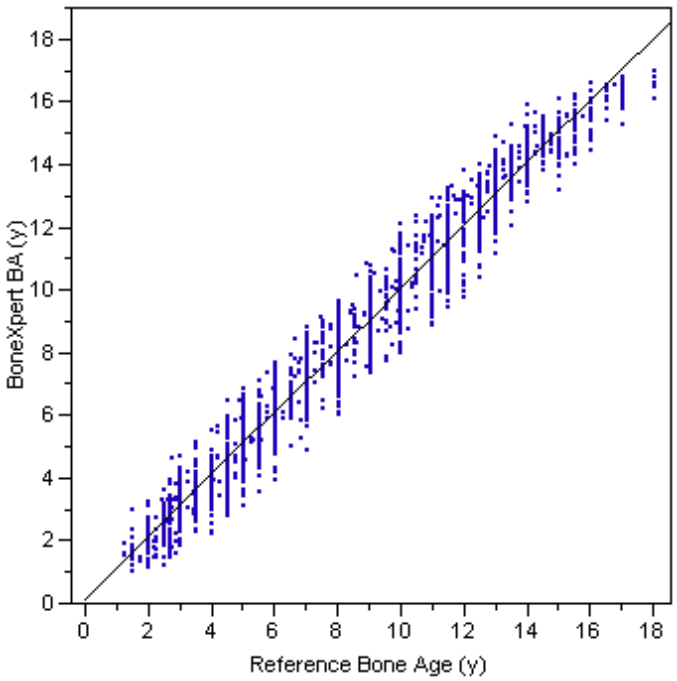


Figure 3.2.1: Relation between BXpBA and RefBA in boys
rms: 0.72 (n 2354), p < 0.001
BoneXpert BA (y) = 0.16 + 1.0 Reference BA (y)

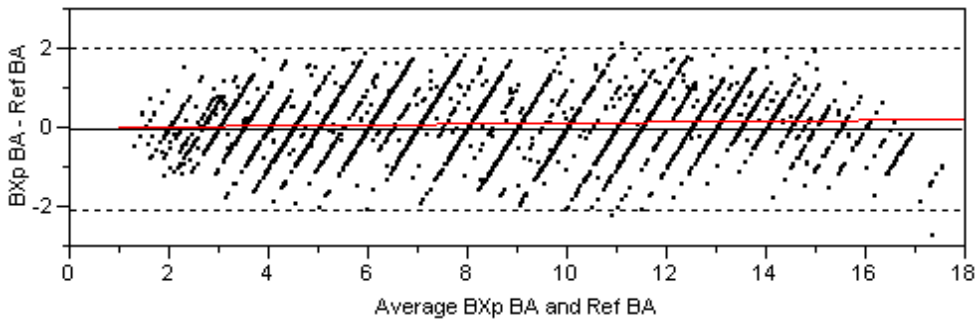


Figure 3.2.2: Difference between BXpBA and RefBA, boys.
Difference plotted vs. the average of the two measurements.

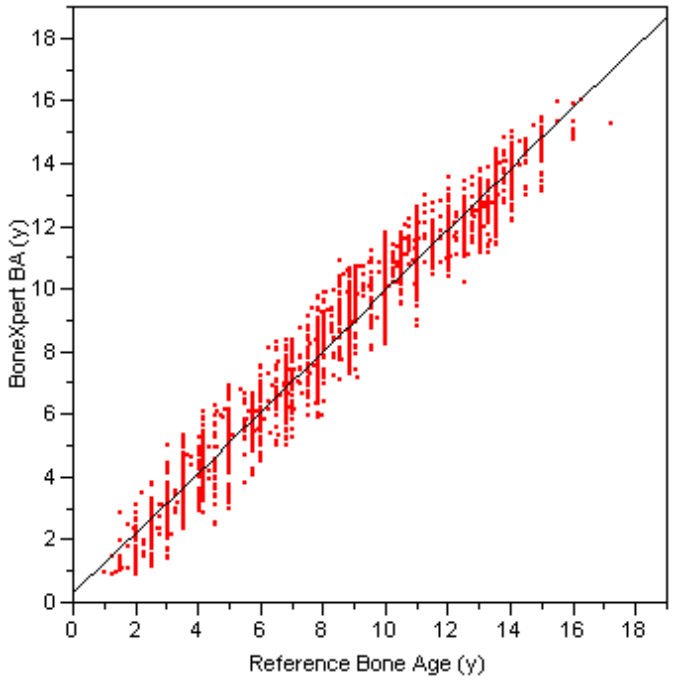


Figure 3.2.3: Relation between BXpBA and RefBA in girls
rms: 0.73 (n 3314), p < 0.001
BoneXpert BA (y) = 0.32 + 0.97 Reference Bone Age (y)

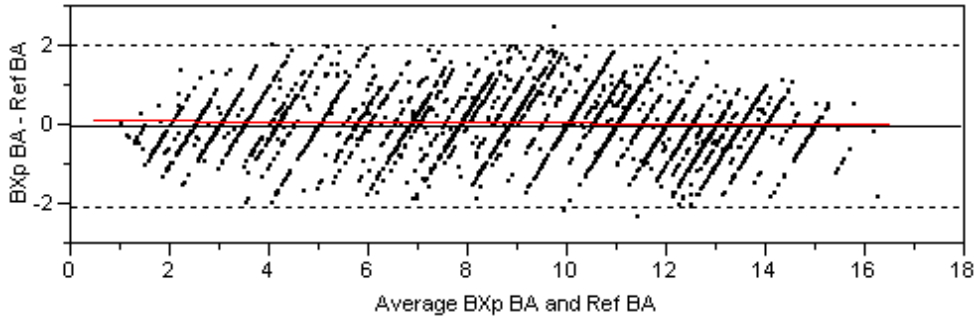


Figure 3.2.4: Difference between BXpBA and RefBA, girls.
Difference plotted vs. the average of the two measurements.

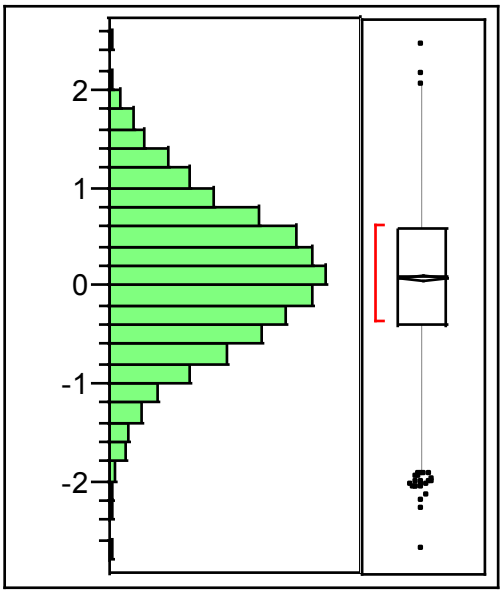


Figure 3.2.5: Distribution of difference between RefBA and autBA, both sexes.
Mean: 0.09, SD 0.73, Std ErrorMean 0.0097, upper 95% Mean 0.11, lower 95% Mean 0.07, N 5668

The ends of the box are the 25th and 75th quantiles. The middle horizontal line identifies the media-sample value and the vertical height of the diamond indicates the 95% CI. The bracket along the box identifies the shortest half, which is the most dense 50% of the observations.

The agreement was not significantly different among the genders and diagnoses. For further details see Figures A.5 to A.21 and Figures A.22a and A.22b (in the appendix).

3.2.2 Accuracy by comparison with manual and semi-automatic (CASAS) bone age determination in patients with Turner's Syndrome

A total of 150 X-rays from girls with untreated Ullrich-Turner Syndrome confirmed by chromosome analysis (age range 0.84 to 18.07 years; mean manBA 8.90 years, mean BA retardation with regard to CA 1.87 years) were rated by BoneXpert. The automatic rating was compared to the CASAS rating and to the original GP rating (manBA). Note that the manual BA in this preliminary version is the raw BA, and not yet the one corrected by Deusch/Ranke/Martin (New manBA), so the end result will be slightly more in accordance with BoneXpert.

BoneXpert BA deviated by more than 1.5 years from the operator BA for 9 X-rays. CASAS-BA deviated by more than 1.5 years from the operator BA for 55 X-rays.

In these 150 X-rays the mean difference between BoneXpert and manual rating was 0.28 years (n.s.), the accuracy (SD of the difference) 0.80 years [0.69; 0.88] 95% CI, bias 0.04 y/y (n.s.), see Figure 3.2.6 below, such as Figure A.2.23 and Figure A.2.24 in the appendix.

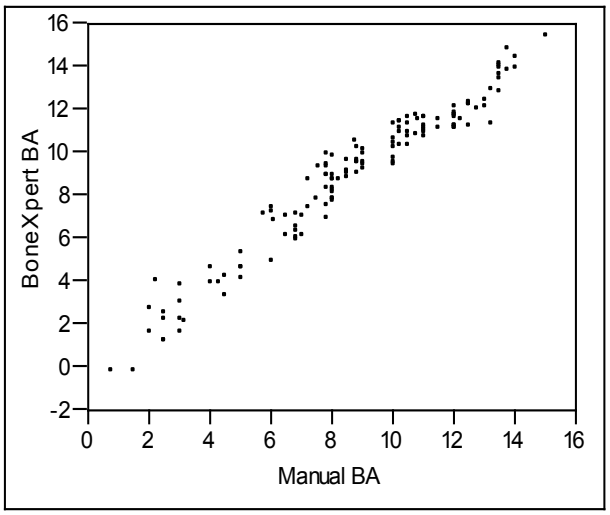


Figure 3.2.6: BoneXpert BA (y) vs. manual BA (y).

The mean difference between CASAS and manual rating was 1.17 years (n.s.), the accuracy (SD of the difference) was 0.95 years [0.85; 1.08] 95% CI, bias 0.1 y/y (n.s.), see Figure 3.2.7 below, such as Figure A.2.25 and Figure A.2.26 in the appendix.

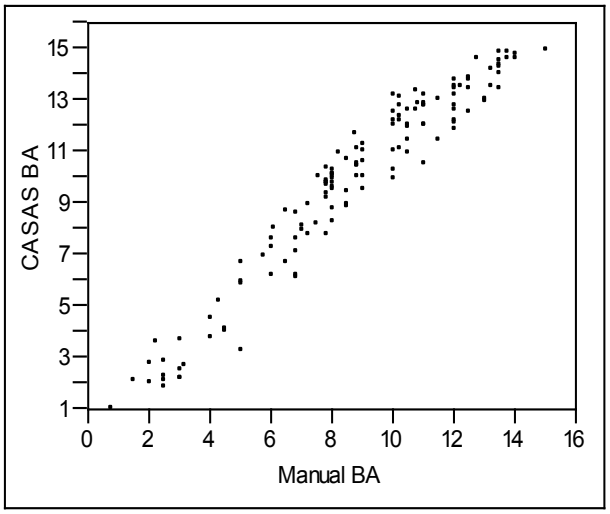


Figure 3.2.7: CASAS BA (y) vs. manual BA (y).

The mean difference between CASAS and BoneXpert rating was 0.87 years (n.s.), the accuracy (SD of the difference) was 0.84 years [0.75; 0.94] 95% CI; bias 0.07y/y (n.s.), see Figure 3.2.8.

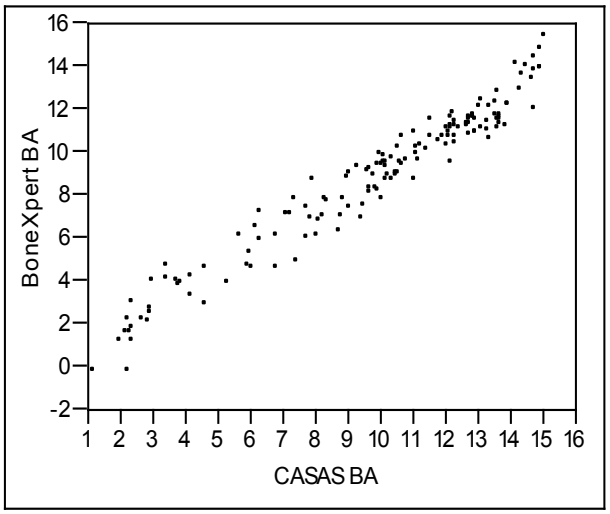


Figure 3.2.8: BoneXpert BA (y) vs. CASAS BA (y).

BoneXpert's GP performance was significantly nearer to the manual GP rating than that of CASAS, while the standard deviations of the differences between the three methods were similar and the biases negligible (see also appendix, Figures A.2.23 to A.2.26).

3.3 Clinical Application of BoneXpert

3.3.1 Bone age during growth hormone treatment in children born SGA

BA ratings between automatic and manual BA of 52 X-rays from 26 children born SGA were compared before and during GH treatment with a dose of 49.8±13.04 µg/kg/d (mean ± SD). Comparing the values at starting point 0 and after 12 months the original manual BA advanced by 1.06±0.62 years (difference between CA and manBA is 0.01 years, P = 0.88) and BXpBA advanced by 1.41±0.60 years (difference between CA and BXpBA is 0.36 years, P = 0.005; difference between manBA and BXpBA is 0.34 years, P = 0.03), see Table 3.2 and Figure 3.3.11.

Table 3.2: Values at the beginning of GH treatment (0/12), after 12 months (12/12) and their difference (Delta △) in children born SGA.
Mean and SD (in brackets), respectively.

	Mean (SD) 0/12	**Mean (SD) 12/12**	**Delta△**
Chronological Age (y)	6,7 (2,8)	7,8 (2,8)	1,0
BoneXpert bone age (y)	5,2 (2,8)	6,6 (2,9)	1,4
Manual bone age (y)	5,3 (3,0)	6,4 (3,0)	1,1
Height SDS	- 3,5 (1,0)	- 2,7 (1,1)	0,8
Weight SDS	- 2,6 (0,8)	- 2,2 (0,9)	0,5
GH dose after 6 months (µg/kg/d)	49,82 (13,04)		
IGF-1 SDS	- 1,6 (1,7)	0,6 (1,7)	1,9
IGF-BP3 SDS	-0,7 (1,0)	0,2 (1,2)	0,8

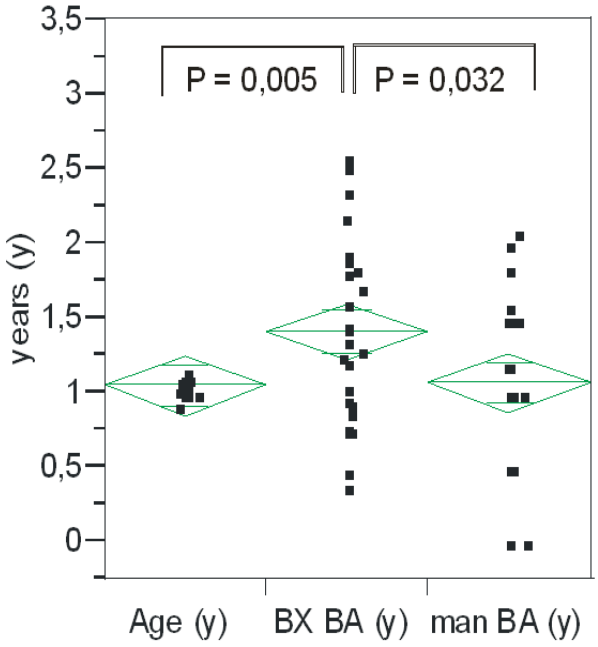


Figure 3.3.1: Change in CA [(Age (y)], automatic BoneXpert rating [BXpBA (y)] and manual GP BA [manBA (y)].
BXpBA shows an acceleration of BA by 1.4 years while manBA doesn't differ significantly from change in CA. The diamonds portray the mean (middle horizontal line) and 95% CI (vertical diamond height).

4 Discussion

The determination of the skeletal maturity is a basic component of the diagnostic and therapeutic measures in children presented with short stature. It is important to have precise, accurate measurements because the determined bone age of a child may influence therapeutic decisions.

However, the intra- and inter-individual variations of manual bone age assessment in clinical and research practice impose a major handicap on its usefulness in clinical practice and research. Time-consuming semiautomatic methods have not found their way into clinical practice.

In this study we surveyed BoneXpert as an automatic method in its coverage, reliability and application in clinical practice.

4.1 Coverage

197 (3%) images of the initial set were rejected. In all cases the reason of rejection was traceable, such as bad image quality, malformation or immature hands. It is acceptable – even desirable – that bad X-rays and hands with abnormal bone structure are rejected by BoneXpert. Then the radiograph should be retaken or the anatomy is abnormal so that a radiologist should evaluate the X-ray. We found no rejection of normal hands.

For the further survey we had 5668 X-rays from 1009 patients.

The re-ratings of 92 images featuring a difference beyond value limit T (2 SD, >2.05 years) between BoneXpert and the observer fell to the advantage of BoneXpert in 61%, indicating that both BoneXpert and the manual raters make

significant errors, but BoneXpert to a lesser extent. After blind re-rating 89 bone age values (97%) of these 92 images were within the limit T. Our original observers were aware of the chronological age, diagnosis and previous ratings when rating the images, which may have led to a bias. Berst et al (7) showed that, depending on the examiner, knowing the chronological age can lead to 18-32% less images being rated as pathological (defined as being outside two standard deviations of chronological age). The Dutch Erasmus study (51) draws upon this hypothesis, too. Our re-ratings show that without knowing chronological age, manBA is more similar to BXpBA. Eliminating this bias might be an additional advantage of BoneXpert and other computerized methods.

Looking at BoneXpert's precision out of its intended bone age range, approximately 3% had a large difference between manBA and BXpBA. But after blind re-rating the manual rating approaches the automatic rating so that all lied within the value limit T. Again, this shows that manual observers are biased knowing the chronological age of the patient and it confirms the advantage of an independent automatic determination. Although BoneXpert shows a good coverage even out of its age range, users should follow the BoneXpert's operator instructions and consider BoneXpert's age range.

BoneXpert was designed to read digital data formats, but is also able to rate printed X-rays. The difference between digital and printed X-rays was insignificant. Therefore BoneXpert is useful on the one hand in clinical praxis of current digital age and on the other hand for research and survey of printed X-rays that have been stored for decades.

4.2 Reliability

The core analysis of this study was the comparison between manual and automatic ratings. BoneXpert was able to determine bone age in various

diagnoses causing short stature. Overall there is a good agreement between manual and automatic bone age determination.

There was no significant deviation between sex and diagnoses.

Comparing BoneXpert with our manual Greulich-Pyle ratings yields an accuracy of 0.52 (0.73/√2) years. The bias (BoneXpert readings were on average 0.24 years behind the manual readings) and the rms deviation (0.73 y) between the automatic and the manual readings are comparable with the inter-observer differences found in the literature (2;4;7;12;13;16;17;19;34-36;52;55).

Compared with the semiautomatic method CASAS (based on Tanner-Whitehouse), BoneXpert's Greulich-Pyle results are significantly nearer to the manual Greulich-Pyle rating than those to CASAS. However, the standard deviations of the differences between the three methods were similar and the biases negligible. These differences are judged to be mainly due to the known intrinsic differences between the Tanner-Whitehouse and the Greulich-Pyle methods (16;17;52).

4.3 Clinical Application

On average BoneXpert’s bone age of the children in our set born SGA was decelerated by about 1.5 years according to chronological age. Via BoneXpert we demonstrated that bone age progresses 1.4 years in first year of growth hormone treatment in children with short stature born SGA, in accordance with results of Arends et al (5). Looking at the manual ratings there were no corresponding significant effects of growth hormone on bone maturation during this first year. This could be interpreted as result of inaccuracy of manual reading within clinical routine. In addition, the radiologist knows – contrary to BoneXpert – that one year passed since the last analysis. On the one hand our

radiologists resort to the last findings; on the other hand they do not know if and when growth hormone treatment began. So they have no reason to estimate that bone maturation processed by more than one year. Those influences and also intra- and inter-individual differences in ratings may be reasons why bone age did not figure in multicentric height prediction models (27;29;30;32).

Strictly controlled studies found relations between final height and difference between bone age and chronological age at the beginning of treatment (14). It would be interesting to apply BoneXpert to studies describing no significant progress of bone age in relation to chronological age while growth hormone treatment (e.g. (6)).

5 Conclusion

In this study we surveyed BoneXpert as an automatic method in its coverage, reliability and validation in clinical practice.

BoneXpert was validated on a very wide variety of images from different kinds of films and with different types of post-processing. BoneXpert's ability to process virtually all images automatically, to avoid errors, and to obtain good agreement with an operator suggests that the method could be efficient and reliable in the spectrum of short children that are presented to endocrine clinics.

In view of our comparison of the manual and BoneXpert ratings of children treated with growth hormone we suggest that all pharmacological studies of hormones and substances that may affect skeletal maturity should include BoneXpert or an automatic program of comparable performance in their study protocol.

A. Appendix

A.1 Tables

Table A.1: The 96 images with >2.05 deviation between manual and automatic BA.

"manBA" is the original manual rating and "BXpBA" is the automatic rating. "New manBA" is the average of three new, blind re-ratings.
In the column with bone age deviations from the New manBA, deviations larger than 2.0 are highlighted in light grey. The dark grey highlights, in the column "manBA", mark where manual bone age is out of BoneXpert's range of validity.
The last column indicates the cases where BoneXpert has larger deviation from the reference (New manBA) than the original manual rating.

id	visit	m = 1 f = 2	Bone age values			Bone age differences		BXp worse
			manBA	BXpBA	New manBA	manBA *minus* New manBA	BXpBA *minus* New manBA	
51	1	2	12	14.08	13.67	-1.67	0.41	
53	4	2	8.75	10.95	9.78	-1.03	1.18	X
53	2	2	8	10.9	9.44	-1.44	1.46	X
63	5	2	6.5	9.72	10	-3.5	-0.28	
72	5	2	10.5	13.05	13.58	-3.08	-0.53	
77	3	2	5.5	9.43	8.13	-2.63	1.3	
80	7	2	8	10.17	9.44	-1.44	0.73	
82	7	2	15	12.72	14	1	-1.28	X
97	3	2	7.5	9.98	8.11	-0.61	1.87	X
123	1	2	5	7.09	6.72	-1.72	0.37	
133	9	2	5	7.06	6.86	-1.86	0.2	
133	7	2	3.75	6.18	5.58	-1.83	0.59	
163	8	2	10	7.26	9.08	0.92	-1.83	X
163	9	2	11	8.55	8.3	2.7	0.25	
183	4	2	9	11.13	10	-1	1.13	X
183	1	2	7.33	10.01	8.31	-0.98	1.7	X
191	5	2	4	6.22	4.53	-0.53	1.69	X
191	7	2	7	9.36	7.94	-0.94	1.42	X
191	6	2	5.75	8.13	6.61	-0.86	1.52	X
408	7	1	19	16.01	18.67	0.33	-2.66	X
562	10	2	13.25	10.66	12	1.25	-1.34	X
562	9	2	13.25	11.16	12.67	0.58	-1.51	X
563	3	1	11	8.89	9.67	1.33	-0.78	
573	1	2	5.75	3.41	4.17	1.58	-0.76	
607	15	1	7	9.32	8.42	-1.42	0.9	
630	4	1	11.5	9.01	9.83	1.67	-0.82	
631	1	1	11.5	8.87	10.25	1.25	-1.38	X
659	5	1	11	7.84	9.5	1.5	-1.66	X

Table A.1 continued

659	6	1	11.5	9.04	9.5	2	-0.46	
667	2	1	9	11.33	9.42	-0.42	1.92	X
667	1	1	7	9.38	8.17	-1.17	1.21	X
667	3	1	10.5	12.96	11.67	-1.17	1.29	X
690	5	2	9	5.94	5.53	3.47	0.41	
690	10	2	14	11.63	13.33	0.67	-1.7	X
711	3	1	3	5.98	4.92	-1.92	1.06	
717	4	1	8	10.32	10.33	-2.33	-0.02	
722	2	2	8.33	6.17	5.84	2.49	0.32	
752	1	2	9	6.8	7.53	1.47	-0.73	
810	3	2	14	11.14	12	2	-0.86	
810	2	2	14	11.77	13	1	-1.23	X
840	2	1	8	10.12	9.83	-1.83	0.29	
845	4	2	11	8.88	8.83	2.17	0.05	
1007	12	1	9.5	11.65	11.17	-1.67	0.48	
1007	9	1	7	9.35	8.42	-1.42	0.93	
1007	11	1	9	11.47	10.25	-1.25	1.22	
1007	10	1	8	10.59	9.17	-1.17	1.43	X
1010	1	1	0.5	3.4	2.58	-2.08	0.81	
1025	6	1	11	13.05	11.67	-0.67	1.38	X
1027	7	1	12.5	9.18	9.42	3.08	-0.24	
1034	6	1	8	10.57	8.58	-0.58	1.99	X
1040	10	1	8.5	11.47	11.33	-2.83	0.14	
1042	8	1	7.83	11.8	10.92	-3.08	0.88	
1043	12	2	15.5	13.25	13	2.5	0.25	
1050	6	1	4.5	6.56	5.25	-0.75	1.31	X
1106	5	2	10	7.8	8.56	1.44	-0.76	
1112	9	1	10	12.54	11.75	-1.75	0.79	
1118	4	1	7.5	9.58	8.25	-0.75	1.33	X
1156	8	2	13.5	10.28	12.53	0.97	-2.25	X
1173	6	1	11.5	9.41	9.92	1.58	-0.51	
1201	7	1	10	12.37	11.17	-1.17	1.2	X
1206	8	1	6	8.09	8	-2	0.09	
1209	1	1	1	4.05	4.08	-3.08	-0.03	
1214	1	1	10	7.76	6.75	3.25	1.01	
1235	15	2	17	14.83	16.33	0.67	-1.5	X
1250	6	1	7	9.06	7.5	-0.5	1.56	X
1262	7	2	7.83	10.03	9.94	-2.11	0.08	
1359	8	1	11	13.1	12.08	-1.08	1.02	
1360	2	2	8.83	6.73	7.81	1.03	-1.07	X
1385	3	1	9	11.85	12	-3	-0.15	
1397	4	1	7	4.63	5.08	1.92	-0.45	
1403	1	1	11.5	9.06	8.67	2.83	0.39	
1600	1	1	11	8.44	8.67	2.33	-0.23	
1609	3	1	9	11.23	10.67	-1.67	0.56	
1613	2	1	2.67	4.75	4.17	-1.5	0.58	
1618	3	1	8	5.92	6.83	1.17	-0.91	
1624	3	1	2.67	4.77	4.2	-1.53	0.57	
1624	5	1	6	8.4	7.33	-1.33	1.07	

Table A.1 continued

					3.5	0.5	1.59	X
1702	3	2	13.33	11.14	12.08	1.25	-0.94	
1702	4	2	13.42	11.23	12.5	0.92	-1.27	X
1850	1	1	11.5	8.46	9.67	1.83	-1.2	
1860	2	1	12.5	10.15	10.33	2.17	-0.18	
1883	3	1	6	8.51	6.67	-0.67	1.85	X
1889	1	2	2.5	5.6	5.93	-3.43	-0.33	
1919	1	1	12	9.84	10.5	1.5	0.66	
4030	13	1	6.5	9.24	7.58	-1.08	1.65	X
4030	14	1	7.5	10.91	9.17	-1.67	1.74	X
4046	5	2	3.25	5.51	5.77	-2.52	-0.25	
33211	6	2	8.5	10.62	9.77	-1.27	0.86	
33212	5	1	10	12.27	10.25	-0.25	2.02	X
33212	3	1	9	11.46	10.42	-1.42	1.05	
100088	3	1	11.5	7.62	8.92	2.58	-1.3	

> 2.05 ("T"; 2 SD)

out of range of validity

A.2 Figures

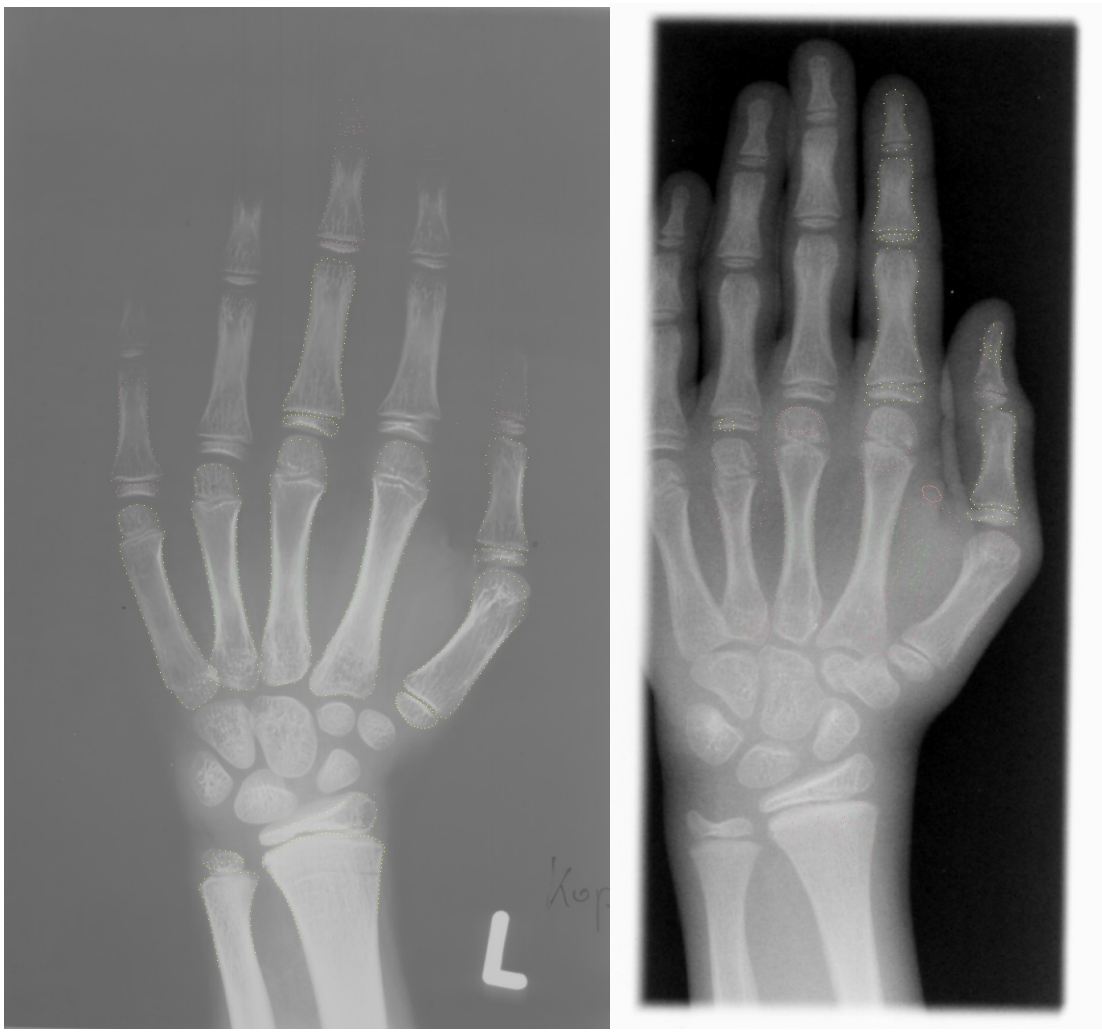


Figure A.2.1: Bad X-ray quality rejected by BoneXpert.

Left: m1034-7 poor exposure, copy of original X-ray.
Right: f1270-3 metacarpal 5 is not fully inside the image.

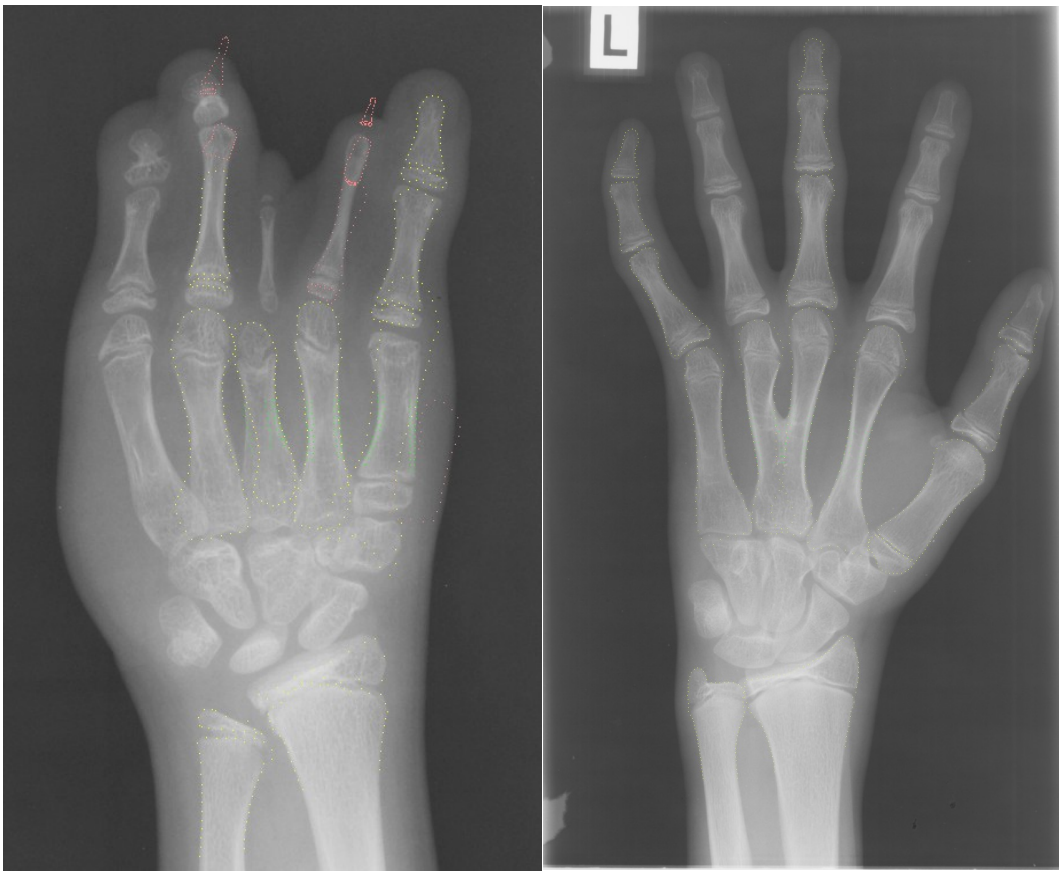


Figure A.2.2: Abnormal bone structure rejected by BoneXpert.

Left: f50-3: Girl with Poland Syndrome. The syndrome occurs with unilateral synbrachydactyly and aplasia of the ipsilateral perctoral muscle.
Right: m4008-17: Boy with isolated GHD in hypolplasia of pituitary plus syndactyly of hands and feet.

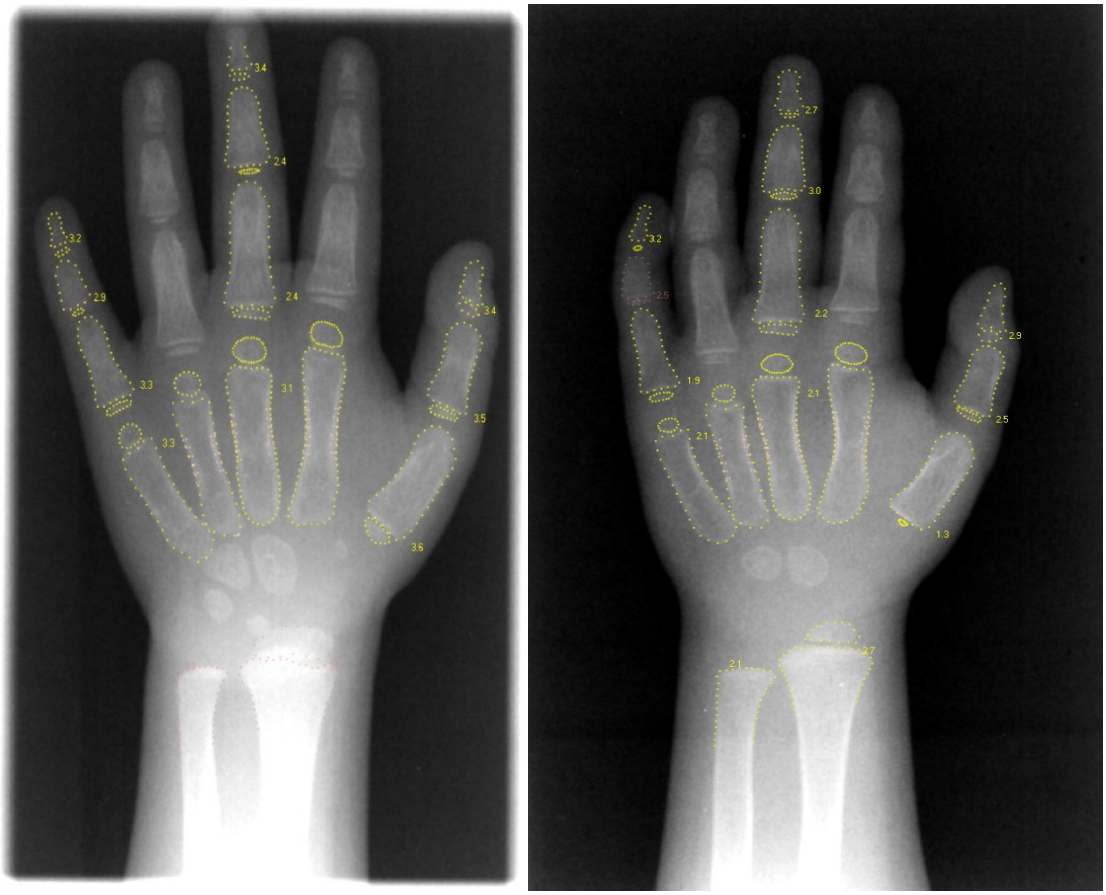


Figure A.2.3: Successful analyses of rejected images due to immature hands after magnification (20%).
Left: f3321-11 (Noonan; manBA 3.5y). After magnification autBA was 3.11y.
Right: f872-1 poor contrast between bone and soft tissue (GHD after chemotherapy/radiation, manBA 2.0y; in addition manBA is close to validity range of BXp). After magnification autBA was 2.48y.

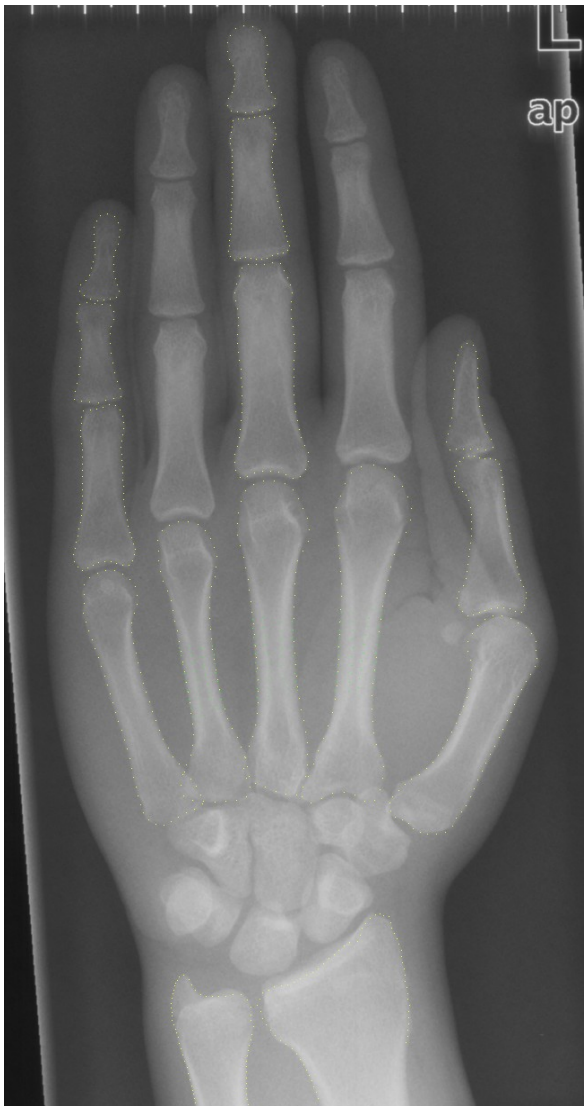


Figure A.2.4: X-ray out of BoneXpert's intended age range, incorrect analyzed by BoneXpert.
Image number m408-7. Finished bone maturation. Computed too young by BoneXpert (autBA 16.01y, manBA 19.0y, New manBA 18.67y).
In the subsequent BoneXpert version this error was eliminated and the X-ray is now rejected as invalid.

Figure A.2.5 to Figure A.2.21:

Correlation of average and difference of autBA and manBA in boys and girls, respectively.

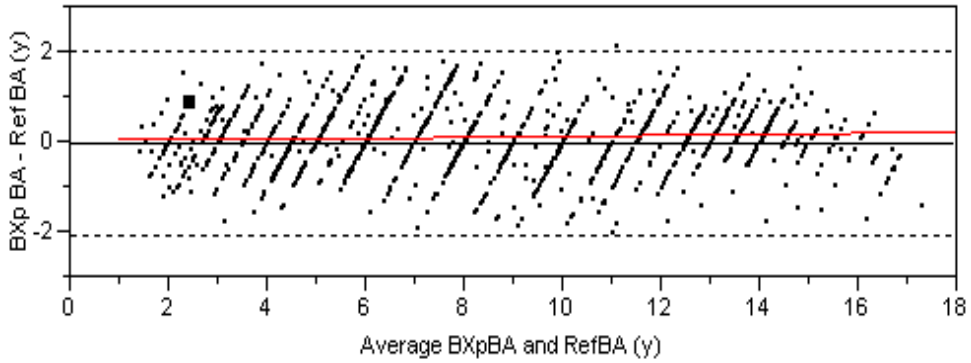


Figure A.2.5: GHD, male patients.
Bivariate Fit of BXpBA - RefBA (y) By Average BXpBA and RefBA (y)
Linear Fit: BXpBA - RefBA (y) = 0.05 + 0.01 Average BXpBA and RefBA (y)

Summary of Fit:

RSquare	0.003
RSquare Adj	0.002
Root Mean Square Error	0.699
Mean of Response	0.125
Observations (or Sum Wgts)	1157

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	1.58	1.58	3.24
Error	1155	564.19	0.49	Prob > F
C. Total	1156	565.77		0.07

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.05	0.05	0.98	0.33

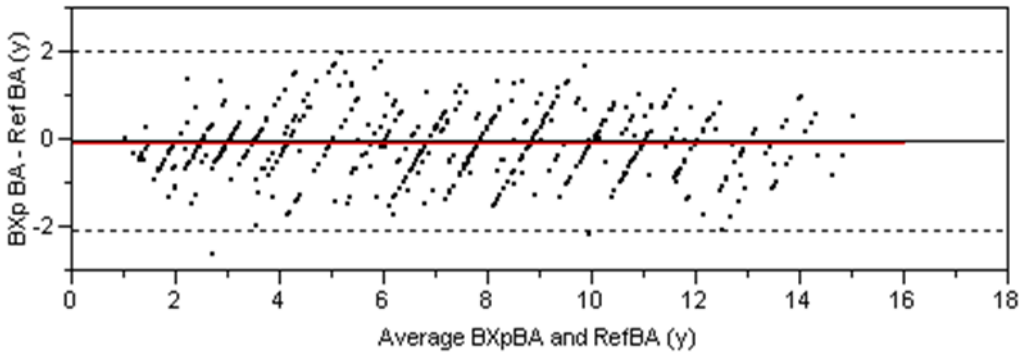


Figure A.2.6: GHD, female patients.
Bivariate Fit of BXpBA - RefBA (y) By Average BXpBA and RefBA (y)
Linear Fit: BXpBA - RefBA (y) = -0.1 + 0.00 Average BXpBA and RefBA (y)

Summary of Fit

RSquare	0.000
RSquare Adj	-0.002
Root Mean Square Error	0.728
Mean of Response	-0.080
Observations (or Sum Wgts)	497

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	0.043	0.043	0.082
Error	495	262.697	0.531	Prob > F
C. Total	496	262.740		0.775

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-0.10	0.076	-1.32	0.187
Average BXpBA and RefBA (y)	0.003	0.009	0.29	0.775

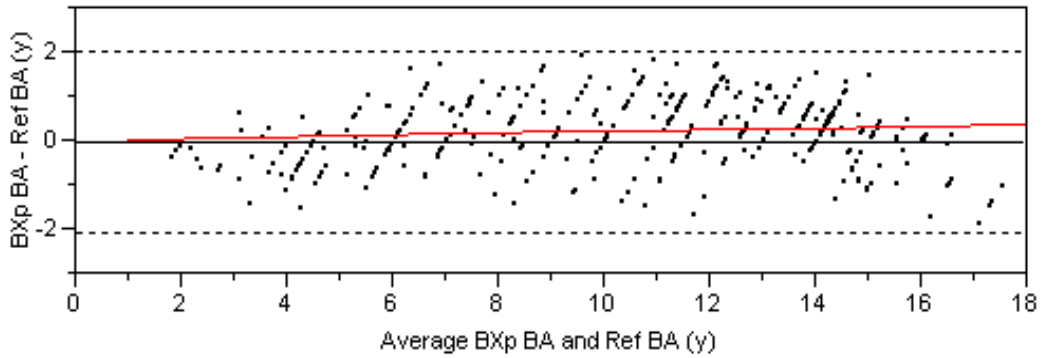


Figure A.2.7: secGHD, male patients.
Bivariate Fit of BXpBA - RefBA (y) By Average BXpBA and RefBA (y)
Linear Fit: BXpBA - RefBA (y) = -0.02 + 0.02 Average BXpBA and RefBA (y)

Summary of Fit

RSquare	0.014
RSquare Adj	0.011
Root Mean Square Error	0.734
Mean of Response	0.225
Observations (or Sum Wgts)	379

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	2.874	2.874	5.33
Error	377	203.289	0.539	Prob > F
C. Total	378	206.163		0.022

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-0.015	0.111	-0.14	0.892
Average BXpBA and RefBA (y)	0.023	0.010	2.31	0.022

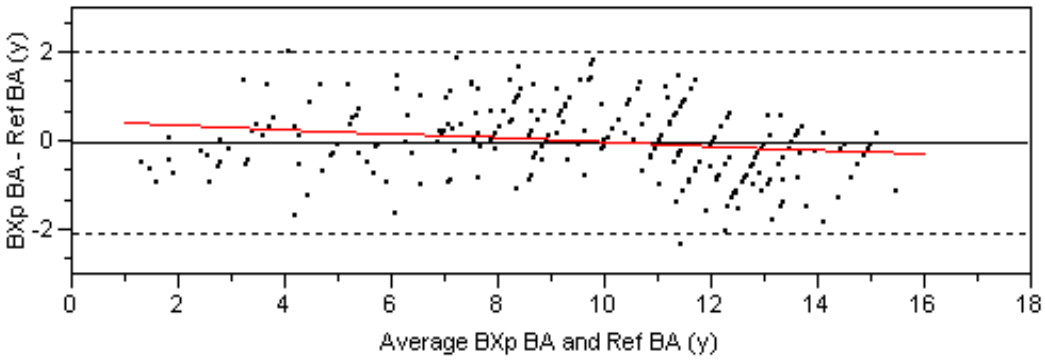


Figure A.2.8: secGHD, female patients
Bivariate Fit of BXpBA - RefBA (y) By Average BXpBA and RefBA (y)
Linear Fit: BXpBA - RefBA (y) = 0.46 - 0.05 Average BXpBA and RefBA (y)

Summary of Fit

RSquare	0.038
RSquare Adj	0.035
Root Mean Square Error	0.809
Mean of Response	0.027
Observations (or Sum Wgts)	260

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	6.730	6.730	10.296
Error	258	168.651	0.654	Prob > F
C. Total	259	175.381		0.002

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.457	0.143	3.20	0.002
Average BXpBA and RefBA (y)	-0.046	0.014	-3.21	0.002

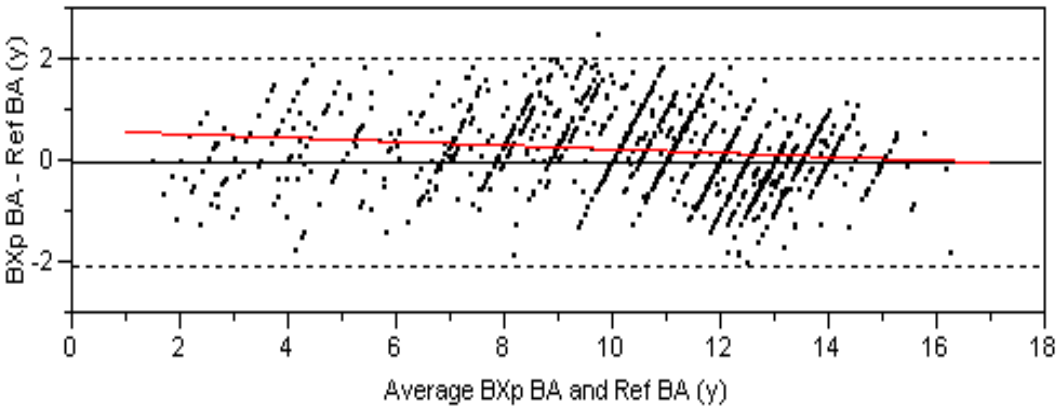


Figure A.2.9: UTS patients
Bivariate Fit of BXpBA - RefBA (y) By Average BXpBA and RefBA (y)
Linear Fit: BXpBA - RefBA (y) = 0.62 - 0.04 Average BXpBA and RefBA (y)

Summary of Fit

RSquare	0.025
RSquare Adj	0.024
Root Mean Square Error	0.715
Mean of Response	0.222
Observations (or Sum Wgts)	1134

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	14.638	14.638	28.604
Error	1132	579.305	0.512	Prob > F
C. Total	1133	593.943		<.0001

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.616	0.077	8.03	<.0001
Average BXpBA and RefBA (y)	-0.038	0.007	-5.35	<.0001

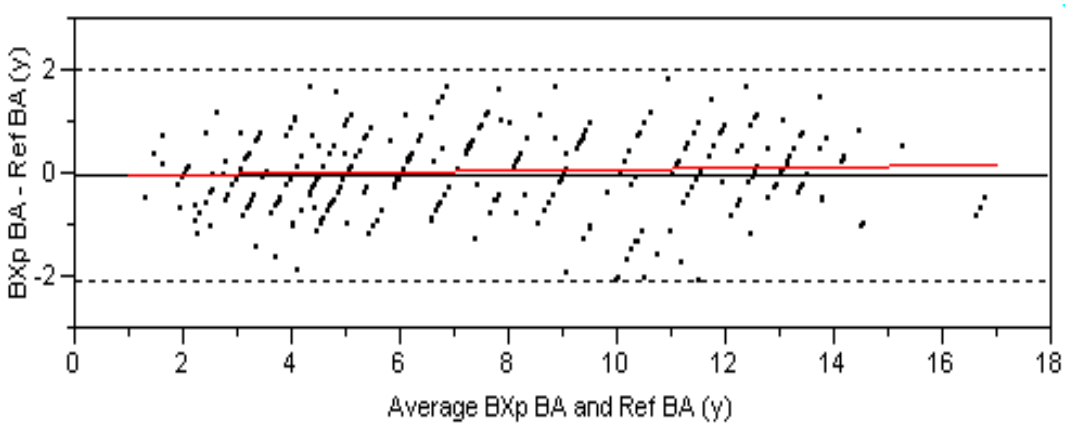


Figure A.2.10: SGA, male patients
Bivariate Fit of BXpBA - RefBA (y) By Average BXpBA and RefBA (y)
Linear Fit: BXpBA - RefBA (y) = -0.06 + 0.01 Average BXpBA and RefBA (y)

Summary of Fit

RSquare	0.004
RSquare Adj	0.001
Root Mean Square Error	0.757
Mean of Response	0.040
Observations (or Sum Wgts)	306

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	0.665	0.665	1.161
Error	304	174.183	0.573	Prob > F
C. Total	305	174.848		0.282

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-0.056	0.010	-0.56	0.573
Average BXpBA and RefBA (y)	0.013	0.011	1.08	0.282

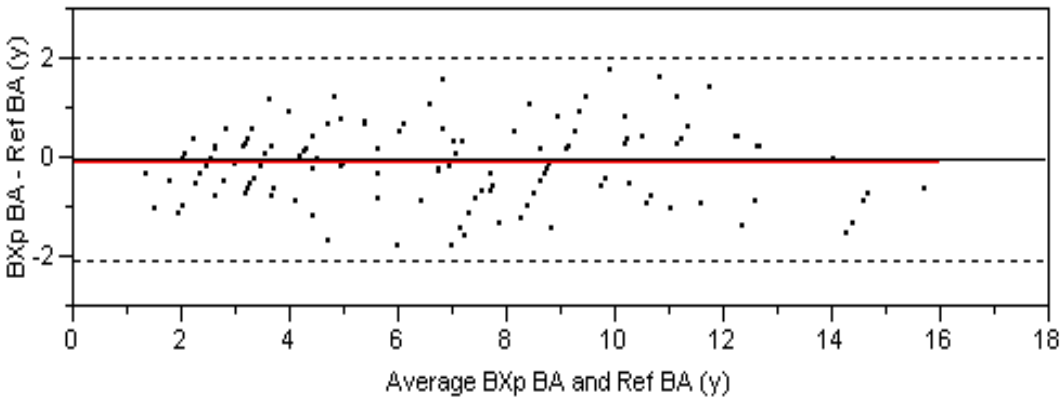


Figure A.2.11: SGA, female patients
Bivariate Fit of BXpBA - RefBA (y) By Average BXpBA and RefBA (y)
Linear Fit: BXpBA - RefBA (y) = -0.06 - 0.00 Average BXpBA and RefBA (y)

Summary of Fit

RSquare	0.000
RSquare Adj	-0.008
Root Mean Square Error	0.763
Mean of Response	-0.077
Observations (or Sum Wgts)	130

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	0.005	0.005	0.009
Error	128	74.496	0.582	Prob > F
C. Total	129	74.502		0.925

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-0.065	0.147	-0.44	0.661
Average BXpBA and RefBA (y)	-0.002	0.019	-0.09	0.925

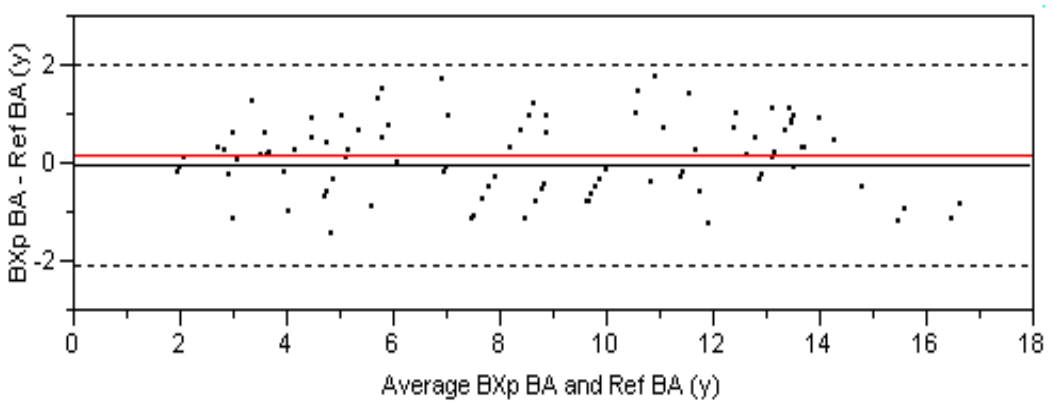


Figure A.2.12: SRS, male patients
Bivariate Fit of BXpBA - RefBA (y) By Average BXpBA and RefBA (y)
Linear Fit: BXpBA - RefBA (y) = 0.19 - 0.00 Average BXpBA and RefBA (y)

Summary of Fit

RSquare	0.000
RSquare Adj	-0.011
Root Mean Square Error	0.769
Mean of Response	0.186
Observations (or Sum Wgts)	95

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	0.001	0.001	0.002
Error	93	54.946	0.591	Prob > F
C. Total	94	54.9473		0.962

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.1944	0.188	1.03	0.305
Average BXpBA and RefBA (y)	-0.001	0.020	-0.05	0.962

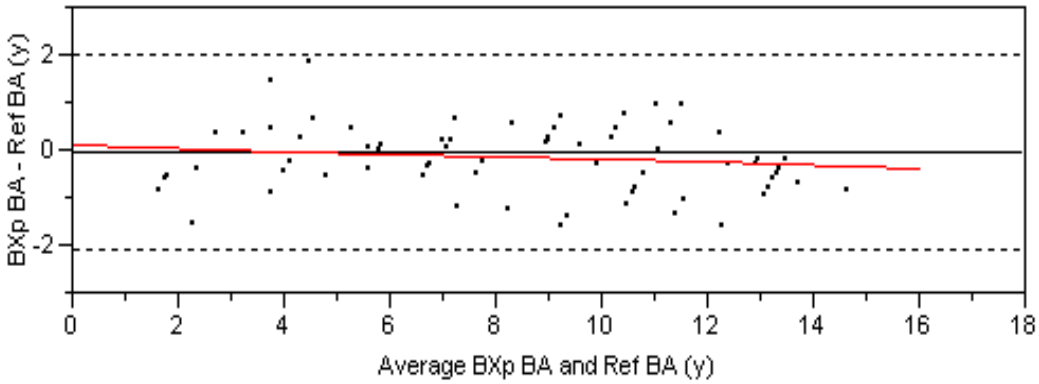


Figure A.2.13: SRS, female patients
Bivariate Fit of BXpBA - RefBA (y) By Average BXpBA and RefBA (y)
Linear Fit: BXpBA - RefBA (y) = 0.12 - 0.03 Average BXpBA and RefBA (y)

Summary of Fit

RSquare	0.024
RSquare Adj	0.010
Root Mean Square Error	0.707
Mean of Response	-0.127
Observations (or Sum Wgts)	70

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	0.843	0.843	1.685
Error	68	34.006	0.500	Prob > F
C. Total	69	34.849		0.199

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.117	0.206	0.57	0.571
Average BXpBA and RefBA (y)	-0.030	0.023	-1.30	0.199

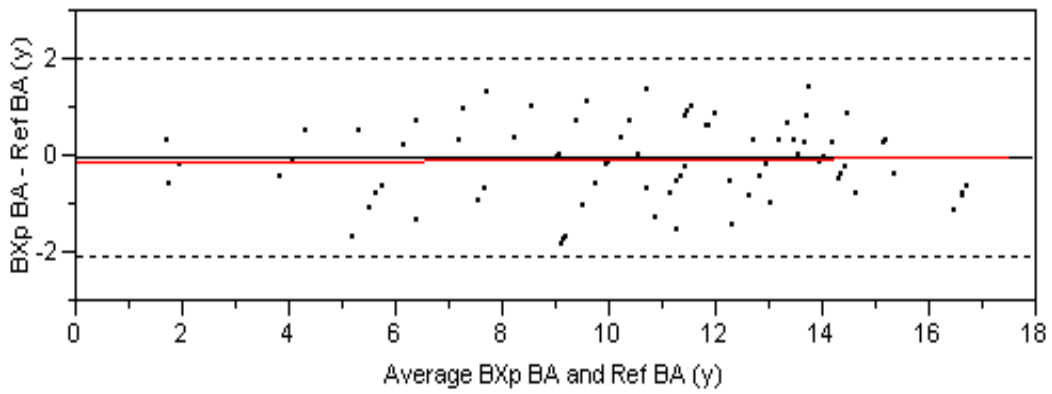


Figure A.2.14 : Noonan, male patients
Bivariate Fit of BXpBA - RefBA (y) By Average BXpBA and RefBA (y)
Linear Fit: BXpBA - RefBA (y) = -0.15 + 0.01 Average BXpBA and RefBA (y)

Summary of Fit

RSquare	0.001
RSquare Adj	-0.012
Root Mean Square Error	0.809
Mean of Response	-0.077
Observations (or Sum Wgts)	79

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	0.045	0.045	0.069
Error	77	50.412	0.655	Prob > F
C. Total	78	50.458		0.794

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-0.146	0.279	-0.52	0.602
Average BXpBA and RefBA (y)	0.007	0.025	0.26	0.794

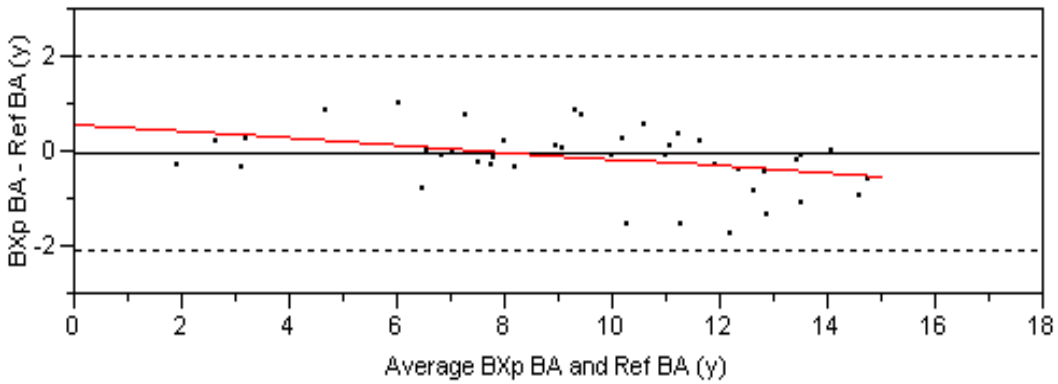


Figure A.2.15: Noonan, female patients
Bivariate Fit of BXpBA - RefBA (y) By Average BXpBA and RefBA (y)
Linear Fit: BXpBA - RefBA (y) = 0.60 - 0.07 Average BXpBA and RefBA (y)

Summary of Fit

RSquare	0.150
RSquare Adj	0.128
Root Mean Square Error	0.617
Mean of Response	-0.109
Observations (or Sum Wgts)	41

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	2.612	2.612	6.853
Error	39	14.8655	0.381	Prob > F
C. Total	40	17.477		0.013

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.5995	0.287	2.08	0.043
Average BXpBA and RefBA (y)	-0.075	0.029	-2.62	0.013

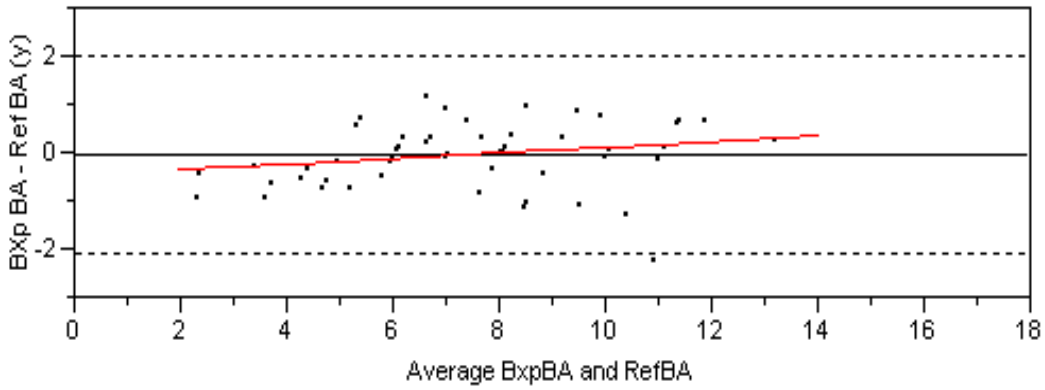


Figure A.2.16 : ISS, male patients
Bivariate Fit of BXpBA - RefBA (y) By Average BxpBA and RefBA
Linear Fit: BXpBA - RefBA (y) = -0.46 + 0.06 Average BxpBA and RefBA

Summary of Fit

RSquare	0.054
RSquare Adj	0.037
Root Mean Square Error	0.645
Mean of Response	-0.026
Observations (or Sum Wgts)	56

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	1.291	1.291	3.105
Error	54	22.451	0.416	Prob > F
C. Total	55	23.741		0.084

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-0.461	0.261	-1.76	0.083
Average BxpBA and RefBA	0.059	0.034	1.76	0.084

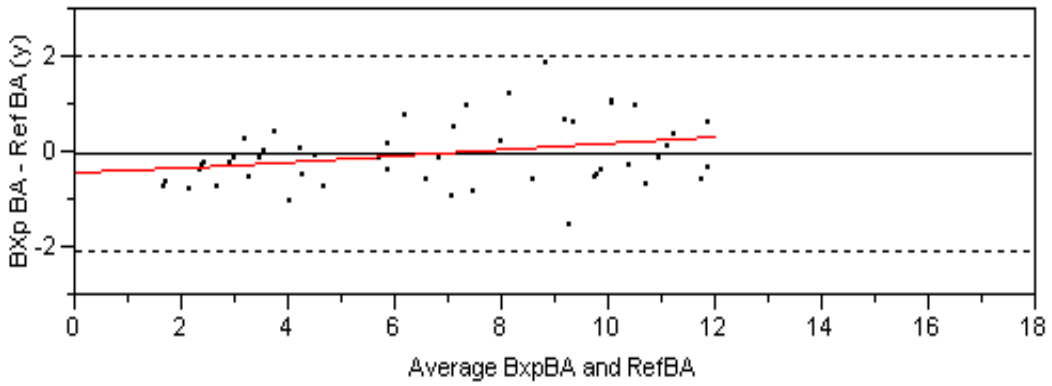


Figure A.2.17 : ISS, female patients
Bivariate Fit of BXpBA - RefBA (y) By Average BxpBA and RefBA
Linear Fit: BXpBA - RefBA (y) = -0.44 + 0.06 Average BxpBA and RefBA

Summary of Fit

RSquare	0.095
RSquare Adj	0.078
Root Mean Square Error	0.639
Mean of Response	-0.026
Observations (or Sum Wgts)	53

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	2.200	2.200	5.381
Error	51	20.852	0.40886	Prob > F
C. Total	52	23.052		0.024

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-0.439	0.199	-2.21	0.032
Average BxpBA and RefBA	0.063	0.027	2.32	0.024

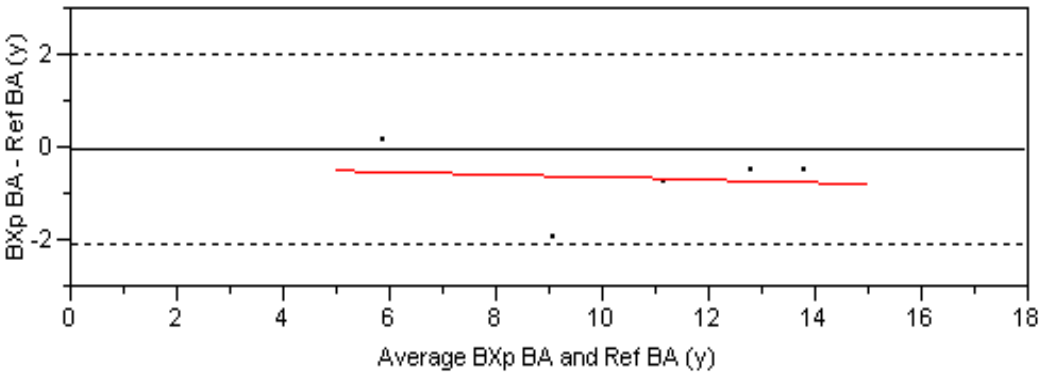


Figure A.2.18 : SHOX, male patients
Bivariate Fit of BXpBA - RefBA (y) By Average BXpBA and RefBA (y)
Linear Fit: BXpBA - RefBA (y) = -0.34 - 0.03 Average BXpBA and RefBA (y)

Summary of Fit

RSquare	0.014
RSquare Adj	-0.315
Root Mean Square Error	0.879
Mean of Response	-0.644
Observations (or Sum Wgts)	5

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	0.0324	0.032	0.042
Error	3	2.320	0.773	Prob > F
C. Total	4	2.352		0.851

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-0.345	1.516	-0.23	0.835
Average BXpBA and RefBA (y)	-0.028	0.139	-0.20	0.851

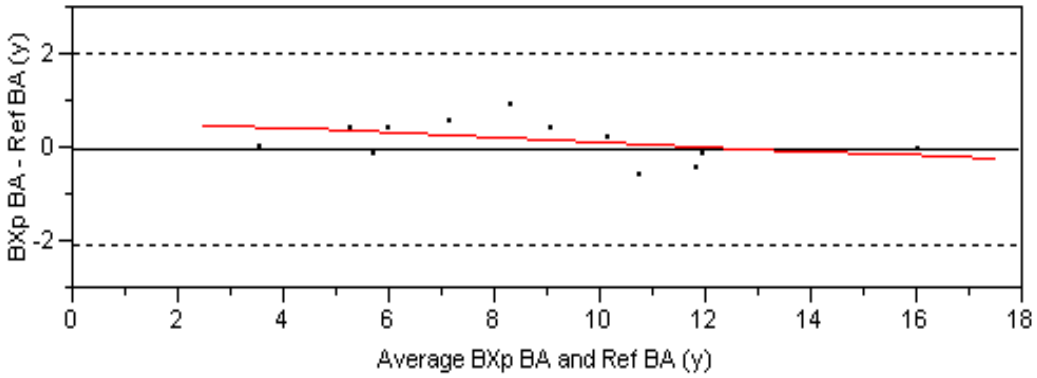


Figure A.2.19 : SHOX, female patients
Bivariate Fit of BXpBA - RefBA (y) By Average BXpBA and RefBA (y)
Linear Fit: BXpBA - RefBA (y) = 0.62 - 0.05 Average BXpBA and RefBA (y)

Summary of Fit

RSquare	0.160
RSquare Adj	0.075
Root Mean Square Error	0.417
Mean of Response	0.189
Observations (or Sum Wgts)	12

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	0.330	0.330	1.898
Error	10	1.740	0.174	Prob > F
C. Total	11	2.071		0.198

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.621	0.336	1.85	0.094
Average BXpBA and RefBA (y)	-0.049	0.036	-1.38	0.198

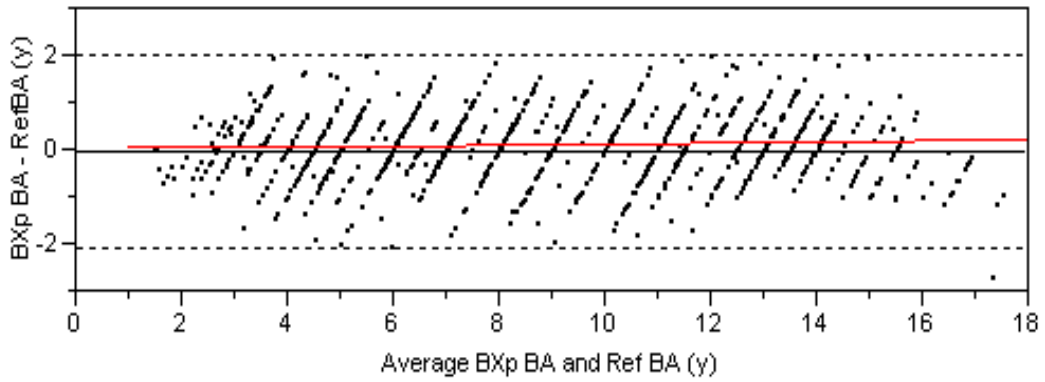


Figure A.2.20 : other, male patients
Bivariate Fit of BXpBA - RefBA (y) By Average BXpBA and RefBA (y)
Linear Fit: BXpBA - RefBA (y) = 0.05 + 0.01 Average BXpBA and RefBA (y)

Summary of Fit

RSquare	0.002
RSquare Adj	0.001
Root Mean Square Error	0.722
Mean of Response	0.128
Observations (or Sum Wgts)	877

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	1.095	1.095	2.102
Error	875	455.859	0.521	Prob > F
C. Total	876	456.953		0.148

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.048	0.060	0.80	0.425
Average BXpBA and RefBA (y)	0.009	0.006	1.45	0.148

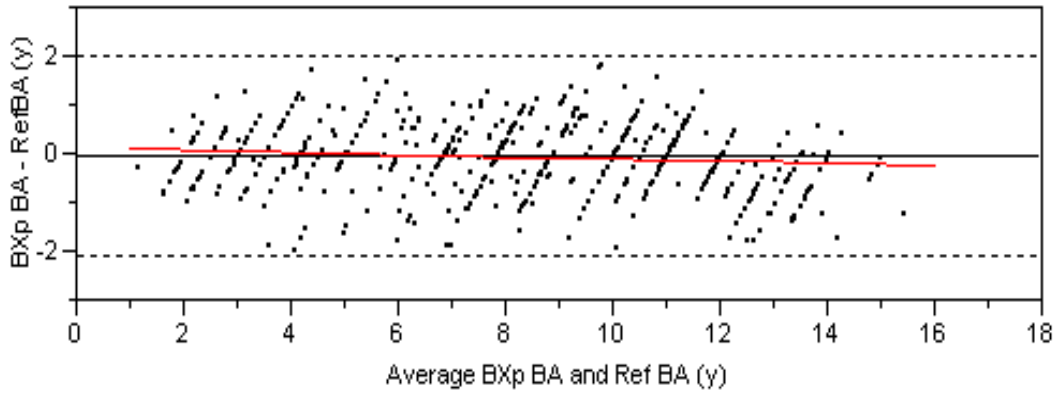


Figure A.2.21 : other, female patients
Bivariate Fit of BXpBA - RefBA (y) By Average BXpBA and RefBA (y)
Linear Fit: BXpBA - RefBA (y) = 0.14 - 0.02 Average BXpBA and RefBA (y)

Summary of Fit

RSquare	0.012
RSquare Adj	0.010
Root Mean Square Error	0.710
Mean of Response	-0.050
Observations (or Sum Wgts)	517

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	3.250	3.250	6.44
Error	515	259.703	0.504	Prob > F
C. Total	516	262.953		0.011

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.144	0.082	1.75	0.081
Average BXpBA and RefBA (y)	-0.023	0.009	-2.54	0.01

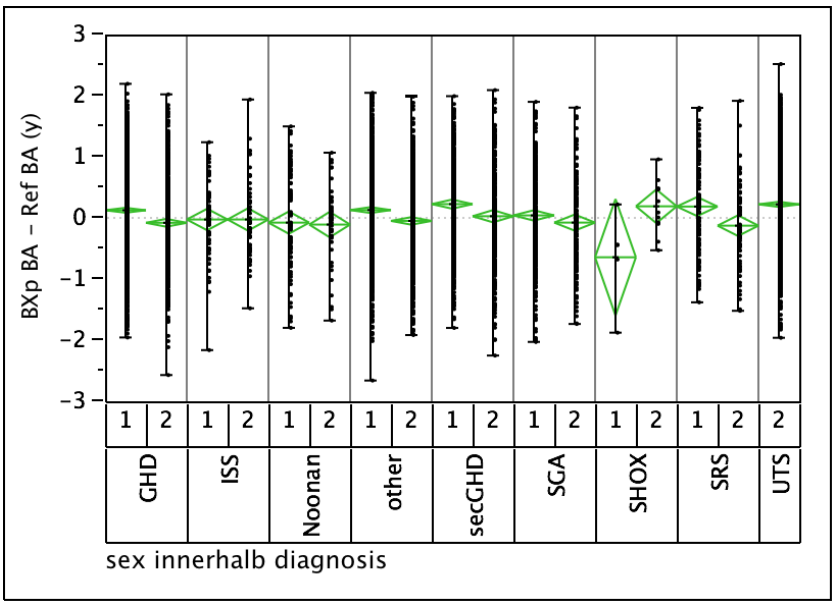


Figure A.2.22a: Analysis of Variance. Difference between autBA and manBA plotted vs. diagnoses and sexes.
The diagrams show that all the none of the mean differences significantly differ from the average mean. The diamonds portray the mean (middle horizontal line) and 95% CI (vertical diamond height). The deviation in male SHOX-patients is caused by small number of subjects.
1 = male, 2 = female

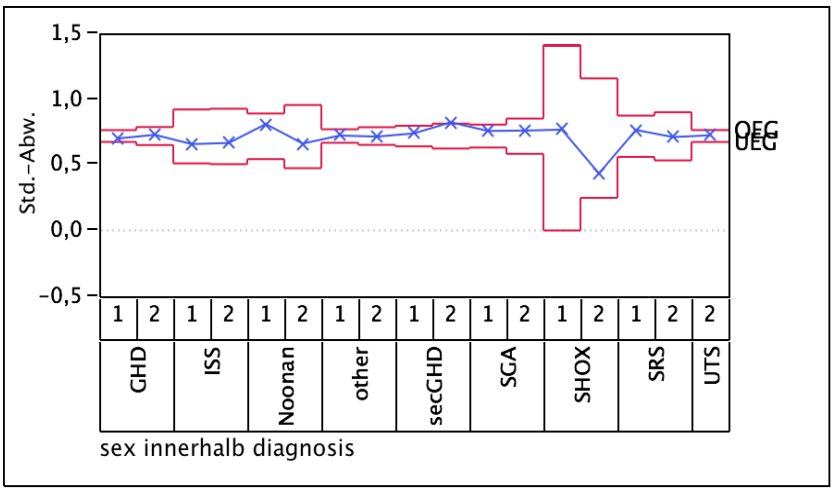
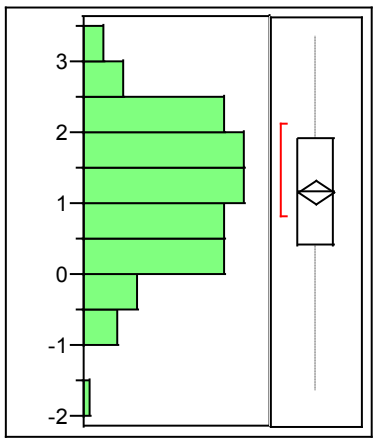


Figure A.2.22 b: Analysis of Variance. Standard deviations plotted vs. diagnoses and sexes.
The red limit markers in the variability diagram show that all the none of the mean SD values significantly differ from the average SD.
1 = male, 2 = female

Figure A.2.23: **Distributions CASAS BA – manBA**



Quantiles

100.0%	maximum	3.320
99.5%		3.320
97.5%		2.921
90.0%		2.290
75.0%	quartile	1.930
50.0%	median	1.160
25.0%	quartile	0.410
10.0%		-0.145
2.5%		-0.666
0.5%		-1.620
0.0%	minimum	-1.620

Moments

Mean	1.1649254
Std Dev	0.9515075
Std ErrorMean	0.0821977
upper 95% Mean	1.3275093
lower 95% Mean	1.0023414
N	150

Figure A.2.24: Bivariate Fit of CASAS BA (“CBA”) - manBA By CA.
X-axis: CBA – manBA, Y-axis: Chronological age (CA, “Alter”)

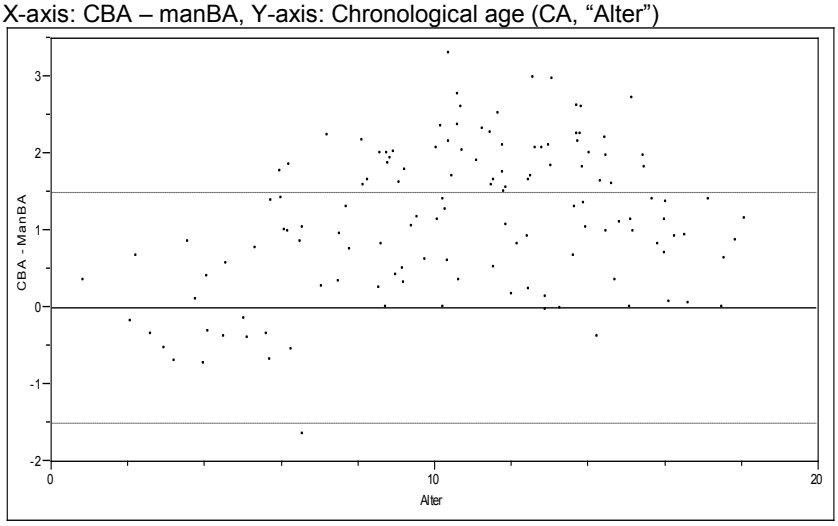
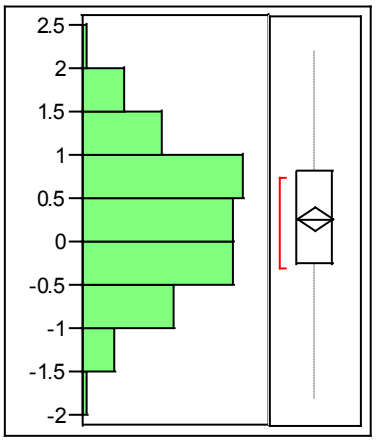


Figure A.2.25: Distribution BXpBA – manBA



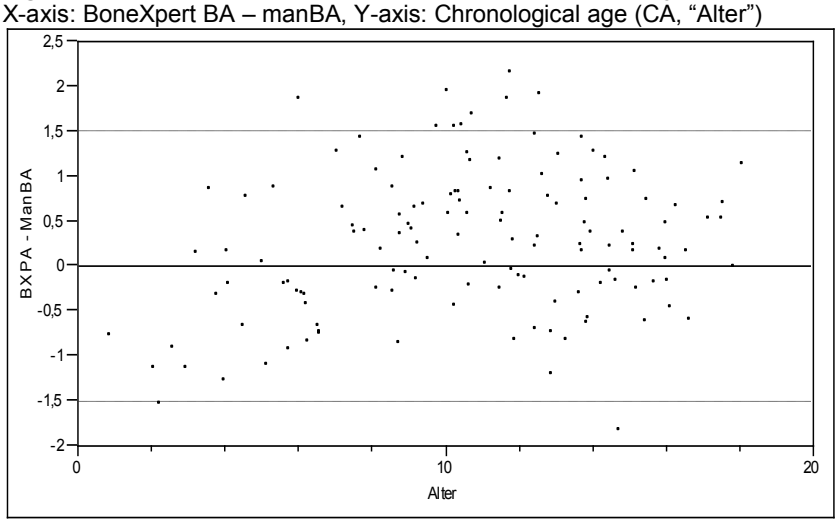
Quantiles

100.0%	maximum	2.180
99.5%		2.180
97.5%		1.918
90.0%		1.295
75.0%	quartile	0.818
50.0%	median	0.255
25.0%	quartile	-0.253
10.0%		-0.770
2.5%		-1.218
0.5%		-1.800
0.0%	minimum	-1.800

Moments

Mean	0.2791045
Std Dev	0.7974692
Std ErrorMean	0.0688908
upper 95% Mean	0.4153679
lower 95% Mean	0.1428411
N	150

Figure A.2.26: Bivariate Fit of BXpBA ("BXPA") - manBA by CA.
X-axis: BoneXpert BA – manBA, Y-axis: Chronological age (CA, "Alter")



B. Reference List

1. Abu-Amero S, Monk D, Frost J, Preece M, Stainer P, Moore GE. The genetic aetiology of Silver-Russell syndrome. J Med Genet 2008;45:193-9.

2. Acheson RM. The Oxford method of assessing skeletal maturity. Clin Orthop 1957;10:19-39.

3. Acheson RM, Vicinus JH, Fowler G. Studies in the reliability of assessing skeletal maturity from x-rays: Part III. Greulich-Pyle and Tanner-Whitehouse method contrasted. Hum Biol 1966;38(3):204-18.

4. Andersen E. Comparison of Tanner-Whitehouse and Greulich-Pyle methods in a large scale Danish Survey. Am J Phys Anthropol 1971;35:373-6.

5. Arends NJT, Boonstra VH, Mulder PGH, Odink RJH, Stokvis-Brantsma WH, Rongen-Westerlaken C et al. GH treatment and its effect on bone mineral density, bone maturation and growth in short children born small for gestational age: 3-year results of a randomized, controlled GH trial. Clin Endocrinol (Oxf) 2003;59(6):779-87.

6. Argente J, Gracia R, Ibanez L, Oliver A, Borrajo E, Vela A et al. Improvement in Growth after Two Years of Growth Hormone Therapy in Very Young Children Born Small for Gestational Age and without Spontaneous Catch-Up Growth: Results of a Multicenter, Controlled, Randomized, Open Clinical Trial. J Clin Endocrinol Metab 2007;92(8):3095-101.

7. Berst MJ, Dolan L, Bogdanowicz MM, Stevens MA, Chow S, Brandser EA. Effect of Knowledge of Chronologic Age on the Variability of Pediatric Bone Age Determined Using the Greulich and Pyle Standards. Am J Roentgenol 2001;176(2):507-10.

8. Bierich JR. Aetiology and Pathogenesis of Growth Hormone Deficiency. Acta Paediatr 1992;79(s370):155-63.

9. Binder G, Ranke M.B., Martin DD. Auxology Is a Valuable Instrument for the Clinical Diagnosis of SHOX Haploinsufficiency in School-Age Children with Unexplained Short Stature. J Clin Endocrinol Metab 2003;88(10):4891-6.

10. Binder G, Seidel AK, Martin DD, Schweizer R, Schwarze CP, Wollmann HA et al. The Endocrine Phenotype in Silver-Russell Syndrome Is Defined by the Underlying Epigenetic Alteration. J Clin Endocrinol Metab 2008;93(4):1402-7.

11. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1(8476):307-10.

12. Bull R, Edwards P, Kemp P, Fry S, Hughes I. Bone age assessment: A large scale comparison of the Greulich and Pyle, and Tanner and Whitehouse (TW2) methods. Arch Dis Child 1999;31:172-3.

13. Cole A, Webb L, Cole T. Bone age estimation: A comparison of methods. Br J Radiol 1988;61:683-6.

14. de Ridder MAJ, Stijnen T, Hokken-Koelega ACS. Prediction Model for Adult Height of Small for Gestational Age Children at the Start of Growth Hormone Treatment. J Clin Endocrinol Metab 2008;93(2):477-83.

15. Greulich W, Pyle S. Radiographic atlas of the skeletal development of the hand and wrist. 2nd ed. Stanford, CA: Stanford University Press; 1959.

16. Johnson GF, Dorst JP, Kuhn JP, Roche AF, Davila GH. Reliability of skeletal age assessments. Am J Roentgenol Radium Ther Nucl Med 1973;118(2):320-7.

17. King D, Steventon D, O'Sullivan M, Cook A, Hornsby V, Jefferson I et al. Reproducibility of bone ages when performed by radiology registrars: an audit of Tanner and Whitehouse II versus Greulich and Pyle methods. The Br J Radiol 1994;67:848-51.

18. Kosho T et al. Skeletal Features and Growth Patterns in 14 Patients with Haploinsufficiency of SHOX: Implications for the Development of Turner Syndrome. J Clin Endocrinol Metab 1999;84(12):4613-21.

19. Loder RT, Estle DT, Morrison K, Eggleston D, Fish DN, Greenfield ML et al. Applicability of the Greulich and Pyle skeletal age standards to black and white children of today. Arch Pediatr Adolesc Med 1993;147(12):1329-33.

20. Marchini A, Rappold G, Schneider KU. SHOX at a glance: from gene to protein. Arch Physiol Biochem 2007;113(3):116-23.

21. Martin DD, Deusch D, Schweizer R, Thodberg HH, Ranke MB. Validation of Automatic Greulich-Pyle Bone Age on GHD, UTS, SGA and Silver-Russell Syndrome Children. Published as abstract for the 47th Annual Meeting of the European Society for Paediatric Endocrinology (ESPE),2007.Horm Res 2007 2007(68(suppl.1)):69.

22. Martin DD, Deusch D, Schweizer R, Binder G, Thodberg H, Ranke M. Clinical application of automated Greulich-Pyle bone age determination in children with short stature. Pediatric Radiology 2009;39(6):598-607.

23. Martin DD, Neuhof J, Jenni OG, Ranke MB, Thodberg HH. Automatic determination of left and right hand bone age in the first Zurich longitudinal study. Horm Res Paediatr. 2010 Jul;74(1):50-5.

24. Noonan JA. Hypertelorism with Turner phenotype. A new syndrome with associated congenital heart disease. Am J Dis Child 1968;116(4):373-80.

25. Ranke M.B., Saenger P. Turner´s Syndrome. Lancet 2001;358:309-14.

26. Ranke M.B., Stubbe P, Majewski F, Bierich JR. Spontaneous Growth in Turner's Syndrome. Acta Paediatr. 1988;77(s343):22-30.

27. Ranke MB, Lindberg A, Cowell C, Wikland KA, Reiter E, Wilton P et al. Prediction of Response to Growth Hormone Treatment in Short Children born small for gestational Age: Analysis of Data from KIGS (Pharmacia International Growth Database). Journal Clin Endocrinol Metab 2003;88(1):125-31.

28. Ranke MB. Towards a Consensus on the Definition of Idiopathic Short Stature. Horm Res 1996;45(Suppl. 2):64-6.

29. Ranke MB, Lindberg A, Bakker B, Wilton P, Albertsson-Wikland K, Cowell CT, Price DA, and Reiter EO. The mathematical model for total pubertal growth in idiopathic growth hormone (GH) deficiency suggests a moderate role of GH dose. J Clin Endocrinol Metab 88, 4748-4753. 2003.

30. Ranke MB, Lindberg A, Chatelain P, Wilton P, Cutfield W, Albertsson-Wikland KA and Price DA. Derivation and validation of a mathematical model for predicting the response to exogenous recombinant human growth hormone (GH) in prepubertal children with idiopathic GH deficiency. J Clin Endocrinol Metab 84(4), 1174-1183. 1999.

31. Ranke MB, Partsch CJ, Lindberg A, Dorr HG, Bettendorf M, Hauffa BP et al. Adult height after GH therapy in 188 Ullrich-Turner syndrome patients: results of the German IGLU Follow-up Study 2001. Eur J Endocrinol 2002;147(5):625-33.

32. Ranke MB, Lindberg A, Chatelain P, Wilton P, Cutfield W, Albertsson-Wikland KA and Price DA. Prediction of long-term response to recombinant human growth hormone in Turner syndrome: development and validation of mathematical models. J Clin Endocrinol Metab 85, 4212-4218. 2000.

33. Rao E, Weiss B, Fukami M, RumpAndreas, Niesler B, Mertz A et al. Pseudoautosomal deletions encompassing a novel homeobox gene cause

growth failure in idiopathic short stature and Turner syndrome. Nat Genet 1997;16(1):54-63.

34. Roche AF, Davila G. The reliability of assessments of the maturity of individual hand-wrist bones. Hum Biol 1976;48(3):585-97.

35. Roche AF, Davila G, Pasternack B, Walton M. Some factors influencing the replicability of assessments of skeletal maturity (Greulich-Pyle). Am J Roentgenol 1970;109(2):299-306.

36. Roche AF, Rohmann C, French N, Davila G. Effect of training on replicability of assessments of skeletal maturity (Greulich-Pyle). Am J Roentgenol 1970;108(3):511-5.

37. Russell A. A syndrome of intra-uterine dwarfism recognizable at birth with cranio-facial dysostosis, disproportionately short arms, and other anomalies (5 examples). Proc R Soc Med 1954;47(12):1040-4.

38. Saenger P, Czernichow P, Hughes I, Reiter EO. Small for Gestational Age: Short Stature and Beyond. Endocr Rev 2007;28(2):219-51.

39. Sato K, Ashizawa K, Anzo M, Otsuki F, Kaneko S, Tanaka T et al. Setting up an automated system for evaluation of bone age. Endocr J 1999;46:S97-100.

40. Schweizer R, Ranke MB, Binder G, Herdach F, Zapadlo M, Grauer ML et al. Experience with Growth Hormone Therapy in Turner Syndrome in a Single Centre: Low Total Height Gain, No Further Gains after Puberty Onset and Unchanged Body Proportions. Horm Res 2000;53(5):228-38.

41. Silver HK, Kiyasu W, George J, Deamer WC. Syndrome of congenital hemihypertrophy, shortness of stature, and elevated urinary gonadotropins. Pediatrics 1953;12(4):368-76.

42. Tanner JM. Assessment of skeletal maturity and prediction of adult height (TW2 method). Academic Press London; 1975.

43. Tanner JM, Gibbons R. A computerized image analysis system for estimating Tanner-Whitehouse 2 Bone Age. Hor Res 1994;42:282-7.

44. Tanner JM, Gibbons R. Automatic Bone Age Measurement Using Computerized Image Analysis. J Pediatr Endocrinol 1994;7(2):141-5.

45. Tanner JM, Healy M, Goldstein H. Assessment of skeletal maturity and prediction of adult height (TW3 Method), 3rd edn. London: WB Saunders. Harcourt Publishers Ltd; 2001.

46. Tanner JM, Landt K, Cameron N, Carter B, Patel J. 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 1983;58:767-76.

47. Tanner J, Gibbons R, Bock R. Advantages of the computer-aided image analysis system for estimating TW skeletal maturity: increased reliability and a continuous scale. Pediatr Res 1993;33(S35).

48. Thodberg HH, Kreiborg S, Juul A, Damgaard-Petersen K. The BoneXpert method for automated determination of skeletal maturity. IEEE Trans Med Imaging 2009;28(1):52-66.

49. Turner HH. A syndrome of infantilism, congenital webbed neck, and cubitus valgus. Endocrinology 1938;23:566.

50. Ullrich O. Über typische Kombinationsbilder multipler Abartungen. Z Kinderheilk 1930;49:271-6.

51. van Rijn R, Lequin M, Thodberg H. Automatic determination of Greulich and Pyle bone age in healthy Dutch children. Pediatric Radiology 2009;39(6):591-7.

52. Vignolo M, Milani S, DiBattista E, Naselli A, Mostert M, Aicardi G. Modified Greulich-Pyle, Tanner-Whitehouse, and Roche-Wainer-Thissen (knee) methods for skeletal age assessment in a group of Italian children and adolescents. Eur J Pediatr 1990;149(5):314-7.

53. Wilson DM. Regular Monitoring of Bone Age Is Not Useful in Children Treated With Growth Hormone. Pediatrics 1999;104(4):1036.

54. Wollmann HA, Kirchner T, Enders H, Preece MA, Ranke MB. Growth and symptoms in Silver-Russell syndrome: Review on the basis of 386 patients. Eur J Pediatr 1995;154(12):958-68.

55. Zachmann M, Frasier S, McLaughlin J, Hurley L, Nessi P. Importance and Accuracy of Bone Age Ratings in a Computerized Growth Evaluation System. Horm Res 1983;18:160-7.

Deutsche Zusammefassung

Zusammenhang: Die Knochenalterbestimmung anhand von Röntgenbildern der Hand spielt eine wichtige Rolle in der Diagnostik und Therapie von kleinwüchsigen Kindern. Die manuelle Ermittlung birgt hierbei einige Probleme. So gibt es beispielsweise hohe inter- und intraindividuelle Differenzen in der Befundung. Um diese zu verringern wurden zunächst halbautomatische Bestimmungsmethoden entwickelt (z.B. CASAS, computer-assisted skeletal age score), die aufgrund ihrer zeitaufwendigen Durchführung allerdings keinen Einzug in die Praxis fanden. Ein vollautomatisches Computerprogramm, wie BoneXpert es darstellt, könnte dem Ruf nach einer zuverlässigen Knochenalterbestimmungsmethode gerecht werden. BoneXpert erstellt völlig automatische Analysen von Röntgenbildern der Hand und bestimmt ein entsprechendes Greulich-Pyle (GP) Knochenalter. Das Programm mustert einzelne Knochen von Bildern aus, wenn diese eine schlechte Bildqualität aufweisen, wenn der Knochenumriss durch das Programm nicht klar definiert werden kann oder es Auffälligkeiten im äußeren Erscheinungsbild des Knochens gibt. Das komplette Röntgenbild wird zurückgewiesen, sollten weniger als 8 der 13 RUS-Knochen analysiert werden können.

Zielsetzung: Das Ziel der Studie ist es, das BoneXpert-Programm zur automatischen Knochenalterbestimmung bei Kindern mit diversen Kleinwuchsdiagnosen zu validieren.
Zunächst untersuchten wir hierzu die allgemeine Präzision von BoneXpert. Wir analysierten zum einen die Ergebnisse der Bilder, welche BoneXpert ermittelte, zum anderen jene, welche durch das Programm ausgemustert wurden. Außerdem untersuchten wir BoneXpert’s Genauigkeit außerhalb seines gültigen Altersbereiches und die Kompatibilität im Gebrauch von digitalen und ausgedruckten Röntgenbildern.

Das Kernstück der Studie lag auf dem Vergleich der Ergebnisse von manueller und automatischer Knochenalterbefundung. Hierzu stellten wir die Ergebnisse von BoneXpert den Knochenalterangaben früherer Befundungen der Radiologen mittels Atlasmethode gegenüber. In einem kleineren Teil der Datenmenge lag uns ein zusätzliches, semiautomatisch ermitteltes, Knochenalter vor. So konnten wir zusätzlich die Übereinstimmung der automatisch ermittelten mit denen der halbautomatisch ermittelten Werte prüfen.
Des Weiteren analysierten wir die Validität von BoneXpert im klinischen Gebrauch. Hierzu untersuchten wir das Voranschreiten der Knochenreife nach dem ersten Jahr Wachstumshormontherapie und stellten auch hier die automatischen Berechnungen den manuell ermittelten Befunden gegenüber.

Methoden: Es wurden 5865 Röntgenbilder der linken Hand von 1152 Patienten einbezogen, welche ein manuell befundetes GP-Alter von 0,5-19 Jahren aufwiesen. Die Bilder stammen aus den Jahren 1973-2005 und wurden im Rahmen der Kleinwuchsdiagnostik und -behandlung in der Universitätskinderklinik Tübingen aufgenommen. Die mit den Patienten assoziierten Diagnosen sind: GHD (29% der Bilder), secGHD (11%), UTS (20%), SGA (8%), SRS (3%), Noonan (2%), ISS (2%), SHOX (< 1%), u.a. Erkrankungen, die sich unter anderem durch Kleinwuchs äußern (25%), wie Prader Willi Syndrom, Hirntumoren und Anormalitäten der Hypophyse (wie Craniopharyngiom oder Panhypopituitarismus), Hyperparathyroidismus, Morbus Crohn, Hypochondrodysplasie, 18p-Syndrome, Pierre-Robin-Syndrome. Diese Bilder wurden im klinischen Alltag von hauptsächlich drei verschiedenen Ärzten befundet.

Ergebnisse: BoneXpert konnte erfolgreich 5668 Bilder von 1009 Patienten analysieren. 198 Bilder (3%) konnten aus dem ursprünglichen Datensatz nicht ausgewertet werden. Diese waren entweder außerhalb des gültigen Altersbereichs (n=125), hatten eine schlechte Qualität (n=51) oder Knochenanomalien (n=3), waren vergrößerte oder verkleinerte Ausdrucke von digitalen Bildern (n=6) oder Bilder, die mit ihrem manuell befundeten

Knochenalter zwar innerhalb des Gültigkeitsbereichs lagen, aber dennoch extrem kleine Hände aufwiesen. Wurden letztere um 20% vergrößert, war BoneXpert in der Lage, ein gültiges Knochenalter zu errechnen.

Die Differenz der Ergebnisse, welche die Analyse von digitalen und entsprechenden Filmabzügen mittels BoneXpert ergab, war nicht signifikant. Der Korrelationskoeffizient zwischen Knochenalterberechnung anhand DICOM (Digital Imaging and Communications in Medicine; Datenformat) und entsprechenden Ausdrucken betrug 0.999; $P < 0.0001$.

Der Hauptteil der Studie bestand darin, die computergenerierten Ergebnisse mit den manuell ermittelten Knochenalterangaben unserer Radiologen zu vergleichen. Das BoneXpert-Knochenalter wich in 1,6% der Bilder (n=92) um mehr als 2,05 Jahre (2 Standardabweichungen, SD) vom originalen, manuellen Befund ab. Diese Ausreißer wurden jeweils von 3 Untersuchern erneut, ohne Wissen von chronologischem Alter, der ursprünglichen manuellen Befundung oder dem Ergebnis der automatischen Analyse, befundet. Beim Vergleich dieser neuen manuellen Befundung (New manBA) mit der automatischen lagen nur noch 3 der 92 Ausreißer außerhalb der Zeitspanne von 2,05 Jahren.

Die Genauigkeit zwischen manueller und automatischer Befundung weist eine Standardabweichung von 0,73 Jahren auf [0,66;0,84] 95% CI, sowie einer mittlere Abweichung von 0,09 Jahren. Die Varianz R^2 beträgt 0,967 für Jungen bzw. 0,957 für Mädchen. Die Übereinstimmung zwischen BoneXpert Ergebnissen und manueller Befundung wurde in einem Bland-Altman plot aufgetragen. Hierbei wurde die Abweichung der beiden Bestimmungsmethoden gegenüber dem Durchschnitt aus beiden aufgetragen. Dies führte zu einer Differenz der Methoden von 0,16 y SD 1,0 für Jungen und 0,32 y SD 0,97 für Mädchen. Die Übereinstimmung zeigte keinen signifikanten Unterschied zwischen den einzelnen Diagnosen.

Außerdem konnten wir die Ergebnisse von BoneXpert mit denen des halbautomatischen Computerprogramms CASAS vergleichen: Bei einem Teil der Bilder der UTS-Patientinnen wurde das Knochenalter zu einem früheren Zeitpunkt bereits mittels CASAS bestimmt. In 150 Bildern von Turner-Syndrom-Mädchen liegen uns somit sowohl manuelles, halbautomatisch bestimmtes

(CASAS) als auch computergeneriertes (BoneXpert) Knochenalter vor. Während des Knochenalter zwischen CASAS und ursprünglichem manuellen Knochenalter in 55 Fällen um mehr als 1,5 Jahre abweicht, sind es im Falle von BoneXpert lediglich 9 Bilder. Verglichen mit der manuellen Knochenalterermittlung nach GP (manBA) ist die Abweichung der GP-Ergebnisse von BoneXpert (autBA) geringer als die der Ergebnisse mittels CASAS (CBA), während die Standardabweichung der Differenzen (Genauigkeit) und die Verzerrung (bias) eine vernachlässigbare Abweichung aufweisen. Die mittlere Abweichung beträgt zwischen manBA und autBA 0,28 Jahre vs. 1,17 Jahre zwischen manBA und CBA, die Genauigkeit (SD der Differenz) jeweils 0,80 Jahre [95% CI 0.69; 0.88], bias 0,04 Jahre/Jahr bzw. 0,95 Jahre [95% CI 0.85; 1.08], bias 0,1 Jahre/Jahr. Im direkten Vergleich zwischen CASAS und BoneXpert ergibt sich eine mittlere Abweichung von 0.87 Jahren (n.s.), die Genauigkeit beträgt 0.84 Jahre [95% CI 0.75; 0.94].

Ein weiteres Ziel der Studie war, die Eignung des computergenerierten Knochenalters in der klinischen Anwendung zu validieren.
Hierzu verglichen wir das Voranschreiten des manuellen bzw. computergenerierten Knochenalters unter Wachstumshormontherapie. Wir wählten die Röntgenbilder aller SGA-Kinder mit kontinuierlicher Wachstumshormonbehandlung aus, bei denen uns Aufnahmen zum Zeitpunkt vor und eines Jahres nach der ersten Dosisgabe vorlagen. Nach der Auswertung dieser Stichprobe zeigte sich mittels BoneXpert ein Fortschreiten des Knochenalters um durchschnittlich 1,4 Jahre. Zum einen stützt dies Aussagen anderer Studien, die ein Fortschreiten des Knochenalters innerhalb des ersten Jahres der Hormonbehandlung beschreiben, zum anderen zeigt es eventuelle Schwächen der manuellen Bestimmung auf. Denn im Durchschnitt zeigte sich in der manuellen Befundung bei denselben Bildern ein Fortschreiten des Knochenalters um lediglich 1,04 Jahre. Vermutlich wurden die Befunder hierbei von äußerlichen Kriterien beeinflusst, denn der entsprechende Radiologie ist nicht über den Therapiestand informiert und wird sich bei der Einordnung des zweiten Bildes zum einen nach dem Vorbefund orientieren,

zum anderen mit eine Fortschritt von einem Jahr rechnen, da das Bild innerhalb dieser Zeitspanne angefertigt wurde.

Zusammenfassung: BoneXpert ist in der Lage, alle Röntgenbilder virtuell zu verarbeiten, welche die entsprechenden äußeren Vorgaben des Computerprogramms erfüllen. Es kann Fehler der inter- und intraindividuellen Inkonsistenz in der Bestimmung vermeiden, und erzielt gute Übereinstimmungen mit der manuellen Knochenalterermittlung. Dies spricht dafür, dass die Methode sowohl effizient als auch reliabel in der Knochenalterbestimmung kleinwüchsiger Kinder angewandt werden kann.

Veröffentlichungshinweis

Auszüge dieser Arbeit wurden veröffentlicht unter:

Pediatric Radiology 2009;39(6):598-607.

Clinical Application of automated Greulich-Pyle bone age determination in children with short stature

David D. Martin ◦ Dorothee Deusch ◦
Roland Schweizer ◦ Gerhard Binder ◦
Hans Henrik Thodberg ◦ Michael B. Ranke

Danksagungen

Sehr herzlich danke ich

Herrn Prof. Dr. Michael B. Ranke für die freundliche Überlassung des Themas und den Ansporn durch sein Lebenswerk, sowie die rasche Durchsicht dieser Arbeit,

Herrn Dr. David D. Martin für die stets hilfsbereite Betreuung bei Fragen und Problemen, sowie die unermüdliche Bereitschaft zur Diskussion. Durch seine Faszination von der Thematik und sein Engagement hat er einen sehr entscheidenden Beitrag zum Entstehen dieser Arbeit geleistet,

meinen Eltern, die mit ihrer Stetigkeit, Unterstützung, Lebensbejahung und ihrem Glauben ein wichtiges Fundament für mein Leben und damit auch für mein Studium und das Vollbringen dieser Arbeit gelegt haben,

meinem geliebten Mann und meinen Freunden, die teils bewusst, oftmals aber auch ohne ihr Wissen, eine Ermutigung und Motivation waren oder den manchmal notwendigen „Blick über den Tellerrand“ ermöglichten. Es ist ein Privileg, den Weg mit dir und euch zu teilen!

Und last but definitely not least Dem Einen, „der da ist und der war und der da kommt.“ (Die Bibel in Offenbarung Kapitel 1, Vers 8)